

Исследование операций

2

Модели и применения

ИЗДАТЕЛЬСТВО
«МИР»

Handbook of Operations Research Models and Applications

Edited by

Joseph J. Moder, Ph. D.
Department of Management Science
University of Miami

and

Salah E. Elmaghraby, Ph. D.

Graduate Program in Operations Research
North Carolina State University

Van Nostrand Reinhold Company
New York, Cincinnati, Atlanta, Dallas, San 1
London, Toronto, Melbourne 1978

519
И-855

Исследование операций

В 2-х томах

2

Модели и применения

Под редакцией
Дж. Моудера,
С. Элмаграби

Перевод с английского
под редакцией

чл.-корр. АН СССР И. М. Макарова,
д-ра техн. наук И. М. Бескровного

ИЗДАТЕЛЬСТВО «МИР»
МОСКВА 1981

ББК 22. 18
И 88
УДК 35.073.5

Авторы: Браун Р., Мэзон Р., Фламгольд Э. и др.

Исследование операций: В 2-х томах. Пер. с англ./Под ред.
И88 Дж. Моудера, С. Элмаграби. — М.: Мир, 1981. — 677 с., ил.

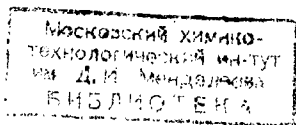
Второй том посвящен применению методов исследования операций для решения задач прогнозирования в промышленности, управления трудовыми ресурсами и запасами, повышения надежности и улучшения ремонта оборудования. Обсуждаются оптимальные способы размещения объектов, составления календарных планов, выбора наилучших проектных решений и разработки вычислительных и информационных систем. Рассматривается эффективность применения этих методов в таких сферах деятельности человека, как транспорт, здравоохранение, управление производственно-технологическими процессами.

Для специалистов в области исследования операций, теории управления, экономистов, инженеров-конструкторов и разработчиков АСУ.

И 30501-144
041(01)-81 144-81, ч.1

1502000000

ББК 22.18



511209

Редакция литературы по новой технике

© by Litton Educational Publishing, Inc., 1978

© Перевод на русский язык, с сокращениями,
«Мир», 1981.

Оглавление

Часть I

ПРИМЕНЕНИЕ МЕТОДОВ ИССЛЕДОВАНИЯ ОПЕРАЦИЙ К ОРГАНИЗАЦИОННО-ЭКОНОМИЧЕСКИМ СИСТЕМАМ

I-1. Прогнозирование, Р. Браун	5
Введение	5
1. Виды временных рядов	5
2. Исходные данные	9
3. Модели	11
4. Система сглаживающих весовых множителей.	16
5. Ошибки прогноза	20
6. Методы, применяемые для уточнения прогноза.	25
7. Изучение сферы сбыта	30
Литература	32
I-2. Управление трудовыми ресурсами, Р. Мезон, Э. Фламгольц	34
Введение	34
1. Комплектование штата организации	34
2. Повышение квалификации персонала	45
3. Распределение трудовых ресурсов	47
4. Использование трудовых ресурсов	54
5. Оценка эффективности работы персонала	57
6. Оплата работы и поощрение персонала	58
7. Модели сложных систем управления трудовыми ресурсами	63
Заключение	65
Литература	66
I-3. Агрегированное планирование производства, А. Хэкс.	71
Введение	71
1. Математические модели агрегированного планирования производства	73
2. Модели с квадратичной функцией затрат	81
3. Модели для определения оптимального размера партии про- дукции (модели с постоянными затратами)	84
4. Модели с произвольной функцией затрат	97
5. Интеграция производственного процесса	113
Литература	116
I-4. Управление запасами, Р. Браун.	121
Введение	121
1. Структура задачи управления запасами.	121

2. Правила принятия решений относительно выбора размера партии продукции для пополнения ее запасов	130
3. Правила принятия решений относительно уровня резервного запаса	144
4. Распределение ограниченных ресурсов	155
5. Архитектура систем управления запасами	157
6. Реализация систем управления запасами	169
Литература	172
4-5. Вычислительные и информационные системы, А. Мертен.	174
Введение	174
1. Вычислительные системы.	176
2. Системы обработки данных.	182
3. Информационные системы.	187
Заключение	189
Литература	189
4-6. Планировка и размещение объектов, Р. Френсис, Дж. Уайт.	192
Введение	192
1. Задачи размещения с непрерывным пространством решений	195
2. Задачи размещения — распределения с использованием непрерывного пространства решений	202
3. Задачи планировки с использованием непрерывного пространства решений	204
4. Задачи планировки с использованием дискретного пространства решений	209
5. Задачи размещения с использованием дискретного пространства решений	213
6. Задачи размещения — распределения с использованием дискретного пространства решений	215
Заключение	226
Литература	227
4-7. Календарное планирование и упорядочение работ, М. Сальвадор	232
Введение	232
1. Классификация задач календарного планирования	232
2. Некоторые общие критерии оценки графика работ	236
3. Детерминированные задачи календарного планирования	237
4. Стохастические задачи календарного планирования	251
5. Перспективы развития теории и методов календарного планирования и упорядочения работ	258
Литература	259
4-8. Выбор проекта. Планирование работ над проектом и руководство проектом, У. Саудер	264
Введение	264
1. Задачи выбора проекта	266
2. Модели выбора проекта	269
3. Сетевое планирование	281
4. Руководство проектом	299
5. Интегрированные системы планирования и руководства	305
Заключение	308
Литература	309

I-9. Теория надежности, Э. Шойер.	314
Общие положения	314
1. Структурные свойства систем	315
2. Надежность связанных систем	320
3. Время безотказной работы и соответствующие случайные величины	324
4. Некоторые оценки надежности	332
5. Интервальная надежность и эксплуатационная готовность	340
Литература	343
I-10. Техническое обслуживание оборудования, Э. Жардин.	344
Введение	344
1. Замена (обновление) оборудования	344
2. Контроль за состоянием оборудования	355
3. Профилактический текущий ремонт и восстановление оборудования	357
4. Организационная структура службы технического обслуживания	359
5. Перспективы развития технического обслуживания оборудования	361
Литература	362

Часть II

ПРИМЕНЕНИЕ МЕТОДОВ ИССЛЕДОВАНИЯ ОПЕРАЦИЙ К СОЦИАЛЬНЫМ И ПРОИЗВОДСТВЕННЫМ СИСТЕМАМ

II-1. Системы городской сферы обслуживания, И. Савае, Г. Грэмме	364
Введение	364
1. Экстренные службы	365
2. Охрана окружающей среды	379
3. Санитарные службы. Службы уличного движения	390
4. Другие внутригородские (местные) сферы обслуживания.	393
5. Обратная связь в системе управления городской сферой обслуживания	395
6. Проблемы применения методов исследования операций в городской сфере обслуживания	396
7. Новые направления в использовании методов исследования операций для решения проблем городской сферы обслуживания	401
Литература	404

II-2. Система здравоохранения, Р. Фиттер.	410
Введение	410
1. Краткая характеристика и проблемы системы здравоохранения	410
2. Анализ проблем системы здравоохранения с помощью методов исследования операций	414
3. Планирование национальных и региональных систем медицинского обслуживания	418
4. Перспективность использования методов исследования операций для решения проблем системы здравоохранения	420
5. Модель распределения ресурсов в системе здравоохранения	420

6. Пути дальнейшего усовершенствования системы медицин- ского обслуживания	423
Литература	423
II-3. Процессы обучения, А. Муд.	439
Введение	439
1. Анализ системы образования	439
Литература	454
2. Аннотированная библиография	455
II-4. Транспорт, Б. Нийл, Э. Кэтенис, Д. Хайндс.	466
Введение	466
1. Прогнозирование потребностей в перевозках людей и грузов	468
2. Анализ распределения возможных видов транспортных средств	476
3. Определение параметров задачи и описание различных ва- риантов решения	480
4. Оценка эффективности возможных вариантов решения	480
5. Разработка макетов транспортной сети и оценка возможных вариантов решения	487
6. Реализация выбранного варианта решения	488
7. Классический пример практического использования общей схемы исследования транспортной системы	489
8. Типичные проблемы, возникающие при исследовании общей схемы анализа транспортной системы	493
Заключение	496
Литература	497
II-5. Военные системы, Э. Квейд.	499
Введение	499
1. Исторические аспекты использования методов исследования операций в военных системах	499
2. Анализ военных небоевых операций	514
3. Анализ боевых операций	517
4. Различные типы анализа состояния вооружения	523
5. Методы анализа организации обороны	527
6. Публикации, посвященные использованию методов исследова- ния операций в военной области	530
Литература	532
II-6. Электроэнергетические системы, Л. Гаргер	537
Введение	537
1. Прогнозирование нагрузки	538
2. Планирование производственных процессов	548
3. Планирование увеличения производства электроэнергии	567
4. Планирование развития сети электропередачи	583
Литература	591
II-7. Производственные и технологические процессы, Г. Блау, Р. Климпел	600
Введение	600
1. Задачи распределения ресурсов при непрерывном производ- стве	602

2. Задачи составления смесей	605
3. Линейные модели технологических операций	608
4. Линейные модели ряда одновременно выполняемых техно- логических операций	610
5. Оптимизация технологических процессов	622
6. Оптимизация проектных решений методом геометрического программирования	630
7. Математические модели технологических операций	635
Литература	640

II-8. Организация досуга (туризм, спорт, развлечения), У. Сюот, Ч. Гиэринг. Т. Ва.	642
Введение	642
1. Применение методов исследования операций при организа- ции развлечений, путешествий и туризма	643
2. Применение методов исследования операций в области спорта	654
Заключение	662
Литература	662
Предметный указатель	669

Часть I

Применение методов исследования операций к организационно-экономическим системам

I-1. Прогнозирование

Р. Браун¹⁾

ВВЕДЕНИЕ

В данной главе дается краткий обзор методов многократного уточнения значений некоторого параметра системы, который используется в качестве выходной переменной модели. В отличие от случая, когда оценка основана только на имеющихся данных и потому проводится лишь один раз, ниже рассматриваются различные случаи оценок, пересматриваемых по мере поступления новой информации, с тем чтобы получать новые значения выходной переменной системы и корректировать выбор варианта решения. Пересмотр и уточнение оценок (прогнозов) могут проводиться либо через одинаковые интервалы времени (например, через неделю или месяц), либо через разные интервалы времени (например, в аналоговых системах, применяемых для контроля за пожароопасными объектами, используются специальные методы последовательной проверки наличия или отсутствия опасности).

1. ВИДЫ ВРЕМЕННЫХ РЯДОВ

Можно выделить два вида прогнозируемых характеристик системы, зависящих от времени: переменные состояния и переменные интенсивности [44]. *Переменная состояния* определяется периодически, и ее значение в течение небольшого интервала времени не зависит от времени, прошедшего с момента начала наблюдения. *Переменная интенсивности* также определяется периодически, но ее значение пропорционально времени, прошедшему с момента предыдущего наблюдения. Такие характеристики системы, как температура, скорость, число подписчиков на журнал или цена, являются примерами переменных состояния. В качестве примера переменной интенсивности можно привести количество выпавших осадков, количество проданных экземпляров или спрос. Если переменная состояния характеризует количество, то переменная интенсивности — скорость его изменения.

¹⁾ Robert G. Brown, Consultant in Materials Management Systems, Box 332, Norwich, Vermont.

Процессы прогнозирования переменных состояния и интенсивности отличаются друг от друга следующими особенностями:

— если измерения характеристик системы проводятся через разные интервалы времени, то величину интервала необходимо учитывать при оценке переменных интенсивности, в то время как при оценке переменных состояния эта величина не имеет значения;

— так как прогнозы обычно осуществляются для нескольких последовательных интервалов времени в пределах некоторого времени упреждения, по истечении которого становятся важными результаты реализации принятых решений, то правильный прогноз переменной состояния должен определять ее значение в конце времени упреждения, а прогноз переменной интенсивности должен представлять собой сумму прогнозов на протяжении времени упреждения;

— функция распределения во времени вероятностей ошибок прогноза для переменной состояния должна соответствовать функции распределения вероятностей ошибок в исходных данных, тогда как для переменной интенсивности закон распределения вероятностей ошибок прогноза во времени стремится к нормальному при любом законе распределения вероятностей ошибок в исходных данных, поскольку эти ошибки представляют собой сумму ошибок прогноза в отдельные интервалы времени.

1.1. Временные интервалы пересмотра и уточнения прогноза

Величина промежутков времени между измерениями входных переменных системы с целью проверки и уточнения ранее сделанных прогнозов о выходных переменных зависит главным образом от длительности времени упреждения и наибольшей частоты циклических изменений в системе, которые должна отражать модель. Поэтому временные интервалы пересмотров прогнозов могут изменяться в широких пределах и измеряются, например, как часами, так и годами. Однако эти интервалы должны быть достаточно велики, чтобы обеспечивалась вероятность осуществления ожидаемых изменений системы.

В течение времени упреждения прогноз чаще всего проверяется один-два раза, но иногда возникает необходимость увеличить количество проверок до десяти. Так, прогнозируемые сроки возобновления лесозаготовок на участке леса с периодом восстановления 80 лет могут уточняться один раз в десятилетие, а изменение потребления электроэнергии в данной местности в зависимости от погодных условий можно наблюдать каждый час.

Когда имеют место какие-то периодические процессы, как, например, при месячных изменениях температуры или изменениях объема корреспонденции, получаемой почтовым отделением в течение недели, частота наблюдений должна быть по крайней мере

вдвое больше частоты изучаемого процесса [20]. Очевидно, что наблюдения с интервалами в один месяц не будут способствовать правильной организации работы почты по понедельникам в отличие от ее работы по средам, а наблюдения с интервалами в сутки не дадут возможности обнаружить различие в интенсивностях работы первой и второй смен.

Если случайная ошибка при определении входных переменных велика по сравнению с измеряемой величиной, интервал уточнения прогноза для переменной интенсивности целесообразно увеличить, усредняя таким образом случайную ошибку. Однако для переменной состояния в аналогичном случае интервал уточнения прогноза лучше уменьшить, что позволяет для выделения полезного сигнала использовать соответствующие методы фильтрации [23].

1.2. Методы прогнозирования

Методы прогнозирования могут быть разделены на три большие группы: статистические (описательные), причинно-следственные и комбинированные. Для изучения исследуемого процесса следует задать закон изменения входных переменных X во времени. Выходные переменные системы могут быть описаны с помощью некоторой модели, значения коэффициентов которой определяются подбором. При этом различные наблюдения учитываются с различными весовыми множителями. По таким моделям, включающим описание предыстории системы, прогноз можно составить путем расчета состояния системы для некоторого будущего момента времени.

В разд. 4—5 настоящей главы рассматриваются различные вопросы, связанные с процессом прогнозирования: исходные данные, модели, методы сглаживания и ошибки прогноза. В разд. 6 показано, что если требуется точный прогноз, то дисперсии ошибок прогноза не всегда являются единственным или даже основным критерием выбора того или иного способа статистического описания временных рядов.

Если удастся построить модель окружающей среды, позволяющую выявить причины изменений в системе (вторая группа методов), то прогноз, полученный с помощью такой модели, *объясняет* будущее системы. Подобные методы охватывают широкий круг моделей от формальных моделей экономики [19], объясняющих, например, почему большое количество товаров покупается при распродаже [8], до очень субъективных моделей типа вербальной модели рынка, обсуждаемой в разд. 7. Статистический прогноз в лучшем случае позволяет делать вероятностные утверждения о возможных появлениях и моментах появления экстремумов в рас-

смаатриваемой временной зависимости переменных системы, а также о значениях переменных в экстремальных точках. Существуют, однако, некоторые очень мощные средства обнаружения экстремумов сразу после того, как они появились [9, 15]. С другой стороны, прогнозы на основе анализа причинно-следственных связей предназначены главным образом как раз для предсказания моментов появления экстремумов и значений переменных в экстремальных точках. Пример такого прогноза представляют кривые жизненного цикла любой продукции [3]. Спрос на любую продукцию имеет фазы начального роста, наивысшего уровня и спада. Во время последней фазы продукция исчезает с рынка. Хотя подобные модели точно описывают историю спроса на продукцию, тем не менее они не дают достаточно твердой основы для предсказания перехода из одной фазы спроса в другую.

Прогнозирование научно-технического прогресса [1] становится все более необходимым и полезным в целях предвидения основных факторов, оказывающих воздействие на окружающую среду. Однако прогнозирование в области экономики не привело пока к созданию эффективного прогноза, который позволил бы предсказать мировой дефицит какой-то продукции, обусловленный противоречиями между возможностями окружающей среды и требованиями экономики.

Если установлен сводный план выпуска определенной продукции, то его можно представить в виде простой суммы отдельных планов, определяющих, например, число и сроки подачи заявок на оборудование, запасные части и сырье (гл. I-3 и I-4). Однако между прогнозируемыми данными и некоторыми основными показателями может существовать и причинная связь. Например, колебание числа заявлений, подаваемых в разные годы в приемные комиссии колледжей, непосредственно связано с изменениями рождаемости, имевшей место соответственно 17 лет назад. В приведенном примере ясно видна причинная взаимосвязь двух показателей, и поэтому упреждающий показатель может быть использован для составления прогноза при правильно выбранном времени упреждения.

Между двумя любыми автокоррелированными временными рядами всегда существует статистическая корреляция [31]. Следовательно, существует серьезная опасность бессмысленного использования множественной регрессии в поисках «хорошего» коэффициента корреляции между прогнозируемыми (выходными) переменными и различными потенциально информативными входными переменными. Известная предыстория представляет собой только элемент позднего временного ряда, и, если искать достаточно долго, связь между переменными будет найдена. Однако, если с увеличением объема информации коэффициенты модели становятся равными нулю, то модель не пригодна для принятия правильно-

го решения, так как именно в критический момент основные показатели перестали быть таковыми.

В последующих разделах данной главы основное внимание уделяется статистическим методам прогнозирования, которые используются при уточнении принятых решений, распространяющихся на относительно короткие промежутки времени (обычно эти промежутки в десять раз больше интервалов между уточнениями предсказаний). Исследование операций включает методы отыскания подходящих моделей и методы уточнения прогнозов по новым наблюдениям. Следует отметить, что наилучшие результаты получаются при использовании комбинации статистических и причинно-следственных методов прогнозирования, что иллюстрируется в разд. 7 на примере исследования состояния рынка.

2. ИСХОДНЫЕ ДАННЫЕ

2.1. Выборочные наблюдения

Исходные данные обычно представляют собой результаты выборочных наблюдений либо переменной интенсивности, такой как спрос на продукцию, либо переменной состояния, такой как цена. Решения, которые должны приниматься в данный момент времени, скажутся в дальнейшем по прошествии некоторого промежутка времени — времени упреждения. Величина этого промежутка времени может быть прогнозируемой переменной.

2.2. Исключение выбросов

Результаты наблюдений регистрируются с большим количеством ошибок, которые возникают как при наблюдениях, так и при регистрации данных. Кроме того, изучаемый процесс может иметь стохастическую природу. Результаты наблюдений могут содержать и аномальные эффекты, например необычайно высокий спрос на продукцию во время забастовок или чрезмерно высокую температуру в помещении в случае перегрева термостата. Поэтому не каждую совокупность зарегистрированных по мере поступления реальных данных следует считать подходящим временным рядом, на основании которого можно составлять прогноз.

Перед тем как подобрать коэффициенты модели по исходным данным, из последних должны быть исключены выбросы, т. е. результаты наблюдений, которые не характеризуют прогнозируемый процесс. Простой и надежный способ такого исключения выбросов может быть проиллюстрирован при помощи рис. 1. Так как наибольший и наименьший результаты наблюдений могут представлять собой выбросы, эти два результата временно исклю-

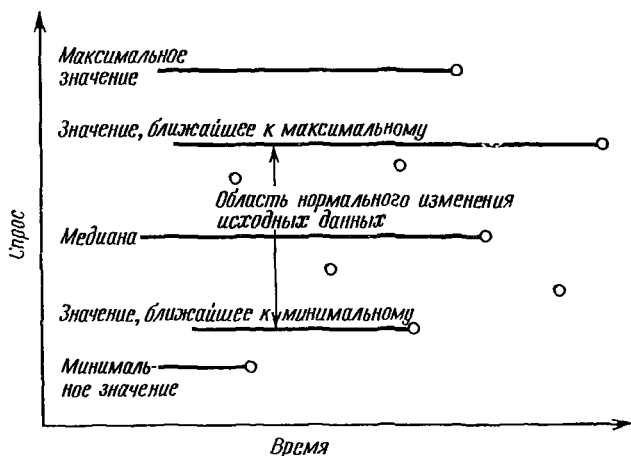


Рис. 1. Схема, иллюстрирующая способ исключения выбросов из исходных данных.

чаются из имеющейся выборки данных. Интервал между крайними оставшимися результатами наблюдений позволяет быстро оценить область нормального изменения исходных данных. Величина интервала между максимальными и следующими по величине (или минимальными и ближайшим к нему) значениями входной переменной может служить индикатором того, что экстремальные результаты являются представительными. Для отрезков временного ряда, содержащих от 20 до 100 наблюдений и превышающих 0,4 ряда, целесообразно проверять исходные данные более тщательно. Очевидно, этот простой тест не позволит исключить ни двух выбросов одного знака, если оба они достаточно велики, ни отклонений в данных, имеющих чисто гармоническую природу при значимом негармоническом периоде. Если разброс данных превышает, например, 4σ , то требуется более тщательный отбор входных переменных, с тем чтобы после подбора коэффициентов получить статистическую модель, правильно описывающую исследуемый процесс.

2.3. Преобразование исходных данных

Экономические данные часто подчиняются экспоненциальным зависимостям, и в таких случаях при построении моделей процесса удобно пользоваться логарифмическим масштабом. Другим видом преобразований является разбиение исходного ряда данных на составные части. Например, полный спрос на какую-либо продукцию трудно прогнозировать с помощью одной общей моде-

ли. Поэтому соответствующий временной ряд целесообразно разбить на отдельные составные ряды, каждый из которых легко описать. Так, хотя решения о пополнении запаса принимаются на основе полного спроса, тем не менее прогнозировать новые заказы, прибыли от продажи, требования, налагаемые сводным планом, правительственные и экспортные заказы, спрос, порождаемый специальными фондами, и т. д. удобнее, используя отдельные компоненты ряда данных.

2.4. Помехи, остатки и ошибки

Для описания стохастических элементов временных рядов и их прогноза используются три различных понятия — помехи, остатки и ошибки. Понятие *помехи* связано с собственной изменчивостью процесса и неопределенностью, вносимой при наблюдении за ним. Следовательно, помеха является составной частью используемых данных. Под *остатками* понимается разность между результатами наблюдений и соответствующими значениями, вычисленными с помощью прогнозирующей их модели. Таким образом, остатки связаны с прошлыми данными и моделью, которая использовалась для их оценок. *Ошибки* прогноза представляют собой разницу между прогнозом, сделанным в настоящее время, и тем, что будет наблюдаться позднее в момент времени, для которого составлен прогноз.

2.5. Условие стационарности процесса

Любой процесс, представленный рядом результатов наблюдений, можно по крайней мере схематически описать набором разностных уравнений (когда процесс дискретен) или дифференциальных уравнений (когда он непрерывен). Если в этих уравнениях коэффициенты не зависят от времени, процесс называется *стационарным*, если же зависят, то *нестационарным*. Эта зависимость может носить вероятностный или регулярный характер. Очевидно, что регулярное изменение коэффициентов во времени может быть описано дополнительными уравнениями.

Для достаточно больших интервалов времени большинство прогнозируемых рядов является нестационарным. Однако их все же можно считать квазистационарными, если прогноз составлять для какого-то одного момента времени.

3. МОДЕЛИ

Пусть в момент времени T задана последовательность результатов наблюдений X_t для некоторого множества моментов времени $t \leq T$. Прогнозирующая модель задает множество выходных переменных $\hat{X}_{T+t}$, которые могут быть выражены в векторной форме

(время $\tau > 0$). В общем виде выражение для модели записывается как $\hat{X}_{T+\tau} = a_T F(\tau)$, где вектор a_T представляет собой коэффициенты модели, получаемые по результатам наблюдений до момента T включительно, а матрица F — набор аппроксимирующих функций. Строки матрицы соответствуют элементам модели, а столбцы — моментам времени. В большинстве практических приложений F определяет время относительно момента самого последнего наблюдения, а значения коэффициентов a зависят от выбора начала отсчета времени.

3.1. Полиномиальное представление временных рядов

Весьма обширный класс моделей процесса составляют модели, которые могут быть представлены с помощью вещественных полиномов F_i : $F_1(t) = 1$, $F_2(t) = t$, $F_3(t) = 1/2t(t-1)$ и т. д. Применение различных методов прогнозирования позволяет получить различные модели, например, за счет использования разностей или суммы функций F_i [5]. Поскольку, однако, решение линейного дифференциального уравнения может быть представлено в виде вещественного полинома, то различные математические описания моделей означают для нас лишь возможность выбора удобной формы анализа, а не фактическое различие в математических свойствах таких моделей.

Довольно часто для анализа локальных изменений в наблюдаемых данных достаточно использовать полином первой степени (прямую линию). Заметим, что константа, т. е. полином нулевой степени, является частным случаем полинома первой степени, в котором коэффициент при линейном члене равен нулю. Подгонка прямой линии для конкретного ряда прогнозируемых данных производится путем варьирования величины второго коэффициента, отражающего темп изменения временного ряда за длительный период времени (вековой тренд). Если величина этого коэффициента мала, то его можно исключить из модели. Следует, однако, заметить, что величины коэффициентов, которые могут быть исключены из модели, в соответствии со статистическим критерием настолько малы, что их исключение не оказывает никакого влияния на краткосрочный прогноз. Затраты на осуществление статистического теста могут быть больше, чем затраты на использование такой малой величины в прогнозе, даже в случае, когда в ее использовании нет никакой необходимости. Если время упреждения более чем в 10 раз превышает величину интервала времени между уточнениями прогноза, исключение трендов из модели становится более важным.

Математические методы позволяют представить прогнозирующую модель в виде полинома любого порядка [7]. Однако, если

нет необходимости в использовании полиномов более высокого порядка (как, например, при отслеживании траектории баллистической ракеты), применение полиномов второго и более высоких порядков для описания прогнозируемых рядов оказывается излишним.

3.2. Сезонные модели

Уровень воды в море изменяется дважды в сутки за счет приливов; температура в области высоких широт ежегодно меняется от высокой до низкой и обратно; потребление электроэнергии ежедневно изменяется по более сложным зависимостям (исключая большие праздники и годовой цикл); спрос на запаасаемую продукцию может меняться по циклам, повторяющимся каждый торговый сезон. С точки зрения выработки правильного решения включение подобных периодических (сезонных) колебаний в общую модель может повысить эффективность прогноза и позволит предсказывать ожидаемые высокие и низкие значения прогнозируемых переменных.

Однако не все ряды, имеющие высокие и низкие значения переменных, обязательно цикличны в указанном смысле. Деловые или экономические «циклы» не повторяются с достаточной воспроизводимостью, позволяющей на практике, исходя из анализа прошлого, делать выводы о каких-то будущих подъемах и спадах.

Для описания сезонных моделей существуют два совершенно разных подхода, каждый из которых имеет множество вариантов.

Если дано N результатов наблюдений за какой-то период времени (например, $N = 12$ для годового цикла или $N = 7$ для недельного цикла), то можно записать N поправочных членов (положительных и отрицательных) или N коэффициентов (каждый из которых больше или меньше единицы), которые затем либо суммируются с результатами прогноза, либо умножаются на него. Значения этих поправочных коэффициентов можно определить по результатам наблюдений в соответствующие моменты времени в прошлых циклах. Например, в некоторых простейших методах такие значения, скажем для сентября месяца, определяются на основе данных для этого же месяца в прошлые годы. В более сложных методах [22] осуществляются усреднение и учет результатов наблюдений, которые предшествовали данному.

При анализе розничной продажи товаров широкого потребления сезонные коэффициенты часто используются в качестве кумулятивных характеристик, которые представляют процентное выражение сделанного к настоящему времени по отношению к тому, что должно быть сделано в течение полного периода.

В любом из названных подходов весь ряд поправочных членов или коэффициентов может быть назван сезонным «сечением». Во

многих случаях оказывается, что существует небольшое семейство сечений, и тогда прогнозы для отдельных временных рядов могут быть получены путем выбора наиболее подходящего представления этого семейства.

Сечения, бесспорно, играют важную роль в методике прогнозирования, так как их сущность легко объяснить, а эффективность их использования легко оценить на практике. Однако существенным недостатком применения сечений является связанная с ними нестабильность прогноза. Если случайная величина x имеет дисперсию σ^2 , то дисперсия обратной величины ($y = 1/x$) порядка σ^4 . В большинстве методов, использующих сезонные или циклические сечения, на некотором этапе вычислений проводится деление на случайную переменную (или ее оценку, имеющую отличную от нуля дисперсию). Если при этом знаменатель имеет тот же порядок величины, что и помеха, то относительно малые колебания исходных данных могут привести к большим изменениям прогноза. К сожалению, подобная нестабильность обычно проявляется как раз тогда, когда принимаемые решения могут иметь самые серьезные последствия (т. е. перед началом следующего цикла).

В работе [15] для уменьшения подобной нестабильности используется предположение, согласно которому дисперсия помехи является функцией только среднего по всем наблюдениям за цикл и не зависит от сезонных изменений. При этом чувствительность метода к колебаниям данных в точках наивысших и наинизших значений прогнозируемых переменных соответственно повышается и понижается.

Пример. Рассмотрим выраженный в процентах прогноз производства товаров широкого потребления. В начале каждого торгового сезона необходимо определить полный объем выпуска товаров и знать положение дел, если не с большей частью имеющихся товаров, то хотя бы с сырьем. Допустим, что 10%-ный уровень сезонной продажи достигается в среднем за первые три недели торговли, в течение которых должен быть заключен окончательный договор на поставки дорогостоящего сырья. Если какой-нибудь оптовый покупатель, как правило, поздно заказывающий товары, в текущем году размстит свой заказ раньше обычного, то учет такого заказа при прогнозировании может привести к значительному завышению оценки общих потребностей в данном сезоне. Так фактически и случилось в США, когда впервые за несколько лет рождественская торговля началась раньше, чем обычно.

Другой подход заключается в представлении циклических изменений прогнозируемой переменной рядами Фурье, т. е. $\hat{X}_t = \sum c_k \cos \omega kt + \sum s_k \sin \omega kt$. Здесь c_k и s_k — коэффициенты, полученные по исходным данным с помощью соответствующей регрессии, $\omega = 2\pi/f$ — основная частота, f — число наблюдений в одном сезонном цикле, t — время, за которое должен быть вычислен прогноз, $k = 1, 2, \dots, N$ и суммирование ведется по всем

частотам вплоть до частоты Найквиста (наивысшей частоты гармонического разложения дискретного ряда, которая определяется половиной интервала между наблюдениями).

Достоинство такой модели [7] состоит в том, что она обеспечивает стабильность прогноза даже в точках цикла с наименьшими значениями прогнозируемой переменной, так как коэффициенты вычисляются путем усреднения всего набора имеющихся данных, а не только результатов наблюдений в пределах одного цикла. На практике, однако, подобная модель оказывается сложной для пользователя. Поэтому были разработаны пакеты программ для ЭВМ [10], с помощью которых редактирование входных и интерпретация выходных данных осуществляются таким образом, что большинство пользователей не испытывает необходимости в понимании существа методов обработки процесса.

Если для составления прогноза используются только первые члены разложения Фурье, то сезонное сечение представляет собой простую синусоиду с некоторой амплитудой и фазой, зависящими от коэффициентов c_1 и s_1 . При увеличении числа членов разложения форма сечения может измениться. В соответствии с теоремой Фурье, если в дискретных рядах величина переменной повторяется через каждые f наблюдений, то такие ряды могут быть представлены точно f членами гармонического разложения.

Амплитуда j -й гармоники, включенной в модель, равна $A_j = \sqrt{c_j^2 + s_j^2}$ (квадрат амплитуды иногда называют *мощностью* данной гармоники). Пусть σ^2 — дисперсия остаточных разностей между исходными данными и результатами расчетов по модели, включающей гармоники вплоть до j -й. Тогда модель будет соответствовать предыстории, охватывающей от 3 до 6 предыдущих циклов, если удовлетворяется соотношение $A_j \geq 1/2\sigma_i$. Проверка на соответствие предыстории обычно проводится по критерию наименьших квадратов. Отметим, что при вычислении σ^2 необходимо учесть потерю степеней свободы некоторым числом переменных в модели. Это означает, что можно достичь минимальной дисперсии, используя модель с числом переменных, которое меньше f . Однако такая модель будет неверной, если мощность включенной в нее высшей гармоники не будет достаточно большой по сравнению с остаточной дисперсией.

3.3. Анализ временных рядов

Допустим, что ряды X_t приведены к рядам с нулевым средним $Y_t = X_t - \hat{X}_t$, причем в модели для переменной $\hat{X}_t$ предусматривается согласование полиномиальных трендов и циклических изменений. Автоковариация σ_t^2 представляет собой ожидаемое значение произведения $Y_t Y_{t-\tau}$, где τ — *время задержки*. Если

модель $\hat{X}_t$ отражает все периодические колебания в данных, то автоковариация в основном равна нулю для всех значений времени задержки. Большие значения автоковариации (положительные или отрицательные) указывают на то, что в рядах Y_t имеется информация, которая может быть внесена в модель. Математические методы отыскания адекватной модели описания временных рядов составляют специальную область статистики — спектральный анализ [17, 21].

3.4. Вероятностные модели

При представлении совокупности результатов наблюдений в виде временных рядов фактически используется предположение о том, что наблюдаемые величины принадлежат некоторому распределению, параметры которого и их изменение можно оценить во времени. По этим параметрам (как правило, по среднему значению и дисперсии, хотя иногда используется и более полное описание) можно построить одну из моделей вероятностного представления процесса.

Другим вероятностным представлением является модель в виде частотного распределения с параметрами p_j для относительной частоты наблюдений, попадающих в j -й интервал. При этом, если в течение принятого времени упреждения не ожидается изменения распределения, то решение принимается на основании имеющегося эмпирического частотного распределения.

4. СИСТЕМА СГЛАЖИВАЮЩИХ ВЕСОВЫХ МНОЖИТЕЛЕЙ

Практически во всех применяемых в настоящее время методах прогнозирования коэффициенты моделей (например, a , $[c_h, s_h]$ или p_j) сначала определяются путем подгонки модели к некоторым данным предыстории, а затем проверяются и уточняются по мере поступления новых данных. Выбор коэффициентов, как правило, осуществляется из условия минимизации суммы квадратов остаточных разностей между данными и результатами расчета по модели с учетом различных весовых множителей, приписываемых остаткам в различные моменты времени. Так, в рамках модели, записанной в общем виде, коэффициенты a выбираются на основе минимизации величины $S = \sum w_j (X_{T-j} - \hat{X}_{T-j})^2$, где суммирование проводится по всем j вплоть до самого последнего наблюдения в момент времени T . При оценке тех или иных достоинств различных систем весовых множителей w_j необходимо учитывать не только достигаемую при их использовании точность прогноза, но и степень сложности соответствующих вычислений.

4.1. Одинаковые весовые множители

Простейший ряд весовых множителей $w_j = 1$ для всех j из интервала $0 \leq j \leq T$ придает одинаковую значимость всем членам временного ряда. Даже если в дальнейшем ожидается, что коэффициенты модели будут уточняться и возможно применение некоторой другой системы весовых множителей, то первоначальные значения коэффициентов, как уже отмечалось, все же получают-ся путем простой подгонки модели по методу наименьших квадратов, при котором все члены ряда имеют одинаковую значимость.

Скользящее среднее представляет собой оценку по методу наименьших квадратов единственной константы для представления исходных данных с одинаковыми весовыми множителями. Этому случаю соответствует простая функция $F(t) = 1$ для всех t . При этом весовые множители $w_j = 1$ для N последних наблюдений в интервале $0 \leq j \leq N - 1$ и $w_k = 0$ для $k \geq N$. Некоторые коэффициенты модели могут определяться с помощью полиномов более высокого порядка путем подгонки модели к результатам N самых последних наблюдений в каждый момент поступления новых данных. Таким образом, в самом общем случае значения коэффициентов в любой заданный момент времени зависят от предыдущих значений коэффициентов, ошибки в прогнозе при использовании самого последнего наблюдения, а также от вида используемого полинома и, конечно, числа N результатов наблюдений.

4.2. Оптимальные весовые множители

Как показано в работе [29], для стационарных временных рядов существует система весовых множителей, позволяющая обеспечить минимальную ошибку прогноза. Эти множители определяются видом автоковариационной функции. В работе [30] предложен метод вычисления таких оптимальных весовых функций для постоянного уровня, тренда и сезонных коэффициентов. В каждом из этих случаев весовые множители экспоненциально уменьшаются по закону α^j , а различные значения α , получаемые для уровня, тренда и сезонных коэффициентов, определяются путем систематического исследования точности прогнозов, получаемых при различных комбинациях весовых функций.

Для достаточно широкого класса моделей предложен метод [5], основанный на сочетании вычисления автоковариационной функции с систематическим исследованием области, в которой при определенном периоде и временной задержке должны находиться оптимальные весовые множители. Данный метод эффективен при большом объеме обрабатываемых данных или если известны основные закономерности изучаемого процесса. Впервые

результаты этой работы были с большим успехом применены в химической промышленности при создании системы управления технологическим процессом, работающей в реальном масштабе времени.

4.3. Весовые множители с экспоненциальным затуханием

Во многих случаях целесообразно использовать последовательность $w_j =: \alpha^j$, $\alpha < 1$, придающую более высокий вес более поздней информации и позволяющую относительно просто оценивать значения коэффициентов даже достаточно сложных моделей, таких, в которых для описания сезонных циклов используются полиномы в сочетании с преобразованиями Фурье (подобное представление можно рассматривать как сложные полиномы). Указанные достоинства экспоненциального сглаживания [6], сделавшие его достаточно популярным, особенно важны в тех случаях, когда уточнение прогноза необходимо проводить многократно, а стоимость вычислений может оказаться довольно большой по сравнению с затратами на составление менее точного прогноза.

В работах [27, 28] для модели экспоненциально взвешенного скользящего среднего предложены способы, с помощью которых в те периоды времени, когда средняя ошибка прогноза близка к нулю (благодаря правильности модели и ее коэффициентов), скорость затухания α может быть увеличена, а в те периоды времени, когда средняя ошибка прогноза существенно отличается от нуля¹⁾ и существует опасность того, что модель может «забыть» старую информацию (в этом случае требуется уточнение прогноза), скорость затухания α может быть уменьшена.

Уточнение прогноза производится по принципу обратной связи — новые прогнозы корректируются на основе учета ошибок в предшествующих прогнозах. Если при выборе весовых множителей в процессе составления прогноза также используется обратная связь, то не только строгий анализ областей устойчивости данной системы, но и любой анализ вообще становится фактически невозможным. Для анализа эффективности какого-либо метода недостаточно привести примеры, подтверждающие его полезность. Необходимо также выявить области (если они существуют), в которых применение рассматриваемого метода невозможно или неэффективно.

Многие методы позволяют отыскать наилучшее значение скорости затухания весовых множителей путем многократного анализа имеющегося ряда данных. При этом в качестве критерия используется достигаемая точность прогноза (минимальная дисперсия ошибки). Однако такой подход содержит и недостатки.

¹⁾ См. разд. 5.2 и 5.3 настоящей главы. — *Прим. ред.*

Во-первых, если в средней ошибке есть значимые разности (они должны быть равны нулю), то более вероятно, что эти разности больше зависят от способа выбора начальных значений коэффициентов модели, чем от различий в скорости затухания весовых множителей или постоянной сглаживания.

Во-вторых, еще более важный источник возможной ошибки можно проиллюстрировать с помощью следующего примера. Будем рассматривать очень длинный ряд чисел как некоторый коррелированный процесс. Этот ряд, стационарность и однородность которого гарантирована самим способом его получения, разделим на короткие отрезки, содержащие достаточно данных для отыскания наилучшего значения скорости затухания весовых множителей. Проанализируем результаты, получаемые для каждого из этих отрезков. Хотя все отрезки относятся к одному и тому же процессу, тем не менее существует широкое распределение соответствующих значений скорости затухания. Определить скорость затухания для данного отрезка можно только после того, как он стал историей. Величина скорости на следующем отрезке, которая может быть другой, будет известна только после его прохождения, и, следовательно, для прогноза такие значения вообще бесполезны.

Разность между точностью, получаемой при постоянном использовании некоторой стандартной скорости затухания весовых множителей, например $\alpha = 0,9$ (существуют предпосылки для теоретического обоснования этой величины), и точностью, которая может быть достигнута при заранее известном оптимальном значении скорости затухания, относительно мала по сравнению с выборочной ошибкой скорости затухания в том случае, когда наилучшее значение весовых множителей для длинного ряда выбирается по данным для коротких отрезков этого ряда (содержащих не менее 50 результатов наблюдений).

4.4. Байесовские прогнозы

Гаррисон и Стивенс [15] разработали строгий подход к прогнозированию временных рядов, включающий в виде частных случаев большинство ранее разработанных методов. При таком подходе каждому наблюдению до его проведения ставится в соответствие ряд первичных вероятностных значений коэффициентов модели. Построение модели можно начинать из состояния полной неопределенности, так как даже в этом случае развитый метод позволяет довольно точно устанавливать вероятностные значения коэффициентов.

После того как получены результаты наблюдений, по правилу Байеса определяют апостериорные вероятности, основываясь на которых вычисляют распределение вероятностей прогнозируе-

мой величины. Такой подход позволяет игнорировать кратковременные изменения прогнозируемой переменной и в то же время четко регистрировать те изменения в основном процессе, которые на графике временной зависимости изображаются в виде ступеней и изменения наклона соответствующих кривых. Многократное повторение приведенной схемы вычислений дает возможность уменьшить число вероятностных распределений до числа состояний изучаемого процесса. В число состояний обычно входят нейтральное состояние, которое характеризуется неизменными значениями всех элементов процесса, и состояния, которые соответствуют различным возможным значениям каждого коэффициента и помех. В таких состояниях дисперсия распределения намного выше, чем в нейтральном состоянии, что указывает на отклонение того или иного коэффициента модели от нормы. Возможность уменьшения числа состояний весьма существенна, так как в противном случае число распределений возрастало бы пропорционально квадрату числа наблюдений.

5. ОШИБКИ ПРОГНОЗА

В любом из методов прогнозирования временных рядов собственно прогноз представляет собой, по существу, оценку ожидаемого распределения результатов наблюдений в будущем. Для того чтобы на основе полученного прогноза можно было принимать решение, в большинстве случаев необходимо знать исходное распределение. Если это распределение описывается стандартной функцией, то, определив один или два его параметра, можно оценить вероятности возможных результатов будущих наблюдений.

5.1. Вид распределения вероятностей

Для целей формального анализа часто удобно принимать предположения о том, что помеха (ошибка) в исходных данных подчиняется нормальному закону распределения. Однако это предположение нельзя считать обоснованным для любых реальных наблюдений. Дело в том, что распределение шума нередко асимметрично и может иметь две моды или большее количество мод. Даже когда распределение симметрично относительно единственной моды, экспесс плотности распределения может сильно отличаться от 3.

Прогнозируемая выходная переменная характеризуется распределением, дисперсия которого зависит от дисперсии шума в исходных данных. Однако, поскольку прогнозы вычисляются на основе результатов многих наблюдений, входная переменная обычно описывается нормальным законом распределения незави-

симо от того, каков вид распределения, описывающего шум в исходных данных.

Распределение ошибок при прогнозировании зависит как от распределения самой выходной переменной, так и от распределения соответствующего ей шума, который обычно считается не зависящим от шума в исходных данных. При прогнозировании суммы значений переменной интенсивности в течение времени упреждения, которое намного больше интервала между наблюдениями, распределение ошибок представляет собой распределение суммы выборок из нескольких таких распределений и с возрастанием времени упреждения все более и более стремится к нормальному.

5.2. Дисперсия ошибок прогноза

В большинстве случаев, представляющих практический интерес, исходное распределение может быть восстановлено из полученного по прогнозу среднего значения и какого-либо другого параметра, характеризующего распределение выходной переменной. В качестве такого параметра, как правило, выбирается стандартное отклонение.

Если имеется какой-то набор прогнозируемых временных рядов, то обычно оказывается, что для каждого ряда существует определенная зависимость между разбросом ошибок и уровнем прогноза. Например, хотя объемы производства пшеницы (a_w) и риса (a_r), прогнозируемые на данный год, могут сильно отличаться друг от друга, между прогнозами производства разных зерновых культур существует четкая корреляция в соответствии с соотношением $\sigma^2 \propto \mu a^2$, где параметры μ и λ представляют собой характеристики всего семейства прогнозов зерновых. Этот же результат справедлив и для прогнозов спроса на различные предметы, относящиеся к одному и тому же классу. Отсюда следует, что, исходя из всего семейства прогнозов, можно оценивать параметры μ и λ и использовать связывающее их соотношение в целях установления стандартного отклонения распределения ошибок для любых отдельных рядов.

Другое соотношение между величиной a и параметрами σ , μ и λ ($\sigma^2 = \mu a + \lambda a^2$) было предложено в работе [11], а основы анализа процессов, описываемых подобными соотношениями, рассмотрены в работе [24].

Среднее абсолютное отклонение представляет собой оценку отклонений, которая имела бы место, если бы прогнозы разрабатывались на основе минимизации суммы абсолютных величин остаточных разностей между результатами наблюдений и медианой распределения вместо обычно проводимой минимизации сум-

мы квадратов отклонений от среднего. При появлении методов исследования операций, когда возможности обработки данных были ограничены, среднее абсолютное отклонение использовалось вместо стандартного отклонения только потому, что его вычисление требовало меньше времени и меньшего объема памяти. Особенно широко этот показатель использовался в программах ЭВМ для прогнозирования спроса [6]. Однако при достаточно высоком быстродействии и большой емкости памяти современных ЭВМ применение среднего абсолютного отклонения нецелесообразно.

Если распределение описывается нормальным законом, то среднее абсолютное отклонение $m = 0,8\sigma$ (точное значение коэффициента при σ равно $\sqrt{2/\pi} = 0,7979$). Для других распределений аналогичное соотношение может иметь иной вид. Так как средняя ошибка прогноза должна быть равна нулю, то дисперсия определяется среднеквадратичной ошибкой. Выбрав подходящую модель процесса, которая математически может быть выражена, например, либо константой, либо через линейные и сезонные составляющие, либо дисперсию, можно пересматривать значения коэффициентов этой модели по среднеквадратичной ошибке каждый раз, когда поступает новая информация и заново измеряется ошибка прогноза.

Если шум результатов наблюдений автокоррелирован незначительно, то дисперсия ошибок прогноза переменной интенсивности в течение времени упреждения будет равна сумме дисперсий ошибок прогнозов для отдельных интервалов, на которые разбивается время упреждения L с целью периодического уточнения прогноза. Следовательно, стандартное отклонение, используемое для принятия решений, приблизительно равно $\sigma\sqrt{L}$. В тех случаях, когда имеет место автокорреляция шума, выражение для стандартного отклонения может быть записано в виде $\sigma_L = L^\delta \sigma_1$, где σ_1 — стандартное отклонение ошибок одного знака, L — время упреждения, измеренное в интервалах между проверками прогноза, и δ — постоянная, которая является характеристикой всего семейства прогнозируемых рядов.

5.3. Метод следящих сигналов

Как уже отмечалось, точность прогноза зависит от шума входных переменных. При этом разумно считать, что прогнозы не содержат систематической ошибки, так как средняя ошибка прогноза должна быть равна нулю. Пока модель соответствует изучаемому процессу, ошибки прогноза должны колебаться около нуля. Если же модель прогноза ошибочна или станет ошибочной из-за того, что резко изменится сам процесс, то появится последователь-

ность «положительных» или «отрицательных» ошибок и, следовательно, средняя ошибка прогноза больше не будет равна нулю.

Метод следящих сигналов состоит в том, что с помощью ЭВМ систематически осуществляется проверка близости средней ошибки прогноза к нулю. Если средняя ошибка превысит некоторый заранее установленный предел, вычислительная машина может сформировать предупреждение, дающее возможность пользователю принять необходимые меры для своевременной корректировки прогноза.

Для осуществления такой корректировки прежде всего необходимо установить конкретную причину смещения прогноза; после этого внести в модель достаточно обоснованные изменения: принять новые значения для одного или большего количества коэффициентов, изменить некоторые или все весовые множители, или, наконец, изменить сам вид модели. В соответствии с одним из существующих в настоящее время научных направлений эта задача должна решаться автоматически с помощью специально создаваемых вычислительных систем, способных заменить труд большого количества людей. В соответствии с другим направлением в решении указанной задачи должны принимать участие специалисты, учитывающие весь комплекс информации, собираемой за длительные промежутки времени. Тот факт, что многие конкретные прогнозы не поддаются автоматическому уточнению, может указывать на необходимость постоянного участия специалистов в процессе прогнозирования, с тем чтобы сократить время отклика запрограммированных решений, позволяющих внести необходимые изменения в прогноз.

Для обнаружения выбросов следящих сигналов и соответственно для выявления недопустимых отклонений средней ошибки прогноза могут быть использованы следующие виды контрольных величин.

Кумулятивные суммы. Величина предложенной в работе [6] кумулятивной суммы ошибок прогноза, задаваемой выражением $Y_t = Y_{t-1} + e_t$, при $Y_0 = 0$ колеблется около нуля, если прогнозы не смещены, и быстро возрастает при появлении последовательности ошибок прогноза одного знака. Для интенсивности затухания весовых множителей β в модели, которая может быть описана полиномом $(n-1)$ -й степени, дисперсия кумулятивной суммы $\sigma_Y^2 = [(1 + \beta)/2 (1 - \beta^{2n})] \sigma_e^2$, где σ_e^2 — дисперсия ошибок прогноза. Выброс может быть зарегистрирован, например, тогда, когда выполняется соотношение $|Y| \geq 3\sigma_Y$. Однако очевидно, что в этом соотношении для уменьшения числа ложных регистраций вместо множителя 3 может быть использован множитель $k > 3$, а для уменьшения вероятности пропуска фактических изменений ошибок прогноза — множитель $k < 3$.

Сглаживание ошибки. Простое экспоненциальное сглаживание алгебраической суммы ошибок прогноза в виде $Y_t = \beta Y_{t-1} + (1 - \beta) e_t$ [26] создает основу для регистрации выброса, например в случае выполнения соотношения $|Y| \geq 0,4 m$ (вместо множителя 0,4 может быть использован и какой-либо другой множитель). При этом среднее абсолютное отклонение m замещается величиной $m_t = \beta m_{t-1} + (1 - \beta) |e_t|$.

Такой подход имеет два весьма существенных преимущества по сравнению с предыдущим. Предположим, во-первых, что произошла единственная большая ошибка, по все же не настолько большая, чтобы вызвать выброс следящего сигнала. Начиная с этого момента при использовании первого метода величина (невзвешенной) кумулятивной суммы колебалась бы не около нуля, а около уровня, соответствующего этой единственной большой ошибке, так что даже при вполне допустимых ошибках последующих прогнозов должны были бы регистрироваться выбросы следящих сигналов. Допустим, во-вторых, что следящий сигнал близок к выбросу (кумулятивная сумма близка к одному из пределов) и мы получили идеальный прогноз. При применении первого метода величина невзвешенной кумулятивной суммы не изменилась бы, однако оценка стандартного отклонения σ_e должна была бы уменьшиться, что привело бы к сжатию области допустимых отклонений и, следовательно, увеличению вероятности регистрации выбросов следящего сигнала. Использование сглаженных ошибок позволяет успешно преодолеть обе указанные трудности.

Дисперсия сглаженных ошибок прогноза может быть выражена соотношением $\sigma_y^2 = [(1 - \beta)/(1 + \beta)^2] \sigma_e^2$.

V-образные маски. Маски подобного типа были разработаны в работе [2] на основе метода последовательного анализа Вальда. V-образная маска помещается вершиной вблизи самого последнего значения (невзвешенной) кумулятивной суммы ошибок прогноза, при этом ее раскрытая часть обращена в сторону предыдущих членов ряда. Если хотя бы одно из предыдущих значений кумулятивной суммы лежит вне маски, то это является признаком значительного смещения прогнозов. Угол наклона и точка пересечения двух прямых, составляющих V-образную форму маски, вычисляются на основе анализа ошибок и вероятностей пропустить значительное отклонение и ошибочно обнаружить какое-то отклонение, когда его на самом деле не было.

Данный метод был обобщен для масок параболической формы [9] при всех возможных уровнях смещения прогноза, которые можно считать существенными. Применение такого подхода к анализу временных рядов результатов наблюдений курса акций на фондовой бирже позволило успешно прогнозировать предстоящие изменения на бирже.

5.4. Коррекция исходных данных

Внешние воздействия и ошибки, имевшие место при регистрации исходных данных, могут привести к тому, что результаты последующих наблюдений нельзя будет использовать для уточнения прогноза. Если отклонение полученного результата от прогнозируемого значения превышает 4σ, то рекомендуется по возможности скорректировать входные переменные и только после этого продолжать прогнозирование на основе выбранной модели.

6. МЕТОДЫ, ПРИМЕНЯЕМЫЕ ДЛЯ УТОЧНЕНИЯ ПРОГНОЗА

Существует много различных методов прогнозирования, которые (при правильном их использовании) должны приводить практически к одному результату — выбору одного и того же варианта из ряда возможных и одинаковой оценке экономических последствий выбранного варианта [12]. Правильное ведение прогнозирования предполагает, в частности, применение соответствующих критериев при пересмотре прогнозов. Известны три вида подходов, широко применяемых для пересмотра и уточнения прогнозов при анализе больших рядов, таких, как спрос на запасные товары, курс акций или стоимость товаров. В данном разделе рассматриваются основные черты каждого из этих подходов, но, чтобы ознакомиться со всеми деталями, которые необходимы для практической работы, целесообразно обратиться к многочисленным оригинальным источникам.

6.1. Метод адаптивного сглаживания

Основы метода адаптивного сглаживания были заложены при разработке метода экспоненциального сглаживания и затем развиты путем включения в рассмотрение сезонных моделей [7, 8]. Адаптивное сглаживание успешно применяется при прогнозировании спроса для управления запасами и поэтому вводится в соответствующие пакеты прикладных программ. Этот метод целесообразно использовать в тех случаях, когда модели спроса достаточно хорошо описываются с помощью постоянного уровня, тренда и сезонных моделей, которые не претерпевают резких изменений для большинства рядов. Во многих случаях адаптивное сглаживание позволяет получать хорошие результаты в течение короткого промежутка времени, что иногда имеет более важное значение, чем получение более точного прогноза в течение длительного времени и с применением значительных усилий.

Математическое описание модели, позволяющей проводить прогнозирование для любых рядов, имеет вид $\hat{X} = aF$ (разд. 3), где a — вектор m коэффициентов, вычисленных отдельно для

каждого временного ряда, $\mathbf{F}$ — матрица соответствующих функций размерности $m \times T$, строки которой соответствуют каждому из m членов модели и столбцы — каждому из T интервалов прогнозирования. Компоненты вектора $\hat{X}$ представляют собой прогнозируемые значения для каждого из интервалов.

Если в течение годового цикла имеется 12 наблюдений, то основная частота равна $\omega = 2\pi/12$ и существуют только семь возможных видов моделей. Простейшая модель включает только уровень и тренд, а наиболее сложная содержит также члены рядов Фурье вплоть до $\cos 6\omega t$. Строки матрицы $\mathbf{F}$ приводятся ниже.

	1	2	3	4	5	6	7
$f_1(t) = 1; f_2(t) = t$	x	x	x	x	x	x	x
$f_3(t) = \cos \omega t; f_4(t) = \sin \omega t$		x	x	x	x	x	x
$f_5(t) = \cos 2\omega t; f_6(t) = \sin 2\omega t$			x	x	x	x	x
$f_7(t) = \cos 3\omega t; f_8(t) = \sin 3\omega t$				x	x	x	x
$f_9(t) = \cos 4\omega t; f_{10}(t) = \sin 4\omega t$					x	x	x
$f_{11}(t) = \cos 5\omega t; f_{12}(t) = \sin 5\omega t$						x	x
$f_{13}(t) = \cos 6\omega t$							x

При других основных частотах, конечно, возможны и другие модели.

Коэффициенты a вычисляются сначала по методу наименьших квадратов в соответствии с N исходными данными ($m \times N$) матрицы $\mathbf{F}$. Гармоники включаются в модель вплоть до той частоты, при которой показатель степени (квадрат амплитуды) составляет по крайней мере 0,25 от величины дисперсии остатков, получаемых при подгонке этой модели с учетом имеющихся для ошибок степеней свободы.

С каждым новым наблюдением вектор a заменяется на вектор $a_T = L'a_{T-1} + he$, где L — матрица перехода ($m \times m$), которая определяется с помощью выражения $f(t) = Lf(t-1)$ и переводит значения соответствующих функций из одного периода времени в следующий. Столбец матрицы $\mathbf{F}$, имеющий номер t , обозначается как $f(t)$, и e представляет собой ошибку (скаляр) самого последнего прогноза, сделанного на один интервал вперед: $e = x_T - \hat{X}_{T-1}$. Вектор h , характеризующий m постоянных сглаживания, определяется следующим образом. Используя з-преобразования и некоторые элементарные тригонометрические соотношения [7, 8], можно вывести выражение для матрицы $\mathbf{H}$ в замкнутом виде: $\mathbf{H} = \sum \beta^j f(-j) f'(-j)$. Так как $\beta < 1$, то значимость более старых данных снижается и увеличивается вес, приписываемый более поздним данным.

Следовательно, вектор сглаживания можно записать как $\mathbf{h} = \mathbf{H}^{-1}\mathbf{f}(0)$. Величины $\mathbf{L}$ и $\mathbf{h}$ необходимо вычислить только один раз, так как при дальнейшем пересмотре прогнозов они не изменяются.

Если прогнозы составляются для переменных интенсивности (например, таких, как спрос на запасенные товары) и результаты наблюдений получены в нерегулярные интервалы времени, то исходные данные преобразуются к величинам, характеризующим скорость их изменения на равномерных интервалах времени. Получаемые в результате прогнозы могут быть приведены в соответствие с величиной будущих интервалов времени. Заметим, что подобное использование сечения какой-либо входной переменной одинаково пригодно для любых рядов и не приводит к такому же увеличению дисперсии, какое имело бы место в случае получения сечения из данных предыстории.

6.2. Метод Бокса — Дженкинса

Основные положения и сам метод рассматриваются в работах [4, 5]. Согласно данному методу, прогнозы составляются на основе всестороннего анализа статистических свойств рядов. Для того чтобы в полной мере использовать его достоинства, требуются достаточно длинные ряды (содержащие не менее 100 результатов наблюдений); это в значительной степени помогает понять изучаемый процесс с точки зрения получения предварительных данных относительно возможных изменений процесса в будущем. С помощью данного метода может быть проанализирован гораздо более широкий круг моделей, чем с помощью метода адаптивного сглаживания. Кроме того, он особенно полезен в тех случаях, когда требуется получить как можно более точный прогноз.

Анализ моделей по методу Бокса — Дженкинса начинается с формирования разностей из исходных данных Z_t и вычисления частных автокорреляционных функций φ для разностных рядов $w_t = \nabla^d z_t$. Затем разностные ряды вводятся в модель с соответствующим числом коэффициентов авторегрессии и скользящего среднего:

$$w_t = \varphi_1 w_{t-1} + \dots + \varphi_p w_{t-p} + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q},$$

где φ_j — частные автокорреляционные функции. Для оценки коэффициентов в моделях такого типа и для вычисления дисперсии шума в результатах наблюдений используются методы наименьших квадратов.

Операторы разности скользящего среднего могут быть разложены таким образом, чтобы выразить модель прогнозов через исходные данные наблюдений:

$$z_t = \Phi_1 z_{t-1} + \dots + \Phi_p z_{t-p} + \theta_0 + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q}.$$

По методу Бокса — Дженкинса сначала прогнозируется изменение преобразованных входных переменных, а затем осуществляется обратное преобразование.

Если необходимо проанализировать сезонные циклы, то вклад от каждого из них вычисляется путем составления разностей для определенных отрезков цикла. Для годового цикла, например, в качестве таких отрезков можно принять месяцы, и тогда прогноз для каких-то месяцев в будущем будет основываться на известных данных для этого месяца в прошлые годы.

Одно из основных применений метода Бокса — Дженкинса связано с управлением процессами в химической промышленности, где имеются ряды температур, давлений, скоростей потоков и т. п. с большим числом известных результатов наблюдений и неплохо известна возможная динамика изменения системы, подлежащей управлению. Цель управления состоит в том, чтобы наиболее рациональным образом менять технологические параметры — температуру, давление или скорость потока. Значительные усилия по разработке точных прогнозов для подобных производственных процессов вполне оправдываются достигаемым при этом экономическим эффектом.

6.3. Байесовские прогнозы

В работе [15] предложена схема разработки прогноза, которая включает основные положения двух предыдущих подходов как частные случаи. Ниже (после небольшого вступления) приводится строгое изложение математических основ этого метода и руководство для его практического использования.

Пока результаты наблюдений согласуются с прогнозами, последние стабильны и ожидаемое распределение ошибок прогнозов ограничено узкими пределами в соответствии с шумом в результатах наблюдений. Появление хотя бы одного большого отклонения повышает чувствительность метода. Если последующие наблюдения подтверждают скачкообразное или постепенное изменение изучаемого процесса, прогнозы быстро корректируются с помощью новой модели. Однако, если указанное отклонение представляло собой случайный выброс, прогноз остается без изменения.

Объем вычислений при использовании рассматриваемого метода гораздо больше, чем при адаптивном сглаживании, так как успешное прогнозирование гарантируется и в тех случаях, когда могут возникнуть непредвиденные и существенные изменения в прогнозируемом процессе. Это справедливо даже в случае изучения спроса на товары широкого потребления, который особенно сильно подвержен влиянию моды или неконтролируемой рыночной конъюнктуры. Если есть предварительные сведения об ожидаемых событиях, то можно начать прогнозирование при весьма

скромных данных или вообще при их отсутствии. Вместе с тем, если имеется большой объем результатов наблюдений, прогнозирование можно начать при отсутствии исходных предпосылок о поведении прогнозируемой величины. Применяемый алгоритм вычислений позволяет построить модель процесса по данным его предыстории.

Типичная модель для результатов текущих наблюдений может быть представлена выражением $d_t = \mu_t + \varepsilon_t$, где μ — уровень полезной составляющей и ε — шум текущего наблюдения. Текущий уровень связан с трендом выражением $\mu_t = \mu_{t-1} + \beta_t + \gamma_t$, где тренд изменяется по закону $\beta_t = \beta_{t-1} + \delta_t$. Предполагается, что скачкообразное изменение γ основного уровня, изменение δ тангенса угла наклона кривой, описывающей зависимость выходной переменной от времени, а также шум ε независимы и распределены по нормальным законам с нулевым средним и известными (но не обязательно постоянными) дисперсиями V_ε , V_γ и V_δ .

В модели, которая описывает много состояний, обычно имеются два состояния для каждого возможного возмущения, причем каждое из них характеризуется своим значением дисперсии V^j . В устойчивом состоянии дисперсия мала, а в состоянии изменения велика. Имеются также априорные вероятности π_j того, что состояние j не изменяется. Ниже приведен пример четырех состояний и соответствующих им условных вероятностей и дисперсий, использованных авторами метода в качестве типичных начальных условий. Дисперсия шума V_0 зависит от соотношения между фактическими изменениями процесса и текущим прогнозом (разд. 5.2) и описывается соотношением $R^{(j)} = V^{(j)}/V_0$.

Характеристики состояния	j	π_j	R_ε	R_γ	R_δ
Стабильное	1	0,900	1	0	0
Скачкообразное изменение уровня выходной переменной	2	0,003	1	100	0
Постепенное изменение тангенса угла наклона	3	0,003	1	0	1
Переходное	4	0,094	101	0	0

Пусть вектор Φ содержит следующие элементы: ожидаемое значение уровня, ожидаемое значение тангенса угла наклона, дисперсию ошибок при определении уровня, ковариацию наклона и уровня, а также дисперсию ошибок при определении тангенса угла наклона. Тогда, если в момент времени $t - 1$ совместное распределение уровня и тангенса угла наклона является двумерным нормальным и характеризуется параметрами Φ_{t-1} , то соответ-

ствующее распределение в момент времени t также будет двумерным нормальным, а его параметры будут определяться вектором Φ_t .

Перемешные, находящиеся в левой и правой частях приведенных ниже уравнений, относятся соответственно к текущей и предшествующей итерациям:

$$m = m \div b \div A_1 e, \quad b = b \div A_2 e,$$

$$v_{\mu\mu} = r_{11} - A_1^2 V_e, \quad v_{\mu\beta} = r_{12} - A_1 A_2 V_e, \quad v_{\beta\beta} = r_{22} - A_2^2 V_e.$$

Пять постоянных сглаживания определяются следующим образом:

$$r_{11} = v_{\mu\mu} + 2v_{\mu\beta} \div v_{\beta\beta} \div V_e + V_\delta,$$

$$r_{12} = v_{\mu\beta} \div v_{\beta\beta} + V_\delta, \quad r_{22} = v_{\beta\beta} \div V_e,$$

$$V_e = r_{11} \div V_e, \quad A_1 = r_{11}/V_e, \quad A_2 = r_{12}/V_e,$$

где e — ошибка, полученная при сравнении результата текущего наблюдения и самого последнего прогноза. Существуют также выражения для коэффициентов сезонных сечений, которые здесь не приводятся.

Если первоначально имелось N возможных состояний, то данному процессу соответствуют N^2 возможных комбинаций. Эти N^2 возможностей описываются N двумерными нормальными распределениями с одинаковыми (взвешенными) первыми и вторыми моментами, так что процедура корректировки прогноза может проводиться без ограничений.

7. ИЗУЧЕНИЕ СФЕРЫ СБЫТА

Ни в одном из статистических методов прогнозирования не может заранее предусматриваться изменение модели прогнозируемого процесса. Как уже отмечалось, имеется ряд изящных методов быстрого обнаружения изменений в последующих процессах и соответствующей реакции на них. Модель прогноза может все более усложняться, когда это экономически оправдано и позволяет глубже проникнуть в механизм наблюдаемых явлений.

Во многих случаях изменения в изучаемом процессе можно предвидеть заранее, но в модель прогноза они тем не менее не включаются. Например, интенсивная кампания по подписке может увеличить число подписчиков на периодические издания; принятие проекта о дополнительном налоге вызывает волну покупок как раз до того, как он входит в действие; массовое ознакомление покупателей с новой продукцией типа Tange или Polaroid SX-10 не только позволяет продать ее, но и создает представление о типичной продукции такого рода и меняет спрос на старую продукцию; директивы Вашингтона радикально меняют соотношение закупок и расхода нефтяных продуктов.

Последствия таких изменений в изделиях, в спросе на них, экономике и в правительственной регулировании не могут быть точно рассчитаны. Тем не менее опытные руководители на основе тщательного анализа различных вариантов могут предсказать характер изменений. И это все же лучше, чем полное игнорирование проблемы из-за отсутствия формальных методов ее решения. В подобной ситуации единственно неверным является прогноз, не предсказывающий никаких изменений [14]. Поэтому руководителям необходимо иметь обзор статистических прогнозов, в который должны регулярно вноситься изменения, отражающие появление новой продукции, ее развитие, действия конкурентов, погодные условия, образование дефицита материалов и т. д.

Одной из функций любой действующей прогнозирующей системы должна быть подготовка отчетов по отдельным вопросам статистических прогнозов с учетом изменений всех параметров модели сферы сбыта. В таких отчетах данные могут быть сгруппированы по различным показателям: типам продукции, рынкам сбыта, периодам времени, — что облегчает их восприятие человеком. Иногда общие объемы продажи какой-либо корпорации или прогнозы доступных рынков не будут соответствовать сумме предсказанных объемов продажи отдельных видов продукции.

Если появляется несоответствие между зонами, заключенными в пределах от минимальных до максимальных значений прогноза, руководители должны рассмотреть причины этого несоответствия и факторы, учтенные при получении прогнозов. Далее, если для некоторой группы товаров требуется скорректировать прогнозы, чтобы привести их в соответствие с достигнутыми показателями продаж, то необходимо не только решить вопрос о том, как осуществить эту коррекцию по всей группе товаров, но также и о том, какие нужно принять специальные меры, чтобы гарантировать реализацию этих прогнозов.

Если решения о производстве и распределении приняты на основе хорошего прогноза и торговые операции гарантируют соответствие действительных показателей продаж этому прогнозу, то издержки, связанные с ошибочностью прогноза, будут минимальны. Торговля при неточном прогнозировании может представлять собой более экономичный путь к правильным решениям, чем сложные попытки точного прогнозирования поведения рынка.

Для корректировки прогноза иногда применяют прием, заключающийся в обеспечении обратной связи. Проиллюстрируем его на конкретном примере.

Допустим, что по статистическому прогнозу объем продаж на март составляет 1000 ед. В то же время по субъективному прогнозу на этот же месяц, принятому ранее с учетом обеспечения достаточного количества товаров, объем продаж составлял 1500 ед. В апреле стало известно, что фактический объем продаж по отчету

там за март равен 1375 ед. Если бы поступление товаров обеспечивалось на основании статистической модели, то имел бы место потенциальный дефицит в 375 ед. (смягченный, конечно, фондом запаса). Если бы компания руководствовалась данными субъективного прогноза, то был бы получен потенциальный избыток в 125 ед. Тогда при наличии потерь, обусловленных недостатком товаров, вдвое более весомых, чем от избытка товаров, чистый выигрыш за счет использования субъективного прогноза составил бы $(2 \cdot 375) - 125 = 625$ ед. Чистый выигрыш может быть выражен в долларах или в виде нормирующего показателя для стандартного отклонения ошибок прогноза.

В любой отдельный период времени существует, очевидно, несколько серий прогнозов, отличных от простого описательного прогноза. Вместе с тем результаты изучения рынка, представленные цифрами значимости изменений, расположенными по порядку от высоких до низких, действуют на людей как положительная обратная связь, побуждающая к дальнейшим значительным и содержательным изменениям. Это также позволяет минимизировать время, затрачиваемое на внесение незначительных изменений, и активно вносить действительно существенные изменения. Независимо от того, вносятся ли изменения в прогноз или в результате прогноза принимаются более решительные действия при продаже, результатом является рост прибыли, повышение эффективности торговли и рост престижа организации.

ЛИТЕРАТУРА

1. Ayres R. U., *Technological Forecasting and Long Range Planning*, McGraw-Hill, N.Y., 1969.
2. Barnard G. A., *Control Charts and Stochastic Processes*, *J. Roy. Stat. Soc.*, 21, Series B, 239—271 (1959).
3. Bass F. M., A New Product Growth for Model Consumer durables, *Management Sci.*, 15, 5, 215—227 (January 1969).
4. Box G. E. P., Jenkins G. M., Some Statistical Aspects of Adaptive Optimization and Control, *J. Roy. Stat. Soc.*, 24, Series B, 297ff (1962).
5. Box G. E. P., Jenkins G. M., *Time Series Analysis Forecasting and Control*, Holden-Day, San Francisco. 1970.
6. Brown R. G., *Statistical Forecasting for Inventory Control*, McGraw-Hill, N.J., 1959.
7. Brown R. G., *Smoothing, Forecasting and Prediction of Discrete Time Series*, Prentice-Hall, Englewood Cliffs, N.J., 1963.
8. Brown R. G., *Decision Rules for Inventory Management*, Holt, Rinehart and Winston, N.Y., 1967.
9. Brown R. G., Detection of Turning Points in a time Series, *Decision Sci.*, 2, 4, 383—403 (October 1971).
10. Brown R. G., APL * Plus 747 Forecast User's Guide. Scientific Time SHARING Corp., Washington, D.C.. 1973.
11. Burgin T. A., Wild A. R., Stock Control — Experience and Usable Theory, *Operational Res. Quant.*, 18, 1, 35—52 (1967).

12. Chambers C. C., Mullick S. K., Smith D. D., How To Choose the Right Forecasting Technique. *Harvard Bus. Rev.*, 65, 45—74 (July — August 1971).
13. Evans M., Macro-economic Activity: Theory, Forecasting and Control, Harper & Row, N.Y., 1969.
14. Forrester J. W., Industrial Dynamics, MIT Technology Press Cambridge, Mass., 1961.
15. Harrison P. J., Stevens C. F., A Bayesian Approach to Short-term Forecasting. *Operational Res. Quart.*, 22, 4, 341—362 (December 1971).
16. Howard R. A., Decision Analysis, Applied Decision Theory, Proceedings of the Fourth International Conference on Operational Research, Wiley, N.Y., 1966, pp. 55—71.
17. Jenkins G. M., Watts D. G., Spectral Analysis and its Applications, Holden-Day, San Francisco, 1969.
18. Kalman R. E., New Methods in Wiener Filtering Theory, Proceeding of the First Symposium on Engineering Application of Random Function Theory and Probability, Wiley, N.Y., 1963.
19. Leontief W., Input-Output Economics, Oxford University Press, N.Y., 1966.
20. Nyquist H., Certain Topics in Telegraph Transmission Theory, *AIEE Trans.*, 41, 614—644 (April 1928).
21. Rosenblatt M. (Ed.), Time Series Analysis, Wiley, N.Y., 1963.
22. Shishkin J., Electronic Computers and Business Indicators, National Bureau of Economic Research Occasional Paper 57, 1957.
23. Sittler R. W., Lectures on Sampled Data Systems Analysis, MIT Lincoln Laboratory Memorandum 2M-0671, 22. chapters 10—12 (August 1957).
24. Stevens C. F. On the Variability of Demand for Families of Items, *Operational Res. Quart.*, 25, 3, 411 (1974).
25. Stevens C. F., Harrison P. J., Bayesian Forecasting, a Series of Reports Issued Jointly by the Statistics Department of University College London and the University of Warwick.
26. Trigg D. W., Monitoring a Forecast System, *Operational Res. Quart.*, 15, 3, 271—274 (September 1964).
27. Trigg D. W., Leach A. G., Exponential Smoothing with an Adaptive Response Rate, *Operational Res. Quart.*, 18, 1, 53—60 (1967).
28. Van Dobben de Bruyn C. S., Smoothing Sales Data by Progressive Correction, *Statistica Neerlandica*, 17, 3, 243—265 (1969).
29. Wiener N., Extrapolation, Interpolation and Smoothing of Stationary Time Series, MIT Technology Press. Cambridge, Mass. 1949.
30. Winters P. R. Forecasting Sales by Exponentially Weighted Moving Averages, *Management Sci.*, 6, 3, 324—342 (April 1960).
31. Yule G. U., Why We Sometimes Get Nonsense Correlations Between Time Series, *J. Roy. Stat. Soc.*, 89 (new series) 1—64 (1926).

I-2. Управление трудовыми ресурсами

Р. Мэзон¹⁾, Э. Фламгольц²⁾

ВВЕДЕНИЕ

Наука о руководстве и методы исследования операций для решения задач управления трудовыми ресурсами до сих пор не получили должного развития и применения. Это объясняется главным образом тем, что при решении задач подобного рода приходится учитывать множество различных по степени важности факторов, существование которых обусловлено взаимными требованиями, предъявляемыми коллективом, отдельными индивидуумами и организацией друг к другу. Из всех возможных требований при анализе и выработке соответствующих рекомендаций в первую очередь необходимо учитывать следующие:

- организация должна иметь эффективную структуру должностей и число сотрудников, которые смогли бы обеспечивать своевременное производство необходимого объема продукции требуемого качества, и влиять на членов коллектива;

- организация должна эффективно использовать трудовые ресурсы, которыми она располагает;

- организация должна удовлетворять основные потребности отдельных индивидуумов и создавать условия для максимального проявления их личных качеств и способностей;

- производство товаров и уровень соответствующей заработной платы должны оказывать влияние на качество работы сотрудников организации и деятельность руководящих органов.

Первое требование, имеющее непосредственное отношение к выбору структуры организации и уровня руководства, связано с планированием и выполнением работ.

Второе требование, касающееся организации производства и управления его циклами, предусматривает изучение длительности и хода их выполнения, определение степени загрузки и распределение трудовых ресурсов, а также выбор средств стимулирования, повышающих производительность труда.

Третье требование связано с разработкой способов удовлетворения личных потребностей с изучением индивидуальных и групповых взаимоотношений с учетом психологических факторов.

Четвертое требование, по существу, сводится к необходимости учитывать социальные и правовые факторы, существенным образом влияющие на эффективность использования организацией трудовых ресурсов, которыми она располагает.

Улучшение качества функционирования некоторой социальной системы возможно только при тщательном анализе перечисленных выше требований, так как именно они определяют вид целевых функций и накладываемых ограничений, а также метод исследования, и в конечном счете от них зависит

¹⁾ Richard O. Mason, Ph. D., Associate Professor of Management and Information Systems, and Assistant Dean of the Graduate School of Management at the University of California, Los Angeles, California.

²⁾ Eric G. Flamholtz, Ph. D., Associate Professor of Accounting and Information Systems in the Graduate School of Management at the University of California, Los Angeles, California.

реализация разработанных рекомендаций. Таким образом, управление трудовыми ресурсами возможно только на научной основе, и именно этой цели и призваны служить наука о руководстве и методы исследования операций.

Для организации улучшение качества функционирования означает повышение эффективности выполнения работы ее исполнителями как в настоящий момент времени, так и в будущем. Для исполнителя улучшение качества функционирования означает увеличение заработной платы с целью повышения уровня своих потребностей, а также обеспечение определенного общественного положения и возможности развития и реализации индивидуальных способностей. Довольно часто эти две точки зрения противоречат одна другой, однако иногда они оказываются взаимосогласованными, т. е. решение, ведущее к улучшению функционирования одной из «конфликтующих» сторон, приводит также к улучшению функционирования и другой. Нахождение таких решений является одной из главных задач науки о руководстве и основной целью большинства исследований в области социально-технических систем, систем планирования и управления трудовыми ресурсами. Реальное улучшение функционирования организации возможно на основе теории принятия решений, причем наиболее доступным способом оценки эффективности применения методов исследования операций в области управления трудовыми ресурсами является определение пространства решений по управлению такими ресурсами. Этот подход позволяет классифицировать методы исследования операций, которые были использованы ранее или могут быть использованы в каждом из подпространств решений. Пространство решений в первую очередь определяется следующими функциями, которые являются основой управления.

1. *Комплектование штата организации*, т. е. набор кандидатов и отбор из их числа лиц, удовлетворяющих требованиям, которые руководство фирмы предъявляет к будущим сотрудникам.

2. *Повышение квалификации персонала*, т. е. общая и профессиональная подготовка и усовершенствование.

3. *Распределение трудовых ресурсов*, т. е. расстановка кадров и назначение на должности (предполагается, что задано расписание работ и список должностей, осуществляется более рациональное закрепление рабочих мест и должностей).

4. *Использование трудовых ресурсов*, т. е. разработка организационной структуры, информационных потоков и схем взаимоотношений между сотрудниками, стиля руководства, инструкций по выполнению работ (предполагается, что штат сотрудников задан или может быть спрогнозировано и осуществлено более рациональное распределение работ и должностей).

5. *Оценка эффективности персонала*, т. е. измерение вклада каждого сотрудника в успешную деятельность организации.

6. *Оплата и стимулирование персонала*, т. е. использование различных (экономических и неэкономических) форм вознаграждения и поощрения для стимулирования каждого сотрудника. Эта функция имеет непосредственное отношение к заработной плате, доходам, продвижению по службе, условиям труда, профессиональному и общественному признанию и социальным стимулам.

1. КОМПЛЕКТОВАНИЕ ШТАТА ОРГАНИЗАЦИИ

1.1. Набор кандидатов

Как уже отмечалось, комплектование — это набор кандидатов и отбор из их числа лиц, наиболее удовлетворяющих требованиям, которые предъявляются к будущим сотрудникам. При выполнении первого из указанных процессов, который, по существу,

представляет собой формирование множества кандидатов на получение работы, требуется решить сразу несколько задач, таких, как определение источников формирования множества кандидатов, определение необходимого числа кандидатов, определение требуемой квалификации и составление расписания набора кандидатов. В этом случае отбор будет состоять в принятии решения, кого из кандидатов следует зачислить на работу.

Обычно те методы исследования операций, с помощью которых могут быть решены задачи набора, предполагают существование относительно большого числа кандидатов, каждый из которых должен выполнять сравнительно небольшое число идентичных операций или заданий. Поэтому изучение задачи набора следует начать с определения реакции организации на численность персонала. Эта реакция отражает взаимосвязь начальных и конечных показателей организации по персоналу и может быть использована для минимизации либо издержек на содержание персонала, либо его численности. Задача выбора числа сотрудников может быть сформулирована следующим образом: минимизация величины N или $C(N)$ при условии, что $R(N) \geq G$, где N — уровень входного потока, выражаемый количеством людей, принимаемых на работу, объемом рабочих часов, числом сделок, заключенных с целью реализации продукции, и т. д., $C(N)$ — затраты на содержание персонала в зависимости от величины N (линейная функция затрат имеет вид $C_1 + C_2N$); G — целевой показатель, или требуемый уровень выходного потока, $R(N)$ — функция, характеризующая выходные показатели функционирования организации в зависимости от входного потока. В тех случаях, когда функция $R(N)$ является линейной, решение задачи тривиально. Однако в большинстве случаев эта функция аппроксимируется S -образной кривой (рис. 4). Начальный участок этой кривой соответствует исходному значению выходного показателя функционирования организации. Прямой участок может быть аппроксимирован линейной функцией и соответствует положительной реакции организации на возрастание уровня входного потока (т. е. любое увеличение входного потока приводит к возрастанию выходного потока). Система сохраняет положительную реакцию до тех пор, пока не будет достигнуто состояние насыщения. После этого дальнейшее увеличение уровня входного потока практически не приводит к возрастанию выходного потока. При достижении состояния перенасыщения при любом увеличении уровня входного потока выходной поток сокращается. Последнее обстоятельство имеет место в тех случаях, когда с ростом числа сотрудников быстро увеличиваются плотность их размещения, информационная перегрузка, возникают дополнительные проблемы взаимоотношений между сотрудниками и др., что в конечном итоге приведет к ограничению производительности труда. Например, в торговле

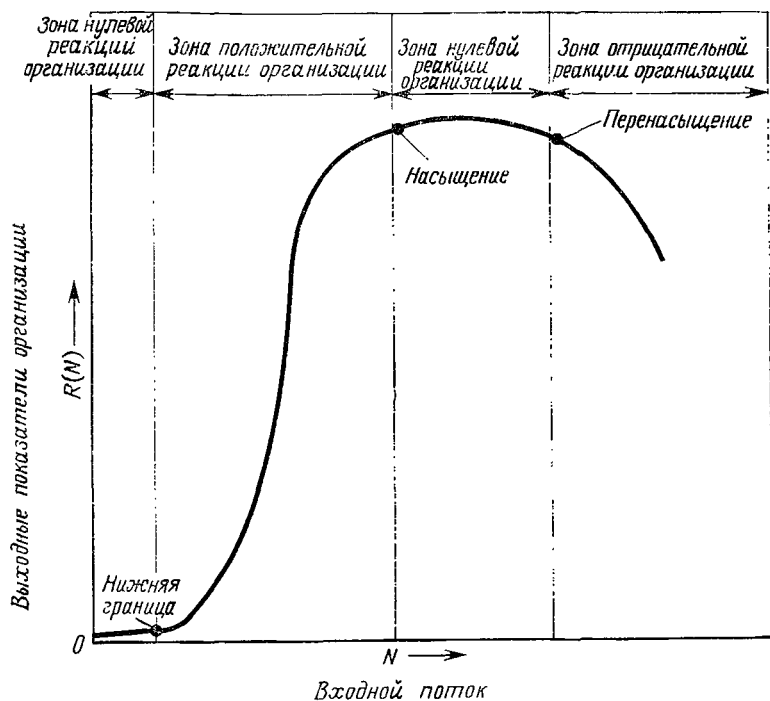


Рис. 1. Реакция организации на численность персонала.

и других сферах обслуживания подобная ситуация может возникнуть вследствие чрезмерных усилий, направленных на реализацию продукции (рост рекламы, организация новых форм торговли и т. д.), что в свою очередь может неблагоприятно повлиять на мнение покупателей и оттолкнуть потенциальных клиентов.

Научный подход к определению числа людей, принимаемых на работу, уровня ресурсов рабочей силы и рабочего времени, а также значений входных переменных предполагает нахождение наилучшего приближения для реакции организации. Такое приближение может быть получено либо путем подбора эмпирической кривой, либо с помощью проведения эксперимента.

Первый метод оказывается эффективным, если мы располагаем рядом данных, которые показывают, какие значения выходных переменных можно получить при различных уровнях входного потока, используемых в некоторой задаче или классе подобных задач. Для оценки функции реакции организации с помощью графических методов или нелинейной регрессии можно использовать в качестве модели кривую Перла—Рида или кривую Гомперт-

ца. Однако, как правило, данные, которыми мы располагаем, относятся к весьма узкому классу работ, и полное определение функции реакции организации возможно осуществить только с помощью дополнительных экспериментов. В любом случае необходимо, чтобы в течение рассматриваемого интервала времени в организации не имели места существенные технологические или организационные изменения.

Если выходные показатели организации (целевая функция) известны, то определить, какое число людей следует принять на работу, в общем случае можно либо с помощью прямых расчетов, либо с помощью методов математического моделирования или графическими методами. Так, Акофф и др. [1, 92] использовали рассматриваемый подход для определения перспективного планирования числа продавцов, которых следовало бы принять на работу в данной сфере обслуживания, и среднего числа сделок, которое необходимо было закрепить за каждым продавцом. В результате проведенного анализа было установлено, что область функционирования одной из организаций в зависимости от численности персонала соответствовала участку кривой (рис. 1), расположенному за точкой насыщения. В связи с этим было принято решение сокращать, а не увеличивать штат. Годовая экономия в результате отказа от предполагаемого дополнительного набора продавцов примерно в 25 раз превысила затраты на изучение этого вопроса.

Несмотря на то что нахождение функции реакции организации является основным этапом решения задач по управлению трудовыми ресурсами, довольно часто эту функцию считают линейной в предположении однородности множества сотрудников. Из всех существующих методов исследования операций при моделировании задач управления трудовыми ресурсами наиболее часто используют многопериодную потоковую модель, входной переменной которой являются люди, принимаемые на работу, при этом существующий и потенциальный персонал представляют в виде дискретного набора групп.

Для каждого данного момента времени разбиение на группы осуществляется по таким важным показателям, как квалификация, возраст, образование, оклад и должность. В результате получают совокупность групп, которая описывает состояние системы в момент времени t . Переход системы из одного состояния в другое, соответствующее моменту времени $t + T$, определяется структурой системы трудовых ресурсов и принятыми к этому моменту времени решениями. В течение любого периода времени люди поступают в организацию и покидают ее, повышают свою квалификацию, стареют; кроме того, изменяются должностные оклады, происходит перемещение по должностям и т. д. Все это приводит к изменению состава групп, т. е. к изменению состояния системы. В зависимости от характера регулярного или случайного

перехода системы из одного состояния в другое потоковые модели, описывающие такие системы, могут быть детерминированными или стохастическими.

Детерминированные модели, как правило, используются на многих этапах работы при планировании трудовых затрат или фонда рабочей силы. В этих случаях для решения поставленных задач применяются линейное программирование, динамическое программирование, принцип максимума и метод оптимального поиска.

Стохастические модели, используемые для предсказания перемещения по должностям и рабочим местам, основаны на предположении о случайном характере подобных перемещений. Наиболее достоверно такие модели описывают состояние системы с явно выраженной иерархической структурой либо состояние системы, которая ежегодно (или через какие-то другие фиксированные интервалы времени) осуществляет прием на работу и перемещение сотрудников внутри организации (к таким системам относятся военные организации, высшие и средние учебные заведения и некоторые промышленные предприятия).

Стохастические модели могут быть использованы для анализа как индивидуального, так и коллективного поведения людей в системе. С помощью такой модели можно определить вероятность того, согласен или нет данный сотрудник оставаться на соответствующем месте работы в той же должности, получать тот же оклад, согласен он или нет с установленным сроком работы и т. д. или намерен как-то изменить свое положение. Определить вероятность этих событий можно на основе данных о случайных переходах внутри системы за фиксированный промежуток времени, которые могут быть использованы для оценки полезности индивидуума для организации и предсказания модели продвижения его по службе. Поскольку предыстория и опыт работы индивидуума могут иметь значение для его продвижения по службе, вероятности перехода внутри системы определяются с учетом всех или многих предыдущих состояний, в которых находился индивидуум.

При анализе коллективного поведения предполагают, что структура организации удовлетворяет требованиям марковского анализа, т. е. что, во-первых, система имеет конечное число различных состояний, во-вторых, условные вероятности перехода системы из одного состояния в другое не зависят от ее предыстории, а зависят только от состояния, в котором система находится в данный момент, в-третьих, вероятности перехода не изменяются во времени (т. е. мы имеем дело со стационарными условиями) и, в-четвертых, известен набор начальных состояний системы в виде вероятностей. Если эти требования выполняются, то задача может быть сформулирована как частный случай линейной

динамической модели размножения в виде

$$X(t+1) = X(t)P + y(t)G + z(t)H,$$

где $X(t)$ — выходные (эндогенные) переменные (или переменные состояния), неуправляемые со стороны лица, принимающего решения, реализация которых определяется матрицей вероятностей перехода P ; $y(t)$ — регулируемые переменные, соотношение между которыми описывается структурной матрицей G ; $z(t)$ — входные (экзогенные) переменные, воздействующие на систему извне, соотношение между которыми описывается структурной матрицей H .

Марковская модель перемещений персонала базируется на использовании матрицы вероятностей перехода (рис. 2), которая составляется на основе существующих данных о системе или с помощью оценок экспертов. Если P_{ij} — вероятность перехода системы из состояния i в состояние j за данный отрезок времени, то, разбив последний на конкретное число интервалов, можно определить вероятность перехода системы из любого данного состояния в любое другое.

X (1970)	P				
		A	B	C	Выход
100	A	0,80	0,40	0,00	0,10
50	B	0,00	0,90	0,05	0,05
10	C	0,00	0,00	0,90	0,10
0	Выход	0,00	0,00	0,00	1,00

Предполагаемое перемещение сотрудников в 1971 г.

$$80 \ 55 \ 12^1 \ 14^1 = X(1971)$$

¹ Приводятся округленные результаты.

Рис. 2. Вероятность перехода сотрудника с должности на должность в период с 1970 по 1971 г. ($n = 4$).

При анализе строго иерархических систем принято составлять матрицу вероятностей с учетом степени важности осуществляемых переходов. В этом случае P_{ii} — вероятность остаться в состоянии i (вероятность продолжения существования), а P_{ij} — вероятность перехода на более важную работу, если $j > i$, и вероятность понижения в должности, если $j < i$. Кроме того, одно из состояний (обычно n -е) представляет выходное состояние. Матрица вероятностей перехода имеет следующие свойства: во-первых, она является квадратичной матрицей размерности n , где n — число состояний, во-вторых, $0 \leq P_{ij} \leq 1$ и, в-третьих, сумма ее каждой строки равна 1,0.

Если матрицу вероятностей перехода возвести в квадрат, то матрица P^2 будет описывать вероятность перехода в каждое из состояний в течение двух последовательных интервалов времени; соответственно матрица P^t будет описывать вероятность перехода в каждое из состояний в течение t последовательных интервалов времени. Упрощенный вариант основной модели, в котором не учитываются зависимости $y(t)$ и $z(t)$, можно представить следующим образом: $X(t+1) = X(t)P$, где $X(t)$ — n -мерный вектор, компоненты которого соответствуют числу сотрудников в каждом состоянии системы i ($i = 1, 2, \dots, n$) в течение интервала времени t ; $X(0)$ — начальный вектор распределения сотрудников; P — марковская матрица.

Модель может быть расширена, если учесть такие показатели, как набор и увольнение сотрудников.

В этом случае имеем $X(t+1) = X(t)P + R_t$, где R_t — n -мерный вектор, компоненты которого соответствуют числу людей, принятых на постоянную или временную работу (положительные коэффициенты), или числу сотрудников, уволенных окончательно или временно (отрицательные коэффициенты), в каждом состоянии в течение интервала времени t .

Используя приведенные выше модели, можно представить многошаговую модель планирования для данного периода T в следующем виде:

$$X(T) = X(0)P^T + \sum_{t=1}^T R_t P^{T-t}.$$

Вообще говоря, руководитель управляет кадрами во времени путем принятия соответствующих решений, предусматривающих набор и увольнение (R_t) или изменение темпов продвижения по службе, в том числе сохранение в той же должности или понижение в должности, а также изменение выходных переменных, вероятность реализации которых описывается матрицей P .

Следовательно, данная модель может быть использована для исследования эффективности различных вариантов решения в течение некоторого периода планирования. Кроме того, она может быть использована и для выбора варианта решения, обеспечивающего достижение цели путем зачисления в штат ($R_0, R_1, \dots, R_{T-1}$), если известны $X(0)$ и некоторое желаемое конечное, или заданное, состояние системы $X(T)$. В последнем случае для выбора наилучшего варианта (выраженного в терминах $n \times T$ — матрицы добавлений и исключений) из тех, которые обеспечивают достижение цели, должны быть применены дополнительные способы выбора решения или методы оптимизации.

Рассматриваемая модель может быть также использована при планировании состава персонала, для прогнозирования ожидаемого числа сотрудников, занимающих каждую из имеющихся

должностей в данный момент времени, или для выбора R_t — варианта решения, предусматривающего прием новых сотрудников.

Для определения степени удовлетворения потребности организации в кадрах полученные значения могут быть сопоставлены с прогнозируемой потребностью организации. Излишки, если они имеются, означают возможные перемещения внутри организации, временное увольнение и т. д.

Примеры применения стохастических моделей для планирования потребностей организации в кадрах можно найти в работах [5, 7, 74, 75, 91], а новые теоретические результаты в этом направлении представлены в работе [19].

В большинстве описанных моделей для прогнозирования будущих потребностей в кадрах используются данные о предыстории системы, и прогнозирование осуществляется с помощью разного рода нелинейных регрессионных моделей. Поскольку такой подход предполагает, что существующая структура системы остается неизменной, то, естественно, он неприменим в тех случаях, когда имеют место непредвиденные структурные изменения. Один из способов преодоления подобных трудностей предусматривает использование метода Дельфи. В соответствии с этим методом [18, 42] сначала формируют группы экспертов. Затем каждый эксперт независимо от других участников группы составляет перечень вопросов относительно возможных потребностей в кадрах. Полученный таким образом «вопросник» поступает к экспертам. Каждый эксперт знакомится с содержанием «вопросника» и на его основе вновь дает свою оценку возможных потребностей в кадрах. Затем составляют уточненный «вопросник», и процесс повторяют до тех пор, пока не будет достигнуто единого мнения или относительно малого расхождения во мнениях экспертов.

Милкович и др. [65] использовали метод Дельфи для прогнозирования потребностей в кадрах крупной государственной организации розничной торговли. В результате пяти циклов работы экспертов, информацию которых получили управляющие 7 компаний, было установлено, что метод Дельфи обеспечивает получение данных, более близких к фактическому состоянию предприятия в будущем, чем каждая из трех регрессионных моделей, основанных на прогнозировании числа точек розничной торговли, учете объема и фактического количества торговых точек.

1.2. Отбор из числа кандидатов лиц, наиболее удовлетворяющих требованиям, которые предъявляются к будущим сотрудникам

Сущность задачи отбора кадров для развития организации состоит в удовлетворении ее требований в части поддержания престижа и возможности выполнения стоящих перед ней задач.

Обычно прием на работу осуществляют, исходя из текущих потребностей организации, хотя в некоторых случаях делаются попытки оценить «потенциал» будущего сотрудника с точки зрения его продвижения по службе. Иногда возникает необходимость найти компромиссное решение между отбором людей с наибольшей производительностью на данный момент и кандидатов с наивысшими возможностями в будущем.

Традиционный подход к комплектованию штата предусматривает прогнозирование категорий сотрудников с помощью корреляционного анализа. Для этого с помощью коэффициентов корреляции строится модель, непосредственно связывающая независимые переменные (т. е. различные меры индивидуальных различий) с так называемыми критериями (т. е. различными оценками последствий, или «успешных исходов» для организации) [26]. После выбора и проверки модели становится возможной классификация претендентов с точки зрения эффективности их работы на потенциально перспективных и потенциально бесперспективных на основе множества независимых переменных, составленного для каждого претендента. Линейная комбинация множества из двух и более переменных может быть получена с помощью коэффициентов множественной корреляции, при этом используются как линейные, так и нелинейные методы оценки.

Исский [51] применил этот подход к исследованию проблемы текучести кадров и обнаружил, что пребывание сотрудника в той или иной должности определяется его различными индивидуальными показателями. Кроме того, он установил, что продолжительность пребывания в определенной должности существенным образом зависит от таких показателей, как обеспеченность жильем, накопленный опыт, возраст в момент поступления на работу, уровень образования (в виде обратной зависимости).

При использовании такого рода корреляционного анализа следует иметь в виду, что, во-первых, корреляция сама по себе не объясняет и не устанавливает причинно-следственных связей. Поэтому необходимо теоретически или экспериментально обосновать выбор независимых переменных, характеризующих реальную ситуацию. Во-вторых, корреляционный анализ предполагает наличие и обработку случайной выборки, что требует тщательного исследования имеющихся данных с целью выявления возможных систематических ошибок. Довольно часто процесс отбора претендентов зависит как от метода отбора, так и от ряда социальных и культурных факторов. В-третьих, при принятии решения относительно отбора будущих сотрудников из множества претендентов следует тщательно проверять справедливость принятых предположений о структуре организации. Это особенно важно в периоды быстрых технологических, организационных или прочих изменений. И наконец, в-четвертых, следует всегда проверять предпо-

ложение о нулевых коэффициентах корреляции, которое, кстати, не всегда гарантирует равенство нулю «реальных» коэффициентов.

Оценка некоторых независимых переменных, используемых в корреляционных моделях, производится с помощью психологического тестирования. Многие организации все чаще прибегают к такого рода тестированию, и некоторые из них получили вполне удовлетворительные результаты. Тем не менее большинство тестов для отбора будущих сотрудников из множества претендентов является вспомогательным средством для организации, а не для претендента, особенно если последнему отказано в приеме [81].

Теоретически отбор будущих сотрудников возможно осуществить с помощью теории принятия решений применительно к индивидуальному выбору. В этом случае оценивается способность некоторого множества кандидатов выполнить конкретную работу, причем оценка производится с точки зрения некоторой условной эффективности каждого претендента. Кроме того, для каждого претендента оценивается вероятность успешного выполнения данной конкретной работы. Следовательно, стратегия отбора основана на выборе претендентов, способных с наибольшей эффективностью выполнить требуемую работу. Наиболее перспективной моделью теории принятия решений оказалась модель, основанная на использовании денежного эквивалента для оценки вклада сотрудника в успех организации [33—35, 57]. Так, например, в специализированных организациях сферы обслуживания оценка вклада сотрудника в денежном выражении может быть представлена в виде отношения ожидаемого его вклада к доходу предприятия (прибыль от реализации услуг, снижение заработной платы и др. затраты).

Существуют два показателя, характеризующие вклад каждого сотрудника в успех работы организации, — это «ожидаемая» условная полезность и «ожидаемая» реализуемая полезность [35]. Первая из них представляет собой потенциальную оценку вклада (дохода), который может внести в успех организации каждый сотрудник при условии, что он остается сотрудником данной организации в течение определенного срока. Ожидаемая реализуемая полезность — это вклад, который фактически может внести сотрудник в успех организации с учетом вероятности его ухода из организации.

В настоящее время многие руководители, ответственные за набор сотрудников, как правило, при принятии решений по отбору будущих сотрудников из множества претендентов используют в качестве критерия ожидаемую условную, а не ожидаемую реализуемую полезность, т. е. они отбирают людей, которые, как им кажется, обеспечивают максимум условной полезности для организации. Такой подход целесообразен в тех случаях, когда у претендента почти нет возможности получить другую работу.

Если же такая возможность существует и для ухода сотрудника из данной организации отсутствуют серьезные препятствия юридического, морального и прочего характера, то вероятность того, что он будет пытаться реализовать такую возможность, достаточно велика. Поэтому эту вероятность следует включать в оценку ожидаемого вклада сотрудника в успех работы организации, и, следовательно, в данном случае более подходящей оценкой вклада является ожидаемая реализуемая полезность.

Рассмотрим случай, когда руководителю необходимо сделать выбор из двух кандидатов на данную должность. Кандидат А имеет ожидаемую условную полезность в 30 000, а кандидат В — 25 000 долл. Используя в качестве критерия максимум ожидаемой условной полезности, лицу, принимающему решение, разумно было бы выбрать кандидата А. Однако если предположить, что вероятности ухода из организации кандидатов А и В соответственно равны 0,3 и 0,4, то лицу, принимающему решение, должно выбрать кандидата В, поскольку значения ожидаемой реализуемой полезности для кандидатов А и В соответственно равны 21 000 и 22 500 долл. Следует иметь в виду, что подобный случай является слишком упрощенным, и в реальных ситуациях вероятность ухода кандидата из организации должна оцениваться на каждой стадии случайного процесса. Использование ожидаемой реализуемой полезности обеспечивает более точное моделирование реального процесса, чем просто учет такого показателя, как текучесть кадров. Тем не менее приведенный пример демонстрирует потенциальную чувствительность решений относительно отбора сотрудников из множества претендентов к оценкам вероятности ухода сотрудников из организации или выполнения ими требуемой работы.

2. ПОВЫШЕНИЕ КВАЛИФИКАЦИИ ПЕРСОНАЛА

Для решения проблемы повышения квалификации персонала важно выяснить, следует ли повышать общий образовательный уровень и профессиональную подготовку (включая обучение на рабочих местах) сотрудников, и если следует, то какую, когда и в течение какого периода. До сих пор решению этих задач с помощью методов исследования операций не уделялось должного внимания.

Авторы работы [3] разработали схему моделирования этой задачи, используя потоковые модели динамического программирования. Простейшим вариантом такой схемы является многопериодная модель программы обучения с минимальными затратами, периодом планирования, состоящим из N временных интервалов равной длительности, и двумя уровнями обучения (программа обучения может быть завершена в течение одного временного ин-

тервала). В этом случае задача формулируется следующим образом:

найти

$$\min \sum_{t=1}^N \theta^{t-1} [C_t(X_t) + g_{1,t}(S_{1,t+1}) + g_{2,t}(S_{2,t+1})]$$

при условиях

$$S_{1,t+1} = S_{1,t} + (P_t - X_t) \div \lambda(X_t) - d_{1,t}, \quad t = 1, 2, \dots, N$$

(условие сохранения численности группы сотрудников 1-го уровня) и

$$S_{2,t+1} = S_{2,t} \div [X_t - \lambda(X_t)] - d_{2,t}, \quad t = 1, 2, \dots, N$$

(условие сохранения численности группы сотрудников 2-го уровня)

и при ограничениях

$$0 \leq X_t \leq P_t, \quad t = 1, 2, \dots, N,$$

т. е. число обучаемых людей не должно превышать общей численности пригодной для этого группы населения.

Здесь X_t — число сотрудников, допущенных для обучения в течение периода t , и лиц, закончивших программу обучения 1-го уровня и переместившихся на 2-й уровень; P_t — численность группы сотрудников, пригодных для обучения (задается); $S_{1,t}$ — численность группы сотрудников, соответствующих 1-му уровню в начале периода t ; $S_{2,t}$ — численность группы сотрудников 2-го уровня в начале периода t ; $\lambda(X_t)$ — число сотрудников, исключенных или прекративших обучение, как функция числа сотрудников, допущенных к обучению в периоде t ; $d_{1,t}$ — внешний спрос или потребность в сотрудниках 1-го уровня, обучающихся в течение периода t ; $d_{2,t}$ — внешний спрос или потребность в сотрудниках 2-го уровня, обучающихся в течение периода t ; θ — учетная ставка (учетный процент); $C_t(X_t)$ — издержка на обучение; $g_{1,t}(S_{1,t+1})$ — убытки, обусловленные недостатком или избытком сотрудников 1-го уровня; $g_{2,t}(S_{2,t+1})$ — убытки, обусловленные недостатком или избытком сотрудников 2-го уровня.

Решение данной задачи находится с помощью методов динамического программирования в предположении, что целевая функция является кусочно-линейной и разрывной. Рассматриваемый подход может быть распространен на случай многоступенчатой модели обучения и позволяет учесть такие факторы, как изменение продолжительности обучения, введение обязательных уровней обучения и 4-уровневых систем обучения (типа колледжа). Для получения решения при использовании подобной модели и общей

многоступенчатой модели применяются методы линейного программирования.

Рассматриваемый подход обладает рядом недостатков, которые обусловлены тем, что некоторые из используемых переменных приняты в ограничениях как заданные. Например, численность группы сотрудников, подходящих для обучения, P_i является функцией решений по приему и отбору и, таким образом, отчасти может быть управляемой переменной для лица, принимающего решение. Кроме того, потребность в кадрах d на разных уровнях может быть в большой степени результатом решений, принятых организацией, а не является объективной реальностью.

Организации заинтересованы в общеобразовательных и профессиональных программах обучения, повышения квалификации персонала с целью увеличения его производительности, расширения диапазона выполняемых работ, перемещения сотрудников на новую работу, продвижения по службе или улучшения морального и материального благополучия. Поэтому соответствующие модели принятия решения по развитию организации должны включать некоторое множество программ обучения и стажировки с учетом стоимости их реализации и получаемой прямой отдачи, потерь от сокращения производства (или некоторых нереализованных возможностей) и в конечном итоге с учетом выгод, получаемых от реализации каждой из программ. К сожалению, получение информации о выгодах вызывает наибольшие трудности. В настоящее время еще мало накоплено знаний о степени влияния общеобразовательных программ и программ профессионального обучения на людей и их производительность.

3. РАСПРЕДЕЛЕНИЕ ТРУДОВЫХ РЕСУРСОВ

Основной задачей распределения являются расстановка кадров и назначение на должность, обеспечивающие выполнение требуемых видов работ. Задачи распределения являются «зеркальным отображением» задач использования. Если в задачах распределения ресурсов структура работ считается заданной и требуется назначить их исполнителей, то в задачах использования ресурсов заданными являются исполнители, их возможности и индивидуальные характеристики и требуется определить структуру работ, которая позволила бы наилучшим образом использовать эти возможности. В большинстве случаев при управлении трудовыми ресурсами решаются одновременно обе задачи.

Задача распределения трудовых ресурсов, для решения которой часто использовались методы исследования операций, решается в организации на двух уровнях управления. На верхнем уровне решается задача агрегированного планирования трудовых ресурсов, целью которой является определение численности групп

сотрудников, необходимых каждому подразделению организации; часто эта численность выражается через потребность в человеко-часах и затратах на рабочую силу. Вторым уровнем, на котором решается задача распределения ресурсов, является уровень непосредственного управления. Результатом решения задачи распределения на этом уровне является фактическое назначение каждого конкретного исполнителя на конкретную работу, т. е. решение задачи, которая иногда называется задачей оптимального назначения, дает ответ на вопрос, как назначить n исполнителей ($I_1, \dots, I_n$) на n работ ($J_1, \dots, J_n$), так чтобы максимизировать общую эффективность такого назначения.

Анализ задач распределения трудовых ресурсов осуществляется с помощью моделей транспортного типа, или моделей назначения, программно-целевых моделей, моделей последовательного назначения и имитационных моделей.

3.1. Модели транспортного типа

Основная модель транспортного типа, или модель назначения, имеет следующее математическое выражение:

найти максимум целевой функции

$$z = \sum_{i=1}^n \sum_{j=1}^n r_{ij} X_{ij}$$

при условиях

$$\sum_{i=1}^n X_{ij} = 1, \quad \sum_{j=1}^n X_{ij} = 1,$$

где $X_{ij} \in \{0, 1\}$ для всех i, j ; $X_{ij} = 1$ означает, что I_i назначен на работу J_j , а r_{ij} есть относительная эффективность каждого исполнителя с точки зрения выполнения им каждого вида работ.

Такое математическое представление модели требует, чтобы каждый исполнитель был назначен на одну и только на одну работу и соответственно на каждую работу должен быть назначен один и только один исполнитель. Сформулированная задача может быть решена с помощью типового алгоритма решения задачи о назначениях или задачи транспортного типа [44].

Для специалиста, использующего эту модель в реальных ситуациях, наиболее трудным моментом является определение соответствующей целевой функции и, следовательно, соответствующей схемы получения оценок величины r_{ij} . При выборе целевой функции возможно несколько подходов:

1. Максимизация суммы оценок назначений, т. е. требуется найти максимум целевой функции

$$z = \sum_{i=1}^n \sum_{j=1}^n r_{ij} X_{ij}.$$

В этом случае величина r_{ij} должна непосредственно выражаться через такие важные для организации показатели, как время выполнения работы, издержки производства, объем выпуска в единицу времени и т. д. Однако довольно часто трудно осуществить это, так как каждое значение r_{ij} есть точечная оценка, при получении которой существенную роль играет фактор распределения ошибок [56]. Более того, значение r_{ij} представляет собой прогнозируемую оценку показателей выполнения j -й работы i -м исполнителем. Таким образом, данный подход основан на прогнозировании данных и не лишен недостатков и неопределенностей. Это может привести к тому, что результаты фактической работы организации при использовании назначений, полученных с помощью модели, могут сильно отличаться от тех, которые должны были бы быть при значениях r_{ij} , принятых в модели.

2. Максимизация вероятности успешного выполнения каждым исполнителем работы, на которую он назначен, т. е. требуется найти максимум целевой функции

$$z = \sum_{i=1}^n \sum_{j=1}^n \log (P_{ij}) X_{ij},$$

P_{ij} есть вероятность того, что исполнитель I_i успешно выполнит работу J_j [56]. Оценки вероятности того, что исполнитель с определенными характеристиками удовлетворительно выполнит каждую из работ, могут быть определены на основе статистических данных. Максимизация вероятности успешного назначения в целом на все работы осуществляется с помощью такого выбора исполнителей для каждой из работ, при котором достигается максимум произведения вероятностей успешного выполнения ими работы. Благодаря свойствам логарифмов целевая функция может быть выражена с помощью приведенной выше формулы. После введения понятия вероятность успеха или неудачи становится возможным рассмотрение других методов оценки значений r_{ij} и подходов к выбору целевой функции. Так, Вейл [97] предложил следующие дополнительные подходы к выбору целевой функции.

3. Минимизация вероятности неуспешного в целом назначения на все виды работ, т. е. требуется найти минимум целевой функции

$$z = \sum_{i=1}^n \sum_{j=1}^n \log (1 - P_{ij}) X_{ij}.$$

4. Максимизация ожидаемого количества успешных назначений, т. е. требуется найти максимум целевой функции

$$z = \sum_{i=1}^n \sum_{j=1}^n P_{ij} X_{ij}.$$

5. Минимизация ожидаемого количества неуспешных назначений, т. е. требуется найти минимум целевой функции

$$z = \sum_{i=1}^n \sum_{j=1}^n (1 - P_{ij}) X_{ij}.$$

6. Максимизация средних (геометрических) значений вероятностей успеха $(P/1 - P)$ для всех назначений, т. е. требуется найти максимум целевой функции

$$z = \sum_{i=1}^n \sum_{j=1}^n \log (P_{ij}/1 - P_{ij}) X_{ij}, \quad 0 < P_{ij} < 1.$$

7. Минимизация средних (геометрических) значений вероятностей неуспеха $[(1 - P)/P]$ для всех назначений, т. е. требуется найти минимум целевой функции

$$z = \sum_{i=1}^n \sum_{j=1}^n \log [(1 - P_{ij})/P_{ij}] X_{ij}, \quad 0 < P_{ij} < 1.$$

Последние два способа являются общепринятыми для решения задач логистики. По мнению Вейла [97], следует использовать не подход, предусматривающий экстремизацию средних значений вероятностей, а подходы 6 и 7, так как они позволяют получить более адекватное распределение вероятностей успеха.

Кроме того, они позволяют осуществить назначения, обеспечивающие максимальное приближение вероятности успеха к единице, в то время как сумма вероятностей дает величину, большую единицы. Однако при этом следует особенно тщательно производить оценку значений r_{ij} , для больших ($> 0,95$) и малых значений ($< 0,05$) вероятностей.

3.2. Модели целевого программирования

Чарнес, Кулер и др. [12] считают, что оценку вклада исполнителей в успех организации следует производить по составляющим этого вклада, а общую оценку, по-видимому, можно осуществить, используя основные представления целевого программирования и вводя следующие определения: r_{ij} — число j -х показателей, необходимых для выполнения i -й работы; a_{sj} — общее число j -х показателей, характеризующих s -го исполнителя; V_{ij} — недостающее число j -х показателей для выполнения i -й работы (слабая

переменная); V_{ij}^+ — избыточное число j -х показателей при выполнении i -й работы (слабая переменная); n — число индивидуумов и работ; m — число принимаемых во внимание показателей, характеризующих исполнителей.

Модель имеет следующее математическое выражение:

найти минимум целевой функции

$$Z = \sum_{i=1}^n \sum_{j=1}^m V_{ij}$$

при условиях

$$\sum_{s=1}^n X_{is} a_{sj} = V_{ij}^+ + V_{ij}^- + r_{ij}, \quad j = 1, 2, \dots, m,$$

$$\sum_{i=1}^n X_{is} = 1, \quad \sum_{s=1}^n X_{is} = 1,$$

где $X_{is} \in \{0, 1\}$, $V_{ij}^-, V_{ij}^+ \geq 0$.

Все члены V_{ij}^+ имеют весовые множители, равные нулю, а все члены V_{ij}^- — единице. В модели минимизируется недостающее суммарное число всех показателей, необходимых для выполнения всех работ.

Имеется несколько модификаций рассматриваемой модели, которые могут быть использованы в зависимости от ситуации. Согласно Чайдсеу и Волфу [16], при решении задач по управлению исследованиями и разработками и других задач управления каждой профессиональную группу необходимо разбить на подгруппы (категории) по квалификационному признаку и приписать каждой подгруппе весовой множитель, с помощью которого учитываются как потребности самой работы, так и ее важность для организации в целом. Авторы предполагают, что эффективность r_{ij} участия исполнителя в проекте равна сумме его относительных способностей (выполнять работу каждой подгруппы), умноженных на значимость соответствующей подгруппы для успешного завершения проекта и значимость проекта для системы в целом.

Для определения r_{ij} могут быть использованы также метод Дельфи и метод расчета трудовых ресурсов. Применение модели целевого программирования для отбора игроков в спортивной команде обсуждается в гл. II-8.

3.3. Модели последовательного назначения исполнителей

В приведенных выше моделях предполагалось, что каждый исполнитель должен выполнить только одну работу, каждая работа поручается только одному исполнителю и все работы распределяются одновременно. Однако на практике подобная ситуация

встречается крайне редко. Обычно имеются работы, выполнение которых еще никому не поручено, и назначение исполнителей в этом случае осуществляется постепенно, по мере поступления работ. Этот аспект задач назначения изучен относительно мало. Первые шаги в этом направлении были сделаны Дерманом, Либберманом и Россом [22]. Авторы предположили, что имеется n исполнителей для выполнения n работ, причем работы появляются в случайном порядке и исполнители этих работ назначаются также в случайном порядке. Подобная ситуация, в частности, характерна для бухгалтерских и юридических контор, служб быта, фирм, разрабатывающих математическое обеспечение ЭВМ, и ряда предприятий, на которых группы исполнителей формируются по мере необходимости (например, при появлении заказов на выполнение проектов и др.). После того как определенному исполнителю поручается выполнение той или иной работы, его кандидатура не рассматривается при последующих возможных назначениях.

Далее, за выполнение каждой j -й работы устанавливается денежное или какое-то иное вознаграждение X_j , имеющее ценность x_j , и для каждого исполнителя i производится оценка вероятности p_i выполнения им работы. При этом предполагается, что эта величина является характеристикой исполнителя и не зависит от характера j -й работы, которую он выполняет, и, кроме того, $0 \leq p_i \leq 1$. Это означает, что если, например, оценка вероятности успешного ведения дел адвокатом равна 1,0, то он всегда выигрывает дело, при оценке 0,0 он всегда проигрывает его, а при оценке 0,5 вероятности выиграть или проиграть дело равны.

При условии $0 \leq p_i \leq 1$ организация должна стремиться к максимизации общего ожидаемого вознаграждения (дохода), т. е. требуется найти максимум целевой функции

$$Z = E \left[\sum_{j=1}^n p_{ij} x_j \right],$$

где $p_{ij} = p_i$, если i -й исполнитель выполняет j -ю работу.

На основании теоремы Харди [41] авторы показали, что если известны функция совместного распределения X , а также число работ и людей n , которые еще должны получить назначение, и произведено упорядочение индивидуумов в соответствии с величинами $p_1 \leq p_2 \leq \dots \leq p_n$, то оптимальным следует считать такое назначение, когда исполнителю с соответствующей ему оценкой p_i поручается выполнение работы, которая приносит доход X , причем значение этого дохода должно находиться в i -м интервале действительной числовой оси. Разбиение числовой оси на интервалы зависит от величины n и распределения X и не зависит от величины p_i . В сущности, мы располагаем аналитическим аппаратом для реализации во времени следующего принципа управле-

ния: исполнителю с большей оценкой вероятности успешного выполнения им работы следует поручать работу с наивысшей оплатой, а исполнителю с более низкой оценкой — работу с наименьшей оплатой.

В ряде работ, посвященных задачам о назначениях, рассматриваются модели, в которых предполагается, что отдача исполнителя, получающего назначение на работу, не зависит от назначений, осуществленных к данному моменту. Это предположение если и выполняется, то в очень редких случаях. Действительно, большинство должностей и работ формируется непосредственно на основе структуры организации и принятой технологии и определяется информационными, энергетическими и материальными потоками, необходимыми для производства продукции и оказания услуг. Результаты социально-технологических исследований, проведенных в ряде областей промышленности, свидетельствуют о существовании явной зависимости между различными работами и задачами системы [29]. Такая зависимость является результатом неконтролируемых отклонений, которые могут иметь место на любом этапе технологического процесса (или участке сборочной линии) и передаваться на последующие этапы. Так, например, если выполнение какой-либо отдельной операции, входящей в заданный комплекс работ, поручается неквалифицированному исполнителю, то его низкая квалификация скажется не только на операции, которую он непосредственно выполняет, но и на последующих операциях. Для определения этих зависимостей используется метод «вариационной матрицы», который успешно применялся при исследовании задач управления динамическими процессами [29].

Вариационные матрицы формируются путем последовательной идентификации этапов процесса, эффективности и показателей функционирования системы, существующих для выполнения каждого этапа. Затем формируется матрица предшествующих операций, элементы которой характеризуют степень влияния отклонения показателя отдельного этапа процесса на показатели последующих его этапов. Цель построения такой матрицы состоит в выявлении в системе точек, в которых могут иметь место неконтролируемые отклонения, так как такие отклонения должны быть учтены технологией процесса, структурой организации или исполнителями последующих этапов.

3.4. Имитационные модели

Во многих случаях применение аналитических методов для решения задачи управления трудовыми ресурсами оказывается невозможным ввиду необходимости учитывать сложные взаимодействия многих переменных, влияющих на конечный результат.

В подобных ситуациях, по-видимому, целесообразно использовать имитационные модели. Нельсон [67], например, построил и провел эксперимент с общей моделью, которая позволяет осуществлять испытания систем с ограниченными трудовыми и машинными ресурсами путем варьирования таких показателей, как дисциплина обслуживания очереди, процедура назначения рабочих, количество рабочих, степень централизации управления назначением рабочих, показатели последовательности выполнения работ на сети, а также реакции каждого исполнителя при использовании им оборудования или выполнении определенного вида работ.

Хотя целью этих экспериментов являлась проверка структурных связей в системе и интерпретация действующих эвристических стратегий, а не моделирование данной реальной системы, тем не менее их результаты позволяют оценить полезность правил распределения трудовых ресурсов.

4. ИСПОЛЬЗОВАНИЕ ТРУДОВЫХ РЕСУРСОВ

Задача использования трудовых ресурсов может быть сформулирована следующим образом: организация должна выполнить определенный набор работ. Для этого необходимо так скоординировать эти работы и найти такое их выполнение, чтобы распределение работ и должностей было эффективным с точки зрения организации и удовлетворительным с точки зрения каждого исполнителя. В такой формулировке задача использования аналогична задаче нахождения оптимальной структуры организации. Поскольку подробное рассмотрение этой задачи в пределах одной главы не представляется возможным, ограничимся некоторыми замечаниями относительно возможных путей ее решения.

Можно выделить три основных подхода к исследованию проблемы использования трудовых ресурсов. Один из них, предложенный и разработанный Ф. Тейлором, известен как подход «паучьего управления». В основе этого подхода лежит следующее положение: «Каждый исполнитель должен учиться отказываться от своих собственных методов работы, приспособляясь к новым порядкам и требованиям, и приучать себя соответствующим образом воспринимать приказы по выполнению работ и подчиняться им, даже если они затрагивают такие аспекты, выполнение которых раньше являлось прерогативой исполнителя» [85]. Таким образом, подход «паучьего управления» предусматривает нахождение «строгих правил» выполнения каждого движения каждым исполнителем и разработку способов совершенствования и стандартизации всех орудий и условий труда. Это в свою очередь требует изучения операций, времени их выполнения и детального моделирования процесса работы (см. «теорию Х» Мак-Грегори [63]).

Согласно другому подходу, функционирование организации

зависит от психологического состояния исполнителей и взаимоотношений между ними, т. е. человеческие отношения в трудовом процессе определяются психологическими, социальными и организационными факторами, а не физическими, физиологическими и экологическими. Это означает, что такие показатели, как удовлетворенность работой, возможность получения данной заработной платы с помощью работы, возможность самовыражения и другие, должны учитываться при формировании целей организации (см. «теорию Y» Мак-Грегори).

И наконец, социально-технический подход, по существу объединяющий наиболее значительные достижения первых двух подходов, рассматривает организацию как открытую систему, которая имеет сильные связи с внешней средой и для которой технология играет важную роль.

Анализ социально-технической системы начинается с предположения, что для выполнения заданного набора задач существует более чем одна структура связей работ. Каждая такая структура имеет различные социальные и психологические показатели и различную технологию выполнения работ. Эти показатели совместно влияют на функционирование системы в целом. Например, Трист и др. [87, 88] в своих оригинальных исследованиях разработок крупных месторождений каменного угля выявили целый ряд структур, которые могут быть классифицированы, с одной стороны, в зависимости от структуры сложных комплексов работ, каждая из которых является простой, а с другой стороны — в зависимости от структур простых комплексов, состоящих из сложных работ. В структурах последнего типа за каждым исполнителем организации в течение нескольких календарных периодов закреплено много различных групп задач из множества основных задач без жесткого распределения труда по выполнению каждой из них. Такая «составная» структура связей работ была признана наилучшей в исследованной ситуации с точки зрения продуктивности и издержек производства, так как она обеспечивала возможность на профессиональном уровне преодолевать трудности добычи полезных ископаемых, связанные с изменением условий добычи, кроме того, способствовала снижению психологического напряжения, уменьшению количества психических заболеваний на работе вследствие несчастных случаев, заболеваний или без уважительных причин. Аналогичные результаты были получены Райсом [73] при исследовании текстильного производства в Ахмадабаде (Индия). Принимая во внимание результаты Райса, можно предположить, что и анализ управленческого персонала даст в общем такие же результаты.

В основе социально-технического подхода лежит положение, которое математически может быть представлено следующим образом: $O = T(I, E)$, где O — выход системы, I — входы системы,

Е — влияние внешней среды на систему и T — преобразование, характеризующее определенную структуру связей между отдельными операциями или задачами.

В простейшей постановке задача использования трудовых ресурсов состоит в нахождении наилучшей структуры преобразования T . Решение задачи использования предусматривает выявление всей задачи организации, для того чтобы сформировать работы и соответствующие им обязанности (обязанности представляют собой набор правил и ожидаемых выгод, которыми руководствуются сотрудники организации при выполнении работы).

Предложенный подход включает рассмотрение окружающей среды для определения степени ее влияния на систему, постановку задачи и отдельных операций, необходимых для достижения требуемого выхода; определение причин основных изменений или отклонений от нормы, которые могут появиться в системе, определение с помощью вариационных матриц, их взаимосвязей, а также анализ социальных систем, метод рассмотрения отклонений (включая понимание сотрудниками своих обязанностей) и систематического наблюдения за воздействием этих факторов на систему в целом [90].

Как уже отмечалось, функция реакции организации характеризует ее структуру и зависит от технологических требований в системе. Основные переменные функции реакции, которые наименее подвержены влиянию внешней среды, составляют «технологическую основу» организации [86].

Анализ технологической основы можно осуществить с помощью двух методов: синтеза («снизу вверх») и анализа («сверху вниз»). Используя первый из них в рамках методов обработки списков с применением предложенного Мак-Карти языка ЛИСР, Джервин [36] подтвердил уже существующие структуры (в частности, он повторил исследование крыла бомбардировщика, проведенное Томпсоном [86]) и при заданном множестве ограничений и критериев определил структуру промышленного предприятия. Согласно методу оценки списка, структура работ представляется в виде ориентированного графа, вершинами которого являются элементарные задачи, образующие основной список, а дуги соответствуют направлению потока работ. При этом определяется взаимозависимость задач (т. е. выясняется, зависят ли задачи одна от другой, или одна из них является продолжением другой, или они образуют непрерывную группу) и сложность задач. Затем с помощью некоторого эвристического метода задачи разбиваются на группы по степени связности, типу взаимозависимости (задачи с обоюдно взаимозависимостью имеют наивысшую оценку значимости) и степени сложности. Получающаяся в результате структура списка интерпретируется как структура системы (или организации).

В основе метода оценки списка лежит принцип декомпозиции [4, 21], который, по-видимому, может быть использован для проектирования структуры организации с помощью модели линейного программирования. Иными словами, если функции организации и все решаемые ею задачи можно описать с помощью модели линейного программирования большой размерности, то, очевидно, элементы матрицы ограничений могут дать ключ к формированию кадров. В той мере, в какой данная задача допускает разбиение на ряд почти независимых подзадач, принцип декомпозиции может быть использован как основа для построения модели децентрализованной структуры управления, что в свою очередь привело бы к необходимости закрепления исполнителей и задач за подразделениями организации. Подобная модель допускает и другую интерпретацию. Например, скрытую (неявную) цену, используемую для связи между задачей управления и ее подзадачами, можно рассматривать как отпускную цену системы.

5. ОЦЕНКА ЭФФЕКТИВНОСТИ РАБОТЫ ПЕРСОНАЛА

Оценка эффективности — это процесс измерения вклада (прошлого и будущего) каждого исполнителя в успех организации. Оценка эффективности может быть произведена руководителем и использована в качестве исходных данных для принятия решений по компенсации вклада исполнителя в успех организации, его назначения и продвижения по службе.

Одним из основных методов оценки эффективности персонала является определение значимости (полезности) показателей исполнителя в соответствии с выбранной шкалой оценок личных качеств и профессионального мастерства.

Другим методом являются психометрические исследования, в частности психологическое тестирование. В этом случае обоснованность полученных оценок определяется путем сравнения результатов тестирования «известной» группы людей с некоторыми показателями успешного функционирования организации, например с размером оклада, определенным распределением исполнителей и работ, мнением руководителя и т. д. Результаты тестирования используются при отборе будущих сотрудников из числа кандидатов, а также при необходимости продвижения сотрудника по службе.

Методология применения теории исследований операций для решения задачи оценки эффективности до сих пор не получила должного развития. В этой связи особого внимания заслуживает работа Фламгольца [33], в которой автор попытался отчасти восполнить этот пробел.

Фламголец предложил стохастическую модель для оценки эффективности, в основе которой лежит представление о стохастичес-

ческом процессе с вознаграждением [47]. Согласно этому представлению, перемещение исполнителей с одного рабочего места на другое в пределах организации можно рассматривать как стохастический процесс, в котором функцию вознаграждения выполняют доходы, получаемые организацией в заданные периоды времени при определенном распределении рабочих мест. Стохастическая модель с вознаграждением учитывает два вида полезности исполнителя для организации — условную $M(CU)$ и реализуемую $M(RU)$ полезность. Оценка этих полезностей включает следующие этапы:

— определение набора m всех взаимоисключающих должностей i , которые может занимать в организации некоторый исполнитель;

— оценку полезности u_i каждой должности для организации;

— оценку ожидаемого времени пребывания каждого исполнителя в данной должности;

— вероятностную оценку $P(u_i, t)$ того, что исполнитель будет занимать каждую из возможных должностей (рабочих мест) в настоящий момент и в будущем (как уже отмечалось, для прогнозирования должностей, которые будут заняты в настоящий момент и в будущем, можно использовать стохастическую модель перемещения).

Используя введенные обозначения, можно записать выражение для условной полезности в виде

$$M(CU) = \sum_{i=1}^n \sum_{t=1}^{m-1} [u_i P(u_i, t) (1+r)^t],$$

где r — коэффициент дисконтирования. Для реализуемой полезности имеем

$$M(RU) = \sum_{i=1}^n \sum_{t=1}^m [u_i P(u_i, t) / (1+r)^t],$$

где суммирование по i проводится с учетом состояния m , соответствующего уходу сотрудника из организации. (Рассмотренная модель используется фирмой Lester Witte and Co. при решении целого ряда задач по оценке эффективности персонала [34].)

6. ОПЛАТА РАБОТЫ И ПОощРЕНИЕ ПЕРСОНАЛА

Решение проблем, связанных с оплатой работы и поощрением персонала, дает, по существу, ответ на вопрос, кто сколько получает и за что. При формулировке этой задачи необходимо иметь в виду, что, во-первых, для успешного функционирования каждая организация должна гарантировать своим сотрудникам активную и плодотворную деятельность и, во-вторых, организация

должна выполнять свои обязательства в части обеспечения соответствующего распределения зарплаты среди своих сотрудников. Для достижения этих целей могут быть использованы затраты, различного рода вознаграждения, штрафы, санкции.

Во многих организациях решение вопроса, кто сколько получает и за что, касается в основном заработной платы рабочих служащих, признания и продвижения по службе. Однако этот вопрос включает и такие психологические факторы, как участие в принятии решений, уровень притязаний на работу, желание учиться, способность сочетать производственную и общественную деятельность, и, естественно, стремление обеспечить свое будущее. Хотя в некоторых работах [20, 43, 59, 60] делается попытка учесть психологические и социальные факторы при решении задач оплаты и поощрения, тем не менее усилия большинства индустриально развитых стран направлены главным образом на разработку соответствующих систем оплаты, а методы исследования операций до сих пор используются в основном для определения денежного вознаграждения за выполненную работу.

В общем виде задача оплаты и поощрения может быть сформулирована как

$$\text{Оплата} = f(A, O, E),$$

где A — показатели, характеризующие исполнителей (например, производительность, профессия, возраст, трудовой стаж, образование, неиспользованные возможности и т. д.); O — переменные, характеризующие организацию с точки зрения общего уровня оплаты сравнимых работ, структуры оплаты в соответствии с номенклатурой должностей, зависимости оплаты от иерархии должностей, потребности организации в отдельных профессиях, общего уровня доходов и т. д.; E — внешние переменные, такие, как трудовое законодательство, правила найма, влияние профсоюзов и т. д. Следует иметь в виду, что при такой формулировке задачи оплаты и поощрения оплата является функцией доходов и множества O и E являются ограничениями, а не переменными.

Одним из методов решения сформулированной задачи является линейный регрессионный анализ, предусматривающий использование данных о предыстории организации. В этом случае размер оплаты или поощрения (в денежном или каком-либо ином выражении) за выполненную j -ю работу может быть представлен в виде

$$P_j = \sum_{i=1}^n w_{ij} a_{ij},$$

где a_{ij} — независимые переменные, или значимость (полезность) i -го показателя исполнителя j -й работы, либо значимость i -го показателя, необходимого исполнителю для выполнения j -й работы; w_{ij} — весовые множители i -го показателя при выполнении

j -й работы. (Если весовые множители одинаковы для всех работ, то индекс j опускается и имеем w_i . Оценку всех весовых множителей можно осуществить с помощью метода наименьших квадратов и методов множественной линейной регрессии [37].) В этом случае данные об исполнителе могут быть сгруппированы по таким показателям, как стаж работы, производительность, способность к руководящей работе, чувство ответственности и другим.

По мнению авторов работы [12], модель оптимального вознаграждения должна удовлетворять определенным требованиям. Прежде всего модель должна быть линейной и совместимой с существующей организационной основой. Иными словами, использование модели не должно нарушать существующей в организации иерархии должностей (например, курьер или опытный продавец не может получать большее вознаграждение, чем руководитель организации). Далее, учитывая возможность конкуренции, модель должна исключить потерю цепных, способных к конкуренции сотрудников, но без создания несогласованности внутри организации. Кроме того, модель не должна содержать отрицательных весовых множителей (например, группирование сотрудников преклонного возраста на низкооплачиваемых должностях не должно расцениваться как штрафование за возраст). И наконец, модель оптимального вознаграждения должна исключить предвзятость решений, связанную с существующим фондом заработной платы организации. Поскольку рассмотренная выше линейная регрессионная модель не удовлетворяет полностью всем перечисленным требованиям (в частности, последнему) и, кроме того, чрезвычайно трудно установить размеры вознаграждений и получить необходимые данные о других организациях (в особенности о конкурирующих), Чарнес и Купер [12] предложили новую формулировку задачи оплаты и поощрения:

найти значения w_i , минимизирующие целевую функцию

$$Z \equiv d_1^+ + d_L^- + \sum_{k=2}^{L-1} (d_k^+ + d_k^-),$$

при ограничениях

$$\sum_{i=1}^n w_i X_{i1} + d_1^+ = P_M,$$

$$\sum_{i=1}^n w_i X_{i1} - \sum_{i=1}^n w_i X_{iL} \geq 0,$$

.....

$$\sum_{i=1}^n w_i X_{i(L-1)} - \sum_{i=1}^n w_i X_{iL} \geq 0,$$

$$\sum_{i=1}^n w_i X_{iL} - d_L^- = P_m,$$

$$\sum_{i=1}^n w_i X_{ik} - d_k^+ - d_k^- = P_k, \quad k \in K,$$

где X_{ij} — значимость i -го показателя при выполнении j -й работы и $j = 1, 2, 3, \dots, L$, при этом работы располагаются в предполагаемом порядке их важности, т. е. $P_1 \geq P_2 \geq \dots \geq P_j \dots \geq P_L$. Здесь P_M , P_m и P_k — соответственно верхний, нижний и промежуточный уровни оплаты (предполагается, что организация должна стремиться к тому, чтобы оплата, насколько это возможно, соответствовала этим установленным уровням), d_k^+ и d_k^- — переменные, характеризующие положительное или отрицательное отклонение от некоторого установленного k -го уровня оплаты, принадлежащего к заданному множеству уровней K . Выбор множества K может осуществляться с учетом таких факторов, как промышленный опыт, конкуренция, правовые требования.

Применение рассматриваемого подхода к решению задачи оплаты и поощрения возможно только при условии, что работы, завершение которых необходимо для достижения организацией стоящих перед нею глобальных и локальных целей, хорошо определены и описаны на уровне, позволяющем организовать их выполнение. Если это условие выполняется, то можно приступить к анализу системы вознаграждений в организации, в процессе которого должны быть выполнены следующие операции.

— Определение показателей, которые должны входить в целевую функцию. Для этого прежде всего необходимо установить, какие показатели исполнителя влияют на степень вознаграждения за выполненную им работу.

— Разработка системы оценок значимости каждого показателя. Обычно оценка каждого показателя осуществляется в соответствии с установленной шкалой оценок (при использовании модели линейного программирования такая шкала должна иметь строго фиксированные предельные значения). Например, для такого показателя, как образование, можно ввести 5-балльную шкалу оценок: доктор философии — 5, магистр — 4, бакалавр — 3, выпускник специального учебного заведения или колледжа — 2, выпускник средней школы — 1.

Предварительные, достаточно объективные оценки часто могут быть даны и самим исследователем систем. Однако оценки более субъективных показателей, например степени ответственности или степени критического отношения к работе, могут потребовать коллективных усилий экспертов. На практике, как правило, число оценок колеблется от 5 до 6, хотя в некоторых случаях оно

может равняться 2, а в некоторых достигать 18. Чтобы преодолеть некоторые трудности, связанные с оценкой показателей исполнителя, определяют число оценок (например, показателю «knowledge» соответствует 14 оценок, т. е. $K_{14} > K_{13} > \dots > K_1$), и каждой оценке полезности показателя приписывают весовой множитель (или относительную значимость) [12]. В этом случае уравнение баланса по каждой из работ записывается в виде суммы взвешенных оценок полезности отдельных показателей.

— Формулировка задачи компенсации с учетом предполагаемого распределения показателей по их важности, иерархического характера организации (предполагается, что структура окладов внутри организации имеет явно выраженный иерархический характер) и других ограничений и требований, а также таких факторов, как конкуренция сотрудников, ценность сотрудника для организации и т. д.

— Получение решения задачи в виде перечня возможных вознаграждений. Если необходимо оценить большое количество работ, то для принятия решения можно использовать выборочный метод.

— Использование полученных результатов совместно с другими исходными данными для выбора системы вознаграждений.

Ограничения выбирают в соответствии с требованиями, согласно которым вклад каждого показателя в общий объем вознаграждения не должен превышать установленного процента, каждая оценка должна иметь положительное значение и должно быть достигнуто приемлемое минимальное значение разности между весовыми множителями двух соседних показателей (например, $K_2 - K_1 \geq 0,07$). Вознаграждение в формулировке задачи представлено не в денежном выражении, а в процентах от размера наивысшего оклада.

Чарпес и Куер успешно применили этот метод к выборке должностей административных и канцелярских работников крупной нефтяной компании и к выборке 21 должности руководящих работников предприятий обслуживания.

В работах [11, 64, 99] рассматривается возможность расширения области применения подхода, предложенного Чарпесом и Купером. Так, автор работы [11] использовал этот подход, чтобы получить шкалу ставок для школ округа, которая учитывала бы ресурсы при определении окончательной структуры вознаграждений, допускала бы гибкость в перекрестиях размера окладов между различными уровнями иерархии организации и отражала бы цели школ и значимость каждой из них для округа. Минимизируя издержки и максимизируя различные показатели, автор получил приемлемую шкалу ставок при каждом из принятых предположений. Такие расчетные шкалы заработной платы использовались в качестве исходных при определении тарифной сетки

долягистных окладов или при выработке коллективных договоров об условиях оплаты труда.

Интересная особенность работы [11] состоит в том, что автор использовал не одну, а целый ряд целевых функций.

7. МОДЕЛИ СЛОЖНЫХ СИСТЕМ УПРАВЛЕНИЯ ТРУДОВЫМИ РЕСУРСАМИ

В предыдущих разделах были рассмотрены примеры применения методов теории исследований операций для решения различных задач управления трудовыми ресурсами. Исследование систем управления, имеющих большую размерность, в целом можно осуществлять с помощью моделей. Один из возможных подходов включает разработку модели функционирования системы и использование имитационных моделей и ЭВМ для оценок эффективности применения различных методов и стратегий управления. Например, Вебер [96] разработал модель сложной системы планирования потребностей в трудовых ресурсах корпораций. Эта модель соответствует подсистеме трудовых ресурсов в организации с иерархической структурой и позволяет изучать поведение исполнителей, процесс принятия решений и влияние различных внешних по отношению к организации факторов на процесс управления (руководства). Для построения модели Вебер, аналогично Боинни [10], использовал теоретические данные, содержащиеся в научной литературе, а не результаты экспериментов или анализа систем. Модель, предложенная Вебером, схематически представлена на рис. 3, где стрелками показаны причинные связи между различными факторами, из которых основными являются показатели, характеризующие исполнителей, стратегии организации, состояние биржи труда.

Вебер использовал модель для исследования стратегии продвижения по службе и увеличения заработной платы (сотрудников со стажем и вновь принятых). При решении задачи компенсации Вебер изменял весовые множители показателей работ и значимости сотрудников. Он установил, что увеличение вознаграждения, пропорциональное объему выполняемой работы, ведет к возрастанию удовлетворенности работой и общему улучшению функционирования организации. Однако в противоположность широко распространенному мнению [48] об эффективности такого подхода последний приводит со временем к уменьшению желания исполнителей и администрации выполнять различные работы, а также к снижению эффективности работы всего персонала. Таким образом, если вознаграждение устанавливается в основном на основании объема выполняемой работы, то возникает проблема эф-

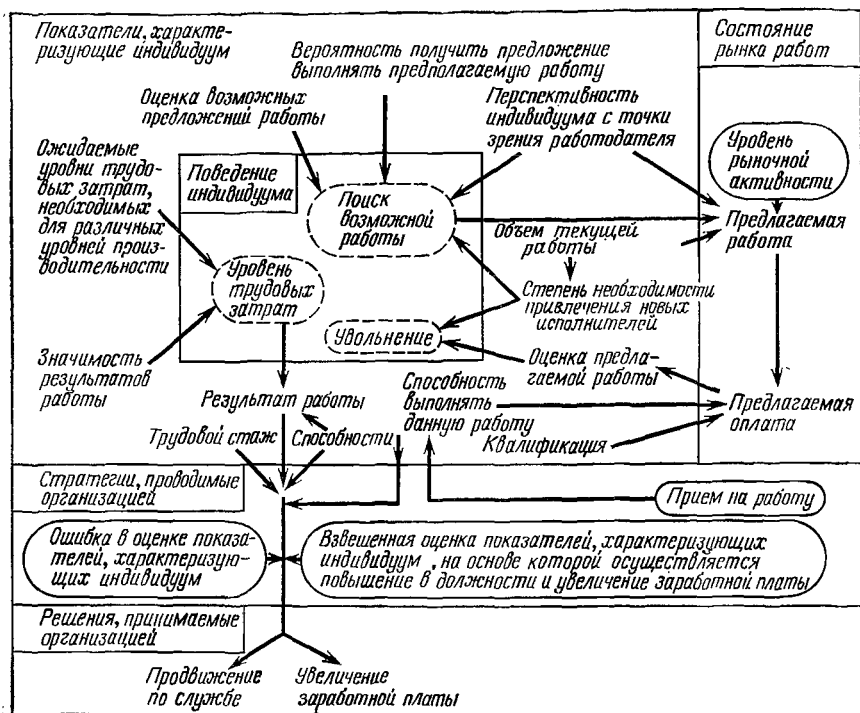


Рис. 3. Схематическое изображение модели, предложенной Вебером. (Последствия решений, принимаемых организацией, непосредственно сказываются на показателях, характеризующих индивидуум.)

фактивности работы всей организации в течение длительного периода времени.

Изучение стратегий найма осуществлялось путем сравнения двух вариантов приема: в одном случае прием осуществлялся так, что уровень индивидуальных способностей был в среднем выше, чем уровень способностей, требуемый для выполнения всех работ, а в другом случае уровень индивидуальных способностей соответствовал требованиям, предъявляемым конкретной работой. Как и предполагалось, в первом случае возрос средний уровень персонала в организации в целом, при этом, однако, снизилась удовлетворенность персонала работой и не улучшилось функционирование организации.

Таким образом, ошибки, связанные с оценкой показателей качества персонала, привели к снижению уровня удовлетворенности работой, высокой текучести кадров и ухудшению качества работы.

Влияние изменений внешних факторов на систему изучалось путем рассмотрения поведения сотрудников (поиск новых мест работы, вероятность ухода сотрудников из организации) в зависимости от колебаний спроса на результаты труда организации. Значительные колебания спроса отражаются главным образом на сотрудниках и руководящих работниках, имеющих высокую степень полезности и более высокую оценку показателей их работы. Издержки на заработную плату при этом снижаются; число уходов с работы и вакансий растет одновременно с ростом числа случаев поиска сотрудниками новых мест работы и снижением степени удовлетворенности своей работой.

Несмотря на то что подобных исследований выполнено не так уж много, тем не менее полученные результаты позволяют предположить, что имитационные модели такого типа могут оказаться полезными для понимания динамики больших систем и влияния на них различных стратегий управления, а также понимания смысла и последствий принимаемых предположений относительно поведения отдельных исполнителей и организации в целом, причем эти предположения могут быть сделаны на основе теоретических или экспериментальных данных, либо на основе установленных канонов, либо интуитивно.

ЗАКЛЮЧЕНИЕ

Проведенный выше обзор подходов к применению методов исследования операций для решения задач управления трудовыми ресурсами позволяет выявить определенные тенденции. Так, например, для принятия решения по комплектованию используются модели планирования, стохастические модели перемещения кадров в пределах организации, для принятия решения по распределению ресурсов — модели назначения; для принятия решений по оплате и премированию — модели линейного программирования. Что касается принятия решений по развитию организации, соответственно трудовым ресурсам и по оценке вклада каждого исполнителя в успех организации, то в этих случаях выбор моделей весьма ограничен. Наиболее вероятной причиной такого положения, по-видимому, является использование базовой потоковой модели. Дело в том, что такая модель оказывается эффективной только в тех случаях, когда имеется конечный список хорошо определенных взаимоисключающих рабочих мест или должностей, которые образуют иерархическую структуру. При этом структура является относительно устойчивой и правила перехода в ней с одного уровня на другой относительно постоянны во времени. Входные и выходные потоки базовой потоковой модели можно сравнить с потоками воды, втекающими в бассейн и вытекающими из него, если предположить, что потоки людей (исполнителей) однородны. Тогда потоки людей могут «сливаться в один бассейн», образуя однородные группы исполнителей, а переход исполнителя из одной однородной группы в другую представляется в виде потока воды, вытекающего из одного бассейна и втекающего в другой. Структура этих потоков и их скорость задаются с учетом существующей иерархической структуры организации.

Этот подход достаточно эффективен в хорошо упорядоченных системах, таких, как военные организации, учебные заведения и некоторые корпорации. По-видимому, он может быть использован и для принятия решений

по комплектованию, распределению трудовых ресурсов и сбалансированным оплате и поощрениям в тех случаях, когда структура работ нечувствительна к резким изменениям во внешней по отношению к системе среде. Некоторые решения по развитию могут быть приняты с помощью потоковой модели. Например, одно из возможных решений предполагает изменение схемы движения и скорости перемещения потока людей через систему. Такое решение, естественно, не учитывает достигнутый в результате получения новых знаний и опыта рост благосостояния отдельных индивидумов и всех сотрудников организации в целом. Однако этот вопрос требует специального рассмотрения.

Необходимо отметить, что в большинстве подходов целевые функции являются зависимыми переменными от показателей целевой эффективности организации, т. е. они, как правило, отражают интересы и цели организации, связанные с эффективностью. Однако помимо этих интересов организация (так же как и отдельные индивидумы) может иметь и другие интересы. К сожалению, мы располагаем недостаточной информацией относительно подходов к оценке иных интересов, чем те, которые связаны с эффективностью организации. Можно только с определенностью утверждать, что использование существующих методов для получения решения в таких многошаговых и многокритериальных моделях оказывается малоэффективным.

Расширение области применения методов исследования операций для решения задач оценки и использования трудовых ресурсов предусматривает более глубокую и детальную разработку концепции системы управления трудовыми ресурсами. В этой связи особую важность приобретают проводимые в настоящее время исследования методов расчета трудовых ресурсов, целью которых является доказательство необходимости применения методов исследования операций для решения задач оценки и использования трудовых ресурсов и создания основы для такого использования.

Специалистам в области теории руководства предстоит решить множество проблем, связанных с управлением трудовыми ресурсами, что, вероятно, с необходимостью потребует выхода за пределы существующих представлений, обусловленных формализмом потоковой модели и концепцией эффективности организации. Нам представляется, что ни один ученый, считающий себя специалистом в области научных методов руководства, не вправе уклониться от решения этих проблем.

ЛИТЕРАТУРА

1. Ackoff R., A Concept of Corporate Planning, Wiley, N.Y., 1970.
2. Arabayre R. P., Fearnly J., Steiger F. C., Teather W., The Airline Crew Scheduling Problem: A Survey, *Transport. Sci.*, 3, 2, 140—163 (May 1969).
3. Balinsky W., Reisman A., Some Manpower Planning Models Based on Levels of Educational Attainment, *Management Sci.*, 18, 12, B-691—B-705 (August 1972).
4. Baumol W. J., Fabian T., Decomposition, Pricing for Decentralization and External Economics, *Management Sci.*, 11, 1, 1—32 (September 1964).
5. Bartholomew D. J., Stochastic Models for Social Processes, Wiley, N.Y., 1967.
6. Bartholomew D. J., Decision-Making in Manpower Planning, *Operational Res. Quart.*, 19 (Session Report) (April 1968) 129—132.
7. Bartholomew D. J., A Mathematical Analysis of Structural Control in a Graded Manpower System. Berkeley, California: University of California Research Program in University Administration, Office of the Vice President-Planning and Analysis, Paper P-3 (1969).
8. Bergstrom G. L., Smith B. E., Multi-Item Production Planning — An Extension of the HMMS Rules, *Management Sci.*, 16, 10, 614—629 (June 1970).

9. Bishop A. B., Rockwell T. H., A Dynamic Programming Computational Procedure for Optimal Manpower Loading in a Large Aircraft Company, *Operations Res.*, 6, 835—848 (1958).
10. Bonini C. P., Simulation of Information and Decision Systems in the Firm, Prentice-Hall, Englewood Cliffs, N.J., 1963.
11. Bruno J. E., Compensation of School District Personnel, *Management Sci.*, 17, 10, B-569 — B-587 (June 1971).
12. Charnes A., Cooper W. W., Ferguson R. O., Optimal Estimation of Executive Compensation by Linear Programming, *Management Sci.*, 1. 2. 138—151 (January 1955).
13. Charnes A., Niebhaus R. J., A Goal Programming Model for Manpower Planning, Carnegie-Mellon University, G.S.I.A., Management Sciences Research Report, No. 115 (December 5, 1967).
14. Charnes A., Sholtz D., An Extended Goal Programming Model for Manpower Planning, Carnegie-Mellon University, G.S.I.A., Management Science Research Report, No. 156 (December 23, 1968).
15. Charnes A., Stedry A., Static and Dynamic Assignment Models with Multiple Objectives and Some Remarks of Organization Design, *Management Sci.*, 15, 8, B-365—B-375 (April 1969).
16. Childs M., Wolfe H., A Decision and Value Approach to Research Personnel Allocation, *Management Sci.*, 18, 6, B-269—B-278 (February 1972).
17. Conrath D. W., Hamilton W. F., The Economics of Manpower Pooling, *Management Sci.*, 18, 2, B-19—B-29 (October 1971).
18. Dalkey N. C., Delphi, Santa Monica, Cal: Rand Corporation, P-3704 (October 1967).
19. Davies G. S., Structural Control in a Graded Manpower System, *Management Sci.*, 20, 1, 76—84 (September 1973).
20. Davis L. E., Taylor J. C., Design of Jobs, Baltimore-Penguin Books, Inc., 1972.
21. Dantzig G. B., Wolfe P., Decomposition Principles for Linear Programs, *Operations Res.*, 8 (February 1960).
22. Derman C., Lieberman G. J., Ross S. M., A Sequential Stochastic Assignment Problem, *Management Sci.*, 18, 7, 349—355 (March 1972).
23. Dill W. E., Garver D. P., Weber W. L., Models and Modelling for Manpower Planning, *Management Sci.*, 13, 4, 142—167 (December 1966).
24. Drucker P. F., The Employee Society, *Am. Soc. Rev.*, 58, 358—363 (1952).
25. Duffett R. H. E., A Quantitative Approach to Company Manpower Planning, *Manpower and Appl. Psy.*, 3, 1 and 2, 11—22 (1969).
26. Dunette M. D., Personal Selection and Placement, Wadsworth Publishing Company, Inc. Belmont, Ca., 1966.
27. Emery F. E., The Democratization of the Workplace, *Manpower and Appl. Psychol.*, 1, 118—129 (1967).
28. Emery F. E., Trist E. L., Socio-Technical Systems, in Emery F. E. (Ed.), Systems Thinking, Penguin Harmondsworth, England, 284—296 (1969).
29. Engelstad P. H., Socio-Technical Approach to Problems of Process Control, in Davis and Taylor, *Design of Jobs*, 328—356 (1972).
30. Fan L. T., Wang C., The Discrete Maximum Principle, Wiley, N.Y., 1964.
31. Fan L. T., Wang C., The Continuous Maximum Principle, Wiley, N.Y., 1966.
32. Fischer F. M., Morton A. S., The Costs and Effectiveness of Reenlistment Incentives in the Navy, *Operations Res.*, 15, 3, 373—387 (May — June 1967).
33. Flamholtz E., A Model for Human Resource Valuation: A Stochastic Process with Service Rewards, *Account. Rev.*, 46, 253—267 (April 1971).
34. Flamholtz E., Human Resource Accounting, Dickenson Publishing Co., Encino, California, 1974, Ch. 6.
35. Flamholtz E., Toward a Theory of Human Resource Value in Formal Organizations, *Account. Rev.*, 47, 666—678 (October 1972).

36. Gerwin D., Cristoffel W., Organizational Structure and Technology: A Computer Model Approach. *Management Sci.*, 20, 12. 1531—1542 (Aug. 1974).
37. Gilmore R. W., Industrial Wage and Salary Control, Wiley. Inc., N.Y., 1956.
38. Glover W. S., Ackoff R. L., Five-Year Planning for an Integrated Operation, in Proceedings of the Conference on Case Studies in Operations Research, Case Institute of Technology, Cleveland. 38—47 (1936).
39. Gifford J. P., Psychometric Methods, Second Edition, McGraw-Hill, N.Y., 1954.
40. Hanssman F., Hess S. W., A Linear Programming Approach to Production and Employment Scheduling. *Management Tech.*, 1, 1. 46—51 (January 1960).
41. Hardy G. H., Littlewood J. E., Polya G., Inequalities, Cambridge University Press, 1934.
42. Helmer O., Social Technology, Basic Books, N.Y., 1968.
43. Herzberg F., One More Time: How do you Motivate Employees?, *Harvard Bus. Res.*, 46, 53—62 (1968).
44. Hillier F. S., Lieberman G. J., Introduction to Operations Research, Holden-Day, Inc., San Francisco, 1967.
45. Holstein W. K., Berry W. L., The Labor Assignment Decision: An Application of Work Flow Structure Information, *Management Sci.*, 18, 7. 390—400 (March 1972).
46. Holt C. C., Modigliani F., Muth T., Simon H. A., Planning Production Inventories and Work Force, Prentice-Hall, Englewood Cliffs, N.J., 1960.
47. Howard R., Dynamic Programming and Markov Processes, The M.I.T. Press, 1960.
48. Humbel J. W., Improving Management Performance, *Person. Management*, 47, 136—142 (March 1965).
49. Hwang C. L., Fan L. T., Application of the Maximum Principle to Industrial and Management Systems, *J. Indust. Eng.*, XVIII, 11, 589—593 (November 1966).
50. Hwang C. L., Optimum Production Planning by the Maximum Principle, *Management Sci.*, 13, 9, 751—755 (May 1967).
51. Inskip G. C., Statistically Guided Employee Selection: An Approach to the Labor Turnover Problem, *Person. J.*, 49, 1, 15—24 (January 1970).
52. Jewett R. F., A Minimum Risk Manpower Scheduling Technique, *Management Sci.*, 13, 10, B-578 — B-592 (June 1967).
53. Jones C. G., Parametric Production Planning, *Management Sci.*, 13, 11, 843—866 (July 1967).
54. Kao R. C., Rowan Th. C., A Model for Personnel Recruiting and Selection, *Management Sci.*, 5, 2, 192—203 (January 1959).
55. Karush W., Vazsonyi A., Mathematical Programming and Employment Scheduling, *Nev. Res. Log. Quart.*, 4, 297—320 (December 1957).
56. King W., A Stochastic Personnel-Assignment Model, *Operations Res.*, 13, 1, 67—81 (January - February 1967).
57. Lev B., Schwartz A., On the Use of the Economic Concept of Human Capital in Financial Statements, *Account. Rev.*, 46, 103—112 (January 1971).
58. Lippman St., Rolfe A. J., Wagner H. M., Yuan J. S. C., Optimal Production Scheduling and Employment Smoothing with Deterministic Demands, *Management Sci.*, 14, 3, 127—158 (November 1967).
59. Maslow A. H., Eupsychian Management: A Journal, Irwin-Dorsey, Homewood, Ill., 1965.
60. Maslow A. H., The Father Reaches of Human Nature, The Viking Press, N.Y., 1971.
61. Maslow A. H., Motivation and Personality, Harper & Row, N.Y., 1954.

62. McCloskey J. F., Handsman F., An Analysis of Stewardess Requirements and Scheduling for a Major Domestic Airline, *Nev. Res. Log. Quart.*, 4, 3, 183--202 (September 1957).
63. McGregor D., The Human Side of Enterprise, McGraw-Hill, N.Y., 1960.
64. Metzger R. W., Elementary Mathematical Programming, Wiley, N.Y., 1958.
65. Milkovich G. T., Annoni A. T., Mahoney Th. S., The Use of Delphi Procedures in Manpower Forecasting, *Management Sci.*, 19, 4, 381--388 (December 1972).
66. Nelson R. T., Labor Assignment as a Dynamic Control Problem, *Operations Res.*, 14, 3, 369--376 (May - June 1966).
67. Nelson R. T., Labor and Machine Limited Production Systems, *Management Sci.*, 13, 9, 648--671 (May 1967).
68. Melson R. T., A Simulation of Labor Efficiency and Centralized Assignment in a Production Model, *Management Sci.*, 17, 2, B-97 - B-106 (October 1970).
69. Nemhauser G. L., Nuttle H. L., A Quantitative Approach to Employment Planning, *Management Sci.*, 11, 8, 155--165 (June 1965).
70. Pappas I., Dynamic Job Assignment for Railway Personnel, *Management Sci.*, 13, 12, B-809 - B-816 (August 1967).
71. Rau J. C., A Model for Manpower Productivity During Organization Growth, *Nev. Res. Log. Quart.*, 18, 4, 543--559 (December 1971).
72. Rehms F. P., Wagner H. M., Applying Linear Programming to Your Pay Structure, *Bus. Horizons*, 6, 4, 89--98 (Winter 1963).
73. Rice A. R., Productivity and Social Organization: The Ahmedabad Experiment. Tavistock, London, 1958.
74. Rowe St. M., Wagner W. G., Weathersby G. B., A Control Theory Solution to Optimal Faculty Staffing, Berkeley, Calif.: University of California: Research Program in University Administration, Office of the Vice President-Planning and Analysis. Paper P 11 (November 1970).
75. Rowland K. M., Sovereign M. G., Markov-Chain Analysis of Internal Manpower Supply, *Ind. Rel.*, 9, 88--89 (October 1969).
76. Root J. G., An Application of Symbolic Logic to a Selection Problem, *Operations Res.*, 12, 519--529 (July - August 1964).
77. Sikes T. W., The Search Decision Rule Applied to Aggregate Planning: Improved Efficiency and Stochastic Extensions, Unpublished Ph. D. Dissertation, University of California, Los Angeles, 1970.
78. Silver E. A., A Tutorial on Production Smoothing and Work Force Balancing, *Operations Res.*, 15, 983--1010 (1967).
79. Sinha N. P., Manpower Planning: A Research Bibliography, Minneapolis: Industrial Relations Center. University of Minnesota, Bulletin 52 (January 1970).
80. Smith R. D., Greenlaw P. S., Simulation of a Psychological Decision Process in Personnel Selection, *Management Sci.*, 13, 8, B-409 - B-419 (April 1967).
81. Taylor E. R., Nevis E. G., Personnel Selection, *Ann. Rev. Psy.*, 12, 389--412 (1961).
82. Taylor F. W., The Principle of Scientific Management, Harper Brothers, N.Y., 1911.
83. Taylor F. W., Shop Management. Harper Brothers, N.Y., 1919.
84. Taubert W. H., Search Decision Rule for the Aggregate Scheduling Problem, *Management Sci.*, 14, 6, 343-359 (February 1968).
85. Taubert W. H., The Search Decision Rule Approach to Operation Planning, Unpublished Ph. D. Dissertation. University of California, Los Angeles, California, 1968.
86. Thompson J. D., Organization in Action, McGraw-Hill, N.Y., 1967.

87. Trist E. L., Bamforth K. W., Some Social and Psychological Consequences of the Longwall Method of Coal-getting, *Human Rel.*, 4, 1, 3—38 (1951).
88. Trist E. L., Murray H., Work Organization at the Coal Face: A Comparative Study of Mining Systems, Tavistock Institute of Human Relations, Doc. 506 (1948).
89. Trist E. L., Higgin G. W., Murray H., Pollack A. B., Organizational Choice, Tavistock, London, 1963.
90. Van Beinum H. J. J., The Design of the New Radial Tyre Factory as an Open Socio-Technical System, Tavistock Institute of Human Relations, Human Resources Centre, Doc. HRC 150 (October 28, 1968).
91. Vroom V. H., MacCrimmon K. R., Toward a Stochastic Model of Management Careers, *Admin. Sci. Quart.*, 13, 1, 26—46 (June 1968).
92. Waid C., Clark D. F., Ackoff R. L., Allocation of Sales Effort in the Lamp Division of the General Electric Company. *Operations Res.*, 4, 629—647 (1956).
93. Wagner H. M., Principles of Management Science, Prentice-Hall, Englewood Cliffs, Inc., N.J., 1970.
94. Walker J. W., Forecasting Manpower Needs. *Harvard Bus. Rev.*, 27, 2, 132—164 (March — April 1969).
95. Warner M. D., Prawda J., A Mathematical Programming Model for Scheduling Nursing Personnel in a Hospital. *Management Sci.*, 19, 4, 411—422 (December 1972).
96. Weber W. L., Manpower Planning in Hierarchical Organizations: A Computer Simulation Approach, *Management Sci.*, 18, 3, 119—144 (November 1971).
97. Weil R. L., Functional Selection for the Stochastic Assignment Model, *Operations Res.*, 15, 6, 1063—1067 (November/December 1967).
98. Weissman C., LISP 1. 5 Primer. Dickenson Publishing Co., Belmont, California, 1967.
99. Welch N., Florian M., An Application of Linear Programming to Wage Analysis, *Canadian Operational Research Society*, 8, 1, 28—37 (November 1970).
100. Young A., Models for Planning Recruitment and Promotion of Staff. *Brit J. Ind. Rel.*, 3, Part 3, 301—310 (1965).
101. Young A., Almond G., Predicting Distribution of Staff, *Computer J.*, 3, 246—250 (1961).

1-3. Агрегированное планирование производства

А. Хэкс¹⁾.

ВВЕДЕНИЕ

Производственное планирование предусматривает определение уровней производства и запасов, а также фонда рабочей силы, необходимых для удовлетворения меняющегося спроса. Обычно для рассматриваемого периода планирования материальные ресурсы предприятия считаются заданными, и целью планирования является обеспечение наиболее эффективного использования этих ресурсов. Необходимость в подобном планировании, как правило, обусловлена несоответствием сроков и объема производства продукции, продиктованных спросом, срокам и объему производства, обеспечивающим эффективное использование ресурсов производства. Планирование эффективного использования ресурсов производства осуществляют в сводных (агрегированных) показателях, если факторы (спрос, различные виды затрат, производственные мощности), влияющие на процесс производства, не стабильны во времени. При этом период планирования определяется характером изменения указанных факторов (например, если спрос изменяется в зависимости от сезона, то период планирования должен охватывать весь цикл изменений). Период планирования может составлять от 6 до 18 месяцев, для большинства систем он равен 12 месяцев.

При исключении этапа агрегированного планирования в составленном плане будут учтены только ближайшие по времени потребности предприятия и не будут учтены возможные в будущем нежелательные последствия реализации этого плана. Подобная ситуация может возникнуть в системе ²⁾ управления запасами, в которой пополнение запасов производится по точке заказа. Если в такой системе каждое изделие рассматривается изолированно от других и не учитывается сезонность спроса, то в начале сезонного пика спрос быстро растет и по большому ассортименту изделий точка заказа достигается одновременно. Это потребует увеличения производства изделий в соответствии с размерами заказов. Руководство предприятия, будучи не в состоянии выполнить все заказы при сохранении соответствующего уровня обслуживания, может пойти на сокращение производственного цикла по каждому изде-

¹⁾ Arnoldo C. Hax, Ph. D., Professor of Management in the Alfred P. Sloan School of Management at the Massachusetts Institute of Technology, Cambridge, Massachusetts.

²⁾ В системе управления запасами по точке заказа пополнение запасов изделия осуществляется всякий раз, когда *наличное количество изделий плюс величина заказа минус невыполненные заказы достигает определенного уровня*, называемого точкой заказа. Точка заказа обычно определяется как *ожидаемый спрос в пределах срока поставки (или производства) изделия плюс резервный запас*. Экономичный размер заказа (ЭРЗ) есть фиксированная величина, минимизирующая общие затраты, включая затраты на хранение запасов, закупку (или производство) заказанного количества изделий. Подробное описание системы управления запасами по точке заказа можно найти в работах [11, 61].

люю путем многократной перестройки производства на выпуск изделий в небольших количествах. Такая перестройка приведет к снижению общей производительности (из-за высокого процента простоя оборудования вследствие большого числа перепалодок), повышению стоимости изделий и ухудшению качества обслуживающих потребителей. Кроме того, поскольку в конце сезона изделия будут производиться все еще в количествах, определяемых размером заказа, то их реализация станет возможной только в начале следующего сезона. (Реализация изделий возможна и до наступления следующего сезона, но по сниженным ценам.) Таким образом, агрегированное планирование можно исключить только в том случае, когда спрос, затраты и цены остаются стабильными и предприятие обеспечено ресурсами в количестве, достаточном для удовлетворения его постоянных потребностей.

Планирование производства каждой единицы (мелкой детали изделия) продукции, да еще в течение продолжительного периода планирования, оказывается практически невозможным, и в этом случае также осуществляют агрегированное планирование. Для этого осуществляется классификация однородных изделий, машин, работ и отдельных потребителей и используются сопоставимые единицы измерения объединяемых элементов. Способ агрегирования определяется и используемой системой планирования, а также технологическими и управленческими характеристиками производственного процесса. После составления агрегированного плана формулируются ограничения задачи планирования каждой из подлежащих производству единиц изделий продукции. Эти ограничения обычно устанавливают общее количество изделий данной группы, которые должны быть изготовлены в течение месяца, а также численность персонала, эффективность использования оборудования, необходимое количество часов сверхурочного времени.

Способы «поглощения» колебаний спроса. Существует несколько способов «поглощения» колебаний спроса:

- изменение фонда рабочей силы путем приема и увольнения. Это позволяет регулировать производительность в соответствии с колебаниями спроса. Однако чрезмерное использование данного способа может привести к возникновению серьезных трудностей;

- сохранение постоянного состава рабочей силы и изменение производительности за счет сверхурочного времени и/или сокращения времени простоя либо путем заключения контрактов с субподрядчиками;

- сохранение постоянным объема производства и накопление продукции, пользующейся сезонным спросом. В большинстве случаев, прежде чем прибегнуть к этому способу, следует оценить соотношение между затратами, связанными с изменением производительности, и хранением таких изделий;

- осуществление политики отсрочки выполнения заказов при условии, что потребитель согласен на такую отсрочку;

- привлечение дополнительных производственных мощностей в зависимости от сезонных колебаний спроса. Этот способ обычно применяется на более высоком уровне планирования, он обеспечивает эффективное использование ресурсов предприятия, но не исключает необходимости в агрегированном планировании.

Затраты, принимаемые во внимание при агрегированном планировании производства. При агрегированном планировании необходимо учитывать следующие виды затрат:

- основные затраты производства, т. е. постоянную и переменную составляющую затрат производства продукции данного типа за данный период времени, прямые и косвенные трудовые затраты и надбавки за регламентированную и сверхурочную работу;

- затраты, связанные с изменением объема производства. Типичными затратами этого вида являются затраты на прием, обучение и увольнение;

- затраты на хранение запасов, основная составляющая которых обусловлена тем, что капитал, соответствующий стоимости запасов, оказы-

вается связанным. Другие составляющие связаны с хранением, страхованием, выплатой налогов, устареванием изделий;

— затраты, связанные с невыполнением заказов в срок.

Обычно затраты этого вида трудно измерить, поскольку они включают затраты на форсирование выполнения заказа, издержки, связанные с потерей доверия потребителей, и убытки от распродажи по сниженным ценам избытка продукции, образующегося в результате задержки выполнения заказа. (Более подробно о затратах см. в работах [39, 63].)

1. МАТЕМАТИЧЕСКИЕ МОДЕЛИ АГРЕГИРОВАННОГО ПЛАНИРОВАНИЯ ПРОИЗВОДСТВА

Математическое моделирование играет важную роль в процессе принятия решений при прогнозированном планировании. В одной из работ, посвященных этой проблеме, перечисляются следующие задачи управления, при решении которых модели оказываются весьма эффективными:

— определение способа количественной оценки показателей и использование неосознанных принципов, которые обычно случайно и эпизодически учитываются в решениях по планированию;

— проведение в регулярном порядке всестороннего анализа факторов, имеющих отношение к задачам планирования. Это включает возможность принятия решений на основании тривиальных и поверхностных оценок;

— отведение каждому решению по планированию соответствующего места в установленной последовательности решений и корректировка ошибок первоначального прогноза, осуществляемая с помощью механизма обратной связи, включенного в систему выбора вариантов решения;

— изменение установившейся практики принятия решений для обеспечения большей свободы действий, в особенности при возникновении чрезвычайных ситуаций.

Рассмотрим различные типы моделей агрегированного планирования, используя классификацию этих моделей в соответствии с предположениями о структуре составляющих затрат.

1.1. Модели с линейной функцией затрат

При разработке первых моделей систем агрегированного планирования предполагалось, что затраты линейно зависят от управляющих переменных модели. В результате этого предположения такие модели могли быть использованы для решения довольно узкого круга задач. Поскольку, однако, оказалось, что нелинейные выпуклые функции затрат могут быть с любой степенью точности аппроксимированы кусочно-линейными функциями, то область применения моделей такого типа существенно расширилась. Подобные модели и сегодня находят широкое применение.

что объясняется возможностью использования эффективных вычислительных методов, разработанных для решения задач линейного программирования.

Рассмотрим два основных типа линейных моделей.

1.2. Модель планирования производства с постоянным фондом рабочей силы

Если предприятие имеет постоянный фонд рабочей силы, то, следовательно, объем производства может меняться только за счет использования сверхурочного времени и «поглощение» колебаний спроса путем приема и увольнения невозможно. Для формального математического описания такой системы используем в качестве *параметров модели*:

- затраты на производство одной единицы продукции i -го типа в течение t -го интервала времени — v_{it} ,
 - затраты на хранение одной единицы продукции i -го типа в течение t -го интервала времени — c_{it} ,
 - стоимость одного человеко-часа трудозатрат штатного сотрудника в течение t -го интервала времени — r_{it} ,
 - стоимость одного человеко-часа сверхурочной работы в течение t -го интервала времени — o_t ,
- а в качестве *переменных модели*:

- число единиц продукции i -го типа, производимой в течение t -го интервала времени — X_{it} ,
- число единиц продукции i -го типа, которое необходимо иметь в запасе к концу t -го интервала времени — I_{it} ,
- количество человеко-часов постоянных сотрудников в течение t -го интервала времени — W_t ,
- количество человеко-часов сверхурочной работы в течение t -го интервала времени — O_t .

Тогда задачу планирования можно сформулировать как минимизацию суммы затрат на производство продукции, хранение запасов, затрат, связанных с оплатой труда постоянных сотрудников, включая и сверхурочную работу, т. е.

найти

$$\min z = \sum_{i=1}^N \sum_{t=1}^T (v_{it}X_{it} + c_{it}I_{it}) + \sum_{t=1}^T (r_t W_t + o_t O_t) \quad (1)$$

при ограничениях

$$X_{it} + I_{i,t-1} - I_{it} = d_{it}, \quad t=1, \dots, T; \quad i=1, \dots, N, \quad (1a)$$

$$\sum_{i=1}^N k_i X_{it} - W_t - O_t = 0, \quad t=1, \dots, T, \quad (1б)$$

$$0 \leq W_t \leq (rm)_t, \quad t=1, \dots, T, \quad (1в)$$

$$0 \leq O_t \leq (om)_t, \quad t = 1, \dots, T, \quad (1г)$$

$$X_{it}, I_{it} \geq 0, \quad i = 1, \dots, N; \quad t = 1, \dots, T, \quad (1д)$$

где d_{ik} — спрос на продукцию i -го типа в течение t -го интервала времени, k_i — количество человеко-часов, необходимых для производства единицы продукции i -го типа, $(rm)_t$ — общее количество человеко-часов трудозатрат штатных сотрудников в течение t -го интервала времени, $(om)_t$ — общее количество человеко-часов сверхурочного времени в течение t -го интервала времени, I_{i0} — уровень начальных запасов продукции i -го типа, W_0 — число постоянных сотрудников, T — период планирования, N — общее число наименований типов продукции.

Если предельные затраты v_{it} на производство одной единицы продукции i -го типа не зависят от времени, то член $v_{it}X_{it}$ можно не учитывать, так как общее количество продукции каждого типа фиксировано. Аналогично, если величина W_t не зависит от времени, то член $r_t W_t$ можно исключить из выражения (1).

Ограничение (1а) представляет собой уравнение баланса по количеству производимой продукции и уровню запасов при условии детерминированного спроса на каждый тип продукции в каждый заданный интервал времени. Одним из способов учета неопределенности прогноза спроса является определение для каждого заданного интервала времени минимального уровня резервного запаса, т. е. введение ограничения вида $I_{it} \geq ss_{it}$, где ss_{it} — резервный (буферный) запас продукции i -го типа в течение времени t ¹⁾. Следует заметить, что ограничения (1а и 1д) означают отсутствие задолженности по выполнению заказов в течение t -го интервала времени.

Ограничение-равенство (1б) определяет допустимое количество трудовых ресурсов, которые могут быть использованы для выпуска продукции i -го типа в каждый заданный интервал времени, причем это — единственное ограничение на резервы производства. Если сохранить предположение о линейной зависимости затрат от управляющих переменных системы, то целевая функция будет учитывать не только трудовые, но и другие ресурсы производства.

Ограничения (1в, 1г), по существу, устанавливают нижний и верхний пределы человеко-часов трудозатрат штатных сотрудников и сверхурочных работ для каждого заданного интервала времени. Как уже было показано, учет неопределенности прогноза спроса приводит к введению ограничения $I_{it} \geq ss_{it}$. Если уровень запасов в конце периода планирования никак не ограничен, то решение, получаемое с помощью модели, всегда обеспечивает

¹⁾ Величина резервного запаса зависит от качества прогнозирования спроса и от обеспечиваемого уровня обслуживания потребителей [10].

$I_{it} = 0$ для всех значений i . В тех случаях, когда полное исчерпание запасов по всем видам продукции нежелательно, в модель должны быть введены соответствующие ограничения на уровень запасов. Дополнительные ограничения должны быть введены и в тех случаях, когда потребности в складировании продукции не могут превышать некоторой заданной величины. Например, ограничение

$$\sum_{i=1}^N I_{it} \leq (sc)_t, \quad t = 1, \dots, T,$$

означает, что суммарный объем запасов всей продукции в каждый заданный интервал времени не может быть большим, чем общая емкость склада $(sc)_t$. В тех случаях, когда необходимо зафиксировать изготовление продукции за различными производственными единицами ограниченной мощности, управляющие переменные выбирают таким образом, чтобы получить соответствующие решения. Например, величина X_{ict} может означать число единиц продукции i -го типа, производимой производственной единицей c в течение t -го интервала времени. Результат подобной замены (или преобразования) управляющих переменных необходимо программировать непосредственно в рамках общей модели.

Следует иметь в виду, что даже в случае очень простой модификации рассматриваемой модели могут возникнуть существенные трудности, связанные с процессом вычисления, если не выполнена укрупненная классификация отдельных видов продукции, производство которых планируется. Если предположить, что мы не будем принимать во внимание ограничения (1в-д), которые лишь устанавливают верхние и нижние пределы значений переменных, то все еще остается $T(N+1)$ существенных ограничений.

На сложных производствах общее число отдельных наименований изделий N может достигать нескольких тысяч. Например, если модель составлена для периода планирования, равного 12, и 5000 наименований изделий, то она должна содержать около 60 000 ограничений.

Это, естественно, превышает возможности линейного программирования.

В большинстве случаев, однако, нерационально планировать распределение производственных ресурсов на таком уровне детализации. Это объясняется следующими причинами. Во-первых, программы детального планирования должны принимать в расчет большое количество показателей технологии производства и системы сбыта, которые не могут быть включены в общую модель, так как имеют явно выраженный качественный характер. Во-вторых, как уже отмечалось, многие результаты планирования, полученные с помощью модели, широко связаны с решением задач

распределения ресурсов и, следовательно, чрезмерно подробная информация будет только затруднять процесс принятия решения, а не способствовать ему. В-третьих, прогнозы, основанные на агрегированной информации, являются более точными, чем прогнозы, основанные на подробной информации. Поэтому обычно осуществляют укрупненную классификацию изделий. Критерии подобной классификации определяются структурой модели: единицы продукции одной группы должны совместно формировать спрос d_{ij} , иметь общие показатели затрат v_{ij} , c_{ij} , r_i , o_i и на изготовление каждой единицы продукции должно затрачиваться одно и то же количество времени k_i . Принятые решения по агрегированному планированию позволяют сформулировать ограничения, которым должен удовлетворять детальный план изготовления изделий.

Рассмотренная модель, как и любая другая модель планирования, требует определения периода планирования и разбиения его на интервалы, длительность которых не обязательно должна быть одинаковой. Для многих систем планирования целесообразно разбивать период планирования на интервалы различной длительности, так как в начале периода мы располагаем более подробной информацией. Из-за неопределенности условий, в которых реализуются решения по планированию, последние обычно оказываются полностью реализованными лишь на начальном этапе планирования. В конце каждого этапа появляется новая дополнительная информация, которая используется для корректировки модели и пересмотра плана для последующих этапов планирования. При формулировании задачи могут быть учтены ограничения, связанные с технологией производства, финансами, организационной структурой, рекламой, системой сбыта. Такая гибкость, характерная для подходов к решению задач линейного программирования, позволяет довольно широко и эффективно использовать этот тип модели.

Простой вариант задачи линейного программирования с фиксированным фондом рабочей силы, имеющей структуру транспортной задачи, впервые был рассмотрен в работе [8].

1.3. Модель планирования с переменным фондом рабочей силы

В тех случаях, когда «поглощение» колебаний спроса оказывается возможным за счет изменения фонда рабочей силы, последний становится переменной модели. Так как подобное изменение осуществляется в течение периода планирования путем приема и увольнения, то затраты, связанные с увольнением и приемом, должны быть учтены в целевой функции. Кроме того, поскольку модель допускает учет недостатка в ресурсах, то, следо-

вательно, необходимо учитывать и издержки, связанные с невыполнением заказов.

В данном случае *переменными модели* являются: X_{it} — число единиц продукции i -го типа, производимой в течение t -го интервала времени; W_t — количество человеко-часов работы штатных сотрудников в течение t -го интервала времени; O_t — количество человеко-часов сверхурочной работы в течение t -го интервала времени; H_t — количество человеко-часов, зачисленных в штат организации в течение t -го интервала времени; F_t — количество неиспользованных человеко-часов в течение t -го интервала времени из-за увольнения штатных сотрудников; I_{it}^+ — число единиц продукции i -го типа, которую необходимо иметь в задаче к концу t -го интервала времени; I_{it}^- — число единиц продукции i -го типа, которой не хватает для удовлетворения спроса потребителя в конце t -го интервала времени.

Используя перечисленные переменные модели и обозначения, принятые для предыдущей модели, определим следующие составляющие затрат в течение времени t :

- переменная затрат производства $v_{it}X_{it}$,
- затраты на хранение запасов $c_{it}I_{it}^+$,
- затраты, связанные с задержкой выполнения заказов потребителей. $b_{it}I_{it}^-$,
- затраты на содержание штатных сотрудников r_tW_t ,
- затраты на оплату сверхурочной работы o_tO_t ,
- затраты, связанные с приемом на работу, h_tH_t ,
- затраты, связанные с увольнением, f_tF_t .

В этом случае задача планирования формулируется следующим образом:

найти

$$\min z = \sum_{i=1}^N \sum_{t=1}^T (v_{it}X_{it} + c_{it}I_{it}^+ + b_{it}I_{it}^-) + \sum_{t=1}^T (r_tW_t + o_tO_t + h_tH_t + f_tF_t) \quad (2)$$

при ограничениях

$$X_{it} + I_{it}^+ - I_{it}^- - I_{it}^+ + I_{it}^- = d_{it}, \quad i=1, \dots, N; \quad t=1, \dots, T, \quad (2a)$$

$$\sum_{i=1}^N k_i X_{it} - W_t - O_t \leq 0, \quad t=1, \dots, T, \quad (2б)$$

$$W_t - W_{t-1} - H_t + F_t = 0, \quad t=1, \dots, T, \quad (2в)$$

$$-pW_t + O_t \leq 0, \quad t=1, \dots, T, \quad (2г)$$

$$X_{it}, I_{it}^+, I_{it}^- \geq 0, \quad i=1, \dots, N; \quad t=1, \dots, T,$$

$$W_t, O_t, P_t, H_t, F_t \geq 0, \quad t=1, \dots, T.$$

Ограничение (2а) представляет собой уравнение баланса по количеству произведенной продукции и уровню запасов. Заметим, что это уравнение эквивалентно уравнению баланса (1а), если принять, что $I_{i,t-1} = I_{i,t-1} - I_{it}^+$ и $I_{it} = I_{it}^+ - I_{it}^-$.

В рассматриваемой модели величина запаса I_t к концу t -го интервала времени может иметь положительное значение ($I_{it}^+ > 0$ означает, что к концу t -го интервала времени остается часть запаса) или отрицательное значение ($I_{it}^- > 0$ означает, что к концу t -го интервала времени происходит накопление невыполненных заказов). Поскольку, однако, как величине I_{it}^+ , так и величине I_{it}^- соответствуют определенные затраты, то эти переменные никогда не будут одновременно принимать положительных значений. Ограничения (2б) отражают тот факт, что объем производства лимитируется фондом трудовых ресурсов.

Ограничения-равенства (2в) определяют изменение фонда рабочей силы в течение времени t , т. е. $W_t - W_{t-1} = H_t - F_t$. Фонд рабочего времени увеличивается, если $H_t > 0$, и сокращается, если $F_t > 0$. Аналогично предыдущему случаю величины H_t , F_t никогда не будут принимать одновременно положительных значений в течение заданного интервала времени, так как им соответствуют определенные затраты. Ограничение (2г) устанавливает верхний предел объема сверхурочной работы в течение времени t как функцию фонда рабочего времени штатных сотрудников, т. е. $O_t \leq pW_t$, где p — допустимая часть фонда рабочего времени штатных сотрудников для выполнения сверхурочных работ.

Модель такого типа впервые была рассмотрена в работе [33], а возможные модификации модели — в работах [69, 87].

Липшман и др. [59] проанализировали вариант оптимального решения задачи планирования производства продукции одного типа в предположении, что функция затрат производства является выпуклой, функция затрат на хранение запасов — возрастающей, а функция затрат, связанных с изменением фонда трудовых ресурсов, V -образной. Авторы работы [60] предложили для частного случая, когда все функции затрат линейны, а спрос либо монотонно убывает, либо возрастает, эффективный алгоритм решения задачи.

Алгоритм представляет собой итеративную процедуру, которая начинается с предположения, что задано значение W_T , которое должно быть достигнуто к концу периода планирования. Это обеспечивает выбор оптимального решения для фиксированного значения W_T и проверку решения на оптимальность. Если оказывается возможным улучшить решение, то алгоритм формирует новое значение W_T , обеспечивающее улучшенное решение, и процесс повторяется. Сходимость алгоритма гарантируется за конечное число итераций.

Если функция затрат линейна и функция спроса неубываю-

щая, существует оптимальное решение, для которого выполняются условия

$$\begin{aligned} W_{t+1} &\geq W_t, \quad t=1, \dots, T-1, \\ O_{t+1} &\geq O_t, \quad t=1, \dots, T-1, \\ \frac{O_{t+1}}{W_{t+1}} &\geq \frac{O_t}{W_t}, \quad t=1, \dots, T-1, \\ (W_T - W_1) O_t &= 0, \quad t=1, \dots, T. \end{aligned}$$

(Этот результат постоянно используется в процессе вычисления.)

Юан [92] обобщил этот подход для решения задачи планирования производства продукции различных типов.

В работах [2, 38] рассматривается модель производства продукции одного типа с линейной функцией затрат, при этом целевая функция учитывает изменения уровня производства. Авторы проанализировали качественные свойства оптимального решения и предложили простые методы поиска этого решения для случая, когда спрос является монотонно возрастающей функцией. Эта работа была обобщена Джоусоном и Данишгом [46].

Модели линейного программирования легко могут быть применены для планирования многоэтапного процесса производства. В работе [44] проведено исчерпывающее обсуждение многошаговых моделей линейного программирования, включающих множество технологий, множество источников снабжения, решения по выбору номенклатуры продукции, а также решения по организации многономенклатурного производства и распределения.

1.4. Преимущества и недостатки моделей с линейной функцией затрат

Основным преимуществом моделей с линейной функцией затрат является возможность «свести» задачи агрегированного планирования к задачам линейного программирования, решение которых можно быстро осуществить с помощью легкодоступных эффективных машинных программ. Подобные программы позволяют быстро и дешево получать решение для моделей с большим числом переменных и ограничений. Кроме того, такие модели хорошо поддаются параметрическому анализу и анализу на чувствительность, что может оказаться полезным в процессе агрегированного планирования. Так, информация о скрытых затратах может быть использована с целью определения благоприятных условий для расширения производственных мощностей, стратегий проникновения в сферу сбыта, целесообразности освоения производства новой продукции и т. д. Что касается предположения о линейной зависимости затрат от переменных, то, во-первых, основные составляющие затрат, как правило, являются линейными функциями

в интервале возможных изменений рассматриваемых переменных и, во-вторых, выпуклые сепарабельные функции могут быть аппроксимированы кусочно-линейными. Более того, некоторые функции, имеющие на первый взгляд нелинейные характеристики, могут быть линеаризованы [33]. Однако модели линейного программирования не позволяют учесть неопределенности спроса, что, несомненно, является их существенным недостатком. В этой связи заслуживает внимания работа [19], в которой сообщается о положительных результатах экспериментов по использованию моделей линейного программирования при достаточно высокой степени неопределенности в динамических условиях.

2. МОДЕЛИ С КВАДРАТИЧНОЙ ФУНКЦИЕЙ ЗАТРАТ

Если для решения задач агрегированного планирования производственных мощностей используются модели с квадратичными функциями затрат, то решающие правила имеют линейную структуру. Это связано с тем, что в результате дифференцирования квадратичной функции получается линейная функция. Первая такая модель была разработана Холтом, Модильяни, Мазом и Саймоном (ХММС) [39] для случая полного агрегирования всех типов продукции в один класс. Это в свою очередь предполагает использование соответствующих составимых единиц измерения продукции.

В модели ХММС используются в основном две переменные: объем выпуска агрегированной продукции в течение t -го интервала времени P_t и фонд рабочей силы в течение t -го интервала времени W_t . Переменная I_t — уровень запаса t -го интервала времени — определяется с помощью переменных P_t и W_t и соответствующих соотношений между этими переменными.

Оптимальное решающее правило определяет для каждого интервала времени необходимый объем производства агрегированной продукции и величину фонда рабочей силы, минимизирующие квадратичную функцию затрат.

Составляющие затрат. Рассмотрим подробнее составляющие затрат, которые учитываются в квадратичной функции затрат.

Затраты на содержание персонала. Предполагается, что эти затраты линейно возрастают с увеличением фонда рабочей силы и описываются величиной $c_1 W_t + c_{13}$, где c_1 и c_{13} — коэффициенты затрат, которые определяются вне модели. Поскольку c_{13} — константа, то она может быть исключена из дальнейшего рассмотрения.

Затраты, связанные с приемом и увольнением. Предполагается, что эти затраты являются квадратичной функцией изменения фонда рабочей силы ($W_t - W_{t-1}$) и описываются величиной

рые должны быть определены, причем константа c_{11} учитывает асимметрию функции затрат.

Затраты, связанные со сверхурочными работами и простоями. При условии что фонд рабочего времени равен W_t , желаемый объем производства определяется величиной $c_4 W_t$. Если объем производства превышает это количество, то будут иметь место затраты, связанные со сверхурочной работой; если же объем производства менее чем величина $c_4 W_t$, то будут иметь место издержки, связанные с простоями. Рассматриваемые затраты характеризуются величиной $c_3 (P_t - c_4 W_t)^2 + c_5 P_t - c_6 W_t + c_{12} P_t W_t$, где последние три члена включены для повышения точности.

Затраты, связанные с хранением запасов. Издержки, обусловленные невыполнением заказов. Предполагается, что затраты, связанные с хранением запасов, описываются величиной $c_7 [I_t - (c_8 + c_9 d_t)]^2$, где d_t — ожидаемый уровень спроса на агрегированный продукт в течение t -го интервала времени.

Планируемый уровень запасов определяется величиной $c_8 + c_9 d_t$. Отклонения от этого уровня приводят к затратам на хранение запасов или к издержкам, обусловленным невыполнением заказов, причем затраты и издержки возрастают пропорционально квадрату отклонений.

В работе [39] значение c_9 было принято равным нулю. Расчет коэффициентов затрат является длительной дорогостоящей процедурой, требующей статистического анализа данных учета и данных о системе управления.

Рядом исследователей была проделана большая работа с целью улучшения качества оценок составляющих затрат [55] и определения функций агрегированных затрат, характеризующих затраты на производство отдельных изделий [54].

Построение модели. Задача агрегированного планирования производства с учетом структуры затрат может быть сформулирована как минимизация ожидаемых затрат:

найти

$$\min z = \sum_{t=1}^T [(c_1 - c_6) W_t + c_2 (W_t - W_{t-1} - c_{11})^2 + c_3 (P_t - c_4 W_t)^2 + c_5 P_t + c_{12} P_t W_t + c_7 (I_t - c_8 - c_9 d_t)^2] \quad (3)$$

при ограничениях

$$P_t + I_{t-1} - I_t = d_t, \quad t = 1, \dots, T, \quad (3a)$$

$$P_t, W_t \geq 0, \quad t = 1, \dots, T. \quad (3б)$$

Одним из возможных подходов является отказ от предположения о детерминированности спроса d_t . В работе [56] показано,

что при песмещенных оценках прогнозируемого спроса решение сформулированной задачи при ограничениях (3а), (3б) дает минимум ожидаемых затрат.

2.1. Линейные решающие правила

Если целевая функция рассматриваемой модели строго выпуклая, то она имеет единственный глобальный минимум. Это условие обычно выполняется в тех случаях, когда имеет место возрастание предельных затрат.

Оптимальное решение задачи находит с помощью метода множителей Лагранжа.

Несколько примеров решения задач агрегированного планирования с помощью конкретных решающих правил описаны в работе [11].

В общем виде линейные решающие правила могут быть представлены уравнениями вида

$$P_t = a_0 d_t + a_1 d_{t+1} + \dots + a_{T-t} d_T + bW_{t-1} + c - dI_{t-1}, \quad (4)$$

$$W_t = e_0 d_t + e_1 d_{t+1} + \dots + e_{T-t} d_T + fW_{t-1} + g - hI_{t-1}. \quad (5)$$

Уравнение (4) устанавливает зависимость объема производимой агрегированной продукции в течение t -го интервала времени от прогнозов будущего спроса, фонда рабочей силы в течение времени $(t - 1)$ и уровня начального запаса.

Агрегированный фонд рабочей силы (5) в течение t -го интервала времени также линейно зависит от прогнозов будущего спроса, фонда рабочей силы в течение $(t - 1)$ -го интервала времени и уровня начального запаса.

Значения весов множителей (a и e), приписываемых прогнозируемым спросам, быстро убывают к концу периода планирования.

2.2. Модификация модели ХММС

Рассмотрим несколько модификаций модели ХММС. Авторы работы [7] сформулировали обобщенную многопродуктовую задачу и предложили модель многономенклатурного производства, целевая функция в которой учитывает сокращение предельного (добавочного) дохода. Эта работа была продолжена Гаусманом и Мак-Клейном [34], которые ввели в модель недетерминированный спрос на изделия. Чапг и Джонс [12] также рассматривали задачу агрегированного планирования производства продукции различных типов и предложили алгоритм решения для случая, когда производство не может быть начато и закончено в заданный период

времени. Сиккенс [79] в качестве вспомогательных переменных модели использовал производственные мощности. Петерсон [70] предложил модификацию модели ХММС, которая позволяет изготовителю с помощью изменения цен на продукцию таким образом распределить выполнение заказов, чтобы свести к минимуму влияние колебаний фонда рабочей силы, объема производства и уровней запасов.

2.3. Преимущества и недостатки моделей с квадратичной функцией затрат

Основные преимущества моделей с квадратичной функцией затрат состоят в том, что они удовлетворяют требованию наибольшего возможного соответствия существующей структуре затрат в процессе планирования, позволяют использовать линейные решающие правила и непосредственно учитывать элементы неопределенности, так как ожидаемые затраты минимизируются при условии несмещенных оценок ожидаемого спроса.

Наиболее серьезными недостатками этих моделей являются принципиальная необходимость в агрегировании, громоздкость процедур оценок, которые требуются для определения численных значений коэффициентов затрат, и вычислительные трудности, возникающие при возрастании числа переменных и ограничений.

Результаты вычислений, полученные в работе [82], свидетельствуют о том, что решающие правила, по-видимому, не очень чувствительны к ошибкам в оценке показателей составляющих затрат, что можно рассматривать, как положительное свойство модели, поскольку практически невозможно получить точные значения затрат.

Несмотря на обнадеживающие результаты применения рассматриваемого типа моделей, в реальных деловых ситуациях они не используются. Вероятно, перечисленные недостатки модели оказываются более существенными, чем недостатки моделей линейного программирования. Кроме того, модели линейного программирования позволяют решать задачи очень большой размерности, что значительно расширяет область их применения.

3. МОДЕЛИ ДЛЯ ОПРЕДЕЛЕНИЯ ОПТИМАЛЬНОГО РАЗМЕРА ПАРТИИ ПРОДУКЦИИ (МОДЕЛИ С ПОСТОЯННЫМИ ЗАТРАТАМИ)

Если производство является серийным (в противоположность поточному), то оно характеризуется затратами на палачочные работы, связанные с перестройкой производства на заданный рабочий цикл. При планировании производства с учетом таких затрат возникает ряд существенных проблем. Во-первых, каждая деталь, требующая определенной перепалачки оборудования (или

некоторая группа деталей, для производства которых необходима комплексная переналадка оборудования), должна выделяться и рассматриваться отдельно. Это настолько увеличивает число переменных и ограничений модели, что для получения решения необходимы специальные вычислительные методы. Во-вторых, включение затрат на переналадочные работы порождает задачу выбора размера неделимой партии продукции, так как производство продукции данного типа связано только с одним видом переналадочных работ.

Все это приводит к тому, что в модель вводятся целочисленные переменные. Кроме того, простои, которые неизбежны при каждой наладке, дополнительно вносят нелинейности в множество ограничений. В результате получается целочисленная модель нелинейного программирования большой размерности, что, естественно, затрудняет получение решения с помощью ЭВМ.

Рассмотрим некоторые из наиболее эффективных подходов к решению этой задачи.

3.1. Модель для определения размера партии продукции (изделия) при отсутствии ограничений на его величину

С помощью хорошо известной формулы Уилсона можно определить наиболее экономичный размер заказа (ЭРЗ) в тех случаях, когда затраты на наладочные работы и хранение запасов сопоставимы с затратами по сбыту продукции. Эта формула не учитывает никаких взаимосвязей, существующих между отдельными выпусками продукции, производство которой планируется, и в частности дефицита производственных мощностей, который накладывает наиболее существенные ограничения при планировании. Кроме того, эта формула предполагает, что спрос должен быть постоянным в течение всего периода планирования. В тех случаях, когда спрос известен, но меняется в течение различных интервалов периода планирования, эта формула дает весьма приближенные значения размеров заказа.

Ватнер и Уайтин [90] предложили для определения наиболее экономичного размера заказа использовать методы динамического программирования. Рассмотрим более подробно предложенный ими подход, поскольку он играет важную роль при обсуждении модели выбора размера партии при наличии ограничений на его величину.

В упрощенном виде задача выбора размера партии при отсутствии ограничений на его величину может быть сформулирована следующим образом:

найти

$$\min z = \sum_{t=1}^T [s_t \delta(X_t) + c_t I_t]$$

при ограничениях

$$X_t + I_{t-1} - I_t = d_t, \quad t = 1, \dots, T,$$

$$X_t \geq 0, \quad t = 1, \dots, T,$$

$$I_t \geq 0, \quad t = 1, \dots, T,$$

где

$$\delta(X_t) = \begin{cases} 0, & \text{если } X_t = 0, \\ 1, & \text{если } X_t > 0, \end{cases}$$

и при этом X_t — объем продукции, производимой в течение t -го интервала времени; I_t —уровень запаса к концу t -го интервала времени; s_t — затраты на наладочные работы в течение t -го интервала времени; c_t — затраты на хранение запасов в течение t -го интервала времени; d_t — спрос в течение t -го интервала времени.

Заметим, что переменные, связанные с производственными затратами, в целевую функцию не включены, так как предполагается, что они постоянны в течение всего периода планирования. Кроме того, считается, что невыполнение заказов в течение t -го интервала времени недопустимо.

Функциональное уравнение, описывающее стратегию минимальных затрат (включающее только затраты на наладочные работы и хранение запасов в течение интервала времени $t \leq T - 1$), имеет вид

$$f_t(I_{t-1}) = \min [s_t \delta(X_t) + c_t(X_t + I_{t-1} - d_t) + f_{t+1}(X_t + I_{t-1} - d_t)], \quad X_t \geq 0, \quad X_t + I_{t-1} \geq d_t,$$

а для последнего интервала периода планирования имеем

$$f_T(I_{T-1}) = \min [s_T \delta(X_T)], \quad X_T \geq 0, \quad X_T + I_{T-1} = d_T.$$

Без ограничения общности можно предположить, что к концу периода планирования запасы должны быть исчерпаны, т. е. $I_T = 0$.

Для вычисления оптимального размера партии продукции в течение периода планирования может быть применен обратный индуктивный процесс. Однако этот путь решения не является самым лучшим.

При разработке эффективного алгоритма решения задачи в предположении, что начальный уровень запасов¹⁾ равен нулю,

¹⁾ Если начальный уровень запаса отличен от 0, то для получения скорректированной потребности в этом периоде необходимо вычесть величину данного запаса из спроса, заявленного в первом периоде. Если начальный запас превышает спрос первого периода, то следует продолжить процесс корректировки до тех пор, пока не будет использован весь запас. Затем перейти к предложенному алгоритму.

авторы работы [90] установили ряд важных свойств оптимального решения.

1. Всегда существует оптимальная стратегия, такая, что $I_{t-1}X_t = 0$ для всех $t = 1, \dots, T$. Это означает, что при наличии переходящего запаса и производства продукции в течение одного и того же интервала времени соответствующие затраты не снижаются.

2. Достаточно рассматривать только такие оптимальные стратегии, при которых для всех t выполняется условие

$$X_t = 0 \text{ или } X_t = \sum_{j=t}^k d_j, \text{ где } t \leq k \leq T.$$

Это означает, что объем производства в течение любого интервала времени равен либо нулю, либо суммарной потребности в продукции в течение нескольких последовательных интервалов времени после заданного, включая и заданный.

Если продолжительность периода планирования равна T временным интервалам, то общее число последовательностей производственных операций будет равно 2^{T-1} .

Элементы такой последовательности (обычно называемой доминирующей последовательностью) отражают факт наличия или отсутствия производства продукции в каждый заданный интервал времени.

Подход к решению, предложенный в работе [90], требует рассмотрения только $T(T-1)/2$ таких последовательностей.

3. Всякий раз, когда оптимальное решение для любого заданного интервала времени t предусматривает выполнение условия $I_t = 0$, последовательности интервалов времени от 1 до t и от $t-1$ до T могут рассматриваться как самостоятельные, не зависящие друг от друга последовательности. Поэтому удобно представить процесс решения задачи динамического программирования в виде прямого индуктивного процесса.

Пусть $f(t)$ — значение затрат при оптимальном решении в течение интервалов времени от первого до t -го. Тогда функциональное уравнение, описывающее прямой индуктивный процесс, имеет вид

$$f(t) = \min \left\{ \min [s_j + \sum_{h=j}^{t-1} \sum_{k=h+1}^t c_h d_k + f(j-1)], s_t + f(t-1) \right\},$$

$$1 \leq j < t, \quad (6)$$

где $f(1) = s_1$, а $f(0) = 0$ и s_j — затраты на наладочные работы в течение j -го интервала времени и $\sum_{h=j}^{t-1} \sum_{k=h+1}^t c_h d_k$ — затраты на хранение запасов в течение времени от начала $(j+1)$ -го интервала

до начала t -го интервала. (Численные примеры реализации описанного процесса проводятся в упомянутой работе [90].)

Наиболее важным результатом работы Вагнера и Уайтина является *теорема о периоде планирования*:

Если в течение t^* -го интервала времени функция достигает минимума при $j = t^{**} \leq t^*$, то из всех интервалов времени $t > t^*$ достаточно рассмотреть только интервалы $t^{**} \leq j \leq t$.

В частности, если $t^* = t^{**}$, то достаточно рассмотреть планы, для которых $X_{t^*} > 0$.

При реализации метода решения, использующего прямой индуктивный процесс, исходная задача может быть сформулирована как ряд последовательных задач меньшей размерности. Это достигается путем нахождения интервалов времени, в которых либо объем выпуска продукции положителен, либо уровень запасов равен нулю.

В работе [89] этот подход развит для случая, когда в течение периода планирования, который может быть разбит на большое число интервалов, меняются затраты на производство или закупку продукции. В работах [24, 93] дается более общая формулировка теоремы о периоде планирования. Запгвилл [98] предложил способ учета затрат, связанных с невыполнением заказов, и общее решение этой задачи. Другой подход, учитывающий возможность невыполнения заказов, был предложен Эльмаграби и Баулом [21]. Авторы рассмотрели задачу определения размера партий изделий при отсутствии ограничений на размер партий, предполагая, что заказ должен состоять не менее чем из двух партий изделий, с учетом и без учета затрат на переналадку.

Идея использования доминирующих последовательностей производственных операций оказалась весьма плодотворной с точки зрения улучшения вычислительных возможностей алгоритмов в тех случаях, когда имеются ограничения на размер партии изделий. В последующих разделах иллюстрируется использование этих возможностей.

3.2. Модель определения размера партии при наличии ограничений на его величину

Необходимость разработки модели с ограничением на величину размера партии возникает при решении задач планирования многономенклатурного производства в условиях изменяющегося спроса в течение всего периода планирования. Из-за ограниченного размера партии затраты на переналадку становятся существенной составляющей минимизируемой функции суммарных затрат.

Сначала рассмотрим задачу с постоянным фондом рабочей силы, т. е. когда увеличение объема производства возможно толь-

ко за счет сверхурочной работы, а затем — задачу с переменным фондом рабочей силы, т. е. когда увеличение общего объема производства возможно за счет приема и увольнения.

В общем виде задача может быть сформулирована следующим образом:

найти

$$\min z = \sum_{i=1}^N \sum_{t=1}^T [s_{it} \delta(X_{it}) + v_{it} X_{it} + c_{it} I_{it}] + \\ + \sum_{t=1}^T (r_t W_t + o_t O_t) \quad (7)$$

при ограничениях

$$X_{it} + I_{i,t-1} - I_{it} = d_{it}, \quad t = 1, \dots, T, \quad i = 1, \dots, N, \\ \sum_{i=1}^N [a_i \delta(X_{it}) + k_i X_{it}] - W_t - O_t \leq 0, \quad t = 1, \dots, T, \quad (8) \\ 0 \leq W_t \leq (rm)_t, \quad t = 1, \dots, T, \\ 0 \leq O_t \leq (om)_t, \quad t = 1, \dots, T, \\ X_{it}, I_{it} \geq 0, \quad i = 1, \dots, N, \quad t = 1, \dots, T,$$

где

$$\delta(X_{it}) = \begin{cases} 0, & \text{если } X_{it} = 0, \\ 1, & \text{если } X_{it} > 0; \end{cases}$$

a_i — время, необходимое для наладки оборудования для производства продукции i -го типа; функция $\delta(X_{it})$ является нелинейной.

Большинство замечаний, сделанных при рассмотрении модели планирования производства с постоянным фондом рабочей силы и линейной функцией затрат, имеют силу и для рассматриваемой модели. Кроме того, эта модель предполагает отсутствие невыполненных заказов. (Такие заказы могут быть учтены, если ввести в модель соответствующие переменные.)

Рассмотрим некоторые возможные методы решения сформулированной задачи.

А. Если время, затрачиваемое на переналадку, невелико, то в выражении (8) $a_i = 0$ и модель для определения размера партии продукции с постоянным фондом рабочей силы превращается в модель с постоянными затратами, которая в свою очередь является моделью линейного программирования. Так как целевая функция модели с постоянными затратами является выпуклой, а ограничения — выпуклыми, то глобальный минимум достигается в граничной точке. Однако в общем случае в граничной точке

достигаются и локальные минимумы. Поэтому решение рассматриваемой задачи не может быть получено с помощью симплекс-метода, который позволяет определить только локальный минимум.

Для решения этой задачи были предложены различные методы (в том числе и точные), из которых следует выделить методы, представленные в работах [31, 65] и использующие упорядочение граничных точек, и метод ветвей и границ. Однако эти методы позволяют решать задачи лишь относительно небольшой размерности и поэтому имеют ограниченное практическое значение.

Отмеченные недостатки точных методов решения задачи породили различные эвристические подходы, обеспечивающие получение решений, близких оптимальным. Эти подходы предполагают выбор хорошей начальной граничной точки и далее путем исследования смежных граничных точек — определение локального минимума. Затем отыскивается новая граничная точка, и вычислительный процесс повторяется до тех пор, пока дальнейшее улучшение решения становится невозможным или не будет выполнено заданное число итераций. Эффективные эвристические методы предложены в работах [6, 14, 18, 72, 78].

Б. Если время a_i , затрачиваемое на переналадку оборудования для производства нового типа продукции, достаточно велико, то модель для определения размера партии продукции является нелинейной и имеет большую размерность. В результате решение непосредственно этой задачи вызывает существенные затруднения, для преодоления которых Манн предложил свести сформулированную задачу к задаче линейного программирования [62]. Этот подход в дальнейшем был развит в работах [49, 20, 57].

Предложенный подход предусматривает учет затрат на наладочные работы при определении множества возможных последовательностей операций. Для продукции заданного типа последовательность производственных операций в течение периода планирования представляет собой множество T положительных целых чисел. Этим числам соответствует определенный объем продукции i -го типа, производимой в данный интервал времени и достаточной для удовлетворения спроса. Как уже отмечалось, для решения задачи достаточно рассмотреть для продукции каждого типа 2^{T-1} доминирующих последовательностей.

Пусть переменная X_{ijt} есть число единиц продукции, подлежащей изготовлению в течение t -го интервала времени посредством выполнения j -й последовательности производственных операций $i = 1, \dots, N$; $j = 1, \dots, J$; $t = 1, \dots, T$, и пусть d_{it} — спрос на продукцию i -го типа в течение t -го интервала времени.

Пример, иллюстрирующий метод построения последовательностей производственных операций. Допустим, что период планирования разбивается на три интервала. Число доминирующих последовательностей для продукции

i -го типа будет равно $2^{3-1} = 4$. Эти четыре последовательности могут быть определены следующим образом:

Последовательность производственных операций j	Объем производства в заданный интервал времени t		
	1	2	3
1	$X_{i11} = d_{i1} + d_{i2} + d_{i3}$	$X_{i12} = 0$	$X_{i13} = 0$
2	$X_{i21} = d_{i1} + d_{i2}$	$X_{i22} = 0$	$X_{i23} = d_{i3}$
3	$X_{i31} = d_{i1}$	$X_{i32} = d_{i2} + d_{i3}$	$X_{i33} = 0$
4	$X_{i41} = d_{i1}$	$X_{i42} = d_{i2}$	$X_{i43} = d_{i3}$

Последовательность производственных операций j	Общие затраты t_{ij}
1	$t_{i1} = s_{i1} + v_{i1}(d_{i1} + d_{i2} + d_{i3}) + c_{i1}(d_{i2} + d_{i3}) + c_{i2}(d_{i3})$
2	$t_{i2} = (s_{i1} + s_{i3}) + v_{i1}d_{i1} + v_{i3}d_{i3} + c_{i1}d_{i2}$
3	$t_{i3} = (s_{i1} + s_{i2}) + v_{i1}d_{i1} + v_{i2}(d_{i2} + d_{i3}) + c_{i2}d_{i3}$
4	$t_{i4} = (s_{i1} + s_{i2} + s_{i3}) + v_{i1}d_{i1} + v_{i2}d_{i2} + v_{i3}d_{i3}$

В общем случае

$$t_{ij} = \sum_{t=1}^T [s_{it}\delta(X_{ijt}) + v_{it}X_{ijt} + c_{it}I_{it}].$$

Общие трудовые затраты, необходимые для производства X_{ijt} единиц продукции, могут быть записаны в виде

$$l_{ijt} = a_i\delta(X_{ijt}) + k_iX_{ijt}.$$

Если предположить для простоты, что в каждом интервале времени необходимый фонд рабочей силы не может превышать величины $(rm)_t$, то задача определения размера партии изделий при постоянном фонде рабочей силы может быть сформулирована следующим образом:

найти

$$\min z = \sum_{i=1}^N \sum_{j=1}^J t_{ij}\theta_{ij} \quad (9)$$

при ограничениях

$$\sum_{i=1}^N \sum_{j=1}^J l_{ijt}\theta_{ij} \leq (rm)_t, \quad t = 1, \dots, T, \quad (9a)$$

$$\sum_{j=1}^J \theta_{ij} = 1, \quad i = 1, \dots, N, \quad (9b)$$

$$\theta_{ij} \geq 0, \quad i = 1, \dots, N, \quad j = 1, \dots, J,$$

где J — общее количество доминирующих последовательностей производственных операций, θ_{ij} — часть объема производства продукции i -го типа, выполняемого в течение i -го интервала времени для j -й последовательности.

Минимизация целевой функции (9) означает минимизацию суммарных затрат, включающих производственные затраты, затраты на переналадку и хранение запасов. Модель можно расширить путем включения в нее затрат на содержание штатных сотрудников, на оплату сверхурочной работы, потери из-за дефицита и, наконец, затраты, связанные с увольнением и приемом.

Ограничение (9а) устанавливает для каждого интервала времени соотношение между имеющимся в наличии и необходимым количеством трудовых ресурсов. Можно ввести ограничения с учетом нескольких типов производственных ресурсов и включить в качестве управляющей переменной фонд рабочей силы с допустимым объемом сверхурочной работы [20, 30].

Ограничение (9б) устанавливает, что для продукции i -го типа сумма θ_{ij} по всем $j \in \tau$ должна быть равна единице и переменные модели не должны быть отрицательными.

По-видимому, было бы естественно потребовать, чтобы θ_{ij} для данной модели были целыми числами, а именно 0 или 1. К сожалению, введение такого ограничения делает крайне трудным решение задачи, хотя сколь угодно близко реальной размерности. Более того, может оказаться, что среди целых θ_{ij} не существует оптимального решения задачи (9) в «чистых стратегиях», а можно получить только субоптимальное решение. Это связано с ограничениями на имеющиеся производственные мощности. (Следует заметить, что такой существенный фактор не учитывался многими исследователями.) Однако, решая задачу линейного программирования, определенную условиями (9) — (9б), обычно получают хорошее приближение к желаемому решению. Действительно, поскольку эта модель включает $T + N$ ограничений, то в оптимальном решении задачи линейного программирования не менее чем $T + N$ переменных будут иметь положительное значение. В то же время по крайней мере одна из этих переменных будет входить в каждое из N ограничений (9б). Поэтому может быть не более чем T случаев, когда переменная с положительным значением включена больше чем один раз в ограничения (9б). Очевидно, что если для продукции i -го типа только одна переменная θ_{ij} положительна, то в соответствии с ограничениями (9б) ее значение должно быть равно 1. Следовательно, всякий раз, когда N намного больше T (случай, который довольно часто имеет место в реальной ситуации), будет всего лишь несколько нецелочисленных значений θ_{ij} , которые могут быть округлены либо до 0, либо до 1. Вносимая при этом погрешность обычно является незначительной.

Как уже отмечалось, одна из модификаций модели может включать ограничения по фонду трудовых ресурсов, а также по любому числу K дефицитных ресурсов. В этом случае расщепление последовательностей производственных операций становится несущественным, если N — число планируемых к выпуску продуктов значительно превышает величину $K \times T$. В практических ситуациях это условие обычно выполняется.

Однако, даже если не принимать во внимание условие целочисленности задачи относительно переменных θ_{ij} , решить сформулированную задачу линейного программирования обычными методами довольно трудно. В некоторых ситуациях приходится планировать производство нескольких тысяч изделий. Получить решение с помощью симплекс-метода с таким количеством строк

не представляется возможным. Кроме того, изготовление каждого изделия предусматривает осуществление $2T^{-1}$ доминирующих последовательностей производственных операций. Если $T = 12$, то в модели будет 2048 переменных для каждого изделия. Следовательно, если модель предназначена для выпуска 1000 изделий в течение 12 интервалов времени, то она включает более двух миллионов переменных θ_{ij} . Для преодоления такого рода трудностей Дзелинский и Гомори [20] предложили использовать метод декомпозиции Данцига — Вульфа [17], согласно которому каждая из подзадач сводится к задаче определения размера партии без ограничения на его величину, т. е. к задаче, аналогичной рассмотренной Вагнером и Уайтином. Эти достаточно просто решаемые подзадачи используются для формирования вводимых в базис улучшающих последовательностей производства без необходимости каждый раз определять с самого начала все базисные значения θ_{ij} . Однако метод декомпозиции для задачи этого типа имеет один существенный недостаток. Хорошо известно, что метод декомпозиции позволяет относительно быстро получить решение, близкое к оптимальному. Для получения же оптимального решения требуется выполнить еще достаточно большое число шагов. Однако во многих практических применениях получение строго оптимального решения не является обязательным.

Для определения момента прекращения процесса решения могут использоваться нижние границы значения целевой функции путем сравнения с ними значений целевой функции текущего допустимого решения. В рассматриваемой нами задаче необходимо получить оптимальное решение. Это вызвано тем, что только в случае достижения оптимального решения удовлетворяются требования целочисленности θ_{ij} и тем самым находится допустимое решение исходной задачи.

Для устранения этого недостатка Лэддон и Терьюнг [57] использовали процедуру формирования столбцов, предложенную Дзелинским и Гомори. Однако вместо построения координирующей задачи они решили исходную задачу линейного программирования с помощью обобщенного метода верхних границ, предложенного Данцигом и Ван Слайком [16], используя тем самым специфику структуры начальной модели.

Поясним, в чем состоит процедура генерации столбцов. Пусть π_t ($t = 1, \dots, T$) — множество двойственных переменных, соответствующих ограничениям (9а), и π_{T+1} ($i = 1, \dots, N$) — множество двойственных переменных, соответствующих ограничениям (9б). Приведенные затраты, соответствующие задаче (9), определяются выражением

$$\overline{t_{ij}} = t_{ij} - \sum_{t=1}^T \pi_t l_{ijt} - \pi_{T+1} i.$$

Для выбора вводимой в базис переменной в соответствии с симплекс-методом необходимо найти

$$\min_i \min_j \overline{t_{ij}}.$$

Подставим в выражение для $\overline{t_{ij}}$ вместо t_{ij} и l_{ijt} соответствующие им введенные выше функции. После простых преобразований можно минимизацию по j представить в виде

$$\min_j \left\{ \sum_{t=1}^T [(s_{it} - \pi_i a_{it}) \delta(X_{ijt}) + (v_{ij} - \pi_i k_i) X_{ijt} + C_{it} I_{it}] \right\}. \quad (10)$$

Так как $\pi_i = 0$, то все коэффициенты в этом выражении положительны.

Теперь задача заключается в минимизации общих затрат, содержащих затраты на наладочные работы, производство и хранение запасов, при условии что объем производства X_{ijt} удовлетворяет спрос на продукцию i -го типа в пределах периода планирования.

Сформулированная задача является задачей определения размера партии изделий при отсутствии ограничений на ее величину. Она может быть решена методом динамического программирования, предложенным Вагнером и Уайтином [90] и Вагнером [89]. В практических ситуациях нет необходимости осуществлять перебор последовательностей, поскольку метод Вагнера — Уайтина обеспечивает получение X_{ijt} оптимального плана производства продукции i -типа в каждый t -й интервал времени.

Для того чтобы определить, какой столбец необходимо включить в базис, вычтем величину $\pi_{T+1,i}$ из оптимального значения выражения (10) для каждой единицы продукции i -го типа. Номер единицы продукции i -го типа, которому соответствует минимальное значение разности, определяет столбец, который необходимо ввести в базис.

Таким образом формулируется новая задача линейного программирования, решение которой находят с помощью стандартного обобщенного метода верхних границ.

Горенштейн [30] использовал аналогичную модель при подготовке долгосрочных плановых решений для компании по производству шин. Кроме того, он связал выходы этой модели с краткосрочным планом — графиком, а также ввел отношение предшествования в производстве полуфабрикатов, готовых шин и их компонент. В работе [53] были предложены иные подходы к решению задачи определения размера партии с ограничением на ее величину и с постоянными затратами.

Метод Ласдона и Терьюнга применим только в случае, когда число изделий N намного больше, чем число интервалов времени T , составляющих период планирования. Это вызвано тем, что

предложенный метод основан на непрерывной аппроксимации задачи целочисленного программирования. Для исключения отмеченного недостатка Ньюсон [68] предложил эвристическую процедуру, которая не зависит от способа формирования столбцов и интерпретирует задачу определения размера партии как задачу поиска кратчайшего пути.

В модели Ньюсона фонд рабочей силы считается переменной величиной. При использовании введенных ранее обозначений задача может быть сформулирована следующим образом:

найти

$$\min z = \sum_{i=1}^N \sum_{t=1}^T [s_{it}\delta(X_{it}) + v_{it}X_{it} + C_{it}I_{it}] + \\ + \sum_{t=1}^T (r_t W_t + o_t O_t + h_t H_t + f_t F_t)$$

при ограничениях

$$X_{it} + I_{i,t-1} - I_{it} = d_{it}, \quad i = 1, \dots, N; \quad t = 1, \dots, T,$$

$$\sum_{i=1}^N [a_i \delta(X_{it}) + k_i X_{it}] - W_t - O_t \leq 0, \quad t = 1, \dots, T,$$

$$W_t - W_{t-1} - H_t + F_t = 0, \quad t = 1, \dots, T,$$

$$-pW_t + O_t \leq 0, \quad t = 1, \dots, T,$$

$$X_{it}, I_{it} \geq 0, \quad i = 1, \dots, N; \quad t = 1, \dots, T,$$

$$W_t, O_t, H_t, F_t \geq 0, \quad t = 1, \dots, T,$$

где

$$\delta(X_{it}) = \begin{cases} 0, & \text{если } X_{it} > 0, \\ 1, & \text{если } X_{it} = 0. \end{cases}$$

Интерпретация этой модели очевидна. В модель легко могут быть включены затраты, связанные с невыполнением заказов, с помощью приемов, использованных при рассмотрении модели с линейными функциями затрат и переменным фондом рабочей силы.

Метод поиска решения в этой модели аналогичен методу, использованному в модели определения размера партии с постоянным фондом рабочей силы, т. е. в тех случаях, когда простои, связанные с перепаладками машин, незначительны, рассматривается модель с постоянными затратами, в противном случае может быть использовано приближение с помощью модели линейного программирования, предложенное Дзелинским и Гомори или Лэсдоном и Терьюнгом.

Ньюсон [68] предложил решать эту задачу в два этапа. На первом этапе отбрасывается детализированный график производства каждого типа для всего периода планирования без учета имеющегося ограничения на фонд трудовых ресурсов. Для продукции данного i -типа задача на этом этапе может быть сформулирована следующим образом:

найти

$$\min z_i = \sum_{t=1}^T [s_{it} \delta(X_{it}) + v_{it} X_{it} + C_{it} I_{it}]$$

при ограничениях

$$\begin{aligned} X_{it} + I_{i,t-1} - I_{it} &= d_{it}, \quad t = 1, \dots, T, \\ X_{it}, I_{it} &\geq 0, \quad t = 1, \dots, T. \end{aligned}$$

Решение задачи первого этапа позволяет определить объем трудовых ресурсов, необходимых для выполнения подробного графика производства для каждого t -го интервала по всем типам продукции

$$\hat{P}_t = \sum_{i=1}^N [a_i \delta(X_{it}) + k_i X_{it}], \quad t = 1, \dots, T.$$

На втором этапе принимаются агрегированные решения об используемых трудовых ресурсах. На этом этапе задача может быть сформулирована следующим образом:

найти

$$\min z(\hat{P}) = \sum_{t=1}^T (r_t W_t + o_t O_t + h_t H_t + f_t F_t)$$

при ограничениях

$$\begin{aligned} W_t + O_t - \hat{P}_t &\geq 0, \quad t = 1, \dots, T, \\ W_t - W_{t-1} - H_t + F_t &= 0, \quad t = 1, \dots, T, \\ -pW_t + O_t &\leq 0, \quad t = 1, \dots, T, \\ W_t, O_t, H_t, F_t &\geq 0, \quad t = 1, \dots, T. \end{aligned}$$

Ньюсон предложил эвристическую итеративную процедуру, которая связывает последовательно модели двух этапов. Эта процедура выполняется до тех пор, пока полученное решение не будет удовлетворять некоторым конечным показателям качества.

3.3. Преимущества и недостатки моделей для определения размера партии продукции (изделий)

Основным преимуществом рассмотренных моделей является возможность в процессе планирования одновременно решать задачу составления графиков производства и задачу определения размера неделимой партии продукции. Однако такой подход требует большого объема детальной информации на протяжении всего периода планирования, что в свою очередь связано со значительными затратами. В качестве возможного подхода, обеспечивающего взаимодействие агрегированного планирования и процесса составления детальных графиков производства, может служить построение иерархических систем планирования, рассмотренных в работе [37].

4. МОДЕЛИ С ПРОИЗВОЛЬНОЙ ФУНКЦИЕЙ ЗАТРАТ

Несмотря на то что рассмотренные линейные, квадратичные модели планирования производства и модели для определения размера партии продукции находят широкое практическое применение, вследствие серьезных ограничений на вид функции затрат они не адекватны реальным системам. Это объясняется тем, что функции затрат в практических производственных ситуациях имеют нелинейный характер и не являются непрерывными. Согласно работе [11], основными причинами, обуславливающими такое поведение функции затрат, являются взаимодействие между спросом и поставками, стратегии производства или закупок, эффект обучения, скачкообразные изменения в издержках в связи с появлением новых возможностей в производстве, технологические изменения и изменения объемов выпуска, спад темпа работы.

Было предложено несколько подходов к решению задач агрегированного планирования, а соответственно и моделей, которые претендовали на более достоверное отображение реальных сложных ситуаций, возникающих в процессе принятия конкретных решений по управлению. Однако, вообще говоря, нет никакой уверенности, что при использовании даже таких подходов будет найдено оптимальное решение. В первом приближении предложенные подходы можно разделить на следующие классы:

- нелинейные аналитические модели, позволяющие математически выразить в общем виде нелинейную структуру функции затрат;

- эвристические решающие правила, целью которых является обогащение интуиции лица, принимающего решения, путем введения практических правил, способствующих получению решения стоящей перед ним задачи;

— поисковые решающие правила, которые состоят в применении методов крутого восхождения по поверхности отклика, определяемой нелинейной функцией затрат и ограничениями задачи;

— решающие правила, основанные на имитационном моделировании, с помощью которых рассматриваемая задача может быть представлена в виде набора программированных инструкций. Лицо, принимающее решение, имеет возможность итеративным путем проверить эффективность различных подходов. При этом результаты каждой итерации позволяют определить способ выбора последующей итерации. Имитационное моделирование, в частности, пригодно для учета неопределенностей, которые могут присутствовать в решении. Рассмотрим каждый из перечисленных подходов.

4.1. Нелинейные аналитические модели

В течение последних 20 лет вопросам аналитической разработки моделей планирования производства с нелинейной зависимостью затрат общего вида от переменных модели было посвящено значительное количество работ. Во многих из этих работ авторы пытались использовать для решения задач многоэтапного агрегированного планирования производства принципы динамического программирования. Однако вследствие сложности этих задач, а также большого объема вычислений, присущего методу динамического программирования, существующие модели в редких случаях могут быть использованы для решения общей задачи на каждом этапе планирования. Тем не менее они оказываются полезными при исследовании качественных свойств оптимальных решений.

В данном разделе рассматриваются в основном две модели однопродуктовых систем. В модели первого типа целевая функция не учитывает изменения объема производства от периода к периоду. Задача может быть сформулирована следующим образом:

найти

$$\min z = \sum_{t=1}^T [v_t(X_t) + c_t(I_t)] \quad (11)$$

при ограничениях

$$X_t + I_{t-1} - I_t = d_t, \quad t = 1, \dots, T, \quad (11a)$$

$$I_t \leq \bar{I}_t, \quad t = 1, \dots, T, \quad (11б)$$

$$\underline{X}_t \leq X_t \leq \bar{X}_t, \quad t = 1, \dots, T, \quad (11в)$$

где $v(X)$ и $c(I)$ — соответственно нелинейные функции затрат на производство и содержание запасов, $\underline{I}_t$, $\bar{I}_t$ и $\underline{X}_t$, $\bar{X}_t$ — соответ-

ственно нижние и верхние пределы общего уровня запасов и производимой продукции в течение t -го интервала времени. Если предположить, что $I_t = 0$ для всех t , то исключается возможность невыполнения заказов.

Оптимальное решение может быть получено с помощью обычного метода динамического программирования.

Модели второго типа учитывают затраты, связанные с изменением объемов производства и зависящие от уровня производства в течение заданного и предшествующего интервалов времени. В этом случае целевая функция имеет вид

$$z = \sum_{t=1}^T [v_t(X_t) + c_t(I_t) + p_t(X_{t-1}, X_t)].$$

Значения функции затрат, связанных с изменением объемов производства p , предполагаются равными нулю при $X_{t-1} = X_t$ и неотрицательными в остальных случаях.

Для характеристики этих издержек был предложен приводимый ниже набор функциональных зависимостей [44]:

$$p_t(X_{t-1}, X_t) = p_t |X_t - X_{t-1}|, \quad (12)$$

$$p_t(X_{t-1}, X_t) = p_t(X_t - X_{t-1})^+ + p'_t(X_t - X_{t-1})^-, \quad (13)$$

$$p_t(X_{t-1}, X_t) = p_t(X_t - X_{t-1})^2, \quad (14)$$

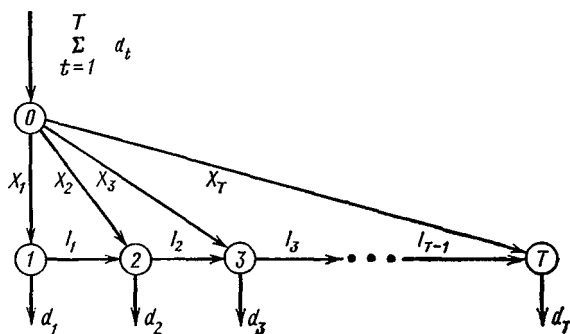
$$p_t(X_{t-1}, X_t) = \begin{cases} p_t, & \text{если } X_t > 0 \text{ и } X_{t-1} = 0. \\ 0 & \text{в остальных случаях,} \end{cases} \quad (15)$$

$$p_t(X_{t-1}, X_t) = \begin{cases} p_t, & \text{если } X_t > 0 \text{ и } X_{t-1} = 0, \\ p'_t, & \text{если } X_t = 0 \text{ и } X_{t-1} > 0 \\ 0 & \text{в остальных случаях.} \end{cases} \quad (16)$$

Функции (12) и (13) могут рассматриваться как линейные функции [33]. Выражение (14) является квадратичной функцией затрат; соотношение (15) определяет модель с дополнительными начальными затратами, связанными с пуском производства продукции в течение t -го интервала времени; соотношение (16) включает дополнительные начальные затраты, связанные как с запуском, так и с прекращением производства в течение t -го интервала времени [76, 96].

Для получения решения в таких моделях невозможно использовать обычный метод динамического программирования, так как имеются две переменные состояния X_t и I_t .

Зангвилл [98] предложил, а затем и обобщил на случай невыполнения заказов следующую сетевую форму записи ограничений (11a—11в):



где $1, 2, \dots, T$ представляют различные интервалы времени, на которые разбит период планирования, 0 — фиктивный источник, обеспечивающий выполнение условия баланса потока.

Используя аналогию с сетевыми моделями, авторы работ [48, 86, 97, 98] получили ряд важных результатов в решении задачи планирования производства.

Для более эффективного обсуждения свойств нелинейных аналитических моделей разобьем их на три основных класса: модели с выпуклыми функциями затрат, модели с вогнутыми функциями затрат и модели оптимального управления с обратной связью. Модели с выпуклыми функциями затрат. На практике получение решения в моделях с выпуклыми функциями затрат не вызывает трудностей. Эти функции могут быть с любой степенью точности аппроксимированы кусочно-линейными функциями, и решение задачи планирования производства может быть получено с помощью моделей линейного программирования. В то же время непосредственное аналитическое исследование моделей с выпуклыми функциями затрат позволяет лучше понимать природу оптимальных решающих правил.

Вейнотт [84] рассмотрел задачу оптимального планирования производства продукции одного типа в течение конечного числа интервалов, целью которой является минимизация выпуклой функции общих затрат на производство и хранение запасов (11) при ограничениях (11a — 11в).

Если общая функция затрат строго выпукла, то существует единственный оптимальный план производства. Вейнотт провел анализ изменений показателей оптимального плана производства в зависимости от колебаний спроса, уровня запасов и предельных

значений объемов производства. На основании этого анализа можно сделать следующие выводы.

1. Оптимальный уровень производства в течение заданного интервала времени является неубывающей функцией следующих факторов:

— спроса в течение любого периода планирования ($d_1, d_2, \dots, d_T$);

— верхнего и нижнего предельных уровней производства в течение заданного интервала времени ($\underline{X}_t, \bar{X}_t$);

— верхнего и нижнего предельных уровней запасов в течение заданного интервала времени и всех последующих интервалов периода планирования ($\bar{I}_k, \underline{I}_k$ для $k = t, t + 1, \dots, T$).

2. Оптимальный уровень производства в течение заданного интервала времени является невозрастающей функцией:

— верхнего и нижнего предельных уровней производства в любом из интервалов периода планирования, отличных от заданного ($\underline{X}_k, \bar{X}_k$ для $k = 1, \dots, t - 1, t + 1, \dots, T$);

— верхнего и нижнего предельных уровней запасов в течение любого интервала, предшествующего заданному ($\bar{I}_k, \underline{I}_k$ для $k = 1, 2, \dots, t - 1$).

Вейнотт использовал эти результаты для разработки простых, эвристических (основанных на интуиции) вычислительных процедур решения задачи для различных значений показателей оптимального плана производства. Каруш [49] для решения рассматриваемой задачи предложил использовать метод динамического программирования.

Джонсон [45] рассмотрел частный случай данной задачи, когда невыполнение заказов является недопустимым, ограничения на предельные значения уровня запасов отсутствуют, функция затрат на хранение запасов линейна, и предложил относительно простое правило получения оптимального решения: *потребности следует удовлетворять наиболее дешевым способом по мере наступления соответствующих сроков.*

Модильяни и Хон [64] исследовали особый случай задачи агрегированного планирования, когда функция затрат производства является выпуклой и неубывающей, а функция затрат на хранение запасов — линейной, при этом ограничения на предельные значения уровней производства запасов отсутствуют. Кроме того, они предположили, что в каждом интервале периода планирования затраты производства остаются неизменными.

В этом случае задача может быть сформулирована следующим образом:

найти

$$\min z = \sum_{t=1}^T [v(X_t) + cI_t] \quad (17)$$

при ограничениях

$$\begin{aligned} X_t + I_{t-1} - I_t &= d_t, \quad t = 1, \dots, T, \\ I_t &\geq 0, \quad X_t \geq 0, \quad t = 1, \dots, T. \end{aligned} \quad (18)$$

Производная функции $v(X_t)$ предполагается неотрицательной, монотонно возрастающей и непрерывной.

Для описания свойств оптимального решения удобно представить совокупный план производства $P_t = \sum_{k=1}^t X_k$ и совокупный спрос $D_t = \sum_{k=1}^t d_k$ в виде кусочно-л.л. ейных функций времени, где

K_t есть верхняя выпуклая оболочка совокупного спроса D_t и $P_0 = D_0 = K_0 = 0$.

1. Если через P_t^* обозначить оптимальный план производства, то можно показать, что $P_T^* = D_T$.

2. Если $K_t = D_t$, то $P_t^* = D_t$ для любого $t = 1, \dots, T$. Интервалы времени, для которых это условие выполняется, называются периодами планирования.

3. Если D_t выпуклая функция, то $P_t^* = D_t$, $1 \leq t \leq T$.

4. $D_t^* \leq K_t$, $t = 1, \dots, T$.

5. Если D_t выпуклая функция, то P_t^* также выпуклая функция.

6. Основным решением задачи является набор последовательностей производственных операций X_t , минимизирующий целевую функцию (17) только при ограничении (18). Основное решение удовлетворяет следующему условию, накладываемому на предельные затраты: $v'(X_{t+1}) = v'(X_1) + tc$.

Модильяни и Хон предложили алгоритм решения задачи, который основан на использовании основных решений и может быть реализован графически. Наиболее важным результатом, полученным этими авторами, является качественное определение свойств периода планирования. Они показали, что общий интервал планирования может быть разбит на подынтервалы, определяемые периодом планирования. Внутри каждого из этих подынтервалов оптимальный план не зависит от уровня спроса и затрат в течение других периодов. Кроме того, если затраты на хранение запасов незначительны, то оптимальным является план, при котором внутри каждого подынтервала уровень производства сохраняется постоянным. Это позволяет сделать важные практические выводы. В тех случаях, когда сильная сезонность сбыта

влияет на принятие решений по управлению, а также когда в течение следующего сезонного цикла уровень сбыта не увеличивается, целесообразно, чтобы при планировании производства период планирования не выходил за периоды высокого уровня сбыта текущего сезонного цикла. Если же выбранный период планирования превосходит текущий цикл, то такое расширение, вероятно, будет касаться и всех последующих циклов.

Чарнес, Купер и Меллон [13] распространил результаты Модильяни и Хона на случай несколько более общей структуры производственных затрат. Они показали, что в модель может быть включен более чем один тип продукции при условии, что вместо переменной объема выпуска будет использовано приемлемое отображение общих затрат (например, трудозатраты).

Клейн [51] использовал результаты работы [38], чтобы ввести в модель кусочно-линейную функцию затрат, связанных с флуктуациями уровня производства в течение различных периодов. В более ранней работе Клейн [50] сделал ряд замечаний относительно общего вида оптимальных графиков производства при различных условиях, накладываемых на затраты и спрос. Липпман и др. [59] изучили вид оптимальных решений задачи агрегированного планирования производства продукции одного типа в предположении, что функция издержек производства является выпуклой и гладкой, а функция затрат на хранение запасов — возрастающей.

Авторы работы [60] предложили вычислительный алгоритм для случая, когда функции затрат линейны, а спрос либо монотонно возрастающий, либо монотонно убывающий. Юан [92] развил предложенный выше алгоритм на случай выпуклой функции затрат производства при любом спросе.

Модели с вогнутыми функциями затрат. Вогнутость функций затрат может быть следствием затрат на наладочные работы, дисконтирования и изменения уровня производительности в процессе производства. Изучению проблем планирования производства при наличии вогнутых функций затрат уделялось большое внимание, поскольку на практике ситуации, приводящие к моделям такого типа, встречаются довольно часто.

Рассмотрим сначала задачу планирования производства продукции одного типа с вогнутой функцией затрат при условии, что не зависящие от спроса ограничения на верхние предельные уровни производства и запасов отсутствуют и не учитываются затраты, связанные с изменением объема производства.

Задача может быть сформулирована следующим образом:
найти

$$\min z = \sum_{t=1}^T [v_t(X_t) + c_t(I_t)]$$

при ограничениях

$$X_t + I_{t-1} - I_t = d_t, \quad t = 1, \dots, T,$$

$$X_t \geq 0, \quad t = 1, \dots, T.$$

Если невыполнение заказов является недопустимым, то следует ввести дополнительное ограничение $I_t \geq 0, t = 1, \dots, T$.

Зангвилл [95] показал, что без ограничения общности можно принять $I_0 = I_T = 0$. Для решения задачи было предложено несколько алгоритмов, в которых использовались эти предельные уровни запасов.

Минимум вогнутой целевой функции при линейных ограничениях находится в граничной точке выпуклого множества, определяемого линейными ограничениями. Поэтому минимум целевой функции может быть найден путем перебора всех граничных точек выпуклого множества, что, вообще говоря, осуществимо лишь в редких случаях. Однако для построения эффективных алгоритмов решения задачи, основанных на методе динамического программирования, может быть использован ряд важных свойств ее оптимального решения.

В тех случаях, когда невыполнение заказов недопустимо, объем производства в течение каждого заданного интервала времени и уровень запасов в течение каждого предшествующего заданному интервала должны удовлетворять условию [90] $X_t I_{t-1} = 0, t = 1, \dots, T$.

Этот результат был распространен на случай невыполнения заказов [83, 98]. Полученное оптимальное решение удовлетворяло следующим условиям: 1) если $I_{t-1}^+ > 0$, то $X_t = 0$; 2) если $X_t > 0$, то $I_t^- = 0$; 3) если $I_{t-1}^+ > 0$, то $I_t^- = 0$. Свойства оптимального решения, а также представления о доминирующих последовательностях и периоде планирования сыграли основную роль при анализе моделей с вогнутой функцией издержек.

Особо важным классом задач планирования производства с вогнутыми функциями издержек являются задачи определения размера партии изделий, в которых целевая функция включает затраты на наладочные работы и линейные переменные затрат на производство и хранение запасов. Как было показано в работе [44], прямой и обратные алгоритмы динамического программирования, используемые для решения задачи, могут быть распространены на случай вогнутой функции общего вида.

Для решения задачи агрегированного планирования производства продукции данного типа с наличием или отсутствием невыполненных заказов Зангвилл [95] предложил обратный алгоритм динамического программирования. При этом предполагается, что задолженность по выполнению заказов должна быть устранена

не более чем через α периодов после назначенного срока, т. е.

$$I_t \geq - \sum_{k=t-\alpha+1}^t d_k,$$

где $d_k = 0$ для $k \leq 0$. Если $\alpha = 0$, то невыполнение заказов является недопустимым. Если $\alpha > T$, то срок задержки с выполнением заказов становится неограниченным, поскольку период планирования сверху ограничен T интервалами времени.

Зангвилл рассмотрел также случай, когда функция затрат производства вогнута, а функция затрат на хранение запасов — кусочно-вогнутая, допуская разрыв последней при нулевом уровне запасов. Предположение о разрыве функции затрат позволяет строго разграничить затраты, связанные с хранением запасов, от затрат, связанных с невыполнением заказов. Результаты были рассмотрены Зангвиллом [94] для случаев последовательного и параллельного соединения агрегатов. Впоследствии Зангвилл [98] и Калиман [48] интерпретировали задачу определения размера партии изделий при наличии множества производственных единиц, как сетевую задачу. Калиман, в частности, исследовал случай древовидной структуры производственной системы, в которой каждая производственная единица связана по входу только с одной непосредственно предшествующей ей единицей. Зангвилл проанализировал возможность использования сетевых методов для однопродуктовых моделей с вогнутой функцией затрат, которые могут быть также распространены на случай моделей многопродуктовой системы с одним источником. Вейнотт [85], используя характеристики граничных точек моделей замен Леонтьева, разработал единую теорию решения задач агрегированного планирования продукции одного типа с вогнутой функцией затрат и обобщил результаты, полученные Зангвиллом, на системы с древовидной многоступенчатой структурой. Добавление в модель с вогнутой функцией затрат ограничений вида $X_t \leq \bar{X}_t$, $t = 1, \dots, T$, создает серьезные вычислительные трудности при использовании алгоритмов, основанных на методе динамического программирования. Эти трудности обусловлены тем, что вводимые ограничения разрушают структуру оптимального решения.

Флорнан и Клейн [25] исследовали модели однопродуктовых систем с вогнутой функцией затрат и ограничениями на объем производства, предполагая при этом, что невыполненные заказы либо отсутствуют, либо их выполнение должно занять не более чем α интервалов. Они проанализировали возможность доработки алгоритма поиска кратчайшего пути, основанного на методе динамического программирования, для решения задачи при одних и тех же ограничениях на объем производства в каждом интервале периода планирования.

Флориан и Робиллард [26] для решения сетевой задачи с ограничениями пропускных способностей дуг с вогнутой функцией затрат предложили вычислительную процедуру, основанную на методе ветвей и границ, которая может быть применена для решения рассматриваемой задачи агрегированного планирования производства.

Ряд работ был посвящен изучению задачи планирования производства продукции одного типа при условии, что функция затрат на хранение — вогнутая, функция затрат от уровня запасов — вогнутая и функции затрат, обусловленные изменением уровня производства от интервала к интервалу, — кусочно-вогнутая. Заигвилл [96] проанализировал случай невыполнения заказов при спросе, не убывающем от интервала к интервалу, т. е. изучил свойства оптимального решения задачи и предложил алгоритмы для частных видов функций затрат. В работе [76] выполнен подобный анализ задачи для случая, когда функции затрат не имеют разрывов в моменты запуска и окончания производства.

Модели оптимального управления с обратной связью. Аналитические исследования и использование принципа максимума Понтрягина [71], теории обратной связи и следящих систем [28, 39, 74] позволили обобщить задачу планирования производства на случай непрерывного производства и определить условия устойчивости решающих правил.

Для агрегированного планирования производства было предложено несколько подходов. Как и при рассмотрении предыдущих аналитических моделей эти подходы, вообще говоря, не приводят к практичным вычислительным процедурам и требуют высокого уровня агрегирования (обычно приводящего к модели однопродуктовой системы). Полезность подобных моделей главным образом состоит в том, что они позволяют определить структуру оптимальных решений. Модели с непрерывным временем оказываются наиболее эффективными в тех случаях, когда требуется непрерывное управление с постоянным контролем за реакцией системы.

Одна из первых работ по исследованию моделей непрерывного производства была выполнена Эрроу, Карлином и Скарфом [5]. Авторы сформулировали задачу планирования непрерывного производства продукции и хранения запасов следующим образом:

найти

$$\min \int_0^t \{v[X(t)] + c[I(t)]\} dt$$

при ограничении

$$I(t) = I(0) + \int_0^t X(t) dt - \int_0^t d(t) dt$$

или

$$\frac{dI(t)}{dt} = X(t) - d(t),$$

где $I(t)$ — уровень запасов в момент времени t , $X(t)$ — объем производства в момент времени t , $d(t)$ — спрос в момент времени t , $v(t)$ — затраты, связанные с производством X единиц продукции; $c(I)$ — затраты, связанные с обеспечением уровня запасов I .

Если дефицит считается недопустимым, то должно быть введено дополнительное ограничение $I(t) \geq 0$.

Эрроу, Карлин и Скарф исследовали эту задачу в предположении, что функция затрат на хранение запасов линейна, а затраты производства пропорциональны скорости изменения уровня производства при его возрастании и равны нулю при его уменьшении. Кроме того, они проанализировали свойства оптимального решения при различных видах функции спроса.

Хванг и Фэн [42], а также Хванг, Фэн и Эрикссон [43] применили принцип максимума Понтрягина для решения задачи планирования непрерывного производства. В этом случае оптимальные управляющие воздействия соответствуют оптимальному уровню производства, а переменные состояния — оптимальным уровням запасов. Рассмотренная ими задача может быть сформулирована следующим образом:

найти

$$\min \int_0^t \{v[X(t) - X^*]^2 + c[I(t) - I^*]^2\} dt$$

при ограничении

$$\frac{dI(t)}{dt} = X(t) - d(t).$$

В целевой функции затраты производства и затраты на содержание запасов аппроксимированы квадратичными функциями. Переменные X^* и I^* — плановые уровни производства и запасов — предполагаются известными и постоянными. Использование принципа максимума Понтрягина приводит к следующему оптимальному решению задачи:

$$I(t) = A_1 e^{\lambda t} + A_2 e^{-\lambda t} + [I(t)]_p,$$

$$X(t) = A_1 \lambda e^{\lambda t} - A_2 e^{-\lambda t} + \frac{dI(t)_p}{dt} + d(t),$$

где $\lambda^2 = c/v$, а A_1 и A_2 определяются начальными условиями задачи, $[I(t)]_p$ — частное решение, которое определяется видом уравнения и значениями I^* и d [43].

Нельсон [67] использовал принцип максимума Понтрягина для получения необходимых и достаточных условий оптимальности решения задачи распределения трудовых ресурсов в производственных системах с ограниченными ресурсами рабочей силы и основного оборудования.

Следует заметить, что в процессе проектирования и эксплуатации систем планирования непрерывного производства большую роль играют такие понятия, как обратная связь, запаздывание, способы управления, устойчивость производственных систем во времени. Теория следящих систем служит основой для формального анализа этих понятий в системах рассматриваемого класса.

Холт и Саймон [40] и Хэлсман [32] показали, что решения по управлению, полученные путем дифференцирования квадратичных функций затрат, могут приводить к существенной неустойчивости системы. Неустойчивость системы возникает при отсутствии в течение периода планирования идеального прогноза спроса и проявляется в возникновении недопустимых колебаний уровней производства и запасов.

Вассиал [83] предложил в подобных ситуациях вводить стабилизирующую обратную связь, а Холт и Саймон [40, 74] показали, как с помощью неустойчивого правила принятия решений можно выбрать способ определения классов устойчивых правил. Однако, по мнению Хэлсмана, в настоящее время отсутствует общая теория, которая способствовала бы полному пониманию устойчивости предложенного множества правил принятия решений, что особенно ощущается при введении в модель планирования производства внешних прогнозов. По-видимому, в настоящее время единственным методом исчерпывающего исследования степени устойчивости правил принятия решений по определению планов производства является имитационное моделирование [39].

4.2. Эвристические правила принятия решений

Боуман [9] разработал метод агрегированного планирования (метод управляющих коэффициентов), позволяющий при планировании объема производства учесть поведение органов управления. Этот метод основывается на следующих предположениях.

С одной стороны, органы управления стремятся выбрать для определения объема производства, уровня запасов и количества необходимых трудовых ресурсов способ, чувствительный к общему объему затрат и влияющий на принимаемые решения.

С другой стороны, органы управления склонны переоценивать значение повседневных отклонений в своей работе, в результате чего эпизодически принимаются либо дорогостоящие, либо ошибочные решения, которые отличаются от усредненной модели предшествовавшей деятельности этих органов. Поскольку, однако,

большинство функций затрат в окрестности оптимума имеют плоскую форму, то, следовательно, малые отклонения значений переменных от оптимума не приводят к значительному изменению целевой функции.

На основании вышеизложенных соображений Боуман пришел к выводу, что правило принятия решения, использующее среднее значение коэффициентов, полученных с помощью анализа решений по управлению, принятых в предшествующие интервалы времени, должно привести к лучшим результатам, чем обычно используемые правила и правила, полученные в результате аналитических исследований.

Структура искомого правила принятия решения может быть получена на основе либо интуитивных, либо аналитических соображений, либо на основе и тех и других одновременно.

Бауман предложил следующий вариант правила принятия решений:

$$P_t = \sum_{i=1}^{t+T-1} a_i S_i + x(P_{t-1} - S_t) + y(I_N - I_{t-1}),$$

где P_t — планируемый объем производства в течение t -го интервала времени; S_t — прогноз сбыта в течение t -го интервала времени; x, y — сглаживающие постоянные; $0 \leq x \leq 1$, $0 \leq y \leq 1$; I_N — нормальный уровень запаса; I_{t-1} — уровень запаса к концу $(t-1)$ -го интервала времени; a_i — весовой множитель, приписываемый прогнозу сбыта S_i , причем $a_t > a_{t+1} > \dots > a_{t+T-1}$; T — период планирования.

Значения весовых множителей a_i , сглаживающих постоянные x и y , вычисляют с помощью регрессионного анализа управленческой деятельности в предшествующие интервалы времени. Боуман опробовал разработанную им модель в четырех отраслях промышленности и получил обнадеживающие результаты.

4.3. Поисковые правила принятия решений

Джонс [47] объединил эвристический подход, определяющий структуру правила принятия решения, и поисковые методы для вычисления значений коэффициентов, присутствующих в решающих правилах. Он разработал метод агрегированного планирования объема производства, который назвал параметрическим планированием производства. Джонс исходил из существования двух линейных правил принятия решений, одно из которых позволяет определить уровень фонда рабочей силы, а другое — объем производства, причем первое имеет вид $W_t = W_{t-1} + A(W_D - W_{t-1})$, где W_{t-1} — текущий уровень фонда рабочей силы; W_t — планируемый уровень фонда рабочей силы в течение

рассматриваемого интервала времени; W_D — желательный уровень фонда рабочей силы, удовлетворяющий заданный прогнозируемый спрос; A — коэффициент, определяющий подлежащую реализации часть разности между желательным и текущим уровнями фонда рабочей силы.

Величина фонда рабочей силы W_D определяется как взвешенная сумма величин фондов рабочей силы, необходимой для обеспечения в течение периода планирования T сбыта продукции, начиная с t -го интервала времени, т. е.

$$W_D = \sum_{i=1}^T b_i K(S_{t+i-1}),$$

где b_i — весовой множитель, приписываемый показателю прогнозируемого сбыта S_{t+i-1} , $K(S_t)$ — число рабочих, требуемых для производства с минимальными затратами S_t единиц продукции.

Джонс по результатам экспериментов с несколькими весовыми функциями предложил следующее выражение для определения значения весовых множителей b_i :

$$b_i = \frac{B^i}{\sum_{i=1}^T B^i},$$

где B — коэффициент значимости прогноза, принимающий значения от 0 до 1.

Отметим, что все весовые множители b_i , $i = 1, \dots, T$, являются функцией одного параметра B . Кроме того, Джонс включил в структуру правила принятия решений член, который исключает нежелательное сокращение или накопление запасов и имеет вид $b_i K(I_t^* - I_{t-1})$, где I_t^* — заданный оптимальный уровень запасов в конце t -го интервала времени.

В результате правила принятия решений по выбору уровня фонда рабочей силы можно записать как

$$W_t = W_{t-1} + A \left[\sum_{i=1}^T b_i K(S_{t+i-1}) - W_{t-1} + b_1 K(I_t^* - I_{t-1}) \right],$$

а правило принятия решений по определению объема производства —

$$P_t = K^{-1}(W_t) + C \left[\sum_{i=1}^T d_i S_{t+i-1} - K^{-1}(W_t) + d_1 (I_t^* - I_{t-1}) \right],$$

где P_t — объем производства, выраженный в единицах продукции; $K^{-1}(W_t)$ — число единиц продукции, которая производится с минимальными затратами в течение t -го интервала времени W_t .

рабочими; C — коэффициент значимости, для которого вводится шкала оценок от 0 до 1 и который определяет необходимую степень увеличения или уменьшения объема производства; d_i — весовой множитель показателя прогнозируемого сбыта S_{t+i-1} в течение $(t + i - 1)$ -го интервала времени.

Весовые множители d_i определяются по формуле, аналогичной той, которая используется для вычисления весовых множителей b_i , т. е.

$$d_i = \frac{D^{-i}}{\sum_{i=1}^T D^i}.$$

Значения коэффициентов A , B , C и D определяются путем исследования зависимости функции полезности от значений этих коэффициентов. Вид функции полезности зависит от общей структуры затрат, учитываемых при принятии решений по выбору объема производства и фонда рабочей силы.

Существует значительное количество различных методов поиска, пригодных для оптимизации значений коэффициентов [91]. Среди них наиболее перспективным является метод прямого поиска, разработанный Хуком и Дживсом [41]. Джонс предположил, что поверхность отклика, определяемая четырьмя коэффициентами, и соответствующая функция оценки полезности являются плоскими, гладкими и унимодальными функциями.

Другой важный подход к решению задачи агрегированного планирования объема производства был предложен Таубертом [80]. Существенная особенность подхода, предложенного Таубертом, состоит в том, что объем производства, уровни фонда рабочей силы и запасов определяются на основе данных о значениях этих показателей в течение каждого интервала времени периода планирования, в то время как Джонс использует только значения четырех коэффициентов (A , B , C и D). Число операций поиска в подходе Тауберта зависит от числа интервалов, на которые разбит период планирования, что порождает дополнительные вычислительные трудности.

Тауберт предложил комбинированный метод поиска агрегированного плана объема производства, согласно которому с помощью метода ветвей и границ производится разбиение множества допустимых решений на подмножества. На этих подмножествах формируются более простые задачи агрегированного планирования, решения которых находятся с помощью вышеупомянутого метода. Комбинированный метод кажется перспективным, однако его возможности еще недостаточно исследованы. В работе [29] рассматривается еще один простой подход к решению задачи агрегированного планирования объема производства.

4.4. Правила принятия решений, основанные на имитационном моделировании

Имитационное моделирование обычно используется в тех случаях, когда аналитические модели либо слишком сложны, либо слишком упрощают представления о реальном процессе [20, 66].

Верджин [86] предложил имитационную модель, позволяющую учитывать специфику каждого конкретного предприятия. Процесс имитационного моделирования начинается с рассмотрения некоторого начального приемлемого плана, полученного экспериментальным путем, или на основании данных, отражающих существующую ситуацию на предприятии. Целевая функция, на структуру которой никаких ограничений не налагается, используется для определения необходимости изменения показателей полученного плана. В процессе перехода от одного плана к другому изменяются значения таких показателей производства, как уровень занятости, объем сверхурочной работы, уровень запасов, объем выполнения работ по субподрядам и др., причем показатели изменяются до тех пор, пока не будет достигнут локальный минимум целевой функции.

Верджин провел исследование на трех промышленных предприятиях, функционирование которых существенным образом зависит от сезонных колебаний спроса. Полученные результаты свидетельствуют о том, что планы, разработанные с помощью имитационной модели, имеют лучшие показатели, чем обычно используемые планы и планы, составленные с помощью линейных правил принятия решений.

4.5. Преимущества и недостатки моделей с произвольной функцией затрат

Одним из основных преимуществ рассмотренных моделей является возможность более точного учета факторов, влияющих на формирование плана производства, в том числе различного рода неопределенностей, специфики структуры функции затрат и ограничений. Кроме того, эти модели с большей степенью достоверности воспроизводят реальный процесс принятия решения. Это приводит к тому, что результаты подобного моделирования оказываются доступными для понимания лицу, принимающему решение. В результате лицо, принимающее решение, может объяснить существо принимаемых решений и доказать необходимость их утверждения. Однако разработка и эксплуатация этих моделей требует значительных затрат, а вычислительные методы, используемые для решения, редко гарантируют достижение глобального оптимума. Более того, некоторые из этих моделей требуют высокой степени агрегирования, что создает существенные трудности

в случае необходимости получить на более низких уровнях дезагрегированное решение. Следует также иметь в виду, что для рассмотренных моделей при наличии большого числа связанных между собой ограничений отсутствуют эффективные методы решения. Аналитические модели используются для качественных исследований свойств оптимальных решений, и они редко порождают практически используемые алгоритмы.

Ли и Хумаваля [58] сравнили результаты практического применения различных подходов к решению задач агрегированного планирования производства. Они пришли к выводу, что модель, предложенная Таубертом, намного эффективнее, чем модель линейного программирования и модели, разработанные Джонсоном и Боуманом, и предложили метод синтеза моделей агрегированного планирования объема производства.

5. ИНТЕГРАЦИЯ ПРОИЗВОДСТВЕННОГО ПРОЦЕССА

Экономисты определяют производство как процесс, в результате которого создаются товары и становится возможным оказание различных услуг. В более узком смысле производство может быть определено как процесс преобразования сырья или полуфабрикатов в конечный продукт. Понятия «сырье» или «конечный продукт» являются относительными, поскольку то, что является конечным продуктом для одного предприятия, может быть сырьем для другого. Эффективное управление процессом производства должно обеспечивать получение в определенные моменты времени заданного объема готовой продукции требуемого качества при приемлемых затратах.

5.1. Структура производственных решений

Процесс управления производством включает принятие решений, касающихся различных организационных уровней. Для того чтобы понять роль операционного анализа при выработке подобных решений удобно систематизировать их, согласно Антони [1], как решения в области стратегического планирования, тактического планирования и оперативного управления [36].

Стратегическое планирование: планирование производственных мощностей. Целью стратегического планирования, как правило, является выработка стратегии управления и разработка способов обеспечения необходимых предприятию ресурсов, причем эти способы должны соответствовать конкретным целям предприятия.

В области управления производством наиболее важными являются решения по планированию ввода новых и расширению старых производственных мощностей, так как именно они опре-

деляют основные капиталовложения. Кроме того, они включают решения по определению местоположения и размера новых объектов (установок), приобретению нового оборудования, определению производственных участков внутри каждого объекта.

Решения, требующие жесткой координации с системой сбыта, связанны с выбором новых видов готовой продукции и планированием обеспечивающих систем (включая определение местоположения и емкости склада, парка транспортных средств и т. д.). Эти решения играют чрезвычайно важную роль, так как от них в значительной степени зависит конкурентоспособность предприятий. Характерная особенность стратегического планирования состоит в том, что реализация принятых на этом уровне решений рассчитана на длительный срок, т. е. рассматривается продолжительный период планирования. Это приводит к необходимости анализировать влияние неопределенностей и риска на процесс принятия решения, что в свою очередь затрудняет использование моделей математического программирования, которые не позволяют провести подобный анализ. Кроме того, решения по выделению капиталовложений на создание новых и расширение существующих производственных мощностей принимаются на высших административных уровнях с учетом как внешней, так и внутренней по отношению к предприятию информации. Таким образом, стратегическое планирование охватывает широкий круг проблем, для решения которых требуется высокой степени агрегирование обрабатываемой информации, чтобы исключить перегрузку персонала высших административных уровней второстепенной информацией.

Тактическое планирование: агрегированное планирование производства. После определения требуемых производственных мощностей основной задачей является эффективное распределение ресурсов (производственных мощностей, объемов складов, наличной рабочей силы, финансовых и управленческих ресурсов и т. д.). Распределение ресурсов должно быть таким, чтобы удовлетворялись спрос на производимую продукцию и технологические требования. Решения, принимаемые на этой стадии, являются весьма сложными, поскольку приходится учитывать наличие нескольких предприятий, большое количество распределительных центров, региональных и местных складов, производства со сложным многоэтапным процессом изготовления и сборки, подверженного случайным и сезонным колебаниям спроса. Эти решения обычно принимаются на небольшом периоде планирования, разделенном на несколько интервалов, при агрегировании выпускаемой продукции.

Решения, которые принимаются на стадии тактического планирования, касаются способа использования регламентированной и сверхурочной рабочей силы, распределения ограниченных агре-

гированных ресурсов для производства продукции, накопления сезонных запасов, определения способов (путей) распределения и перевозок продукции.

На этапе тактического планирования используются исключительно модели, обеспечивающие составление агрегированного плана производства.

Оперативное управление: детальный план производства. После составления агрегированного плана распределения производственных мощностей принимаются оперативные решения и составляются графики производства на основе полного дезагрегирования информации, поступающей от органов управления более высоких уровней. В результате разрабатываются методы управления, которые реализуются в повседневной деятельности организации.

На стадии оперативного управления принимаются решения по выделению конкретного оборудования для выполнения заказов и упорядочению этих заказов в ходе производственного процесса, по учету запасов товарно-материальных ценностей и управлению запасами, по организации диспетчерского управления, обработке заказов, составлению графиков перевозок и т. д.

5.2. Иерархическая система принятия решений

При реализации рассмотренной в разд. 5.1 структуры принятия производственных решений возникает ряд трудностей. Прежде всего следует иметь в виду, что решение по выбору объема капиталовложений, их размещению и распределению, а также по составлению календарных планов не могут приниматься изолированно друг от друга, поскольку они взаимно влияют друг на друга. Поэтому для получения оптимальных решений необходим интегрированный подход. Однако интегрированный подход не может быть реализован без той или иной декомпозиции задачи в рамках иерархической системы, в которой решения нижних уровней иерархии согласованы с решениями верхних уровней. При этом решения, принятые на верхних уровнях иерархии, накладывают ограничения на решения, принимаемые на нижних уровнях. Такой иерархический подход определяет степень вмешательства органов управления различных уровней, границы действия решения, степень агрегирования необходимой информации, интервалы времени, в течение которых должно быть принято решение. По-видимому, было бы серьезной ошибкой пытаться получить решения всех задач одновременно с помощью одной математической модели. Если бы даже мощность ЭВМ и методологические возможности машинного программирования смогли бы обеспечить решение большой детализированной интегрированной производственной модели (что, очевидно, в настоящее время не реально), этот подход был бы некорректным из-за нечувствитель-

ности решений к потребностям органов управления каждого уровня организации и исключил бы взаимодействие между моделями и органами управления на каждом организационном уровне. Поэтому при создании систем принятия общих решений по управлению производством необходимо выявить возможные способы расчленения процесса принятия решений на несколько этапов, выбрать адекватные модели выработки отдельных решений на каждом уровне иерархии, построить механизмы передачи результатов решений от верхних уровней иерархии к нижним, включающие способы дезагрегирования информации, обеспечить количественную оценку результирующих отклонений от оптимального функционирования организации на каждом уровне иерархии. В ряде работ изложены некоторые соображения относительно реализации иерархического подхода. Хэкс [35] описал применение таких систем для планирования непрерывных производственных процессов, а Хэкс и Мейл [37] указали на полезность иерархических систем при планировании серийного производства. Армстронг и Хэкс [3], а также Швиммер [73] исследовали их применение для планирования деятельности производственных цехов. В работах [45, 77, 81] обсуждается возможность интеграции процесса управления производством с функциями управления, связанными со сферой сбыта, финансами и т. д. (часть этих работ была выполнена для ВМС США согласно контракту N00014-67-A-0204-0076).

ЛИТЕРАТУРА

1. Anthony R. N., *Planning and Control Systems: A Framework for Analysis*, Harvard University Graduate School of Business Administration, Boston, 1965.
2. Antosiewicz H., Hoffman A. J., A Remark on Smoothing Problem. *Management Sci.*, 1, 92—95 (1954).
3. Armstrong R. J., Hax A. C., A Hierarchical Approach for a Naval Tender Job Shop Design. Operations Res. Center, MIT, September 1974.
4. Arrow K. J., Harris T. H., Marschak J., Optimal Inventory Policy, *Econometrica*, 19, 250—272 (1951).
5. Arrow K. J., Karlin S., Scarf H., *Studies in the Mathematical Theory of Inventory and Production*, Standford University Press, 1958.
6. Balinski M. L., Fixed Cost Transportation Problem, *Nav. Res. Log. Quart.*, 8, 41—54 (1961).
7. Bergstrom G. L., Smith B. E., Multi-Item Production Planning—An Extension of the HEMS Rules, *Management Sci.*, 16, B614-B629 (1970).
8. Bowman E. H., Production Scheduling by the Transportation Method of Linear Programming, *Operations Res.*, 4, 100—103 (1956).
9. Bowman E. H., Consistency and Optimality in Decision Making, *Management Sci.*, 9, 310—321 (1963).
10. Brown R. G., *Decision Rules for Inventory Management*, Holt, Rinehart and Winston, N.Y., 1967.
11. Buffa E. S., Taubert W. H., *Production-Inventory Systems: Planning and Control*, Richard D. Irwin, Inc., 1972.

12. Chang R. H., Jones C. M., Production and Workforce Scheduling Extensions, *AIIE Trans.*, 2, 326--333 (1970).
13. Charles A., Cooper W. W., Mellon B., A Model for Optimizing Production by Reference to Cost Surrogates, *Econometrica*, 23, 307--323 (1955).
14. Cooper L., Drebes C., An Approximate Solution Method for the Fixed Charge Problem, *Nav. Res. Log. Quart.*, 14, 101--113 (1967).
15. Damon W. W., Schramm R., A Simultaneous Decision Model for Production, Marketing and Finance, *Management Sci.*, 19, 161--172 (1972).
16. Dantzig G. B., Van Slyke R. M., Generalized Upper Bounding Techniques, *J. Computer and System Sci.*, 1, 213--226 (1967).
17. Dantzig G. B., Wolfe P., Decomposition Principles for Linear Programs, *Operations Res.*, 8, 101--111 (1960).
18. Denzler D. R., An Approximate Algorithm for the Fixed Charge Problem, *Nav. Res. Log. Quart.*, 16, 411--416 (1969).
19. Dzielinski B. P., Baker C. T., Manne A. S., Simulation Tests of Lot Size Programming, *Management Sci.*, 9, 229--258 (1963).
20. Dzielinski B. P., Gomory R. E., Optimal Programming of Lot Sizes, Inventory and Labor Allocations, *Management Sci.*, 11, 874--890 (1965).
21. Elmaghraby S. E., Bawle V. Y., Optimization of Batch Ordering Under a Deterministic Variable Demand, *Management Sci.*, 18, 508--517 (1972).
22. Emshoff J. R., Sisson R. L., Computer Simulation Models, MacMillan, N.Y., 1970.
23. Eppen G. D., Gould F. J., A Lagrangean Application to Production Models, *Operations Res.*, 16, 819--829 (1968).
24. Eppen G. D., Gould F. J., Pashinian B. P., Extensions of the Planning Horizon Theorem in the Dynamic Lot Size Model, *Management Sci.*, 15, 268--277 (1969).
25. Florian M., Klein M., Deterministic Production Planning with Concave Costs and Capacity Constraints, *Management Sci.*, 18, 12--20 (1971).
26. Florian M., Robillard P., An Implicit Enumeration Algorithm for the Concave Network Flow Problem, *Management Sci.*, 18, 184--193 (1971).
27. Ford L. R., Fulkerson D. R., Flows in Networks, Princeton University Press, Princeton, N.Y., 1962. [Имеется перевод: Форд Л., Фалкерсон Д. Поток в сетях. — М.: Мир, 1966.]
28. Forrester J., Industrial Dynamics, MIT Press, 1961. [Имеется перевод: Форрестер Дж. Кибернетика промышленного предприятия: индустриальная динамика. — М.: Прогресс, 1964.]
29. Goodman D. A., A New Approach to Scheduling Aggregate Production and Work Force, *AIIE Trans.*, 5, 135--141 (1973).
30. Gorenstein S., Planning the Production, *Management Sci.*, 17, B72-B81 (1970).
31. Gray P., Exact Solution of the Fixed-Charge Transportation Problem, *Operations Res.*, 19, 1529--1537 (1971).
32. Hansmann F., Operations Research in Production and Inventory Control, Wiley, N.Y., 1962.
33. Hansmann F., Hess S. W., A Linear Programming Approach to Production and Employment Scheduling, *Management Tech.*, 1, 46--51 (1960).
34. Hausman W. H., McClain J. D., A Note on the Bergstrom-Smith Multi-Item Production Planning Model, *Management Sci.*, 17, 783--785.
35. Hax A. C., Integration of Strategic and Tactical Planning in the Aluminum Industry, Operations Research Centre, MIT, Technical Paper O. R. 026-73, September 1973.
36. Hax A. C., The Design of Large Scale Logistics Systems: A Survey and an Approach, Chapter 6 in Modern Trends in Logistics Research, W. H. Marlow, Ed., MIT Press, 1976.
37. Hax A. C., Meal H. C., Hierarchical Integration of Production Planning and Scheduling, in TMS Studies in Management Science, Vol. 1, Logistics, Murray Geisler, Ed., 53--69 (January 1975).

38. Hoffman A. J., Jacobs W., Smooth Patterns of Production, *Management Sci.*, 1, 86—91 (1954).¹
39. Holt C. C., Modigliani F., Muth J. F., Simon H. A., Planning Production Inventories and Work Force, Prentice-Hall, Inc., Englewood Cliffs, N.Y., 1960.
40. Holt C. C., Simon H. A., Optimal Decision Rules for Production and Inventory Control, Proc. Conf. on Operations Res. in Production and Inventory Control, Case Institute of Technology, January 1954.
41. Hooke R., Jeeves T. A., Direct Search Solution of Numerical and Statistical Problems, *J. Assoc. Computing Machinery*, 8, 212—229 (1961).
42. Hwang C. L., Fan T., The Application of the Maximum Principle to Industrial and Management Systems, *J. Indus. Eng.*, 17, 589—593 (1966).
43. Hwang C. L., Fan T., Erickson L. E., Optimum Production Planning by the Maximum Principle, *Management Sci.*, 13, 751—755 (1967).
44. Johnson L. A., Montgomery D. C., Operations Research in Production Planning, Scheduling, and Inventory Control, Wiley, N.Y., 1974.
45. Johnson S. M., Sequential Production Planning Over Time at Minimum Cost, *Management Sci.*, 3, 435—437 (1957).
46. Johnson S. M., Dantzig G. B., A Production Smoothing Problem, Proc. Second Symposium in Lin. Programming, Washington, D.C., January 1955.
47. Jones C. H., Parametric Production Planning, *Management Sci.*, 13, 843—866 (1967).
48. Kalymon B. A., A Decomposition Algorithm for Arborescence Inventory Systems, *Operations Res.*, 20, 860—874 (1972).
49. Karush W., On a Class of Minimum Cost Problems, *Management Sci.*, 4, 136—155 (1958).
50. Klein M., Some Production Planning Problems, *Nav. Res. Log. Quart.*, 4, 269—286 (1957).
51. Klein M., On Production Smoothing, *Management Sci.*, 7, 286—293 (1961).
52. Kolenda J. F., A Comparison of Two Aggregate Planning Models, unpublished master's thesis, Wharton School of Finance and Commerce, The Univ. of Pennsylvania, 1970.
53. Kortanek K. O., Sodaro D., Soyster A. L., Multi-Product Scheduling via a Extreme Point Properties of Linear Programming, *Nav. Res. Log. Quart.*, 15, 287—300 (1968).
54. Krajewski L. J., Mabert V. A., Thompson H. E., Quadratic Inventory Cost Approximations and the Aggregation of Individual Products, *Management Sci.*, 19, 1229—1240 (1973).
55. Kriebel C. H., Coefficient Estimation on Quadratic Programming Models, *Management Sci.*, 13, B437-B486 (1967).
56. Kuhn H. W., Baumol W. J., An Approximate Algorithm for the Fixed-Charge Transportation Problem, *Nav. Res. Log. Quart.*, 9, 1—16 (1962).
57. Lasdon L. S., Terjung R. C., An Efficient Algorithm for Multi-Item Scheduling, *Operations Res.*, 19, 946—969 (1971).
58. Lee W. B., Khumawala B. M., Simulation Testing of Aggregate Production Planning Models in Implementation Methodology, *Management Sci.*, 20, 903—911 (1974).
59. Lippman S. A., Rolfe A. J., Wagner H. M., Yuan J. S. C., Optimal Production Scheduling and Employment Smoothing with Deterministic Demands, *Management Sci.*, 14, 127—158 (1967).
60. Lippmann S. A., Rolfe A. J., Wagner H. M., Algorithm for Optimal Production Scheduling and Employment Smoothing, *Operations Res.*, 15, 1011—1029 (1967).
61. Magee J. F., Boodman D. M., Production Planning and Inventory Control, McGraw-Hill, N.Y., 1967.
62. Manne A. S., Programming of Economic Lot Sizes, *Management Sci.*, 4, 115—135 (1958).

63. McGarrah R. E., Production and Logistics Management, Wiley, N.Y., 1963.
64. Modigliani F., Hohn F. E., Production Planning over Time and the Nature of the Expectation and Planning Horizon, *Econometrica*, 23, 46—66 (1955).
65. Murty K. C., Solving the Fixed Charge Problem by Routing Extreme Points, *Operations Res.*, 16, 268—279 (1968).
66. Naylor T., et al., Computer Simulation Techniques, Wiley, N.Y., 1966.
67. Nelson R. T., Labor Assignment as a Dynamic Control Problem, *Operations Res.*, 14, 369—376 (1966).
68. Newson E. F. P., Lot Size Scheduling to Finite Capacity, unpublished doctoral thesis, Sloan School of Management, MIT, 1971.
69. O'Malley R. L., Elmaghraby S. E., Jeske J. W., An Operational System for Smoothing Batch-Type Production, *Management Sci.*, 12, B433-B449 (1966).
70. Peterson R., Optimal Smoothing of Shipments in Response to Orders, *Management Sci.*, 17, 597—607 (1971).
71. Понягин Л. С., Болтянский В. Г., Гамкрелидзе Р. В., Мищенко Е. Ф. Математическая теория оптимальных процессов.— М.: Наука, 1961.
72. Rousseau J. M., A Cutting Plane Method for the Fixed Cost Problem, unpublished doctoral thesis, Sloan School of Management, MIT, August 1973.
73. Shwimer J., Interaction Between Aggregate and Detailed Scheduling in a Job Shop, unpublished Ph. D. thesis, Sloan School of Management, MIT, 1972.
74. Simon H. A., On the Application of Servomechanism Theory in the Study of Production Control, *Econometrica*, 20, 247—268 (1952).
75. Simon H. A., Dynamic Programming under Uncertainty with a Quadratic Criterion Function, *Econometrica*, 24, 74—81 (1956).
76. Sobel M. J., Smoothing Start-up and Shut-down Costs: Concave Case, *Management Sci.*, 17, 78—91 (1970).
77. Starr M. K., Production Management-Systems and Procedures, Prentice-Hall, Englewood, Cliffs, N. J., 1972.
78. Steinberg D. I., The Fixed Charge Problem, *Nav. Res. Log. Quart.*, 17, 217—237 (1970).
79. Syppkens H. A., Planning for Optimal Plant Capacity, unpublished master's thesis, Sloan School of Management, MIT, 1972.
80. Taubert W. H., A Search Decision Rule for the Aggregate Scheduling Pattern, *Management Sci.*, 14, B343—B359 (1968).
81. Thomas J., Linear Programming Models for Production-Advertising Decisions, *Management Sci.*, 17, B474—B484 (1971).
82. Van dePanne C., Bosje P., Sensitivity Analysis of Lost Coefficient Estimates: The Case of Linear Decision Rules for Employment and Production, *Management Sci.*, 9, 82—107 (1962).
83. Vassian H. J., Application of Discrete Variable Servo Theory to Inventory Control, *Operations Res.*, 3, 272—282.
84. Veinott A. F., Production Planning with Convex Costs: A Parametric Study, *Management Sci.*, 10, 441—460 (1964).
85. Veinott A. F., Minimum Concave-Cost Solution of Leontieff Substitution Models of Multi-Facility Inventory Systems, *Operations Res.*, 17, 262—291 (1969).
86. Vergin R. C., Production Scheduling Under Seasonal Demand, *J. Indust. Eng.*, 7, 260—266 (1966).
87. Von Lanzenaue C. H., Production and Employment and Scheduling in Multistage Production Systems, *Nav. Res. Log. Quart.*, 17, 193—198 (1970).
88. Von Lanzenaue C. H., A Production Scheduling Model by Bivalent Linear Programming, *Management Sci.*, 17, 105—111 (1970).
89. Wagner H. M., A Postscript to Dynamic Problems in the Theory of the Firm, *Nav. Res. Log. Quart.*, 7, 7—12 (1960).

90. Wagner H. M., Whitin T. M., A Dynamic Version of the Economic Lot Size Model, *Management Sci.*, 5, 89—96 (1958).
91. Wilde D. J., Optimum Seeking Methods, Prentice-Hall, Englewood, Cliffs, N.J., 1964. [Имеется перевод: Уайлд Д. Дж., Методы поиска экстремума. — М.: Наука, 1967.]
92. Yuan S. C., Algorithms and Multi-Product Model in Production Scheduling and Employment Smoothing, Technical Report No. 22, NSF-GS-552, Stanford University, Stanford, California, August 1967.
93. Zabel E., Some Generalizations of an Inventory Planning Horizon Theorem, *Management Sci.*, 10, 465—471 (1964).
94. Zangwill W. I., A Deterministic Multiproduct, Multifacility Production and Inventory Model, *Operations Res.*, 14, 486—507 (1966).
95. Zangwill W. I., A Deterministic Multi-Period Production Scheduling Model with Backlogging, *Management Sci.*, 13, 10—119 (1966).
96. Zangwill W. I., Production Smoothing of Economic Lot Sizes with Non-Decreasing Requirements, *Management Sci.*, 13, 191—209 (1966).
97. Zangwill W. I., Minimum Concave Cost Flows in Certain Networks, *Management Sci.*, 14, 429—450 (1968).
98. Zangwill W. I., A Backlogging Model and Multi-Echelon Model of a Dynamic Economic Lot Size Production System — a Network Approach, *Management Sci.*, 15, 506—527 (1969).

I-4. Управление запасами

Р. Браун¹⁾

ВВЕДЕНИЕ

Одной из сфер производства, в которой успешно применяются методы исследования операций, является сфера управления запасами. Система управления запасами представляет особый интерес для исследователя систем. Это объясняется, во-первых, сложностью задач, которые приходится решать в этой сфере управления, а во-вторых, возможностью решать проблемы, которые являются общими для организаций по крайней мере трех типов. Иными словами, в разработке системы управления запасами в первую очередь заинтересованы предприятия поточно-массового производства, оптовой и розничной торговли и обрабатывающей промышленности.

Первые системы управления запасами были разработаны применительно к запасным частям для предприятий крупных компаний и оказались успешными. При разработке систем предполагалось, что спрос на продукцию формируется большим числом потребителей и определяется главным образом дефицитом (а не ростом потребностей), что отсутствует необходимость замены одной детали другой, а затраты и время, которое требуется для выполнения заказов, известны с достаточной степенью точности.

Несмотря на то что первые работы в области систем управления запасами носили в основном теоретический характер, для практического внедрения разработанные методы были изложены в виде точных эвристических правил на языке, понятном пользователю. Развитию теории систем управления запасами способствовали исследования, проводимые в двух направлениях. Одно из них было связано с абстрактным моделированием процесса формирования запасов [1, 9, 14], который рассматривался с чисто экономической точки зрения, т. е. как процесс минимизации затрат при ограничениях на уровень запасов, необходимых для удовлетворения спроса. Второе направление носило прагматический характер [12] и предусматривало разработку способа определения уровня запасов, а не простое введение функциональной зависимости $f(x)$. Уайтинг [23] был одним из первых, кто попытался объединить эти два направления.

1. СТРУКТУРА ЗАДАЧИ УПРАВЛЕНИЯ ЗАПАСАМИ

Основная проблема, возникающая при решении задачи управления запасами, состоит в создании эффективной и надежной системы управления. Общая система управления запасами представляет собой три системы разного уровня.

Система первого уровня предусматривает обработку, ведение учета и хранение информации о запасах. Основные сведения, не-

¹⁾ Robert G. Brown, Consultant in Materials Management Systems, Box 332, Norwich, Vermont.

обходимые для успешного функционирования системы управления запасами, — это уровень паличных запасов и запасов, которые будут созданы за счет размещения заказов, а также задел по заказам потребителя, если такие заказы имеются. Кроме того, учет информации относительно любой имеющейся в запасах продукции включает такие показатели, как ее стоимость, время упреждения, единица измерения, источник получения и др. Хотя на первый взгляд количество различных данных о производимых или необходимых операциях невелико (потребность, выпуск, расход, возврат, передача, корректировка данных и т. п.), однако объем поступающей информации оказывается существенным. Значительно повысить эффективность обработки и хранения информации удалось путем автоматизации процесса обработки данных при условии, что правила обработки любого документа точно определены. Оперативный цикл системы первого уровня равен одному дню.

Система второго уровня предполагает разработку правил принятия решения, на основе которых устанавливаются срок и размер заказа, необходимого для пополнения запасов. Каждое правило принятия решений является результатом решения некоторой частной задачи оптимизации управления запасами. Задачи, которые приходится решать пользователю в практических условиях, могут быть различного характера, так как конкретные системы имеют разные параметры. Поэтому исследователю систем необходимо знать условия, при которых одно правило принятия решений оказывается предпочтительнее другого. Формально эти решения являются оптимальными, тем не менее для успешного применения необходимо учитывать общее мнение органов управления среднего уровня, обладающих соответствующим опытом, поскольку не всегда оправдано представление о «единственно правильном» решении. Оперативный цикл системы второго уровня постоянен и равен одному месяцу. Следует иметь в виду, что при разработке правил принятия решений более важным является соответствующий учет реальной обстановки, чем достижение максимального показателя эффективности системы.

Система третьего уровня позволяет на основе разработанных правил принятия решений построить модель системы управления и в соответствии с этой моделью определить стратегию финансирования на следующий год. Текущий контроль за выполнением плана дает возможность органам управления решать такие вопросы, как целесообразно ли вмешательство в функционирование системы, и если да, то каким образом и когда необходимо осуществлять управление системой в целом, а не запасами конкретной продукции.

Управление финансовой деятельностью предприятия в соответствии с разработанным планом часто осуществляется через

систему бухгалтерского учета. При этом важное значение имеет контроль не только финансовых показателей, но и таких, например, как уровень обслуживания заказчиков. Соответствующие способы моделирования и система контрольных показателей, по-видимому, в значительной степени зависят от методов работы каждого конкретного органа управления.

1.1. Правила принятия решений по управлению запасами

Как уже отмечалось, система второго уровня предполагает разработку правил принятия решений по восстановлению уровня первоначального запаса и сохранению уровня резервного запаса, что осуществляется с помощью прогнозирования, точность которого определяется степенью отклонения данных прогноза от фактических данных. Каждое правило принятия решений содержит по крайней мере одну управляющую переменную, значение которой обычно задается органами управления. Это делается с целью влияния на перераспределение капитала между затратами, связанными с запасами, и эксплуатационными затратами. Кривые распределения позволяют определить способ выбора необходимого варианта правила принятия решений из всех возможных, а также установить величину управляющей переменной в соответствии с выбранным вариантом. Для ряда систем необходимо разработать правило принятия решений по распределению ограниченного ресурса (не обязательно дефицитного). Чтобы повысить эффективность повседневной практики принятия решений, последние с помощью различных методов могут быть встроены в систему управления.

1.2. Классификация и систематизация запасов

Распределение запасов продукции различного типа в зависимости от годового объема ее реализации обычно подчиняется логарифмически-нормальному закону и может быть представлено в виде двух множеств точек, нанесенных на график с логарифмической шкалой по оси абсцисс. При этом одно из множеств характеризует объем продукции, реализуемой в течение года, а другое — часть общего годового дохода (или стоимости), связанную с реализацией данной продукции. Если распределение является логарифмически-нормальным, то оно имеет вид, показанный на рис. 1: две параллельные прямые, отношение тангенсов угла наклона которых есть «стандартное отношение». В этом случае объем реализации около 16% всей продукции превышает средний объем на величину большую, чем одно стандартное отклонение, а объем реализации 84% продукции превышает среднее

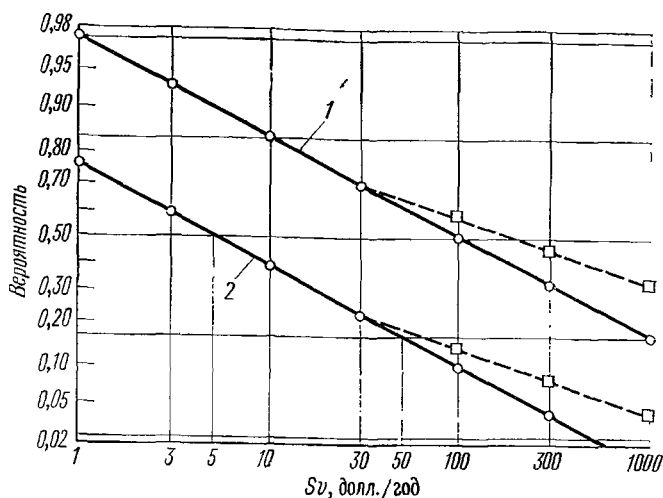


Рис. 1. Логарифмически-нормальное распределение.

1—доля совокупного дохода от продажи продукции с годовым объемом реализации, превышающим Sv ; 2—доля продукции с годовым объемом реализации, превышающим Sv .

значение минус стандартное отклонение (эти границы отмечены сплошными горизонтальными линиями на соответствующих уровнях).

Параметр σ (не являющийся стандартным отклонением) логарифмически-нормального распределения (т. е. нормального распределения натурального логарифма объема реализуемой продукции Sv) может быть вычислен для объема реализации 16 и 50% продукции как $\sigma = \ln 50 - \ln 5 = \ln 50/5 = \ln 10 = 2,3$. Отношение $50/5 = 10$ есть стандартное отношение. Если средний объем реализуемой продукции в год обозначить через m и $J = e^{1/2\sigma^2}$, то медиане распределения соответствует объем реализации, равный m/J , половина которого обусловлена реализацией продукции объемом большим, чем величина mJ [2].

Многие отрасли промышленности имеют в запасе продукцию более одного типа. С точки зрения потребителя этих запасов подобная ситуация соответствует двум логарифмически-нормальным распределениям, имеющим разные значения средних объемов реализации, но одинаковые стандартные отношения. В этом случае форма распределения отличается от логарифмически-нормального (штриховые линии на рис. 1). Существование подобной зависимости может быть проиллюстрировано на следующих примерах. Так, наличие гостиниц различного разряда определяет не только стремлением повысить стоимость номера, но и целями, а также возможностями людей, которые пользуются услугами гостиниц; показатели рынка промышленных инструментов отличаются от показателей рынка инструментов для кустарной

работы, хотя как те, так и другие потребители стремятся приобрести одинаковую продукцию. Учет показателей различия между отдельными совокупностями потребителей имеет важное значение для усовершенствования системы управления запасами и требует от исследователя систем определенного навыка и богатого воображения.

Распределение запасов более удобно представить в виде зависимости годового объема реализованной продукции (выраженного в процентах) от процентного содержания данной продукции в общем уровне запасов (так называемое распределение Парето). В результате составляется перечень продукции в порядке убывания годового объема (выраженного в долларах) продукции каждого типа. Если годовой объем реализации 20% имеющихся запасов продукции составляет 80% (эти цифры редко выдерживаются), то стандартное отношение равно примерно 5 и могло бы служить показателем спроса на потребительские товары в розничной торговле. Стандартное отношение для готовой продукции обычно составляет ~ 10 , а для такой высокотехнологической продукции, как комплектующие узлы самолетов или вычислительных машин, оно может быть равно 25. Для поиска стандартного отношения и установления возможности неоднородных совокупностей удобно представить распределение запасов в виде зависимости дохода от доли соответствующего типа продукции в запасах, используя вероятностную шкалу по обеим осям координат (рис. 2) [5].

Если распределение является логарифмически-нормальным, то точки располагаются вдоль прямой линии, имеющей наклон 45° . Как видно из рис. 2, реализация 50% продукции дает 99% дохода. Иными словами, 50% продукции определяют некоторую часть общего дохода $F(\sigma)$ (в данном случае равную 0,99). Поскольку функциональная зависимость графически изображается в виде кумулятивной кривой нормального распределения, то можно найти, что $\sigma = 2,33$, так как стандартное отношение равнялось $\rho = e^\sigma = 10,15$. Полученное значение параметра σ хорошо согласуется с вычисленным выше значением на основе данных, приведенных на рис. 1.

Продукция нескольких типов, которая составляет верхнюю часть перечня и реализация которой дает 50% дохода, обычно называется продукцией класса А. Для хранения такой продукции необходима наиболее совершенная система управления запасами и распространение формальных правил принятия решений на случаи, которые раньше не принимались во внимание. Нижняя часть перечня содержит продукцию класса С, реализация которой дает всего лишь 1% общего дохода. Для управления запасами подобной продукции может быть использована только очень простая система, так как издержки, связанные с при-

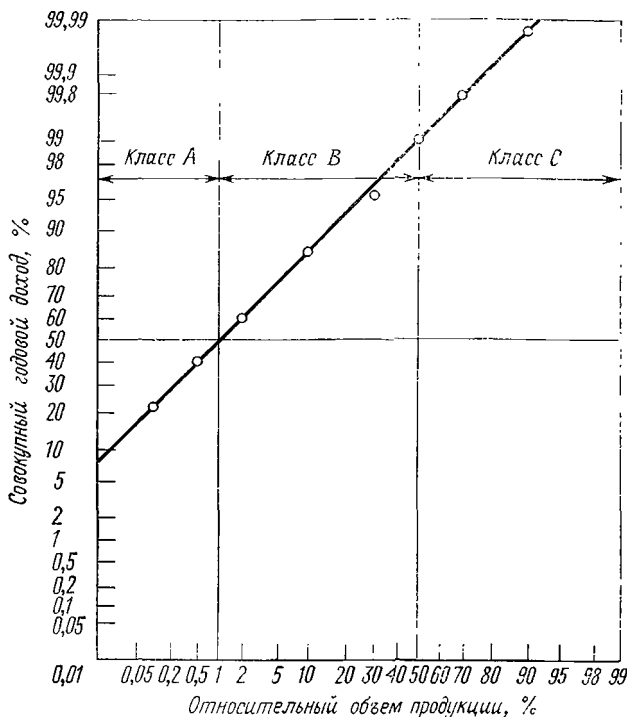


Рис. 2. Распределение продукции по значимости.

ближенностью управления, могут быть намного меньше, чем экономия от использования простых процедур для такого большого объема продукции. Средняя часть перечня охватывает почти половину всей продукции (класс В), реализация которой дает около половины всего дохода. В этом случае целесообразно создание добротной системы, но с минимальным вмешательством органов управления в функционирование существующей системы.

Для более точного управления запасами последние могут быть классифицированы следующим образом:

- материалы (сырье, полуфабрикаты), запас комплектующих изделий, готовая продукция на складе предприятия, готовая продукция, поступающая в систему распределения;

- продукция собственного производства, продукция, поставляемая независимыми поставщиками или другими подразделениями фирмы;

— продукция, которую заказчик может получить на основе выбора из ряда конкурирующих, продукция поставляемая единственной компанией.

В качестве показателя, по которому может быть осуществлена классификация запасов, используют чистый доход (или чистую прибыль), «возраст» продукции, т. е. время с момента ее появления на рынке, и степень необходимости производства данной продукции (спроса).

Следует иметь в виду, что под «запасом» мы понимаем запасы единиц продукции, имеющих аналогичные характеристики в рамках установленной классификации. Предприятие может иметь от трех до десяти видов запасов, и для каждого из них можно разработать соответствующие правила принятия решений и выбрать необходимые значения управляющих переменных.

1.3. Целесообразность создания запасов

Существуют две точки зрения относительно необходимости создания запасов. Согласно одной из них, наличие запасов позволяет быстро удовлетворять запросы потребителей; согласно другой, наличие запасов позволяет поставщику нейтрализовать колебания спроса в условиях равномерного производства продукции.

В системе управления запасами, ориентированной на потребителя, решение об увеличении размера запаса обычно принимается с учетом количества заявок на данную продукцию. Если в течение времени T на некоторую продукцию поступило по крайней мере N заявок, то создается запас этой продукции. При такой стратегии органы управления, ответственные за сферу сбыта, могут составить таблицу решений, предусматривающих затраты при заданных значениях (N, T) , путем классификации запасов. Существует продукция определенного типа (запасные части для нового оборудования и основные запасы для производства новой продукции), которую целесообразно запасать до поступления первых заявок. В ряде случаев более оправданными является создание запасов неходовой продукции (исключительно запасных частей), т. е. не имеющие прямого функционального назначения, а не продукции, пользующейся постоянным спросом (например, детали, участвующие в сборке изделия).

Поскольку доход от реализации имеющейся в запасе продукции в течение длительного времени может быть лишь частично оценен в денежном выражении (и подобная оценка, как правило, является весьма субъективной), то, по-видимому, не имеет смысла разрабатывать сложные модели системы управления запасами с учетом затрат на хранение дополнительного объема продукции. Органы управления, принимающие решение, должны иметь пред-

ставление об экономическом эффекте принимаемого решения, однако модель, используемая для такого прогноза, должна быть предельно простой и учитывать планируемый объем инвестиций в товарно-материальные запасы, размеры складских помещений, а также затраты на функционирование системы управления первого уровня (системы учета).

В системе управления запасами, ориентированной на поставщика, существует реальная экономическая основа для принятия решения относительно того, запастись ли данную продукцию или производить, т. е. организовывать ли собственный ее выпуск или закупать на стороне, если она пользуется спросом. Предположим, что в год реализуется N единиц продукции. Издержки на производство партии продукции составляют A долл. (не считая затрат на запуск в производство или размещение заказа), стоимость одной единицы продукции — v долл. и затраты на хранение в течение одного года запасов стоимостью в 1 долл. составляют r долл. Если размер партии продукции для однократного пополнения запасов равен p ед., то годовые затраты определяются как $NA/p + \frac{1}{2}rv(p - 1)$. Если выполняется соотношение $N \leq rv/A$, то экономия средств в случае отсутствия необходимости хранить запасы продукции ($p = 1$) гораздо меньше, чем затраты на производство двух единиц продукции, одна из которых используется сразу, а другая хранится в течение времени реализации первой. Следовательно, необходимо стремиться дорогостоящую продукцию или изделия с очень низким значением издержек на запуск в производство выпускать по заказам, а высокий объем реализации дешевой продукции или продукции с большими затратами на запуск в производство поддерживать из запасов.

Моделирование систем управления запасами с необходимостью предполагает разработку подходов к решению задачи экономически целесообразного размещения данной продукции на складах распределительной сети. Одним из возможных подходов является следующий. Сначала определяются тип и объем продукции, которая может быть размещена на каждом складе. Затем составляется перечень всей продукции, подлежащей хранению, с учетом (в порядке убывания) числа заявок (а не выраженной в долларах годовой потребности). Продукция, занимающая верхнюю часть перечня, покрывает 50% всех заявок и должна быть распределена по всем складам. Это позволит удовлетворить наибольшее число заказов при относительно небольших затратах на создание запасов, так как последние будут охватывать небольшой объем продукции. Середина перечня — это продукция, которая может накапливаться по региональному принципу для сокращения числа необходимых учетных операций, а также для экономии транспортных расходов, связанных с доставкой больших партий

продукции в районы сбыта, и извлечения выгод от приближения пунктов сбыта имеющейся продукции к потребителю.

Продукция, перечисленная в нижней части списка, должна размещаться только на одном складе для сокращения затрат, связанных с многократными расчетами по одинаковым видам запасов. Задержки с поставками таких материалов заказчикам не могут приводить к серьезным потерям для сбыта в целом из-за незначительности доли этого спроса и, как следствие, незначительности транспортных издержек.

В последнее время изучается возможность выделения «в запас» производственных мощностей. Это обеспечило бы быстрее производство небольших объемов продукции, пользующейся малым спросом. Следовательно, исключалась бы необходимость создания ее запасов. Для осуществления такого производства необходимы, по-видимому, специализированный производственный участок, имеющий совершенно самостоятельную технологию производства и снабженный станками с числовым программным управлением, а также бригада обслуживающих механиков. Такой участок может оказаться существенно менее дорогим, чем физическое хранение запасов готовой продукции, пользующейся малым спросом, который однако должен быть удовлетворен.

Для отрасли промышленности, выпускающей товары широкого потребления, характерна ситуация, когда на складе имеется много сходных типов продукции, любая из которых могла бы удовлетворить запросы потребителей. В этом случае в большинстве подходов (связанных с определением величины партии продукции, необходимой для пополнения запасов и сохранения уровня резервных запасов) теория управления запасами рассматривает всю агрегированную группу взаимозаменяемой продукции как элемент планирования. Таким образом, пока существует общий запас подходящей продукции, потребитель будет выбирать и приобретать что-либо из того, что имеется на складе.

Результаты некоторых экспериментов позволяют предполагать, что существует определенное нормируемое (в широких пределах) число вариантов целесообразного ассортимента задаваемой продукции. Если ассортимент продукции довольно ограничен, то потребитель вряд ли станет приобретать какую-либо из них, даже если эта продукция пользуется широкой популярностью в качестве заменителя; если ассортимент слишком велик, потребитель может оказаться в затруднительном положении и не сумеет сделать окончательного выбора. Естественно, что разные компании имеют различные точки зрения на необходимую для потребителя широту ассортимента товаров, имеющихся на складе. Поэтому, вероятно, ассортимент продукции, имеющейся в запасе, влияет на ожидаемое количество потребителей данного пункта сбыта.

2. ПРАВИЛА ПРИНЯТИЯ РЕШЕНИЙ ОТНОСИТЕЛЬНО ВЫБОРА РАЗМЕРА ПАРТИИ ПРОДУКЦИИ ДЛЯ ПОПОЛНЕНИЯ ЕЕ ЗАПАСОВ

При необходимости пополнить запасы продукции на первом уровне системы управления решается задача определения размера заказа (или размера партии) продукции. Восстановление уровня запасов может осуществляться либо путем выполнения заказа на изготовление некоторого фиксированного объема продукции, зарегистрированного системой первого уровня, либо путем выполнения заказа на изготовление объема продукции, равного разнице между уровнем располагаемого запаса (который равен: Наличный запас + Объем заказов — Ожидаемый расход) и максимальным уровнем запасов, данные о котором также хранятся в системе учета. Иногда размер заказа по каждому типу продукции представляет соответствующую часть, ограниченную допустимыми объемом, весом и стоимостью общего заказа для всей совокупности производимой продукции.

Рассмотрим решающие правила, которые разрабатываются системой второго уровня для определения экономически обоснованного оптимального размера заказа, которые могут быть использованы при определении установленного (запланированного) или максимального уровня запаса для конкретного типа продукции или совокупностей продукции.

2.1. Исходные предположения относительно условий функционирования системы управления запасами

Различные подходы к решению задачи определения экономического размера заказа должны базироваться на тех или иных возможных предположениях относительно таких показателей, как способ поставки заказанной продукции, интенсивность реализации продукции, затраты на освоение производства новой продукции, себестоимость единицы продукции и текущие затраты на хранение запасов [17]. Вначале для каждого из перечисленных факторов делается наиболее простое предположение, а затем рассматриваются условия, при которых возможные предположения оказываются более предпочтительными, чем первоначальное. Поскольку предположения относительно одного из факторов не связаны с предположениями относительно других факторов, то возможны по крайней мере 120 комбинаций, приводящих к различным решающим правилам. Из такого множества решающих правил можно выбрать наиболее приемлемое, однако целесообразность того или иного выбора определяется с учетом конкретной ситуации, в которой эти правила будут реализованы.

Рассмотрим некоторые возможные предположения относительно указанных факторов.

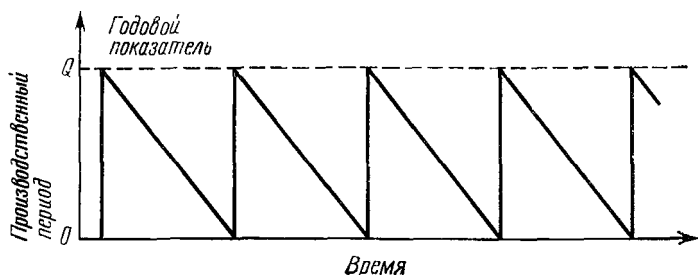


Рис. 3. Наличный запас при получении заказа одной партией.

Способ поставки заказанной продукции. а) Вся заказанная продукция поставляется в виде одной партии. Согласно модели, наличный запас возрастает от минимального уровня таким образом, что к началу следующего цикла достигает размера партии (рис. 3).

б) Продукция поступает на склад с момента начала производственного цикла в течение интервала времени, достаточного для реализации продукции до поставки всей партии. Как видно из рис. 4, если годовой темп поступления продукции на склад равен P , а годовой темп ее потребления — S , то максимальный уровень наличного запаса к концу производственного периода равен $Q(1 - S/P)$. Заметим, что суммарный годовой объем поступления продукции на склад будет иметь величину порядка S . Так как P — темп поступления продукции в течение производственного периода, то готовая продукция производится только в продолжение $(1 - S/P)$ рассматриваемого периода.

Подобное предположение является обоснованным, если $P < 2S$, и в этом случае выполняется соотношение $(1 - S/P) > 0,5$. Если $P < S$, то подразумевается, что имеется несколько производственных единиц, каждая производительностью P , причем

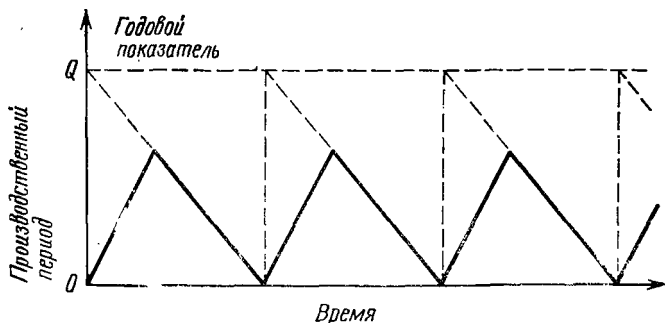


Рис. 4. Распределение во времени запаса производимой продукции.

($n - 1$) из них функционируют непрерывно (т. е. без переналадок), а n -я имеет переналадки, чтобы удовлетворить эффективный годовой спрос, соответствующий $S - (n - 1)P$ единиц продукции в год.

Интенсивность и сроки реализации продукции.

а) Продукция реализуется с постоянной интенсивностью в течение неограниченного времени. Это потребует многократного пополнения запасов за счет заказов одинакового размера, причем наличный запас будет уменьшаться по линейному закону (рис. 3 и 4). Практически предположение о неограниченном времени реализации означает, что необходимо по крайней мере более чем три пополняющие партии продукция.

б) Продукция реализуется с известной или прогнозируемой интенсивностью, изменяющейся во времени. (Это случай, которому в течение последних 15 лет уделяется большое внимание в литературе по исследованию операций.) Для получения точного решения требуется применение методов динамического программирования [22]. Было предложено большое количество возможных приближенных решений задачи с учетом рассматриваемого предположения, но при этом не всегда приводились оценки степени их приближения к оптимальному решению.

Если потребность в продукции, пользующейся высоким спросом, приходится на конец какого-то интервала периода планирования (как это бывает с продукцией, быстро выходящей из моды, или с новой продукцией, для которой предполагается экспоненциальный рост спроса), то нельзя использовать ни один из предложенных приближенных методов. Если в течение периода планирования имеют место высокий и низкий спрос, то различия между возможными приближениями незначительны и оценки их обычно делаются на основе отдельных случайных пробных выборок, а не на основе принципиальных различий.

в) При значительных случайных колебаниях спроса минимальный уровень запаса отличен от нуля (а не равен нулю, как это показано на рис. 3 и 4). В этом случае существует резервный запас. Если размер партии заказываемой продукции увеличивается по сравнению с тем, которое соответствует случаю выполнения заказов, то вероятность дефицита продукции в течение года уменьшается. Следовательно, резервный запас должен быть таким, чтобы обеспечить желаемый уровень обслуживания потребителей. Общая экономия в результате одновременного определения уровня резервного запаса и размера партии продукции для его пополнения осуществляется в тех случаях, когда размер партии продукции, вычисленный при условии отсутствия дефицита, меньше, чем стандартное отклонение ошибок прогноза на интервале упреждения.

Рассмотрим случай, когда слишком дорогая продукция, освое-

ние производства которой требует малых затрат, может быть выпущена одновременно в объеме недельной поставки, но реализация ее оказывается столь неравномерной, что стандартное отклонение эквивалентно размеру месячной поставки. Точное решение задачи требует итеративного процесса вычислений. Практически удовлетворительным приближением решения является увеличение размера партии по крайней мере до величины стандартного отклонения. Однако подобное решение предполагает выполнение значительного объема вычислений, и поэтому дальнейшее улучшение результата может оказаться неоправданным с точки зрения затрат, связанных с вычислительными работами.

г) Продукция реализуется до тех пор, пока не появится непредвиденная необходимость полного списания оставшегося наличного запаса [6]. В этом случае следует принимать во внимание технологические изменения, которые планируются на будущее, с тем чтобы постепенно свести к нулю текущие запасы. Для этого последняя поставляемая партия изделий должна иметь объем, достаточный для потребления лишь в пределах планового периода, предшествующего моменту запланированного нововведения. Однако возможно, что федеральное правительство сочтет некоторую продукцию опасной и подлежащей немедленному снятию с производства. При такой ситуации будут увеличиваться текущие затраты на хранение запасов с учетом вероятности того, что такое изменение произойдет в течение следующего года.

д) Если интенсивность реализации продукции в течение некоторого интервала времени периода планирования снизится до нуля (окончание контракта, изменение модели изделия, нарушение непрерывности и т. д.), то размер последней партии должен быть таким, чтобы компенсировать разницу между наличным запасом и общим неудовлетворенным спросом. Для решения вопроса о том, экономично ли изготовление изделий в количестве, удовлетворяющем оставшийся спрос, одной или несколькими партиями, требуется использование моделей динамического программирования. Если предполагается удовлетворение спроса путем изготовления нескольких партий изделий, то необходимо определить размер первой партии. Разница между размером первой партии в оптимальной последовательности и размером партии без учета спадов существенна лишь при поставке для удовлетворения оставшегося спроса не более трех партий.

Затраты на запуск в производство. Эти затраты могут включать затраты на наладку производственной линии, организацию выполнения календарного графика производства и затраты на размещение заказа, которые в конечном счете зависят от количества выполненных в течение года заказов. В этом случае возможны следующие предположения.

а) Освоение производства продукции каждого нового типа независимо от выпускавшихся ранее требует новых затрат.

б) Основные затраты связаны с размещением заказа на изготовления совокупности изделий, и незначительные затраты связаны с включением в текущий заказ дополнительных единиц продукции этой же совокупности. Примерами, иллюстрирующими эти случаи, могут служить перечень товаров, приобретенных у одного поставщика, различные продукты, разливаемые на конвейере в бутылки одинаковой емкости, или детали, изготавливаемые путем штамповки с использованием сменяемых штампов, но из одинаковых имеющихся на складе рулонов материала постоянной толщины и ширины.

Вариант объединенных заказов оказывается эффективным, когда основные затраты на освоение производства новой продукции превышают общие дополнительные затраты по всей совокупности изделий.

Себестоимость единицы продукции. При определении объема капиталовложений в запасы учитываются соображения относительно себестоимости продукции, что приводит к целесообразности выпуска более крупных партий изделий, чем это требуется в данный момент. Здесь возможны следующие предположения.

а) Может быть произведен любой объем продукции при постоянной ее себестоимости.

б) Существуют непредвиденные изменения цен на продукцию. В данном случае, если затраты на размещение будущих заказов могут возрасти по сравнению с существующей себестоимостью продукции, то текущий размер заказываемой партии должен быть больше того, который определяется величиной вероятности невыполнения заказов потребителей; это позволяет отдалить точку увеличения расходов на приобретение изделий по повышенной цене. Экономически целесообразнее увеличение объема разовой поставки пропорционально ожидаемому относительному увеличению этой цены. На коэффициент пропорциональности основное влияние оказывают текущие управленческие расходы или значение коэффициента приведения будущих затрат к текущему моменту времени. Эффективность такого решения может оказаться гораздо выше, чем при выборе стратегии повторения заказов равного размера, определяемого из условия отсутствия дефицита без учета изменения окружающих условий.

в) Дисконтирование объема заказа. В тех случаях, когда увеличение размера партии изделий сверх определяемого вероятностью невыполнения заказов потребителей приводит к снижению себестоимости продукции, можно определить экономичный размер заказа на основе соотношения общих затрат при нехватке запасов с затратами в каждый момент изменения цены. Можно

рекомендовать удобное эвристическое правило, согласно которому экономически целесообразно каждые M месяцев заказывать партию продукции, размер которой определяется из условия отсутствия дефицита запасов, если снижение удельных затрат (себестоимости) будет не менее чем на $2M\%$ (исходя из типичного положения, когда ежегодные текущие управленческие расходы лежат в пределах $20-24\%$).

Текущие затраты на хранение запасов. Подобные затраты являются управляющей переменной, которая может быть увеличена, чтобы снизить объем капиталовложений в запасы за счет выполнения большего количества заказов в год. Величина текущих затрат, которая могла бы теоретически обеспечить максимальную отдачу капиталовложений, равна текущему значению прибыли фирмы на ее нетто-активы, т. е. должна находиться в пределах установленного диапазона $20-24\%$ годовых затрат. Однако реальные текущие затраты должны быть выше этого значения, если органы управления считают, что вложение капитала в запасы связано с большим риском или меньшей ликвидностью, чем по другим объектам капиталовложений фирмы. Отсюда следует, что для какой-нибудь крупной корпорации текущие затраты на хранение запасов сырья могут быть меньше, чем на хранение запасов готовой продукции, а текущие затраты на хранение периферийных запасов будут больше, чем на хранение запасов готовой продукции на складе предприятия. Возможны следующие варианты.

а) Текущие затраты определяются политикой капиталовложений, основанной на учете риска возникновения дефицита запасов.

б) Учитывается занимаемое пространство. Если имеются изделия относительно крупных габаритов и наряду с ними изделия, которые располагаются на складе достаточно компактно (с плотностью, выраженной в долларах на m^3), то органы управления могут в качестве второй характеристики стратегии выбрать затраты, связанные с занимаемым продукцией пространством склада, и это должно привести для крупногабаритных изделий к уменьшению размера партии по сравнению с тем, который был бы в случае отсутствия дефицита. Следует отметить, что, например, в случае управления запасами стеклянных бутылок или изолирующих материалов, используемых в строительстве, когда все объекты имеют одинаковые объемы, нет необходимости иметь в математической модели специальный член, учитывающий занимаемые объемы: в данном случае высокие капитальные затраты требуют сокращения запасов до уровня, определяемого имеющимся свободным пространством.

В определенных границах текущие затраты на хранение запасов являются управляемой переменной, характеризующей ли-

нию поведения органов управления. Если органы управления выбирают величину таких затрат несколько большей, чем прибыли фирмы на нетто-активы, то размер партии заказываемых изделий будет меньше и частота пополнения заказов более высокой. Выбор более низкой величины текущих затрат приводит к увеличению капиталовложений в запасы, но зато уменьшает ежегодные затраты, связанные с размещением заказов на пополнение запасов, которые в этом случае более редки.

В этом смысле существует максимальное практически целесообразное значение текущих затрат. Когда для поддержания нужного уровня обслуживания заказчиков применяются резервные запасы, стратегия более мелких и более частых пополняющих заказов требует дополнительного увеличения резервных запасов. Начиная с некоторого значения величины текущих затрат на хранение запасов, уменьшение объема текущих запасов сопровождается увеличением требуемого страхового запаса, и поэтому за пределами указанной точки увеличение текущих затрат ничего не дает, кроме увеличения издержек размещения заказов без компенсации их за счет снижения общего объема капитальных затрат.

2.2. Случай поставки изделий партиями, определяемыми с учетом риска нехватки запасов

Если интенсивность реализации продукции равна S ед/г, объем партии продукции, поставляемой одновременно для пополнения запаса, — Q , затраты на запуск в производство заказанной партии продукции — A , себестоимость одной единицы продукции — v долл., а стратегия управления состоит в том, чтобы требовать получения отдачи r долл. на 1 долл. вложенных в запасы денежных средств в год, то экономичный размер заказа равен

$$Q = \sqrt{2AS/rv}.$$

Эта формула впервые была опубликована в 1915 г. Харрисом [41] и является одной из первых моделей, оптимизирующих затраты.

Годовые затраты, связанные с размещением заказов на пополнение запасов, составят $C_1 = AS/Q$, стоимость хранения запасов — $C_2 = 1/2 rvQ$ и общие затраты — $C(Q) = C_1 + C_2$ (рис. 5). Заметим, что в окрестности точки оптимума кривая общих затрат довольно пологая. Если фактически заказанный объем продукции находится в пределах от $Q/\sqrt{2}$ до $Q\sqrt{2}$, то общие годовые затраты будут менее чем на 6% превышать минимальное значение. Следовательно, ошибка в оценке любого из входящих в формулу параметров не будет серьезно влиять на эффективность практического использования формулы. Эффективность применения

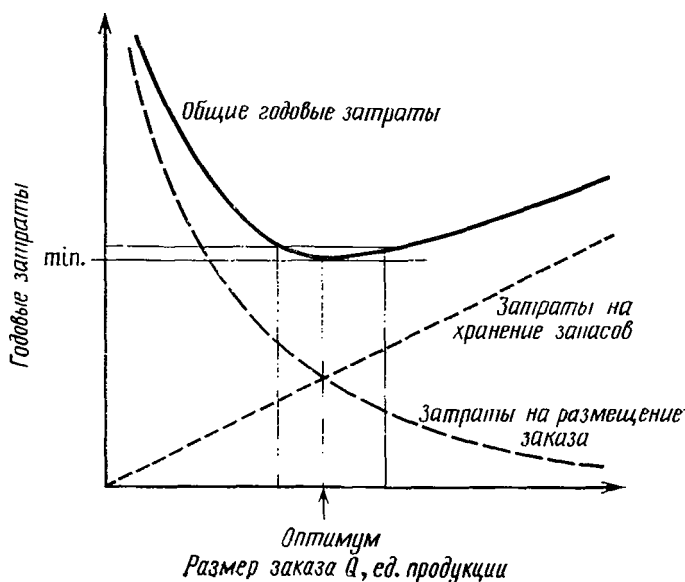


Рис. 5. Общие годовые затраты на создание и хранение запасов.

формулы может снизиться, если, например, при оценке параметров будут допущены ошибки в сторону завышения значения обоих параметров в числителе и в сторону занижения значения в знаменателе. Однако 2-кратное изменение параметров является столь широким допуском, что большинство способов их оценки позволяет уложиться в эти пределы.

2.3. Общий случай

Большинство из возможных предположений могут быть включены в более общую формулу. Такие случаи, как известный и случайный спрос, ограниченный неудовлетворенный спрос и изменения цены в зависимости от времени или размера заказываемой партии, будут рассмотрены отдельно.

Выразим величину интервала, в течение которого будет реализовываться партия продукции, как $T = Q/S = \sqrt{2A/rSv(1 - S/P)}$. Заметим, что если P превосходит S , то влияние параметра S/P становится незначительным. Если существует отличная от нуля вероятность p того, что наличный запас будет неожиданно списан в следующем году, то $T = \sqrt{2A/(r + p)Sv}$. Если $p = 0$, то рассматриваемый случай сводится к случаю дефицита запасов. Определим теперь груп-

повые издержки по запуску заказов в производство. Пусть A — основные затраты на освоение производства заказанной продукции. a_j — второстепенные издержки, связанные с включением в заказ j -й продукции. Если a_j зависит от последовательности изготовления типов продукции, то для нахождения оптимальной последовательности производства отдельных единиц продукции может быть сформулирована задача коммивояжера и решена, например, методом ветвей и границ. На практике органы управления всегда знают одну из возможных приближенно-оптимальных последовательностей и поэтому не будут рассматривать никакую из явно неэкспонимых последовательностей. Таким образом, нецелесообразно искать оптимальную последовательность, так как возможные улучшения не смогут оправдать затрат на их получение. Во всяком случае, значения a_j являются второстепенными издержками в оптимальной (или близкой к оптимальной) последовательности. Если в каждом цикле производится вся продукция, то наилучший интервал между выпуском партий изделий будет равен $T = \sqrt{2(A + \sum_j a_j)/r \sum_j S_j v_j}$, где j — индекс изделия, входящего в группу. Если медленно реализуемая дешевая продукция производится в каждом k -м цикле и пополнение запасов будет происходить в течение $k_j T$ лет, то наилучший интервал между циклами выпуска партий продукции становится равным

$$T = \sqrt{2(A + \sum a_j/k_j)/r \sum k_j S_j v_j}.$$

Для определения оптимальной частоты включения в заказ медленно реализуемой продукции можно использовать уравнение [4] вида

$$k_j^2 = (a_j/S_j v_j) \left(\sum_i k_i S_i v_i / (A + \sum_i a_i/k_i) \right).$$

Решение уравнений начинается с предположения, что $k_i = 1$ для всех правых частей, и уравнения решаются относительно k_j в левых частях.

Полученные значения подставляются в правые части уравнений, и такой процесс продолжается до тех пор, пока он не сойдется к предполагаемому решению.

Целое значение k_j может быть получено на каждом шаге округлением в большую сторону дробной части квадратного корня, если эта дробная часть больше 0,4.

Другой метод определения оптимальной частоты включения в заказ медленно реализуемой продукции дан в работе [16].

Все эти варианты могли бы быть объединены в одну на первый взгляд громоздкую формулу. Но так как в случае дефицита запасов каждая из переменных принимает конкретное значение, то эта формула упрощается.

2.4. Частные случаи изменения годовых затрат

На рис. 6 приведены кривые затрат на размещение заказов на пополнение запасов, включающих капиталовложения в запасы и общие затраты в случае, когда себестоимость единицы продукции изменяется при очередном заказе. В этом случае может быть несколько решений $Q = \sqrt{2AS/rv}$ для каждого значения себестоимости v_i , но имеется только одно непротиворечивое решение, для которого Q_i находится в пределах, обеспечивающих достижение себестоимости изделия v_i . Это — количество, определяемое из условия отсутствия дефицита запасов. Минимальное значение годовых затрат должно достигаться либо при таком объеме продукции, либо при некотором большем значении Q , если себестоимость продукции снижается. Выбор значения (Q^*, v^*) , которому соответствуют наименьшие годовые затраты, может быть сделан на основе сопоставления затрат для ограниченного числа возможных решений.

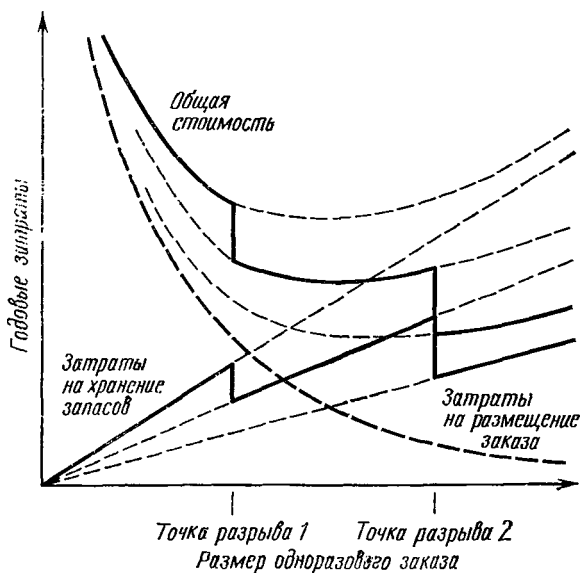


Рис. 6. Общие затраты на создание и хранение запасов при наличии скидок.

Если ожидается, что себестоимость единицы продукции в будущем будет выше, чем в настоящий момент (например, в случаях, когда продажа изделия закончена или когда производственная линия переоборудуется в конце года на выпуск только запасных частей для модели, снимаемой с производства), то полезно знать приведенное к настоящему моменту значение $W(Q)$ будущего удорожания для возможных вариантов решения относительно текущего и будущего размеров партии продукции, т. е. $W(Q) = (A + Qv)/(1 - e^{-rQ/S})$. Если все параметры определены при условии отсутствия дефицита запасов, то, дифференцируя это выражение, можно найти оптимальный размер партии при отсутствии дефицита. Это будет стратегия, обеспечивающая наименьшие затраты, в которой параметр r — это коэффициент, используемый органами управления для дискотирования будущих затрат при сравнении возможных вариантов.

Если возможен выпуск любого объема продукции Q_1 без изменения себестоимости v_1 единицы продукции, а будущее производство характеризуется более высокой себестоимостью v_2 , то оптимальное текущее значение размера партии определяется из формулы $Q_1 = (S/r) \ln(r/Sv_1) (A_2 + Q_2v_2 + Sv_2/r)$, где Q_2 — размер партии при условии отсутствия дефицита и увеличении себестоимости единицы продукции до значения v_2 .

В общем случае, когда затраты A на освоение производства новой продукции намного меньше величины (Qv) , имеем следующее приближенное соотношение:

$$\frac{Q_1v_1 - Q_2v_2}{Sv_1} = (1/r)(v_2 - v_1)/v_1 + (2A_2S/rv_1 - Q_1^2)/2S^2.$$

Если второй член в правой части равенства близок к нулю (разность квадратов объемов постановки), то увеличение размера партии сверх экономичного размера заказа пропорционально относительному возрастанию цен на продукцию и коэффициент пропорциональности эквивалентен затратам на хранение запасов или коэффициенту дискотирования затрат.

2.5. Частные случаи изменения статистической характеристики интенсивности использования продукции

Если истинная интенсивность известного, планируемого или прогнозируемого спроса существенно меняется в течение интервала времени, который является внешним по отношению к будущему периоду планирования, то точное решение задачи определения экономичного размера заказа может быть получено с помощью динамического программирования [22]. Этот процесс может быть наглядно продемонстрирован путем построения кривых суммарных

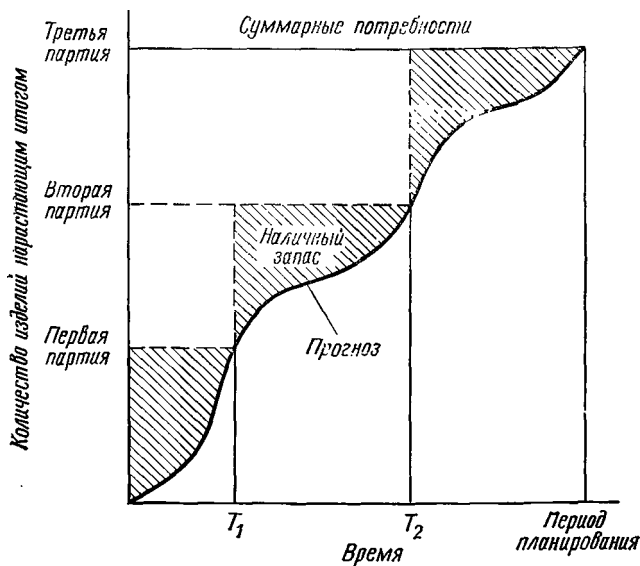


Рис. 7. Удовлетворение суммарных потребностей за счет поставок трех партий продукции.

потребностей (рис. 7). Рассмотрим в качестве одного из возможных вариантов случай, когда удовлетворение суммарных потребностей достигается за счет поставок трех партий продукции. Затененная область на рисунке представляет общий уровень запасов продукции, хранимой в течение времени T_1 и T_2 , причем моменты времени T_1 и T_2 соответствуют полному использованию как первой, так и второй партий продукции и поступлению следующей партии для пополнения запаса. Размер третьей партии продукции, которая предназначена для пополнения запаса, очевидно, определяется из условий баланса потребностей уже за пределами периода планирования. Можно выбрать такие интервалы времени T_1 и T_2 , в течение которых общий уровень запасов, а следовательно, и затраты, связанные с удовлетворением потребностей за счет лишь только трех партий продукции, достигнут минимума. Аналогичный анализ можно сделать для любого числа $N = 1, 2, \dots$ партий продукции. Если уменьшение затрат на хранение минимального запаса, достигаемое благодаря минимальному их объему при переходе от N к $N + 1$ партий изделий, меньше, чем дополнительные затраты на освоение производства еще одной партии продукции, то в этом случае известно оптимальное количество партий, а значит, и последовательность размеров партий при оптимальной стратегии.

Если суммарная кривая будущих потребностей строго выпуклая (вверх или вниз, указывая на то, что высокий спрос целиком приходится на окончание одного из интервалов планового периода), то удовлетворительным компромиссным решением является использование искусственной интенсивности потребления S , определяемой как средняя плановая потребность за пределами периода планирования. Затем следует определить оптимальный интервал между выпуском партий продукции любым из обсуждавшихся выше способов, подходящим для рассматриваемых внешних условий. Фактически размер первого заказа на пополнение запасов равен сумме планируемых потребностей в течение оптимального интервала времени.

Иногда целесообразно дальнейшее улучшение этого приближения. Если планируемые потребности в течение последнего периода (недели, месяца и т. д.) первого интервала превышают среднее значение величин S , то следует планировать выпуск партий изделий с охватом интервала времени, который на один период меньше оптимального. В этом случае ожидаемое изъятие запасов будет приходиться на начало второго интервала, а не на конец первого, в результате чего сокращаются капиталовложения в запасы.

Рассмотрим две приближенные модели: модель наименьшей себестоимости единицы продукции и модель частичного балансирования периода планирования [8] (рис. 8). При использовании

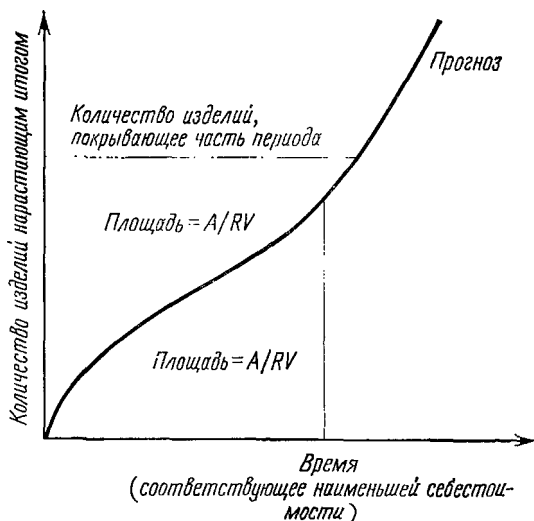


Рис. 8. Данные анализа затрат, полученные с помощью методов наименьшей себестоимости единицы продукции и частичного балансирования периода планирования.

первой модели размер партии выбирается таким, чтобы соответствующая ему на графике ордината и кривая суммарного спроса ограничивали область площадью A/rv . В случае второй модели эта область также должна иметь площадь равную A/rv . Если спрос близок к постоянной годовой интенсивности реализации S продукции, то площадь выделенных на графике областей и, следовательно, размеры партий для обоих приближений равны ЭРЗ. Если спрос отличается от величины S , то различие результатов, получаемых с помощью двух алгоритмов, весьма чувствительно к форме кривой суммарных потребностей в конце периода реализации первой партии продукции.

Рассмотрим случай, когда спрос снижается экспоненциально, т. е. $S_t = S_0 e^{-at}$. Если наступает момент заказа последней партии, то объем продукции должен быть таким, чтобы удовлетворить все невыполненные заказы, т. е. должен быть равен $Q = S_0/a$. Если общий спрос целесообразно удовлетворить с помощью n партий, то используется следующее рекуррентное соотношение для момента времени T_j , когда j -я партия будет израсходована и следующая партия изделий пойдет на образование запасов:

$$a(T_j - T_i) = 1 - \exp[-a(T_{j+1} - T_j)].$$

Так как величина T_n равна бесконечности, то имеем $a(T_{n-1} - T_{n-2}) = 1$, а поскольку первая партия должна быть изготовлена к данному моменту времени, то $T_0 = 0$ и $a(T_1) = 1 - \exp[-a(T_2 - T_1)]$. Размер первой партии, определяемый с помощью этих соотношений, существенно отличается от экономичного размера заказа только при условии, если $n \leq 3$.

2.6. Производство циклами

В тех случаях, когда производственные мощности обеспечивают производство продукции каждого типа в объеме ее годового потребления, и при этом осуществляется последовательный выпуск каждого типа изделий, целесообразно продолжать выпуск продукции до тех пор, пока не будет исчерпан запас продукции некоторого другого типа или на производство данной продукции не будет израсходована слишком большая часть запасов, в результате чего возникает опасность одновременного истощения запасов продукции двух других типов. Модели экономичных интервалов между выпусками партии изделий для случая, когда пополнение запаса производится с конечной скоростью, позволяют определить максимальный размер запаса (удвоенный средний для любого изделия) [5] как

$$I_j / \sum_i I_i = 2S_j(1 - S_j/P_j) / \sum_i S_i(1 - S_i/P_i).$$

Если запас j -й продукции составляет именно такую часть от общего уровня запаса, выпуск этой продукции прекращается и переходят к выпуску продукции с минимальной величиной разовой поставки. Таким образом обеспечивается производство всей продукции достаточно экономичными партиями без риска создания дефицита продукции двух типов одновременно.

3. ПРАВИЛА ПРИНЯТИЯ РЕШЕНИЙ ОТНОСИТЕЛЬНО УРОВНЯ РЕЗЕРВНОГО ЗАПАСА

Решение относительно подачи заказа на пополнение запасов принимается системой управления первого уровня и зависит от результатов сравнения текущего состояния запаса (Наличный запас + Заказанный объем продукции — Заявки заказа) с максимальным значением разумных потребностей на интервале выполнения, которое хранится как точка заказа в информационных массивах первого уровня. Отсчет времени выполнения поставки изделия в запасы (время упреждения) начинается с момента получения заявки, которая уменьшает наличный запас изделия ниже точки заказа. Время поставки включает собственно время обработки заказа, время, необходимое поставщику на выполнение заказа, и некоторое требуемое в дальнейшем время на получение материалов и передачу их для использования.

Максимальный разумный спрос представляет собой сумму прогнозируемого спроса в течение времени упреждения и резервного запаса. Резервный запас равен произведению коэффициента резервного запаса на стандартное отклонение ошибок прогнозирования спроса в течение времени выполнения заказа на пополнение запаса, т. е. $k\sigma$.

Предположим, что системы прогнозирования, рассмотренные в гл. I-4, обеспечивают получение оценки стандартного отклонения σ . Если время упреждения известно с достаточной степенью точности, то величина σ может быть оценена с учетом ошибок прогнозирования спроса. Если время упреждения изменяется непредсказуемо, то для свертывания распределения этого времени с распределением ошибок прогнозирования опасно использовать любую модель. Результирующее распределение весьма чувствительно к предположениям относительно статистической зависимости (часто предполагают независимость), существующей между этими распределениями.

Можно показать, что между этими распределениями существует положительная или отрицательная корреляция; можно также доказать, что реальные данные практически не дают оснований для предположения о независимости этих распределений. Интервалы упреждения для продукции разного типа, естественно, различны. Первая задача состоит в том, чтобы найти причину раз-

личный и, если можно, устранить ее. Многие компании договариваются с поставщиком о конкретных датах поставок, благодаря чему становится известно время упреждения. Это в свою очередь позволяет определить необходимый уровень резервного запаса.

Если практически невозможно управлять временем упреждения, то следует оценивать спрос непосредственно в течение этого времени. Согласно определению, момент подачи заказа на пополнение запаса является началом отсчета времени упреждения. В этот момент заявки, поступившие к моменту подачи запаса с начала года, должны быть зафиксированы системой учета открытых заказов. Момент, когда поставка заказанных материалов осуществлена и переведена в запас, по определению является концом интервала упреждения. В этот момент нужно вычесть из объема заявок, поступивших до него, объем, зафиксированный в момент подачи заказа. Полученная разность характеризует спрос в течение времени упреждения и является переменной, для которой требуется определить стандартное отклонение без допущений относительно существования (или отсутствия) зависимости между спросом и временем упреждения.

В рассматриваемых ниже случаях будем предполагать, что распределение ошибок прогноза является нормальным с нулевым средним значением. Хороший прогноз имеет распределение ошибок с нулевым средним, и если интервал между последовательными проверками прогноза намного меньше, чем одна треть времени упреждения, то распределение близко к нормальному. Однако в тех случаях, когда время упреждения меньше, чем интервалы между прогнозами, и, кроме того, уровни запасов близки к запросам потребителей, целесообразно определить вид распределения ошибок прогноза для данного момента времени, так как с течением времени оно может существенно меняться. Результаты использования рассмотренных правил принятия решений, которые вырабатываются системой второго уровня, применимы ко всем случаям, если пользоваться таблицами корректировки распределения вероятностей. Для удобства обычно достаточно представить спрос как непрерывную, а не дискретную случайную переменную.

3.1. Показатели уровня обслуживания

Резервный буферный запас, добавляемый к запасу, определяемому в соответствии с прогнозом времени упреждения до пополнения запаса, выражается через $k\sigma$, где k — коэффициент резервного запаса, который обычно ограничен неотрицательными значениями, определяемыми при выборе стратегии. Однако для

балансирования запасов в конце сезона сбыта может потребоваться допущение отрицательности k . Вероятность того, что спрос в течение следующего интервала упреждения превысит разумный максимум, выражается через функцию распределения вероятностей $F(k)$ в единицах $(x - \mu)/\sigma$. Иными словами, если спрос $(x = \mu + k\sigma)$ равен заказу в точке заказа, то выполняется соотношение $(\mu + k\sigma - \mu)/\sigma = k$. Значения функции распределения вероятностей получаются интегрированием соответствующей плотности распределения $p(k)$ в направлении справа налево (в противоположность принятому в статистике), т. е. в области значений вероятностей от разумного спроса к бесконечности.

Введем понятие частичного математического ожидания, т. е.

$$E(k) = \int_k^{\infty} (t - k) p(t) dt.$$

Поскольку ордината правой части кривой распределения вероятностей пропорциональна вероятности возникновения дефицита продукции и недостающий объем продукции выражается как $\sigma(t - k)$, то ожидаемый объем недостающей продукции на один цикл пополнения будет равен $\sigma E(k)$.

Шлейфер [19] назвал функцию $E(k)$ «линейной функцией потерь», а в справочниках фирмы IBM она называется «функцией обслуживания». Значения функций, описывающих нормальное, экспоненциальное, равномерное и логарифмически-нормальное распределения, приводятся в работе [35].

Вероятность возникновения дефицита продукции сразу после получения партии, пополняющей запас, практически равна нулю. Если же уровень запаса соответствует нижнему пределу размера партии непосредственно перед получением новой партии, то возникает дефицит продукции и число возможных случаев нехватки запаса в течение года равно $(S/Q) F(k)$, а значение предполагаемого неудовлетворенного спроса в денежном выражении — $(S/Q) \sigma v E(k)$ долл. Таким образом, оба показателя уровня обслуживания зависят от количества пополнений запасов в год (S/Q) и, следовательно, коэффициент резервного запаса также зависит от частоты пополнения запасов.

Если спрос, который не может быть удовлетворен за счет запаса, удовлетворяется после получения следующей партии продукции от поставщика, то средний объем капиталовложений в резервный запас составит $k\sigma v$. Однако, если спрос, который не может быть удовлетворен из наличных запасов продукции, в дальнейшем не принимается во внимание, то средний объем капиталовложений будет выше и равен $v[k + E(k)]\sigma$.

3.2. Кривые распределения

Каждое правило принятия решений по управлению запасами содержит ту или иную управляющую переменную, которая может быть использована для регулирования капиталовложений в резервные запасы (выраженных в долларах) в зависимости от некоторого показателя уровня обслуживания потребителей, имеющего размерность [1/год]. Таким показателем может служить годовое число случаев дефицита запасов продукции, годовое количество (в долларах) учтенных, но не выполненных заказов потребителей или годовое количество заказов потребителей, которое не может быть удовлетворено из наличных запасов. Выбор конкретного значения управляющей переменной в одном случае может привести к высокому объему капиталовложений и высокому уровню обслуживания потребителей, а в другом случае — к уменьшению объема капиталовложений и ухудшению уровня обслуживания.

Одной из важных особенностей системы управления третьего уровня являются так называемые кривые распределения, которые показывают (графически или в табличной форме — в зависимости от способности органов управления воспринимать информацию визуально или на слух) результаты выбора возможных значений управляющих переменных. Ордината каждой из точек кривой соответствует $\sum k\sigma v$, где суммирование производится по всей продукции данного типа (с очевидным дополнением в случае, когда неудовлетворенный спрос означает экономические потери), а абсцисса представляет собой годовые неудовлетворенные, но учтенные заказы — $\sum (S/Q) v \sigma E(k)$ или число случаев дефицита запасов в течение года — $\sum (S/Q) F(k)$.

На рис. 9 показаны кривые, полученные для двух возможных вариантов правил принятия решений по управлению запасами одной и той же продукции. Для общности отображенные на рисунке правила будем называть правилами «А» и «В». Очевидно, что если объем учтенного, но не удовлетворенного спроса отражает уровень обслуживания, то решающее правило «А» оказывается предпочтительнее решающего правила «В»; если же число случаев дефицита запасов — фактор более важный, то порядок предпочтения меняется на обратный.

Учтенные, но невыполненные заказы (или экономические потери) чаще всего являются основными показателями уровня обслуживания так называемых систем управления «конечными» запасами, которые удовлетворяют спрос потребителей и в которых запасы относительно изолированы от поставщика. Число случаев нехватки запасов может быть более подходящей оценкой для систем, в которых фактический спрос может настолько превышать

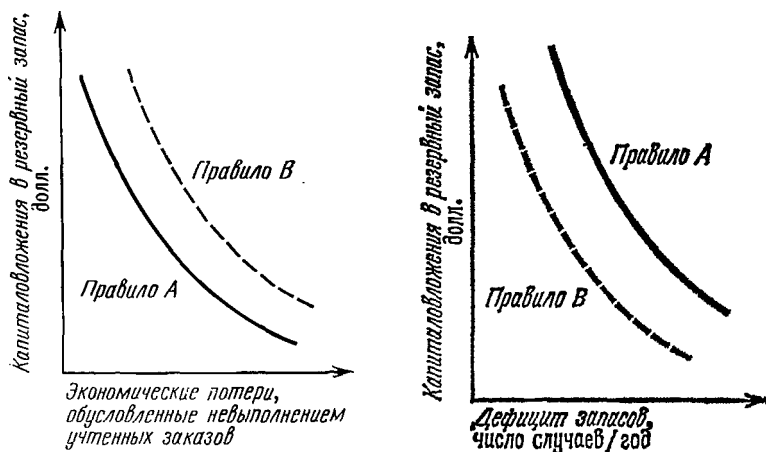


Рис. 9. Два возможных варианта правил принятия решений по управлению запасами одной и той же продукции.

прогнозируемый, что приемлемый резервный запас оказывается недостаточным, но имеется возможность быстро изготовить необходимый объем продукции и получить материалы ранее установленного срока поставки.

Для того чтобы решить, какие меры должны быть приняты в каждом конкретном случае, органы управления любого предприятия должны учитывать рыночные и производственные возможности работы вообще без запасов. Симпсон [21] показал, что для минимизации общего объема капиталовложений в запасы в многоуровневой системе материально-технического обеспечения риск невыполнения заказов потребителей допустим лишь на последнем уровне управления. На всех остальных уровнях необходимо выполнять все требования. Система управления базовыми (нормальными) запасами, описанная в работе [12], позволяет учесть колебания спроса. Располагая подобной информацией, можно с помощью эффективной системы доставки организовать получение необходимых материалов раньше, чем это планировалось, и таким образом демпфировать колебания спроса, даже если на промежуточных уровнях отсутствуют резервные запасы.

Кривые распределения затрат строятся для некоторых фиксированных возможных значений управляющих переменных. Как показано на рис. 10, сравнение текущих капиталовложений в резервный запас и текущего фактического уровня обслуживания помогает органам управления своевременно принять решение относительно будущих объема капиталовложений и уровня

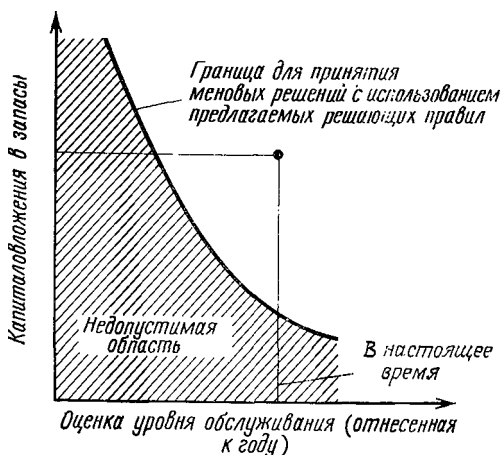


Рис. 10. Кривая распределения затрат.

обслуживания. Так как кривые распределения затрат наглядно демонстрируют физическую сущность различных вариантов решений, то именно они могут быть использованы для преодоления противоречий между органами управления, ответственными за финансы, и органами управления, ответственными за сбыт.

Определение необходимого объема капиталовложений для эффективности сбыта является прерогативой руководителя, идущего на риск, и этот объем не имеет «идеального» значения, которое можно было бы объективно оценить с помощью методов исследования операций. Ниже приводятся некоторые общие решающие правила определения значения коэффициента k резервного запаса по всей запасаемой продукции и даются рекомендации относительно применения каждого из них.

3.3. Некоторые общие правила определения коэффициента резервного запаса

Правило трехнедельного запаса. Во многих существующих системах управления запасами (обычно достаточно сложных) резервным считается трехнедельный запас, при этом число 3 является конкретным выбранным значением управляющей переменной. Это правило крайне слабо учитывает тот факт, что показатель колебания спроса (отношение стандартного отклонения к среднему значению прогноза) изменяется в четыре раза даже для продукции с примерно одинаковыми ожидаемыми объемами реализации. Следовательно, фактическое значение коэффициента

резервного запаса для одной продукции может оказаться крайне малым и в результате приведет к низкому уровню обслуживания, а для другой продукции оно может оказаться излишне большим и приведет к неоправданно высоким капиталовложениям в запасы. Правило трехнедельного страхового запаса широко используется консультантами по операционному анализу, так как с его помощью удастся без больших усилий продемонстрировать значительный выигрыш в эффективности управления запасами.

Правило постоянного коэффициента резервного запаса. В тех случаях, когда требуется, чтобы правила принятия решений по управлению запасами были просты в разработке и внедрении, удобны с точки зрения их документирования, текущего использования и обучения персонала пользователей, предпочтительным является правило, согласно которому резервный запас должен составлять 1,1σ для всех типов продукции, составляющей данный уровень запасов. Значение коэффициента резервного запаса $k = 1,1$ является управляющей переменной. Поскольку данное правило предусматривает точный расчет стандартного отклонения ошибок прогноза, то при его использовании требуется значительно меньший объем капиталовложений для любого заданного уровня обслуживания по сравнению с предыдущим правилом. Основная сложность применения данного правила состоит в оценке величины стандартного отклонения, которая является частью системы прогнозирования и требуется всегда, когда нужно улучшить эффективность функционирования системы управления запасами.

Правило постоянства коэффициента резервного запаса часто является полезным промежуточным этапом реализации варианта системы управления, так как позволяет сотрудникам организации освоить способы вычисления резервных запасов с учетом стандартного отклонения, а не только собственного прогноза.

Правило минимального числа невыполненных заказов. Рассмотрим задачу определения коэффициента страхового запаса каждого типа продукции при заданном уровне запаса, минимизирующего число невыполненных заказов в год при условии, что объем капиталовложений в страховой запас равен I_0 . Впервые эта задача была формализована Джерсоном [10] путем перевода вполне реалистичных целей органов управления с «управленческого» языка на язык математики.

Ожидаемый объем (долларах) невыполненных заказов в год равен $\sum (S/Q) \text{ } \forall E(k)$ причем каждая из переменных характеризует конкретные типы продукции, составляющей данный уровень запасов при ограничении объема капиталовложений в резервный запас — $\sum k \sigma v = I_0$. Задача без ограничений состоит в ми-

минимизации функции Лагранжа

$$H(k_i, \lambda) = \sum_i S_i/Q_i v_i \sigma_i E(k_i) + \lambda (\sum k_i v_i \sigma_i - I_0),$$

где λ — множитель Лагранжа или управляющая переменная, характеризующая число невыполненных заказов в год.

Предположим, что частная производная по коэффициенту резервного запаса каждого типа продукции $\partial H/\partial k_j$ равна 0. Так как все члены суммы выражения для $H(k, \lambda)$ не зависят от k_j , то имеем уравнение $(S_j/Q_j) v_j \sigma_j F(k_j) = \lambda v_j \sigma_j$, откуда следует решающее правило, согласно которому $F(k_j) = \lambda/(S_j/Q_j)$. Таким образом, задача состоит в том, чтобы гарантировать одно и то же ожидаемое число невыполненных заказов на продукцию всех типов с учетом частоты пополнения запасов. Значения управляющей переменной могут быть установлены органами управления на основе анализа соответствующей кривой распределения и использования данных анализа либо для удовлетворения ограничения на объем капиталовложений (или занимаемое пространство), либо для достижения некоторой цели в сфере сбыта.

Если данное правило применяется при управлении запасами продукции класса А, которые пополняются часто, то получаем большие значения коэффициента резервного запаса. В случае продукции класса В, которая может производиться при необходимости в количестве одно- или двухлетнего запаса, формула может давать отрицательные значения коэффициентов резервного запаса, часто произвольно ограничиваемых снизу нулем, т. е. для продукции, запас которой пополняется через длительные интервалы времени, достигается более высокий уровень обслуживания по сравнению со случаем, когда вообще отсутствует резервный запас. Однако, если в наличии имеется дополнительный капитал, который может быть вложен в запасы, то увеличение резервного запаса продукции класса А приведет к более значительному сокращению числа невыполненных заказов, чем увеличение резервного запаса продукции класса С при том же объеме дополнительных капиталовложений.

Правило минимального числа случаев дефицита запасов. Если задачей управления запасами является минимизация числа случаев дефицита запасов при заданном объеме капиталовложений, то она формулируется аналогично предыдущей задаче с заменой ожидаемого числа случаев дефицита на ожидаемое значение потерь из-за дефицита, выраженных в долларах. В связи с этим решающее правило предусматривает нахождение (если он существует) такого коэффициента резервного запаса k_j , который удовлетворял бы уравнению $p(k_j) = \lambda v_j \sigma_j / (S_j/Q_j)$, где управляющая переменная (множитель Лагранжа) λ имеет размерность [1/долл.] и может быть интерпретирована как предельная величина потерь,

которых можно было бы избежать с помощью бесконечно малого приращения объема капиталовложений в запасы. Это приращение представляет собой устанавливаемые, а не калькулируемые затраты. Выбираемое руководством тактическое значение λ должно ежегодно пересматриваться с учетом кривой распределения.

3.4. Другие решающие правила

Большое разнообразие решающих правил обусловлено выбором различных целевых функций, подлежащих минимизации в пределах некоторых бюджетных ограничений. Например, органы управления могут считать целесообразным минимизировать количество таких заказов потребителей, которые не могут быть удовлетворены за счет запаса. Браун [7] показал, что оптимальное решение этой задачи заключается в содержании текущих запасов продукции, имеющей низкую отпускную цену, и в допущении риска дефицита запасов для продукции с высокой отпускной ценой партии Q_v .

В другой задаче рассматриваются расходы по управлению в расчете на одну заявку и доплаты, связанные с транспортными затратами на единицу веса материалов, которые приходится нести при удовлетворении спроса из другого источника, когда данной продукции нет на складе. В этом случае решающее правило аналогично правилу минимального числа невыполненных заказов и состоит в том, чтобы стремиться к уменьшению коэффициента резервного запаса для материалов с малым весом или продукции, которая пользуется большим спросом.

3.5. Резервный запас и частота его пополнения

Ранее предполагалось, что число случаев дефицита запасов S/Q определяется средним размером заказа и объемом годовой реализации S продукции. Общие годовые затраты, связанные с размещением заказов на пополнение запасов с хранением рабочего (величиной $1/2Q$) и резервного запасов, могут быть представлены как

$$C = \sum AS/Q + r \sum v \left(k\sigma + \frac{1}{2} Q \right),$$

причем если невыполненный заказ не фиксируется, то добавляется член $v\sigma E(k)$. В результате минимизация общих годовых затрат означает минимизацию целевой функции вида

$$H(Q_j, k_j, \lambda) = C + \lambda \left(\sum (S_j/Q_j) v_j \sigma_j E(k_j) - B \right)$$

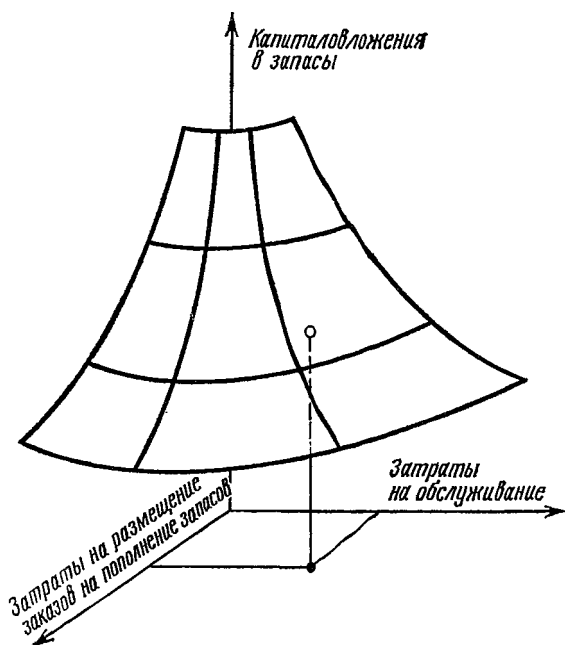


Рис. 11. Поверхность трехмерного распределения.

при ограничениях на общую стоимость B (долл.) учтенных, но невыполненных заказов. После соответствующих преобразований получаем

$$F(k_j) = (r/\lambda)/(S_j/Q_j), \quad (1)$$

$$Q_j = \sqrt{2(A_j + \lambda v_j \sigma_j E(k_j)) S_j / r v_j}. \quad (2)$$

Из уравнений (1) и (2) следует, что как затраты r на хранение запасов, так и множитель Лагранжа λ являются управляющими переменными. На рис. 11 представлена поверхность трехмерного распределения. Практически оказывается довольно трудно одновременно определить коэффициент резервного запаса k и размер партии продукции Q с помощью этих двух уравнений, и поэтому обычно они оцениваются с помощью следующей итеративной процедуры. Сначала предполагается, что значение Q равно размеру партии продукции при условии отсутствия дефицита, и решается уравнение (1) относительно k . Подставляя полученное значение k в уравнение (2), решают его относительно Q . Подобная процедура повторяется несколько раз. Обычно примерно три итерации дают допустимое приближение. Для того чтобы приве-

денные выше результаты были наилучшими в смысле величины общих затрат, необходимо, чтобы размер партии, вычисленной при условии отсутствия дефицита и используемый в качестве начального значения Q , был не меньше стандартного отклонения. Поэтому для экономии затрат на итеративные вычисления полученный начальный размер партии необходимо округлить до величины не меньшей, чем одно стандартное отклонение. Эта величина размера партии не является, вообще говоря, той, которая будет получена с помощью приведенных выше решающих правил, но если она достаточно большая, то не имеет смысла вычислять ее точное значение.

Заметим, что член $\lambda \sigma E(k)$ характеризует дополнительные ожидаемые затраты, связанные с удовлетворением учтенных, но не выполненных заказов, что имеет место один раз в каждом цикле подачи заказа, — аналогично затратам A , связанным с запуском продукции в производство. Таким образом, получаемая величина размера заказа превосходит вычисленную при условии отсутствия дефицита, что приводит к уменьшению частоты подачи заказов на пополнение запасов, которая в свою очередь сокращает уровень резервного запаса настолько, что становится оправданным увеличением рабочего запаса.

Если любой из коэффициентов в этом выражении равен нулю, то величина размера партии становится равной экономичному размеру заказа.

3.6. Запасы комплектующих изделий

Рассмотрим случай, когда в течение недели должна быть выполнена сборка определенного числа грузовых автомобилей в соответствии с планом производства. Каждый автомобиль имеет свои особенности, о которых потребитель может судить на основе перечня предлагаемых моделей изделий. При этом конкретный набор заказов, которые должны быть выполнены в течение недели, заранее не известен. Поэтому необходимо иметь в наличии достаточный объем каждой продукции, с тем чтобы предусмотреть некоторую разумную совокупность разнородных заказов. Предположим, что цель управления запасами состоит в минимизации числа случаев дефицита запаса любой продукции с учетом ограничения на общий объем капиталовложений в ее запасы. Если вероятность заказа i -й продукции равна p_i , то распределение совокупности разнородных заказов задается полиномиальным распределением. Безусловное распределение для i -й продукции является биномиальным и при больших объемах продукции может быть аппроксимировано нормальным распределением (распределением Гаусса). Решение относительно уровня запаса уменьшает число степеней свободы на единицу, так как общее количество

подлежащих выполнению заказов должно быть сведено к общему числу запланированных к сборке автомобилей. В данном случае решающее правило аналогично правилу минимального числа случаев дефицита запаса. Основное отличие состоит в том, что мы делаем выводы на основании вычисления стандартного отклонения с помощью формулы $\sigma = \sqrt{Npq}$, а не на основании оценки ошибок прогноза

4. РАСПРЕДЕЛЕНИЕ ОГРАНИЧЕННЫХ РЕСУРСОВ

Существует несколько способов распределения ограниченных ресурсов при выполнении различных операций в системе управления запасами. Например, партия краски или зерна должна быть расфасована таким образом, чтобы имелись упаковки различных размеров; запасы готовой продукции должны распределяться по множеству периферийных складов; объем железнодорожного вагона должно распределяться между различными грузами одной станции назначения. Каждый из рассмотренных примеров является частным случаем общей задачи распределения.

Правило принятия решений по определению объема ресурсов, которые целесообразно выделить для каждого процесса с учетом объема израсходованных ресурсов, обычно разрабатывается с учетом максимизации некоторой целевой функции каждого процесса при ограничении на общий объем наличных ресурсов. Часто это приводит к тому, что для какого-то процесса выделено ресурсов больше, чем в случае справедливого распределения. В этой ситуации имеющиеся ресурсы и потребности рассматриваются вновь и устанавливается распределение для меньшего числа процессов. В результате получается конечное число итераций, причем каждый вид ресурсов должен быть выражен хотя бы для одного процесса.

4.1. Критерии распределения ресурсов

Рассмотрим некоторые возможные критерии оптимального запаса материалов для каждого процесса. По-видимому, простейшим для понимания и использования решающим правилом является выделение для каждого процесса одинакового количества единиц запаса. Пусть дополнительное количество запаса, которое должно быть распределено между процессами, равно Q . Каждый процесс характеризуется текущей интенсивностью потребления S_i и текущими наличными запасами a_i . Общий запас в системе будет сохраняться в течение времени $T = (Q + \sum a_i) / \sum S_i$. Следовательно, количество единиц запаса, которые должны быть выделены для i -го процесса, равно $q_i = TS_i - a_i$, если

эта разность положительна. Обычно непрактично разгружать контейнеры, которые перегружены сверх нормы, и неэкономично привозить назад продукцию с переполненных складов. Поэтому, если величина q_j отрицательна, то необходимо из сумм в выражении для времени T исключить величины S_j и a_j и определить T' , которое будет меньше, чем T . Майер (см. работу [4]) дал логическое обоснование для распределения, при котором минимизируется ожидаемый остаточный запас. Предположим, что партия продукции поставляется в виде упаковок разных размеров и что новая партия требуется всякий раз, когда истощается запас в какой-либо одной упаковке (емкости). Запас, имеющийся в данный момент в емкости, называется остаточным. Запас должен быть распределен между емкостями таким образом, чтобы вероятность того, что в первую очередь передается запас j -й емкости, равнялась отношению $S_j / \sum S_i$. Если заявка детерминированная, то остаточный запас будет отсутствовать. Работа с приведенным выше выражением для определения вероятности того, что в первую очередь истощится запас в j -м контейнере, встречает затруднения. Хорошие результаты могут быть получены при использовании в выражении для времени T среднего значения времени истощения запаса по каждой продукции. Прогнозируемый объем реализации продукции $S_i(T)$, так же как и стандартное отклонение $\sigma_i(T)$, зависит от интервала времени обеспечения запасом. Для каждого процесса коэффициент резервного запаса должен быть таким, чтобы функция $F(k_j)$ была пропорциональна отношению $\alpha S_j / \sum S_i$. В этом случае целевая функция имеет вид $S_j(T) + k_j \sigma_j(T)$. Коэффициент пропорциональности α находится в пределах $0,5 < \alpha < 0,693$ и является корнем полинома

$$G(\alpha) = \prod_j (1 - \alpha S_j / \sum S_i) = 0,5.$$

В качестве других критериев распределения продукции может быть минимизация ожидаемого общего числа невыполненных заказов или числа случаев нехватки запасов.

Следует иметь в виду, что правила решения относительно уровня резервных запасов, рассмотренные в разделах 3.5–3.9, были получены для распределения ограниченного бюджета капиталовложений среди продукции различного типа при заданном уровне запасов.

4.2. Оценка эффективности решающих правил

Поскольку каждое решающее правило при дальнейшем уточнении постановки задачи может связываться с достижением множества целей, то при оценке возможных правил следует учитывать побочные эффекты, такие, как средний запас в системе в целом,

число решений по распределению ресурсов в течение года, ожидаемое число случаев дефицита запасов и величина этого дефицита. Может оказаться, например, что такое простое правило, как выделение равного числа единиц запаса для каждого процесса, будет иметь совершенно ясный физический смысл и давать хорошие результаты сразу во всех аспектах, в то время как некоторое, более точное правило даст возможность детально учитывать какой-либо один фактор, но будет приводить к нежелательным побочным явлениям. Такие «невыгодные» правила могут быть исключены с помощью имитационной модели системы третьего уровня, если моделирующий алгоритм будет настроен на оценку сразу нескольких аспектов функционирования системы при использовании того или иного репашующего правила.

5. АРХИТЕКТУРА СИСТЕМ УПРАВЛЕНИЯ ЗАПАСАМИ

Комбинируя описанные выше элементарные операции управления, можно создать различные системы управления запасами. Достаточно глубокое понимание теоретических основ разработки правил принятия решения относительно пополнения запасов может быть достигнуто при исследовании моделей системы управления запасами с нулевым временем поставки заказа или с детерминированным спросом. При использовании систем управления запасами, как правило, приходится иметь дело со случайным спросом в течение времени поставок заказа. Большинство теоретических результатов получено в предположении, что данное решающее правило будет применяться многократно, однако в действительности реальные интервалы поставок материалов, издержки и распределение спроса меняются во времени. Таким образом, принятое в данный момент решение относительно пополнения запасов является наилучшим лишь для этого момента.

В настоящем разделе мы рассмотрим некоторые возможные методы решения задач системы первого уровня, т. е. задач размещения очередного заказа на пополнение запаса некоторой продукции и определения размера этого заказа.

5.1. Подача независимых заказов на пополнение запасов

В тех случаях, когда можно заказать любую продукцию в любое время, не учитывая состояние дел с продукцией других типов, можно использовать следующие модели для определения объема и сроков заказа.

Модель точки заказа и размера партии. Точка заказа или уровень запаса, при котором подается заказ на пополнение,

прогнозируемой потребности в ней на интервале времени поставки пополнения и некоторого резервного запаса. Последний равен произведению стандартного отклонения распределения ошибок прогноза в течение времени поставки на коэффициент резервного запаса. Коэффициент страхового запаса может быть вычислен в результате применения любого из решающих правил, рассмотренных в разд. 3, или любой пригодной их модификации.

Система первого уровня следит за состоянием запаса, который равен: Уровень наличного запаса $+$ Объем заказанных поставок $-$ Объем учтенных, но не выполненных заявок или Объем продукции по другим обязательствам. В тех случаях, когда уровень наличного запаса снижается в результате удовлетворения спроса до или ниже точки заказа, подается заказ на пополнение запаса. Объем заказа вычисляется на основании подходящего решающего правила, которое разрабатывается системой второго уровня (разд. 2). Полученный объем заказа может быть округлен до величины, удобной с точки зрения получения поставок или кратного числа упаковок, поставок или единиц оборудования без существенного сокращения экономической выгоды. Решающие правила второго уровня обычно проверяются через определенные интервалы времени, например один раз в месяц, и результаты фиксируются в информационных массивах системы первого уровня.

Такой простейший для понимания и применения способ довольно часто оказывается весьма эффективным. Например, в производстве красок и лекарств размер цистерн физически ограничивает объем заказа, так что имеет смысл лишь predeterminedный заранее объем заказываемой партии.

Модели типа максимум/минимум, или (s, S). Минимальный уровень запасов s вычисляется аналогично тому, как вычисляется точка заказа в предыдущей модели. Максимальный уровень S равен сумме минимального уровня и объема заказа, получаемого с помощью соответствующего решающего правила определения размера партии (разд. 2). Значения этих уровней вычисляются через одинаковые промежутки времени в системе второго уровня, и результаты фиксируются системой учета первого уровня.

Система первого уровня непрерывно следит за уровнем наличного запаса, и если он снижается до величины s , то заказывается объем продукции, равный разности между максимальным и фактическим уровнями запаса. Если имеется возможность сделать новый заказ после удовлетворения каждой единицы спроса (в результате чего фактический запас снижается точно до минимального уровня), то система функционирует совершенно аналогично преды-

дущей. Однако, если информация о наличии запаса передается через большие интервалы времени или если существует одиночный спрос на большой объем продукции, то существует возможность того, что к моменту распознавания необходимости пополнения запаса уровень наличного запаса снизится ниже точки заказа и система устранит этот дефицит. Этот способ интенсивно исследовался Эрроу [1] и является более предпочтительным по сравнению с моделью «точка заказа и размер партии» при наличии существенного дефицита в течение времени выполнения заказа на пополнение запасов. В этом случае объем заказываемой партии также может быть определен на основании предположения о его кратности числу упаковок, числу поставок или количеству единиц оборудования.

Модель с фиксированным интервалом. Если поставщик принимает заказы только в течение определенного интервала времени, то прием заказов между этими интервалами отсутствует. Например, машиностроительный завод может принимать заказы только до 24 числа каждого месяца, поскольку к этому времени заканчивается формирование плана производства на очередной месяц. Другой пример — судно, снабжающее корабли некоторой флотилии и прибывающее к ним один раз в 10 дней. В обоих указанных примерах потребитель вынужден ждать, пока поставщик будет готов к приему нового заказа. В этом случае точка заказа или минимальный уровень запаса вычисляется на основе прогноза потребностей в течение времени поставки пополняющего запаса и дополнительного интервала ожидания возможности подачи заказа с учетом, как и ранее, резервного запаса, компенсирующего ошибки прогноза потребностей в течение всего рассматриваемого интервала. Если наличный запас в момент проверки ниже, чем новая точка заказа, то пополняющий заказ выполняется либо в заранее установленном объеме, либо в объеме запаса, определяемого разностью между максимальным и наличным уровнями. Заметим, что минимальный объем заказываемой продукции должен быть достаточным по крайней мере для того, чтобы в интервале ожидания возможности подачи заказа на пополнение запаса он полностью не исчерпывался. Эта схема используется также в ряде случаев как аппроксимация объединенных заказов, когда необходимо регулирование запасов сразу нескольких видов. В общем случае значения точки заказа или минимального уровня запаса и объем заказа или максимальный уровень запаса всякий раз при пересмотре прогноза должны быть пересчитаны. В некоторых системах точка заказа пересматривается с каждым новым прогнозом, но объем заказа или максимальный уровень запаса вычисляются только непосредственно перед размещением заказа.

5.2. Частный случай системы типа «точка заказа и размер партии»

Система с двумя бункерами является частным случаем системы типа «точка заказа и размер партии» и требует минимального числа учетных операций и весьма эффективна для продукции класса «С». В этой системе объем запаса, равный точке заказа, выделяется физически и размещается в резервном бункере, или в запломбированной емкости. Основной запас размещается в рабочем бункере. При этом потребителю разрешается в случае необходимости получать продукцию из рабочего бункера без соответствующей регистрации. Когда рабочий бункер освобождается, вскрывается резервный бункер и из него извлекается путевой лист для внесения отметок о пополнении запаса. Для гарантии того, что путевой лист изъят из резервного бункера и передан в производственный отдел или отдел снабжения, где он может использоваться в качестве учетного документа, необходимы высокая сознательность и дисциплина. Одновременно поставщик может использовать путевой лист как накопительный и отгрузочный документ, который возвращается потребителю при каждой новой поставке. После этого резервный бункер вновь заполняется до уровня точки заказа полученной продукцией, закрывается и пломбируется. Основной запас помещается в открытый рабочий бункер. Объем продукции, заказанной в какой-либо момент времени, должен обеспечивать наличие ненулевого запаса в течение времени на одного более продолжительного, чем интервал времени ожидания поступления пополняющего запаса, с тем чтобы в системе в любой момент времени подавался не более чем один заказ на пополнение. Это утверждение справедливо для продукции класса «С» в большинстве видов запасов.

Следует иметь в виду, что при такой системе управления запасами возможны некоторые потери и несанкционированное сокращение запаса, но исключение всяких регистрационных операций, кроме заполнения путевого листа, может дать экономию связанных с управлением средств, которая значительно превысит затраты на несовершенный контроль.

Информация об интенсивности потребления может быть получена из путевого листа, так как он содержит сведения об объеме продукции в рабочем бункере и относительно даты последнего пополнения. Поскольку запас в резервном бункере тоже известен, то интенсивность потребления равна размеру рабочего запаса, деленному на время, прошедшее с момента его пополнения до открытия бункера с резервным запасом. Для оценки размера пополняющего запаса используется экономичный размер запаса, определяемый на основе текущей закупочной цены. Используя результаты одноразового исследования, имеющего целью установ-

ление соотношения между средней интенсивностью потребления и стандартным отклонением ошибок прогноза спроса в течение времени поставки заказа, и правило «3 σ », можно составить таблицу значений резервных запасов. Эта таблица дает возможность определить точки заказа, основанные на текущей оценке интенсивности потребления.

5.3. Неравномерность спроса

Спрос на большинство видов запасов (в частности, запасных частей для машин) может носить неравномерный характер, т. е. наблюдаются периоды различной продолжительности, в течение которых спрос практически равен нулю, незначителен или резко возрастает. Если резкое возрастание спроса (которое может быть следствием специальных проектов или контрактов) может быть предсказано, то в этом случае возможно спланировать поставку продукции (разд. 5.4), в противном случае применение классических методов управления запасами (в частности, прогнозирования) оказывается весьма затруднительным. Большим числом корпораций в этой ситуации успешно была использована следующая система.

Запас пополняется в соответствии с максиминной системой (s, S). При определении размера заказа, используемого для определения максимального уровня запаса, учитывается взаимосвязь между размерами партии и резервными запасами (разд. 3.8), при этом размер партии должен быть не меньше, чем стандартное отклонение σ . Точка заказа должна превышать средний минимум на величину ожидаемого дефицита, или на среднюю величину разности точки заказа и наличного запаса в моменты подачи пополняющих заказов. Оценка величины этого ожидаемого дефицита равна $\langle d^2 \rangle / 2 \langle d \rangle$, где $\langle \cdot \rangle$ обозначает ожидаемое значение, а d — число единиц продукции в отдельной заявке или в партии, отправляемой на склад. Заметим, что если все заявки одинакового объема (равномерный спрос), то ожидаемый дефицит сокращается до величины, равной половине ожидаемого объема спроса. Для случая равномерного спроса этот дефицит значительно больше.

Таким образом, в случае обычной продукции система управления с неравномерным спросом превращается в классическую минимаксную систему. Неравномерность спроса оценивается с помощью коэффициента вариации (σ/μ), а не абсолютного уровня спроса. Если коэффициент вариации (или любая другая удобная оценка относительной изменчивости спроса) больше единицы, то продукция может рассматриваться как продукция, на которую существует равномерный спрос. В связи с этим продукция в зависимости от характера спроса на нее должна пересматриваться не чаще чем один раз в год при условии, чтобы обычная продукция

становилась продукцией неравномерного спроса только в том случае, когда коэффициент вариации превышает 1,2, и, наоборот, продукция неравномерного спроса становится «обычной», если коэффициент вариации падает ниже 0,8. Способ управления запасом изделий с отношением σ/μ , близким к единице, в основном не будет изменяться.

На основе анализа текущего спроса невозможно получить информацию о том, оправдан ли частый пересмотр максимальных уровней. Именно в этом состоит существенное отличие данной системы управления от системы второго уровня для обычного случая, когда частота такого пересмотра известна. Можно собирать статистические данные о средней величине спроса и среднем квадрате величин спросов, которые пересматриваются после получения каждой заявки по методу экспоненциального сглаживания. Спрос в течение времени поставки заказа может фиксироваться после прихода каждого заказа, для того чтобы корректировать статистические данные о средних значениях и средних квадратах. После поступления не менее чем 30 заказов или через календарный год (в зависимости от того, что произойдет раньше) эти статистические данные используются для пересчета параметров максимальной системы.

5.4. Планирование, согласованное во времени

Описанные выше системы управления запасами следят только за общим балансом запасов по всей продукции, на которую поданы пополняющие заказы. Приход этого баланса определяется новыми заказами на пополнение запасов, а расход — выдачами продукции со склада (с соответствующими дополнительными операциями по очень быстрому или длительному поставкам). Будущие потребности выражаются как общий прогноз спроса в течение времени упреждения с учетом резервного запаса.

В некоторых ситуациях целесообразно следить за спросом на продукцию и наличием ее запасов не только в пределах, но и далеко за пределами времени упреждения. Например, если необходимо выполнить какие-то обязательства, касающиеся программы организации сбыта продукции, реконструкции оборудования потребителя, выдачи со склада партии продукции, равной размеру пополняющего заказа или каких-либо контрактов с жесткими сроками, спрос может учитываться и удовлетворяться независимо от прогноза обычного спроса. Так, при выполнении заказов на модные товары с доставкой по почте или при сбыте кондиционеров допускается возврат поставщику ранее проданной продукции. Естественно, что в данном случае необходим прогноз относительно возврата проданной продукции на некотором интервале времени, равно как и поступления новой продукции из источника снабже-

ния. Хранение более детальных сведений в системе учета первого уровня целесообразно по двум причинам. Во-первых, если зависимый спрос на более низких уровнях сборки определяется детализацией планов производства на более высоких уровнях, то потребности должны планироваться с некоторым опережением во времени с учетом необходимости временных задержек на каждом уровне сборки. Во-вторых, в тех случаях, когда с практической точки зрения выгоднее ускорять выполнение заказов, чем хранить резервные запасы, информация, относящаяся к конкретным моментам времени, дает более точную картину состояния дел, позволяющую судить о том, какой из запущенных в производство заказов должен быть обязательно выполнен ценой пересмотра графика, а какие можно снять с производства в целях высвобождения производственных мощностей.

Типичная запись учета (рис. 12) представляет собой одну колодку для каждого периода времени и одну строку для каждого независимо планируемого спроса на продукцию и вида источников ресурсов. Обычно существует так называемый период «замораживания» (сплошная черная линия между 8 и 9 периодами), в течение которого объем заявок на ресурсы считается фиксированным во времени. Заказы, появляющиеся за пределами этого периода, который может быть менее или более продолжительным, чем фактический интервал упреждения, являются условными. Они, например, используются для планирования потребности в материалах и трудозатратах, однако их объем и срок выполнения зависят от текущего спроса и уровня наличных запасов.

Условные заказы определяются следующим образом. Всем условным заказам приписываются объемы вне периода замораживания, равные нулю. Объем планируемых общих потребностей в рассматриваемом периоде вычисляется как алгебраическая сумма каждого вида потребностей, отраженных в записи. Резервный запас принимается равным произведению соответствующего коэффициента резервного запаса на величину стандартного отклонения с ошибок прогноза общего спроса в течение периода замораживания.

В последней строке записи указывается остаточный запас, имеющийся в каждом периоде. В первом периоде этот запас равен превышению фактически имеющегося у потребителя запаса над запланированным резервным запасом. (Следует иметь в виду, что эта разность может быть отрицательной.) В каждом периоде остаточный запас вычисляется как сумма остаточного запаса в конце предыдущего периода и заказа фирмы за вычетом общего запланированного спроса.

Таким образом, мы имеем дело с условным заказом, когда кончается период замораживания или когда остаточный запас в конце предыдущего периода становится равным нулю либо отри-

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	.	.	.	.	
	Прогнозы нормального спроса																											
	30	30	30	35	35	35	35	40	40	40	40	40	35	35	35	35	30	30	30	30	25	25	25	.	.	.	.	
	Потребности, выявляемые путем стабилизации потребностей более высоких уровней																											
			600			100				500					500				700									
	Специальные контракты и проекты																											
	Прогнозы возврата товаров																											
	0	0	0	0	0	0	0	0	0	0	0	5	5	5	10	10	10	10	15	15	15	15	15	.	.	.	.	
	Суммарные запланированные потребности																											
Резервный запас	200	30	30	630	35	35	135	35	40	40	340	40	1035	30	30	525	25	20	20	15	15	10	10	10	.	.	.	.
	Фиксированные заказы											Условные заказы																
Наличный запас	250		750			500					125			585		55		40			.	.	.	.	.	.	.	
	Чистый запас																											
	50	20	740	110	75	40	405	370	350	700	175	50	1035	0	555	525	0	40	20	0	25	10	0	.	.	.	.	

Рис. 12. Формат записи для периода планирования, согласованного во времени (посчитаемого, например, в неделях).

цательным. Иначе говоря, допускается падение уровня наличного запаса до или ниже запланированного уровня резервного запаса. Подобная ситуация на рис. 12 соответствует периоду 9, который находится вне периода замораживания и для которого прогнозируемый остаточный запас падает до -50 единиц. Размер этого условного заказа равен сумме фактических потребностей в течение оптимального числа плановых периодов (разд. 2.5) или некоторому фиксированному объему, который может определяться уровнем качества продукции или способом производства. В рассматриваемом примере первый условный заказ равен 1125 и обеспечивает удовлетворение общего запланированного спроса в течение последующих трех периодов и, кроме того, покрывает возникший в периоде замораживания дефицит в 50 ед. В течение 12-го периода остаточный запас падает до нуля, и последующие три периода обеспечиваются за счет условного заказа в периоде 13.

Условный заказ, поступающий в первом после интервала замораживания периоде, рассматривается как фиксированный заказ и не должен изменяться в последующие периоды при повторяющемся процессе планирования. Заметим, что фиксированный заказ может быть и нулевым.

5.5. Перечень планируемой к выпуску продукции

Иногда целесообразно начать производство какой-то продукции до того, как в ней появится потребность. Это позволяет обеспечивать полную загрузку имеющегося производственного оборудования и тем самым использовать его наиболее эффективным образом. Один из методов, применяемых для достижения этой цели, предусматривает составление перечня условных заказов по всем видам продукции данного предприятия. Перечень составляется с указанием таких сроков выполнения заказов, при которых достигается более высокая производительность труда, не приводящая, однако, к простоям оборудования.

5.6. Запасы продукции, пользующейся сезонным спросом

В любой отрасли промышленности, где экономически оправдана определенная интенсивность производства продукции, пользующейся сезонным спросом, необходимо планировать создание запасов такой продукции в течение сезонов неинтенсивного спроса. Для этого сначала следует отобрать продукцию, которая может храниться в течение длительного времени, а затем составить перечень отобранной продукции в порядке предпочтительности сезонного спроса на нее. Обычно предпочтение выражается как отношение трудоемкости производства продукции (в человеко-часах или машино-часах) к нормативной стоимости материалов.

Целью такой системы является минимизация затрат на хранение запасов трудоемкой продукции. С началом сезона неинтенсивного спроса, когда интенсивность производства продукции превосходит спрос на нее, осуществляется выпуск продукции, стоящей первой в перечне, в объеме, достаточном для удовлетворения спроса следующего пикового сезона. Если при этом остаются свободные производственные мощности, то начинают выпуск продукции, стоящей второй в перечне. Простая в изготовлении, но дорогостоящая продукция производится в течение всего года для удовлетворения текущего спроса. Трудоемкая, но дешевая продукция выпускается в виде одной партии, что обеспечивает хорошую загрузку оборудования при достаточно малых капиталовложениях в запасы.

5.7. Объединенные заказы

Если при заказе сходной продукции существуют общие издержки, не зависящие от типа продукции, то следует заказывать всю такую продукцию одновременно. Например, материалы, которые покупаются у одного поставщика, могут поставляться со скидкой в цене, если заказанное количество будет превышать некоторый минимум, и с меньшими транспортными расходами, если соблюдаются транзитные нормы; различные жидкие продукты могут разливаться в бутылки на одной линии упаковки с единственной ее переналадкой; множество различных деталей может штамповаться из одного и того же рулона материала путем последовательной смены штампов и т. п.

Подачу заказов на пополнение запасов для всей продукции можно осуществлять несколькими способами. Одним из них является размещение заказа через фиксированные интервалы времени, в другом случае такой заказ делается всякий раз, когда хотя бы для одного типа продукции достигается точка заказа. В последнем случае образуются значительные остаточные запасы (разд. 4.1). Третий способ предусматривает размещение заказа на пополнение запаса тогда, когда агрегированный прогнозируемый дефицит достигнет некоторого критического уровня.

Оценить возможный уровень запаса в конце интервала упреждения можно с помощью формулы $I_i(L) = I_i(0) - L\mu_i$, где L — общее время упреждения для группы изделий и μ_i — прогнозируемая интенсивность потребления i -й продукции. Для некоторой продукции величина $I_i(L)$ может быть отрицательной. Фактический коэффициент резервного запаса — $k_i = I_i(L)/\sigma_i$. Ожидаемый общий дефицит для группы изделий равен $\sum v_i \sigma_i E(k_i)$. Если для некоторой управляющей переменной λ значения этого выражения больше чем $\lambda \sum \mu_i$, то наступает момент подачи пополняющего заказа.

В случае когда типы продукции независимы, точку заказа не принято связывать с ожидаемым дефицитом. Тем не менее это вполне допустимо и широко используется на практике.

Заказ всей продукции производится на основе единого времени наличия запаса T . Для продукции, производящейся в каждом цикле, заказ должен иметь объем, достаточный для удовлетворения спроса в течение интервала $T\mu_i + k_i\sigma_i$. Для продукции, которая должна производиться только в каждом n -м цикле, заказываемая партия должна иметь размер, позволяющий удовлетворять спрос в течение периода $n_iT\mu_i + k_i\sigma_i$. Однако если запас, имеющийся к настоящему моменту, не будет исчерпан, то по крайней мере в течение времени T данную продукцию не следует заказывать совсем. Если подача заказа на группу продукции производится тогда, когда запас любой из них достигает точки заказа, то согласованный заказ для других типов продукции может быть получен путем сравнения наличного запаса с «возможной» точкой заказа, которую принято полагать равной части разности между обычной точкой заказа и максимально возможным запасом [20].

5.8. Многоуровневое физическое распределение запасов

Многоуровневым системам управления запасами было посвящено несколько научных работ, и в частности работа [18]. Во многих распределительных сетях государственного и промышленного назначения была успешно применена система типа «центр — периферия» и этим обеспечено значительное увеличение возможностей и сокращение денежных вложений в запасы. В такой системе каждый продукт, запас которого хранится более чем в одном пункте (см. разд. 1.4), имеет один базовый склад, определяемый как главный. Главные склады для различной продукции не обязательно совпадают по местоположению. В одних отраслях промышленности они размещаются ближе к предприятию или порту назначения, в других — в центре района сбыта готовой продукции. Функцией главного склада является поддержание центрального резервного запаса для перераспределения его между периферийными складами. Часто такую же функцию выполняет периферийный склад, который удовлетворяет спрос покупателей, находящихся в обслуживаемой им зоне торговли.

Результирующий запас в системе равен: Запас, находящийся на складе + Объем заказов фирмы на поставку — Объем невыполненных обязательств перед потребителями — Сумма запасов, уровень каждого из которых равен точке заказа соответствующего периферийного склада — Некоторый неперераспределяемый избыток продуктов. Новые заказы поставщику могут определяться путем сравнения этого наличного чистого запаса с точкой заказа центрального резервного запаса или посредством согласованного

планирования во времени. В любом случае общий резервный запас определяется стандартным отклонением распределения ошибки прогноза агрегированного спроса и коэффициентом резервного запаса.

При получении материала из источника снабжения на главном складе откладывается общий резервный запас, а остаток поставки распределяется между периферийными складами.

Каждый периферийный склад имеет точку заказа, определяемую прогнозом спроса на обслуживаемой им территории в течение времени упреждения перераспределения запасов главным складом с учетом резервного запаса, равного произведению стандартного отклонения распределения ошибок прогнозов на коэффициент резервного запаса. Наличный запас периферийного склада равен: Запас, находящийся на складе + Поставка, отправленная в его адрес, но еще не полученная, — Запасы, идущие на покрытие невыполненных обязательств перед местными потребителями. При достижении наличным запасом точки заказа производится перераспределение запасов. В этот момент становится допустимым распределение центрального резервного запаса (он больше не сохраняется, как это делалось при начальном распределении после получения продуктов на главный склад) и общий запас системы распределяется между всеми складами.

Складам отгружается только тот объем продукции, в котором они действительно нуждаются. Остаток сохраняется на главном складе до тех пор, пока не возникнет необходимость перераспределения продукции по заказу какого-либо другого периферийного склада. Если в процессе вычисления объемов распределения продуктов окажется, что некоторый периферийный склад уже имеет продукции больше, чем ему положено, то образовавшийся на нем избыток является неперераспределяемым избытком и вычитается из результирующего наличного запаса системы при принятии решения о том, следует ли делать дополнительные заказы поставщику. Если этот избыток велик, то можно временно назначить данный периферийный склад главным для распределения материалов другим нуждающимся в них складам, и тогда, разумеется, избыток перестает быть неперераспределяемым. В подобных случаях возможны различные варианты. В некоторых ситуациях расчет начального распределения производится до транспортировки материалов на главный склад, так что поставщик может поставлять продукцию по заказу непосредственно на каждый периферийный склад. Вся территория, рассматриваемая как одна система, может иметь склады только ограниченной зоны обслуживания, для таких, например, продуктов, как краска и кирпич, которые не транспортабельны на большие расстояния; центральные склады, предназначенные для обслуживания всей страны, могут располагаться только к западу от Скалистых гор. Может быть и такая

продукция, перераспределение запасов которой нецелесообразно, в этом случае при дефиците нужных запасов на складе, куда поступил заказ, последний переадресуется на другой склад. Могут существовать также основания для завышенного распределения каких-либо изделий, с тем чтобы гарантировать, что даже в условиях образования дефицита в системе в целом на конкретной обслуживаемой территории снабжение сохранится на должном уровне; такая ситуация может, например, складываться с медикаментами, используемыми небольшим числом крупных центров медицинского обслуживания. Если интервалы упреждения производства оцениваются в месяцах и кварталах, а интервалы упреждения перераспределения — в днях и неделях, то продолжительное время упреждения демпфируется только одним государственным центральным резервным запасом. Обычно предположение о том, что ошибки прогноза продаж на ограниченной территории являются статистически независимыми, не оправдывается, но замечено, что стандартное отклонение ошибок прогноза общего спроса совокупности территорий заметно меньше, чем сумма стандартных отклонений ошибок отдельных территорий. Перераспределение позволяет с помощью единого государственного резервного запаса эффективно защитить от дефицита территорию всей страны.

Результаты применения рекомендованных решающих правил определения двух модификаций коэффициентов резервного запаса свидетельствуют о том, что имеется тенденция накопления большей части резервного запаса продукции класса А на периферийных складах, а класса С — на главном центральном складе.

Вопросам дальнейшей детальной разработки плана транспортировки и хранения материалов в рассматриваемой системе управления посвящена работа [13].

РЕАЛИЗАЦИЯ СИСТЕМ УПРАВЛЕНИЯ ЗАПАСАМИ

Современные теоретические основы управления запасами получили широкое признание как в США, так и в других странах мира. Тем не менее многие попытки реализовать хорошо разработанные системы не увенчались успехом. Одна из причин подобных неудач состояла в том, что основное внимание уделялось использованию ЭВМ для непосредственного решения задач управления запасами в конкретной системе, а не решению задач управления запасами вообще с применением ЭВМ, когда это целесообразно. Другая причина была связана с отсутствием поэтапности разработки и внедрения системы в действие. Как правило, осуществлялись разработка и внедрение всей системы в целом, и в результате тратилось ценное рабочее время персонала (от служащих до вице-президентов фирм) на тщетные усилия убедить его в реальных преимуществах радикально новых систем. В то же время поэтап-

ное внедрение хорошо разработанной системы дает возможность персоналу постепенно свыкнуться с производимыми изменениями, анализировать первые успехи внедрения, что, несомненно, облегчает дальнейший ход этого процесса, наблюдать реальные результаты уже на ранних этапах и, таким образом, убеждаться в действенности теории и реальности получения выгод для организации, пусть даже вначале незначительных и, наконец, принимать окончательные решения относительно использования некоторых возможностей системы после того, как будет накоплен достаточный опыт.

6.1. Переход от старой системы управления запасами к новой

Имитационные модели системы управления третьего уровня позволяют оценивать уровень обслуживания, количество заказов для сохранения определенного уровня запасов и уровень капиталовложений в запасы для установившегося состояния системы управления. Однако, прежде чем спрогнозированные результаты будут отражены в отчетах системы, должен пройти определенный период времени (переходный период), иногда довольно продолжительный. В связи с этим на каждом этапе реализации системы управления запасами необходима имитационная модель объекта управления, позволяющая спрогнозировать ход изменения запасов и уровня обслуживания месяц за месяцем, начиная от текущего состояния и кончая моментом внедрения новой системы. Существует временная характеристика, постоянная для каждой конкретной системы управления запасами, учитывающая интервал времени от подачи пополняющего заказа до поступления заказанной партии в наличные запасы и распределение интервалов времени, в течение которых текущий запас достигает новой точки заказа. Браун [4] установил математические зависимости, включающие указанные параметры решений относительно пополняющих заказов с учетом логарифмически-нормального распределения интенсивности потребления при условии частичной реализации системы как функции времени, прошедшего с момента начала нововведений. Однако на практике бывает проще использовать для этих целей грубое имитационное моделирование.

В сети взаимосвязанных предприятий органы управления, у которых выработалось определенное представление о временных параметрах, характеризующих некоторую сферу их деятельности, могут испытывать тревогу по поводу грандиозных отличий, характеризующих какую-либо другую сферу бизнеса. Надлежащим образом построенные имитационные модели системы третьего уровня управления запасами могут помочь им уточнить свои представления и исключить тем самым непредвиденные осложнения.

6.2. Избыточные запасы

Некоторые из имеющихся запасов неминуемо оказываются лишними если не при данной системе управления, то в условиях предполагаемых изменений этой системы и значений управляющих переменных. Поэтому важно уметь обнаруживать такие излишки запасов и предпринимать меры по устранению наиболее существенных диспропорций.

Максимально оправданный уровень наличного запаса целесообразно считать равным суммарному объему заказа на пополнение запасов всей продукции плюс удвоенный резервный запас (в предположении, что минимально возможный уровень запаса равен нулю, а средний уровень наличного запаса — резервному). Любой реальный запас, превышающий этот уровень, считается избыточным.

Общий располагаемый запас (наличный плюс заказанный) может достигать максимально допустимого уровня или уровня, равного точке заказа плюс размер заказанной партии. Располагаемый запас сверх этого уровня тоже является избыточным. Чтобы исключить случаи тривиальных излишков, обычно не обращают внимания на те избыточные запасы, которые при прогнозируемой интенсивности потребления могут быть реализованы за короткий период времени (например, за 3—4 мес), а также на избыточные запасы, незначительные по своему уровню (например, стоящие менее 1000 долл.).

Излишки, которые не могут быть реализованы в течение 3 мес или стоят больше 1000 долл., могут быть занесены в список, упорядоченный по убыванию их стоимости. В результате, если невозможно принять решение относительно использования сразу всех излишков, то первыми будут рассматриваться наиболее существенные из них. Если избыток вызван новым заказом, необходимо его аннулировать, изменить срок доставки или уменьшить размер заказа. Следует поступать именно так, поскольку даже уплата неустойки, связанная с аннулированием заказа, может оказаться ничтожной по сравнению с потерями, которые понесет фирма в результате возможного списания излишков в будущем. Если же избыток относится к наличным запасам, то требуется более тщательный поиск наилучшего способа его реализации.

Согласно мнению специалистов бухгалтерского учета, списание излишков необходимо производить в соответствующий момент. Торговые работники придерживаются иного мнения, так как опасаются, что передача излишков в другие инстанции может создавать конкуренцию выпускаемой ими доходной продукции. В этой связи они способны выдвигать остроумные предложения относительно организации специальных мероприятий по распродаже излишков.

В тех случаях, когда спрос увеличивается или сохраняется в течение длительного времени, экономический анализ затрат на хранение избыточных запасов в сравнении с их использованием показывает, что лучше продолжать хранить излишки, чем списывать их. Однако, если наличный запас превышает общие потребности, прогнозируемые на весь оставшийся период жизненного цикла продукции, то все излишки должны быть списаны независимо от возникающих при этом потерь.

Для запасных частей к автомобилям и других изделий машиностроения достаточно обоснованным является предположение [3], что спрос на каждую модель автомобиля затухает по экспоненциальному закону и пятилетний запас деталей учитывает все возможные потребности. Это означает, что суммарное остаточное потребление некоторой детали в период затухания спроса в любой момент времени равно пятикратному ее фактическому потреблению в истекшем году.

ЛИТЕРАТУРА

1. Arrow K. J., Karlin S., *Studies in the Mathematical Theory of Inventory and Production*, Stanford University Press, Stanford, California, 1958.
2. Brown R. G., *Statistical Forecasting for Inventory Control*, McGraw-Hill, N.Y., 1959.
3. Brown R. G., *Smoothing, Forecasting and Prediction*, Prentice-Hall, Englewood Cliffs, N.Y., 1963.
4. Brown R. G., *Decision Rules for Inventory Management*, Holt, Rinehart and Winston, N.Y., 1967.
5. Brown R. G., *Management Decision for Production Operations*, Dryden, 1971.
6. Brown R. G., Economic Order Quantities for Materials Subject to Engineering Changes, *Production and Inventory Management*, 12, No. 2, 89—91 (1971).
7. Brown R. G., Improved Customer Service at the HAL Division, *Proc. Am. Production and Inventory Control Conf.*, Toronto, 152—177 (October 1972).
8. deMatteis J. J., Mendoza A. O., Part-Period Algorithm, *IBM Syst. J.*, 7, No. 1, 30—46 (1968).
9. Dvoretzky A., Kiefer J., Wolfowitz J., The Inventory Problem, *Econometrica* XX, 187—222 and 450—466 (1952).
10. Gerson G., Brown R. G., Decision Rules for Equal Shortage Policies, *Nav. Res. Log. Quart.*, 17, No. 3, 351—358 (September 1970).
11. Harris F. W., *Operations and Cost*, A. W. Shaw, Chicago, 1915.
12. Magee J. F., *Production Planning and Inventory Control*, McGraw-Hill, N.Y., 1958.
13. Magee J. F., *Industrial Logistics*, McGraw-Hill, N.Y., 1968.
14. Mannell R. F., Early History of the Economic Lot Size, *Am. Production and Inventory Control Soc. Quart. Bull.*, 2, No. 2, 19—22 (April 1961).
15. Moran P. A. P., *The Theory of Storage*, Methuen, London, 1959.
16. Nocturne D. H., Economic Ordering Frequency for Several Items Jointly Replenished, *Management Sci.*, 19, No. 9, 1093—1096 (May 1973).
17. Raymond F. E., *Quantity and Economy in Manufacturing*, McGraw-Hill, N.Y., 1931.

18. Scarf H. E., Guilford D. M., Shelly M. W. (Eds.), Multi-Stage Inventory Models and Techniques, Standford University Press, Stanford, California, 1963.
19. Schlaiffer R., Introduction to Statistics for Business Decisions, McGraw-Hill, N.Y., 1961.
20. Silver E. A., Schaack J. P., A Procedure, Involving Simulation, for Selecting the Control Variables in an: S, c, s) Joint Ordering Strategy, *INFOR J.*, 10, 2 (May 1972).
21. Simpson K. F., In-Process Inventories, *Operations Res.*, 6, 863—873 (September 1958).
22. Wagner II., Whitin T. M., Dynamic Version of the Economic Lot Size Model, *Management Sci.*, 5, 89—96 (1958).
23. Whitin T. M., The Theory of Inventory Management, Princeton University Press, Princeton, N.J., 1953.

I-5. Вычислительные и информационные системы

А. Мертен¹⁾

ВВЕДЕНИЕ

С начала сороковых годов в деятельности промышленных предприятий, научных и учебных организаций существенную роль стали играть вычислительные машины. Аппаратурное и программное обеспечение таких машин, а также специалисты, участвующие в разработке и эксплуатации вычислительных систем, в значительной степени определяют потенциальные возможности любой системы управления. Поэтому органам управления и специалистам приходится решать огромное количество задач, связанных с наиболее рациональным размещением и использованием вычислительной техники. Для этого сначала осуществляется моделирование вычислительных систем, а затем их анализ с помощью методов исследования операций [14].

В настоящей главе дается классификация вычислительных и информационных систем, определяются задачи в области проектирования, эксплуатации, оценки функционирования вычислительных систем и результатов их использования для управления, рассматриваются направления дальнейшего применения методов исследования операций для решения конкретных задач.

Методы исследования операций применяются для анализа самых различных элементов и функциональных возможностей цифровых вычислительных машин. Для удобства описания таких исследований целесообразно классифицировать вычислительные системы по уровню их организации. Первый, самый нижний, уровень составляют *аппаратные средства*, в состав которых, как правило, входят процессор, оперативное и внешнее запоминающие устройства, периферийные устройства и средства управления. По вопросам использования методов исследования операций при разработке аппаратных средств опубликовано сравнительно мало работ.

Наиболее часто объектом системного анализа являются *вычислительные системы*, определяемые как сочетание аппаратного и программного обеспечения, такого, как операционная система и лингвистические процессоры (компиляторы, ассемблеры и т. д.). Операционная система представляет собой совокупность программ, предназначенных для управления большим числом разнообразных прикладных программ. Операционная система может использоваться и для выполнения других функций, таких, как выделение и учет ресурсов вычислительной машины во время ее работы. Большая часть приложенных методов исследования операций в вычислительной технике и теории информации приходится на область разработок отдельных элементов операционных систем и способов сопряжения операционных систем с различными аппаратными средствами.

Многие предприятия и организации не вносят или почти не вносят никаких изменений в аппаратуру и программное обеспечение приобретаемых ими вычислительных машин и обычно используют эти готовые средства для разработки прикладных программ и систем файлов, отвечающих их специфиче-

¹⁾ Alan G. Merten, Ph. D., Associate Professor of Management Science in the School of Business Administration at the University of Michigan in Ann Arbor, Michigan.

ским требованиям к обработке информации. Совокупность прикладных программ, аппаратных средств и системного программного обеспечения часто называется *автоматизированной системой обработки данных*. Такие системы являются системами более высокого уровня по сравнению с уже рассмотренными.

До недавнего времени методы исследования операций редко применялись для нахождения эффективных способов разработки прикладных программ, рассчитанных на использование всех возможностей вычислительной системы. В большинстве случаев эти методы применялись лишь для проектирования системы управления файлами пользователя.

Вычислительная машина является устройством, которое обеспечивает ввод, хранение и обработку массивов данных, используемых человеком. Во многих организациях автоматизированная система обработки данных является неотъемлемой частью *информационной системы*. Такая информационная система может быть отнесена к системам самого высокого уровня. Она включает много функций, выполнение которых не автоматизировано и осуществляется вручную. В их число входят функции, связанные с подготовкой исходных данных для ввода в вычислительную машину и использованием результатов ее работы специалистами как в области техники, так и в области управления. Методы исследования операций применялись для проектирования и оценки информационных систем, а также для определения наиболее эффективных способов использования вычислительных машин в этих системах.

При рассмотрении вопросов использования методов исследования операций в вычислительной технике и теории информации необходимо учитывать, с какой целью эти методы применяются в той или иной конкретной системе — для оптимизации проектирования, эксплуатации, оценки эффективности или организации функционирования системы. Это в равной степени относится к системам, представляющим все четыре рассмотренных класса.

Можно считать, что процесс создания вычислительной системы любого из рассматриваемых уровней включает следующие стадии: решение о необходимости создания системы, разработка логической части системы, разработка аппаратной части системы, монтаж и наладка системы, испытания системы, освоение и оценка функционирования системы, текущее обслуживание и усовершенствование системы.

Итак, прежде всего надо убедиться в необходимости создания вычислительной системы, затем установить, какие требования к ней предъявляются, и составить подробный перечень работ, которые должны быть выполнены. Построение системы ведется с использованием необходимых для этого компонент. Во время испытаний система проходит проверку в различных условиях — и идеальных, и производственных. Затем она в течение некоторого времени эксплуатируется, после чего оценивается эффективность ее функционирования и производится, в случае необходимости, устранение недостатков и совершенствование системы. На этой последней стадии процесса создания системы другие стадии могут многократно уточняться и повторяться.

До сих пор методы исследования операций использовались, как правило, на стадии разработки аппаратных средств. В некоторых случаях эти методы применялись для оценки и сравнения различных вариантов конструкции системы, а также для выбора способов оценки качества работы вычислительной системы, основываясь на которых можно выполнять необходимые усовершенствования.

На отдельных стадиях создания систем перечисленных выше классов использовались различные методы исследования операций. Трудность сравнения этих методов друг с другом заключается в том, что довольно часто невозможно отделить простое использование математических символов и элементарного количественного анализа от применения методов исследования операций при анализе вычислительной системы. В большинстве случаев при

таким анализе применяются следующие три группы методов исследования операций: методы массового обслуживания, теории оптимизации и имитационного моделирования с помощью ЭВМ. Модели массового обслуживания, относящиеся к различным компонентам операционных систем вычислительных машин, проанализированы в большом количестве работ. При анализе, как правило, вычислительная система рассматривается как один или большее число ресурсов, которые запрашиваются пользователями одного или разных типов. Подобные исследования были выполнены применительно к распределению ресурсов внешней (дисковой) памяти и к компонентам операционной системы.

В течение последних десяти лет при разработке вычислительных систем стали использоваться методы математического программирования. В большинстве таких работ формулируются задачи оптимизации, касающиеся совместного использования машинного времени и памяти вычислительной системы несколькими пользователями, и решение задач осуществляется с помощью методов математического программирования, например целочисленного. Некоторые из этих задач касаются распределения ресурсов машины при ее работе в многопрограммном режиме, организации оперативной памяти и внешней памяти, а также проектирования средств управления файлами для системы автоматизированной обработки данных [24].

В качестве вспомогательных средств при разработке различных систем создаются имитационные модели ЭВМ и машинные имитационные модели. Если с помощью имитационных моделей, как правило, осуществляется исследование различных элементов аппаратуры и операционных систем, то машинные имитационные модели помогают пользователю сделать выбор вычислительной системы в соответствии с заранее сформулированными требованиями.

В настоящей главе приводятся примеры применения методов исследования операций в области вычислительных и информационных систем. Примеры сгруппированы по классам систем; используемый метод описывается достаточно подробно, а соответствующая стадия создания системы, на которой этот метод применяется, становится понятной из описания.

1. ВЫЧИСЛИТЕЛЬНЫЕ СИСТЕМЫ

Разработчик программного обеспечения ЭВМ (операционных систем, компиляторов, систем управления данными и т. п.) хотя и ограничен аппаратурными возможностями машины, тем не менее стремится создать наиболее эффективное, экономически оправданное программное обеспечение, которое будет использоваться разработчиками прикладных программ и файлов конкретной вычислительной системы.

Приложение методов исследования операций к проектированию вычислительных систем обычно имеет целью рациональное распределение памяти и времени работы процессора между одновременно выполняемыми программами. Опубликованные в этом плане работы можно разделить на следующие три категории:

- разработка схем планирования заданий и выделения памяти применительно к отдельному запоминающему устройству;
- разработка схем размещения файлов пользователя в памяти машины при наличии нескольких типов памяти;
- разработка схем планирования заданий в многопрограммном режиме разделения времени.

Фрэнк [12] разработал модели внешних запоминающих устройств (магнитных дисков) с подвижными головками. В качестве основных характеристик накопителя на дисках он использовал ожидаемое время поиска головкой дорожки записи и ожидаемую задержку считывания нужной записи после выхода головки на дорожку. При моделировании рассматривались следующие вопросы, связанные с работой запоминающих устройств: зависимость времени поиска необходимой дорожки записи от скоростных характеристик механизма; влияние вероятностного распределения запросов информации, записанной на дорожках; особенности работы устройства в случае переполнения дорожки записями; способы установления дисциплины очередей; возможность уменьшения числа оборотов, необходимых для выполнения ряда обращений к диску, при помощи буферного считывания и использования устройства с многими головками. Разработанные аналитические стохастические модели позволили определить характеристики переходных процессов в запоминающих устройствах рассматриваемого типа.

На основе этих моделей Фрэнку удалось получить ряд важных результатов. Так, оказалось, что среднее время поиска записи может быть уменьшено, если часто используемые записи располагать в центральной части диска. Значительное уменьшение времени поиска записи достигается также за счет установления определенной дисциплины очередей заявок вместо последовательного выполнения заявок в порядке их поступления. Если объем записи превышает емкость дорожки, то для уменьшения общего времени доступа могут быть использованы буферные запоминающие устройства, позволяющие записывать последовательно несколько дорожек (в пределах одного цилиндра). Такой способ исключает задержки, обусловленные ожиданием выхода головки на маркер начала дорожки и неизбежные в тех случаях, когда на каждой дорожке запись осуществляется независимо от других. Некоторые новые результаты были получены и для случая применения нескольких независимых дисковых головок.

Симен и др. [33] разработали модель массового обслуживания для анализа различных способов хранения файлов прямого доступа во внешних запоминающих устройствах. Авторы вывели формулу для вычисления времени реакции системы и коэффициентов использования модулей файлов и каналов. Кроме того, для оценки эффективности использования системы индексации в больших последовательных файлах они рассмотрели различные модификации основной модели.

Авторы работы [1] применили теорию массового обслуживания для анализа функционирования внешнего запоминающего устройства в режиме реального времени. Изучалась зависимость между производительностью обслуживающего устройства, работающего в стохастическом режиме, и временем его реакции на каждую

заявку, требующую обслуживания. При этом использовалось дисковое запоминающее устройство типа IBM 2314. Результаты анализа были представлены в виде серии графиков, демонстрирующих зависимость времени реакции системы файлов от производительности для различных распределений длины записи и заявок пользователей. Согласно полученным данным, работа любого накопителя на дисках, у которого характеристика времени поиска нужной информации аппроксимируется кусочно-линейной непрерывной функцией, может быть проанализирована предложенными методами.

Лауэр [20] разработал аналитическую модель для сравнения ЭВМ типа IBM 360/67, в которой для организации постраничной виртуальной памяти используются магнитные барабаны, с такой же ЭВМ, но имеющей для этих целей запоминающее устройство большой емкости на магнитных сердечниках. В результате анализа было установлено, что магнитные барабаны обеспечивают необходимую интенсивность постраничного обмена, однако при этом резко возрастает стоимость системы.

Деннинг [10] исследовал зависимость производительности накопителей на дисках или барабанах от дисциплины очередей заявок и провел сравнение производительности устройств с подвижными головками при трех разных дисциплинах обслуживания. В качестве базовой дисциплины, с которой сопоставлялись две другие, использовалась дисциплина последовательного выполнения заявок в порядке их поступления. С целью уменьшения пути, проходимого головкой, и ожидаемого времени поиска информации были проанализированы дисциплины, при которых каждая ожидающая заявка проверяется с точки зрения целесообразности ее обслуживания в следующем цикле. Одна из этих дисциплин состояла в том, что очередное обслуживание проходит заявка, для выполнения которой требуется перемещение головки на кратчайшее расстояние (время поиска минимально). Другая дисциплина представляла собой модификацию первой. Деннинг показал, что эти методы организации обслуживания позволяют уменьшить ожидаемое время поиска информации и, следовательно, ожидаемую длительность обслуживания. Однако дисперсия времени окончания обслуживания часто бывает при этом очень большой из-за длительных задержек, возникающих при выполнении заявок на ту информацию, которая хранится во внутренних и внешних зонах запоминающего устройства. Для моделирования перемещения головки Деннинг использовал ряд значений непрерывной случайной величины, математическое ожидание которой являлось критерием при сравнении эффективности различных дисциплин обслуживания заявок. Авторы работ [25, 42] развили и модифицировали подход Деннинга, предложив аналитические и имитационные модели для сравнения различных дисциплин обслуживания заявок по сред-

нему значению и дисперсии рабочих характеристик систем обращения к файлам, хранящимся в дисковой памяти.

Кофман [8] разработал математические модели дисциплин обслуживания заявок для запоминающего устройства на магнитных барабанах. Предполагалось, что заявки поступают последовательно и произвольным образом. На основе анализа этих моделей были предложены: величина, с помощью которой можно характеризовать степень использования барабана; производящая функция для вычисления распределения вероятностей длины очереди в установившемся состоянии; способ вычисления средней длины очереди и среднего времени ожидания.

Количественные методы оценки разных способов распределения оперативной памяти между заданиями, выполняемыми вычислительной машиной, применялись различными авторами. Келли [17] разработал модели размещения подпрограмм в свободной оперативной памяти, сформулировав задачи размещения на основе требований к объему памяти, времени выполнения операций и функциональным характеристикам, связанным с различными способами размещения. В одной из моделей минимизировался объем памяти при фиксированных значениях времени выполнения операций, в другой модели — время при фиксированном объеме памяти. Предложенные модели являются моделями целочисленного и динамического программирования. Хотя автор отмечает, что в рамках этих моделей переменные должны принимать значения либо 0, либо 1, тем не менее он не утверждает, что рассматриваемые задачи относятся к задачам дискретного программирования и что они не могут быть решены обычными классическими методами линейного программирования.

Кнут [19] сообщил о результатах имитационного моделирования различных алгоритмов перераспределения оперативной памяти. Им было показано, что общепринятая схема выделения участков памяти для обслуживания заявок хуже предложенной им схемы, которая более удобна для реализации. Другим результатом моделирования было установление связи между полным объемом имеющейся памяти и объемом памяти, необходимым для удовлетворения отдельной заявки.

Пишкертон [30] разработал стохастические модели для исследования проблемы увеличения производительности вычислительной системы за счет улучшения распределения оперативной памяти.

В названных выше моделях решалась задача распределения памяти применительно к запоминающим устройствам одного уровня. В вычислительной системе третьего поколения проектировщику доступны различные уровни памяти. Известен ряд результатов по применению исследования операций для решения задач распределения памяти и управления программами и файлами в вычислительных системах с несколькими видами памяти.

Рамамурти и Чанди [32] разработали модель для выбора типа запоминающего устройства и его емкости, при которых среднее время обращения к блоку информации минимально при заданных ограничениях на стоимость устройства. Программа разделилась ими на одинаковые по размеру блоки, и для каждого из них определялась относительная частота использования. Были разработаны также аналитические модели и модели математического программирования для привязки программы и данных (блоков информации) к отдельным компонентам иерархической структуры памяти. Авторы рассмотрели также случай, когда память состоит из неделимых модулей. При этом задача организации иерархической структуры памяти была сформулирована как задача целочисленного программирования, для решения которой авторы предложили использовать алгоритм метода ветвей и границ.

Хеллерман и Смит [15] предложили аналитическую модель для определения рабочих характеристик вычислительной системы при различных конфигурациях каналов и памяти. Авторы ставили перед собой цель подробно исследовать возможные пределы пропускной способности системы и получить уравнения, пригодные для проверки результатов имитационного моделирования. В модели было принято, что ресурсы вычислительной системы включают один центральный процессор, главную память, вспомогательную память, состоящую из одного или более устройств, и каналы управления информационным обменом между главной и вспомогательной памятью.

Энекер и Ванг [3] исследовали проблему иерархической памяти, анализируя потоки данных между различными уровнями памяти. Ими был предложен метод оценки быстродействия системы обработки данных и определения характеристик относительного использования ее запоминающих устройств при различных конфигурациях системы и многократных загрузках программ. Статистика, касающаяся потоков данных в иерархической памяти, была получена путем анализа множества последовательностей фиксировавшихся адресов выполняемых программ. Для анализа работы процессора и памяти было предложено пользоваться имитационными моделями в сочетании с методами теории массового обслуживания. Модели были использованы для анализа двух машинных программ: исходной программы, написанной на Фортране, и программы в объектных кодах, полученной в результате трансляции исходной программы.

Гексей и др. [13] предложили количественные методы, позволяющие уменьшить трудоемкость использования имитационных моделей для оценки иерархических структур памяти, и разработали способы выбора рациональной частоты обращения к каждому уровню памяти. Эти способы применялись для широкого класса алгоритмов перераспределения памяти, используемых при реше-

нии вопроса о том, какую программу или какой файл данных следует перевести на более дешевое запоминающее устройство с меньшим быстродействием, чтобы освободить место для другой программы или другого файла.

Различные операционные системы третьего поколения основываются на концепции виртуальной памяти. В этом случае программист может писать машинные программы без учета требований к объему памяти, необходимой для реализации этих программ. Другими словами, в то время как объем имеющейся физической памяти ограничен, объем виртуальной памяти, доступной программисту, является неограниченным. Операционная система управляет перемещением блоков программы пользователя в главную память и составляет схему распределения виртуальной памяти по реальным запоминающим устройствам, выделенным для реализации программы.

Депинг в работе [11] изложил принципы функционирования вычислительных систем с виртуальной памятью и показал, как в этом случае надо применять разные аналитические методы и методы моделирования для сопоставления различных схем распределения памяти. Он предложил ряд аналитических выражений для выбора размера страниц, оптимального способа перераспределения страниц памяти и способа организации информационного обмена между основной и вспомогательной памятью, описал механизмы реализации виртуальной памяти и правила использования этих механизмов. Основными механизмами явились сегментация, при которой все пространство адресов разделяется на сплошные сегменты различной величины, и постраничное разбиение памяти, при котором все пространство адресов организуется в виде сплошных страниц фиксированных размеров. Автор предложил аналитические модели для сравнения и сопоставления различных способов организации и реализации виртуальной памяти.

Ахо и др. [2], решая задачу оптимального перераспределения страниц памяти в вычислительных системах с виртуальной памятью, пытались найти способ организации постраничного разбиения памяти, при котором суммарное время ожидания страницы при выполнении программы минимально. Задача была поставлена как задача динамического программирования.

Маккипни [23] дал обзор работ, в которых рассматривались модели массового обслуживания применительно к системам с разделением времени. В обзоре приведены параметры, с помощью которых можно характеризовать различные аналитические модели. Особое внимание уделено моделям, использованным для анализа различных классических дисциплин обслуживания (например, модели дисциплины с возвращением в очередь), и многоуровневым моделям массового обслуживания с приоритетами или без

приоритетов. Статья содержит прекрасную аннотированную библиографию.

Мур [27] и Бузен [5] применили сетевые модели массового обслуживания для анализа работы систем в многопрограммном режиме с разделением времени. Мур представил систему с разделением времени как совокупность независимых ресурсов. На основании этого допущения он интерпретировал процессы образования очередей к ресурсам как непрерывный марковский процесс. Модель была проверена на операционных системах, реализующих режим разделения времени, и использована для сравнения двух схем распределения памяти, применяемых в системах с виртуальной памятью (с постраничной организацией и с сегментированием).

Бузен использовал сетевые модели массового обслуживания для разработки операционных систем с оптимальной емкостью буферных накопителей в устройствах последовательного действия. В дальнейшем некоторые из моделей применялись для оптимального распределения заявок на обслуживание по множеству функционально эквивалентных периферийных процессоров, а также для выбора оптимального многопрограммного режима в системах с постраничным разбиением памяти.

2. СИСТЕМЫ ОБРАБОТКИ ДАННЫХ

Промышленные предприятия, а также органы управления, самостоятельно не занимающиеся разработкой вычислительных систем, приглашают программистов и специалистов по системному анализу для создания *систем обработки данных* (точнее, для разработки прикладных программ и файлов пользователей) на основе серийных ЭВМ и системного программного обеспечения, поставляемых изготовителями машин. Примерами, характеризующими существенную часть таких систем обработки данных, являются программы и файлы бухгалтерского учета и управления запасами, а также программы и файлы систем учета, действующих в учебных заведениях.

При проектировании программ и файлов пользователей методы исследования операций применялись с учетом ограничений на ресурсы аппаратных средств и операционной системы. При разработке программ применялось также имитационное моделирование [16, 39].

Имитационные модели представляют собой пакеты специальных программ, которые имитируют различную нагрузку (программы и файлы) упомянутых выше вычислительных систем (аппаратных средств, операционной системы и, возможно, системы управления файлами). Гипотетическая рабочая нагрузка системы задается с помощью совокупности документов установленной формы или на языке спецификаций. Характеристики файлов включают

такую информацию, как коэффициент блочности, количество записей, длина записей, плотность записей. Каждая отдельная программа, работающая в пакетном режиме или режиме реального времени, описывается в форме, которая не связана с техническими характеристиками какой-либо конкретной машины. Описание программ включает указание требуемого вида обработки и ожидаемой интенсивности использования формируемых файлов. Технические характеристики аппаратных и программных средств ряда машин организуются в виде библиотеки, которую можно использовать в рамках любой имитационной модели. Характеристики аппаратных средств содержат описание центрального процессора, внешних устройств и коммуникационного оборудования. Характеристики программных средств включают описание операционных систем, процессоров, проблемно-ориентированных языков, программ сортировки данных и генераторов отчетов. Эти данные вводятся в машинную модель, моделируется выбранная нагрузка и конфигурация, и затем печатаются выходные данные, которые содержат краткие сведения об интенсивности использования файлов, отдельных устройств и времени реакции системы.

Тайс [43], Дей [9] и Чу [7] рассмотрели задачи организации и размещения файлов как задачи математического программирования. Для выбора файла, к которому следует обращаться первым при поступлении группы запросов, Дей применил целочисленное программирование. Задача сводилась к минимизации общего времени поиска нужной информации на магнитной ленте при условии, что каждая заявка должна быть удовлетворена. Дей принял допущение, согласно которому в любой данный момент времени отыскивается только один файл и он вызывается целиком, если даже для удовлетворения данного запроса достаточно было бы найти только некоторую часть файла. Он предложил также модификацию указанной модели, позволяющую обращаться к двум файлам данных одновременно.

Чу [7] разработал модель размещения группы файлов, к которым в вычислительной системе совместно обращается несколько ЭВМ. Модель учитывает затраты на хранение и передачу данных, размеры и интенсивность использования различных файлов, максимально допустимое время доступа к файлам каждой вычислительной машины и их объем памяти. Критерием оптимального размещения информации являются минимальные общие затраты на хранение и передачу данных. Рассматриваемая задача была сформулирована как задача целочисленного программирования с булевыми переменными и нелинейными ограничениями, которые могут быть сведены к линейным.

Тайс [43] исследовал проблему оптимального блокирования записей в группе файлов с целью минимизации общего времени ввода-вывода информации для серии машинных расчетов. Он

принял, что файлы размещаются на магнитной ленте и что обработка данных происходит без буферизации операций ввода и вывода. Он допустил также, что в серии машинных расчетов число записей данных в каждом входном и выходном файле задано так же, как и объем внешней памяти, доступной входным и выходным данным, находящимся во внутренней памяти машины. Дополнительное ограничение состояло в том, что число символов в записи не может быть изменено и ни одна запись не может быть расплечена. Тайс поставил эту задачу как задачу нелинейного программирования с линейными ограничениями и нелинейной целевой функцией. Ослабив допущения, принятые в основной модели, он получил возможность моделирования буферных накопителей, периферийных ЭВМ, лентопротяжных механизмов различных типов и т. п. Для решения целочисленной задачи использовались общие алгоритмы нелинейного программирования с отбрасыванием нецелочисленных решений и оставлением лишь целочисленных коэффициентов блочности. С целью минимизировать время считывания и записи для каждого прогона Тайс исследовал разные варианты назначения файлов на ленточные запоминающие устройства различных типов. После определения коэффициентов блочности он предложил модель для выбора ленточного запоминающего устройства, обеспечивающего максимальную скорость перемещения файлов. В результате первоначальная задача была сведена к транспортной задаче.

Имитационная модель файла FOREM, написанная на Фортране, была предложена Сенко и др. [34]. В качестве входных данных в ней используются таблицы, описывающие логические файлы, записи, зоны, параметры устройств и способы организации поиска информации. Заявки на обслуживание генерируются на основании спецификации запросов, а результаты ответов на запросы определяются описаниями, даваемыми в форме таблиц. В модели имеются имитационные блоки, отображающие специфические условия и функционизирующие совместно с блоками имитации запросов. Время обработки оценивается при помощи соответствующих уравнений, введенных в блоки модели. Обширные материалы, касающиеся моделирования индексно-последовательного метода доступа, изложены Ламом и др. [21]. Мартин [22] рассмотрел использование языка GPSS для моделирования способа организации доступа к файлам.

Ниссен [28], Бихаймер и др. [4] провели сравнительную оценку достоинств индексно-последовательной и прямой организации файла как в случае пакетной обработки, так и при обработке в реальном масштабе времени. В работе [4] оценивался также эффективный метод актуализации файла прямого доступа в режиме пакетной обработки. Бихаймер и др. разработали имитационную модель для генерации таблиц, которые могут использоваться для выбора

рационального способа организации файлов и режима обработки на основе спецификации требований пользователя и описаний файлов.

Карденасом [6] предложена имитационная модель базы данных. Входом ее служат структурные модули файлов, выходом — приблизительная общая стоимость хранения данных и среднее время доступа при заданной конкретной системе сбора данных, заданных характеристиках запросов и требованиях к устройствам. Моделирование различных структур файлов осуществлялось при различных совокупностях данных без учета действий по актуализации файлов.

Необходимость использования обобщенных моделей файлов для оценки и выбора их структур была отмечена Северансом в работе [36]. До этой работы в имитационных моделях отображали лишь структуры файлов и способы поиска, причем и те и другие характеризовались небольшим числом параметров. В каждом конкретном случае такая модель настраивалась на генерацию ряда структур файлов путем изменения параметров обобщенной модели. Модель, созданная Северансом, не имитировала операций по постановке запроса в очередь или изменению данных, организованных в файлы. Вместо этого имитировались действия разработчика при выборе наилучшей структуры файлов. На основе использования моделей принятых структур файлов машинная программа Северанса четко оценивает множество возможных структур файлов с точки зрения требований пользователя, вводимых в машину. По критерию, выбранному пользователем, который определяет относительные веса основной и вспомогательной памяти и задает время обработки, программа оценивает те структуры файлов, которые могут представлять интерес для разработчика, поскольку некоторые структуры файлов могут вообще не рассматриваться как не удовлетворяющие основным требованиям. В итоге программа воспроизводит на экране дисплея выбранные структуры файлов.

В целях моделирования и анализа различных структур файлов многие авторы разрабатывали точные модели данных, подлежащих хранению, работы пользователя с этими данными, а также аппаратных и программных средств, используемых для обеспечения хранения и текущего ведения информационных массивов. Для демонстрации компонентов большинства таких моделей рассмотрим более детально модель, предложенную Северансом и Мертенсом [35].

Эта модель включает точную спецификацию данных, действий пользователя и параметров памяти. Набор данных описывается путем указания числа записей, количества полей данных в каждой записи и конкретных полей, которые могут использоваться для однозначной идентификации записи. Например, набор данных о структурах какой-либо организации может содержать 1000 логи-

ческих записей, состоящих каждая из 10 пар «признак — значение», которые составляют описание характеристик одного сотрудника. В каждой логической записи может присутствовать номер пропуска, принадлежащий только этому сотруднику.

Действия пользователя описываются с помощью сведений об исключении, дополнении и обновлении записей, а также о требуемых выборках из совокупностей данных (примером обновления может служить ежемесячное изменение пяти телефонных номеров). Описание выборок, указанных пользователем, предусматривает указание ключа критерия упорядочения и частоты обращения (например, может потребоваться, чтобы система ежемесячно находила фамилии и номера пропусков всех служащих отделения А, располагая их последовательно в порядке возрастания номеров).

Запоминающее устройство описывается посредством задания параметров, представляющих собой характеристики аппаратных средств (например, скорость вращения диска) и программных средств (например, способы установления логической связи двух элементов хранимых данных). Различные способы организации файлов оцениваются на основе анализа линейной комбинации следующих составляющих затрат: объема памяти, требуемого для размещения файлов; времени, необходимого для осуществления выборки из файлов; времени, необходимого для текущего обслуживания (ведения) системы файлов.

Кеннеди [18] предложил модель для построения структур файлов в виде деревьев с раздвоением ветвей, учитывающую частоту обращения к данным и обеспечивающую минимизацию взвешенной длины пути в дереве. Для случая, когда путям приписываются равные весовые множители, приближенное решение было дано Сассенгатом [38], а точное решение — Стеффелом [37], поставившим задачу как задачу целочисленного программирования. Было показано, что эта задача соответствует задаче минимизации двоичного дерева и задаче о станках, известной в теории расписаний. Для получения решения был использован метод ветвей и границ.

Рамамурти и Блевинс [31] разработали эвристический алгоритм для изучения проблемы размещения данных в памяти с последовательным доступом в соответствии с частотой обращения к ним. Такой подход применим для широкого класса целевых функций, отражающих время поиска данных, а именно монотонно-возрастающих и кусочно-линейных. В результате использования указанного алгоритма удастся определить границы расположения данных, обеспечивающие достижение локального оптимума, получить доказательство существования решения и достижимости его за конечное число шагов, сравнить различные модификации алгоритма и осуществить имитационное моделирование. Методы, использованные в данном случае, основаны на применении аппарата ориентированных графов.

Но [44] разработал систему проектирования, обеспечивающую выбор наиболее рационального варианта организации файлов. Согласно этой системе, необходимые для проектирования параметры определяются на основе описаний данных, действий пользователя и характеристик ЭВМ. Выходом системы проектирования является целесообразная структура файлов; детали этой структуры могут уточняться посредством имитационного моделирования или иными способами. Работа модели основана на том, что при любом способе выборки осуществляются постепенное сужение поиска и иерархическая декомпозиция пространства поиска. Траектории доступа к информации базы данных для большинства структур файлов моделировались с помощью древовидной схемы выборки. Эта схема представляется двумя группами переменных: одна из них характеризует конфигурацию дерева, а другая — его ветви. Модель древовидной схемы выборки выполняет преобразование пространства структур файлов в n -мерное пространство переменных. Проектные требования учитываются в функции затрат, определяемой на этом множестве переменных. С помощью алгоритма нелинейного программирования определяется множество значений переменных, обеспечивающее минимум функции затрат, что в свою очередь приводит к структуре файлов минимальной стоимости.

3. ИНФОРМАЦИОННЫЕ СИСТЕМЫ

Предприятия и организации используют имеющуюся у них информацию в связи с различными сторонами их деятельности. Например, отчеты о продаже товаров и модели сбыта используются для установления маршрутов передвижения коммивояжеров и определения размеров комиссионных вознаграждений за их работу. Машинные программы и файлы, обеспечивающие хранение и обработку данных, содействуют планированию наиболее эффективной деятельности предприятий или организаций. Однако для создания потока информации и управления им требуется применение ручных операций подготовки данных и других неавтоматизированных действий, например преобразование первичных документов для ввода содержащихся в них данных в машину и доставка результатов машинной обработки ответственному руководителю.

Ньюмэйкер [29], а также Тичроус и Сейзли [41] разработали способы применения методов исследования операций для автоматизации проектирования информационных систем. Основная концепция этих исследований состоит в том, что вся система может генерироваться вычислительной машиной по предварительному описанию требований к обработке информации в будущей системе. Методология проектирования включает использование ряда частных моделей, на основе которых осуществляется поиск решения

методами математического программирования и теории графов с применением эвристических алгоритмов. Используемый алгоритм представляет собой многоуровневую модель принятия решений, в которой управляющие переменные одного уровня являются ограничениями для следующего уровня.

Для формулирования требований к информационной системе, не зависящих от способа обработки данных, предложен формальный язык. С помощью модели выбирается комплект аппаратных средств и генерируются спецификации программ и файлов, удовлетворяющих требуемым временным характеристикам и ограничениям на объем оперативной памяти и на емкость внешних запоминающих устройств; при этом минимизируется стоимость аппаратных средств. Применение модели ограничивается проектированием однопрограммных систем пакетной обработки с последовательной организацией внешней памяти, линейными структурами данных и выбором однопроцессорных ЭВМ. Модели математического программирования используются для оптимизации показателей блочности файлов, для определения числа вспомогательных запоминающих устройств и их типов, для распределения файлов по запоминающим устройствам и для формирования текущего графика последовательности запуска программных модулей.

Тичуров и др. [40], Мертен и Тичуров [26] развили концепцию использования ЭВМ как вспомогательного средства проектирования автоматизированных информационных систем. Суть этой концепции состоит в том, что в процессе создания системы описание требований пользователя должно быть как можно раньше записано в машиночитаемой форме.

Описание требований к системе обработки информации подразумевает определение и задание многих элементов и связей между ними. Во-первых, должны быть установлены элементы системы, которые участвуют в обработке, хранении или использовании информации. При этом должны быть определены физические устройства для передачи или хранения данных. Затем следует определить элементы данных и способы их обработки. Динамика поведения системы описывается посредством формулирования условий или ситуаций, которые вызывают определенные события, и действий, которые за этими событиями следуют. Наконец, должны быть описаны масштабы проектируемой системы путем указания соответствующих параметров.

Модели, основанные на формализованном описании требований пользователя, создаются для того, чтобы помочь им в разработке машинных программ, файлов и баз данных. В частности, алгоритмы, предназначенные для группировки и представления требований пользователя в виде, соответствующем входным спецификациям моделей проектирования файлов, обсуждались в предыдущем разделе.

ЗАКЛЮЧЕНИЕ

В этой главе приведены классификация и обзор приложений методов исследования операций в вычислительной технике и информатике. Основные достижения в этой области касаются приложений количественных методов к планированию работы со вспомогательной памятью большой емкости. Публикации последних лет содержат сведения об определенных успехах, достигнутых в применении методов исследования операций к задачам, непосредственно связанным с учетом требований пользователей, т. е. с проектированием информационно-вычислительных систем.

Успехам в этой области препятствуют разные факторы. Во-первых, большинство проектировщиков и специалистов в области вычислительных и информационных систем не в состоянии четко формулировать целевые функции и/или ограничения, которые необходимо учитывать при решении поставленных перед ними задач. Кроме того, иногда необходимо проводить оптимизацию системы сразу по нескольким критериям.

Во-вторых, многие задачи, связанные с проектированием вычислительных и информационных систем, являются задачами расписания, упорядочения или распределения. В большинстве случаев эти задачи интерпретируются как задачи дискретного программирования, для которых отсутствуют эффективные вычислительные методы. Наконец, как правило, трудно согласовать уже существующие методы реализации задач распределения с моделями, которые появляются на более поздних этапах создания системы. Это — часто встречающаяся проблема, по наиболее резко она проявляется именно в рассматриваемой области применения методов исследования операций. Основная трудность здесь заключается в широком диапазоне значений временных интервалов, с которыми сталкиваются проектировщики. Например, в системах с разделением времени, эффективность которых зависит от действий пользователя за терминалом и особенностей функционирования конкретного процессора, временные характеристики пользователей часто измеряются в микрутах или секундах, в то время как характеристики работы процессора измеряются в микро- или даже в наносекундах.

Основные направления работ по использованию методов исследования операций при проектировании вычислительных систем должны быть сосредоточены на решении рассмотренных проблем. Предстоит также интенсифицировать исследования в сфере проектирования, связанные с требованиями возможных пользователей ЭВМ. Проектирование информационно-вычислительных систем также остается еще нерешенной проблемой.

ЛИТЕРАТУРА

1. Abate J., Dubner H., Weinberg S., Queuing Analysis of the IBM 2314 Storage Facility, *J. ACM*, **15**, No. 4, 577—589 (October 1968).
2. Aho A. V., Denning P. J., Ullman J. D., Principles of Optimal Page Replacement, *J. ACM*, **18**, No. 1, 80—93 (January 1971).
3. Anacher W., Wang C. P., Performance Evaluation of Computing Systems with Memory Hierarchies, *IEEE Trans. Electronic Comput.*, **EC-16**, No. 6, 764—772 (December 1967).
4. Behymer J., Merten A., Ogilvie R., Analysis of Indexed Sequential and Direct Access File Organizations, ACM-SIGFIDFT Workshop on Data Description, Access and Control (May 1—3, 1974).
5. Buzen J. P., Computational Algorithm for Closed Queuing Networks with Exponential Servers, *Communications ACM*, **16**, No. 9, 527—531 (September 1973).
6. Cardenas A. F., Evaluation and Selection of File Organization — A Model and System, *Communications ACM*, **16**, No. 9, 540—548 (1973).

7. Chu W. W., Optimal File Allocation in a Multicomputer Information System, Proc. IFIP Conf., North-Holland, Amsterdam, F30—F85, 1968.
8. Coffman E., Analysis of a Drum/Output Queue Under Scheduled Operation in a Paged Computer System, *J. ACM*, 16, No. 1, 73—90 (January 1969).
9. Day R. H., On Optimal Extracting from a Multiple File Data Storage System: An Application of Integer Programming, *Operations Res.*, 13, No. 3, 482—492 (1965).
10. Denning P. J., Effects of Scheduling on File Memory Operations, Spring Joint Comput. Conf., 1967, pp. 9—21.
11. Denning P. J., Virtual Memory, *Computing Surveys*, 2, No. 3, 152—189 (September 1970).
12. Frank H., Analysis and Optimization of Disc Storage Devices for Time-Sharing Systems, *J. ACM*, 16, No. 4, 602—620 (October 1969).
13. Gecsei J., Slutz D. R., Traiger I. L., Mattson R. L., Evaluation Techniques for Storage Hierarchies, *IBM Syst. J.*, 9, No. 2, 78—117 (1970).
14. Hanssmann F., Ed., Operational Research in the Design of Electronic Data Processing Systems, Proceedings of a NATO Conference, Munich, West Germany, The English Universities Press, Ltd., London, Crane, Russak and Co., Inc., N.Y., 1971.
15. Hellerman H., Smith H. J., Throughput Analysis of Some Idealized Input, Output and Complete Overlap Configurations, *Computing Surveys*, 2, No. 2, 111—118 (June 1970).
16. Ihrer F. C., Computer Performance Projected Through Simulation, *Comput. and Autom.*, 17, 22—27 (April 1967).
17. Kelley J. E., Jr., Techniques for Storage Allocation Algorithms, *Communications ACM*, 4, No. 10, 449—454 (October 1961).
18. Kennedy S. R., The Use of Access Frequencies in Data Base Organization, Ph. D. Dissertation, Cornell University, 1973.
19. Knuth D., The Art of Computer Programming. Vol. 1 Addison-Wesley, Reading Massachusetts, 1968, pp. 435—455.
20. Lauer H. C., Bulk Core in an IBM 360/67 Time Sharing Systems. Fall Joint Computer Conference, 1967.
21. Lum V. Y., Ling H., Senko M. E., Analysis of a Complex Data Management Access Method by Simulation Modeling, Proc. Fall Joint Computer Conference, 1970.
22. Martin J., Design of Real-Time Computer System, Prentice-Hall, Englewood Cliffs, N.Y., 1967.
23. McKinney J. M., A Survey of Analytical Time Sharing Models, *Computing Surveys* K, 2, 105—116 (June 1969).
24. McWhorter A., Jr., A Survey of the Application of Mathematical Programming to the Optimization of Information Processing Systems unpublished paper, 1971.
25. Morten A. G., Some Quantitative Techniques for File Organization. Ph. D. Thesis, Technical Report No. 15, University of Wisconsin Computing Center, 1970.
26. Merten A., Teichroew D., Impact of Problem Statement Languages in Software Evaluation, Fall Joint Computer Conference, 1972.
27. Moore Ch. G., III, Network Models for Large-Scale Time Sharing Systems, Technical Report 71—4, ISDOS Research Project (April 1971).
28. Nyssen G. M., Indexed Sequential versus Random, *IAG J.*, 4, 1, 29—37 (March 1971).
29. Nunamaker J., A Methodology for the Design and Optimization of Information Processing Systems, Spring Joint Computer Conference, AFIPS Press, Montvale, N.J., 1971.
30. Pinkerton T., Program Behavior and Control in Virtual Storage Computer Systems, Ph. D. Thesis, The University of Michigan, 1968.

31. Ramamoorthy C. V., Blevins P. R., Arranging Frequency Dependent Data on Sequential Memories, Spring Joint Computer Conference, 1971.
32. Ramamoorthy C. V., Chandy K. M., Optimization of Memory Hierarchies in Multiprogrammed Systems, *J. ACM*, 17, 3, 426—445 (July 1970).
33. Seaman P., Lind R., Wilson T., An Analysis of Auxiliary Storage Activity, *IBM Syst. J.*, 5, 3, 158—170 (1966).
34. Senko M. E., Lum V., Owens P., A File Organization Evaluation Model (FOREM), Proc. Intl. Federation of Info. Processing Congress 68, 1968.
35. Severance D., Merten A., Performance Evaluation of File Organizations through Modeling, Proc. ACM, 1972, pp. 1061—1072.
36. Severance D. G., Some Generalized Modeling Structures for Use in Design of File Organizations, Ph. D. Dissertation, The University of Michigan, 1972.
37. Stanfel L. E., Tree Structures for Optimal Searching, *J. of ACM*, 17, 3, 508—517 (July 1970).
38. Sussenguth E. H., Use of Tree Structures for Processing Files, *CACM*, 65 272--279 (May 1963).
39. Sutherland J. W., Tackle System Selection Systematically, Comput. Decisions, 3, 4, 14—19 (April 1971).
40. Teichroew D., Hershey E. A., Bastarache M., An Introduction to PSL, PSA, ISDOS Working Paper 86, University of Michigan, 1974.
41. Teichroew D., Sayani H. Automation of System Building, Datamation 25—30 (August 15, 1971).
42. Teorey T., Pinkerton T., A Comparative Analysis of Disk Scheduling Policies, *Communications ACM*, 15, 3, 177—184 (March 1972).
43. Thiess H. E., Mathematical Programming Techniques for Optimal Computer Use, *Proc. ACM*, 501—512 (1965).
44. Yao B., Evaluation and Optimization of File Organizations Through Analytic Models, Ph. D. Thesis, The University of Michigan, 1974.

I-6. Планировка и размещение объектов¹⁾

Р. Френсис²⁾, Дж. Уайт³⁾

ВВЕДЕНИЕ

Данная глава посвящена вопросам использования методов исследования операций для решения проблем планировки и размещения объектов. Первые попытки анализа подобных задач с целью нахождения приемлемого решения относятся к XVII столетию. Однако только при появлении теории исследования операций эти задачи стали предметом тщательного и всестороннего исследования с участием представителей самых различных профессий, в том числе архитекторов, экономистов, инженеров, географов, специалистов по управлению, исследованию операций и системному анализу.

Термин «объект» можно трактовать очень широко. Например, объектом может быть космический корабль, автобус, который развозит детей в школы, станок, больница, удаленный терминал ЭВМ, депо пожарной охраны, карета скорой помощи, очистное сооружение, книга в библиотеке, аэропорт, полицейский патрульный автомобиль, радиолокационная станция, программа для ЭВМ, которая должна быть записана на магнитную ленту, товарный склад, дисплей или служебные помещения в административном здании. Именно вследствие столь широкого толкования указанного термина работы по проблеме планировки и размещения объектов публикуются в различных специальных журналах.

Анализируя многочисленные исследования, выполненные начиная с 1960 г., нетрудно заметить, что формулировка почти каждой задачи планировки и размещения объектов содержит одни и те же основные показатели, которые могут быть использованы для классификации этих задач. К таким показателям относятся характеристики новых объектов и размещение существующих, взаимодействие новых и существующих объектов, пространство решений, а также мера расстояния между объектами (или метрика пространства перемещений) и критерии оценки возможных вариантов решения. Рассмотрим каждый из перечисленных показателей более подробно.

Основным показателем, характеризующим новые объекты, является их число. Кроме того, в зависимости от размеров каждый новый объект можно рассматривать либо как точку, либо как некий протяженный объект. В последнем случае управляющей переменной является форма объекта (или форма занимаемой им площади) и задача размещения сводится к задаче планировки. (В некоторых задачах размещения число новых объектов является управляющей переменной, а не параметром задачи.) Наконец, размещение каждого нового объекта может зависеть или не зависеть от размещения других новых объектов.

¹⁾ В основу данной главы положен материал, опубликованный в работе [35]. Перепечатка произведена с разрешения издательства Prentice-Hall, Englewood Cliffs, New Jersey.

²⁾ Richard L. Francis, Ph. D., Professor of Industrial and Systems Engineering at the University of Florida in Gainesville, Florida.

³⁾ John A. White, Ph. D., Professor of Industrial and Systems Engineering at the Georgia Institute of Technology in Atlanta, Georgia.

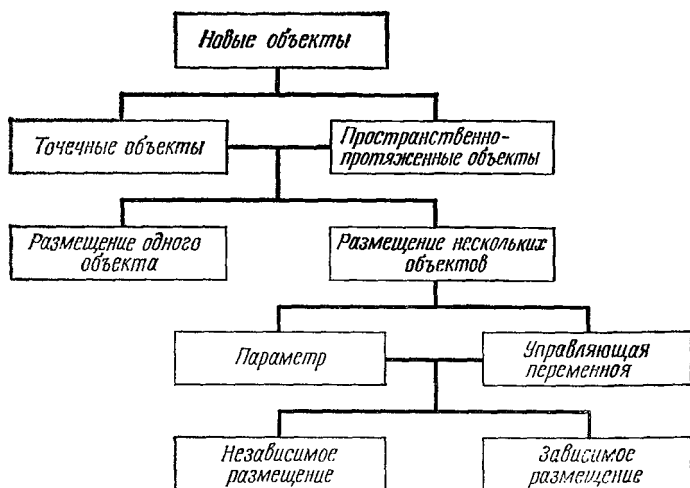
Что касается существующих объектов, то они также в зависимости от размеров могут рассматриваться либо как точка, либо как протяженный объект. Кроме того, размещение существующих объектов может быть статическим или динамическим, детерминированным или стохастическим. Если размещение существующих объектов является управляющей переменной, то мы имеем дело с задачей перераспределения, а если, кроме того, управляющей переменной является форма занимаемой ими площади, то и с задачей перепланировки.

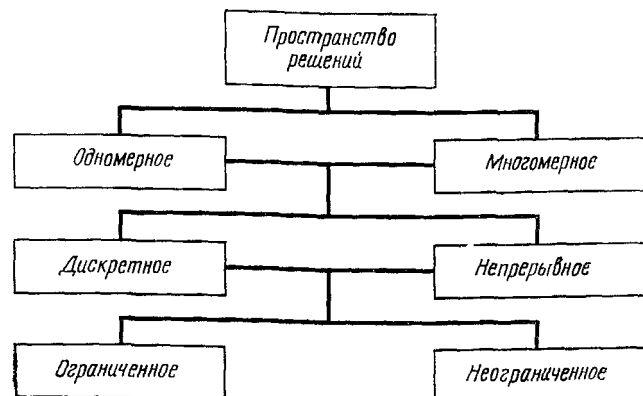
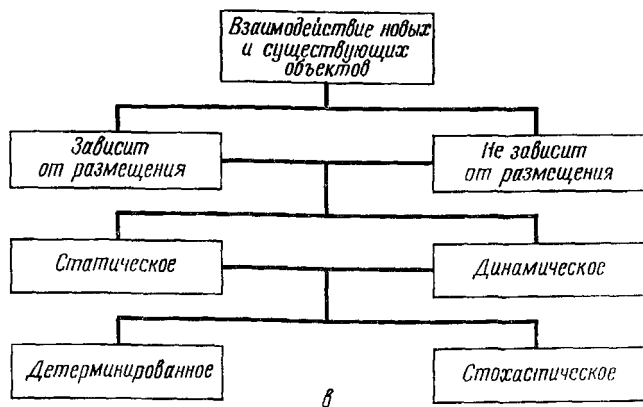
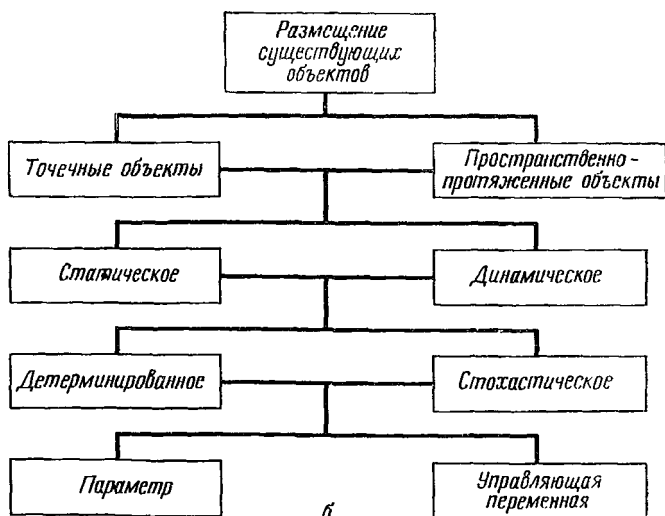
Взаимодействие новых и существующих объектов может выполнять функцию параметра задачи или управляющей переменной. В ряде случаев степень этого взаимодействия зависит от размещения новых объектов, причем характер взаимодействия может быть статическим или динамическим, детерминированным или стохастическим (рис. 1).

Пространство решений может быть одномерным, двух- или трехмерным (чаще всего мы имеем дело с двумя последними случаями). Кроме того, оно может быть дискретным или непрерывным. В первом случае для размещения новых объектов имеется конечное число мест, в то время как во втором случае, т. е. когда пространство предполагается непрерывным, существует бесконечное число мест для размещения новых объектов.

Мера расстояния (или метрика пространства перемещения) также учитывается при формулировке задач размещения. Иногда в качестве приближенной оценки фактических расстояний вычисляются евклидовы расстояния, в других случаях приходится измерять и хранить для последующих вычислений фактические расстояния.

Выбор критерия оценки возможных решений существенным образом зависит от того, с каким сектором мы имеем дело: частным или государственным. В первом случае оптимальным считается решение, приводящее к минимизации общих затрат или максимальных затрат для множества пар объектов; во втором случае — максимизация государственной выгоды. За исключением задач о покрытии, задачи планировки и размещения, как правило, сводятся к минимизации взвешенной суммы расстояний или минимизации максимального взвешенного расстояния между объектами [16, 24—26, 31, 32, 44, 75, 97].





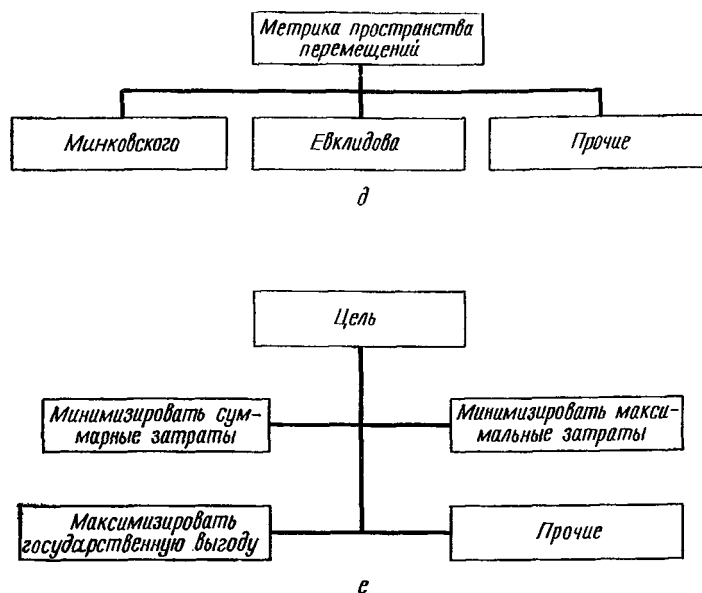


Рис. 1. Классификация задач размещения и планировки объектов.

В последние годы разработано большое число программ решения задач планировки заводов с помощью ЭВМ, в том числе ALDEP [84], CORELAP [63], диалоговая система CORELAP [74], CRAFT [2], PLANET [1] и др. [35].

Задача размещения оборудования на сети, которая впервые была рассмотрена в работе [46], анализируется в работах [17, 36, 39, 41, 42, 45, 49, 64, 72, 94].

1. ЗАДАЧИ РАЗМЕЩЕНИЯ С НЕПРЕРЫВНЫМ ПРОСТРАНСТВОМ РЕШЕНИЙ

Рассмотрим обобщенную задачу размещения, известную как задача Ферма (иногда ее называют задачей Стейнера, задачей Вебера или задачей Стейнера — Вебера). Математически эта задача формулируется следующим образом.

Пусть m существующих объектов размещены в различных точках $P_1, \dots, P_m$ плоскости, а n новых объектов — в точках $X_1, \dots, X_n$. Расстояние между точками расположения j -го нового и i -го существующего объектов обозначим как $d(X_j, P_i)$, расстояние между точками расположения j -го и k -го новых объектов — как $d(X_j, X_k)$. Обозначим годовые удельные затраты (т. е. затраты на единицу расстояния) на перевозки между j -м новым и i -м существующим объектами через w_{ji} , а аналогичные затраты на перевозки между j -м и k -м новыми объектами — через v_{jk} . Тогда

общие годовые транспортные затраты, связанные с размещением новых объектов в точках $X_1, \dots, X_n$, определяются формулой

$$f(X_1, \dots, X_n) = \sum_{1 \leq j < k \leq n} v_{jk} d(X_j, X_k) + \sum_{j=1}^n \sum_{i=1}^m w_{ji} d(X_j, P_i). \quad (1)$$

Многоэлементная задача размещения может быть сформулирована как задача выбора такого расположения новых объектов в точках $X_1^*, \dots, X_n^*$, при котором минимизируются общие годовые транспортные затраты.

1.1. Размещение объектов на плоскости с метрикой Минковского

Рассмотрим размещение объектов на плоскости, когда кратчайшее расстояние между объектами определяется по формуле

$$d(X_j, X_k) = |x_j - x_k| + |y_j - y_k|, \quad (2)$$

$$d(X_j, P_i) = |x_j - a_i| + |y_j - b_i|, \quad (3)$$

где $X_j = (x_j, y_j)$ и $P = (a_i, b_i)$.

Подставляя выражения (2) и (3) в формулу (1), получим

$$f(X_1, \dots, X_n) = f_1(x_1, \dots, x_n) + f_2(y_1, \dots, y_n), \quad (4)$$

где

$$f_1(x_1, \dots, x_n) = \sum_{1 \leq j < k \leq n} v_{jk} |x_j - x_k| + \sum_{j=1}^n \sum_{i=1}^m w_{ji} |x_j - a_i|, \quad (5)$$

$$f_2(y_1, \dots, y_n) = \sum_{1 \leq j < k \leq n} v_{jk} |y_j - y_k| + \sum_{j=1}^n \sum_{i=1}^m w_{ji} |y_j - b_i|. \quad (6)$$

В дальнейшем будем считать, что составляющие f_1 и f_2 общих затрат определяются с помощью формул (5) и (6). Используя значения f_1 и f_2 , можно определить общие затраты на перевозки в направлении координатных осей x и y .

Из формулы (4) следует, что

$$\min f(X_1, \dots, X_n) = \min f_1(x_1, \dots, x_n) + \min f_2(y_1, \dots, y_n).$$

Таким образом, оптимальные x -координаты размещения новых объектов могут быть определены независимо от y -координат. Кроме того, поскольку функции f_1 и f_2 имеют один и тот же вид, то любая процедура, разработанная для минимизации функции f_1 , может быть применена также и к функции f_2 при замене x_j на y_j и a_i на b_i .

Переход к минимизации целевых функций f_1 или f_2 , по существу, означает сведение рассматриваемой задачи к эквивалентной задаче линейного программирования, любое оптимальное решение которой будет давать оптимальную x -координату размещения нового объекта.

Рассмотрим постановку эквивалентной задачи линейного программирования в предположении, что выполняется соотношение $|a - b| = p + q$, где a, b, p и q — заданные числа, $a - b - p + q = 0$, $p \geq 0$, $q \geq 0$ и $pq = 0$. Отсюда следует, что минимизация функции f_1 эквивалентна минимизации целевой функции

$$\sum_{1 \leq j < k \leq n} v_{jk} (p_{jk} + q_{jk}) + \sum_{j=1}^n \sum_{i=1}^m w_{ji} (r_{ji} + s_{ji})$$

при ограничениях

$$\begin{aligned} x_j - x_k - p_{jk} - q_{jk} &= 0, \quad 1 \leq j < k \leq n, \\ x_j - r_{ji} - s_{ji} &= a_i, \quad i = 1, \dots, m, \quad j = 1, \dots, n, \\ p_{jk}, q_{jk} &\geq 0, \quad 1 \leq j < k \leq n, \\ r_{ji}, s_{ji} &\geq 0, \quad i = 1, \dots, m \text{ и } j = 1, \dots, n, \\ x_j &\text{ не ограничено,} \quad j = 1, \dots, n, \\ p_{jk} \cdot p_{jk} &= 0, \quad 1 \leq j < k \leq n, \\ r_{ji} \cdot s_{ji} &= 0, \quad i = 1, \dots, m \text{ и } j = 1, \dots, n. \end{aligned}$$

Заметим, что если исключить два последних ограничения, то сформулированная задача является задачей линейного программирования (выпеприведенную задачу будем называть задачей Р0, а эквивалентную ей задачу линейного программирования — задачей Р1). Поскольку задача Р1 имеет меньше ограничений, чем задача Р0, то минимальное значение целевой функции в первом случае будет по крайней мере таким же, как во втором, или меньше. Если минимальное допустимое решение задачи Р1 удовлетворяет всем ограничениям задачи Р0, то оно является также минимальным допустимым решением задачи Р0.

Согласно теории линейного программирования, при решении задачи Р1 некоторое базисное допустимое решение будет минимальным допустимым решением. Для любого базисного допустимого решения, если p_{jk} принадлежит базисному допустимому решению, q_{jk} не будет ему принадлежать, и наоборот; аналогично, если r_{ji} принадлежит базисному допустимому решению, s_{ji} не принадлежит ему, и наоборот. Поскольку переменные, не входящие в базисное допустимое решение, равны нулю, то при любом базисном допустимом решении будут удовлетворяться последние два ограничения задачи Р0. Таким образом, задачи Р0 и Р1 являются эквивалентными задачами оптимизации, и для решения задачи размещения на плоскости с принятой метрикой можно применять линейное программирование.

Если число новых и существующих объектов велико, решение задачи Р1 стандартными методами линейного программирования может потребовать больших затрат времени. В частности, из ана-

лиза задачи P0 следует, что для определения x -координат размещения новых объектов может потребоваться $n^2 + 2mn$ переменных и $n(n-1)/2 + mn$ ограничений. Естественно, что в зависимости от значений величин v_{ji} и w_{ji} фактическое число переменных и ограничений может быть значительно меньше приведенных выше; однако очевидно, что для задач большой размерности желательно было бы иметь более эффективный метод решения. В работе [7] формулируется задача, аналогичная задаче P1, и для получения эффективного решения она сводится к задаче о потоке на сети.

В тех случаях, когда рассматривается только один новый объект, для решения задачи размещения используется довольно простой метод: расстояние между новым и i -м существующим объектом выражается весовым множителем w_i . Тогда можно показать, что оптимальной x -координатой (или y -координатой) размещения нового объекта является медиана величины x или y . Это соответствует такому размещению, при котором не более половины общей суммы весовых множителей имеют объекты, расположенные слева (ниже) и не более половины — справа (выше) точки размещения нового объекта.

Помимо решения данной задачи мы можем получить графическое изображение функций общих затрат в виде семейства линий одинаковых затрат и таким образом составить представление о ее форме, что в свою очередь позволяет оценить возможные варианты размещения новых объектов. Рассмотрим процесс построения линий одинаковых общих затрат для задачи размещения объектов на плоскости с принятой метрикой. Этот метод будет использован в дальнейшем при решении задачи планировки складских помещений.

1. Нанесем на график точки с координатами $(a_1, b_1), \dots, (a_m, b_m)$ и проведем через каждую точку линии, параллельные осям x и y (рис. 2).

2. Пронумеруем вертикальные линии от 1 до p слева направо, а горизонтальные линии — от 1 до q снизу вверх.

3. Обозначим точку пересечения оси x с j -й вертикальной линией через c_j , а точку пересечения оси y с i -й горизонтальной линией — d_i . Обозначим через $[i, j]$ область, ограниченную j -й и $(j+1)$ -й вертикальными линиями и i -й и $(i+1)$ -й горизонтальными линиями. (Для того чтобы все области были пронумерованы, будем считать, что слева от вертикальной линии 1 имеется линия 0, справа от вертикальной линии p — линия $p+1$, ниже горизонтальной линии 1 — линия 0 и выше горизонтальной линии q — линия $q+1$).

4. Примем величины C_j и D_i в виде суммы весовых множителей, приписываемых j -й вертикальной и i -й горизонтальной линиям соответственно. [Например, если точкам $(5, 3)$ и $(5, 10)$ сопоставлены весовые множители 6 и 8 соответственно и вторая вертикальная

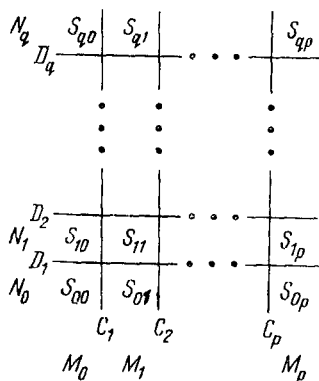


Рис. 2. Схема построения линий общих затрат.

линия пересекает ось x в точке 5, то $C_2 = 14$.] Удобно записать числа C_j под соответствующими вертикальными линиями, а числа D_i — слева от соответствующих горизонтальных линий.

5. Найдем следующие величины:

$$M_0 = - \sum_{j=1}^p C_j = - \sum_{i=1}^m w_i, \quad N_0 = - \sum_{i=1}^q D_i = - \sum_{i=1}^m w_i,$$

$$M_1 = M_0 + 2C_1, \quad N_1 = N_0 + 2D_1$$

$$M_2 = M_1 + 2C_2, \quad N_2 = N_1 + 2D_2$$

$$M_p = M_{p-1} + 2C_p = \sum_{i=1}^m w_i, \quad N_q = N_{q-1} + 2D_q = \sum_{i=1}^m w_i$$

и нанесем их на график, как показано на рис. 2.

6. Определим наклон S_{ij} в виде любой линии одинаковых общих затрат, проходящей через область $[i, j]$, по формуле

$$S_{ij} = -M_j/N_i.$$

Если $N_i = 0$, то линия одинаковых общих затрат в квадрате $[i, j]$ вертикальна.

7. Найдем координаты точки (x^*, y^*) , в которой минимизируются общие затраты. Рассмотрим 4 случая:

а) Если $M_{i-1} < 0$, $M_i > 0$, $N_{s-1} < 0$, $N_s > 0$, то $x^* = c_i$, $y^* = d_s$.

б) Если $M_{i-1} < 0$, $M_i = 0$, $N_{s-1} < 0$, $N_s > 0$, то в качестве x^* можно взять любую точку между c_i и c_{i+1} , а $y^* = d_s$.

в) Если $M_{i-1} < 0$, $M_i > 0$, $N_{s-1} < 0$, $N_s = 0$, то $x^* = c_i$, а в качестве y^* можно взять любую точку между d_s и d_{s+1} .

г) Если $M_{t-1} < 0$, $M_t = 0$, $N_{s-1} < 0$, $N_s = 0$, то x^* — любое значение между c_t и c_{t+1} , а y^* — любое значение между d_s и d_{s+1} .

8. Используя полученные выше данные, можно построить линию одинаковых общих затрат, берущую начало из любой точки, за исключением точки минимума. (Следует иметь в виду, что линия одинаковых общих затрат является замкнутой.)

1.2. Размещение объектов на плоскости с евклидовой метрикой

В случае евклидовой метрики имеем

$$d(X_j, X_k) = [(x_j - x_k)^2 + (y_j - y_k)^2]^{1/2}, \quad (7)$$

$$d(X_j, P_i) = [(x_j - a_i)^2 + (y_j - b_i)^2]^{1/2}, \quad (8)$$

где $X_j = (x_j, y_j)$ и $P_i = (a_i, b_i)$.

Подставляя выражения (7) и (8) в формулу (1), получим, что задача размещения нескольких объектов на плоскости с евклидовой метрикой состоит в *минимизации целевой функции*

$$f(X_1, \dots, X_n) = \sum_{1 \leq j < k \leq n} v_{jk} [(x_j - x_k)^2 + (y_j - y_k)^2]^{1/2} + \sum_{j=1}^n \sum_{i=1}^m w_{ji} [(x_j - a_i)^2 + (y_j - b_i)^2]^{1/2}. \quad (9)$$

(В данном случае используются те же обозначения, что и при решении задачи размещения объектов на плоскости с метрикой Минковского.) Необходимым условием оптимальности размещения новых объектов является равенство нулю (или перемена знака) частных производных функции $f(X_1, \dots, X_n)$ по $X_1, \dots, X_n$.

Частные производные $f(X_1, \dots, X_n)$ по x_j и y_j равны соответственно

$$\frac{\partial f}{\partial x_j} = \sum_{\substack{k=1 \\ k \neq j}}^n v_{jk} (x_j - x_k) / D_{jk} + \sum_{i=1}^m w_{ji} (x_j - a_i) / E_{ji}, \quad j = 1, \dots, n, \quad (10)$$

$$\frac{\partial f}{\partial y_j} = \sum_{\substack{k=1 \\ k \neq j}}^n v_{jk} (y_j - y_k) / D_{jk} + \sum_{i=1}^m w_{ji} (y_j - b_i) / E_{ji}, \quad j = 1, \dots, n, \quad (11)$$

где $D_{jk} = [(x_j - x_k)^2 + (y_j - y_k)^2]^{1/2}$, $E_{ji} = [(x_j - a_i)^2 + (y_j - b_i)^2]^{1/2}$.

Заметим, что если j -й и k -й новые объекты расположены в одной точке ($D_{jk} = 0$) и j -й новый и i -й существующий объекты расположены в одной точке ($E_{ji} = 0$), то частные производные $\partial f / \partial x_j$ и $\partial f / \partial y_j$ не определены.

Геометрически каждая составляющая целевой функции (9) представляет собой уравнение прямого кругового конуса. Следовательно, уравнение (9) представляет собой сумму конусов, вершины которых являются точками разрыва производных, образующих усеченную поверхность [92]. Так как конус есть предельная форма гиперboloида, то, заменяя конусы гиперboloидами, получим гладкую аппроксимирующую функцию $\hat{f}$. Более того, поскольку функции, описывающие гиперboloиды, являются строго выпуклыми, то функция $\hat{f}$ также строго выпуклая при условии, что по крайней мере один весовой множитель (коэффициент) w_{ji} больше нуля для каждого j -го объекта.

Уравнение гиперboloида с центром в точке (a_i, b_i) на плоскости $x - y$ может быть записано в виде

$$\hat{f}_{ji} = w_{ji} [(x_j - a_i)^2 + (y_j - b_i)^2 + \varepsilon]^{1/2}.$$

Можно видеть, что добавление константы ε соответствует замене вершины конуса на гладкую гиперболическую поверхность, и следовательно, частные производные существуют всюду. Кроме того, чем меньше величина ε , тем точнее гиперboloид аппроксимирует конус.

Полагая, что

$$\hat{D}_{jk} = [(x_j - x_k)^2 + (y_j - y_k)^2 + \varepsilon]^{1/2},$$

$$\hat{E}_{ji} = [(x_j - a_i)^2 + (y_j - b_i)^2 + \varepsilon]^{1/2},$$

можно сформулировать новую задачу оптимизации как задачу минимизации целевой функции вида

$$\hat{f}(X_1, \dots, X_n) = \sum_{1 \leq j < k \leq n} v_{jk} \hat{D}_{jk} + \sum_{j=1}^n \sum_{i=1}^m w_{ji} \hat{E}_{ji}. \quad (12)$$

где $\lim_{\varepsilon \rightarrow 0} \hat{f}(X_1, \dots, X_n) = f(X_1, \dots, X_n)$.

При очень малых значениях константы ε решение уравнения (12) почти аналогично решению уравнения (9).

Беря частные производные $\hat{f}$ по x_j и y_j , приравнявая их нулю и разрешая относительно x_j и y_j , получаем следующие итерационные формулы:

$$x_j^{(h+1)} = \frac{\sum_{\substack{k=1 \\ k \neq j}}^n v_{jk} x_k^{(h)} / \hat{D}_{jk}^{(h)} + \sum_{i=1}^m w_{ji} a_i / \hat{E}_{ji}^{(h)}}{\sum_{\substack{k=1 \\ k \neq j}}^n v_{jk} / \hat{D}_{jk}^{(h)} + \sum_{i=1}^m w_{ji} / \hat{E}_{ji}^{(h)}}, \quad (13)$$

$$y_j^{(h+1)} = \frac{\sum_{k=1}^n v_{jk} y_k^{(h)} / \hat{D}_{jk}^{(h)} + \sum_{i=1}^m w_{ji} b_i / \hat{E}_{ji}^{(h)}}{\sum_{k=1}^n v_{jk} / \hat{D}_{jk}^{(h)} + \sum_{i=1}^m w_{ji} / \hat{E}_{ji}^{(h)}}, \quad (14)$$

где верхние индексы обозначают номер итерации.

Итерационные выражения (13) и (14) в дальнейшем будем называть методом гиперболической аппроксимации [27]. Известно, что если решение сходится, то это решение оптимально. Хотя не доказано, что выражения (13) и (14) всегда дают сходящееся решение, однако при большом числе экспериментальных решений, полученных методом гиперболической аппроксимации, не наблюдалось ни одного случая, когда бы решение не сходилось. Было показано, что выражения (13) и (14) сводятся к итерационным выражениям, используемым в градиентных методах поиска, однако величина шага при гиперболической аппроксимации не оптимальна. (Более подробно метод гиперболической аппроксимации обосновывается в [27].)

Для случая одного объекта и при $\varepsilon = 0$ выражения (13) и (14) сводятся соответственно к виду

$$x^{(h+1)} = \frac{\sum_{i=1}^m w_i a_i / E_i^{(h)}}{\sum_{i=1}^m w_i / E_i^{(h)}}, \quad (15)$$

$$y^{(h+1)} = \frac{\sum_{i=1}^m w_i b_i / E_i^{(h)}}{\sum_{i=1}^m w_i / E_i^{(h)}}, \quad (16)$$

где $E_i^{(h)} = [(x^{(h)} - a_i)^2 + (y^{(h)} - b_i)^2]^{1/2}$.

Сходимость выражений (15) и (16) к оптимальному размещению исследовалась рядом авторов, в том числе Купом [61]; использование формул (15) и (16) для решения задачи размещения одного объекта на плоскости с евклидовой метрикой можно найти в работах [9-11, 60 и 71].

2. ЗАДАЧИ РАЗМЕЩЕНИЯ — РАСПРЕДЕЛЕНИЯ С ИСПОЛЬЗОВАНИЕМ НЕПРЕРЫВНОГО ПРОСТРАНСТВА РЕШЕНИЙ

При анализе задач размещения многих объектов предполагалось, что весовые множители v_{jk} и w_{ji} не зависят от размещения объектов. Теперь рассмотрим случай, когда все весовые

множители v_{jk} равны нулю, а весовые множители w_{ji} являются искомыми переменными наряду с координатами размещения новых объектов, т. е. так называемую задачу размещения — распределения, причем будем использовать непрерывное пространство решений [9—11].

В самой общей постановке задача размещения — распределения состоит в определении числа новых объектов и координат их размещения, а также в распределении перевозок между новыми и существующими объектами. Одним из примеров является задача размещения оптовых баз, получающих товары от производственных предприятий и распределяющих их между розничными и оптовыми торговыми точками; другим примером является задача размещения ряда отделений банка в крупном городе. Вместо отделений банка и складов легко представить себе в качестве новых объектов, например, больницы или бакалейные магазины. В случае отделений банка, больниц или бакалейных магазинов в состав существующих объектов включаются места проживания потребителей.

Задача размещения — распределения может быть сформулирована следующим образом:

найти

$$\min \Psi = \sum_{j=1}^n \sum_{i=1}^m z_{ji} w_{ji} d(X_j, P_i) + g(n) \quad (17)$$

при условии

$$\sum_{j=1}^n z_{ji} = 1, \quad i = 1, \dots, m, \quad n = 1, 2, \dots, m,$$

где w_{ji} — удельные затраты в единицу времени при взаимодействии j -го нового и i -го существующего объектов; z_{ji} равно 1, если j -й новый объект функционально связан с i -м существующим объектом, и равно 0 в противном случае; $X_j = (x_j, y_j)$ — координаты размещения j -го нового объекта; $P_i = (a_i, b_i)$ — координаты расположения i -го существующего объекта; $g(n)$ представляет собой затраты на перевозки в единицу времени при установке n новых объектов и Ψ — общие затраты на перевозки в единицу времени.

Управляющими переменными в выражении (17) являются n , z_{ji} , x_j и y_j . Каждое из ограничений означает, что любой из существующих объектов взаимодействует лишь с одним новым объектом. Поскольку не задано никаких ограничений на производительность объектов, предполагается, что новый объект удовлетворяет всем требованиям, предъявляемым к взаимодействию с существующими объектами. Кроме того, предполагается, что весовые множители $w_{j1} = w_{j2} = \dots = w_{ji}$, поскольку речь идет о размещении идентичных новых объектов. К сожалению, оптимальное решение

уравнения (17) получить нелегко. Трудность заключается в определении управляющих переменных n и z_{ji} .

Один из подходов, который используется для решения этой задачи, состоит в том, что при фиксированной величине n рассматриваются все возможные комбинации z_{ji} . Для каждого набора значений z_{ji} оптимальное решение получается путем решения соответствующей задачи размещения одного объекта. Затем определяется набор значений z_{ji} для данного n , обеспечивающий минимум затрат. И наконец, получают оптимальное решение путем поиска величины n .

Описанный подход невозможно использовать для решения задач большой размерности, так как слишком велико число комбинаций значений z_{ji} , которые необходимо рассмотреть для каждого значения n и которые определяются формулой Стирлинга второго рода:

$$S(n, m) = \sum_{k=0}^n \frac{(-1)^k (n-k)^m}{k! (n-k)!}.$$

Напомним, что подсчет числа комбинаций основан на предположении, что все новые объекты идентичны; если это не так, то пришлось бы рассматривать еще больше комбинаций.

В связи с тем что для точного решения задачи размещения распределения могут потребоваться значительные вычислительные затраты, был разработан ряд эвристических методов получения хороших решений [10, 11]. Авторы работы [59] приводят алгоритм метода ветвей и границ, а в работе [86] дается алгоритм билинейного программирования для определения оптимального распределения и размещения n новых объектов.

3. ЗАДАЧИ ПЛАНИРОВКИ С ИСПОЛЬЗОВАНИЕМ НЕПРЕРЫВНОГО ПРОСТРАНСТВА РЕШЕНИЙ

В тех случаях, когда новый объект не является точечным, возникает необходимость в его планировке. Рассмотрим задачу планировки товарного склада, причем в дальнейшем в качестве новых объектов будем рассматривать товары, предназначенные для хранения, а в качестве существующих объектов — приемные и погрузочные устройства.

Для простоты изложения введем дополнительные обозначения. В частности, обозначим число различных товаров, которые должен вмещать склад, через m ; множество точек на плоскости, занимаемой i -м товаром, — через S_i и площадь, соответствующую этому множеству, — через A_i . Далее предположим, что для любых i -го и j -го товаров множества S_i и S_j не пересекаются. Поскольку множества $S_1, \dots, S_m$ определяют расположение всех товаров, то набор $\{S_1, \dots, S_m\}$ является планом размещения. Для любого

плана размещения каждое множество S_i располагается в известной области L , которая может быть представлена, например, в виде плана площади, на которой должен располагаться склад. Множество всех планов размещения обозначим через $\Pi_m (L : A)$. Допустим теперь, что склад имеет n погрузочных платформ с известным расположением $P_1, \dots, P_n$. Для любого заданного плана размещения $\{S_1, \dots, S_m\}$ с равной вероятностью помещения товара в каждую точку $|X - P_j|$ представляет собой расстояние между X и P_j , и среднее расстояние, на которое перемещается i -й товар к j -й погрузочной платформе и от нее, может быть представлено формулой

$$\int_{S_i} (1/A_i) |X - P_j| dX.$$

Предположим, что стоимость перевозки i -го товара к j -й платформе и от нее за данный период времени прямо пропорциональна среднему расстоянию с неотрицательным коэффициентом пропорциональности w_{ij} . Тогда общая стоимость перемещения i -го товара к j -й платформе и от нее за данный период времени определяется формулой

$$w_{ij} \int_{S_i} (1/A_i) |X - P_j| dX,$$

а общая стоимость перемещения всех товаров — выражением

$$F(S_1, \dots, S_m) = \sum_{i=1}^m \sum_{j=1}^n w_{ij} \int_{S_i} (1/A_i) |X - P_j| dX. \quad (18)$$

Если для элементов матрицы $W = (w_{ij})$ выполняется соотношение $w_{ij} = c_i w_j$, где $c_i = \sum_{j=1}^n w_{ij}$ для $i = 1, \dots, m$, и

$$w_j = \sum_{i=1}^m w_{ij} / \sum_{i=1}^m \sum_{j=1}^n w_{ij},$$

то выражение (18) принимает следующий вид:

$$F(S_1, \dots, S_m) = \sum_{i=1}^m (c_i/A_i) \int_{S_i} f(X) dX, \quad (19)$$

где

$$f(X) = \sum_{j=1}^n w_j |X - P_j|. \quad (20)$$

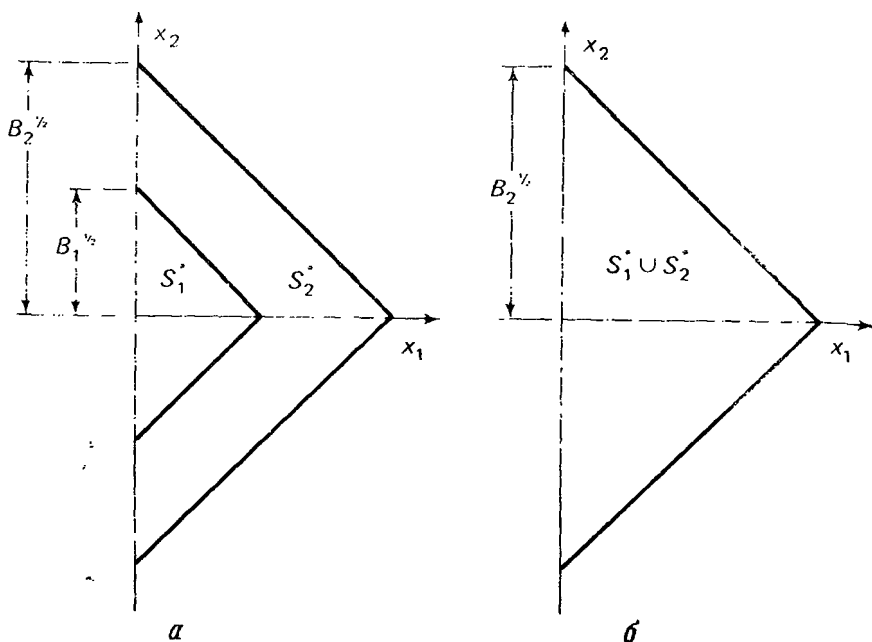


Рис. 3. Планировка зоны складирования (а) и проектирование склада (б) для случая размещения нескольких товаров и одной погрузочной платформы.

Предположение, сделанное относительно элементов матрицы W , удовлетворяется всякий раз, когда общее количество (w_{ij}) i -го товара, перевозимого между складом и j -й платформой в единицу времени, может быть получено путем умножения общего количества (c_i) i -го товара, перевозимого на склад и вывозимого со склада в единицу времени, на долю всех товаров w_j , перевозимых между складом и j -й платформой.

Прежде чем получить условия оптимальной планировки склада, введем некоторые новые обозначения. Для любого данного плана размещения $\{S_1, \dots, S_m\}$ объединение $S_1, \dots, S_q$, определяющее совокупность всех точек, принадлежащих по крайней мере одному из множеств $S_1, \dots, S_q$, обозначим через $\bigcup_{i=1}^q S_i$,

где $q = 1, 2, \dots, m$. В нашем случае $\bigcup_{i=1}^q S_i$ определяет местоположение всех товаров от 1 до q , а занимаемая ими площадь B_q определяется формулой $B_q = A_1 + \dots + A_q$, где $= 1, \dots, m$ (рис. 3).

Достаточным условием минимизации затрат на перевозки является следующее. Если план размещения $\{S_1^*, \dots, S_m^*\}$ есть один из множества $H_m(L : A)$ и если $c_1/A_1 \geq \dots \geq c_m/A_m > 0$, то

$$F(S_1^*, \dots, S_m^*) \leq F(S_1, \dots, S_m),$$

где $\{S_1, \dots, S_m\}$ — любой другой план размещения из множества $H_m(L : A)$, при этом размещение $\{S_1^*, \dots, S_m^*\}$ должно удовлетворять требованию, согласно которому для $q = 1, \dots, m$ существуют такие $k_1 \leq \dots \leq k_m$, что $f(X) \leq k_q$ для каждой точки X из $\bigcup_{i=1}^q S_i^*$ и $f(X) \geq k_q$ для каждой точки X , не принадлежащей $\bigcup_{i=1}^q S_i^*$.

В этом случае план размещения можно получить путем построения линий одинаковых общих затрат (19). Если использовать терминологию разд. 1.2, то погрузочные платформы — это существующие объекты, а w_j — весовые множители, приписываемые товарам, которые подлежат хранению на складе, и j -му существующему объекту. Линии одинаковых затрат, соответствующая величине k_1 , ограничивает область $B_1 = A_1$, площадь которой используется для хранения товара 1; линия, соответствующая величине k_2 , ограничивает область $B_2 = A_1 + A_2$, площадь которой используется для хранения товаров 1 и 2 и т. д. Подлежащие хранению товары нумеруются в соответствии с величиной отношения c/A , причем дефицитные товары, как правило, располагают ближе к погрузочным платформам. Таким образом, если рассматривается прямолинейное перемещение товаров, то линии одинаковых общих затрат могут быть построены методом, описанным в разд. 1.1.

Если требуется получить величину общих затрат для проекта оптимальной планировки склада, целесообразно вывести единое интегральное выражение для планов размещения, обеспечивающих наименьшие затраты. Для этого определим функцию $q(k)$ как площадь множества $\{X \in L : f(X) \leq k\}$. При довольно слабых ограничениях функция $q(k)$ является строго возрастающей и имеет обратную функцию $r(t)$, такую, что $q(r(B_i)) = B_i$ и $r(q(k_i)) = k_i$. Кроме того, $r(B_0)$ и k_0 соответствуют минимальному значению функции $f(X)$.

Предположим, что существуют такие $k_0 < k_1 < \dots < k_m$, что план размещения $\{S_1^*, \dots, S_m^*\}$ является одним из множества $H_m(L : A)$, причем

$$S_i^* = \{X \in L : k_{i-1} \leq f(X) \leq k_i\}, \quad i = 1, \dots, m.$$

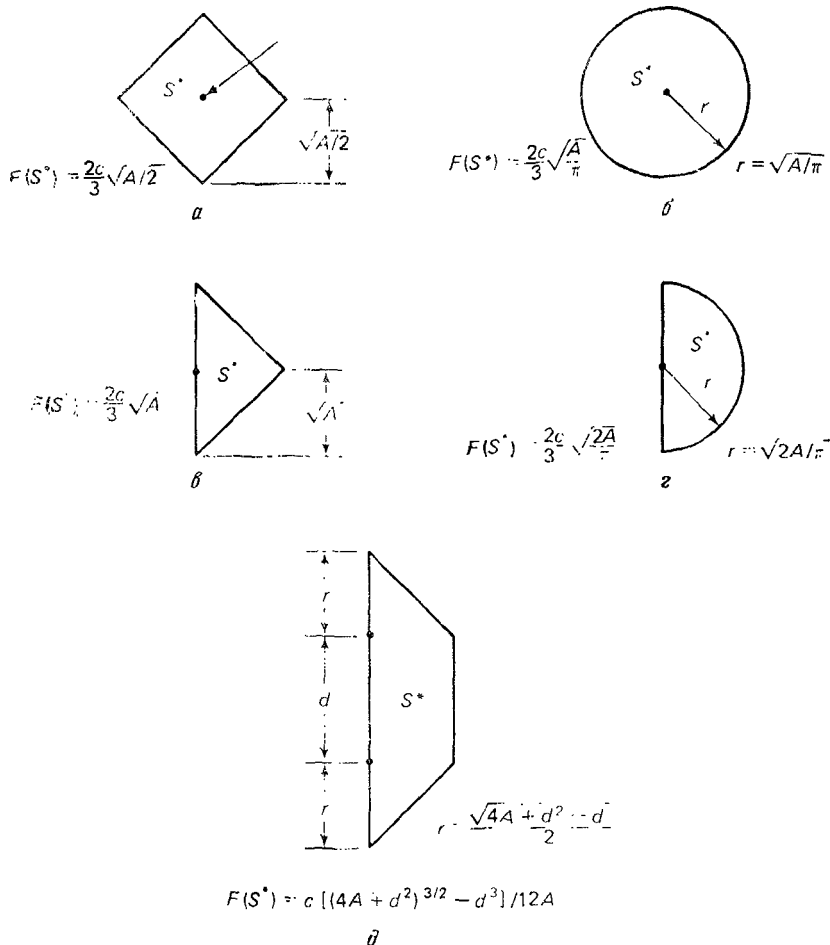


Рис. 4. Варианты планировки склада для хранения однотипных товаров.

Тогда для $i = 1, \dots, m$ имеем $k_i = r(B_i)$ и

$$\int_{S_i^*} (1/A_i) f(X) dX = (1/A_i) \int_{r(B_{i-1})}^{r(B_i)} q'(z) z dz, \quad (21)$$

где $q'(z)$ — первая производная $q(z)$ по z .

Принимая во внимание полученный вывод, можно вычислить ожидаемое расстояние, на которое за единицу времени перемещается i -й товар. Таким образом, ожидаемые общие затраты

на перемещение всех товаров могут быть даны формулой

$$F(S_1^*, \dots, S_m^*) = \sum_{i=1}^m (c_i/A_i) \int_{r(B_{i-1})}^{r(B_i)} g'(z) z dz. \quad (22)$$

В качестве иллюстрации использования полученных выводов рассмотрим оптимальную планировку склада, предназначенного для хранения однородных товаров (рис. 4) [29, 30]. Согласно планировкам, представленным на рис. 4, а и б, единственная погрузочная платформа располагается в центре склада, что позволяет минимизировать ожидаемое перемещение товара в единицу времени. Если перемещение товаров осуществляется в пространстве Минковского, то оптимальной конфигурацией склада будет ромб, а если перемещение товаров осуществляется в евклидовой плоскости, то круг. В тех случаях, когда платформа располагается на периферии склада, планировки для пространства Минковского и евклидова пространства будут иметь вид, представленный на рис. 4, в и г. Если используются две платформы, а перемещение осуществляется в пространстве Минковского и $w_1 = w_2 = 1/2$, то получаемая планировка будет такой, как показано на рис. 4, д. На рис. 4 приведены также формулы ожидаемых общих затрат $F(S^*)$ для каждой планировки.

Аналогичная задача планировки и размещения может быть также сформулирована и для случая, когда новые объекты рассматриваются как точки, а существующие объекты — как точки либо как протяженные объекты [98].

В некотором смысле данная задача планировки может рассматриваться как частный случай более общей задачи проектирования региона [13]. Последняя может быть сформулирована как задача определения оптимальных областей на плоскости, удовлетворяющих некоторым заданным критериям, т. е. задача заключается в разбиении данного региона на подобласти. Примерами подобных задач являются деления штата на избирательные округа, деление округа на школьные микрорайоны, разделение данного региона на зоны обслуживания телефонными станциями, размещение предприятия, имеющего несколько филиалов на территории заданного региона, размещение автостоянки в зависимости от размещения объектов прибывания людей и проектирование склада в зависимости от размещения погрузочных платформ.

4. ЗАДАЧИ ПЛАНИРОВКИ С ИСПОЛЬЗОВАНИЕМ ДИСКРЕТНОГО ПРОСТРАНСТВА РЕШЕНИЙ

Проанализируем задачу планировки, имеющую дискретное пространство решений, в предположении, что между новыми объектами нет взаимодействия. В качестве примера рассмотрим задачу планировки склада.

Предположим, что L — некоторая область складского помещения, предназначенная для хранения товаров, разбита на n квадратов одинакового размера, пронумерованных любым удобным образом от 1 до n . Предположим также, что на складе должны храниться m товаров, и обозначим общее число квадратов, которые будет занимать i -й товар, через A_i (например, если товар 3 занимает 4 квадрата, то $A_3 = 4$). Если склад имеет p погрузочных платформ, размещение которых известно, то соответствующим образом определенное расстояние между k -й платформой и центром j -го квадрата будет d_{kj} . Обозначим через S_i множество квадратов, которые занимает i -й товар [например, если i -й товар занимает 4 квадрата, имеющих номера 12, 13, 29 и 30, то $S_i = \{12, 13, 29, 30\}$]. При предположении, что для данной зоны складирования товара он будет с равной вероятностью перемещаться между k -й платформой и любым квадратом, занимаемым i -м товаром, среднее расстояние перемещения i -го товара между k -й платформой и зоной его складирования может быть записано в виде

$$\sum_{j \in S_i} (1/A_i) d_{kj},$$

т. е. среднее расстояние является суммой расстояний между k -й платформой и всеми квадратами, занимаемыми i -м товаром, деленной на A_i . Наконец, пусть w_{ik} представляет собой известные общие удельные затраты (на единицу расстояния) на перемещение i -го товара между k -й погрузочной платформой и зоной его хранения за заданный период времени. Обычно, если товары хранятся, в упаковках, величина w_{ik} прямо пропорциональна числу загруженных i -м товаром упаковок, перемещенных между k -й платформой и зоной хранения i -го товара за некоторый заданный период времени. Таким образом, величина

$$w_{ik} \sum_{j \in S_i} (1/A_i) d_{kj} \quad (23)$$

является общими средними затратами на транспортировку i -го товара между k -й погрузочной платформой и зоной складирования i -го товара. Общие средние затраты за рассматриваемый период времени на транспортировку товаров на склад и со склада определяются как

$$F(S_1, \dots, S_m) = \sum_{i=1}^m \sum_{k=1}^p w_{ik} \left[\sum_{j \in S_i} (1/A_i) d_{kj} \right]. \quad (24)$$

Задача планировки склада состоит в том, чтобы таким образом разместить на складе каждый товар, предназначенный для хранения (при условии, что в любом квадрате, на которые разбит склад,

может находиться не более одного вида товара), чтобы общие затраты, определяемые выражением (24), были минимальными.

Чтобы преобразовать задачу планировки в обобщенную задачу размещения, определим управляющие переменные x_{ij} следующим образом: управляющие переменные равны единице, если товар i ($i = 1, \dots, m$) занимает квадрат j ($j = 1, \dots, n$), и равны нулю в противном случае. Сделаем временное допущение, согласно которому общее количество квадратов, занимаемых товарами, равно общему числу квадратов, на которые разбит склад, т. е. $\sum_{i=1}^m A_i = n$. Так как i -й товар занимает всего A_i квадратов, то, следовательно,

$$\sum_{j=1}^n x_{ij} = A_i \quad \text{для } i = 1, \dots, m.$$

Так как каждый квадрат занят только одним товаром, то имеем

$$\sum_{i=1}^m x_{ij} = 1 \quad \text{для } j = 1, \dots, n,$$

Если переменные x_{ij} определены указанным образом, то S_i — это совокупность всех j -х квадратов, для которых $x_{ij} = 1$, т. е. $S_i = \{j : x_{ij} = 1\}$. Это означает, что выражение (23) эквивалентно выражению

$$w_{ik} \sum_{j=1}^n (1/A_i) d_{kj} x_{ij}$$

и формула для общих затрат (24) эквивалентна выражению

$$\sum_{i=1}^m \sum_{k=1}^p w_{ik} \left[\sum_{j=1}^n (1/A_i) d_{kj} x_{ij} \right]. \quad (25)$$

Если ввести величину

$$c_{ij} = (1/A_i) \sum_{k=1}^p w_{ik} d_{kj}, \quad \text{для } i = 1, \dots, m \text{ и } j = 1, \dots, n,$$

то выражение (25) для общих затрат можно записать в виде

$$\sum_{i=1}^m \sum_{j=1}^n c_{ij} x_{ij}.$$

Теперь обобщенная задача размещения, эквивалентная рассматриваемой задаче планировки, может быть сформулирована следующим образом:

найти

$$\min \sum_{i=1}^m \sum_{j=1}^n c_{ij} x_{ij}$$

при ограничениях

$$\sum_{j=1}^n x_{ij} = A_i, \quad i = 1, \dots, m,$$

$$\sum_{i=1}^m x_{ij} = 1, \quad j = 1, \dots, n,$$

$$x_{ij} \in \{0, 1\}, \quad i = 1, \dots, m, \quad j = 1, \dots, n.$$

В некоторых случаях число товаров и квадратов настолько велико, что обобщенная задача размещения оказывается слишком большой размерности и практически ее невозможно решать на ЭВМ с помощью существующих программ. Однако, если сделать соответствующее допущение, которое во многих случаях является вполне оправданным, то задача планировки может быть решена с помощью более простого метода, чем метод решения обобщенной задачи размещения. Допущение делается относительно величины w_{ik} , для которой должно выполняться соотношение $w_{ik} = c_i w_k$ для $i = 1, \dots, m$ и $k = 1, \dots, p$. В этом случае функционал $F(S_1, \dots, S_m)$ можно представить в виде

$$F(S_1, \dots, S_m) = \sum_{i=1}^m (c_i/A_i) \sum_{j \in S_i} f_j,$$

$$\text{где } f_j = \sum_{k=1}^p w_k d_{kj} \quad \text{для } j = 1, \dots, n.$$

Если товары пронумерованы таким образом, что

$$c_1/A_1 \geq c_2/A_2 \geq \dots \geq c_m/A_m,$$

рассматриваемая задача сводится к нахождению такого плана размещения $\{S_1^*, \dots, S_m^*\}$, при котором $F(S_1^*, \dots, S_m^*) \leq F(S_1, \dots, S_m)$, где $\{S_1, \dots, S_m\}$ — любой другой план размещения.

Теперь перейдем к нахождению плана размещения, обеспечивающего минимальные затраты на перевозки [35]. Если для величин $j = 1, \dots, n$ известна функция

$$f_j = \sum_{k=1}^p w_k d_{kj},$$

причем $j_1, j_2, \dots, j_n$ есть такая последовательность чисел $1, \dots, n$, что $f_{j_1} \leq f_{j_2} \leq \dots \leq f_{j_n}$, то минимальные затраты

возможны при следующем размещении $\{S_1^*, \dots, S_m^*\}$ из множества $H_m(L: A)$:

$$\begin{aligned} S_1^* &= \{j_1, \dots, j_{B_1}\}, \\ S_2^* &= \{j_{B_1+1}, \dots, j_{B_2}\}, \\ &\vdots \\ S_m^* &= \{j_{B_{m-1}+1}, \dots, j_{B_m}\}, \end{aligned}$$

$$\text{где } B_i = \sum_{h=1}^i A_h \quad \text{для } i = 1, \dots, m.$$

5. ЗАДАЧИ РАЗМЕЩЕНИЯ С ИСПОЛЬЗОВАНИЕМ ДИСКРЕТНОГО ПРОСТРАНСТВА РЕШЕНИЙ

В тех случаях, когда между множеством пар объектов допускается существование определенной взаимосвязи, задача размещения таких объектов носит дискретный характер и обычно называется задачей размещения с квадратичной функцией, поскольку может быть сведена (хотя мы не будем этого делать) к задаче с квадратичной целсовой функцией.

Обозначим через $a(j)$ номер места размещения j -го объекта, и пусть $\mathbf{a} = (a(1), a(2), \dots, a(n))$ есть вектор размещения, где n — число возможных мест для размещения n новых объектов. (Если число возможных мест размещения превышает число новых объектов, то вводится соответствующее число фиктивных объектов.) Например, вектор размещения $\mathbf{a} = (4, 1, 2, 3)$ означает, что объект 1 помещается на место 4, объект 2 — на место 1, объект 3 — на место 2 и объект 4 — на место 3.

После того как разработан план размещения новых объектов, необходимо определить общие затраты на его реализацию. Для этого расстояние между местами расположения объектов обозначим как $d(g, h)$, где g и h могут принимать значения от 1 до n , и предположим, что $d(g, h) = d(h, g)$ для всех g и h . Кроме того, введем константу пропорциональности v_{jh} , с помощью которой расстояние между новыми j -м и h -м объектами для всех $j < h$ преобразуется в затраты, т. е. если j -й объект размещается в месте $a(j)$, а h -й объект — в месте $a(h)$, то общие затраты, связанные с новыми j -м и h -м объектами, равны $v_{jh} d(a(j), a(h))$. Аналогично пусть расстояние между существующим i -м объектом и местом h размещения нового объекта равно $c(i, h)$ и w_{ji} есть константа пропорциональности, с помощью которой расстояние между существующим i -м и новым j -м объектами преобразуется в затраты, т. е. если j -й объект размещается в месте $a(j)$, то общие затрата-

ты для j -го и i -го объектов равны $w_{ji}c(a(j), i)$. Тогда общие затраты на реализацию плана размещения равны

$$TC(a) = \sum_{1 \leq j < k \leq n} v_{jk} d(a(j), a(k)) + \sum_{j=1}^n \sum_{i=1}^m w_{ji}c(a(j), i). \quad (26)$$

Таким образом, задача размещения с квадратичной функцией назначения состоит в определении вектора размещения a , при котором функционал (26) имеет минимальное значение.

Если возникают дополнительные затраты, не пропорциональные расстоянию, то имеем

$$TC(a) = \sum_{1 \leq j < k \leq n} [v_{jk} d(a(j), a(k)) + p(a(j), a(k))] + \sum_{j=1}^n \sum_{i=1}^m w_{ji}c(a(j), i) + \sum_{j=1}^n q(a(j)), \quad (27)$$

где $p[a(j), a(k)]$ — затраты на размещение новых j -го и k -го объектов соответственно в местах $a(j)$ и $a(k)$, не зависящие от расстояния, а $q[a(j)]$ — затраты на размещение нового j -го объекта на месте $a(j)$, не зависящие от расстояния. Для упрощения последующего изложения будем рассматривать выражение (26), а не (27). Однако результаты могут быть распространены и на случай формулы (27).

Для решения рассматриваемой задачи был разработан ряд точных и эвристических методов, причем точные методы, за исключением метода полного перебора, являлись, как правило, модификацией метода ветвей и границ, т. е. неявного перебора. Поскольку, однако, точные методы оказываются неприемлемыми при $n > 15$, то предпочтение отдается эвристическим методам. (Что касается метода полного перебора, то он исключается сразу, так как в этом случае требуется рассмотреть все $n!$ вариантов размещений.)

Одним из наиболее удачных эвристических методов решения квадратичной задачи размещения является метод наискорейшего спуска или попарных перестановок, который обычно называют методом CRAFT¹⁾ [2]. В этом случае задается начальный вектор размещения a и рассматриваются изменения величины затрат при взаимной перестановке каждой пары объектов. Среди всех возможных перестановок выбирается и фиксируется та, при которой достигается наибольшее снижение величины затрат. Полученной величиной затрат соответствует новый вектор размещения a' , для

¹⁾ Computerized Relative Allocation of Facilities Technique

которого весь процесс повторяется, в результате чего получается второе улучшение назначения a'' , причем

$$TC(a) > TC(a') > TC(a''). \quad (28)$$

Этот процесс продолжается до тех пор, пока ни одна из возможных перестановок не даст дальнейшего снижения затрат. Полученное таким методом размещение считается решением квадратичной задачи.

Для решения квадратичной задачи размещения могут быть использованы и другие эвристические методы, сравнение которых проводится в работах [50, 51, 76].

Рассматриваемые эвристические методы — это процедуры улучшения: задается некоторое начальное размещение (решение), и целью последующих итераций является улучшение этого решения. Другой используемый подход состоит в том, что процесс конструирования начинают с нулевого вектора и производят назначения одного из объектов на одно из мест расположения до тех пор, пока все объекты не будут размещены [51].

Для получения точного решения квадратичной задачи назначения был разработан ряд подходов, основанных на методе ветвей и границ [35, 37, 40, 62].

6. ЗАДАЧИ РАЗМЕЩЕНИЯ — РАСПРЕДЕЛЕНИЯ С ИСПОЛЬЗОВАНИЕМ ДИСКРЕТНОГО ПРОСТРАНСТВА РЕШЕНИЙ

Задачи подобного рода, как правило, приходится решать при размещении производства и складов.

6.1. Задачи размещения предприятий ¹⁾

Задача размещения предприятий обычно включает в себя определение их числа, места расположения и мощности заводов и складов, необходимых для снабжения потребителей продукцией. При этом предполагается, что известно конечное число мест возможного размещения предприятий и расположения потребителей.

В работе [3] была впервые сформулирована задача размещения предприятий с дискретным пространством решений как задача смешанного целочисленного программирования. Для постановки задачи смешанного программирования введем следующие обозначения: m — число потребителей, n — число возможных размещений предприятий, y_{ij} — доля или часть потребностей i -го потре-

¹⁾ Авторы выражают свою глубокую признательность Леону Ф. Макджиннису за помощь в подготовке данного раздела. Обзор результатов по исследованию задачи размещения предприятий см. в работе [70].

бителя, которая удовлетворяется предприятием, расположенным в j -м месте, причем $i = 1, \dots, m$; $j = 1, \dots, n$; c_{ij} — стоимость полного удовлетворения потребностей i -го потребителя предприятием, расположенным в j -м месте, и f_j — постоянные затраты, обусловленные размещением предприятия в j -м месте.

Управляющие переменные x_j равны единице, если предприятие располагается в j -м месте, и нулю в противном случае.

Используя приведенные выше обозначения, можно сформулировать задачу размещения Р0, называемую задачей размещения одного предприятия, следующим образом:

найти

$$\min z = \sum_{i=1}^m \sum_{j=1}^n c_{ij} y_{ij} + \sum_{j=1}^n f_j x_j \quad (29)$$

при ограничениях

$$\sum_{i=1}^m y_{ij} \leq m x_j, \quad j = 1, \dots, n, \quad (30)$$

$$\sum_{j=1}^n y_{ij} = 1, \quad i = 1, \dots, m, \quad (31)$$

$$y_{ij} \geq 0 \text{ для всех } i, j, x_j = (0, 1) \text{ для всех } j.$$

Целевая функция (29) определяет затраты при размещении $\sum_{j=1}^n x_j$ предприятий в местах, соответствующих положительным значениям x_j . Ограничения (30) означают, что общая доля потребностей потребителя, удовлетворяемая предприятием, расположенным в j -м месте, либо равна нулю, если $x_j = 0$, либо не может превосходить числа потребителей, если $x_j = 1$. Из выражения (31) следует, что все потребности i -го потребителя должны быть удовлетворены некоторым набором предприятий. Обычно величины c_{ij} представляют собой транспортные расходы. Если бы не было постоянных затрат, то оптимальное решение состояло бы в том, чтобы строить предприятие в каждом из возможных мест. С другой стороны, если бы не было затрат на транспортировку, оптимальное решение состояло бы в постройке одного предприятия в месте, для которого постоянные затраты минимальны. Таким образом, необходимо найти решение, обеспечивающее определенное соотношение между транспортными и постоянными затратами.

В более общем случае величина y_{ij} определяется как часть потребности i -го потребителя, удовлетворяемой предприятием, расположенным в j -м месте, при этом, как правило, на производственную мощность предприятия накладываются ограничения. С учетом этих факторов задача Р0 сводится к задаче размещения

предприятия с ограничением производительности (P1) и формулируется следующим образом:

найти

$$\min z = \sum_{i=1}^m \sum_{j=1}^n c_{ij} y_{ij} + \sum_{j=1}^n g_i \left(\sum_{i=1}^m y_{ij} \right)$$

при ограничениях

$$\sum_{i=1}^m y_{ij} - Q_j x_j \leq 0,$$

$$\sum_{j=1}^n y_{ij} = d_i, \quad y_{ij} \geq 0, \quad x_j = (0, 1),$$

где Q_j — производительность объекта, расположенного в j -м месте; c_{ij} — стоимость транспортировки единицы продукции из j -го места в i -е; $g_i(\cdot)$ — функция затрат для предполагаемого объекта, размещаемого в j -м месте, которые могут включать также постоянные затраты; y_{ij} — объем поставок i -му потребителю из j -го места; d_i — количество продукции, которая должна поставляться i -му потребителю.

Модификации задачи P1 учитывают случаи нелинейных функций $g_i(\cdot)$, а также нелинейность функции затрат на распределение продукции. Задача P1 может быть также сформулирована при дополнительных ограничениях на взаимное расположение предприятий. Ограничения подобного рода могут, например, включаться, чтобы предотвратить размещение предприятия в j -м месте, если какое-то предприятие размещается, например, в k -м месте. Различные варианты подобных задач рассматриваются в работах [20, 22, 23, 82, 88, 89].

В матричной форме задача P1 может быть записана следующим образом (задача P2):

найти

$$\min (fx + CY)$$

при ограничениях

$$A_1 x + A_2 Y = b, \quad Y \geq 0, \quad x \in X, \quad x_j = (0, 1),$$

где X — множество допустимых комбинаций мест размещения предприятий, определяемое любыми ограничениями на взаимное расположение; A_1 и A_2 — соответствующим образом определенные матрицы, а b — вектор-столбец.

Задача P2 является задачей смешанного целочисленного программирования с целочисленными переменными, принимающими значения 0 и 1. Предлагаемые методы для решения задачи P2 относятся к методам перечисления и включают «релаксационные

алгоритмы» и «алгоритмы разбиения». Для обоих алгоритмических подходов характерен упорядоченный поиск оптимального значения булевых переменных x_j , который можно представить в виде дерева поиска.

Каждая вершина дерева поиска данной задачи P2 однозначно соответствует разбиению индексов булевых переменных, т. е.

$$K_0 = \{j: x_j \text{ присвоено значение } 0\},$$

$$K_1 = \{j: x_j \text{ присвоено значение } 1\},$$

$$K_2 = \{j: x_j \text{ не присвоено значения}\}.$$

Любое x_k для $k \in K_2$ называется свободной переменной. Из любой вершины v дерева исходят по крайней мере две ветви. Две порожденные вершины соответствуют выбору в вершине v свободной переменной x_k и образованию двух новых разбиений, считая поочередно, что $k \in K_0$ и $k \in K_1$. Эта операция называется ветвлением по x_k .

Задачу P2, в которой опущено требование целочисленности, обозначим как RP. Если все свободные переменные в оптимальном решении задачи линейного программирования целочисленны, то это решение является допустимым для задачи P2; в противном случае для ветвления выбирается одна из нецелочисленных переменных x_k .

Аналогично задачу P2, в которой переменным присвоены некоторые значения $x \in X$, обозначим как SP. В частности, фиксированному значению $x \in X$ соответствует оптимальное значение Y , обозначаемое Y_x , и задача SP может быть сформулирована следующим образом:

найти

$$\min CY$$

при ограничении

$$A_2 Y = b - A_1 x, Y \geq 0$$

Алгоритмы разбиения заключаются в поиске по X . Переменные x вычисляются путем решения задачи SP, и найденное решение называется планом. Оценка эффективности плана x равна $Z(x) = fx + CY_x$.

Различие между релаксационными алгоритмами и алгоритмами разбиения заключается в характере задач RP и SP. Если в задаче RP имеются некоторые неопределенные булевы переменные и считается, что они непрерывные, то в задаче SP всем булевым переменным присвоено значение, так что единственные переменные, которые нужно определить, — это непрерывные переменные y_{ij} . Кроме того, задача RP может решаться в каждой

вершине дерева, тогда как задача SP решается только в конечных вершинах, т. е. в вершинах, для которых K пусто.

Прежде чем перейти к рассмотрению процесса решения задач RP и SP, целесообразно остановиться на методе решения задачи P2, предложенном в работе [4]. В алгоритме Бендерса требуется, чтобы декомпозиция задачи была произведена следующим образом (задача BP):

найти

$$\begin{array}{ll} \min (fx + & \min CY) \\ \text{при ограничениях} & \\ x \in X, & A_2 Y = b - A_1 x, \\ x_j = 0, 1, & Y \geq 0. \end{array}$$

Как можно заметить, та часть задачи BP, которая обведена пунктирной линией, является задачей SP. Двойственная к SP задача DSP может быть сформулирована следующим образом:

найти

$$\max u (b - A_1 x)$$

при ограничениях

$$u A_2 \leq C, \quad u \geq 0,$$

где u — вектор двойственных переменных. Если решение задачи BP существует и не ограничено, то она может быть преобразована к задаче BP1, которая может быть сформулирована следующим образом:

найти

$$\begin{array}{ll} \min fx + & \max u (b - A_1 x) \\ \text{при ограничениях} & \\ x \in X, & u A_2 \leq C, \\ x_j = 0, 1, & u \geq 0. \end{array}$$

Для обоснования этой процедуры, используемой при решении задачи P2, предположим, что для некоторого данного значения x решение задачи DSP не ограничено. Следовательно, существует двойственная угловая точка u^t и двойственный вектор допустимого направления u^r , такие, что $(u^t + \theta u^r) A_2 \leq C$ и $(u^t + \theta u^r) (b - A_1 x) \rightarrow \infty$ при $\theta \rightarrow \infty$. Отсюда следует, что

$$u^r (b - A_1 x) > 0.$$

Если решение задачи DSP не ограничено для некоторого x , то задача SP не имеет допустимого решения. Следовательно, x должен удовлетворять условию

$$u^r (b - A_1 x) \leq 0, \quad \forall r \in R,$$

где R — индексное множество направлений экстремальных лучей ограничения $uA_2 \leq C$.

Пусть T — индексное множество всех экстремальных точек для $uA_2 \leq C$. Тогда задачу BP1 можно сформулировать следующим образом (задача VIP):

найти

$$\min x_0$$

при ограничениях

$$x_0 \geq fx + u^t (b - A_1 x), \quad t \in T, \quad (32)$$

$$0 \geq u^r (b - A_1 x), \quad r \in R, \quad (33)$$

$$x \in X, \quad x_j = 0, 1, \quad \forall j.$$

Задача VIP является задачей смешанного целочисленного программирования с одной непрерывной переменной x_0 . При решении задачи VIP множества T и R первоначально пусты, т. е. экстремальные лучи и экстремальные точки задачи DSP еще не определены. Если выбрано некоторое x и задача DSP решена, то определяются либо некоторая экстремальная точка, либо экстремальный луч. Соответствующее ограничение (32) или (33) добавляется при формулировке задачи VIP, и эта задача решается вновь с целью получения нового значения x с помощью любого метода смешанного целочисленного программирования.

Подобная итерационная процедура поочередного решения задач VIP и DSP продолжается до тех пор, пока не будет получено некоторое x , удовлетворяющее условию

$$x_0 - fx = \max u (b - A_1 x)$$

при ограничениях

$$uA_2 \leq C, \quad u \geq 0.$$

Конечность этой процедуры очевидна, так как существует только конечное число экстремальных точек и экстремальных лучей. Более детально итерационный метод решения рассматривается в работе [38].

Возвращаясь к релаксационным алгоритмам, сформулируем задачу RP^v , возникающую из задач $P1$ и $P2$ в вершине v дерева поиска, следующим образом:

найти

$$\min \sum_{j \in K_1 \cup K_2} \sum_i c_{ij} y_{ij} + \sum_{j \in K_1} f_j + \sum_{j \in K_2} f_j x_j$$

при ограничениях

$$\sum_{j \in K_1 \cup K_2} y_{ij} = d_i, \quad \sum_i y_{ij} - Q_j x_j \leq 0, \quad j \in K_2,$$

$$\sum_i y_{ij} \leq Q_j, \quad j \in K_1, \quad y_{ij} \geq 0,$$

$$0 \leq x_j \leq 1, \quad j \in K_2, \quad x \in X.$$

Если имеются ограничения на взаимное расположение (дополнительные ограничения), то задачу RP^v следует решать как задачу линейного программирования. Однако, если дополнительных условий нет, то оптимальным будет решение уравнения вида

$$\sum_i y_{ij} - Q_j x_j = 0, \quad j \in K_2.$$

Выражая x_j через y_{ij} , получим

$$x_j = \frac{1}{Q_j} \sum_i y_{ij},$$

при этом $0 \leq x_j \leq 1$. Принимая во внимание полученное выражение для x_j , можно сформулировать задачу RP^v относительно только одних y_{ij} (задача $RP1^v$):

найти

$$\min \sum_{j \in K_1} \sum_i c_{ij} y_{ij} + \sum_{j \in K_2} \sum_i \left(c_{ij} + \frac{f_j}{Q_j} \right) y_{ij} + \sum_{j \in K_1} f_j$$

при ограничениях

$$\sum_{j \in K_1 \cup K_2} y_{ij} = d_i, \quad i = 1, \dots, m,$$

$$\sum_i y_{ij} \leq Q_j, \quad j \in K_1 \cup K_2,$$

$$y_{ij} \geq 0, \quad j \in K_1 \cup K_2, \quad i = 1, \dots, m.$$

(Авторы работ [15, 82] решили задачу $RP1^v$ как транспортную задачу.)

Если задача размещения производства не имеет дополнительных ограничений и мощность предприятий не ограничена, так что задача $P2$ сводится к задаче $P0$, то удобно переопределить

y_{ij} и c_{ij} , обозначая через y'_{ij} долю потребности i -го потребителя, удовлетворяемую предприятием, размещенным в j -м месте, и c'_{ij} — стоимость удовлетворения всей потребности i -го потребителя предприятием, размещаемым в j -м месте. Таким образом, если $y'_{ij} = y_{ij}/d_i$, $c'_{ij} = c_{ij} d_i$ и $Q_j = \sum_{i=1}^m d_i$, то задача RP1^v сводится к задаче RP2^v, которая может быть сформулирована следующим образом:

найти

$$\min \sum_{j \in K_1} \sum_{i=1}^m c'_{ij} y'_{ij} + \sum_{j \in K_2} \sum_{i=1}^m \left(c'_{ij} + \frac{f_j}{m} \right) y'_{ij}$$

при ограничениях

$$\begin{aligned} \sum_{j \in K_1 \cup K_2} y'_{ij} &= 1, \quad \forall i, \\ y'_{ij} &\geq 0, \quad \forall i, j \in K_1 \cup K_2. \end{aligned}$$

Определив новую величину g_i как

$$g_i = \begin{cases} 0, & \text{если } j \in K_1, \\ f_j, & \text{если } j \in K_2, \end{cases}$$

получим решение задачи RP2^v путем проверки.

Для каждого i выберем такое j^* , что

$$m c'_{ij^*} + g_{j^*} = \min (m c'_{ij} + g_j),$$

и пусть

$$y'_{ij} = \begin{cases} 1, & \text{если } j = j^*, \\ 0 & \text{в противном случае.} \end{cases}$$

Оптимальные значения x_j получаются тогда из соотношения

$$x_j = \frac{1}{m} \sum_{i=1}^m y'_{ij}.$$

В работе [18] впервые предложено применить описанную процедуру в алгоритме метода ветвей и границ для решения задачи P0. Впоследствии автор работ [55, 56] разработал ряд правил выбора вершин и правил ветвления, обеспечивающих вычислительную эффективность решения задачи P0.

Заметим, что если при использовании алгоритмов разбиения для решения задачи P2 определить полный план, например $x^{(h)}$, таким образом, что K_2 — множество свободных булевых пере-

менных — пусто, то задача SP сводится к SP (k), которая может быть сформулирована следующим образом:

найти

$$\min \sum_{j \in K_1} \sum_{i=1}^m c_{ij} y_{ij}$$

при ограничениях

$$\sum_{j \in K_1} y_{ij} = d_i, \quad i = 1, \dots, m,$$

$$\sum_{i=1}^m y_{ij} \leq Q_j, \quad j \in K_1,$$

$$y_{ij} \geq 0, \quad j \in K_1, \quad i = 1, \dots, m.$$

Кроме того, поскольку ограничения в задаче SP (k), те же, что и в задаче RP1^v, когда K_2 пусто, и дополнительные ограничения отсутствуют, то все замечания к задаче RP1^v и ее модификациям справедливы и для задачи SP (k).

Алгоритм неявного перебора для решения задачи с дополнительными условиями без ограничений на производительность приводится в работах [88, 89], алгоритм для задачи без ограничений на производительность, но с ограничениями на конфигурацию — в работах [20, 21], алгоритм для задачи с ограничениями на производительность, но без ограничений на конфигурацию — в работе [6].

6.2. Задачи о покрытии множества

Одним из вариантов задачи размещения производства является задача о покрытии множества, т. е. задача определения числа и мест размещения. Например, задача размещения складов, при котором расстояние от склада до каждого потребителя не превышает 100 км, может быть сформулирована как задача о покрытии множества. Другим примером задачи о покрытии множества является задача определения числа и размещения на территории штата студенческих общежитий, при котором каждый студент тратит на дорогу до колледжа не более одного часа, или задача размещения пожарных команд, при котором расстояние до любой точки города покрывается за 5 мин.

Задача P3 о покрытии множества может быть сформулирована следующим образом:

найти

$$\min z = \sum_{j=1}^n c_j x_j$$

при ограничении

$$\sum_{j=1}^n a_{ij}x_j \geq 1, \quad i = 1, \dots, m,$$

$$x_j = (0, 1), \quad j = 1, \dots, n.$$

Величины a_{ij} называются коэффициентами покрытия и принимают значения, равные единице, если потребитель находится в пределах j -й области (т. е. покрывается j -й областью), в противном случае a_{ij} равны нулю. Аналогично x_j принимает значение, равное единице, если в j -й области расположен некоторый объект, и равное нулю в противном случае. Ограничения в задаче РЗ требуют, чтобы каждый из m потребителей был «покрыт» по крайней мере одним из n объектов. Цель в этом случае состоит в том, чтобы «покрыть» потребителей с минимальными затратами, причем c_j — стоимость помещения объекта в j -ю область.

Поясним смысл термина «покрывать». Если имеется ряд жилых строений и решается задача размещения пожарных команд, то i -е жилое строение считается покрытым, если пожарная команда находится в пределах пяти минут езды от этого строения; аналогично, если существует i -й потребитель и речь идет о размещении заводов, один из которых должен удовлетворять потребности этого потребителя, то последний считается «покрытым», если завод расположен, например, в местах 1, 2 или 3. Таким образом, $a_{i1} + a_{i2} + a_{i3} = 1$ и все остальные a_{ik} равны нулю для $k \neq 1, 2, 3$.

Поскольку РЗ является также задачей целочисленного линейного программирования, то любой приемлемый метод решения задач целочисленного программирования может быть использован для решения задачи РЗ. Однако благодаря особой структуре задачи РЗ специально для ее решения был разработан ряд алгоритмов. Вообще говоря, в литературе описаны четыре подхода к решению задач о покрытии: методы неявного перебора, методы секущей плоскости, методы отсечения и эвристические методы. Общий обзор задачи о покрытии и ее применениях дан в работе [38].

Обычно задача о покрытии множества при решении проблемы размещения состоит в определении минимального количества объектов, необходимых для удовлетворения (покрытия) потребностей некоторого множества потребителей. В подобной ситуации РЗ сводится к так называемой задаче о полном покрытии Р4, которая получается из задачи РЗ, если положить все c_j равными единице. В [91] показано, что ряд задач размещения экстренных служб могут быть сформулированы как задачи о полном покрытии.

Помимо задачи о полном покрытии возможна постановка задачи о частичном покрытии. Если задача о полном покрытии состоит в определении минимального числа и мест расположения объектов, при котором удовлетворяются все потребители, то задача о частичном покрытии связана с определением размещения заданного числа объектов, при котором удовлетворяется максимальное число потребителей. Практически не всегда возможно обеспечить такое количество объектов, которое полностью удовлетворяло бы всех потребителей. Обычно количество имеющихся в распоряжении объектов достаточно только для частичного «покрытия» множества потребностей. В таких ситуациях целесообразно задачу размещения свести к задаче о частичном покрытии.

Математически задача о частичном покрытии Р5 может быть сформулирована следующим образом:

найти

$$\max \tilde{Z} = \sum_{i=1}^m \max a_{ij} x_j, \quad 1 \leq j \leq n,$$

при ограничениях

$$\sum_{j=1}^n x_j \leq K, \quad x_j = (0, 1), \quad j = 1, \dots, n,$$

где K — максимальное количество объектов, подлежащих размещению, величины a_{ij} и x_j те же, что и в задаче Р3.

Из вида функционала $\max a_{ij} x_j$, используемого в целевой функции в задаче Р5, следует, что если некоторое место расположения потребителя покрывается более чем одним размещаемым объектом, то при вычислении $\tilde{Z}$ учитывается только максимальная величина a_{ij} . Ограничение в задаче Р5 показывает, что в лучшем случае имеется K объектов для размещения. Заметим: если $\tilde{Z}$ равно m — числу потребителей, то это значит, что K достаточно велико, чтобы полностью удовлетворить всех потребителей. Таким образом, задача о полном покрытии (Р4) может быть сведена к задаче о частичном покрытии (Р5) для различных значений K и решение задачи Р5 с наименьшим значением величины K , для которого $\tilde{Z}$ равно m , будет оптимальным решением задачи Р4. Однако обычно задачу Р4 не решают таким образом, так как существуют более прямые эффективные методы ее решения.

В общем виде задача о частичном покрытии может быть сформулирована следующим образом:

найти

$$\min \tilde{Z} = \sum_{i=1}^m \min_{j \in \theta(x)} a_{ij}$$

при ограничениях

$$\sum_{j=1}^n x_j \leq K, \quad x_j = (0, 1), \quad j = 1, \dots, n,$$

где $\theta(x)$ — множество мест, в которых размещены объекты, $\theta(x) = \{j : x_j = 1\}$. Если a_i равно числу потребителей в i -й группе и d_{ij} есть расстояние между i -й областью и местом размещения j -го объекта, то a_{ij} в задаче Р6 равно $a_i d_{ij}$. Иными словами, задача состоит в размещении самое большое K объектов таким образом, чтобы суммарное расстояние между группами потребителей и объектами было минимальным. Из выражения для a_{ij} следует, что если некоторая группа может обслуживаться более чем одним объектом, то она будет обслуживаться самым близким к ней объектом (заметим, что множество $\theta(x)$ не должно быть пустым).

В зависимости от исходных данных задача Р6 может быть центральной задачей размещения объектов [78], задачей p -медианы [47] или задачей размещения складов [85]. В первом из перечисленных вариантов каждое место размещения потребителей рассматривается как потенциальное место размещения объекта, т. е. предполагается, что n равно m .

Точное решение задачи Р6 может быть получено с помощью методов динамического программирования, метода ветвей и границ и двойственных методов. Однако они не приемлемы с вычислительной точки зрения при $n \geq 20$ и $K \geq 10$, и поэтому для решения задачи Р6 был разработан ряд эвристических методов [54, 68, 85, 90].

ЗАКЛЮЧЕНИЕ

Ввиду огромного количества работ по исследованию задач размещения и планировки объектов невозможно дать исчерпывающее изложение этого вопроса в пределах одной главы. В связи с этим мы ограничились тем, что продемонстрировали постановку различных задач размещения и планировки, чтобы проиллюстрировать не только многообразие возможных вариантов такой постановки, но и ту роль, которую играют методы исследования операций для их решения.

До настоящего времени основное внимание уделялось либо новой постановке задач размещения и планировки объектов, либо разработке методов решения уже сформулированных задач. Хотя многие исследователи систем занимаются решением практических задач размещения и планировки объектов с помощью методов исследования операций, тем не менее результаты подобных работ почти не публикуются. Примеры применения количественных методов, представленных в данной главе, при размещении и/или планировке заводов, складов, пожарных частей, больниц, отделений банков, почтовых отделений и кладовых, а также более подробное изложение состоящих данной проблемы можно найти, например, в работе [35].

В заключение отметим, что в течение последнего десятилетия достигнут существенный прогресс в разработке аналитических моделей задач размещения и планировки объектов, а также в разработке эффективных методов решения многих из них.

ЛИТЕРАТУРА

1. Apple J. M., Deisenroth M. P., A Computerized Plant Layout Analysis and Evaluation Technique (PLANET), Proceedings, American Institute of Industrial Engineering, 23rd Annual Conference and Convention, Anaheim, California, 1972.
2. Armour G. C., Buffa E. S., A Heuristic Algorithm and Simulation Approach to the Relative Location of Facilities, *Management Sci.*, 9, No. 2, 294—309 (1963).
3. Balinski M. L., On Finding Integer Solution to Linear Programs, *Mathematica*, Princeton, N.J., 1964.
4. Benders J. F., Partitioning Procedures for Solving Mixed-Variables Programming Problems, *Numerische Mathematik*, 4, 238—252 (1962).
5. Bindschedler A. E., Moore J. M., Optimal Location of New Machines in Existing Plant Layouts, *J. Ind. Eng.*, 12, No. 1, 41—48 (1961).
6. Bulfin R. L., Unger V. E., Investigation of Surrogate Constraints in a Plant Location Algorithm, Georgia Institute of Technology, Atlanta, Georgia, 1973.
7. Cabot A. V., Francis R. L., Stary M. A., A Network Flow Solution to a Rectilinear Distance Facility Location Problem, *AIIE Trans.*, 2, No. 2, 132—141 (1970).
8. Christofides N., Viola P., The Optimum Location of Multi-Centres on a Graph, *Operational Res. Quart.*, 22, No. 2, 145—154 (1971).
9. Cooper L., Location-Allocation Problems, *Operations Res.*, 11, No. 3, 331—334 (1963).
10. Cooper L., Heuristic Methods for Location-Allocation Problems, *SIAM Rev.*, 6, No. 1, 37—52 (1964).
11. Cooper L., Solution of Generalized Locational Equilibrium Models, *J. Regional Sci.*, 7, No. 1, 1—18 (1967).
12. Cooper L., The Transportation-Location Problem, *Operations Res.*, 20, No. 1, 94—108 (1972).
13. Corley H. W., Jr., Roberts S. D., A Partitioning Problem with Applications in Regional Design, *Operations Res.*, 2, No. 5, 1010—1019 (1972).
14. Curry G. L., Skeith R. W., A Dynamic Programming Algorithm for Facility Location and Allocation, *AIIE Trans.*, 1, No. 2, 133—138 (1969).
15. Davis P. S., Ray T. L., A Branch-Bound Algorithm for the Capacitated Facilities Location Problem, *Nav. Res. Log. Quart.*, 16, No. 3, 331 (1969).
16. Dearing P. M., Francis R. L., A Network Flow Solution to a Multi-Facility Minimax Location Problem Involving Rectilinear Distances, *Transport. Sci.*, 8, No. 2, 126—141 (1974).
17. Donath W. E., Statistical Properties of the Placement of a Graph, *SIAM J. Appl. Math.*, 16, No. 2, 439—457 (1968).
18. Eifroymson M. A., Ray T. L., A Branch-Bound Algorithm for Plant Location, *Operations Res.*, 14, No. 3, 361—368 (1966).
19. Eilon S., Watson-Gandy C. D. T., Christofides N., Distribution Management: Mathematical Modeling and Practical Analysis. Hafner Publishing Company, N.Y., 1971.
20. Ellwein L. B., Gray P., Solving Fixed Charge Allocation Problems with Capacity and Configuration Constraints, *AIIE Trans.*, 3, No. 4, 290—298 (1971).

21. Elshafei A. N., Optimal Location of Facilities, Ph. D. Thesis, University of Birmingham, England, 1973.
22. Elshafei A. N., On Solving the Capacitated Facilities Location Problem with Concave Cost Functions—Part I: Features of Search Procedure, OR Report No. 92, North Carolina State University at Raleigh, Raleigh North Carolina, 1974.
23. Elshafei A. N., On Solving the Capacitated Facilities Location Problem with Concave Cost Functions—Part II: Algorithmic Aspects and Computational Experience, OR Report No. 93, North Carolina State University at Raleigh, Raleigh North Carolina, 1974.
24. Elzinga J., Hearn D., Geometrical Solutions for Some Minimax Location Problems, *Transport. Sci.*, 6, No. 4, 379—394 (1972).
25. Elzinga J., Hearn D., The Minimum Covering Sphere Problem, *Management Sci.*, 19, No. 1, 96—104 (1972).
26. Elzinga J., Hearn D., A Note on Minimax Location Problem, *Transport. Sci.*, 7, No. 1 (1973).
27. Eyster J. W., White J. A., Wierwille W. W., On Solving Multifacility Location Problems Using Hyperboloid Approximation Procedure, *AIIE Trans.*, 5, No. 1, 1—6 (1973).
28. Francis R. L., A Note on the Optimum Location of Machines in Existing Plant Layouts., *J. Ind. Eng.*, 14, No. 1, 57—59 (1963).
29. Francis R. L., Sufficient Conditions for Some Optimum-Property Facility Design, *Operations Res.*, 15, No. 3, 448—446 (1967).
30. Francis R. L., On Some Problem of Rectangular Warehouse Design and Layout, *J. Ind. Eng.*, 18, No. 10, 595—604 (1967).
31. Francis R. L., Some Aspects of a Minimax Location Problem, *Operations Res.*, 15, No. 6, 1163—1168 (1967).
32. Francis R. L., A Minimax Facility Configuration Problem Involving Lattice Points, *Operations Res.*, 21, No. 2, 101—111 (1973).
33. Francis R. L., Cabot A. V., Properties of a Multifacility Location Problem Involving Euclidean Distances, *Nav. Res. Log. Quart.*, 19, No. 2, 335—353 (1972).
34. Francis R. L., Goldstein J. M., Location Theory: a Selective Bibliography; *Operation Res.*, 22 (1974).
35. Francis R. L., White J. A., Facilities Layout and Location: an Analytical Approach; Prentice-Hall, Inc., Englewood Cliffs, N.J., 1974.
36. Frank H., Optimum Location on a Graph with Probabilistic Demand, *Operations Res.*, 14, No. 3, 409—421 (1966).
37. Gavett J. W., Plyter N. V., The Optimal Assignment of Facilities to Locations by Branch and Bound, *Operations Res.*, 14, No. 2, 210—232 (1966).
38. Garfinkel R. S., Nemhauser G. L., Integer Programming, Wiley, N.Y., 1972.
39. Gilbert E. N., Pollack H. O., Steiner Minimal Trees, *J. Soc. Ind. and Appl. Math.*, 16, No. 1, 1—29 (1968).
40. Gilmore P. C., Optimal and Suboptimal Algorithms for the Quadratic Assignment Problem, *SIAM Journal*, 10, No. 2, 305—313 (1962).
41. Goldman A. J., Optimal Locations for Centers in a Network, *Transport. Sci.*, 3, No. 4, 352—360 (1969).
42. Goldman A. J., Optimal Center Location in Simple Networks, *Transport. Sci.*, 5, No. 2, 212—221 (1971).
43. Goldman A. J., Approximation Localization Theorems for Optimal Facility Placement, *Transport. Sci.*, 6, No. 2, 195—201 (1972).
44. Goldman A. J., Minimax Location of a Facility on a Network, *Transport. Sci.*, 6, No. 4, 407—418 (1972).
45. Goldman A. J., Witzgall C. J., A Localization Theorem for Optimal Facility Placement, *Transport Sci.*, 4, 406—408 (1970).

46. Hakimi S. L., Optimum Locations of Switching Centers and the Absolute Centers and Medians of Graph, *Operations Res.*, 12, No. 3, 450—459 (1964).
47. Hakimi S. L., Optimum Distribution of Switching Centers in a Communication Network and Some Related Graph Theoretic Problems, *Operations Res.*, 13, No. 3, 462—475 (1965).
48. Hakimi S. L., Steiner's Problem in Graphs and Its Implications, *Network*, 1, No. 2, 113—134 (1972).
49. Hakimi S. L., Maheshwari S. N., Optimum Locations of Centers in Networks, *Operations Res.*, 20, No. 5, 967—973 (1972).
50. Hanan M., Kutzberg J., A Review of the Placement and Quadratic Assignment Problem, *SIAM Rev.*, 14, No. 2, 324—342 (1972).
51. Hiller F. S., Connors M. M., Quadratic Assignment Problem Algorithms and the Location of Indivisible Facilities, *Management Sci.*, 13, No. 1, 43—57 (1966).
52. Järvinen P., Rajala J., Sinerro J., A Branch-and-Bound Algorithm for Seeking the p-median, *Operations Res.*, 20, No. 1, 173—178 (1972).
53. Katz N., On the Convergence of a Numerical Scheme For Solving Some Locational Equilibrium Problems, *SIAM J. Appl. Math.*, 17, No. 6, 1224—1231 (1969).
54. Khumawala B. M., Whybark D. C., A Comparison of Some Recent Warehouse Location Techniques, *Log. Rev.*, 7, 3—19 (1971).
55. Khumawala B. M., An Efficient Branch and Bound Algorithm for the Warehouse Location Problem, *Management Sci.*, 18, No. 12, 718—734 (1972).
56. Khumawala B. M., An Efficient Heuristic Procedure for the Uncapacitated Warehouse Location Problem, *Nav. Res. Log. Quart.*, 20, No. 1, 109—122 (1973).
57. Khumawala B. M., An Efficient Algorithm for the p-median Problem with Maximum Distance Constraints, *Geograph. Anal.*, 5, No. 4, 309—321 (1974).
58. Koopmans J. C., Beckman M., Assignment Problems and the Location of Economic Activities, *Econometrica*, 25, No. 1, 53—76 (1957).
59. Kuenne R. E., Soland R. M., The Multisource Weber Problem: Exact Solutions by Branch and Bound, IDA Economic Papers, H. Williams (Ed.), Program Analysis Division, 400 Army-Navy Drive, Arlington, Virginia, 1971.
60. Kuhn H. W., Kuenne R. E., An Efficient Algorithm for the Numerical Solution of the Generalized Weber Problem in Spatial Economics, *J. Regional Sci.*, 4, No. 2, 21—33 (1962).
61. Kuhn H. W., A Note on Fermat's Problem, *Math. Prog.*, 4, No. 1, 98—107 (1973).
62. Lawler E. L., The Quadratic Assignment Problem, *Management Sci.*, 9, No. 4, 586—599 (1963).
63. Lee R. C., Moore J. M., CORELAP—Computerized Relationship Layout Planning, *J. Ind. Eng.*, 18, No. 3, 195—200 (1967).
64. Levy J., An Extended Theorem for Location on a Network, *Operational Res. Quart.*, 18, No. 4, 433—442 (1967).
65. Love R. F., Locating Facilities in Three-Dimensional Space by Convex Programming, *Nav. Res. Log. Quart.*, 16, No. 4, 503—516 (1969).
66. Love R. F., A Computational Procedure for Optimally Locating a Facility with Respect to Several Rectangular Regions, *J. Regional Sci.*, 12, No. 2, 233—242 (1972).
67. Mallett A. J., Francis R. L., A Generalized Assignment Approach to Optimal Facility Layout, *AIIE Trans.*, 4, No. 2, 144—147 (1972).
68. Maranzana F. E., On the Location of Supply Points to Minimize Transport Costs, *Operational Res. Quart.*, 15, No. 3, 261—270 (1964).
69. Marsten R. E., An Algorithm for Estimating the Medians of a Weighted Graph Subject to Side Constraints and An Application to Rural Hospitals

- in British Columbia, Unpublished Ph. D. Thesis, University of British Columbia, Vancouver, 1971.
70. McGinnis L. F., A Survey of Recent Results for a Class of Facilities Location Problems, *AIIE Trans.*, 9, No. 1, 11—18 (1977).
 71. Miehele W., Link-Length Minimization in Networks, *J. Operations Res.*, 6, No. 2, 232—243 (1958).
 72. Minioka E., The m-Center Problem, *SIAM Rev.*, 12, No. 1, 138—139 (1970).
 73. Moore J. M., Optimal Locations for Multiple Machines, *J. Ind. Eng.*, 12, No. 5, 34—38 (1961).
 74. Moore J. M., Computer Program Evaluates Plant Layout Alternatives, *Ind. Eng.*, 3, No. 8, 19—25 (1971).
 75. Nair, K. P. K., Chandrasakaran R., Optimal Location of a Single Service Center of Certain Types, *Nav. Res. Log. Quart.*, 18, 503—510 (1971).
 76. Nugent C. E., Vollmann T. E., Ruml J., An Experimental Comparison of Techniques for the Assignment of Facilities to Locations, *Operations Res.*, 16, No. 1, 150—173 (1968).
 77. Pierce J. F., Crowston W. B., Tree-Search Algorithms for Quadratic Assignment Problems, *Nav. Res. Log. Quart.*, 18, No. 1, 1—36 (1971).
 78. Revelle C., Marks D., Liebman J. C., An Analysis of Private and Public Sector Location Models, *Management Sci.*, 16, No. 11, 692—707 (1970).
 79. Revelle C., Swain R., Central Facilities Location, *Geograph. Anal.*, 2, No. 1, 30—42 (1970).
 80. Rojeski P., Revelle C., Central Facilities Location under an Investment Constraint, *Geograph. Anal.*, 2, No. 3, 343—350 (1970).
 81. Roth R., Computer Solutions to Minimum-Cover Problems, *Operations Res.*, 17, No. 3, 455—465 (1969).
 82. Sa G., Branch-and Bound and Approximate Solutions to the Capacitated Plant Location Problem, *Operations Res.*, 17, No. 6, 1005 (1969).
 83. Scott A. J., Location-Allocation Systems: A Review, *Geograph. Anal.*, 2, No. 2, 95—119 (1970).
 84. Seehof J. M., Evans W. O., Automated Layout Design Program, *J. Ind. Eng.*, 18, No. 12, 690—695 (1967).
 85. Shannon R. E., Ignizio J. P., A Heuristic Programming Algorithm for Warehouse Location, *AIIE Trans.*, 2, No. 4, 334—339 (1970).
 86. Sherali A. D., Shetty C. M., The Rectilinear Distance Location-Allocation Problem, *AIIE Trans.*, 9, No. 2, 136—143 (1977).
 87. Soland R. M., Optimal Plant Location with Concave Costs, presented to the 39th National Meeting of the Operations Research Society of America, Dallas, Texas, 1971.
 88. Spielberg K., Algorithms for the Simple Plant-Location with Some Side Conditions, *Operations Res.*, 17, No. 1, 85—111 (1969).
 89. Spielberg K., Plant Location with Generalized Search Origin, *Management Sci.*, 16, No. 3, 165—178 (1969).
 90. Teitz M. B., Bart P., Heuristic Methods for Estimating the Generalized Vertex Median of a Weighted Graph, *Operations Res.*, 16, No. 5, 901—1092 (1968).
 91. Toregas V., Swain R., Revelle C., Bergman L., The Location of Emergency Service Facilities, *Operations Res.*, 19, No. 6, 1363—1373 (1971).
 92. Vergin R. C., Rogers J. D., An Algorithm and Computational Procedure for Locating Economic Facilities, *Management Sci.*, 13, No. 6, 3240—3254 (1967).
 93. Vollman T. E., Nugent C. E., Zartler R. L., A Computerized Model for Office Layout, *J. Ind. Eng.*, 19, No. 7, 324—327 (1968).
 94. Wendell R. E., Hurter A. P. Jr., Optimal Locations on a Network, *Transport. Sci.*, 6, 18—33 (1972).
 95. Wendell R. E., Location Theory, Dominance and Convexity, *Operations Res.*, 21, No. 1, 314—320 (1973).

96. Wesolowsky G. O., Location in the Continuous Space, *Geograph. Anal.*, 5, No. 2, 95—112 (1972).
97. Wesolowsky G. O., Rectangular Distance Location under the Minimax Optimality Criterion, *Transport Sci.*, 9, No. 2, 103—113 (1972).
98. Wesolowsky G. O., Love R. F., The Optimal Location of New Facilities using Rectangular Distances, *Operations Res.*, 19, No. 1, 124—130 (1971).
99. Wesolowsky G. O., Location of Facilities with Rectangular Distances Among Point and Area Destinations, *Nav. Res. Log. Quart.*, 18, No. 1, 83—90 (1971).
100. Wesolowsky G. O., A Nonlinear Approximation Method for Solving a Generalized Rectangular Distance Weber Problem *Management Sci.*, 18, No. 11, 656—663. (1972).
101. White J.A., Case K.E. On Covering Problems and the Central Facilities Location Problem, *Geograph. Anal.*, 6, No. 3, 281—293 (1974).

I-7. Календарное планирование и упорядочение работ

М. Сальвадор¹⁾

ВВЕДЕНИЕ

Термины «календарное планирование» и «упорядочение работ» широко используются при решении задач методами исследования операций. Строго говоря, упорядочение — это планирование только порядка выполнения некоторого набора работ или заданий, тогда как календарное планирование — это, по существу, планирование сроков выполнения работ, т. е. установление периода времени (или момента времени), в течение которого (или к которому) должна быть выполнена данная работа.

В условиях полной определенности календарное планирование и упорядочение часто взаимозаменяемы, так как все параметры системы известны априори и работы начинают выполняться, как только это становится возможным. В результате порядок работ определяется графиком их выполнения, и наоборот.

В тех случаях, когда в отношении параметров системы существует какая-то неопределенность (например, относительно времени, необходимого для выполнения некоторой работы), то календарное планирование неизбежно сводится к установлению последовательности работ.

Несмотря на указанное различие, общепринято многие задачи, рассматриваемые в данной главе, считать задачами «календарного планирования», хотя, строго говоря, они являются задачами «упорядочения работ». Подобная терминологическая непоследовательность свойственна научно-технической литературе, посвященной вопросам теории и методам календарного планирования и упорядочения. Чтобы избежать влияния подобной терминологической путаницы, необходимо четко представлять, в чем состоит принципиальное различие календарного планирования и упорядочения работ.

1. КЛАССИФИКАЦИЯ ЗАДАЧ КАЛЕНДАРНОГО ПЛАНИРОВАНИЯ

Основная трудность разработки единой схемы классификации (или таксономии) задач календарного планирования обусловлена их разнообразием. Поэтому необходимо не только определить логически обоснованные рамки, в пределах которых можно было бы осуществить основную классификацию задач, но и использовать соответствующую терминологию.

1.1. Процесс производства по специальным заказам

Основой классификации задач календарного планирования является представление о производственном участке, выполняющем работы по заказам. Многие задачи календарного планирова-

¹⁾ Michael S. Salvador, Ph. D., Manager, Ernst & Ernst in Charlotte, North Carolina.

ния, методы решения которых уже известны, являются, по существу, частными случаями извечной задачи составления графика (расписания) выполнения работ производственным участком. Впервые процесс выполнения работ производственным участком был всесторонне проанализирован в работе [20], основные положения и выводы которой приводятся ниже.

Фундаментальным элементом процесса выполнения работ производственным участком является операция. Действительно, любая работа может быть представлена как конечный набор необходимых операций. При этом каждая операция должна выполняться определенной машиной и за определенное время (время обслуживания), специфическое для данной работы и машины.

Производственный участок — это машины, которые предназначены для выполнения заданного множества работ, или, что то же самое, множества операций. Понятие процесса выполнения работ производственным участком включает машины, работы, операции и формулировку физических ограничений, определяющих, какие операции могут быть выполнены каждой машиной.

Обобщенная задача планирования графика выполнения работ производственным участком заключается в нахождении порядка работ, при котором достигается минимизация или максимизация некоторой измеримой функции такого порядка при условии, что операции выполняются каждой машиной производственного участка в соответствии с техническими требованиями, предъявляемыми к процессу выполнения работ на данном участке (равно как и любыми другими ограничениями). В любом случае для решения задачи необходимо выполнение по крайней мере следующих условий:

- каждая машина эксплуатируется в течение непрерывного периода времени, т. е. исключается остановка или поломка машины;

- последовательность операций строго упорядочена, т. е. существует, самое большее, еще только одна операция как непосредственно до, так и непосредственно после данной операции;

- каждая операция может выполняться только машиной одного типа из имеющихся на участке;

- на участке имеется только одна машина данного типа;

- процесс выполнения операции некоторой машиной должен быть непрерывным, т. е. следующая операция на этой же машине может выполняться только после полного завершения предыдущей;

- любая работа в каждый момент времени может выполняться только одной машиной;

- каждая машина в данный момент времени выполняет не более одной операции.

На практике любое из этих условий (ограничений) может не выполняться, и поэтому большая часть аналитических результатов теории календарного планирования отражает идеальную ситуацию.

Помимо перечисленных ранее показателями процесса выполнения работ производственным участком являются также любые дополнительные характеристики работ и операций (например, количество работ, способ их поступления на выполнение); любые дополнительные характеристики машин (например, их количество); дополнительные ограничения на порядок выполнения операций машинами; критерий оценки результирующего расписания. Именно в соответствии с данными показателями и проводятся классификация задач календарного планирования и разработка методов их решения. При этом предполагается, что указанные условия (или ограничения) имеют место во всех рассматриваемых случаях.

Следует иметь в виду, что понятие «процесс выполнения работ производственным участком» не ограничивает наше рассмотрение планированием деятельности только промышленных предприятий, выполняющих производственные заказы, или деятельности предприятий бытового обслуживания. Задача календарного планирования работ может быть сформулирована и для системы уголовного судопроизводства. В этом случае «работа» — это «полная процедура обработки» дел обвиняемых, в качестве операций могут выступать предварительное заключение, судебное разбирательство, освобождение на поруки, условное освобождение и т. д., а роль «машин» отводится следственным или обычным тюрьмам, судам или судьям, поручителям, советам по опеке и т. д.

1.2. Схема классификации задач календарного планирования

В зависимости от характера показателей процесса выполнения работ производственным участком задачи календарного планирования могут быть разделены на детерминированные и стохастические (рис. 1). В первом случае все параметры системы (число работ, моменты их готовности к выполнению, длительность выполнения на машинах, технологические маршруты и т. д.) и все показатели производственного участка (число машин, их доступность и т. д.) постоянны и заранее известны. Во втором случае каждый параметр системы и показатель производственного участка могут меняться случайным образом.

При использовании нескольких машин можно выделить следующие основные задачи календарного планирования:

— все работы требуют выполнения только одной операции,

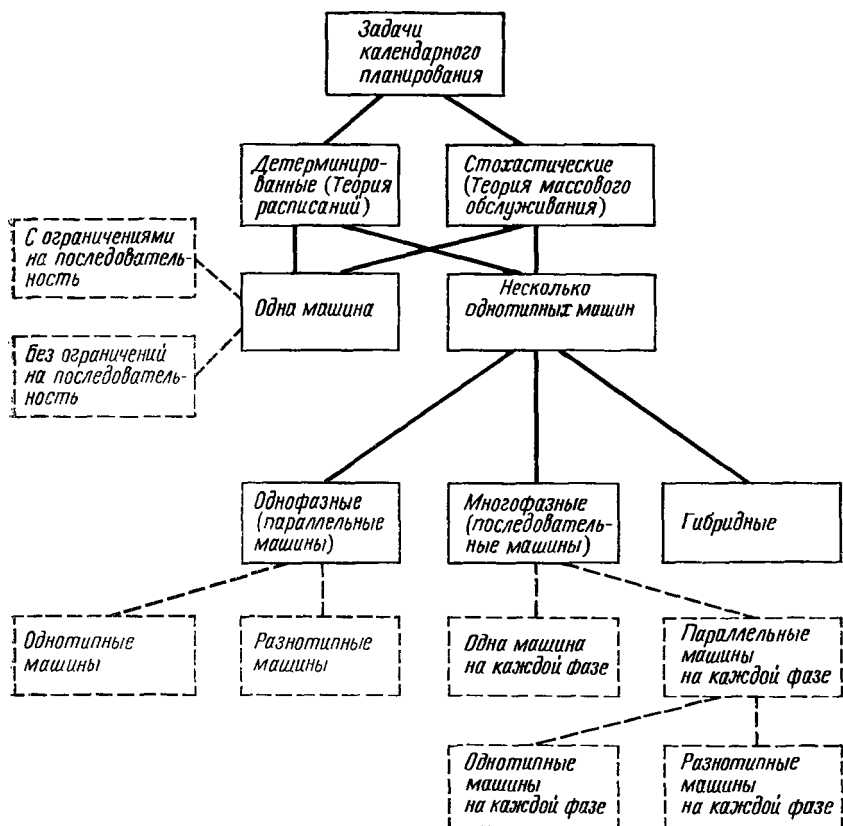


Рис. 1. Схема классификации задач календарного планирования.

которая может быть выполнена любой из имеющихся машин (главным образом однофазная система параллельно работающих машин);

— все работы состоят из одних и тех же операций, последовательность которых строго фиксирована для всех работ (многофазная система, или конвейер);

— все прочие (так называемые гибридные системы). Последующая классификация задач календарного планирования осуществляется в соответствии с используемым критерием оценки графика работ. Планирование производственного процесса при наличии одной машины имеет ряд особенностей и будет рассмотрено позже.

2. НЕКОТОРЫЕ ОБЩИЕ КРИТЕРИИ ОЦЕНКИ ГРАФИКА РАБОТ

При решении любой задачи календарного планирования фактором более критичным, чем ограничения, накладываемые на процесс выполнения работ производственным участком, является способ (критерий) оценки выбранного решения (в данном случае — графика или расписания). Тем не менее обычно в первую очередь учитываются ограничения, т. е. возможность реализации предполагаемого расписания, а не его эффективность с точки зрения критерия оценки. (Следует отметить, что возможность реализации расписания часто более необходима, чем оценка его эффективности.)

Критерий оценки эффективности графика, как правило, выражается через показатели работы, значения которых определяются конкретным графиком или дисциплиной обслуживания. К числу показателей, характеризующих работу, относятся следующие:

- время завершения, т. е. момент времени, к которому должна быть завершена последняя операция;
- время прохождения, т. е. общее время, в течение которого работа «находится» в пределах производственного участка;
- время ожидания, т. е. часть времени прохождения, в течение которого работа «ожидает» выполнения операций;
- время задержки, т. е. разность между фактическим и плановым сроками завершения работы;
- время запаздывания, т. е. время, эквивалентное времени задержки, если оно положительное, и равное нулю в противном случае.

Из показателей, характеризующих производственный участок, рассматриваются, как правило, только два, а именно коэффициент использования (или загрузки) машин, т. е. отношение времени, в течение которого машина занята выполнением работы, к общему времени, в течение которого эта машина может быть использована (это время обычно предполагается равным максимальному времени выполнения работ), и незавершенное производство, т. е. объем работ, процесс выполнения которых еще не завершен (эта величина может измеряться различными способами, в частности числом работ, выполняемых производственным участком, и общей продолжительностью выполнения всех работ на участке).

Независимо от того, включает критерий оценки расписания один или несколько из перечисленных показателей, конкретная задача состоит в минимизации или максимизации среднего, общего, максимального или минимального значения этих показателей, обусловленного расписанием. Наиболее распространенной задачей является минимизация длительности производственного цикла, т. е. минимизация максимального времени прохождения работ. Другая, не менее распространенная задача — это минимизация

времени запаздывания, так как большинство реальных задач календарного планирования имеет дело с работами, сроки выполнения которых строго фиксированы, и за нарушение этих сроков (или за задержку) налагается штраф. Если основной интерес представляют затраты, связанные с данным расписанием, то задача сводится к минимизации таких затрат. В этом случае для оценки эффективности расписания каждому показателю приписываются постоянные удельные затраты, и критерий эффективности, за редким исключением, выражается как линейная сумма таких оценок. (Например, при решении задач минимизации незавершенного производства, максимизации использования машин и минимизации общих задержек в качестве критерия оценки используются затраты, которые в этих случаях прямо пропорциональны перечисленным выше показателям.)

Можно показать, что между различными критериями оценки графика работ существует определенная зависимость [29]. Так, например, среднее время прохождения работы прямо пропорционально среднему объему незавершенного производства, измеряемому числом работ; среднее время прохождения работы отличается от среднего времени запаздывания и среднего времени ожидания на постоянную величину. Расписание, удовлетворительное с точки зрения среднего времени прохождения одной работы, среднего времени доступности работы (равного разности между плановым сроком и временем ее поступления на производственный участок, которое в статическом случае равно нулю), числа машин или среднего коэффициента их использования является одновременно удовлетворительным и с точки зрения других показателей. Расписание, согласно которому минимизируется среднее время прохождения работы, обеспечивает также минимизацию среднего времени запаздывания, среднего времени ожидания и среднего объема незавершенного производства (измеряемого числом работ). Однако подобная взаимосвязь не всегда очевидна. Например, расписание, согласно которому минимизируется среднее время прохождения работы, обеспечивает и минимизацию среднего времени запаздывания, а расписание, согласно которому минимизируется максимальное время прохождения работ, не обязательно обеспечивает минимизацию максимального времени запаздывания.

3. ДЕТЕРМИНИРОВАННЫЕ ЗАДАЧИ КАЛЕНДАРНОГО ПЛАНИРОВАНИЯ

Как уже отмечалось, основное различие между детерминированными и стохастическими задачами календарного планирования состоит в том, что в первом случае параметры системы считаются известными. В результате большинство детерминированных задач календарного планирования сводится к задачам упорядочения.

Это означает, что если мы имеем дело с задачами подобного типа, то достаточно знать момент начала первой операции первой из работ и последовательность выполнения операций каждой машиной. Если предположить, что каждая операция начинает выполняться тотчас, как только это становится возможным, то расписание работ может быть определено однозначно. Именно по этой причине детерминированные задачи календарного планирования называются в литературных источниках детерминированными задачами упорядочения.

В данной главе приводятся некоторые наиболее важные результаты анализа детерминированных задач календарного планирования и дается краткий перечень наиболее интересных, с точки зрения автора, работ по календарному планированию и упорядочению.

3.1. Планирование работ для одной машины

На первый взгляд подобные задачи слишком элементарны. Тем не менее их изучение представляет определенный интерес, особенно если принять во внимание следующие факторы. Во-первых, многие функционирующие системы могут быть представлены с помощью одомашиной модели. Во-вторых, всегда существует вероятность того, что результаты анализа простой задачи могут оказаться полезными при решении более общих, сложных задач календарного планирования.

Следует иметь в виду, что при решении задач упорядочения чаще всего предполагается, что все подлежащие упорядочению работы поступают в систему одновременно. Подобное предположение оказывается оправданным и в том случае, когда одна машина выполняет работы, время поступления которых в систему различно. Как показали авторы работы [20], в этом случае требуется рассматривать прерывания операций или простои. Однако, если прерванные операции могут быть возобновлены с момента прерывания, то задача упорядочения, по существу, сводится к задаче упорядочения для случая одновременного поступления всех работ.

Решения некоторых детерминированных задач упорядочения работ, выполняемых одной машиной, довольно просты: выбор оптимального порядка работ сводится к упорядочению работ по некоторому набору значений показателей этих работ. Перечислим наиболее важные выводы, полученные в результате анализа таких задач.

Если перечень работ составлен в порядке неубывания времени выполнения каждой работы, то минимизируются такие показатели, как среднее время задержки работы, среднее время ожидания, минимальное время прохождения работы, минимальное время

завершения работы, минимальное время ожидания и средний объем незавершенного производства. Подобная дисциплина упорядочения обычно называется правилом «наименьшего времени обслуживания».

Если перечень работ составлен в порядке невозрастания установленных сроков выполнения каждой работы, то минимизируются такие параметры, как максимальное время задержки и максимальное время запаздывания.

Если каждой работе приписывается весовой множитель и перечень работ составлен в порядке неубывания отношения времени выполнения каждой работы к ее весовому множителю, то минимизируются такие показатели, как взвешенное среднее время прохождения, взвешенное среднее время запаздывания и взвешенное среднее время ожидания.

Естественно, что возможны и более сложные задачи упорядочения работ, выполняемых одной машиной. В частности, такая задача возникает при выполнении смежных работ. Если выполнение смежных работ требует переналадки оборудования, то для этого необходимо вполне определенное время (время простоя). Поскольку время простоя будет зависеть не только от характера работы, но и от ее места в перечне работ, то, следовательно, критерий оценки расписания будет функцией не одной переменной. В этом случае главной целью упорядочения является минимизация общего времени простоя (переналадок). Для решения этой задачи, которая, по существу, является вариантом известной «задачи коммивояжера», был разработан целый ряд методов, таких, как динамическое программирование [40, 59] и метод ветвей и границ [76]. Авторы работ [74, 75] для решения одной частной задачи создали эффективный алгоритм неявного перебора, основанный на построении дерева решений.

Ниже перечисляются наиболее интересные, с точки зрения автора, задачи календарного планирования для одной машины и указываются дополнительные ограничения, накладываемые на процесс выполнения работ производственным участком, а также приводятся ссылки на соответствующие литературные источники.

Минимизация времени прохождения при заданных несвязанных фиксированных последовательностях работ (называемых «цепочками») [102].

Минимизация времени прохождения при ограничениях на последовательность выполнения работ [20].

Минимизация общего штрафа при выполнении «взаимосвязанных» работ [29].

Минимизация общих затрат на обработку, если машина характеризуется одним параметром состояния [48].

Минимизация общих потерь при невыполнении работ в установленные сроки при условии, что стоимость задержки есть произ-

вольная неубывающая линейная функция времени задержки для данной работы [29, 82, 96].

Минимизация числа переналадок при условии, что работами являются партии продукции, размеры которых определяются графиком спроса [49].

Минимизация общих потерь при невыполнении работ в установленные сроки при условии, что эти потери являются непрерывной, ограниченной, неубывающей функцией времени выполнения работы [75].

Минимизация максимальных потерь при невыполнении работ в установленные сроки [86].

Минимизация общих затрат на производство и хранение запасов при условии, что работы представляют собой партии продукции, размер которых определяется графиком спроса [8, 11, 15, 23, 24, 27, 36, 78, 79, 91].

Минимизация числа работ, которые не выполняются в установленные сроки [86].

Минимизация общего времени задержки работ [100].

Минимизация взвешенного общего времени задержки и взвешенного общего времени прохождения работ [45].

Вариантами задач упорядочения работ, выполняемых одной машиной, являются вырожденные случаи однофазной системы параллельно работающих машин, а также многофазной системы машин. Поэтому любой новый метод решения, который будет приведен в последующих разделах настоящей главы, может рассматриваться как еще один, дополнительный метод решения задачи упорядочения работ, выполняемых одной машиной.

3.2. Однофазная система параллельно работающих машин

Обобщенной задачей упорядочения работ, выполняемых одной машиной, является случай, когда все работы состоят только из одной операции, но имеется несколько машин, которые могут выполнять каждую работу. При решении этой задачи предполагается, что все работы поступают в систему одновременно и выполняются перечисленные ранее условия, за исключением, естественно, условия, согласно которому на производственном участке имеется только одна машина данного типа.

Наиболее простой вариант рассматриваемой задачи, т. е. использование идентичных машин, был проанализирован в работе [82], автор которой показал, что в данном случае упорядочение может быть осуществлено по правилу «наименьшего времени обслуживания». Если число (m) машин больше числа (n) работ, то n работ назначаются на любые n машин. В противном случае сначала составляется перечень работ в порядке неубывания времени выполнения каждой работы, а затем назначаются первые m

работ из этого перечня на первые m машин (предполагается, что машины пронумерованы произвольным образом). Последующие работы, включенные в перечень, назначаются на освобождающиеся машины. При таком подходе минимизируются среднее время прохождения работы, среднее время ожидания, среднее время задержки и среднее время запаздывания. Однако это вовсе не означает, что обязательно минимизируется максимальное время прохождения и средний объем незавершенного производства.

Следует отметить, что автор статьи [82] не только предложил назначать работы на любые из m имеющихся идентичных машин, но и заложил основы для составления расписания, минимизирующего максимальное время выполнения работ. Кроме того, он показал, что при нахождении оптимального решения задачи минимизации суммарных штрафов за задержку выполнения работ также может потребоваться назначение работ на любые из m имеющихся идентичных машин. Однако, если все плановые сроки равны нулю, подобное исключается, так как время задержки эквивалентно времени завершения и, следовательно, задача сводится к минимизации общих затрат (или потерь при невыполнении работ в установленные сроки), которые являются линейной функцией времени завершения работ.

Определению нижнего уровня минимальных потерь при невыполнении работ в установленные сроки посвящена статья [26], авторы которой пришли к следующему важному выводу: при наличии m идентичных машин уровень общих потерь при невыполнении работ в заданные сроки будет ниже, если при упорядочении работ предполагается, что они назначаются не на m имеющихся идентичных машин, а на одну машину, причем скорость работы этой машины в m раз больше, чем скорость любой из исходных машин. Из определения потерь, связанных с невыполнением работ в установленные сроки, следует, что подобный вывод справедлив также и для общего и среднего времени завершения работ, времени прохождения и времени ожидания, если предположить, что все работы поступают в систему одновременно. По-видимому, именно этим можно объяснить тот факт, что при решении задач упорядочения работ, выполняемых параллельно работающими машинами, главное внимание уделяется минимизации потерь, связанных с невыполнением работ в установленные сроки.

На основе результатов, полученных в работе [82], для решения данной проблемы разработан ряд сложных алгоритмов. Ниже перечисляются некоторые задачи упорядочения работ, выполняемых параллельно работающими машинами, для которых получены алгоритмы решения и приводятся ссылки на соответствующие литературные источники.

Минимизация общих потерь при невыполнении работ в уста-

новленные сроки при условии, что время прохождения для всех работ одинаково [73].

Минимизация общего времени задержки при условии, что плановые сроки выполнения работ одинаковы [92].

Минимизация суммы потерь, связанных с невыполнением работ в установленные сроки, затрат на переналадку оборудования и на обработку в случае, когда время выполнения каждой из работ может быть разным для всех машин [93].

Минимизация суммы потерь, связанных с невыполнением работ в установленные сроки, затрат на переналадку и обработку [2].

Минимизация взвешенного общего времени прохождения работ [9].

Минимизация суммарного штрафа за задержку выполнения работ при условии, что все работы поступают в систему одновременно, плановые сроки их выполнения одни и те же, а штрафы являются линейной функцией времени задержки [37].

Минимизация суммы (линейной) затрат на прохождение и ожидание [56].

Минимизация длительности производственного цикла при равных длительностях выполнения работ [61].

Минимизация длительности производственного цикла при одинаковой длительности выполнения всех работ и заданных ограничениях на последовательность их выполнения [19, 39, 50, 52].

Следует еще упомянуть работы [42, 99], в которых исследовалась задача определения допустимых расписаний при назначении m работ на n идентичных машин при наличии ограничений на последовательность выполнения работ, а также на имеющиеся в распоряжении ресурсы. Ряд исследований [41, 50, 51] посвящен проблеме определения минимальной длительности производственного цикла.

Результаты многочисленных исследований алгоритмов решения задач календарного планирования для параллельно работающих машин приводятся в работе [34].

3.3. Система многофазного конвейера

В тех случаях, когда для выполнения данной работы необходимо осуществить более чем одну операцию, мы имеем дело с календарным планированием выполнения работ на конвейере. Задачам подобного типа уделялось большое внимание, но многие из предложенных методов их решения являются эвристическими и не гарантируют оптимального решения с учетом возможностей ЭВМ. При детерминистическом подходе к решению таких задач предполагается, что все работы поступают в систему одновре-

менно, и целью календарного планирования является минимизация производственного цикла максимальной длительности.

Следует отметить, что в основе почти всех решений задач календарного планирования работ на конвейере лежит одно из наиболее интересных, с точки зрения исследователя систем, свойств такого производства. Оказывается, что при календарном планировании работы m -фазного конвейера при выполнении всех необходимых условий достаточно рассмотреть только расписание, согласно которому устанавливается одинаковый порядок работ для 1-й и 2-й машин и одинаковый порядок работ для $(m - 1)$ -й и m -й машин (номера машины соответствуют заданной последовательности операций). Следовательно, для конвейера с числом фаз не более трех оптимальная последовательность операций на всех машинах обязательно будет одна и та же. Это в свою очередь означает, что достаточно исследовать множество перестановок работ, не рассматривая возможность существования различных оптимальных последовательностей операций на каждой машине.

Если количество фаз больше трех, то число потенциально оптимальных расписаний резко возрастает, поскольку следует рассматривать последовательность работ на каждой машине. Например, задача с 5 машинами и 10 работами требует рассмотрения $(10!)^5$, т. е. $4,78 \cdot 10^{19}$ возможных расписаний. В результате для решения таких m -фазных задач при $m \geq 4$ используются эвристические методы ограниченного перебора, аналогичные методам поиска приближенного решения.

Таким образом, для решения проблемы упорядочения работ многофазной статической системы имеются простые методы, обеспечивающие точное решение для двух- и трехфазных задач. Методы, применимые к m -фазным ($m \geq 4$) системам, обычно используют специальные предположения и являются либо неэффективными с вычислительной точки зрения, либо дают только приближенное решение. Конечно, многие из методов, предложенных для m -фазных систем, могут применяться для двух- и трехфазных систем, хотя первоначально они были разработаны для систем большей размерности.

Подобно тому как работы Мак-Поутона положили начало исследованиям в области однофазных систем параллельно работающих машин, так и работа Джонсона [69] явилась прелюдией к интенсивным исследованиям задач календарного планирования подобного типа. Прежде чем привести результаты, полученные Джонсоном при решении задачи календарного планирования работ, выполняемых двумя последовательно работающими машинами, введем следующие обозначения: A_i — время выполнения (включая настройку, если она необходима) первой операции i -й работы и B_i — время выполнения (включая настройку, если она необходима) второй операции i -й работы. Тогда, если все работы

поступают в систему одновременно и целью планирования является минимизация длительности производственного цикла, то оптимальная последовательность (напомним, что она должна быть одинаковой на обеих машинах) удовлетворяет условию

$$\min(A_j, B_{j+1}) < \min(A_{j+1}, B_j)$$

для любой пары j -й и $(j + 1)$ -й смежных работ данной последовательности. По-видимому, логически оправданным было бы распространить этот результат на случай трех машин, тем более что оптимальная последовательность работ должна быть одинаковой на всех трех машинах. Однако, как показал Джонсон, это справедливо только тогда, когда выполняется условие

$$\min(A_j + B_j, C_{j+1} + B_{j+1}) < \min(A_{j+1} + B_{j+1}, C_j + B_j)$$

и $\min A_i \geq \max B_j$ или $\min C_i \geq B_j$ (C_j — время выполнения j -й работы на третьей машине). Поскольку последнее условие ни в коем случае нельзя считать тривиальным, то именно это и побудило многих исследователей искать иное решение данной проблемы [47, 62, 77 и 81]. К наиболее выдающимся работам в этой области можно отнести работу [47], в которой дается модель целочисленного программирования, и работы [62, 77, 81], авторы которых разработали алгоритмы метода ветвей и границ.

Необходимо подчеркнуть еще раз тот факт, что если число последовательно работающих машин больше трех, то задача канонического планирования практически не имеет точного решения. Согласно опубликованным данным, для ее решения используются в основном три подхода: комбинаторный анализ, математическое программирование и управляемый перебор по методу ветвей и границ. Фактически все исследования сводятся к минимизации длительности производственного цикла. Поскольку, однако, почти во всех случаях предполагается, что последовательности выполнения работ на всех машинах одинаковы (т. е. «обгон» запрещен), то, естественно, это исключает возможность оптимального решения, так как подобное предположение справедливо только для системы, состоящей менее чем из четырех последовательно работающих машин.

Авторы работы [25] предложили комбинаторное решение рассматриваемой задачи, которое, однако, оказалось исключительно сложным. В связи с этим Кэран [72], а затем Смит и Дудек [101] попытались найти другое решение. Разработанный ими алгоритм действительно был более эффективным, но, к сожалению, не менее сложным, чем разработанный в работе [25].

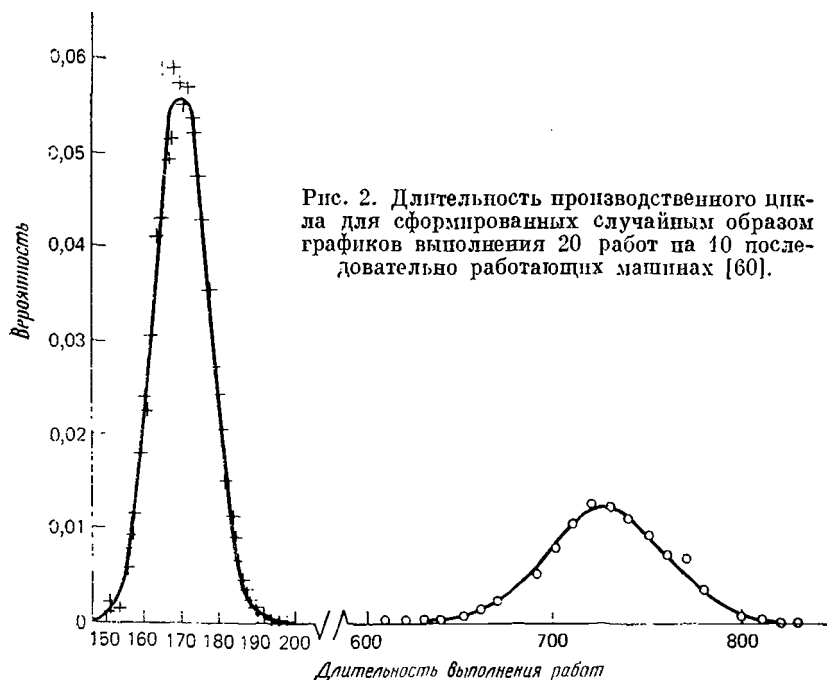
Возможность решения данной задачи как задачи целочисленного программирования была проанализирована в работе [106]. Последующие исследования в этом направлении показали, что при таком подходе практически невозможно реализовать раз-

работанные модели из-за громоздкости и сложности вычислительных процедур.

На основе метода ветвей и границ разработано не менее шести алгоритмов [3, 4, 6, 55, 87], результаты сравнения пяти из которых после их реализации на ЭВМ типа IBM 360/50 приведены в работе [3]. На основе полученных данных (в частности, затраченного машинного времени) был сделан вывод о предпочтительности метода, предложенного в работе [14]. Висмер [109] и Сальвадор [95] показали, что если промежуточное накопление работ между смежными фазами запрещено, то задача календарного планирования может быть сведена к задаче коммивояжера. Кемпбел и др. [17] предложили разделить задачу на p ($p \leq m - 1$, где m — число фаз) вспомогательных подзадач для n работ и двух машин, для которых может быть применен алгоритм Джонсона. Гупта [54] разработал алгоритм решения, который использует метод, называемый «лексикографическим поиском».

Все эти методы, однако, используют предположение о сохранении последовательности работ на всех машинах. Критерием оценки расписания является минимизация длительности производственного цикла. На первый взгляд создается впечатление, что это предположение делается для удобства, например чтобы уменьшить число возможных расписаний с $(n!)^m$ до $n!$ (где n — число работ, а m — число машин) и сделать возможным применение комбинаторных методов и схем прямого перебора или составить расписание, более удобное с точки зрения его реализации, но все это можно сделать только за счет потери оптимальности. Подобные рассуждения не являются полностью неоправданными, так как отклонение этого предположения действительно делает задачу неразрешимой с точки зрения любых практических целей. Тем не менее исследования Хеллера [60] показали, что такое предположение может быть не только удобным, но дает также возможность получать решения, близкие к оптимальному.

Для системы из десяти последовательных машин Хеллер сгенерировал случайным образом 100 работ, время выполнения которых на каждой машине представляет собой целое число в диапазоне от 0 до 9 ед. Затем, используя только первые 20 работ из 100, он сформировал 9037 расписаний с независимыми перестановками целых чисел от 1 до 20, определяющих порядок прохождения каждой работы на каждой машине, и нанес на график длительность производственного цикла для каждого расписания. Далее он составил еще 12 000 расписаний для этих же 20 работ с произвольно заданным одинаковым порядком выполнения работ на каждой машине (рис. 2). Оказалось, что при сохранении заданной последовательности работ не только уменьшается длительность производственного цикла, но и разница между максимальной длительностью производственного цикла для первого



случая и минимальной длительностью для второго составляет более 400 ед. времени. Полученный результат, безусловно, является веским аргументом в пользу обоснованности предположения о сохранении последовательности выполнения работ на многофазной системе, не говоря уже о том, что он наглядно демонстрирует практическую целесообразность такого предположения.

3.4. Гибридные системы

Наибольший интерес с точки зрения изучения гибридных систем представляет обобщенная задача календарного планирования работ на многофазном конвейере с параллельными машинами на каждой фазе. При этом все работы характеризуются одинаковым числом и одинаковой последовательностью операций, каждая из которых может выполняться на одной из нескольких машин (их число не обязательно одинаково для каждой операции). Оптимальное решение подобной задачи при условии, что все работы поступают в систему одновременно, ни одна из них не может задерживаться между фазами и последовательность работ на выходе фазы одна и та же для всех фаз, было получено в рабо-

те [95]. Соответствующий алгоритм основан на использовании методов ветвей и границ и динамического программирования. Хотя результаты вычислений оказались не слишком обнадеживающими для больших производственных систем, тем не менее этот метод был успешно применен в обрабатывающей и химической промышленности.

Не менее интересной является общая задача календарного планирования работы производственного участка, описанная в разд. 3.1. Каждая работа имеет специфический набор операций, которые должны быть выполнены, и имеется только одна машина, способная выполнять операцию каждого вида. Решение задачи для производственного участка с двумя машинами, когда ни одна работа не содержит более двух операций, было дано Джексоном [63]. Существует также графический метод решения задачи для двух работ и m машин, впервые предложенный Эйкерсом и Фридманом [1]. Более эффективное решение предложили Шварц [104] и Хардгрейв и Немхаузер [57].

Алгоритмы для решения общей задачи планирования работ участка без ограничений были сформулированы как модели целочисленного программирования в работах [12, 80, 106], а алгоритмы, основанные на методе ветвей и границ, были даны в работе [13]. К сожалению, для задач достаточно большой размерности все эти алгоритмы оказались не очень эффективными. Так, например, при таком подходе задача с 4 машинами и 10 работами формулируется как задача целочисленного программирования с 220 переменными и 390 ограничениями [80].

По-видимому, единственными методами решения детерминированной задачи календарного планирования работы производственного участка, в общем случае дающими какую-либо надежду на получение практических (не обязательно оптимальных) результатов, являются эвристические методы, из которых наиболее известным и популярным является метод построения графика Ганта.

Другие подобные методы используют правила диспетчирования, которые в своей основе довольно разумны для локальных задач календарного планирования, например для планирования операций, выполняемых на одной машине, либо просто операций, относящихся к одной работе, при котором оптимизируется некоторая локальная оценка расписания. Более сложные процедуры предусматривают составление локальных расписаний во взаимодействии с человеком, принимающим участие в их создании.

Для выяснения потенциальных возможностей различных процедур построения расписаний хотя бы с точки зрения приближенной оптимизации, были проведены обширные исследования с помощью моделирования. Результаты оказались обнадеживающими, но слишком громоздкими, чтобы их можно было изложить доста-

точно подробно. Полный анализ существующих эвристических методов и их пригодность для решения задачи календарного планирования работы производственного участка в общем случае можно найти в работе [20].

3.5. Общие методы решения детерминированных задач календарного планирования

Согласно результатам большинства исследований, детерминированная задача календарного планирования является задачей упорядочения. Кроме того, поскольку детерминированное планирование использует известные априори показатели, характеризующие работы и машины, то множество последовательностей, из которого должно быть выбрано расписание, также известно и фиксировано. Как отмечает автор работы [29], выбор оптимального или «близкого» к оптимальному расписания до сих пор осуществляется с помощью одного из четырех подходов: комбинаторного, математического программирования, эвристического и статистического моделирования.

Комбинаторный подход в данном случае сводится к целенаправленной взаимной перестановке пар работ в некоторой исходной последовательности, пока не будет получено оптимальное решение. Основы этого подхода, обобщенные в работе [102], получили дальнейшее развитие [40, 48, 63, 69, 70, 82, 85].

Что касается методов математического программирования, то они включают линейное, динамическое, квадратичное, выпуклое и целочисленное программирование, а также некоторые методы теории сетей и метод неопределенных множителей Лагранжа.

Наиболее перспективными, по-видимому, являются эвристические методы, использующие направленный перебор, такие, как метод ветвей и границ. Подобные методы обычно находятся ближе к процедурам полного перебора всех возможных последовательностей, которые, очевидно, являются наименее эффективными для получения оптимальной последовательности. Однако они обладают единственным в своем роде свойством, позволяющим с помощью определения границ критерия оценки, допустимости или доминирования исключать из рассмотрения определенные последовательности еще до того, как для этих последовательностей был вычислен критерий оценки расписания.

Последний из указанных методов — это случайный поиск потенциально оптимальной последовательности с помощью статистического моделирования. Хеллер [60] использовал этот метод и сформировал 3000 случайных последовательностей для участка с десятью последовательно работающими машинами. В случае когда время выполнения операций на каждой машине предполагалось распределенным по равномерному закону, длительность

производственного цикла оказалась нормально распределенной. Хотя случайная длительность операций — это прерогатива стохастической теории расписаний, тем не менее, как показано в работе [29], подобный метод может быть полезен в детерминированных задачах календарного планирования общего вида благодаря использованию статистических свойств последовательностей (как, например, распределения длительности производственного цикла) для управления итеративным процессом выбора решения. Объем выборки можно было бы ограничивать таким заранее заданным числом испытаний, которое позволило бы делать статистически достоверные выводы относительно величины производственного цикла для «наилучшей» выбранной последовательности, либо он мог бы определяться некоторым привходящим фактором, таким, как стоимость статистического моделирования. Когда предельные затраты на моделирование превышают предельную величину выигрыша от улучшения расписания, используется расписание с наименьшей длительностью производственного цикла, выбранное до этого момента. В зависимости от назначения конкретной задачи календарного планирования можно было бы использовать другие «правила остановки».

Естественно, что необходимы методы решения, которые дают «хорошие» расписания относительно эффективным способом, не взирая на масштаб задачи. Из указанных выше подходов только два последних подходят в этом отношении какую-то надежду. Основные принципы этих подходов предполагают следующую последовательность действий при анализе детерминированных задач календарного планирования, точное решение которых либо невозможно, либо затруднительно.

— Сокращение множества всех допустимых последовательностей выполнения работ обеспечивается за счет исключения явно недопустимых последовательностей. Многие прикладные задачи связаны с такими процессами, в которых технологические ограничения автоматически исключают определенные последовательности. Некоторые из них очевидны (например, нельзя нарезать резьбу в отверстии до того, как его просверлили), а другие не настолько тривиальны (например, химические ограничения процесса окраски). В некоторых случаях нетрудно построить множество допустимых последовательностей (вместо того чтобы выявлять недопустимые), если ограничения достаточно строгие. Если сокращенное число последовательностей невелико, то можно прибегнуть к полному перебору.

— Исключение последовательностей с явно плохими значениями критерия оценки расписания. Для этого определяется нижняя граница меры оптимального расписания, подобно тому как это было сделано в работе [26] для случая параллельно работающих машин. Так же как и в предыдущем случае, желательно,

хотя и не всегда возможно, построить потенциально оптимальные последовательности. Джонсон [69] предложил подобное решение для задач календарного планирования работ, выполняемых двумя и тремя последовательно работающими машинами.

— Опробование какой-либо схемы направленного перебора, типа метода ветвей и границ или общего метода [46], если ни одна из указанных выше операций не позволяет избежать полного перебора, т. е. число допустимых решений остается все еще очень большим. Последовательности могут систематически оцениваться таким образом, чтобы можно было неявно отбрасывать большое число последовательностей путем исследования непогих. Подобные схемы иногда легко реализуются ввиду особенностей рассматриваемых задач.

— Использование менее строгих или введение дополнительных ограничений. Наиболее ярким примером такого подхода является конвейер из четырех и более машин. Так, чтобы сделать задачу более обозримой, авторы работ [62, 77, 81, 95, 101] предположили, что последовательность выполнения работ на каждой машине должна быть одинаковой, хотя для получения оптимального решения это не обязательно.

Также было показано [95], что в практических задачах календарного планирования работ, выполняемых на конвейере, добавление ограничений об отсутствии запасов между фазами не только делает задачу разрешимой и приводит к улучшению расписания, но и значительно снижает объем капиталовложений благодаря исключению потребности в межоперационных запасах. Следовательно, прежде всего необходимо установить, какие из существующих ограничений следует ослабить либо какие новые ограничения наложить, чтобы сделать задачу более просто решаемой. Далее надо определить условия (например, ограничения на длительность операций), необходимые для того, чтобы выполнялись (или нарушались) ослабленные ограничения либо вновь вводимые ограничения. Иногда эти условия соответствуют реально существующим, что оправдывает изменение постановки задачи.

— Разложение задачи на подзадачи. Например, можно рассмотреть возможность поиска оптимальных расписаний для каждой машины в отдельности или для некоторых подгрупп машин.

— Построение расписания с помощью случайного поиска, т. е. формирования случайным образом множества всех возможных последовательностей, или, что более предпочтительно, только множества допустимых последовательностей, если их можно выделить. При помощи вычислительной машины можно легко оценить тысячи расписаний и составить статистику оценок расписания. Затем можно, используя метод статистического модели-

рования, как это описывалось ранее, построить приемлемые расписания. К подобному методу прибегают в тех случаях, когда все перечисленные ранее способы не дают желаемого результата.

4. СТОХАСТИЧЕСКИЕ ЗАДАЧИ КАЛЕНДАРНОГО ПЛАНИРОВАНИЯ

В отличие от детерминированных задач, в которых все параметры системы известны априори, стохастические задачи календарного планирования относятся к системам, параметры которых меняются случайным образом, т. е. по некоторым законам распределения вероятностей. При анализе подобных задач не накладываются никакие ограничения, хотя, как правило, исследуются задачи со случайным моментом поступления работ и неопределенным временем их выполнения. Кроме того, могут быть случайными размер производственного участка (число машин), технологические маршруты работ и плановые сроки окончания.

Одно из существенных различий между детерминированным и стохастическим подходами к решению задач календарного планирования состоит в выборе критерия оценки эффективности расписания. Если в первом случае управляющие переменные, относящиеся к работам и производственному участку, могут быть измерены точно и целью планирования является оптимизация некоторого обобщенного показателя, то во втором случае эти обобщенные показатели работ и производственного участка выражаются в статистических терминах и процедура оптимизации имеет совершенно иную природу. При решении стохастических задач календарного планирования сначала получают распределение вероятности интересующего нас обобщенного показателя, например времени прохождения работы, а затем оценивают характеристики этого распределения, например среднее время прохождения. Кроме того, стохастические модели календарного планирования являются в основном дискретивными, тогда как в детерминированном случае желательно строить нормативные модели. Заметим, однако, что последние достижения в применении теории оптимизации к задачам массового обслуживания привели также к появлению нормативных вероятностных моделей.

Кроме того, существенно отличаются методы построения расписания. Для детерминированных задач определяется последовательность поступления работ на машины. В стохастическом случае, когда работы поступают случайным образом, обычно имеют дело с постоянной очередью работ, ожидающих выполнения на каждой машине. Последовательность, в которой эти работы должны обрабатываться машиной, может быть определена только некоторой дисциплиной выбора из очереди (этот термин часто используется в теории массового обслуживания наряду с термином «правило приоритетов»), определяющей, какая одна из работ,

ожидающих в очереди, должна обрабатываться, когда данная машина доступна для обработки.

Следует иметь в виду, что стохастические задачи календарного планирования, по определению, являются задачами теории массового обслуживания. Результаты исследований стохастических задач приводятся ниже в соответствии с представленной ранее схемой классификации. При этом используются следующие обозначения:

P — время выполнения работ, W — время ожидания, F — время прохождения работы, N — число работ в системе в установившемся состоянии.

Кроме этого, все результаты, которые будут приведены, базируются на следующих общих предположениях относительно моментов поступления работ и времени их выполнения:

— процесс поступления работ описывается пуассоновским распределением с параметром λ , т. е. в течение любого интервала времени t вероятность того, что число (A) возникающих работ равно n , определяется формулой

$$P\{A=n\} = e^{-\lambda t} \frac{(\lambda t)^n}{n!}, \quad n=0, 1, \dots$$

(Возможны и другие распределения моментов поступления работ.)

— средняя скорость $1/\xi(P)$ выполнения работ, где P — случайная переменная, характеризующая длительность выполнения работ, всегда больше средней скорости λ поступления работ. Другими словами, предполагается, что $\lambda \xi(P) < 1$; в противном случае длина очереди возрастала бы монотонно и неограниченно вне зависимости от дисциплины обслуживания. Величину $\lambda \xi(P)$ будем обозначать ρ .

4.1. Планирование работ для одной машины

Исследованию задач календарного планирования работ, выполняемых одной машиной, или, используя терминологию теории массового обслуживания, одноканальной системы массового обслуживания, посвящено довольно большое число работ. Исчерпывающее рассмотрение этой задачи как с точки зрения теории расписаний, так и теории массового обслуживания можно найти соответственно в работах [20] и [94].

Основная масса исследований по теории расписаний обычно придерживается ограничений простого процесса выполнения работ производственным участком, о котором говорилось ранее, за исключением тех моделей, при разработке которых делаются различного рода допущения о возможности прерывания работ.

Ниже перечислены дисциплины обслуживания очереди (последовательности работ), представляющие наибольший интерес для

стохастической задачи календарного планирования при использовании одной машины. Здесь также даются ссылки на соответствующие работы. В некоторых случаях приводятся формулы для моментов результирующих распределений вероятностей времени ожидания, времени прохождения и числа работ в системе.

— Обслуживание в порядке поступления, т. е. прямой порядок обслуживания («первым прибыл — первым обслужен»): первой выполняется работа с большим временем ожидания. В этом случае имеем

$$\begin{aligned} \mathcal{E}(W) &= \frac{\lambda \mathcal{E}(P^2)}{2(1-\rho)}, \\ \mathcal{E}(W^2) &= \frac{\lambda \mathcal{E}(P^3)}{3(1-\rho)} + \frac{[\lambda \mathcal{E}(P^2)]^2}{2(1-\rho)^2}, \\ \mathcal{E}(F) &= \mathcal{E}(P) + \frac{\lambda \mathcal{E}(P^2)}{2(1-\rho)}, \\ \mathcal{E}(F^2) &= \mathcal{E}(P^2) + \frac{\rho \mathcal{E}(P^2)}{1-\rho} + \frac{\lambda \mathcal{E}(P^3)}{3(1-\rho)} + \frac{[\lambda \mathcal{E}(P^2)]^2}{2(1-\rho)^2}, \\ \mathcal{E}(N) &= \rho + \frac{\lambda^2 \mathcal{E}(P^2)}{2(1-\rho)}. \end{aligned}$$

(Аналогичные результаты были получены многими исследователями, например в работе [5].)

— Обслуживание в обратном порядке («первым прибыл — последним обслужен»): первой выполняется работа, которая ожидает меньше всего. При этой дисциплине обслуживания среднее время ожидания и среднее время прохождения такие же, как и в предыдущем случае, однако

$$\begin{aligned} \mathcal{E}(W^2) &= \frac{\lambda \mathcal{E}(P^3)}{3(1-\rho)^2} + \frac{[\lambda \mathcal{E}(P^2)]^2}{2(1-\rho)^3}, \\ \mathcal{E}(F^2) &= \mathcal{E}(P^2) + \frac{\rho \mathcal{E}(P^2)}{1-\rho} + \frac{\lambda \mathcal{E}(P^3)}{3(1-\rho)^3} + \frac{[\lambda \mathcal{E}(P^2)]^2}{2(1-\rho)^3}. \end{aligned}$$

— Обслуживание в соответствии с установленным приоритетом работ без их прерывания: работы, поступающие в систему, относятся к одному из конечного числа классов по приоритету; первой выполняется работа, которая из класса с наивысшим приоритетом имеет наибольшее время ожидания [18].

— Дисциплина обслуживания кратчайшей заявки: работы, поступающие в систему, относятся к одному из классов по приоритету, который устанавливается в зависимости от распределения времени выполнения данной работы; первой выполняется работа с наивысшим приоритетом (с наименьшим ожидаемым временем выполнения) [89].

— Обслуживание в соответствии с установленным приоритетом с прерыванием и возобновлением обслуживания: если посту-

пает работа с более высоким приоритетом, чем у работы, которая выполняется, выполнение последней прерывается и она возвращается в начало очереди своего класса [5, 44, 58, 107].

— Обслуживание в соответствии с установленным приоритетом с прерыванием и возобновлением прерванной работы с точки прерывания [58, 84, 107].

— Обслуживание по принципу «наименьшего оставшегося времени обслуживания»: прерывание допускается, и длительность дообслуживания после прерывания работы равна лишь времени, необходимому для ее завершения [83, 95].

Ниже приводятся наиболее интересные результаты, полученные при сравнении приведенных дисциплин по среднему времени прохождения с помощью моделирования. Эксперименты [20] проводились при различных скоростях поступления работ и экспоненциальном и равномерном распределении времени их выполнения.

Дисциплина обслуживания по принципу наименьшего оставшегося времени обслуживания во всех случаях обеспечивает минимум среднего времени прохождения. Другие дисциплины, зависящие от времени выполнения, по-видимому, также позволяют сократить среднее время прохождения до значений более низких, чем при прямом порядке обслуживания.

Дисциплины с прерыванием обслуживания не обладают особыми преимуществами по сравнению с другими дисциплинами и к тому же трудны для реализации на практике.

Из дисциплин по крайней мере с двумя классами приоритетов в случае экспоненциального распределения времени выполнения работы дисциплина без прерывания, по-видимому, дает лучшие показатели, чем дисциплина с прерыванием. Ситуация меняется при возрастании числа классов.

При прямом порядке обслуживания и экспоненциальном распределении времени выполнения работ среднее время прохождения всегда больше, чем при равномерном распределении времени выполнения. В случае дисциплины «наименьшего оставшегося времени обслуживания» наблюдается обратная зависимость. При использовании других дисциплин обслуживания очереди результаты зависят от скорости поступления работ.

Основным вкладом в исследование стохастических задач календарного планирования для одной машины с учетом затрат времени на перепалдку и плановых сроков окончания работ явилась работа [44]. Автор работы рассмотрел случай прямого порядка обслуживания при условии случайного времени перепалдки.

Предположим, что имеется r классов настроек, так что время перепалдки (случайная переменная S) учитывается только в случае, когда некоторая работа следует непосредственно за

работой другого класса. Предположим также, что скорость поступления работ класса r равна λ/r для всех классов. Тогда время, необходимое для выполнения работы, равно P с вероятностью $1/r$ и $(P + S)$ с вероятностью $(r - 1)/r$. Когда работы выполняются в порядке их поступления, средние времена ожидания и прохождения для всех классов равны

$$\mathcal{E}(W) = \frac{\lambda}{2(1-U)} \left\{ \mathcal{E}(P^2) + \left(\frac{r-1}{r} \right) [2\mathcal{E}(P)\mathcal{E}(S) + \mathcal{E}(S^2)] \right\},$$

$$\mathcal{E}(F) = \mathcal{E}(W) + U/\lambda,$$

где

$$U = \rho + \left(\frac{r-1}{r} \right) \lambda \mathcal{E}(S).$$

Анализ плановых сроков окончания работ, который приводит к рассмотрению запаздывания и задержек, значительно более сложен, чем любой из приведенных до сих пор результатов. Все немногое, что известно о задачах этого класса (и то главным образом предположительно), можно найти в одной из работ [65, 67, 68].

4.2. Однофазная система параллельно работающих машин

Стохастическая задача календарного планирования работ для параллельно работающих машин до сих пор подробно не исследовалась. Это справедливо и в отношении эквивалентной проблемы теории массового обслуживания — однофазной многоканальной системы массового обслуживания. Судя по имеющимся результатам, большинство авторов исследовали рассматриваемую задачу при прямом порядке обслуживания очереди, т. е. когда на освободившуюся машину назначается первая из очереди работа.

Более интересной является задача назначения работ в одну из нескольких очередей к машинам с возможным изменением очереди перед выполнением. При решении задач подобного типа, как правило, предполагается существование общей очереди, прямой дисциплины обслуживания, наличие идентичных машин с отрицательным экспоненциальным распределением времени выполнения работ, имеющим параметр μ . Пусть m — число параллельно работающих машин. Обозначим для удобства записи $r = \lambda/\mu$ (заметим, что в данном случае $\rho = \lambda/m\mu$). Тогда имеем

$$\mathcal{E}(W) = \frac{P_0 r^{m+1}}{\lambda(m-1)!(m-r)^2},$$

$$\mathcal{E}(F) = \mathcal{E}(W) + (1/\mu), \quad \mathcal{E}(N) = \lambda \mathcal{E}(W) + r,$$

где

$$P_0 = \left\{ \left[\sum_{n=0}^{m-1} \frac{r^n}{n!} \right] + \left[\frac{r^m}{m! \left(1 - \frac{r}{m} \right)} \right] \right\}^{-1}.$$

4.3. Система многофазного конвейера

Важнейшим результатом немногочисленных исследований стохастической задачи планирования выполнения работ на многофазном конвейере является метод декомпозиции, разработанный автором работы [90] в предположении, что выполняются следующие условия:

- Работы поступают к машинам либо извне (в соответствии с пуассоновским распределением), либо от машины, выполнявшей предыдущую операцию.

- Работа с известной вероятностью может покинуть производственный участок после любой машины.

- Дисциплина выбора из очереди для любой машины не зависит от времени прохождения.

При выполнении этих условий каждую машину участка можно рассматривать независимо от других. Интенсивность пуассоновского потока поступления работы на каждую машину определяется как композиция интенсивности потока поступления работ извне и вероятностей их выхода из системы.

Поскольку перечисленные условия не являются необычными для стохастических систем, то метод декомпозиции позволяет использовать для анализа многофазного конвейера методы планирования работ для одной машины. Различные варианты этой задачи, такие, как задача с ограничением на длину очереди между фазами или случай, когда каждая работа должна обрабатываться на каждой машине, исследовались в работах [6, 35, 38].

4.4. Гибридные системы

Основным методом решения общей стохастической задачи календарного планирования является метод декомпозиции, который был разработан Джексоном [64] в предположении, что «сеть массового обслуживания» или гибридная стохастическая система удовлетворяют следующим условиям.

- Работы поступают на производственный участок из внешнего источника в соответствии с пуассоновским распределением.

- Время прохождения для каждой машины имеет экспоненциальное распределение.

- Работы распределяются между машинами либо покидают производственный участок в соответствии с заданным набором вероятностей.

- Дисциплина обслуживания очереди для любой машины не зависит от времени прохождения.

При выполнении этих условий можно считать, что для каждой машины на производственном участке имеется независимая задача календарного планирования (массового обслуживания) с одной

машиной. Соответственно при анализе гибридной системы, удовлетворяющей этим условиям, можно полностью использовать результаты, полученные для многочисленных моделей с одной машиной. Так же как и в случае стохастического конвейера, возможность декомпозиции приводит к тому, что сложный гибридный производственный участок практически всегда анализируется по частям.

4.5. Общие методы решения стохастических задач календарного планирования

В силу природы стохастических задач календарного планирования зависимость между различными дисциплинами обслуживания очереди и критериями оценки расписания, за исключением случая, когда рассматривается одна машина, не может быть установлена аналитически. Экспериментальное определение подобной зависимости связано со значительными затратами денежных средств и времени, поэтому единственным эффективным методом исследования таких задач оказывается комплексное моделирование, в частности статистическое моделирование с использованием ЭВМ для хранения, обработки и поиска данных. Статистическое моделирование, как правило, предполагает выполнение следующих основных (сравнительно простых) операций.

— Все переменные состояния системы хранятся в ЭВМ. В число переменных входят, например, состояние каждой работы (где она «находится»), число работ в очереди к машине и определяющие их характеристики, время, когда освободится каждая машина, и общее время ожидания для каждой работы.

— Процесс наступления случайных событий, таких, как поступление работы в систему, завершение выполнения работы и распределение работ на следующие машины, моделируется с помощью случайного выбора в соответствии с функцией распределения вероятностей, описывающей действительный процесс возникновения этих событий.

— Наступление детерминированных событий, таких, как выбор работы из очереди к машине, когда машина становится доступной, контролируется с помощью хронологического перечня этих событий. Этот перечень постоянно обновляется при возникновении моделируемых случайных событий и изменении переменных состояния системы. Данные события осуществляются просто путем обновления переменных состояния системы в соответствии с влиянием, которое события оказывают на систему. Например, когда машина становится доступной, некоторая работа выбрасывается из очереди в соответствии с заданной дисциплиной обслуживания. Затем пересчитывается время ожидания для этой работы, время, когда машина освободится в следующий раз, и т. д.

Когда такой процесс заканчивается, переменные состояния системы, соответствующие определенному интервалу времени, будут содержать интересующую исследователя информацию, например среднее время прохождения работ, среднее число работ в очереди, среднее время задержки работ, для которых установлены плановые сроки и т. д.

При решении задачи обычно принимается предположение относительно простого процесса выполнения работ на производственном участке, пуассоновского распределения времени поступления работ в систему и вероятности выбора каждой машины на участке в качестве «следующей» для выполнения работы $p < 1$. Благодаря общности этого подхода отсутствуют какие-либо ограничения на показатели работ и производственный участок, которые могут изменяться случайным образом, а также на моделируемую дисциплину обслуживания очереди или контролируемый критерий оценки расписания. Исчерпывающий обзор исследований в этой области приведен в работе [22].

5. ПЕРСПЕКТИВЫ РАЗВИТИЯ ТЕОРИИ И МЕТОДОВ КАЛЕНДАРНОГО ПЛАНИРОВАНИЯ И УПОРЯДОЧЕНИЯ РАБОТ

Результатом многочисленных исследований в области календарного планирования и упорядочения являются модели, алгоритмы, процедуры, и т. п., включая простейшие графики Ганта и сложнейшие модели оптимизации. Очевидно, что круг теоретических проблем, обусловленных этими исследованиями, столь же обширен. Поскольку, однако, большинство задач календарного планирования и упорядочения является по своей природе комбинаторными, то и методы их решения оказываются громоздкими, как правило, итеративными и дорогими, требующими больших затрат машинного времени. Не исключено, что некоторые из этих «решений» никогда не будут реализованы.

Согласно результатам работ по «теории сложности» [24, 74, 105], определенные комбинаторные задачи могут быть подразделены на задачи «P-временные» и «NP-временные» в зависимости от того, ограничено их «время» решения полиномиальной функцией или нет. Применяя эту теорию к задачам календарного планирования и упорядочения, Брукер и др. [16] показали, что некоторые из этих задач относятся к NP-многообразию (так называемые NP-завершающие задачи), следовательно, в данном случае оправданы только методы решения, основанные на переборе в виде обобщенного метода ветвей и границ.

Очевидно, что разработка и использование аналогичных методов является наиболее перспективным направлением дальнейших исследований в области календарного планирования выполнения работ. Нам представляется, что поиск наиболее эффективных ме-

тодов определения «хороших» графиков, а не самых неэффективных из числа возможных, по обещающих оптимальное решение,— вот основная задача, стоящая перед исследователями систем, на успешное решение которой так надеются практики.

ЛИТЕРАТУРА

1. Akers S. B., Jr., Friedman J., A Non-Numerical Approach to Production Scheduling Problems, *Operations Res.*, 3, No. 6, 429—442 (1955).
2. Arthanari T. S., Ramamurthy K. G., A Branch and Bound Algorithm for Sequencing n Jobs on m Parallel Processors, *Opsearch*, 7, No. 3, 147—156 (1970).
3. Ashour Said, Quraishi M. N., Investigation of Various Procedures for Production Scheduling Problems, *Int. J. Prod. Res.*, 7, No. 13, 249—252 (1969).
4. Ashour Said, A Branch and Bound Algorithm for Flow Shop Scheduling Problems, *AIIE Trans.*, 2, No. 2, 172—176 (1970).
5. Avi-Itzhak B., Preemptive Repeat Priority Queues as a Special Case of the Multi-Purpose Server Problem—I and II, *Operation Res.*, 11, No. 4, 597—617 (1963).
6. Avi-Itzhak B., A Sequence of Service Stations with Arbitrary Input and Regular Service Times, *Management Sci.*, 11, No. 5, 553—564 (1965).
7. Avi-Itzhak B., Maxwell W. L., Miller L. W., Queuing with Alternating Priorities, *Operations Res.*, 13, No. 2, 306—318 (1965).
8. Baker K. R., On Madigan's Approach to the Deterministic Multi-Production and Inventory Problem, *Management Sci.*, 12, No. 9, 636—638 (1970).
9. Baker K. R., Merten A. G., Scheduling with Parallel Processor and Linear Delay Costs, *Nav. Res. Log. Quart.*, 20, No. 4, 193—804 (1973).
10. Bellman R., Dynamic Programming Treatment of the Traveling Salesman Problem; *J. Assoc. Comp. Mach.*, 9, No. 1, 61—63 (1962).
11. Bomberger Earl, A Dynamic Programming Approach to a Lot Size Scheduling Problem, *Management Sci.*, 12, No. 11 (1966).
12. Bowman E. H., The Schedule-Sequencing Problem, *Operations Res.*, 7, No. 1, 621—624 (1959).
13. Brooks G. H., White C. R., An Algorithm for Finding Optimal or Near-Optimal Solutions to the Production Scheduling Problem, *J. Ind. Eng.*, 16, No. 1, 34—40 (1965).
14. Brown A. P. G., Lomnicki Z. A., Some Application of the 'Branch-and-Bound' Algorithm to the Machine Scheduling Problem, *Operational Res. Quart.*, 17, No. 2, 173—186 (1966).
15. Brown R. G., Decision Rules for Inventory Management Holt, Rinehart and Winston, N. Y., 1967, pp. 46—55.
16. Brucker P., Lenstra J. K., Rinnooy Kan A. H. G., Complexity of Machine Scheduling Problems, Dept. of Math. Dec. Sc., BW 43/75, Mathematisch Centrum, Amsterdam, 1975.
17. Campbell Herbert G., Dudek R. A., Smith M. L., A Heuristic Algorithm for the n Job Sequencing Problem, *Management Sci.*, 16, No. 10, B-630—B-637 (1970).
18. Cobham A., Priority Assignment in Waiting Line Problems, *Operations Res.*, 2, No. 1, (1954).
19. Coffman E. G., Jr., Graham R. L., Optimal Scheduling for Two-Processor Systems, *Acta Informatica*, 1, 213—299 (1972).
20. Conway R. W., Maxwell W. L., Miller L. W., Theory of Scheduling, Addison Wesley Publishing Company, Reading, Massachusetts, 1967.

[Имеется перевод: Конвей Р. В., Максвелл В. Л., Миллер Л. В. Теория расписаний.— М.: Наука, 1975.]

21. Cook S. A., The Complexity of Theorem-Proving Procedures, in Proc. 3rd Annual ACM Symp. on Computing, 1971.
22. Day J. E., Hottenstein M. P., Review on Sequencing Research, *Nav. Res. Log. Quart.*, 17, No. 1, 11—40 (1970).
23. Doll C. L., Whybark D. C., An Iterative Procedure for the Single-Machine Multi-Product Lot Scheduling Problem, *Management Sci.*, 20, No. 1, 50—55 (1973).
24. Dooley J. E., Ben Bernholtz, A Many-Product Single-Machine Recursive Scheduling Procedure, *CORS J.*, 7, No. 3, 165—176 (1969).
25. Dudek R. A., Teuton O. F., Jr., Development of M-Stage Decision Rule for Scheduling n Jobs Through m Machines, *Operations Res.*, 12, No. 3, 471—497 (1964).
26. Eastman W. L., Even S., Issacs I. M., Bounds for the Optimal Scheduling of n Jobs on m Processors, *Management Sci.*, 11, No. 2, 268—279 (1964).
27. Eilon S., Elements of Production Planning and Control Macmillan, N.Y., Chapter 14, 1962.
28. Elmaghraby S. E., On the Sequencing of n Job on One Machine to Minimize the Number of Jobs, *Management Sci.*, 18, No. 7, 389 (1972).
29. Elmaghraby S. E., The One Machine Sequencing Problem with Delay Costs, *J. Ind. Eng.*, 19, No. 2, 105—108 (1968).
30. Elmaghraby S. E., The Sequencing of n Jobs on m Parallel Processors with Extensions to the Scarce Resources Problem of Activity Networks, Presented at the Inaugural Conference, The Scientific Computation Center., Cairo Egypt, December 17—20, 1969.
31. Elmaghraby S. E., The Machine Sequencing Problem-Review and Extensions, *Nav. Res. Log. Quart.*, 15, No. 2, 205—232 (1968).
32. Elmaghraby S. E., The Sequencing of n Jobs on m Parallel Processors, Research Memorandum North Carolina State University at Raleigh, N.C., January 1968.
33. Elmaghraby S. E., The Sequencing of 'Related' Jobs, *Nav. Res. Log. Quart.*, 15, No. 1, 23—32 (1968).
34. Elmaghraby S. E., Elshafei A. N., The Scheduling of Jobs on Parallel Processors: a Survey and Annotated Bibliography, presented at the Logistics Research Conference, George Washington University, Washington, D.C., 1975.
35. Elmaghraby S. E., Ginsberg A. S., A Dynamic Model for the Optimal Loading of Linear Multi-Operation Shops, *Management Tech.*, 4, No. 1, 47—58 (1964).
36. Elmaghraby S. E., Mallik A., The Scheduling of a Multi-Product Facility, Proceedings of the Symposium on the Theory of Scheduling and Its Application, Springer-Verlag, Berlin, 1973, pp. 244—277.
37. Elmaghraby S. E., Park S., On the Scheduling of Jobs on a Number of Identical Machines, *IE Trans.*, 6, No. 1, 1—13 (1974).
38. Friedman H. D., Reduction Methods for Tandem Queuing Systems, *Operations Res.*, 13, No. 1, 121—131 (1965).
39. Fujii M., Kasami T., Ninomiya K., Optimal Sequence of Two Equivalent Processors, *SIAM J. Appl. Math.*, 17, No. 4, 784—789 (1969).
40. Gapp W., Mannkekar P. S., Mitten L. G., Sequencing Operations to Minimize In-Process Inventory Costs, *Management Sci.*, 11, No. 3, 476—484 (1965).
41. Garey M. R., Graham R. L., Bounds for Multiprocessor Scheduling with Resource Constraints, Bell Telephone Labs., Murray Hill, N.J., 1974.
42. Garey M. R., Johnson D. S., Complexity Results for Multiprocessor Scheduling Under Resource Constraints, Bell Telephone Labs., Murray Hill, N.J., 1974.

43. Garey M. R., Scheduling Tasks with Nonuniform Deadlines on Two Processors Bell Telephone Labs., Murray Hill, N.J., 1975.
44. Gaver D. P., Jr., A Comparison of Queue Disciplines when Service Orientation Times Occur, *Nav. Res. Log. Quart.*, 10, No. 3, 219—235 (1963).
45. Gelders L., Kleindorfer P. R., Coordinating Aggregate and Detailed Scheduling Decisions in the One-Machine Job Shop: Part I, Theory, *Operations Res.*, 22, No. 1, 46—60 (1974).
46. Giffler B., Thompson G. L., Algorithm for Solving Production-Scheduling Problems, *Operations Res.*, 8, No. 4, 487—513 (1960).
47. Giglio R. J., Wagner H. M., Approximate Solutions to the Three-Machine Scheduling Problems, *Operations Res.*, 12, No. 2, 305—324 (1964).
48. Gilmore P., Gomory R. E., Sequencing of One State-Variable Machine, *Operations Res.*, 12, No. 5, 655—679 (1964).
49. Glassco C. R., Scheduling Several Products on One Machine to Minimize Changeovers, *Operations Res.*, 16, No. 2, 342—352 (1968).
50. Graham R. L., Bounds on Multiprocessing Timing Anomalies, *SIAM J. Appl. Math.*, 17, 416—429 (1969).
51. Graham R. L., Bounds on Multiprocessing Timing Anomalies, *Bell Syst. Tech. J.*, 45, 1563—1581 (1966).
52. Greham R. L., Bounds on Multiprocessing Anomalies and Related Packing Algorithm, AFIPS Conf. Proc., 1972.
53. Gupta J. N. D., A Functional Heuristic Algorithm for the Flow Shop Scheduling Problem, *Operations Res. Quart.*, 66, No. 1, 39—47 (1974).
54. Gupta J. N. D., A General Algorithm for the $n \times m$ Flow Shop Scheduling Problem, *Int. J. Prod. Res.*, 7, No. 3, 244—247 (1969).
55. Gupta J. N. D., M-Stage Flow Shop Scheduling by Branch and Bound, *Opsearch*, 7, No. 1, 37—43 (1970).
56. Gupta J. N. D., Walvakar A., Sequencing n -Jobs on m -Parallel Processors, *Opsearch*, 6, No. 3, 295—298 (1970).
57. Hardgrave W. W., Nemhauser G. L., A Geometric Model and A Graphical Algorithm for a Sequencing Problem, *Operations Res.*, 11, No. 6, 889—900 (1963).
58. Heathcote C. R., A Single Queue with Several Preemptive Priority Classes, *Operations Res.*, 8, No. 5, 630—638 (1960).
59. Held M., Karp R. M., A Dynamic Programming Approach to Sequencing Problems, *SIAM J. of Appl. Math.*, 10, No. 2, 196—210 (1962). [Имеется перевод: Хелд М., Карп Р. М. Применение динамического программирования к задачам упорядочения. — Кибернетический сборник, 1964, № 9.]
60. Heller J., Some Numerical Experiments for an $M \times J$ Flow Shop and Its Decision-Theoretical Aspects, *Operations Res.*, 8, No. 2, 178—184 (1960).
61. Hu T. C., Parallel Sequencing and Assembly Line Problems, *Operations Res.*, 9, No. 6, 841—848 (1961).
62. Ignal E., Schrage L., Application of the Branch and Bound Technique to Some Flow Shop Scheduling Problems, *Operations Res.*, 17, No. 4, 400—412 (1965).
63. Jackson J. R., An Extension of Jonson's Results on Job-Lot Scheduling, *Nav. Res. Log. Quart.*, 7, No. 3, 201—204 (1956).
64. Jackson J. R., Jobshop-Like Queueing Systems Research Report 81, Management Sciences Research Project, UCLA, 1963.
65. Jackson J. R., Queues with Dynamic Priority Discipline *Management Sci.*, 8, No. 1, 18—34 (1961).
66. Jackson J. R., Scheduling a Production Line to Minimize Maximum Tardiness, Research Report 43, Management Sciences Research Project, UCLA, 1955.
67. Jackson J. R., Some Problems in Queueing with Dynamic Properties, *Nav. Res. Log. Quart.*, 7, No. 3, 235—249 (1960).

68. Jackson J. R., Waiting-Time Distribution for Queues with Dynamic Priorities, *Nav. Res. Log. Quart.*, 9, No. 1, 31-36 (1962).
69. Johnson S. M., Optimal Two- and Three-Stage Production Schedules with Setup Times Included, *Nav. Res. Log. Quart.*, 1, 61-68 (1954).
70. Johnson S. M., Discussion; Sequencing n Jobs on 2 Machines with Arbitrary Time Lags, *Management Sci.*, 5, No. 3, 299-303 (1959).
71. Karp R. M., Raducibility Among Combinatorial Problems, in: Complexity of Computer Computations, Miller and Thatcher (Eds.), Plenum Press, N.Y., 1972, pp. 85-103.
72. Karush W., A Counterexample to a Proposed Algorithm for Optimal Sequencing of Jobs, *Operations Res.*, 13, No. 2, 323-325 (1965).
73. Lawler E. L., On Scheduling Problems with Deferral Costs, *Management Sci.*, 11, No. 2, 280-288 (1964).
74. Held M., Karp R. M., The Traveling Salesman Problem and Minimum Spanning Trees, *Operations Res.*, 18, No. 6, 1138-1162 (1970).
75. Held M., Karp R. M., The Traveling Salesman Problem and Minimum Spanning Trees, *Math. Prog.*, 1, No. 1, 6-25 (1971).
76. Little J. D. C., Murty K. G., Sweeney D. W., Karel C., An Algorithm for the Traveling Salesman Problem, *Operations Res.*, 11, No. 6, 972-989 (1963). [Имеется перевод: Литл Дж. и др. Алгоритм для решения задачи о коммивояжере.— Экономика и математические методы, 1965, т. 1, № 1.]
77. Lomnicki Z. A., A 'Branch-Bound' Algorithm for the Exact Solution of the Three-Machine Scheduling Problem, *Operational Res. Quart.*, 16, No. 1, 89-100 (1965).
78. Madigan J. G., Scheduling a Multi-Project Single Machine System for an Infinite Planning Period, *Management Sci.*, 14, No. 11, 713-719 (1968).
79. Magee J. F., Production Planning and Inventory Control, McGraw-Hill, N.Y., 1958, pp. 44-66.
80. Manne A. S., On the Job-Shop Scheduling Problem, *Operations Res.*, 8, No. 2, 219-223 (1960).
81. McMahon G. B., Burton P. G., Flow-Shop Scheduling with the Branch-and-Bound Method, *Operations Res.*, 15, No. 3, 473-481 (1967).
82. McNaughton R., Scheduling with Deadlines and Loss Functions, *Management Sci.*, 6, No. 1, 1-12 (1959).
83. Miller L. W., Schaage L., The Queue M/G/1 with the Shortest Remaining Processing Time Discipline; RAND Paper P-3263, 1965.
84. Miller R. G., Jr., Priority Queues, *Ann. Math. Stat.*, 31, No. 1, 86-103 (1960).
85. Mitten L. G., A Scheduling Problem, *J. AIEE*, 10, No. 2, 131-135 (1959).
86. Moore J. M., An N Job One Sequencing Algorithm for Minimizing the Number of Late Jobs, *Management Sci.*, 15, No. 1, 102-109 (1968).
87. Nabeshima I., On the Bound of Makespans and Its Application in M Machine Scheduling Problem, *J. Operations Res. Soc. Japan*, 9, No. 3 and 4, 6-44 (1967).
88. Palmer D. S., Sequencing Jobs Through a Multi-Stage Process in the Minimum Total Times, *Operational Res. Quart.*, 16, No. 1, 101-108 (1965).
89. Phipps T. E., Jr., Machine Repair as a Priority Waiting-Line Problem, *Operations Res.*, 4, No. 1, 76-86 (1956).
90. Reich E., Waiting Times When Queues are in Tandem, *Ann. Math. Stat.*, 68, No. 3, 768-772 (1957).
91. Rogers J., A Computational Approach to the Economic-Lot Scheduling Problem, *Management Sci.*, 4, No. 3, 264-291 (1958).
92. Root J. G., Scheduling with Deadlines and Loss Functions in k Parallel Machines, *Management Sci.*, 11, No. 3, 460-475 (1965).
93. Rothkopf M., Scheduling Independent Tasks on Parallel Processors, *Management Sci.*, 12, No. 5, 437-477 (1966).

94. Saaty T. L., Elements of Queuencing Theory, McGraw-Hill Book Company, Inc., N.Y., 1961. [Имеется перевод: Саати Т. Элементы теории массового обслуживания и ее приложения.—М.: Сов. радио, 1971.]
95. Salvador M. S., A Solution to a Special Class of Flow-Shop Scheduling Problems, Proceedings of the Symposium on the Theory of Scheduling and Its Applications, Springer-Verlag, Berlin, 1973, pp. 83—91.
96. Schild A., Fredman I. J., Scheduling Tasks with Linear Loss Functions, *Management Sci.*, 7, No. 3, 280—285 (1971).
97. Schild A., Fredman I. J., Scheduling Tasks with Deadlines and Non-linear Loss Functions, *Management Sci.*, 9, No. 1, 73—81 (1962).
98. Schrage L. E., A Survey of Priority Queuing, Master's Thesis, Cornell University, 1965.
99. Schrage L. E., Solving Resource-Constrained Network Problems by Implicit Enumeration—Nonpreemptive Case, *Operations Res.*, 18, No. 2, 263—278 (1970).
100. Schwimmer J., On the N-Job, One-Machine, Sequence-Dependent Problem with Tardiness Penalties: a Branch-and-Bound Approach, *Management Sci.*, 18, No. 6, B-301—B-313 (1972).
101. Smith R. D., Dudak R. A., A General Algorithm for Solution of the n-Job m-Machine Sequencing Problem of Flow-Shop, *Operations Res.*, 15, No. 1, 71—82 (1967).
102. Smith W. E., Various Optimizers for Single-State Production, *Nav. Res. Log. Quart.*, 3, No. 1, 59—66 (1956).
103. Stankard M. F., Gupta S. K., A Note on Bomberger's Approach to Lot Size Scheduling: Heuristic Proposed, *Management Sci.*, 15, No. 7, 449—452 (1969).
104. Swarc W., Solution of the Akers-Friedman Scheduling Problem, *Operations Res.*, 8, No. 6, 782—788 (1960).
105. Ullman J. D., Polynomial Complete Scheduling Problems, *Operating Systems Review*, 7, No. 4, 96—101 (1973).
106. Wagner H. M., An Integer Linear-Programming Model for Machine Scheduling, *Nav. Res. Log. Quart.*, 6, No. 2, 131—140 (1959).
107. White H., Christic L. S., Queueing with Preemptive Priorities or Break-down, *Operations Res.*, 6, No. 1, 79—95 (1958).
108. Wishart D. M. G., Queueing Systems in Which the Discipline is Last-Come, First-Served, *Operations Res.*, 8, No. 5, 591—599 (1960).
109. Wismer D. A., Solutions of the Flowshop-Scheduling Problem with No Intermediate Queues, *Operations Res.*, 20, No. 3, 689—697 (1972).

I-8. Выбор проекта.

Планирование работ над проектом и руководство проектом

У. Саудер¹⁾

ВВЕДЕНИЕ

Любой проект — это комплекс взаимосвязанных работ, для выполнения которых выделяются соответствующие ресурсы и устанавливаются определенные сроки. В последние годы особое внимание уделяется проблеме эффективного руководства проектом, решение которой на данном этапе требует участия специалистов в самых различных областях науки и техники. На смену классическому ученому-одиночке, деятельность которого ограничивалась стенами лаборатории, пришел коллектив ученых и инженеров, причем работа такого коллектива должна планироваться с учетом деятельности других коллективов, подрядчиков и т. д.

Проект, в широком понимании этого термина, начинается свой путь в виде туманной, неясной идеи, которая затем приобретает вполне реальные очертания и оформляется в виде проекта, изделия и т. п. После этого проект, изделие и т. п. передается другому коллективу, выполняющему иные функции (например, потребителю, или поступает в сферу сбыта и т. д.) (рис. 1).

Эффективное руководство коллективом, работающим над проектом, и его окружением требует учета и оптимизации психологических, экономических, финансовых и организационных факторов. Некоторые из этих факторов могут быть проанализированы с помощью методов исследования операций. Так, например, Рубинштейн [40, 41] изучал генерацию (формирование) идей и возможность моделирования проектов. На основании полученных результатов он пришел к выводу, что качество генерируемых идей существенным образом зависит от системы поощрения и четкости формулировки целей проекта и ограничений. «Оптимальное» решение, однако, пока не найдено, и к моменту написания данной главы практически отсутствовала литература по использованию методов исследования операций для нахождения такого решения.

Этапы разработки проекта, связанные с оценкой сформированных идей (рис. 1), обычно касаются рациональной оценки полезности идей относительно установленных целевых критериев. Для оказания помощи руководителям в этом вопросе был разработан ряд моделей анализа, например модель Сетрона [10]. Поскольку, однако, оценка сгенерированных идей по своей природе имеет эвристический и качественный характер, определение оптимума в смысле классического исследования операций обычно невозможно.

Выбор проекта и распределение ресурсов включают определение того, какие из соответствующих идей следует принять за основу проекта. Процесс выбора обычно тесно связан с планированием работ над проектом, т. е. календарным планированием выполнения заданий и задач, предусмотренных реализацией проекта. Как будет показано, именно в этом заключается основной вклад системного анализа в данной области. Планирование работ над проектом предусматривает также детальное определение организационных

¹⁾ William E. Souder, Ph. D., Associate Professor of Industrial Engineering, Systems Management Engineering and Operations Research at the University of Pittsburgh, in Pittsburgh, Pennsylvania.

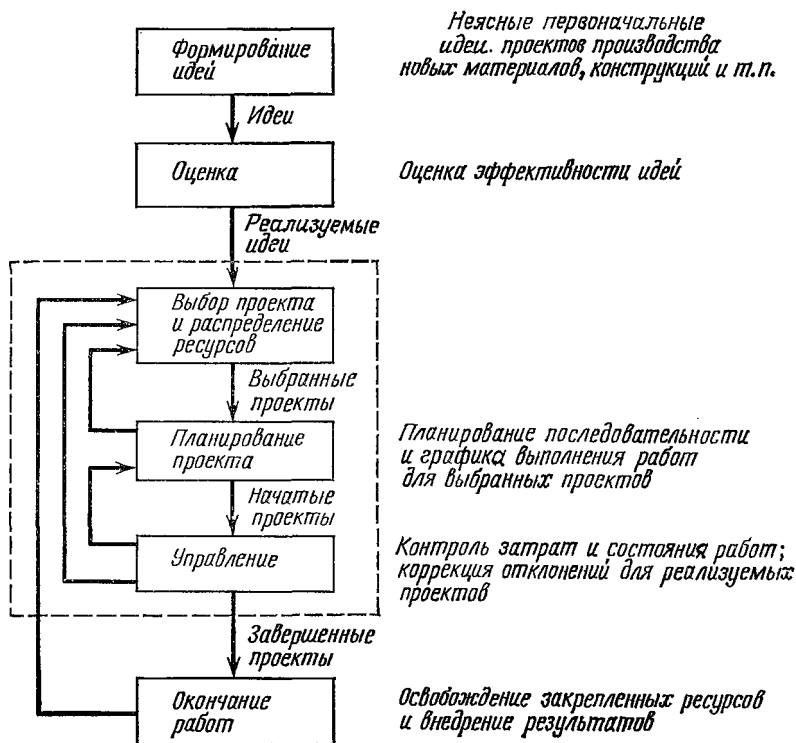


Рис. 1. Процесс создания проекта (штриховой линией отмечена область при-
менения методов исследования операций).

взаимоотношений в проектной организации (см., например, работу [15]).

Руководство проектом в процессе его осуществления состоит в сопоставлении текущих затрат и состояния проекта с запланированными затратами и состоянием и корректировке проекта при больших отклонениях (например, путем перепланирования или выбора нового варианта). Управление процессом выполнения проектом тесно связано с блоком планирования (рис. 1), который определяет затраты и сроки выполнения этапов работы, и с блоком выбора проекта и распределения ресурсов, который перераспределяет фонды между проектами в соответствии с соотношением состояния выполнения проекта и затрат и изменением этого соотношения во времени.

Этап окончания работ над проектом тесно связан с этапом выбора проекта, так как после завершения проекта становится возможным перераспределение ранее выделенных для него ресурсов. Этап внедрения технологии, т. е. передача результатов проекта отделу сбыта, потребителю и т. п., имеет, по существу, организационный характер, и поэтому использование для его анализа точных методов исследования операций связано с определенными трудностями [21].

В настоящее время методы исследования операций в сфере управления проектами используются в основном для решения трех проблем: выбора проекта, планирования выполнения работ над проектом и руководства проектом.

1. ЗАДАЧИ ВЫБОРА ПРОЕКТА

До второй мировой войны многие промышленные фирмы не считали выбор проекта делом особой важности. Вплоть до 50-х годов фирмы не придавали большого значения систематическому планированию капиталовложений в такие рискованные мероприятия, как научно-исследовательские и опытно-конструкторские работы, крупные инженерные разработки (например, в области гражданской авиации) и поиск полезных ископаемых (например, нефтеразведка).

Начало развития методов систематического планирования проектов научно-исследовательских и опытно-конструкторских работ относится к 1955 г. С этого момента термин «выбор проекта» приобретает самые различные значения.

В узком смысле термин «выбор проекта» означает определение того, какой из нескольких возможных проектов (задач, работ, программ и т. д.) должен быть начат, какой будет следующий и т. д., при условии что существуют проекты, которые могут быть начаты одновременно и претендуют на имеющиеся ресурсы. В самом широком смысле «выбор проекта» означает больше, чем выбор следующего проекта, который должен быть начат, или установление приоритетов для данного множества имеющихся проектов. Процесс выбора проекта может также включать определение «оптимального» (в некотором смысле) распределения ограниченных ресурсов между работами, ведущимися над проектами. Таким образом, задачу установления приоритетов проекта можно рассматривать как двоичную задачу (выбрать проект или нет) в общей задаче распределения ресурсов, цель которой — определить, сколько средств следует выделить на каждый проект.

Общий процесс выбора — распределения можно рассматривать как процесс последовательного заполнения портфеля заказов. Лицу, принимающему решение, необходимо знать, какие средства можно расходовать на каждый из нескольких возможных проектов в каждый из периодов времени. В конце каждого периода времени состав портфеля заказов изменяется с учетом проектов, которые в этот момент имеются. Множество имеющихся проектов состоит из проектов, которые в текущий момент выполняются, и проектов, которые находятся в резерве. Определение таких оптимальных проектов исключительно важно для организации, поскольку решения по портфелю заказов влекут за собой большие организационные обязательства, выполнение которых может быть связано со значительными затратами. Таким образом, важно знать, что будут выбраны только «наилучшие» проекты и средства для их выполнения распределены «наилучшим» из всех возможных способов. Однако оптимизация решений по выбору проектов в реальной ситуации исключительно сложна, поскольку для этих

решений часто характерно разделение ответственности за выбор проекта и распределение ресурсов и рассредоточение организационной ответственности за выбор проектов.

При разделении ответственности выбор проектов и установление их приоритетов производятся на одном уровне организации, тогда как ресурсы выделяются на другом уровне организации. Так как два организационных уровня могут иметь различные мнения относительно целей и множества ограничений, то это, естественно, препятствует достижению общего оптимума. Например, нет ничего необычного в том, что при выполнении функции долгосрочного планирования организация устанавливает перечень желательных программ, из которого производственные отделения могут выбирать частные проекты. Однако при определении выделяемых на эти проекты средств учитывается связь между общими бюджетами отделений, бюджетами отделов внутри этих отделений и бюджетами на разработку проектов в пределах отделений. Такая структура может быть представлена в виде

$$D^- \leq \sum_{j \in D} x_j \leq D^+, \quad d^- \leq \sum_{j \in d} x_j \leq d^+, \quad p_j^- \leq x_j \leq p_j^+,$$

где D^+ и D^- — верхний и нижний пределы бюджета отделения, d^+ и d^- — верхний и нижний пределы бюджетов отделов, p_j^+ и p_j^- — верхний и нижний пределы бюджетов на разработку проектов и x_j — величина затрат на j -й проект. Следовательно, фонды, выделяемые на разработку проекта, существенно зависят от ограничений на верхний предел бюджета. Кроме того, бюджеты для уровней отделения и отделов часто устанавливаются на весь год (исходя главным образом из численности отделов и отделений или на другой основе) без учета новых проектов, которые могут быть созданы в течение этого года.

Таким образом, решение о выборе проекта могут приниматься на различных уровнях организации и отличаться одно от другого. На стратегическом уровне, например на уровне планирования для всей фирмы, осуществляется выбор возможной стратегии поведения, или технологии, или вида выпускаемой продукции [52]. На тактическом уровне (например, на уровне отдела) подобный процесс выбора может осуществляться относительно способов выполнения проекта [51]. На обоих этих уровнях в процессе принятия решений по выбору проекта участвует большое число людей. Некоторые из них ответственны за обеспечение информацией, например отдел изучения сферы сбыта дает информацию о потенциальной прибыли от реализации данного проекта; другие ответственны за информацию об ограничениях.

В работах [5, 42] рассматривается схема ответственности при выборе проектов и разработке бюджетов в случае большой организации. Процесс разработки бюджета для такой организации,

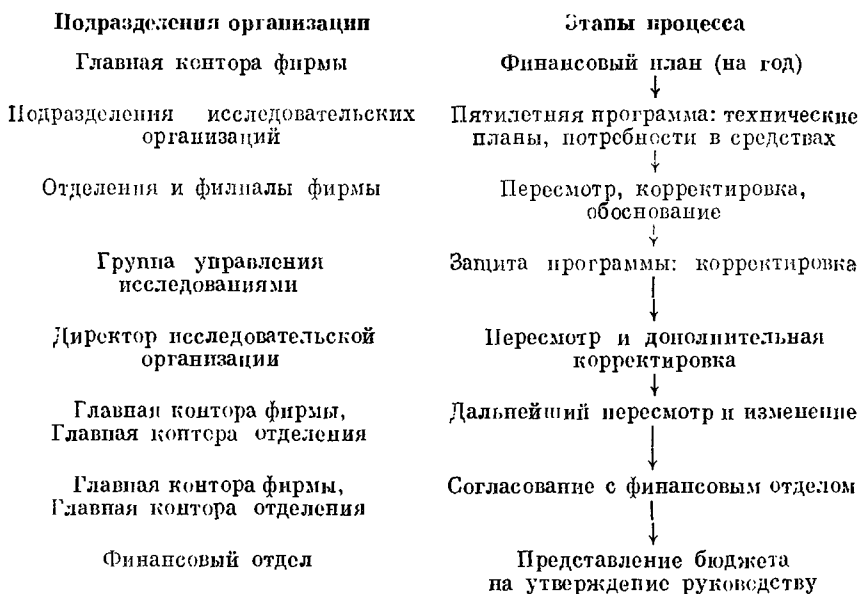


Рис. 2. Рассредоточенный процесс составления бюджета.

изображенный на рис. 2, имеет последовательный характер. Это означает, что указания по составлению бюджета передаются от каждого верхнего организационного уровня непосредственно на следующий, более низкий (подчиненный) уровень. Указания по бюджету выдаются одновременно с указаниями технического характера. Например, лаборатория получает указания относительно своего общего бюджета, программ и показателей работы. Лаборатория передает указания подчиненным ей подразделениям и своим проектным группам. Как только эти указания достигнут самого нижнего организационного уровня, направление потока информации меняется на обратное: каждый подчиненный излагает непосредственному руководителю свои соображения относительно распределения бюджета (если, конечно, ему это разрешено). Эти предложения на каждом уровне собираются и обрабатываются и затем передаются на следующий, более высокий уровень. Такой цикл передачи информации вверх — вниз — вверх может повторяться много раз, пока внешний уровень организации не получит такие предложения по распределению бюджета, которые либо будут соответствовать заранее заданным цифрам, либо будут приемлемы по другим причинам. Окончательно принятый бюджет затем передается в виде директив, имеющих содержание, аналогичное первоначально заданным показателям. Этот бюджет при-

ходится периодически пересматривать в течение года, чтобы учесть новые факторы, которые появились после его разработки. Соответственно, несколько раз в течение года каждый уровень организации сталкивается с необходимостью принимать решения по распределению ресурсов, для которых характерно большое число бюджетных ограничений, определяемых как характером соответствующего подразделения организации, так и особенностью его деятельности [42]. Эмпирическое исследование этого процесса для одного цикла выработки бюджета показывает, что большое число лиц на всех уровнях организации вносят многочисленные небольшие изменения в распределение бюджета. Это проявляется в том, что окончательный бюджет не преследует целей какого-либо одного из участвующих в этом процессе подразделений организации (отделения, филиала, агентства и т. н.).

2. МОДЕЛИ ВЫБОРА ПРОЕКТА

Назначение модели выбора проекта состоит в том, чтобы помочь его руководителю установить приоритет и осуществить окончательные подходящих проектов. Существует три основных типа «моделей» этого процесса, в основе которых лежат оценки экспертов, экономические показатели и объем распределения капиталовложений. Первые два типа моделей были разработаны и используются главным образом руководителями-практиками для решения задачи установления приоритетов проектов; что касается моделей распределения капиталовложений, то они в основном являются прерогативой исследователей систем и используются для решения задачи оптимального распределения средств.

2.1. Экспертные оценки

Если проекты имеют в основном исследовательский характер, как, например, проекты фундаментальных научных исследований, то, как правило, мы располагаем только качественной информацией и управляющие переменные могут иметь большую степень неопределенности. В этом случае предлагаемые проекты часто принимаются или отвергаются на основании мнений, высказанных несколькими учеными или другими компетентными лицами, работающими в данной области техники [25, 26, 35, 60]. Их точка зрения относительно использования набора критериев или характеристик, которым должен удовлетворять принимаемый проект, оказывается полезной при принятии решений.

Перечень некоторых часто используемых критериев, а также показатели их эффективности (или значимости) для гипотетического проекта приведены в табл. 1. По этим оценкам (показателям)

Таблица 1

Экспертные оценки гипотетического проекта

Критерии	Оценка		
	высокая	средняя	низкая
Патентоспособность	×		
Возможность сбыта			×
Вероятность успешного выполнения		×	
Затраты на производство	×		
Затраты на исследования и разработки		×	

составляется систематизированная сводка, которая имеет первоочередное значение на стадии «оценки».

Если мы располагаем существенной информацией относительно сравнительной значимости критериев и проекты могут «соизмеряться» по крайней мере путем упорядочения, то в этом случае нетрудно перейти от оценок экспертов к разработке оценочных моделей. В моделях каждый из возможных вариантов проектов $j = 1, \dots, n$ оценивается в баллах по некоторой шкале для каждого из критериев $i = 1, \dots, m$. Эти балльные оценки критериев для каждого проекта комбинируются в соответствии с приписываемым каждому критерию весовым множителем W_i , и получается суммарная оценка T_j в баллах для каждого проекта. Проекты могут быть потом проранжированы в соответствии со значениями оценок T_j .

В качестве простой оценочной модели (табл. 2) можно, например, использовать

$$T_j = \sum_i s_{ij} W_i,$$

где s_{ij} — баллы i -го критерия для j -го проекта.

Процедура использования оценочных моделей состоит в следующем. Вначале, как правило, определяется график работ над проектом и составляется перечень критериев или факторов для оценки возможных проектов (для составления таких перечней автор работы [57] успешно применял метод Дельфи). Затем разрабатывается шкала оценок для характеристик проекта, по которым оцениваются возможные проекты. Для количественных критериев, таких, как затраты, прибыль и т. п., шкала оценок может разрабатываться либо на основе данных прошлого опыта, либо на основе ожидаемых в будущем характеристик. В качестве примера рассмотрим критерий «стоимость проекта» [2].

горий был принят чисто статистический подход: средние показатели, число которых составляет 213 от общего числа данных (оценка равна 2), приняты за «типичные». Выбор принципа выделения категорий большей частью осуществляется потребителями и определяется в соответствии с их требованиями [35]. Для качественных критериев, например вероятности создания успешного проекта, используются качественные оценки, такие, как «низкая», «средняя» и т. д., которые преобразуются в баллы, например 1, 2 и т. д.

Таким образом, оценочные модели позволяют комбинировать качественные и количественные показатели.

2.2. Экономические показатели

Установление приоритета возможных вариантов проекта на стадиях «оценка» и «выбор проекта и распределение ресурсов» (рис. 1) осуществляется на основе различных экономических показателей [31]. Рассмотрим некоторые из них.

Показатель Энсофа [1], или

$$\text{Показатель качества проекта} = \frac{r \, dp \, (T + B) \, E^*}{\text{Суммарные капиталовложения}},$$

где r — вероятность успешного завершения работ над проектом, d — вероятность успешного внедрения, p — вероятность успешной реализации, T и B — технические и экономические показатели и E^* — приведенная величина дохода от проекта.

Показатель Ольсена [см. дополнительную литературу в конце главы]:

$$\text{Значимость проекта} = \frac{r \, dp \, S \, P \, n}{\text{Стоимость проекта}},$$

где S — годовой объем продажи продукции (изделия), P — доход от реализации единицы продукции (изделия), n — долговечность проекта (изделия) в годах, остальные переменные — как и в случае показателя Энсофа.

Показатель Харта [27]:

$$\text{Возврат капитала} = pG^* / [(R^*) + (D^*) + (F^*) + W],$$

где G^* — приведенная величина валовой прибыли, R^* — приведенные прямые затраты на исследовательские работы, D^* — приведенные прямые затраты на внедрение, F^* — приведенные прямые затраты основного капитала, W — оборотный капитал, p определено выше.

Показатель Виллера [61]:

$$\text{Индекс проекта} = r dp \frac{E^* - R^*}{\text{Затраты}},$$

где условные обозначения те же, что и выше.

Показатель Дисмана [20]:

$$\text{Оправданные максимальные капиталовложения} = r p (v^* - X^*),$$

где v^* — приведенный доход от проекта в долл., X^* — приведенные затраты на производство, сбыт и техническое обслуживание, параметры r и p указаны выше.

Показатель Дипа и Сепгупта [47]:

$$V = \sum_i^n [c_i (1+r)^{-i}],$$

где V — приведенная мера возможности проведения исследований, c_i — движение чистой денежной наличности в i -й год, r — ожидаемая годовая норма прибыли, i — индекс времени в годах и n — общее число лет, в течение которых ожидается прибыль.

Оценка значений c_i и r является субъективной и основывается на прошлом опыте и предполагаемом будущем компании. Этот показатель может быть использован также в случае переменной нормы прибыли.

Перечисленные показатели благодаря доступности понимания и простоте определения используются многими административными руководителями. Хотя подобные показатели позволяют оценить возможные варианты проекта при более чем одном уровне финансирования, оценка нескольких проектов для нескольких возможных уровней ассигнований вручную становится трудоемкой. Поэтому модели экономических показателей используются только для быстрого установления приоритетности конкурирующих проектов для одного уровня ассигнований и дают весьма ограниченную информацию о «наилучшей» приоритетности проектов. Если пользователь не учитывает, что полученные приоритеты относятся лишь к одному из многих возможных уровней ассигнования, то существует вероятность того, что установленные приоритеты будут распространены на все уровни финансирования проектов.

2.3. Модели распределения капиталовложений

В моделях распределения капиталовложений устанавливается неявная приоритетность возможных вариантов проекта в соответствии с размерами выделяемых средств. В общем виде задача может быть сформулирована следующим образом:

найти

$$\max \sum_j v_j(x_j) \quad (1)$$

при ограничении

$$\sum_j x_j \leq B, \quad (2)$$

где x_j — затраты на проект, B — величина общего бюджета для всех возможных проектов $j = 1, \dots, n$ и $v_j(x_j)$ — целевая функция, которая может быть нелинейной, линейной либо однозначной. В последнем случае (т. е. при одном постоянном значении v_j и одной величине затрат x_j для каждого из j -го проекта) модель распределения капиталовложений становится моделью экономических показателей, причем оценки затрат v_j являются показателями приоритета.

Поскольку в выражении (1) могут использоваться разнообразные показатели ценности, то в случае ожидаемых значений показателей ценности оно принимает вид

$$\max \sum_j v_j p_j(x_j),$$

где $p_j(x_j)$ — вероятность получения оценки v_j . Другие модели используют суммарные оценки, выраженные в баллах, например значение T_j из оценочной модели. Помимо ограничения (2) обычно вводится еще ограничение вида

$$b_j^- \leq x_j \leq b_j^+,$$

где b_j^- и b_j^+ — нижний и верхний пределы затрат на проект. Если рассматривается несколько периодов времени, то задача распределения капиталовложений может быть сформулирована, например, следующим образом:

найти

$$\max \sum_{ij} v_{ij}(x_{ij})$$

при ограничении

$$\sum_{ij} x_{ij} \leq B,$$

где $i = 1, \dots, m$ — периоды времени.

В работе [18] анализируется возможность использования математического программирования в случае оценочной модели. Согласно такому подходу, необходимо найти такой набор x_j ($x_j = 0$ или 1) для каждого из n проектов, при котором достигается

$$\max \sum_j x_j T_j$$

при ограничении

$$\sum_j x_j R_j \leq R,$$

где R_j — ресурсы, выделенные j -му проекту, R — общее количество ресурсов; значение x_j (переменная выбора) зависит от того, выбран ($x_j = 1$) или нет ($x_j = 0$) j -й проект. Для решения этой задачи используются методы целочисленного программирования.

В работе [7] дается модель линейного программирования, которая может быть модифицирована в соответствии со скоростью выполнения проекта, например большой, средней, малой. i -й вариант j -го проекта определяется булевой переменной x_{ij} . Чтобы каждый проект выбирался не более чем один раз, вводят дополнительное ограничение

$$\sum_{i=1}^{m_j} x_{ij} \leq 1$$

для возможных вариантов проекта $j = 1, \dots, n$, составляющих портфель проектов, из которых производится отбор и для которых распределяются средства; m_j — число вариантов j -го проекта. В этом случае задача сводится к пахождению

$$\max \sum_{j=1}^n \sum_{i=1}^{m_j} b_{ij} x_{ij},$$

где b_{ij} — ожидаемая прибыль от i -го варианта j -го проекта.

Уоттерс [62] использовал метод целочисленного программирования с булевыми переменными для максимизации суммарной ожидаемой полезности от всех проектов.

$$E(u) = \sum_j (\mu_j - K\sigma_j^2) x_j, \quad j = 1, \dots, n; \quad x_j = 0, 1,$$

где μ_j и σ_j^2 — математическое ожидание и дисперсия чистой прибыли от j -го проекта и K — коэффициент риска. Максимизация производится при ограничениях на размер финансирования для каждого периода, согласно которым вероятность превышения предела B_q не должна быть больше заданной величины β_q , т. е.

$$P\{t_q > B_q\} \leq \beta_q,$$

где $q = 1, 2, \dots, Q$ — планируемые периоды и t_q — переменная, характеризующая общие затраты на все проекты, выполняемые в q -й период.

Модель Розена — Саудера [39] использует динамическое программирование для максимизации чистой прибыли, получаемой

от научных исследований и разработок, т. е. задача сводится к нахождению

$$\max \sum_j [G_j P_j(x_j) - x_j]$$

при ограничении

$$\sum_j x_j \leq B, \quad j = 1, 2, \dots, n,$$

где G_j — приведенная валовая прибыль в случае успешного завершения проекта, $P_j(x_j)$ — вероятность успеха, B — общий бюджет и x_j — капиталовложения в j -й проект. При этом вводятся следующие дополнительные ограничения:

$S_{\max}^j$ = Максимальные общие затраты на проект,

$S_{\min}^j$ = Минимальные общие затраты на проект,

x_j = Затраты в любой период времени выполнения проекта.

Рекуррентное уравнение имеет вид

$$f_2^j(y) = \max [G_j P_j(x_j) - x_j + \rho(1 - P_j(x_j)) f_1^j(x_j + y)], \quad x_j \geq 0,$$

где $f_2^j(y)$ — максимальная ожидаемая прибыль от j -го проекта, осуществляемого за два периода времени, и ρ — коэффициент дисконтирования. Следует заметить, что подобная модель динамического программирования была предложена еще в работе [28], автор которой рассмотрел функцию вида $P_n(x_n, y_n)$, где x_n — бюджет проекта в n -й период и y_n — взвешенная сумма затрат на исследования и разработку проекта за периоды, предшествующие n -му.

Согласно литературным данным, было разработано большое число моделей распределения капиталовложений. В 1964 г. был опубликован обзор работ по таким моделям и приводились результаты анализа 10 моделей. В 1967 г. Сетрон и др. сообщили о результатах своих исследований 30 подобных моделей, а в 1970 г. Саудер [51] провел анализ 41 модели. Они отличаются друг от друга максимизируемыми переменными, а также параметрами и математическими методами, которые используются для описания реальных систем и могут быть классифицированы как линейные, нелинейные, целочисленные, оценочные модели прибыльности и модели полезности. Линейные и нелинейные модели имеют соответственно линейные и нелинейные целевые функции. Целочисленные модели используют целочисленные переменные и методы целочисленного программирования, оценочные модели — дискретные оценки, модели прибыльности максимизируют единый пока-

затель прибыльности, модели полезности максимизируют субъективную полезность, получаемую от вложенных затрат. Перечень работ, посвященных рассмотрению перечисленных видов моделей, приведен в конце данной главы.

Таблица 3

Критерии эффективности моделей и их показатели

Критерий	Показатели
Реалистичность	Многокритериальность Множественность ограничений Ортогональность переменных Учет торгового риска Учет технического риска Наличие ограничений по трудовым ресурсам Наличие ограничений по оборудованию Учет неопределенности данных Учет неопределенности исходных посылок
Гибкость	Применимость к прикладным проблемам Применимость к фундаментальным исследованиям Возможность устанавливать приоритетность проектов Применимость к выработке решения о начале разработки Применимость к задачам распределения бюджета Применимость к планированию размера затрат на проект
Диапазон применимости	Применимость для прогнозирования хода проекта Применимость для оптимизации Применимость для моделирования Применимость для выбора проектов Применимость для распределения ресурсов Применимость для календарного планирования
Простота использования	Низкая чувствительность к точности исходных данных Ориентация на существующий (привычный) набор данных Дискретность переменных Отсутствие необходимости в использовании ЭВМ Отсутствие необходимости привлечения специалистов Отсутствие необходимости специальных средств и методов интерпретации результатов Небольшой объем исходных данных Легкость получения исходных данных Легкость оценки параметров
Экономичность	Низкая стоимость настройки Низкая стоимость использования Низкие затраты на персонал Небольшая потребность в машинном времени Малые затраты на сбор исходных данных

Оценки критерия реалистичности по его

Автор модели	Значения оценок для				
	Многокритериальность	Множественность ограничений	Ортогональность переменных	Риск, связанный с рынком	Технический риск
Эшер	0	1	1	0	1
Хансмайн	0	1	1	0	1
Натт	0	1	0	0	1
Паунд	0	0	1	0	1
Собин	0	0	1	0	1
Саудер	0	0	0	1	1
w_{ij}	10	9	8	7	6

2.4. Оценка моделей

В работах [53 и 54] на основе методологии оценочных моделей была произведена оценка рассмотренных моделей с точки зрения их применимости. В принятой методике c_{ij} обозначает i -й показатель j -го критерия качества (табл. 3), w_{ij} — весовой множитель (или относительную значимость показателя c_{ij}), где $w_{ij} = 1, 2, \dots, \lambda$, x_{ij} — переменную, принимающую по определению значение 1, если данная модель обладает i -м показателем j -го критерия, и значение 0 в противном случае [53]. Тогда абсолютная оценка качества модели по j -му критерию

$$S_{1j} = \sum_i w_{ij} x_{ij}$$

и относительная оценка качества модели по j -му критерию

$$S_{2j} = S_{1j} / \sum_i w_{ij}$$

Обозначая относительную ценность j -го критерия качества через r_j , получим

$$S_j = r_j S_{2j},$$

где S_j — взвешенная оценка качества модели по j -му критерию. Таким образом, общая оценка применимости модели может быть выражена как

$$T = \sum_j [(\sum_i w_{ij} x_{ij} / \sum_i w_{ij}) r_j].$$

Таблица 4

показателям для шести линейных моделей

показателей x_{ij}						
Ограниченность трудовых ресурсов	Ограниченность оборудования	Ограниченность бюджета	Неопределенность данных	Неопределенность посылок	s_{1j}	s_{2j}
1	0	0	0	0	28	0,51
0	1	0	0	0	27	0,49
1	1	1	0	0	27	0,49
0	0	1	0	0	17	0,31
1	1	1	1	0	28	0,51
0	1	0	0	0	17	0,31
5	4	3	2	1	$\sum_i w_{ij} = 55$	

Для критериев, приведенных в табл. 3, весовым множителям были приписаны следующие значения: реалистичности — 4, гибкости — 3, диапазону применимости и простоты использования — по 2, экономичности — 1. Критерии, их показатели и значения r_j выбирались эмпирически.

В табл. 4 и 5 приведены результаты, полученные для одного класса моделей. Общая оценка применимости T была определена для каждой из рассмотренных 41 модели и для всех классов моделей. Оказалось, что линейные, нелинейные и целочисленные модели имеют более высокие показатели реалистичности, гибкости и диапазона применимости, тогда как оценочные модели и модели прибыльности имеют лучшие оценки простоты использования

Таблица 5

Оценка критериев применимости линейных моделей

Автор модели	Реалистичность	Гибкость	Диапазон применимости	Простота применения	Экономичность	Суммарная оценка
Эшер	2,0	2,6	1,0	0,88	0,66	7,1
Хансманн	2,0	1,7	1,3	0,88	0,60	6,5
Натт	2,0	2,6	1,3	0,88	0,20	6,9
Паунд	1,2	1,7	0,66	1,5	1,0	6,1
Собин	2,0	2,1	1,0	1,3	1,0	7,4
Саудер	1,2	2,6	1,0	1,3	0,60	6,7

и экономичности. За исключением модели полезности, нелинейные модели имеют самый низкий показатель простоты использования и наихудшее значение показателя экономичности.

2.5. Применение моделей распределения капиталовложений

Модели распределения капиталовложений, по-видимому, необходимы для руководящих органов, ответственных за распределение находящихся в их распоряжении ограниченных ресурсов между различными имеющимися проектами. Целью подобного распределения является максимизация общей эффективности используемых ресурсов. Решения по распределению средств обычно приводят к тому, что некоторые проекты не финансируются и, следовательно, пополняют портфель невыполненных проектов, которые, возможно, будут финансироваться в один из будущих периодов. С течением времени у руководителя может неоднократно возникать желание пересмотреть распределение ресурсов между имеющимися проектами (как находящимися в стадии выполнения, так и входящими в портфель невыполненных проектов) с учетом изменившихся условий и поступившей новой информации. Модели распределения капиталовложений могут быть использованы в процессе заполнения портфеля проектов для принятия решения относительно того, какая часть общего годового бюджета на следующий финансовый год должна быть выделена каждому из имеющихся в данный момент проектов и как следует перераспределить ресурсы между всеми имеющимися проектами на следующий финансовый год с учетом поступления новой информации [56].

Автор работы [56], исследуя аналитические методы оценки эффективности четырех различных типов моделей выбора проектов, показал, что выбор типа модели зависит от целей, которые преследует руководитель проекта, и состояния имеющихся проектов. Эффективность модели в данном случае определяется качеством формируемого с ее помощью портфеля заказов (проектов). Если целью является максимизация ожидаемой прибыли, то могут быть использованы как линейные, так и нелинейные модели. Если целью является максимизация ожидаемой прибыли с учетом затрат на близкие к неудачному завершению проекты, то более целесообразно применять целочисленные модели или модели оценки полезности. Когда целью являлась максимизация ожидаемой прибыли и требуется обеспечить оптимальность уже заполненного портфеля проектов (например, с помощью перенесения средств с потенциально неудачных проектов на успешные), результаты показывают, что полезными при принятии решений могут быть линейные, целочисленные модели или модели оценки полезности. Практика показывает, что высокую эффективность имеют линейные модели, так как они дают более низкие ожидаемые затраты на

выполнение неудачных работ [57]. Тем не менее все семь участников исследования отвергли модель линейного типа, поскольку получаемые с ее помощью цифры являются «чрезмерными», т. е. значения прибыли, достигаемые с помощью этой модели, лежат за пределами приемлемого уровня, который обеспечивается без применения модели. Кроме того, составленные с помощью модели линейного типа портфели проектов не удовлетворяли менеджеров из-за некоторых присущих этим портфелям свойств.

Таким образом, результаты работ [56 и 57] свидетельствуют о том, что, хотя некоторые типы моделей являются, по-видимому, пригодными, эффективность их использования при распределении капиталовложений, выделенных на научные исследования и разработки, может быть невелика. Этот вывод согласуется с данными других работ, например [4, 50, 54], в которых показано, что модели распределения капиталовложений не имеют пока широкого применения. Несколько сообщений об их применении (например, работы [20, 39, 46 и 55]) показывают, что после короткого периода использования эти модели были заброшены ввиду отсутствия интереса к ним со стороны менеджеров. Однако несомненно, что в рассмотренных случаях использование моделей оказывалось весьма эффективным, в результате достигалась лучшая взаимосвязь между отделами, создавалась обстановка, способствующая анализу последствий принимаемых решений и уменьшению степени неопределенности при принятии решений [43—45]. Тем не менее проблема разработки аналитической модели распределения капиталовложений, которая была бы эффективной (т. е. позволяла формировать «наилучший» портфель проектов по нескольким критериям) и в то же время реалистичной и приемлемой для организации, в настоящее время еще не решена [6, 47, 56, 57].

3. СЕТЕВОЕ ПЛАНИРОВАНИЕ

После принятия решения о начале работ над проектом обычно необходимо решить задачу завершения проекта (составить график выполнения) за заданное время с использованием выделенных ресурсов. Для решения этой задачи в период 1956—1958 гг. были разработаны два метода. Один из них — метод критического пути, или МКП (Critical Path Method — CPM), — впервые был использован компанией DuPont Co. и получил дальнейшее развитие в работах фирмы Mauchly Associates [30, 32]. Другой метод — метод оценки и пересмотра проектов, или ПЕРТ (Project Evaluation and Review Technique — PERT), — был разработан для министерства военно-морских сил США в соответствии с программой создания подводных лодок, оснащенных ракетами «Поларис» [8, 23]. Характерным для этих методов является изображение проекта

в виде сети взаимосвязанных работ. В настоящее время создано большое число модификаций и второе поколение сетевых методов [32].

3.1. Представление проекта в виде сети

Сетевое планирование начинается с составления перечня работ и оценок их продолжительности. При этом работы изображаются стрелками, направление которых указывает продвижение ра-

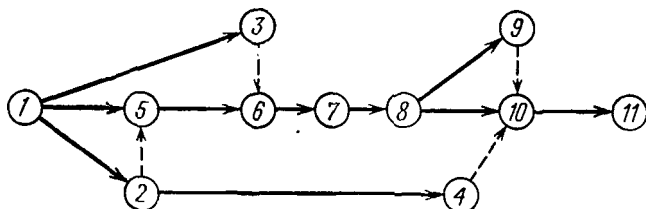
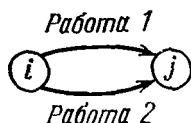


Рис. 4. Сетевой график проекта А.

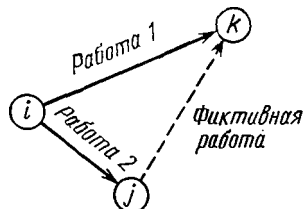
бот по проекту (рис. 4). События, соответствующие началу и завершению работ (или моменты времени), изображаются в виде узлов сети, которые нумеруются соответствующим образом.

Основные правила построения сети:

1. Никакие две работы не могут быть идентифицированы одними и теми же событиями. Это означает, что участок сети вида

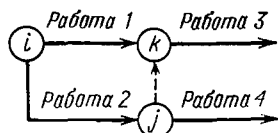


неверно отображает две одновременно завершающиеся работы. В такой ситуации участок сети должен иметь вид

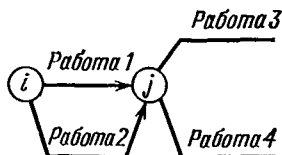


Фиктивная работа не требует ни времени, ни ресурсов; она вводится только для целей однозначности событий, связанных с

завершением работ. Такой прием используется в ситуациях, когда работы 3 и 4 должны следовать за работой 2, но работа 1 не обязательно должна предшествовать работе 4, т. е.

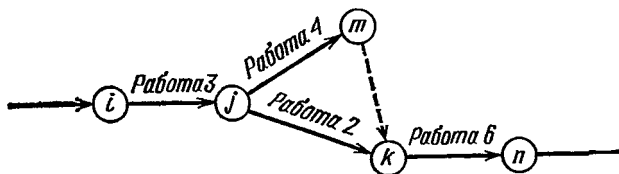


Из участка сети

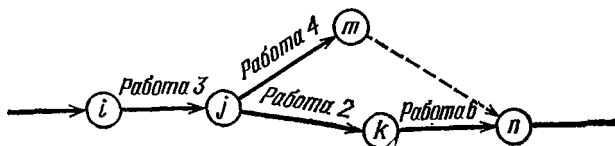


следует, что в случае, когда работы 1 и 2 имеют различное время окончания, они обе должны завершиться прежде, чем смогут начаться работы 3 и 4.

2. Соотношение предшествования — следования должны соблюдаться на всем протяжении сети. Предположим, например, что работа 6 следует за работами 4 и 2, которые в свою очередь следуют за работой 3. Тогда участок сети



является правильным только в том случае, если работа 4 будет завершена прежде, чем может начаться работа 6. Если требование состоит лишь в том, что и работа 4, и работа 6 заканчиваются прежде, чем может начаться последующая работа, то данный участок сети будет иметь вид



Эти два правила иллюстрируются с помощью сети, представленной на рис. 4. Работа 1-3 (общепринято обозначать работу номерами событий, соответствующих ее началу и концу) должна закончиться прежде, чем начнется работа 6-7, работа 1-2 должна закончиться прежде, чем начнется работа 5-6, и т. д. Когда момент окончания предшествующей работы совпадает с моментом начала последующей, они считаются одним и тем же событием.

3.2. Критический путь

Критическими считаются работы, задержка которых приводит к эквивалентной задержке окончания всего проекта. Путь через сеть, который состоит целиком из таких работ,— это критический путь.

В небольших сетях критический путь может быть легко определен, если заданы времена наступления всех событий и все работы характеризуются самым ранним (допустимым) временем начала.

В более крупных системах критический путь приходится определять как путь с нулевым резервом времени. Резерв времени — это количество времени, в течение которого работа может затягиваться, не вызывая увеличения времени наступления события окончания проекта (на рис. 5 это время равно 34 ед. времени). Например, работа 1-2 имеет резерв времени, равный 8 ед. времени, работа 1-3—14 ед. времени и т. д. Работа 2-4 имеет резерв времени 22 ед., если работа 1-3 выполняется без задержки.

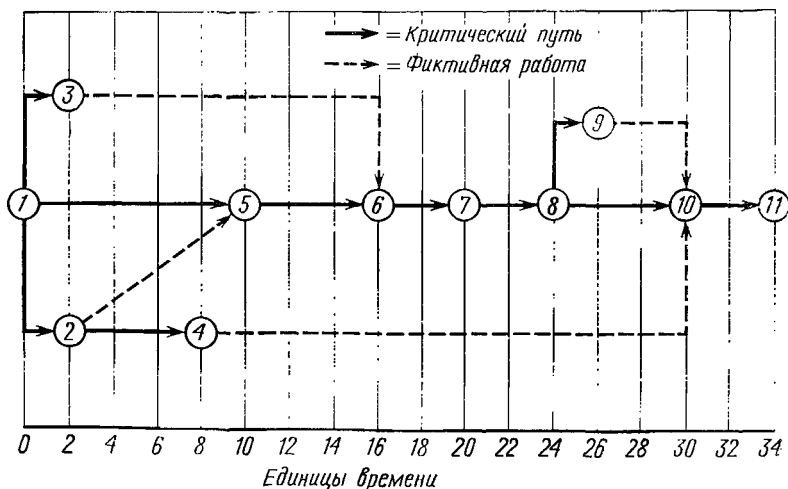


Рис. 5. Сетевой график проекта А (во времени).

Однако, если резерв времени работ 1-2 достигнет максимально возможной величины 8 ед. времени (резерв определяется событием 5), работа 2-4 будет иметь резерв только 14 ед.

Чтобы вычислить резервное время, следует сначала выполнить просчет сети от начала до конца и таким образом получить самые ранние времена начала каждой работы в каждом узле сети. Затем, чтобы определить время самого позднего окончания каждой работы, осуществляют просчет сети в обратном направлении — от конца к началу. Общий резерв времени для каждой работы находится как разность между располагаемым временем и длительностью работы. В частности, вычисления при прямом проходе сети производятся по формуле

$$ES_i = \max_k \{ES_k + t_{ki}\}, \quad i = 2, \dots, n, \quad (3)$$

где i — номер узла, t_{ki} — продолжительность k — 1-й работы, $ES_1 = 0$ и ES_i обозначает самое раннее время начала для всех работ, которым предшествует i -й узел. Например, из уравнения (3) следует, что нужно взять максимальную из величин $\{(24 + 6), (24 + 2 + 0), (8 + 0)\}$, чтобы получить значение $ES_{10} = 30$ (рис. 5), т. е. максимальная длина пути равна либо $ES_8 + t_{8,10}$, либо $ES_8 + t_{8,9} + t_{9,10}$, либо $ES_4 + t_{4,10}$, (заметим, что длительности работ $t_{9,10}$ и $t_{4,10}$ равны нулю). Аналогично производятся вычисления при обратном проходе:

$$LF_j = \min_k \{LF_k - t_{jk}\}, \quad j = 1, 2, \dots, n-1, \quad (4)$$

где j — узлы, LF_j — самое позднее (допустимое) время окончания всех работ, заканчивающихся в j -м узле, и $LF_n = ES_n$ для узла сети, соответствующего завершению проекта. Вычисление общего резерва времени производится по формуле

$$TF_{ij} = LF_j - ES_i - t_{ij}. \quad (5)$$

По определению, $LS_{ij} = LF_j - t_{ij}$, $EF_{ij} = ES_i + t_{ij}$, где LS_{ij} — самое позднее время начала и EF_{ij} — самое раннее время окончания i — j -й работы и, следовательно, равенству (5) эквивалентны равенства

$$TF_{ij} = LS_{ij} - ES_i, \quad (6)$$

$$TF_{ij} = LF_j - EF_{ij}. \quad (7)$$

Результаты прямого и обратного просчетов, а также резервное время для сети проекта *A* приведены в табл. 6 и могут быть легко получены с помощью рис. 5. Заметим, что общее резервное время присваивается работам расчетным путем, однако оно не накапливается для последовательных работ, составляющих некоторый частный путь через сеть. Например, как видно из рис. 5,

Таблица 6

Временные характеристики сетевого графика для проекта А

Работа ¹⁾	Длительность работы	Самое раннее время		Самое позднее время		Общий резерв времени
		начало работы	окончание работы	начало работы	окончание работы	
1-3	2	0	2	14	16	14
1-5	10	0	10	0	10	0 ²⁾
1-2	2	0	2	8	10	8
2-4	6	2	8	24	30	22
5-6	6	10	16	10	16	0 ²⁾
6-7	4	16	20	16	20	0 ²⁾
7-8	4	20	24	20	24	0 ²⁾
8-9	2	24	26	28	30	4
8-10	6	24	30	24	30	0 ²⁾
10-11	4	30	34	30	34	0 ²⁾

¹⁾ Фиктивные работы не приведены.

²⁾ Эти работы, по определению, принадлежат критическому пути.

весь путь 1-2-4 не может иметь резерва времени, равного сумме резервов времени для работ 1-2 и 2-4 (которая равна 30 ед. времени), без задержки события 10. Как указывалось выше, если работа 1-2 использует какую-то часть из своего резервного времени в 8 ед., то резервное время для работы 2-4 соответственно уменьшается. Таким образом, работа 2-4 имеет $22 - 8 = 14$ ед. «свободного» резервного времени. Общее резервное время TF_{ij} получается при независимом рассмотрении каждой работы.

3.3. Процедуры выравнивания потребления, перераспределения ресурсов и нахождение компромиссного решения относительно времени выполнения и стоимости проекта

Резервное время играет важную роль при «выравнивании» потребления ресурсов. Выравнивание потребления ресурсов позволяет сгладить профиль функции использования ресурсов при сохранении заданной продолжительности проекта либо минимизировать длительность проекта при выполнении ограничений на ресурсы. Такая задача может возникнуть, когда накладываются ограничения на трудовые ресурсы, бюджет, квалификацию персонала и т. п.

Предположим, что для выполнения каждой из работ 1-3, 1-5, 1-2 и 2-4 (рис. 5) требуется один исполнитель, который будет занят полный рабочий день. Вместе с тем имеются только два исполнителя. Ограничение, накладываемое на трудовые ресурсы (два

исполнителя), может быть удовлетворено без продления времени завершения проекта, если задержать на 8 ед. времени начало работы 1-3 (оно может быть задержано до 14 ед. времени без задержки завершения проекта). Другой возможный вариант состоит в выполнении каждой из работ 1-3 и 1-2 с меньшей скоростью (например, соответствующей половине возможностей человека) при условии, что время их завершения увеличится вдвое. Имеются алгоритмы выравнивания потребления ресурсов, применимые для больших сетей [64]. Хотя эти алгоритмы называются программами оптимизации, в большинстве ситуаций возможно только эвристическое распределение ресурсов. Обзор методов выравнивания потребления ресурсов содержится в работах [14, 15, 19].

Не менее важная роль отводится резервному времени при сокращении длительности критического пути (путей). Предположим, что допускается перемещение исполнителей работ 1-3 и 1-2 на работу 1-5 (рис. 5). В результате этого произошла бы задержка работ 1-3 и 1-2, тогда как продолжительность работы 1-5 сократилась. Оптимальное решение достигается либо когда ни один исполнитель не может быть перемещен на другую работу, так как все резервное время ликвидировано, либо когда нельзя больше уменьшить критический путь за счет перераспределения ресурсов с работы, имеющей резерв времени, на критическую. Для этих целей имеются программы для ЭВМ, аналогичные тем, что используются для выравнивания ресурсов [15]. Однако подобные перемещения возможны только в рамках заданных ограничений на ресурсы, в противном случае возникает необходимость выравнивания ресурсов.

Единственный способ сокращения критического пути, после того как перераспределение всех ресурсов между работами, имеющими резервное время, и работами критического пути уже выполнено, заключается в использовании дополнительных ресурсов. Выполнение некоторых работ критического пути, по-видимому, может быть ускорено путем увеличения общего бюджета, например путем привлечения дополнительных ресурсов из внешних по отношению к данному проекту источников.

Предположим, что некоторая конкретная работа критического пути, которая в нормальных условиях стоит C_n и занимает t_n ед. времени, может быть выполнена за t_i времени при стоимости C_i , где $t_i < t_n$ и $C_i > C_n$. Используя приращения стоимости, т. е. $\Delta C = C_i - C_n$, можно получить кусочно-линейную аппроксимацию функции $C = f(t)$, которая обычно имеет наклон вправо (уменьшение t с увеличением C).

Предположим далее, что такие функции затраты — время определены для каждой работы критического пути сети. Сравнение этих функций по величине приращений $\Delta t = (t_n - t_i)$ и ΔC укажет порядок, в котором работы критического пути должны полу-

чать дополнительные ресурсы, чтобы достигнуть большего сокращения общей длительности проекта на единицу дополнительных затрат. Этой задаче уделялось основное внимание в первоначальных разработках МКП (Келли [30], Фалкерсон [24]). В этих первых постановках для работ критического пути рассматривались точки нормального (t_n , C_n) и аварийного (t_c , C_c) выполнения работ (последняя из упомянутых точек — это точка, по достижении которой дальнейшее сокращение работы невозможно). Затем для всех работ критического пути данной сети был вычислен «наклон» $(C_c - C_n)/(t_n - t_c)$, в соответствии с величиной которого они упорядочивались. Таким образом, анализ основывался на величине средней стоимости сокращения длительности работ на единицу времени. Эта процедура предполагает сокращение длительности тех работ, которые дают сокращение длительности проекта на единицу при минимальном увеличении общих прямых затрат. Поскольку сокращение длительности проекта приводит к уменьшению косвенных затрат на проект, этот процесс обычно продолжается до тех пор, пока сумма изменяющихся в противоположных направлениях прямых и косвенных затрат не достигнет минимума, который определяет оптимальный (относительно общих затрат) график выполнения проекта. Возможности использования подобной процедуры в реальных условиях, однако, довольно ограничены, поскольку она исходит из предположения о неограниченности имеющихся в наличии ресурсов.

Таблица 7

Показатели работ над проектом

Номер работы	Потребность в трудовых ресурсах	Нормальная длительность работы	$\Delta 1)$
1	2 чел.	6	4
2	2 чел. ²⁾	7	2)
3	1 чел.	5	2
4	Внешний отдел ³⁾	4	—
5	2 чел.	6	1
6	Субподрядчик ³⁾	4	—
7	Субподрядчик ³⁾	3	—
8	2 чел.	6	1
9	Субподрядчик ³⁾	2	—
10	2 чел.	2	3
11	2 чел.	5	2

1) Перемещать исполнителей можно только полностью; данные в этой графе показывают изменение длительности работы (увеличение или уменьшение) при увеличении или уменьшении числа исполнителей на одного человека.

2) Людей с этой работы перемещать нельзя.

3) Эти люди не подчиняются руководителю проекта.

Приведенная процедура по мере ее реализации становится довольно сложной, так как при уменьшении длительности образуется несколько критических путей. Математически оптимальные алгоритмы для этой задачи, разработанные Келли [30] и Фалкерсоном [24], основываются на интерпретации двойственной задачи линейного программирования как задачи о потоке на сети.

В табл. 7 приводятся данные для случая выравнивания потребности в трудовых ресурсах, который исследовался автором данной главы. Был построен сетевой график с певыравненным потреблением ресурсов, удовлетворявший все заинтересованные стороны (рис. 6). Однако, прежде чем пачались работы над проектом, возникли новые обстоятельства, вследствие которых в каждую единицу времени могли работать только четыре исполнителя. Потребность в трудовых ресурсах как функция времени для невыравненного графика показана в нижней части рис. 6. Приемлемым вариантом плана работ оказался график, представленный на

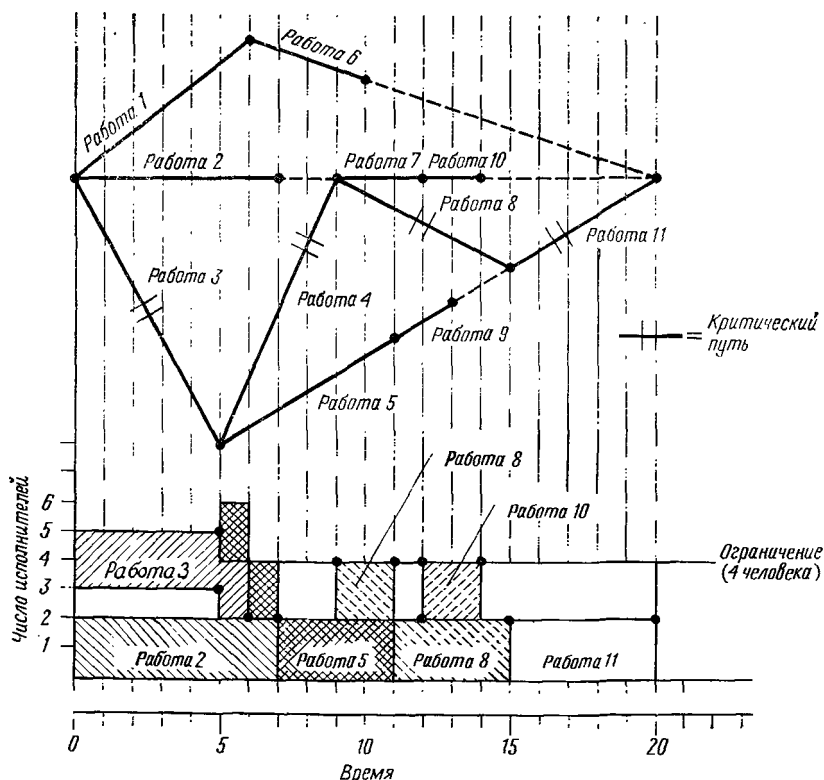


Рис. 6. Сетевой график проекта с невыравненной потребностью в ресурсах,

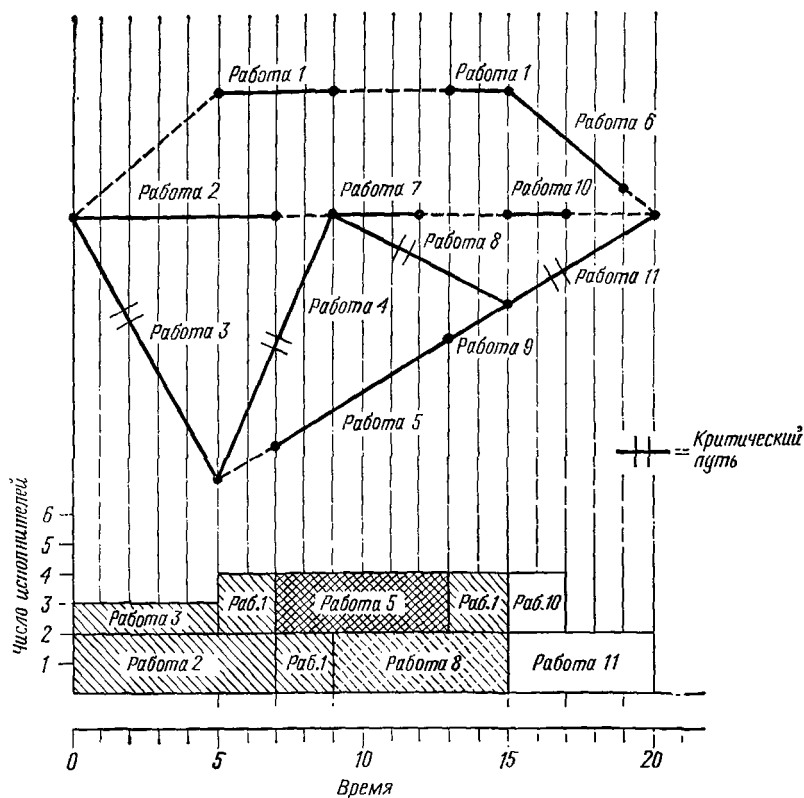


Рис. 7. Сетевой график проекта с выравненной потребностью в ресурсах.

рис. 7. Ограничение по трудовым ресурсам было удовлетворено за счет задержки начала работ 1, 5, 9 и 10 и разделения работы 1 на два этапа, которые выполнялись с разрывом в четыре недели. Таким образом, чтобы сократить длительность некоторых работ критического пути, не потребовалось никакого перераспределения людей между работами и не пришлось использовать данные о зависимости длительности выполнения работы от числа исполнителей, приведенные в табл. 7.

Сокращение общей длительности выполнения проекта, которая на рис. 7 равна 20 нед, возможно за счет дополнительных затрат, как это показано в табл. 8. Например, общая длительность проекта может быть сокращена до 17 нед, если затратить дополнительно 2000 долл. на работу 11 и 200 долл. на работу 8. Принимая во внимание затраты, такой вариант более эффективен, чем вариант, согласно которому для достижения такого же сокращения дли-

Таблица 8

Показатели «время — затраты»

Номер работы	Сокращение длительности выполнения работы, нед	Дополнительные затраты, долл.
6	1 ¹⁾	1000 ¹⁾
6	2 ¹⁾	2000 ¹⁾
11	1 ²⁾	1000 ²⁾
11	2 ²⁾	2000 ²⁾
11	3 ²⁾	3000 ²⁾
5	1 ²⁾	500 ²⁾
8	1 ²⁾	200 ²⁾

1) Сокращение достигается за счет привлечения дополнительных субподрядчиков.

2) Сокращение достигается за счет передачи некоторой части работ сторонней организации (для остальных частей работы число исполнителей остается без изменения).

тельности проекта необходимо затратить 3000 долл. на работу 11. При этом, однако, нарушается ограничение на число одновременно работающих людей (т. е. требуется использовать более четырех исполнителей). Используя данные, приведенные на рис. 7 и в табл. 8, можно построить график выполнения проекта за 17 нед, который удовлетворял бы этому ограничению. Один из вариантов такого графика учитывает затраты указанных выше 2200 долл. на работы 11 и 8, 1000 долл. на работу 6 и 500 долл. на работу 5. Заметим, что сокращение длительности работ 11 и 8 (рис. 7) требует одновременного сокращения длительности работ 6 и 5, если должны по-прежнему соблюдаться ограничения на трудовые ресурсы и длительность проекта. Этот вариант более эффективен по затратам (3700 долл.), чем вариант, предусматривающий дополнительные затраты в 2000 долл. на работу 6 и 3000 долл. на работу 11.

Таблица 9

Перераспределение трудовых ресурсов

Число исполнителей	Снятие с работы	Направление на работу	Сокращение длительности, нед
1	1	3	2
1	5	8	1
1	10	11	2
			5

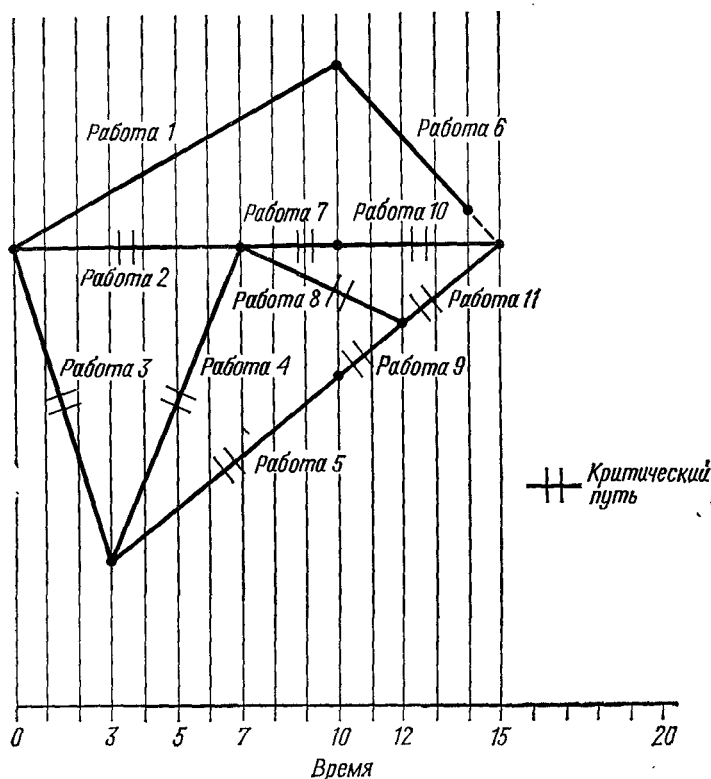


Рис. 8. Сетевой график после перераспределения трудовых ресурсов.

График работ для одного из возможных вариантов перераспределения исполнителей между работами в пределах имеющихся ресурсов приведен на рис. 8 (исходные данные приведены в табл. 7, а произведенные перемещения исполнителей — в табл. 9). При таком графике ограничение на трудовые ресурсы (не более четырех исполнителей) также нарушается.

Таким образом, при наличии ограничений на трудовые ресурсы решение задачи разработки «оптимального» (по соотношению затрат и времени) графика выполнения проекта вызывает определенные трудности.

3.4. Алгоритмы распределения ресурсов

В данном случае задачами распределения ресурсов, по существу, являются задачи определения сроков исполнения проекта, при которых возможно либо выравнивание потребности в ресур-

сах при соблюдении ограничений на длительность выполнения проекта, либо минимизация длительности выполнения проекта при ограничениях на трудовые ресурсы, либо минимизация общей стоимости ресурсов и штрафов за задержку выполнения проекта. Решение последней из перечисленных задач с помощью моделирования приводится в работе [64]. Наибольшее внимание, однако, уделялось первым двум задачам. Ввиду их комбинаторного характера возможность получения оптимальных решений средствами математического программирования очень ограничена. Автор данной главы на собственном опыте убедился в неэффективности использования алгоритмов целочисленного и булевого программирования, так как они не сходятся при разумной длительности счета. Были предложены алгоритмы, основанные на методе ветвей и границ, применимые, однако, только для сетей с числом узлов меньше 50 [15, 36].

Ввиду отсутствия успеха в области создания методов оптимизации основное внимание уделялось развитию эвристических методов, позволяющих получить приемлемое решение при некоторых разумных правилах установления приоритета работ, выполняемых с использованием ресурсов, на которые существуют ограничения. Эти методы делятся на «последовательные» и «параллельные» в зависимости от того, устанавливается ли приоритет до начала составления графика выполнения проекта или постепенно в процессе составления графика. Хотя большинство опубликованных исследований посвящено параллельным методам, относительная ценность каждого из них все еще не установлена. Дэйвис [14—16] дал подробный обзор работ в этой области и оценку большого числа эвристических методов установления приоритета работ в процессе календарного планирования. Его основной вывод заключается в том, что, хотя ни один из эвристических подходов не может всегда давать наилучший график, правило упорядочения, в соответствии с которым первой выполняется работа с наименьшим резервом (или эквивалентное правило минимизации самого позднего времени начала), в среднем дает наилучшие результаты.

Следует отметить, что многие организации, в том числе IBM, CEIR, McDonnell Automation, Control Data Corp., имеют программы для ЭВМ, способные планировать графики для больших (до 1000 работ) проектов со многими ресурсами. Эти программы, как правило, включают средства для расчленения работ, составления расписания в пределах допустимого колебания уровня ресурсов, учета стоимости работ, различные средства составления отчетов, несколько эвристических правил упорядочения и выравнивания ресурсов. Подробное описание таких программ приведено в работе [64].

Подводя итог, отметим, что трудность решения проблемы формализации задачи оптимального распределения ограниченных ре-

сурсов обусловлена отсутствием точно сформулированного общего критерия «оптимального» использования ресурсов. Кроме того, трудно предопределить, каким образом ограничения на упорядочение работ могут влиять на потребность и наличие ресурсов. Эвристические правила построения графика, без сомнения, еще в течение некоторого времени останутся основой всех практических систем. Однако в связи с возрастающими возможностями вычислительных систем, работающих в режиме разделения времени, все больше внимания уделяется разработке итеративных процедур, предложенных в работе [37].

3.5. Стохастические подходы

Часто возникает необходимость оценить вероятность того, будет ли отдельная работа, некоторая совокупность работ или проект в целом закончен к определенному сроку. Подобная формулировка задачи, которая была основой при первоначальной разработке метода ПЕРТ, предполагает использование трех видов оценок длительности каждой работы в сети: t_0 — оптимистической оценки длительности, t_m — модальной оценки, т. е. оценки наиболее вероятной длительности, и t_p — пессимистической оценки длительности. Предполагается, что гипотетическое распределение фактической длительности выполнения работы описывается β -распределением [12]. Математическое ожидание t_e и дисперсия v_i для β -распределения, определяющиеся выражениями $t_e = (t_0 + 4t_m + t_p)/6$, $v_i = [(t_p - t_0)/6]^2$, могут быть вычислены для каждой работы сети. В этом случае ожидаемое время завершения T_e всего проекта находится как сумма величин t_e (которые аналогичны длительностям работ при использовании метода критического пути) для работ, составляющих путь наибольшей длины через сеть. Далее для любого подмножества критических работ S могут быть вычислены доверительные границы длительности по следующей формуле (которая основана на центральной предельной теореме для суммы независимых случайных величин):

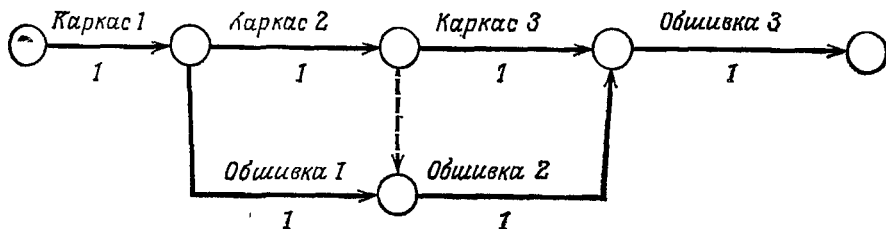
$$T_{c, s} = \sum_{j \in S} (t_e)_j \pm K \sqrt{\sum_{j \in S} (v_i)_j},$$

где K — константа, зависящая от степени достоверности ($K = 3$ для степени достоверности 99,7%, $K = 2$ для степени достоверности 95% и т. д.). Обычно «оптимальным» считается такой план выполнения проекта, для которого $T_{c, s} \leq DC_s$, где DC — желательный срок завершения множества работ S . Для сетевых графиков, которые не удовлетворяют этому критерию, можно прибегнуть к перераспределению ресурсов с ненапряженных работ на критический путь, к увеличению затрат или более радикальным способам перепланировки проекта, пока не будет удовлетворено требование по этому критерию.

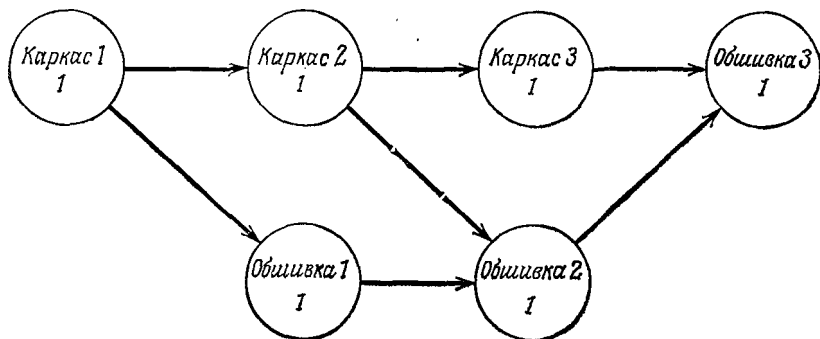
Следует заметить, что все сказанное выше основывается на предположении, что цель заключается в оценке времени завершения проекта. В большинстве случаев это соответствует действительности. Однако, если руководитель проекта считает, что заданный срок окончания легко достижим, например $T_{C, S}$ значительно меньше DC_S , тогда кроме обычного резервного времени (определяемого критическим путем) возникает дополнительное «свободное» резервное время, обусловленное тем, что $T_{C, S}$ меньше DC_S [12, 61].

3.6. Другие методы построения сетей

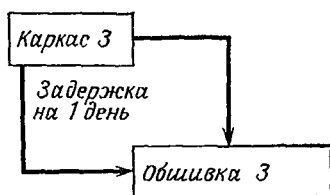
В настоящее время возрастает значение некоторых видов сетей, отличных от описанных в разд. 3.1 [33]. Иногда проще нарисовать сеть, если поменять ролями дуги и узлы. В этом случае работы указываются в узлах сети. Один из вариантов обобщения метода сетей, в котором развивается логика типа ПЕРТ и критического пути, заключается в построении схемы, отражающей предшествование операций [29]. Подобные методы построения сетей используются в тех случаях, когда две работы должны начаться одновременно, либо вместе закончиться, либо производиться параллельно с заданным сдвигом во времени (ситуация, которая часто возникает в строительных работах). Например, типичная сеть с работами, указанными на дугах, имеет вид



Если работы изображены в виде узлов, то та же самая сеть принимает вид



При использовании схемы предшествования эту сеть следовало бы изобразить следующим образом:



Многие проекты научных исследований и разработок часто подвержены действию случайных факторов, т. е. состав работ, выполняемых после завершения некоторого этапа, может изменяться в зависимости от некоторых условий. Например, химик, занимающийся синтезом и испытанием нового инсектицида, должен сознавать, что возможно несколько результатов: неудача, испытания прошли нормально, получен уникальный результат. Каждый из этих результатов может привести к необходимости выполнения ряда последующих работ. Поэтому, если используется метод ПЕРТ или критического пути, результирующая сеть, изображающая все связи между работами над первичным проектом, будет большой и громоздкой. При этом существует вероятность того, что с помощью одной схемы не удастся полностью отобразить проект. Задачи этого типа можно изобразить с помощью сетей типа дерева с точками принятия решений [22] (рис. 9). Заметим, что резервы времени 4-6 и 5-6 являются условными: если в узле 3 получен результат такой, что выбирается работа 3 или работа 5, то весь проект сокращается на соответствующую одному из имеющихся резервов величину. Все прочие расчеты сети производятся так же, как описано ранее, за исключением того, что результаты этих вычислений должны учитывать вероятность соответствующего результата. Например, на рис. 9 ожидаемое время завершения в узле

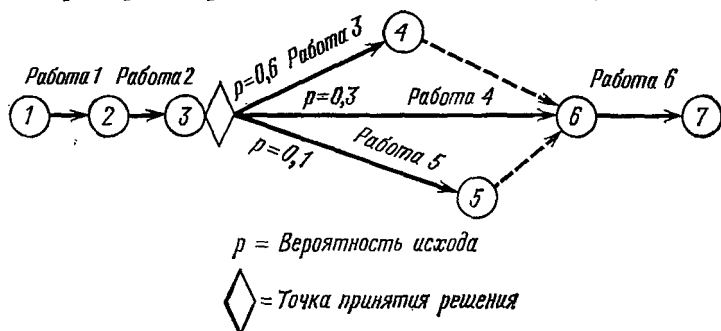


Рис. 9. Сеть с возможными путями.

7 следует вычислять по формуле: $t_e^6 + (0,6t_e^3 + 0,3t_e^4 + 0,1t_e^5) + t_e^2 + t_e^1$, где t_e^j — математическое ожидание длительности j -й работы. Для проблем этого типа специально предназначен язык моделирования GERT [38].

3.7. Комплексные проекты

Рассмотренные выше сетевые методы могут быть применены к планированию нескольких проектов, руководство которыми осуществляется одной организацией. Рассмотрим, например, портфель из четырех проектов научных исследований (рис. 10). Как видно из рисунка, проект A должен быть завершён до того, как сможет начаться проект B (например, проект A должен произвести сырьё для проекта B). Проект X и проект Y могут начаться и закончиться в любое время в интервале между моментами a и e . Таким образом, пунктирные линии на рис. 10 отображают лишь фиктивные зависимости и служат для того, чтобы указать общий диапазон времени для всех четырех проектов. На этих фиктивных связях, если они имеются, можно проставить длительности выполнения проектов и таким путем получить ранний срок начала и поздний срок окончания проекта. Резервное время для отдель-

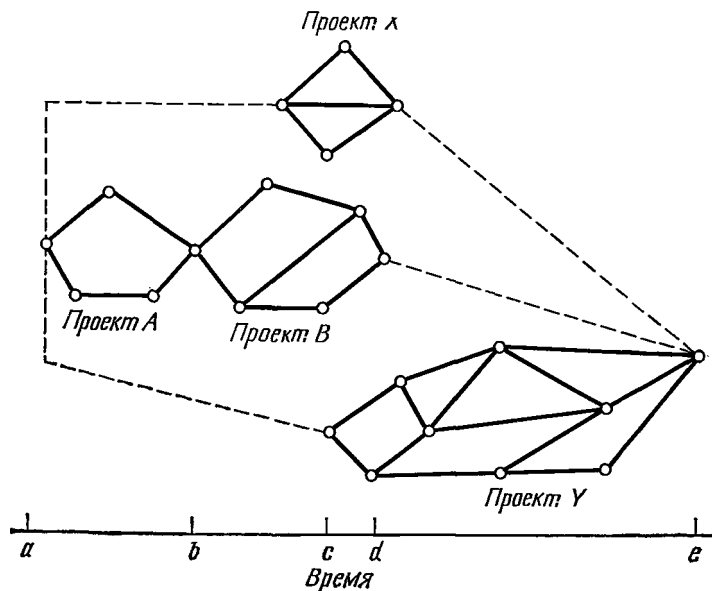


Рис. 10. Сетевой график для комплексной программы.

ных проектов, образующееся за счет этих фиктивных работ, можно использовать так же, как фиктивные работы в сетях для автономного проекта.

Предположим, например, что проекты *B* и *X* используют одни и те же ресурсы и что потребность в этих ресурсах превышает их наличие в случае необходимости одновременной реализации проектов *B* и *X*. Из рис. 10 видно, что начало проекта *X* можно задержать до момента *d*, сосредоточив все ресурсы полностью на проекте *B*. После того как проект *B* будет завершен, эти ресурсы можно будет использовать для выполнения проекта *X*.

Другая возможность состоит в том, чтобы оба проекта использовали эти ресурсы с пониженной интенсивностью и при этом оба проекта будут задерживаться (на столько, чтобы не выйти за пределы момента *e*). Процедуры ускорения выполнения работ за счет повышения затрат, описанные в разд 3.3., могут быть применимы и к отдельным проектам. Таким образом, методы сетевого планирования для комплексных проектов полностью аналогичны сетевым методам, применяемым для отдельных проектов [63].

Существует, однако, еще один важный аспект планирования комплексных работ — приоритет отдельных проектов. Предположим, что проект *X* (рис. 10) считается наиболее «важным» из всех проектов. Могут, например, существовать веские основания для того, чтобы руководитель считал необходимым сначала начать данный проект, а не любой другой. Одной из таких причин могла бы быть опасность конкуренции. В разработанном советом по экономическим вопросам и промышленным исследованиям (США) вычислительном алгоритме RAMPS (распределение ресурсов и календарное планирование комплексных проектов) приоритетность отдельных проектов используется в качестве весового множителя при календарном планировании и распределении ресурсов между конкурирующими проектами в системах комплексных проектов [34].

Вообще говоря, последовательные изменения в сетевых графиках проектов позволят руководителям на уровне проекта и на уровне отдела создавать интегрированные планы. Руководитель каждого проекта может представить на рассмотрение оптимальный сетевой график. Эти графики могут быть затем объединены в одну сеть комплексного проекта [65], на основе чего разработано несколько графиков для комплексного проекта при различных предположениях относительно приоритетов, ресурсов и т. д. Эти варианты графиков выполнения комплексного проекта могут быть затем обсуждены на совещаниях сотрудников, проводимых каждым руководителем отдельного проекта и руководителем комплексного проекта, например заместителем директора фирмы по научно-исследовательской работе. После этого, исходя из результатов обсуждения всех составных частей и высказанных при этом заме-

чений, может быть выбран наилучший сетевой график комплексного проекта. Прежде чем будет разработан приемлемый общий план работ, может потребоваться несколько последовательных итераций разработки графика работ между уровнем управления отдельным проектом и уровнем комплексного проекта.

4. РУКОВОДСТВО ПРОЕКТОМ¹⁾

Можно утверждать, что основной функцией любого руководителя проекта является оперативное руководство последним. Оперативное руководство — это процесс принятия таких решений, в результате которых затраты на осуществление проекта не превышают плановых лимитов, т. е. оперативное руководство состоит из оценки степени технической готовности работ над проектом и корректировки отклонений от планового графика.

На практике функцию оперативного руководства трудно отделить от функции планирования, поскольку оперативное руководство уже предполагает существование некоторого плана, который служит эталоном для измерения и оценки степени продвижения работ. Кроме того, в процессе управления используется сравнение первоначального плана с рядом работ, что позволяет вносить соответствующие коррективы в планы. Таким образом, планирование, оперативное управление и корректировка плана являются взаимосвязанными операциями, выполнение которых постоянно является в обязанности руководителя проекта. Однако сам процесс оперативного руководства не сводится к простому регулированию и часто требует некоторого планирования, как, например, при перераспределении исполнителей и бюджета в пределах установленных сводным планом показателей в целях корректировки. Мы будем использовать термин «планирование» применительно к процессу выбора операций, построения графиков работ и распределения ресурсов, а также прогнозирования будущих операций, а термин «оперативное руководство» — к оценке соответствия хода работ плану и к проведению небольших корректировок плана с целью поддержания в установленных пределах графика выполнения запланированных операций и затрат на их выполнение [47, 49, 52].

Для информации руководителя проекта было предложено и используется несколько способов, начиная от простых форм отчетности до количественных моделей типа затраты — время. Примером информации чисто описательного характера являются периодические отчеты о ходе работ. Они обычно состоят из кратко

¹⁾ Часть приведенного в этом разделе материала опубликована в *IEEE Trans. on Eng. Management*, March 1968, и в *R & Management*, October 1972.

сформулированных пунктов, в которых отмечается, что было сделано и какие работы завершены [49].

Модель управления проектом будем считать эффективной, если она дает обоснованные прогнозы о дальнейшем ходе выполнения проекта путем информирования руководителя о любых отклонениях от плана по состоянию работ и затратам в функции времени, указания необходимых корректирующих действий по управлению и доведения этих сведений до руководителя проекта. Под прогнозом создания проекта понимается информация, необходимая руководителю для оценки сведений о возникновении критического состояния или об ожидаемой неудаче в его выполнении, которая может произойти в будущем. Чтобы удовлетворить этим критериям, модель управления проектом должна прежде всего учитывать как сведения о ходе работ, так и сведения о затратах во времени. Естественно, что ни описательные отчеты, ни модели типа затраты — время не могут обеспечить подобной информации.

1.4. Управление соответствием затрат и достигнутых результатов

Предположим, что для i -го периода планирования построен в масштабе времени оптимальный в некотором смысле сетевой план. Пусть

t — период времени, $t = 1, 2, \dots, n$;

E_t — фактические общие затраты в долларах к концу интервала времени t ;

$\hat{E}_t$ — прогнозируемая величина общих затрат в долларах к концу интервала времени t ;

ΔE_t — величина отклонения по затратам в течение интервала времени t ,

$$\Delta E_t = \hat{E}_t - E_t; \quad (25)$$

ϕ_t — суммарный фактический объем выполненных работ к концу интервала времени t ;

$\hat{\phi}_t$ — прогнозируемый суммарный объем выполненных работ к концу интервала времени t ;

$\Delta \phi_t$ — отклонение объема выполненных работ для интервала времени t ,

$$\Delta \phi_t = \phi_t - \hat{\phi}_t; \quad (26)$$

$\hat{C}_t$ — величина планируемых общих затрат для фактически выполненного к концу интервала времени t объема работ;

ΔC_t — отклонение затрат для выполненного объема работ для интервала времени t ,

$$\Delta C_t = \hat{C}_t - E_t. \quad (27)$$

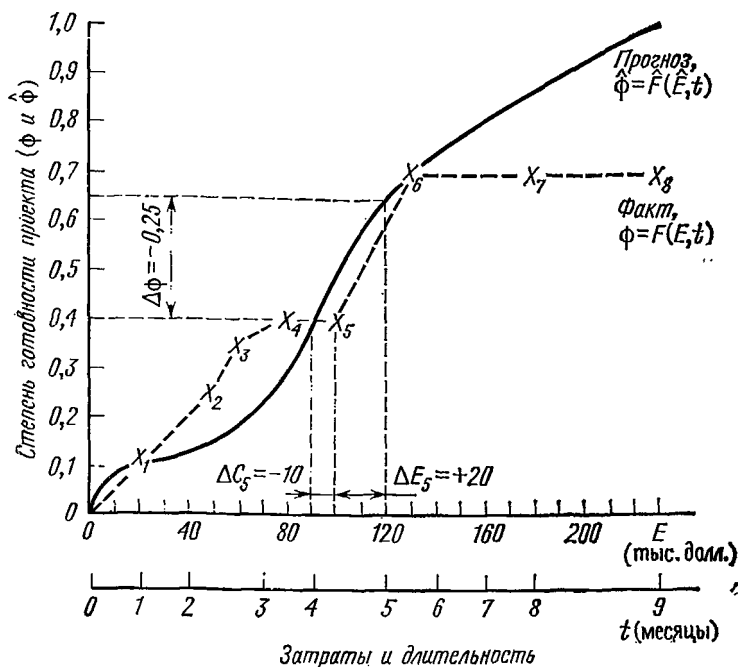


Рис. 11. Модель контроля за реализацией проекта.

Объем выполненных работ измеряется в процентах завершенных узлов сетевого графика. Тогда величина $\hat{\phi} = \hat{F}(\hat{E}, t)$ представляет собой прогноз для t -го интервала периода планирования, т. е. $\hat{\phi}$ — есть прогнозируемое накопленное значение или прогнозируемый суммарный объем выполненных работ, причем $\hat{\phi}$ является функцией прогнозируемых затрат по времени, т. е. функцией бюджета. Если t последовательно принимает значения от 1 до n , то можно построить траекторию $\phi = F(E, t)$ таким образом, что $\{E - \hat{F}\}$ будет представлять собой информационно-контролирующую систему типа затраты — результаты. Иными словами, ряд сопоставимых разностей между значениями функций, отражающих зависимость реализованных и прогнозируемых затрат от объема выполненных работ, оцениваемых по нескольким параметрам, представляют собой информацию для принятия решения руководителем работы.

Рассматриваемая модель иллюстрируется на рис. 11. Пусть, например, было определено, что к концу пятого месяца фактический суммарный объем выполненных работ по проекту составил 0,40 (было достигнуто 40% узлов сети) и фактические суммарные

затраты составили 100 000 долл., что на графике (рис. 11) соответствует точке X_5 с координатами $E = 100$ и $\phi = 0,4$ (при этом шкала времени не рассматривается). Значения параметров $\hat{E}_5$, E_5 , $\hat{\phi}_5$, ϕ_5 и $\hat{C}_5$ были занесены в табл. 10, и величины соответствующих отклонений подсчитаны по формулам (25—27). Подобным же образом (рис. 11) были нанесены точки X_1 , X_2 , X_3 , ..., X_n для 1, 2, 3, ..., n месяца и аналогично определены отклонения (табл. 10).

Таблица 10

Анализ бюджета

t	Затраты			Достигнутые результаты			Соотношение затрат и результатов	
	E_t	$\hat{E}_t$	ΔE_t	ϕ_t	$\hat{\phi}_t$	$\Delta \phi_t$	$\hat{C}_t$	ΔC_t
1	20	20	0	0,10	0,10	0	20	0
2	50	40	-10	0,25	0,15	+0,10	70	+20
3	60	70	+10	0,35	0,25	+0,10	85	+25
4	75	90	+15	0,40	0,40	0	90	+15
5	100	120	+20	0,40	0,65	-0,25	90	-10
6	130	140	+10	0,70	0,75	-0,05	130	0
7	180	160	-20	0,70	0,85	-0,15	130	-50
8	230	180	-50	0,70	0,87	-0,17	130	-100

Рис. 11 — это другой способ отображения во времени сетевого графика (оптимизированного по времени и затратам) и регистрации фактических затрат и достигнутых результатов. Шкалы времени и затрат на рис. 11 взяты из сетевого графика, например, шкала затрат получена путем суммирования затрат для каждого момента времени (или для каждого узла сети). Какую из шкал — затрат или времени — изображать линейной, безразлично. Однако, поскольку речь идет главным образом о соотношении затрат и достигнутых результатов, здесь линейной является шкала затрат.

Эта модель позволяет руководству проектом фиксировать возможные отклонения в тех случаях, когда они имеют место, устанавливать их причины, определять, какие управляющие действия необходимо предпринять (например, как и когда пересматривать план). Так, непропорционально высокие затраты по сравнению с достигнутыми результатами всегда являются для руководителя проекта достаточным основанием для пересмотра детального проекта. Все такие непропорционально высокие затраты отображаются в данной модели в виде отрицательных отклонений затрат по сравнению с достигнутыми результатами (величины ΔC_t в табл. 10 отрицательны). Согласно модели, в тех случаях, когда полученные

результаты ниже запланированных, даже если они пропорциональны затратам, возникает необходимость в пересмотре проекта. Это означает, что, если значение $\Delta\phi_i$ отрицательно (табл. 10), даже когда ему соответствует нулевое значение ΔC_i , проект необходимо пересмотреть. Следует попытаться найти объяснение, почему интенсивность затрат оказалась ниже планируемого уровня. Если считать, что первоначальный план был оптимальным, то подобное отклонение затрат в положительную сторону свидетельствует об отклонении от оптимальности, что может в конце концов вызвать весьма нежелательные отрицательные отклонения затрат по сравнению с достигнутыми результатами. В качестве иллюстрации такого положения могут служить данные для 6—8 месяцев в табл. 10 и на рис. 11. Величина отклонения затрат по сравнению с достигнутыми результатами для шестого месяца (ΔC_6) равна нулю, поскольку уменьшение объема выполненных работ по сравнению с плановым ($\Delta\phi_6 = -0,05$) компенсируется экономией затрат ($\Delta E_6 = +10$). В этом гипотетическом примере можно предположить, что недоиспользование бюджета, начиная с третьего месяца, объяснялось экономией ресурсов для данного проекта с целью их использования для конкурирующего проекта. Строго говоря, это изменение требует повторного проведения процесса оптимизации и построения новой кривой прогноза на рис. 11, которая отражала бы изменившийся размер выделяемых данному проекту ресурсов. Допустим, однако, что подобная корректировка плана не была проведена, так как руководство предполагало, что упущенное время можно компенсировать за счет увеличения ассигнований. Предположим далее, что такое усиленное финансирование осуществлялось с целью подогнать состояние работ к запланированному уровню и проводилось в течение седьмого и восьмого месяцев. Оказалось, что пропорциональное увеличение результатов достигнуто не было и отклонение затрат от заданного уровня стало становиться все более отрицательным. Фактически многие проекты могут аналогичным образом иметь эффективные по соотношению затрат и длительности работ графики, которые, возможно, исключительно трудно перестраивать, если они однажды будут нарушены. Эффективное планирование на длительное время зависит от хорошего контроля, который должен указывать, когда и сколько раз следует производить корректировку плана.

Следовательно, любые положительные и отрицательные отклонения величины затрат в ходе работ и все отрицательные отклонения соотношений затрат и достигнутых результатов должны вызывать пересмотр детального проекта и коррекцию плана его осуществления.

Остается выяснить, как поступить при перерасходе бюджета, т. е. при отрицательных отклонениях величины затрат (отрицательные значения ΔE_i табл. 10)? Согласно модели, подобные от-

клонения не обязательно требуют осуществления управляющих воздействий. Так, например, руководитель проекта не должен беспокоиться в связи с тем, что в конце второго месяца (табл. 10, рис. 11) имеет место перерасход, т. е. $\Delta E_2 = -10$, так как отклонение соотношения затрат и достигнутых результатов положительно $\Delta C_2 = +20$. Это означает, что повышенные расходы дали более чем пропорциональное увеличение результатов в ходе выполнения проекта (рис. 11). Другими словами, в такой ситуации работа с перерасходом бюджета не вызывает опасений, так как перерасход окупается положительными результатами.

Рассмотрим теперь данные для третьего, четвертого и пятого месяцев в табл. 10. Прирост абсолютного уровня выполненных работ падает с возрастающей скоростью, на что указывают числа в графе Δf_t , изменяющиеся от $+10$ до 0 и далее до $-0,25$. Уровень объема выполненных работ на единицу затрат также падает, на что указывают числа в графе ΔC_t . С каждым прошедшим месяцем уменьшается не только прогресс в ходе работ над проектом, но также и результаты в расчете на доллар затрат. Но ни один из этих тревожных сигналов не проявляется в величине отклонений от бюджета или затрат (графа ΔE_t табл. 10). Фактически демонстрируется благоприятное состояние проекта с точки зрения бюджетных ограничений (положительные числа в графе ΔE_t) в течение всех этих трех месяцев.

Следовательно, контроль за ходом реализации проекта только по данным о перерасходе или недоиспользовании бюджета может привести к серьезным ошибкам. Отклонения по затратам (бюджету) и по состоянию работ могут либо балансироваться, что видно по данным на конец шестого месяца в табл. 10 и на рис. 11, либо они могут взаимно усиливать друг друга, что видно по результатам на конец седьмого и восьмого месяцев. В любом случае сочетание отклонений по затратам и по состоянию работ определяет фактическое состояние проекта. Это простой факт часто упускают из вида при руководстве проектом. В результате неоправданно большое значение обычно придается фактам перерасхода средств на проект, тогда как проекты, затраты на осуществление которых меньше предусмотренных бюджетом, часто не привлекают крайне необходимого для них внимания. Даже при фиксированном распределении трудовых ресурсов могут приниматься идущие в разрез с первоначальным планом решения, так как руководство может, например, предпочесть сэкономить на «доле» внешних услуг, в частности на услугах специалистов по исследованию операций, чтобы затраты были ниже планового уровня, либо оно может недорасходовать средства в надежде накопить некоторые ресурсы для других, более важных проектов. Описанная выше модель позволяет обнаруживать такие отклонения.

Предложенные идеи и применение модели контроля хода реа-

лизации проекта иллюстрируются в работах [47—49]. Стандартные пакеты программ сетевого планирования ПЕРТ/МКП обычно предусматривают формирование данных о затратах и состоянии проекта в виде обновляемых оценок времен завершения проектов и затрат [61]. Однако формирование интегрированных сводок, сопоставляющих затраты и достигнутые результаты, обычно не входит в состав пакетов программ ПЕРТ/МКП.

5. ИНТЕГРИРОВАННЫЕ СИСТЕМЫ ПЛАНИРОВАНИЯ И РУКОВОДСТВА

Исходные предпосылки и информация, первоначально используемые при планировании, часто весьма скудны и подвержены быстрому изменению. Техническое состояние проекта, находящегося в стадии реализации, довольно быстро меняется (например, в результате возникновения поломок), конкуренты могут начать разработку подобного проекта, могут измениться цены на основное сырье и т. п. Изменения в исходных данных могут привести к необходимости вновь выполнить функции планирования и руководства, чтобы приспособиться к этим изменениям. Динамиче-

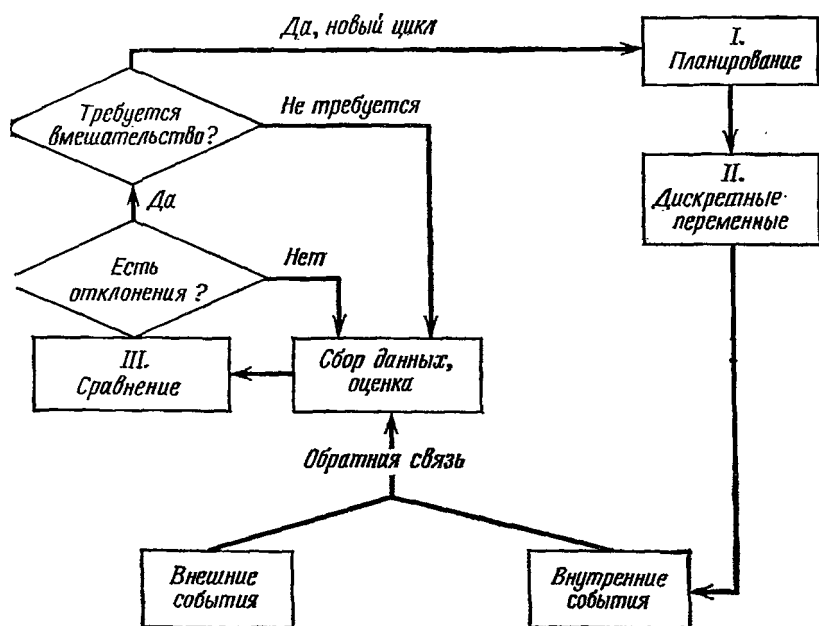


Рис. 12. Интегрированная система планирования и управления реализацией проекта.

ское влияние этих изменений на другие имеющиеся программы (предполагаемая реакция конкурентов, поставщиков и т. д.) могут вызвать дальнейшие изменения, которые требуют дополнительных адаптивных действий по планированию и руководству. Все это приводит к тому, что с течением времени циклы планирования и управления периодически повторяются (рис. 12). Исходные представления (информация), полученные руководителем, образуют основу для построения первоначального плана (блок I), характерные черты и детали которого зависят в первую очередь от существующей технологии и от того, насколько четко руководство представляет себе возможности ее использования. Этот план реализуется путем изменения (уточнения) некоторых (на усмотрение руководства) переменных (дискретных переменных), например затрат (блок II). Результаты подобного уточнения будут влиять на некоторые выходные (эндогенные) переменные, такие, как техническое состояние проекта. Кроме того, возможны изменения ряда входных (экзогенных) переменных, например конъюнктуры. Полученные результаты сравниваются с исходными данными и первоначально ожидаемыми результатами (блок III). Затем принимается решение относительно вмешательства в руководство проектом. Если решение будет положительным, то это может привести к пересмотру плана. Такой цикл динамической адаптации, включающий планирование — оперативное руководство — корректировку плана, может повторяться непрерывно [55].

5.1. Адаптивное планирование работ над проектом и руководство проектом

Схема адаптивного планирования и руководства представлена на рис. 13. «Модель» планирования (т. е. выбора проекта, распределения ресурсов и сетевого планирования) может использоваться многократно, пока не будет составлен удовлетворительный план. После того как разработанный план начнет осуществляться, могут произойти события, которые потребуют внесения изменений в первоначальные прогнозы и оценки, т. е. появятся все основания для пересмотра плана и последовательного изменения модели планирования. Обстоятельства могут измениться до такой степени, что придется составлять новый набор проектов, вновь осуществлять распределение ресурсов между ними и строить новые сетевые графики. В результате это может привести к прекращению работ над некоторыми проектами, перераспределению средств между имеющимися проектами, выбору дополнительных проектов из портфеля заказов, перераспределению затрат во времени или к некоторой комбинации этих действий. Таким образом, цикл планирование — управление повторяется, причем в качестве вспомо-



Рис. 13. Планирование и управление реализацией проекта с динамической адаптацией.

могательного средства руководство повторно использует описанным выше образом аналитические модели планирования и управления.

Итак, модель планирования определяет планируемый порядок выполнения работ $\hat{\phi}_j \leftarrow \hat{F}_j(\hat{E}_j, t)$ для каждого из $j = 1, \dots, s$ выбранных и финансируемых проектов из множества $j = 1, \dots, n$ проектов, имеющих в портфеле (условные обозначения см. в разд. 1). Аналогично модель оперативного руководства определяет текущее выполнение работ $\phi_j = F_j(E_j, t)$ для каждого проекта $j = 1, \dots, s$. Если $\pm \Delta f_j$ есть пределы «управления», то j -й проект является «управляемым» только в том случае, если

$$\hat{F}_j - \Delta f_j \leq F_j \leq \hat{F}_j + \Delta f_j.$$

Если j -й проект становится «неуправляемым», то требуется пересмотр плана. Руководящие действия, такие, как перераспределение ресурсов проекта с ненапряженных работ на работы критического пути, могут осуществляться в любое время, чтобы «удерживать» проект в «управляемых» пределах. Как уже отмечалось в разд. 4.1, руководящие действия осуществляются на основе анализа величины отклонений ΔC_t и ΔE_t . Метод определения соответ-

ствующего значения $\pm \Delta f_j$; зависит от характера проекта и намерений руководства. Во многих успешно завершенных программах отклонения составляют 10—20% величины F_j ; [47, 49].

В рассматриваемой системе комбинируется централизованное и децентрализованное принятие решений. Оперативное руководство проектом является децентрализованным, поскольку оно осуществляется на самом низком из возможных уровней руководства проектом. Руководитель проекта «удерживает» проект в допустимых пределах с помощью небольших изменений в рамках установленного бюджета. Эти изменения рассматриваются соответствующим уровнем руководства проектом, и на данном этапе контроль на более высоком уровне за ходом проекта оказывается нецелесообразным. Однако изменения в общих бюджетах проектов, изменения численности или состава персонала или набора проектов, а также любые серьезные пересмотры планов, которые влияют на другие проекты, требуют подключения руководства более высокого уровня, чем оперативное руководство. В этом случае требуется централизованное принятие решений. Таким образом, модель планирования — это средство, которое обеспечивает взаимодействие планирования и управления на уровне отдельного проекта с планированием и контролем на уровне комплексной программы.

ЗАКЛЮЧЕНИЕ

Несмотря на то что с помощью методов исследования операций разработано большое количество моделей выбора проектов и распределения капиталовложений, создание подходящих и эффективных моделей распределения капиталовложений, которые были бы приемлемы для руководителей проектов, все еще остается проблемой. Эта проблема усложняется тем, что, как это свойственно многим организациям, процессы принятия решений по выбору проектов и распределению ресурсов имеют рассредоточенный и распределенный характер, что способствует квазиоптимизации, осуществляемой по частям. Именно этим объясняется тот факт, что для выбора проектов, как правило, используются модели экономических показателей и оценок экспертов.

Хотя сетевые методы планирования широко приняты, их естественное расширение в виде контроля и отчетности по соотношению затрат и достигнутых результатов не получило заслуженного внимания.

Одной из основных проблем все еще остается разработка практически применимых моделей выбора проектов и разработка комбинированных эвристически оптимальных методов распределения ресурсов при календарном планировании проектов. Что касается проблемы контроля, то здесь все усилия должны быть направлены на разработку более адекватных мер оценки достигнутых результатов и состояния проектов, чем простой подсчет числа завершенных работ сетевого графика, так как во многих проектах степень достижения цели проекта, по-видимому, не пропорциональна проценту завершенных работ. Во времени написания данной главы не были решены полностью также проблемы контроля в сетях, в которых длительности работ являются случайными переменными.

ЛИТЕРАТУРА

1. Ansoff H. I., Evaluation of Applied Research in Business Firm, in: Technical Planning on the Corporate Level, J. R., Bright (Ed.), Harvard University Press, Cambridge, Massachusetts, 1962.
2. Baker N. R., Scoring Models., in: Seminar at Airle House, Document 70/58, Program of Research on the Management of Research and Development, Northwestern University, Evanston, Illinois, 1970.
3. Baker N. R., Moore J. R., Computational Analysis of Scoring Models for R & D Project Selection, *Management Sci.*, 16, No. 4, 124-134 (1969).
4. Baker N. R., Pound W. H., R & D Project Selection: Where We Stand, *IEEE Trans. on Eng. Management*, EM-11, 124-134 (1964).
5. Baker N. R., Souder Wm. E., Shumway C. R., Maher P. M., Rubenstein A. H., A Budget Allocation Model for Large Hierarchical R & D Organizations, *Management Sci.*, 23, No. 1, 59-70 (1976).
6. Baker N. R., R & D Project Selection Models: An Assessment; *IEEE Trans. on Eng. Management*, EM-21, No. 4, 165-172 (1974).
7. Bell D. C., Chilcott J. E., Read A. W., Solway R. A., Application of a Research Project Selection Method in the Northern Region Scientific Services Department, R & D Department, Central Electricity Generating Board (U.K.), RD/H/R2/, 1967.
8. Bigelow C. G., Bibliography on Project Planning and Control by Network Analysis, *Operations Res.*, 10, 728-731 (1962).
9. Brockhoff K., Some Problems and Solutions in the Selection of an R & D Portfolio, paper presented at the 1969 Conference of the IFORS Vancie, Italy, 1969.
10. Cetron M. J., Quest Status Report, *IEEE Trans. on Eng. Management*, EM-14, No 1, 51-62 (1967).
11. Cetron M. J., Roepcke L. H., The Selection of R & D Program Content, *IEEE Trans. on Eng. Management*, EM-14, 4-13 (1967).
12. Clark C. E., The PERT Model for the Distribution of an Activity Time, *Operations Res.*, 10, 405-406 (1962).
13. Cleland D. I., King W. R., Systems, Organization Analysis, Management: A Book of Readings, McGraw-Hill, N.Y., 1970.
14. Davis E. W., Ressours Allocation in Project Network Models—A Survey, *J. Ind. Eng.*, 17, No. 4, 33-41 (1966).
15. Davis E. W., Project Scheduling under Resource Constraints—Historical Review and Categorization of Procedures, *AIIE Trans.*, 5, No. 4, 297-313 (1973).
16. Davis E. W., Patterson J. H., A Comparison of Heuristic and Optimum Solutions in Resource Constrained Project Scheduling, *Management Sci.*, 21, No. 8, 944-955 (1975).
17. Dean B. V., Sengupta S. S., On A Method for Determining Corporate Research and Development Budgets, in Management Science Models and Techniques, Churchman C. W. and Verhulst M. (Eds.), Pergamon Press, N.Y., 1960, pp. 210-225.
18. Dean B. V., Nishry M. J., Scoring and Profitability Models for Evaluating and Selecting Engineering Projects, *Operations Res.*, 13, No. 4, 13-21 (1965).
19. Dewitte L., Manpower Leveling of PERT Networks, Data Proc. for Sci./Eng., March-April 1964, pp. 29-37.
20. Disman S., Selecting R & D Project for Profit, *Chem. Eng.*, 11, 87-90 (December 1962).
21. Douds C. F., The State of the Art in the Study of Technology Transfer, *R & D Management*, 1, No. 3, 125-131 (1971).
22. Eisner H., A Generalized Network Approach to the Planning and Control of a Research Project, *Operations Res.*, 10, 115-125 (1962).

23. Fazar W., The Origin of PERT, *The Controller* (December 1962), pp. 598-602, 618-621.
24. Fulkerson D. R., A Network Flow Computation for Project Cost Curves, *Management Sci.*, 8, No. 2, 51-60 (1961).
25. Gee R. E., A Survey of Current Project Practices, *Res. Management*, 14, No. 5, 38-45 (1971).
26. Harris J. S., New Product Profile Chart, *Chem. Eng. News*, 39, No. 16, 110-118 (1961).
27. Hart A., Evaluation of Research and Development Projects, *Chem. and Ind.*, No. 13, 549-554 (1965).
28. Hess S. W., A Dynamic Programming Approach to R & D Budgeting and Project Selection, *IRE Trans. on Eng. Management*, EM-9, 170-179 (1962).
29. IBM, Project Control System/360 (360A-CP-06X) Program Description and Operations Manual, No. H20-0376-0, IBM Corporation, White Plains, N.Y., 1967.
30. Kelley J. E., Critical-Path Planning and Scheduling: Mathematical Basis, *Operations Res.*, 9, No. 3, 296-320 (1961).
31. Kiefer D. M., Industrial Chemical Research, *Chem. Eng. News*, 42, No. 12, 88-109 (1964).
32. Lerda-Olberg S., Bibliography on Network-Based Project Planning and Control Techniques: 1962-1965, *Operations Res.*, 15, 925-931 (1966).
33. Moder J. J., Phillips C. R., Project Management with CPM and PERT, Van Nostrand Reinhold Company, N. Y., 1970.
34. Moshman J., Johnson J., Larsen M., RAMPS — a Technique for Resource Allocation and Multi-Project Scheduling. C-E-I-R, Inc., Arlington, Virginia, 1963.
35. Moore J. R., R & D Project Selection: Theoretical and Computational Analysis of a Project Scoring Model, Ph. D. Dissertation, Purdue University, Lafayette, Indiana, 1968.
36. Patterson J. H., Huber W. D., A Horizon-Varying, Zero-One Approach to Project Scheduling, *Management Sci.*, 20, No. 6, 990-998 (1974).
37. Paulson B. C., Jr., Man-Computer Concepts for Project Management, Technical Report No. 148, Department of Civil Engineering, Stanford University, Stanford, California, 1971.
38. Pritsker A., Happ W., GERT: Graphical Evaluation and Review Technique—Part I, Fundamentals, *J. Ind. Eng.*, 17, No. 5, 267-274 (1966).
39. Rosen E. M., Souder W. E., A Method for Allocating R & D Expenditures, *IEEE Trans. on Eng. Management*, EM-12, pp. 87-93 (1965).
40. Rubenstein A. H., Studies of Project Selection in Industry, in B. V. Dean (Ed.), *Operations Research in Research and Development*, Wiley, N. Y., 1963, pp. 189-205.
41. Rubenstein A. H., Program Summary 1967-1970, Document 70/31, POROMRAD, Department of Industrial Engineering, Northwestern University, Evanston, Illinois, 1970.
42. Shumway C. R., Maher P. M., Baker N. R., Souder W. E., Rubenstein A. H., Diffuse Decision Making: An Empirical Examination, *Management Sci.*, 21, No. 6, 697-707 (1975).
43. Souder Wm. E., R & D Project Selection: a Budgetary Approach, *Trans. CCDA*, 1966, pp. 25-43.
44. Souder Wm. E., Planning R & D Expenditures with the Aid of Computer, *Budgeting*, XIV, 25-32 (1966).
45. Souder Wm. E., Solving Budgeting Problems with O. R., *Budgeting*, XIV, 9-11 (1967).
46. Souder Wm. E., Selecting and Staffing R & D Projects Via Op Research, *Chem. Eng. Progress*, 63, 27 (November 1967).
47. Souder Wm. E., Experiences with an R & D Project Control Model, *IEEE Trans. on Eng. Management*, EM-15, 39-49 (March 1968).

48. Souder Wm. E., Cost/Progress — a Breakthrough in Operation Budgeting *Managerial Plan.*, XVII, 1-9 (January/February 1969).
49. Souder Wm. E., The Validity of Subjective Probability of Success Forecasts by R & D Managers, *IEEE Trans. on Eng. Management*, EM-16, No. 1, 34-49 (February 1969).
50. Souder Wm. E., Budgeting R & D, *Bus. Horizons*, XIII, 31-38 (June 1970); reprinted in Readings in Business Policy, B.P. Coleman and J. W. Bonge, MacMillan, 1971; reprinted in Product Planning, A. E. Spitz, Auerbach, 1972, pp. 114-123.
51. Souder Wm. E., Suitability and Validity of Project Selection Models, Ph. D. Dissertation, St. Louis University, St. Louis, Missouri, August 1970.
52. Souder Wm. E., R²: Some Results from Studies of the Research Management Process, Pro. AMIF, March 1971, pp. 121-130.
53. Souder Wm. E., A Comparative Analysis of Risky Investment Planning Algorithms, *AIIE Trans.*, 4, No. 1, 56-62 (March 1972).
54. Souder Wm. E., A Scoring Methodology for Assessing the Suitability of Management Science Models, *Management Sci.*, 18, No. 10, 526-543 (June 1972).
55. Souder Wm. E., An R & D Planning and Control Servosystem: A Case Study, *R & D Management*, 3, No. 1, 5-12 (October 1972).
56. Souder Wm. E., Effectiveness of Mathematical Programming Models for Project Selection: A Computational Evaluation, *Management Sci.*, 19, No. 8, 907-923 (April 1973).
57. Souder Wm. E., Acceptability and Utility of Project Selection Models in Development R & D, *Management Sci.*, 19, No. 12, 1384-1394 (August 1973).
58. Souder Wm. E., Autonomy, Gratification and R & D Outputs: A Small Sample Field Study, *Management Sci.*, 20, No. 8, 1147-1156 (April 1974).
59. Souder Wm. E., Achieving Organizational Consensus with Respect to R & D Project Selection Criteria, *Management Sci.*, 21, No. 6, 669-681 (February 1975).
60. Sullivan C. I., CPI Looks at R & D Project Evaluation, *Ind. and Eng. Chem.*, 53, No. 9, 42A-46A (September 1961).
61. Villers R., Research and Development: Planning and Control, Financial Executives Research Institute, Inc., 1964, pp. 30-38.
62. Watters L. D., Research and Development Project Selection: Interdependence and Multiperiod Probabilistic Budget Constraints, Ph. D. Dissertation, Arizona State University, Tempe, 1967.
63. Wiest J. D., Some Properties of Schedules for Large Projects with Limited Resources, *Operations Res.*, 12, No. 3, 395-418 (May-June 1964).
64. Wiest J. D., A Heuristic Model for Scheduling Large Projects with Limited Resources, *Management Sci.*, 13, No. 6, B359-B377 (February 1967).
65. Woodworth B. M., Dane C. W., Multiproject Network Analysis with Resource Leveling: State of Art and a Governmental Application, unpublished paper, Dept. of Business Administration, Oregon State, Corvallis, Oregon, 1975.

ДОПОЛНИТЕЛЬНАЯ ЛИТЕРАТУРА

Линейные модели распределения капиталовложений

- Asher D. T., A Linear Programming Model for the Allocation of R & D Efforts, *IRE Trans. on Eng. Management*, EM-3 (December 1962).
- Hanssman F. T., Operations Research Techniques for Capital Investment, N.Y., Wiley, 1968.
- Nutt A. B., An Approach to Research and Development Effectiveness, *IEEE*

- Trans. on Eng. Management*, EM-10, 103-112 (September 1965).
- Pound W. H., Research Project Selection: Testing a Model in the Field, *IEEE Trans. on Eng. Management*, EM-9, 16-22 (March 1964).
- Sobin D., Proposal Generation and Evaluation Methods in Research and Exploratory Development. Research Analysis Corporation Report RAC-R-11, Roanoke, Virginia, 1965.
- Souder W. E., Operations Research in R & D, The Monsanto Company, St. Louis, Missouri, 1966.

Нелинейные модели распределения капиталовложений

- Burton R., A Model for Project in Developing Countries, The Boeing Company, Renton, Washington, 1969.
- Bobis A. H., et al., A Funds Allocation Method to Improve the Odds of Research Success, *Research Management* (March 1971), pp. 34-49.
- Cooper L., A Method for Selecting R & D Projects, The Monsanto Company, February 1970.
- Daubin S. C., The Allocation of Development Funds, *Naval Res. Log. Quart.*, 3, 263-276 (September 1958).
- Hass S. W., A Dynamic Programming Approach to R & D Budgeting and Project Selection, *IRE Trans. on Eng. Managements*, EM-9, 170-179 (December 1962).
- Matheny C., A Budget Model for Procurement of Army Equipment, Technical Report Processing Center Report ARO-13-D493-A, Radford, Virginia, 1964.
- Scherer F. M., Research and Development Resource Allocation Under Rivalry, *The Quarterly J. of Economics*, 81, 359-394 (August 1967).
- Souder W. E., Rosen E. M., A Method for Allocating R & D Expenditures, *IEEE Trans. on Eng. Management*, EM-11, 87-93 (September 1965).
- Washburn A. R., A Mathematical Approach to Program Planning, The Boeing Company, Renton, Washington, 1969.

Целочисленные модели распределения капиталовложений

- Beged-Dov A. G., Optimal Assignment of R & D Project in a Large Company Using an Integer Programming Model, *IEEE Trans. on Eng. Management*, EM-12, 138-142 (December 1965).
- Freeman R. J., A Stochastic Model for Determining the Size and Allocation of the Research Budget, *IRE Trans. on Eng. Management*, EM-7, 2-7 (March 1960).
- Hanssman F. T., Operations Research Techniques for Capital Investment, N.Y., Wiley, 1968, pp. 104-134.
- Minkes A. L., Samuels J. M., Allocation of Research and Development Expenditures in the Firm, *J. of Management Studies*, 3, 62-72 (February 1966).
- Sengupta S. S., Dean B. V., On a Method for Determining Corporate Research and Development Budgets, in: *Management Science Models and Techniques*, Churchman C. W., Verhulst M. (Eds.), N.Y., Pergamon Press, 1960.
- Soewarso M., Long Range Planning in a National Telephone System, Ph. D. Dissertation, Munich Technical University, Munich, 1967.
- Bakanas V., An Analytical Method to Aid in the Choice of Long Range Study Tasks, Ithaca, N.Y., Cornell Aeronautic Laboratory, May 1964.
- Dean B. V., Project Evaluation: Methods and Procedures, American Management Association Report No. 39, N.Y., 1970.
- Gargiulo G. R., Hannon J., Hertz D. B., Zang J., Developing Systematic Procedures for Directing Research Programs, *IRE Trans. on Eng. Management*, EM-8, 24-29 (March 1971).

- Hertz D. B., Selection, Evaluation and Control of R & D Projects, Operations Research in Research and Development, Dean B. V. (Ed.), N.Y., Wiley, 1963.
- Harris J. S., Evaluating New Project Proposals, *Chem. and Eng. News*, 15, 14-18 (April 17, 1961).
- Mottley C. M., Newton R. D., The Selection of Projects for Industrial Research, *Operations Res.*, No. 6, 740-751 (1959).
- Nishry M. J., Evaluation Selection and Assignment of Industrial Research Projects, The Boeing Company, Renton, Washington, 1964.

Оценочные модели распределения капиталовложений

- Ansoff H. I., Evaluation of Applied R & D in A Firm, Technological Planning on the Corporate Level, Bright J. R. (Ed.), Cambridge, Harvard University Press, 1964, pp. 12-19.
- Disman S., Selecting R & D Projects for Profit, *Chem. Eng.*, 11, 87-90 (December 1962).
- Gloskey C., Research on a Research Department: An Analysis of Economic Decisions on Projects, *IRE Trans. on Eng. Management*, 11, 166-172 (December 1960).
- Hirsch J. H., Fisher E. K., The Alternative Service Concept in R & D Evaluation, Gulf Research and Development Company, Pittsburgh, Pennsylvania, 1962.
- Nyland H. V., Towle G. R., How We Evaluate Return from Research: Experience of an Oil Company, *National Association of Cost Accounts Bulletin*, 37, 1092-1099 (May 1956).
- Olsen J. J., Winds of Change in Chemical Research, *Chem. and Eng. News*, 12, 95-96 (March 1964).
- Pappas G., McLaren D., An Approach to Research Planning, *Chem. Eng. Progress*, 57, 65-69 (May 1964).
- Sobelman D., A Model for R & D, *Nav. Res. Quart.*, 6, No. 3 (1966).

Модели прибыльности

- Cramer R., Smith B., Decision Models for the Selection of Research Projects, *The Engineering Economist*, 9, 1-20 (January/February 1964).
- Green P. E., Risk Attitudes and Chemical Investment Decisions, *Chem. Eng. Progress*, 59, 35-40 (January 1963).
- Markowitz H., Portfolio Selection, N.Y., Wiley, 1960.
- Sharpe W. F., A Simplified Model for Portfolio Analysis, *Management Sci.*, 9, No 2, 277-293 (1963).

I-9. Теория надежности

Э. Шойер¹⁾

ОБЩИЕ ПОЛОЖЕНИЯ

В основе теории надежности лежит представление о вероятности безотказной работы устройства, которое случайно может выйти из строя. При этом вероятность определяется как с учетом, так и без учета срока службы (эксплуатационной долговечности или «возраста») устройства. Для характеристики эксплуатационной пригодности устройства вводится понятие индикатора неисправности, или индикатора состояния, который является случайной величиной X . Если устройство这件 пригодно к эксплуатации, то $X = 1$, в противном случае $X = 0$. Следовательно, вероятность безотказной работы устройства есть не что иное, как математическое ожидание случайной величины X , т. е. $P(X = 1)$.

Пример 1. Если вероятность исправности, т. е. пригодности устройства к эксплуатации, равна p , а вероятность его неисправности — $(1 - p)$, то вероятность безотказной работы устройства выражается как

$$1 \cdot p + 0 \cdot (1 - p) = p. \quad (1)$$

Пример 2. Распределение продолжительности срока службы T устройства описывается функцией F . Если исправность устройства означает, что оно безотказно работает в течение времени t , и если определена индикаторная функция X

$$X(u) = \begin{cases} 0, & u < t, \\ 1, & u \geq t, \end{cases}$$

то функция надежности (или вероятность безотказной работы как функция времени) имеет вид

$$\int_0^t 0 \cdot dF(u) + \int_t^\infty 1 \cdot dF(u) = 1 - F(t). \quad (2)$$

¹⁾ Ernest M. Scheuer, Ph. D., Professor of Management Science in the School of Business Administration and Economics at the California State University at Northridge.

Заметим, что полученный результат следует из результата примера 1, так как вероятность безотказной работы, обозначенная в примере 1 через p , в данном случае есть $1 - F(t)$ [2].

Устройство может представлять собой элемент системы, взаимозаменяемый или сменный блок, либо целую систему. Надежность системы определяется надежностью ее элементов и их взаимосвязью и структурой системы. Структура системы может быть настолько сложной, что задача точного определения надежности системы оказывается практически неразрешимой. В таких случаях иногда можно ограничиться определением доверительных границ (или интервалов) оценок надежности или прибегнуть к моделированию на ЭВМ. Надежность системы может быть повышена путем резервирования элементов, использования запасных элементов (или модулей), профилактического ремонта и замещения элементов (или модулей) до либо во время отказа.

Поскольку законы распределения вероятностей безотказной работы элементов системы могут быть неизвестны, то возникает необходимость в оценке ее надежности на основании имеющихся статистических данных. К сожалению, в пределах одной главы не представляется возможным рассмотреть все аспекты этой проблемы. Поэтому мы ограничимся исследованием связи между состоянием системы и ее элементами, и в первую очередь определением зависимостей, существующих между характеристиками, определяющими исправность системы и ее элементов.

1. СТРУКТУРНЫЕ СВОЙСТВА СИСТЕМ

Рассмотрим систему, состоящую из n элементов $c_1, c_2, \dots, c_n$, с индикаторами состояния $X_1, X_2, \dots, X_n$,

$$X_i = \begin{cases} 1, & \text{если элемент } c_i \text{ исправен,} \\ 0, & \text{если элемент } c_i \text{ неисправен.} \end{cases}$$

Индикаторная функция φ системы определяется следующим образом:

$$\varphi = \begin{cases} 1, & \text{если система работоспособна,} \\ 0, & \text{если система неработоспособна.} \end{cases}$$

Предположим, что информации о системе вполне достаточно, чтобы определить ее состояние через состояния всех входящих в нее элементов, тогда

$$\varphi = \varphi(\mathbf{x}),$$

где $x = (X_1, X_2, \dots, X_n)$ — вектор состояния системы. Функция $\varphi(x)$ называется структурной функцией системы.

Пример 3. Система последовательно соединенных элементов $c_1, c_2, \dots, c_n$ работоспособна тогда и только тогда, когда каждый элемент системы исправен. В этом случае структурная функция системы имеет вид

$$\varphi(x) = \prod_{i=1}^n X_i = \min(X_1, X_2, \dots, X_n). \quad (3)$$

Пример 4. Система параллельно соединенных элементов работоспособна тогда и только тогда, когда по крайней мере один из элементов системы исправен. Структурная функция системы имеет вид

$$\varphi(x) = \max(X_1, X_2, \dots, X_n) = 1 - \prod_{i=1}^n (1 - X_i) \equiv \bigvee_{i=1}^n X_i. \quad (4)$$

Пример 5. Система элементов, соединенных по схеме (k из n), работоспособна тогда и только тогда, когда по крайней мере k из n элементов системы исправны. [Таким образом, последовательное соединение элементов является соединением типа (n из n), а параллельное — соединением типа (1 из n).] Структурная функция системы элементов, соединенных по схеме (k из n), имеет вид

$$\varphi(x) = \begin{cases} 1, & \text{если } \sum_{i=1}^n X_i \geq k, \\ 0, & \text{если } \sum_{i=1}^n X_i < k, \end{cases} \quad (5)$$

или

$$\varphi(x) = \bigvee X_{i_1}, X_{i_2}, \dots, X_{i_k} = \max(X_{i_1}, X_{i_2}, \dots, X_{i_k}),$$

где операция $\bigvee$ и нахождение максимального значения структурной функции системы осуществляются по всем k , выбранным из первых n целых чисел.

Структурная функция системы элементов, соединенных по схеме (2 из 3), может иметь различные представления, например

$$\varphi(X_1, X_2, X_3) = \begin{cases} 1, & \text{если } X_1 + X_2 + X_3 \geq 2, \\ 0, & \text{если } X_1 + X_2 + X_3 < 2, \end{cases}$$

$$\begin{aligned} \varphi(X_1, X_2, X_3) &= X_1 X_2 \bigvee X_2 X_3 = \\ &= 1 - (1 - X_1 X_2) (1 - X_1 X_3) (1 - X_2 X_3) = \\ &= X_1 X_2 X_3 + X_1 X_2 (1 - X_3) + X_1 (1 - \\ &\quad - X_2) X_3 + (1 - X_1) X_2 X_3. \end{aligned}$$

1.1. Связанные системы

Прежде чем перейти к рассмотрению связанных систем, сформулируем основную теорему декомпозиции структурной функции и введем понятие «связанная система».

Основная теорема декомпозиции ¹⁾. Любая структурная функция n -го порядка может быть представлена в виде

$$\varphi(x) = X_i \varphi(1_i, x) + (1 - X_i) \varphi(0_i, x) \quad (6)$$

для всех x и всех $i = 1, 2, \dots, n$,

или

$$\varphi(x) = \sum_y \prod_{j=1}^n X_j^{y_j} (1 - X_j)^{1-y_j} \varphi(y), \quad (7)$$

где суммирование выполняется по всем 2^n векторам y n -го порядка.

Если надежность (работоспособность системы) не зависит от исправности некоторого элемента, то такой элемент называется *несущественным* (в противном случае он называется *существенным*). В частности, i -й элемент является несущественным для системы со структурной функцией φ , если $\varphi(1, x) = \varphi(0_i, x)$ для всех x . Если система не имеет несущественных элементов и ее структурная функция является возрастающей ²⁾ по каждому аргументу, то такая система называется *связанной*.

Для структурной функции n -го порядка связанной системы выполняется следующее соотношение:

$$\prod_{i=1}^n X_i \leq \varphi(x) \leq \bigvee_{i=1}^n X_i, \quad (8)$$

смысл которого состоит в том, что связанная система, состоящая из n элементов, по крайней мере не менее надежна, чем система последовательно соединенных элементов, и не более надежна, чем система параллельно соединенных элементов.

¹⁾ Символы $(1_i, x)$ и $(0_i, x)$ обозначают вектор x , i -компонента которого заменена соответственно на 1 или 0, т. е. $(1_i, x) = (X_1, \dots, X_{i-1}, 1, X_{i+1}, \dots, X_n)$ и $(0_i, x) = (X_1, \dots, X_{i-1}, 0, X_{i+1}, \dots, X_n)$. Теоремы, которые приводятся в данной главе, даются без доказательств.

²⁾ На протяжении всей этой главы термины «возрастающая интенсивность отказов» и «убывающая интенсивность отказов» употребляются в нестрогом смысле, т. е. они означают «неубывающая» и «невозрастающая» соответственно.

Для любой структурной функции связанной системы можно записать

$$\varphi(x \vee y) \geq \varphi(x) \vee \varphi(y), \quad (9)$$

$$\varphi(x \cdot y) \leq \varphi(x) \cdot \varphi(y), \quad (10)$$

при этом

$$x \vee y = (X_1 \vee Y_1, \dots, X_n \vee Y_n),$$

$$x \cdot y = (X_1 Y_1, \dots, X_n Y_n).$$

Напомним, что для булевых переменных справедливы следующие соотношения:

$$x \vee y = \max(x, y) \text{ и } x \cdot y = \min(x, y).$$

Выражение (9) означает, что параллельное соединение элементов более эффективно, чем параллельное соединение систем. Из выражения (10) следует, что последовательное соединение систем эффективнее последовательного соединения элементов.

Для иллюстрации первого положения рассмотрим систему, состоящую из двух последовательно соединенных элементов (рис. 1, а) с вектором состояния $x = (X_1, X_2)$, и пару элементов с вектором состояния $y = (Y_1, Y_2)$ (на рисунке эти два элемента обозначены соответственно через 1' и 2'). Система А будет работоспособной, когда $x = (1, 0)$, $y = (0, 1)$ и когда $x = (0, 1)$, $y = (1, 0)$. Система В ни для одной из этих пар векторов состояния не будет работоспособной. Для всех других пар векторов состояния x , y показатели надежности систем А и В идентичны.

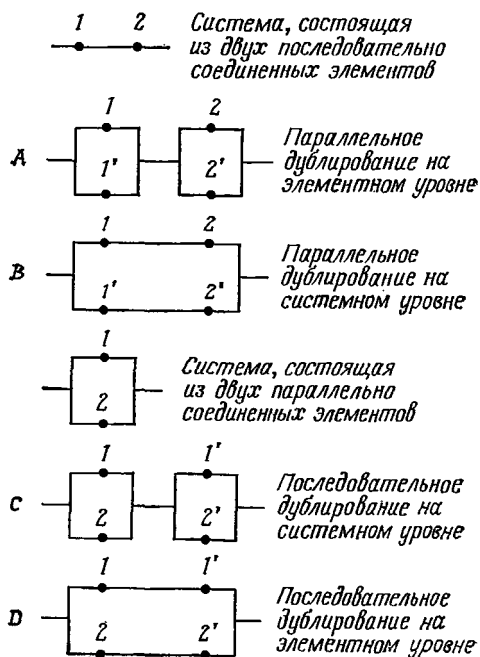
Второе положение можно проиллюстрировать на примере системы, состоящей из двух параллельно соединенных элементов (рис. 1, б). Как видно из рисунка, структура системы С совпадает со структурой системы А, а структура системы D — со структурой системы В. Поскольку система А всегда не хуже системы В, то система С по крайней мере не хуже системы D.

Прежде чем перейти к оценке надежности связанной системы, введем некоторые понятия и дополнительные обозначения.

Рассмотрим структурную функцию φ связанной системы с множеством элементов $\{c_1, \dots, c_n\}$ и вектором состояния $x = (X_1, \dots, X_n)$.

Если $\varphi(x) = 1$, то вектор x называется *вектором пути* и соответствующий ему путь выражается как $C_1(x) = \{c_i \mid X_i = 1\}$. Если вектор x является вектором пути и если для любого вектора y при $y < x$ функция $\varphi(y) = 0$, то вектор x есть *вектор минималь-*

Рис. 1. Различные формы дублирования систем последовательно и параллельно соединенных элементов.



ного пути. Соответствующий путь $C_1(x)$ является минимальным путем. Следует иметь в виду, что минимальный путь — это минимальное множество элементов, необходимых для нормального функционирования системы.

Если функция $\varphi(x) = 0$, то вектор x называется *вектором сечения* и соответствующее ему сечение обозначается как $C_0(x)$, при этом $C_0(x) = \{c_i \mid X_i = 0\}$. Если вектор x является вектором сечения и если для произвольного вектора y при $y > x$ функция $\varphi(y) = 1$, то вектор x есть вектор минимального сечения, а соответствующее ему сечение является *минимальным сечением*. Следует иметь в виду, что минимальное сечение — это минимальное множество элементов, отказ которых приводит к потере работоспособности системы.

Обозначим число минимальных путей через r и j -й минимальный путь через P_j , число минимальных сечений — через k и j -е минимальное сечение — через K_j . Определим структурную функцию, соответствующую системе элементов P_j , как

$$\alpha_j(x) = \prod_{c_i \in P_j} X_i. \quad (11)$$

Очевидно, что функция α_j есть структурная функция системы, у которой все элементы j -го минимального пути соединены последовательно, т. е. если все элементы P_j исправны, то $\alpha_j = 1$, в противном случае $\alpha_j = 0$. Поскольку связанная система будет работоспособной тогда и только тогда, когда по крайней мере один из минимальных путей системы работоспособен, то такая система может быть представлена как параллельное соединение

минимальных путей и структурная функция может быть записана как

$$\varphi(\mathbf{x}) = \bigvee_{j=1}^p \alpha_j(\mathbf{x}) = 1 + \prod_{i=1}^p [1 - \alpha_j(\mathbf{x})]. \quad (12)$$

Аналогично структурная функция системы элементов K_j определяется как

$$\beta_j(\mathbf{x}) = \bigvee_{c_i \in K_j} X_i \quad (13)$$

и принимает значение, равное нулю, если все элементы K_j вышли из строя, и равное единице, если хотя бы один элемент работоспособен. Таким образом, она является структурной функцией системы параллельно соединенных элементов K_j . Поскольку связанная система отказывает тогда и только тогда, когда отказывает по крайней мере одно из ее минимальных сечений, она может быть представлена как последовательное соединение таких сечений, т. е.

$$\varphi(\mathbf{x}) = \prod_{j=1}^h \beta_j(\mathbf{x}). \quad (14)$$

2. НАДЕЖНОСТЬ СВЯЗАННЫХ СИСТЕМ

Рассмотрим сначала случай, когда элементы системы статистически независимы друг от друга (т. е. случай независимости отказов элементов). Обозначим надежность элемента c_i через p_i , надежность системы — через h . Как было показано, эти величины могут быть представлены в виде

$$p_i = P(X_i = 1) = E(X_i) \text{ и } h = P[\varphi(\mathbf{x}) = 1] = E[\varphi(\mathbf{x})].$$

Если элементы системы взаимно независимы, то надежность системы есть функция надежности элементов и, следовательно, $h = h(p)$, где $h(p)$ — функция надежности. (Следует иметь в виду, что если элементы системы статистически взаимно независимы, то надежность системы зависит не только от величины p .) Из основной теоремы декомпозиции (уравнение 6) с необходимостью следует, что

$$h(p) = p_i h(1_i, p) + (1 - p_i) h(0_i, p). \quad (15)$$

(Можно показать, что $h(p)$ — возрастающая функция по каждому из аргументов p_i при $0 < p_i < 1$.)

Используя понятие минимального сечения (или мини-сечения) и минимального пути (или мини-пути), а также основную теорему декомпозиции структурной функции, можно вычислить надеж-

ность связанной системы независимых элементов. Из уравнений (12) и (13) получаем

$$h(p) = E[\varphi(x)] = E\left[\bigvee_{j=1}^p \prod_{c_j \in P_j} X_i\right]. \quad (16)$$

Если теперь воспользоваться выражениями (13) и (14), то имеем

$$h(p) = E\left[\bigvee_{j=1}^h \prod_{c_i \in K_j} X_i\right], \quad (17)$$

а с учетом уравнения (7) функция $h(p)$ может быть записана в виде

$$h(p) = \sum_y \varphi(y) \prod_{j=1}^n p_j^{y_j} (1-p_j)^{1-y_j}, \quad (18)$$

где суммирование производится по всем 2^n векторам y с координатами 0 и 1. Преобразования правых частей уравнений (16) и (17) выполнялись с учетом соотношения $X_i^2 = X_i$ и независимости элементов. Принимая во внимание последнее, математическое ожидание произведения будет равно произведению математических ожиданий. В ряде случаев вычисление (точное определение) надежности системы может оказаться трудоемкой процедурой даже при упрощающем предположении о независимости элементов. В связи с этим рассмотрим возможные способы оценки надежности систем. Однако сначала введем понятие «связанная случайная величина».

2.1. Связанные случайные величины

Случайные величины $T_1, \dots, T_n$ (не обязательно булевы) являются связанными, если для всех пар возрастающих булевых функций Γ и Δ выполняется соотношение

$$\text{cov}[\Gamma(T), \Delta(T)] = E[\Gamma(T) \Delta(T)] - E[\Gamma(T)] E[\Delta(T)] \geq 0.$$

Связанные случайные величины обладают следующими свойствами:

любое подмножество связанных случайных величин является связанным;

если два множества связанных случайных величин взаимно независимы, то их объединение есть множество связанных случайных величин;

множество, состоящее из одной случайной величины, является связанным;

независимые случайные величины являются связанными;

возрастающие функции связанных случайных величин являются связанными.

Последнее свойство означает, что монотонные системы, построенные из связанных (в частности, независимых) элементов, сами являются связанными. Структурные функции минн-пути (и минн-сечения) связанной системы являются возрастающими относительно индикаторных переменных $X_1, X_2, \dots, X_n$ состояния элементов. Если эти переменные являются связанными, то и структурные функции минн-путей (и минн-сечений) являются связанными.

2.2. Пределы надежности систем

Если $X_1, X_2, \dots, X_n$ — связанные случайные булевы величины, то

$$P\left(\bigcap_{i=1}^n X_i = 1\right) \geq \prod_{i=1}^n P(X_i = 1), \quad (19)$$

$$P\left(\bigvee_{i=1}^n X_i = 1\right) \leq \bigvee_{i=1}^n P(X_i = 1). \quad (20)$$

Выражение (19) означает, что нижняя граница оценки надежности системы, состоящей из последовательно соединенных связанных элементов, основана на представлении о независимости элементов; выражение (20) означает, что при вычислении верхней границы оценки надежности системы, состоящей из параллельно соединенных связанных элементов, также используется предположение об их независимости. Грубая оценка надежности системы связанных элементов может быть получена непосредственно из выражений (8), (19) и (20), т. е. для связанной системы связанных элементов, надежности которых соответственно равны $p_1, p_2, \dots, p_n$, имеем

$$\prod_{i=1}^n p_i \leq P[\varphi(x) = 1] \leq \bigvee_{i=1}^n p_i. \quad (21)$$

Ниже приводятся наиболее достоверные границы оценки надежности системы, структурная функция которой представляется в виде минн-путей и минн-сечений ¹⁾.

¹⁾ Хотя можно предположить, что нижняя оценка в выражениях (22) или (23) (основанных фактически на представлении структуры φ минн-путями) должна быть всегда лучше, чем тривиальная нижняя оценка в выражении (21) (которая вообще не использует информации о φ), по это не так. Например, для системы (3 из 4), составленной из независимых элементов, надежность каждого из которых равна p , тривиальная оценка на самом деле лучше, чем основанная на минн-сечениях нижняя оценка. Однако это исключительный случай, и, как правило, нижняя оценка с помощью минн-сечений всегда лучше для «не слишком малых» значений p .

Пусть q — структурная функция связанной системы, состоящей из связанных элементов; $\alpha_1(x), \dots, \alpha_p(x)$ — элементы минипутей и $\beta_1(x), \dots, \beta_k(x)$ — элементы минисечений. Тогда получаем

$$\prod_{j=1}^h P[\beta_j(x) = 1] \leq P[q(x) = 1] \leq \bigvee_{j=1}^p P[\alpha_j(x) = 1]. \quad (22)$$

Если элементы системы независимы, то имеем

$$\prod_{j=1}^h \bigvee_{c_j \in K_j} p_i \leq P[q(x) = 1] \leq \bigvee_{j=1}^p \prod_{c_j \in P_j} p_i. \quad (23)$$

Границы оценки надежности, которые дает выражение (23), являются более точными, но менее сильными, чем границы, которые устанавливаются с помощью выражения (22).

2.2. Модульная декомпозиция

Более доступным методом вычисления точного значения и улучшения доверительных границ оценки надежности систем является метод декомпозиции, т. е. разложение системы на отдельные подсистемы, или модули [8]. В этом случае надежность системы вычисляется или оценивается на основании надежности ее модулей.

Анализ надежности сложной системы может быть выполнен следующим образом. Систему разбивают на отдельные функциональные подсистемы, каждую из которых в свою очередь разбивают на отдельные элементы, а каждый элемент — на отдельные составляющие. Затем, используя выражения (22) и (23), определяют нижнюю границу оценки надежности элемента на основании данных о надежности его составляющих. Полученные значения надежности элементов позволяют определить нижнюю границу оценки надежности каждой подсистемы, что в свою очередь используется для установления нижней границы оценки надежности системы. Можно показать, что такой метод позволяет получить нижнюю границу оценки надежности столь же доверительную, что и метод представления системы в виде минисечений.

В работе [8] приводятся также верхние границы оценки надежности системы, полученные методом декомпозиции. Общие рекомендации относительно получения доверительных границ оценки надежности можно найти в работе [3].

3. ВРЕМЯ БЕЗОТКАЗНОЙ РАБОТЫ И СООТВЕТСТВУЮЩИЕ СЛУЧАЙНЫЕ ВЕЛИЧИНЫ

Обозначим через F интегральную функцию распределения времени безотказной работы некоторого, случайным образом выходящего из строя устройства и через $\bar{F} = (1 - F)$ — вероятность безотказной работы, или функцию надежности устройства. Поскольку время безотказной работы является неотрицательной случайной величиной, то, следовательно, $F(-0) = 0$.

Условная вероятность того, что устройство, проработавшее в течение времени t , проработает еще в течение некоторого времени x , равна

$$F(x | t) = \frac{\bar{F}(t+x)}{\bar{F}(t)}, \quad \bar{F}(t) > 0. \quad (24)$$

Условная вероятность того, что устройство, проработавшее в течение времени t , выйдет из строя в период между моментами времени t и $t + x$, равна

$$F(x | t) = 1 - \bar{F}(x | t) = \frac{F(t+x) - F(t)}{\bar{F}(t)}. \quad (25)$$

Условная функция интенсивности отказов r определяется следующим образом:

$$r(t) = \lim_{x \rightarrow 0} F(x | t)/x. \quad (26)$$

Если интегральной функции распределения F соответствует плотность f , то

$$r(t) = f(t) / \bar{F}(t). \quad (27)$$

Очевидны следующие соотношения:

$$r(t) = -\frac{d}{dt} \log [\bar{F}(t)], \quad (28)$$

$$\bar{F}(t) = \exp \left[- \int_0^t r(x) dx \right], \quad (29)$$

$$f(t) = r(t) \exp \left[- \int_0^t r(x) dx \right]. \quad (30)$$

Для малых Δt величина $r(t) \Delta t$ может быть интерпретирована как условная вероятность отказа в интервале времени $(t, t + \Delta t)$ устройства, проработавшего время t .

3.1. Экспоненциальное распределение времени безотказной работы

Если условная вероятность безотказной работы устройства в течение дополнительного времени x не зависит от его «возраста» t , то такое устройство называется стохастически не стареющим и для всех x и $t \geq 0$ имеем

$$\bar{F}(x | t) = \bar{F}(x), \quad (31)$$

или

$$F(t + x) = \bar{F}(t) \bar{F}(x). \quad (32)$$

Единственной непрерывной функцией надежности, удовлетворяющей этому уравнению, является функция вида

$$\bar{F}(x) = \exp(-\lambda x), \quad \lambda > 0, \quad x \geq 0, \quad (33)$$

т. е. в данном случае мы имеем дело с показательным распределением. Это — единственное распределение, которое характеризуется постоянной интенсивностью отказов λ . Для X , имеющей показательное распределение с интенсивностью отказов λ , $E(X^n) = n!/\lambda^n$, среднее значение времени безотказной работы равно $1/\lambda$ и дисперсия равна $1/\lambda^2$. Плотность показательного распределения задается как

$$f(x) = \begin{cases} \lambda \exp(-\lambda x), & x \geq 0, \\ 0, & x < 0. \end{cases} \quad (34)$$

Распределение времени безотказной работы системы, состоящей из n последовательно соединенных независимых элементов с показательным распределением времени безотказной работы, причем i -й элемент имеет интенсивность отказов λ_i , подчиняется экспоненциальному закону и характеризуется интенсивностью отказов $\sum \lambda_i$ и средним значением $1/\sum \lambda_i$. Установить это можно следующим образом. Пусть T — время безотказной работы системы и T_i — время безотказной работы i -го элемента. Тогда

$$\begin{aligned} P(T > t) &= P(\min(T_1, \dots, T_n) > t) = \\ &= P(T_1 > t, \dots, T_n > t) = \prod_{i=1}^n P(T_i > t) = \\ &= \prod_{i=1}^n \exp(-\lambda_i t) = \exp(-\sum \lambda_i t). \end{aligned}$$

Распределение времени безотказной работы системы, состоящей из n параллельно соединенных независимых элементов с показательным распределением времени безотказной работы, не подчиняется экспоненциальному закону. В этом случае распре-

деление времени безотказной работы и его среднее значение могут быть представлены в виде

$$P(T \leq t) = P[\max(T_1, \dots, T_n) \leq t] = P[T_1 \leq t, \dots, T_n \leq t] = \\ = \prod_{i=1}^n P(T_i \leq t) = \prod_{i=1}^n [1 - \exp(-\lambda_i t)]. \quad (35)$$

$$E(t) = \sum_{i=1}^n (1/\lambda_i) - \sum_{i,j=1}^n [1/(\lambda_i + \lambda_j)] + \\ + \sum_{i,j,k=1}^n [1/(\lambda_i + \lambda_j + \lambda_k)] - \dots \pm 1/\sum_{i=1}^n \lambda_i. \quad (36)$$

В частном случае, когда $\lambda_1 = \lambda_2 = \dots = \lambda_n = \lambda$,

$$E(t) = \frac{1}{\lambda} \left(1 + \frac{1}{2} + \dots + \frac{1}{n} \right). \quad (37)$$

3.2. Распределение времени безотказной работы с возрастающей и убывающей интенсивностью отказов

Если условная вероятность безотказной работы устройства уменьшается с его «возрастом», то такое устройство называется стареющим, т. е. функция надежности

$$\bar{F}(x | t) \text{ убывает по } t \text{ при } t \geq 0 \text{ для любого } x \geq 0. \quad (38)$$

Если распределение F имеет плотность, то это означает, что интенсивность отказов r возрастает во времени. Следовательно, если распределение удовлетворяет условию (38), то оно является распределением с возрастающей интенсивностью отказов.

Если условная вероятность безотказной работы устройства увеличивается с его «возрастом», то такое устройство называется нестареющим. В этом случае функция надежности

$$\bar{F}(x | t) \text{ возрастает по } t \text{ при } t \geq 0 \text{ для любого } x \geq 0. \quad (39)$$

Если распределение F имеет плотность, то это означает, что интенсивность отказов убывает во времени. Следовательно, если распределение F удовлетворяет условию (39), то оно является распределением с убывающей интенсивностью отказов. Композиция¹⁾ показательных распределений дает распределение с убывающей

¹⁾ Композиция конечного, или счетного, числа распределений $F_1, F_2, \dots$ является распределением продолжительности срока службы устройства $F(t) = \sum p_i F_i(t)$, где все p_i неотрицательны и $\sum p_i = 1$. В более общем виде: пусть $\{F_\alpha\}$ обозначает семейство функций распределения, индексированных параметром α , который сам имеет функцию распределения $G(\alpha)$. Тогда результирующая смесь семейства распределений $\{F_\alpha\}$ определяется выражением $F(t) = \int_{-\infty}^{\infty} F_\alpha(t) dG(\alpha)$.

вающей интенсивностью отказов. Некоторые механические детали с «возрастом» увеличивают свою прочность («механическое упрочнение»), и распределение времени безотказной работы таких деталей является распределением с убывающей интенсивностью отказов. Убывающая интенсивность отказов характерна и для начальных периодов функционирования некоторых форм живых организмов, а также электронных элементов.

Класс распределений с возрастающей интенсивностью отказов является замкнутым относительно множества систем последовательно соединенных элементов. Аналогично класс распределений с убывающей интенсивностью отказов замкнут относительно систем последовательно соединенных элементов.

В ряде случаев функция интенсивности отказов вначале убывает (период выхода из строя на начальной стадии эксплуатации), затем в течение некоторого времени остается постоянной (период нормального функционирования) и наконец возрастает (период изнашивания).

3.3. Некоторые типы распределений времени безотказной работы с монотонно меняющейся интенсивностью отказов

Рассмотрим следующие типы распределений времени безотказной работы с монотонно меняющейся интенсивностью отказов во времени.

Распределение Вейбулла. Функция распределения и функция интенсивности отказов соответственно имеют вид

$$F(t) = 1 - \exp [-(\lambda t)^\alpha] \text{ для } t \geq 0 \text{ и } \lambda, \alpha > 0, \quad (40)$$

$$r(t) = \alpha \lambda (\lambda t)^{\alpha-1} \text{ для } t > 0. \quad (41)$$

Таким образом, если $\alpha \geq 1$, то мы имеем распределение Вейбулла с возрастающей интенсивностью отказов, если $0 < \alpha \leq 1$, то с убывающей интенсивностью отказов и при $\alpha = 1$ — показательное распределение. Среднее время безотказной работы и дисперсия соответственно равны

$$(1/\lambda)^{1/\alpha} \Gamma\left(\frac{1}{\alpha} + 1\right) \text{ и } (1/\lambda)^{2/\alpha} \left[\Gamma\left(\frac{2}{\alpha} + 1\right) - \Gamma^2\left(\frac{1}{\alpha} + 1\right) \right].$$

Гамма-распределение. Плотность гамма-распределения может быть выражена как

$$g(t) = \frac{\lambda^\alpha t^{\alpha-1}}{\Gamma(\alpha)} \exp(-\lambda t), \quad t \geq 0, \\ \lambda, \alpha > 0, \quad (42)$$

где λ — масштабный параметр, а α — параметр формы распределения.

Для положительных целочисленных значений α функция гамма-распределения может быть записана следующим образом:

$$G(t) = 1 - \sum_{i=0}^{\alpha-1} \frac{(\lambda t)^i}{i!} \exp(-\lambda t), \quad t \geq 0. \quad (43)$$

(Заметим, что в такой записи функция распределения выражается через функцию распределения Пуассона.)

Для всех α таких, что 2α — положительное целое, функция $F(2\alpha t)$ является функцией хи-распределения с m ($m = 2\alpha$) степенями свободы. Наиболее полно возможные значения функций гамма-распределения представлены в работах [12, 19]. Сумма n независимых показательно распределенных случайных величин, каждая с интенсивностью λ , имеет гамма-распределение с масштабным параметром λ и параметром формы n . Среднее значение и дисперсия гамма-распределения равны α/λ и α/λ^2 соответственно.

Величина, обратная функции интенсивности отказов для гамма-распределения, может быть записана как

$$1/r(t) = \int_0^{\infty} (1+u/t)^{\alpha-1} \exp(-\lambda u) du. \quad (44)$$

Отсюда следует, что если $0 < \alpha \leq 1$, то мы имеем гамма-распределение с убывающей интенсивностью отказов, если $\alpha \geq 1$, то с возрастающей интенсивностью отказов и при $\alpha = 1$ — показательное распределение.

Усеченное нормальное распределение. Плотность усеченного нормального распределения выражается как

$$f(t) = \frac{1}{a\sigma\sqrt{2\pi}} \exp[-(t-\mu)^2/(2\sigma^2)], \\ t \geq 0, \quad \sigma > 0, \quad (45)$$

где

$$a = \int_0^{\infty} \frac{1}{\sigma\sqrt{2\pi}} \exp[-(t-\mu)^2/(2\sigma^2)] dt.$$

Константа a является нормирующей постоянной, т. е. $\int_0^{\infty} f(t) dt = 1$. Если среднее μ более чем в три раза превышает величину σ , то значение a близко к единице и в большинстве случаев может быть опущено. В результате выражение (44) соответствует обычному (неусеченному) нормальному распределению со средним μ и дисперсией σ^2 . Усеченное нормальное распределение является

распределением с возрастающей интенсивностью отказов. Этот факт непосредственно следует из свойств класса функций, называемых функциями плотности Поля второго порядка (сокращение PF_2) [2, 3, 13].

3.4. Другие типы распределений времени безотказной работы

Естественно было бы предположить, что связанные структуры, состоящие из элементов с возрастающей интенсивностью отказов, будут также характеризоваться возрастающей интенсивностью отказов. Однако на самом деле это не так. Как показано в работе [7], подобное предположение оказывается справедливым только в тех случаях, когда функционал $(1/t) \log \bar{F}(t)$ возрастает при $t \geq 0$. Если распределение F имеет плотность, то

это означает, что функционал $(1/t) \int_0^t r(x) dx$ возрастает при $t \geq 0$,

и такой тип распределения называется распределением с возрастающей средней интенсивностью отказов. Заметим, однако, что распределение F не обязательно должно иметь плотность, для того чтобы относиться к такому типу распределений.

Поскольку показательное распределение, по существу, есть распределение с возрастающей интенсивностью отказов, которое в свою очередь является распределением с возрастающей средней интенсивностью отказов, то связанные структуры, состоящие из элементов с экспоненциальной или возрастающей интенсивностью отказов, имеют возрастающую среднюю интенсивность отказов.

Аналогично, если функционал $(1/t) \log \bar{F}(t)$ при возрастании t убывает, распределение является распределением с убывающей интенсивностью отказов. Такой тип распределения не замкнут относительно множества связанных структур и замкнут относительно операции смешения и относительно множества систем последовательно соединенных элементов.

Рассмотрим теперь распределение типа «Новое лучше Существующего», «Новое хуже Существующего», «Новое в среднем лучше Существующего» и «Новое в среднем хуже Существующего». (Подобные типы распределений возникают естественным образом при учете стратегий замещения.)

Распределение F является распределением типа «Новое лучше Существующего» (или «Новое хуже Существующего»), если

$$\bar{F}(x+y) \leq (\geq) \bar{F}(x) \bar{F}(y), \quad x, y \geq 0. \quad (46)$$

Это эквивалентно требованию, согласно которому условная вероятность $\bar{F}(y | x) = \bar{F}(x+y) / \bar{F}(x)$ того, что устройство без-

отказавшее проработавшее время x , прослужит еще дополнительное время y , меньше (больше), чем вероятность $\bar{F}(y)$ того, что новое устройство прослужит время y . Равенство в соотношении (45) имеет место только для показательного распределения.

Распределение является распределением типа «Новое в среднем лучше Существующего» (или «Новое в среднем хуже Существующего»), если, во-первых, распределение F имеет конечное (или бесконечное) среднее μ и, во-вторых, выполняется следующее соотношение:

$$\int_0^{\infty} \bar{F}(x) dx \leq (\geq) \mu \bar{F}(t), \quad t \geq 0. \quad (47)$$

Это эквивалентно требованию, согласно которому среднее значение оставшегося срока службы устройства «возраста» t , $\int_0^{\infty} [\bar{F}(x)/\bar{F}(t)] dx$, меньше (больше), чем средний срок службы μ нового устройства.

Рассмотренные типы распределений могут быть представлены в виде следующих цепочек строгих включений:

Распределение с возрастающей интенсивностью отказов $\subset$ Распределение с возрастающей средней интенсивностью отказов $\subset$ «Новое лучше Существующего» $\subset$ «Новое в среднем лучше Существующего», (48)

Распределение с убывающей интенсивностью отказов $\subset$ Распределение с убывающей средней интенсивностью отказов $\subset$ «Новое хуже Существующего» $\subset$ «Новое в среднем хуже Существующего».

3.5. Процессы восстановления работоспособности систем

Рассмотрим непрерывный процесс, в котором отказавшее устройство немедленно заменяется новым устройством с тем же распределением продолжительности срока службы, что и у предшествующего, и срок службы одного не зависит от срока службы другого. Если $F(t)$ обозначает интегральную функцию распределения продолжительности срока службы, а $F^{(k)}(t)$ — k -кратную свертку $F(t)$, то для неотрицательных случайных величин имеем

$$F^{(1)}(t) = F(t),$$

$$F^{(k)}(t) = \int_0^t F^{(k-1)}(t-x) dF(x). \quad (49)$$

Функция $F^{(k)}(t)$ — это распределение случайных величин $X_1 + \dots + X_k$, где $X_1, \dots, X_k$ независимы и имеют общую функцию распределения $F(t)$. Пусть $S_1, S_2, \dots$ — моменты последовательных заменений (т. е. последовательные интервалы функционирования — это $S_1, S_2 - S_1, S_3 - S_2, \dots$). Пусть $N(t)$ есть число заменений, выполненных в период $[0, t]$; $N(t)$ — неотрицательная целочисленная случайная величина, такая, что

$$P[N(t) = 0] = P[X_1 > t] = \bar{F}(t). \quad (50)$$

$$\begin{aligned} P[N(t) = k] &= P[X_1 + \dots + X_k \leq t, X_1 + \dots + X_{k-1} > t] \\ &= F^{(k)}(t) - F^{(k+1)}(t), \quad k \geq 1. \end{aligned} \quad (51)$$

Случайный процесс $\{N(t) : t \geq 0\}$ называется простым (или дискретным) процессом восстановления. Математическое ожидание величины $N(t)$ обозначается как $M(t)$ и называется функцией восстановления. Легко показать, что

$$M(t) = \sum_{k=0}^{\infty} k P[N(t) = k] = \sum_{k=1}^{\infty} F^{(k)}(t). \quad (52)$$

Распределение Пуассона с интенсивностью λ может быть определено как простой (дискретный) процесс восстановления для распределения времени безотказной работы, которое подчиняется экспоненциальному закону и имеет интенсивность отказов λ . Если $F(t)$ является показательным распределением с интенсивностью λ , то k -кратная свертка $F^{(k)}(t)$ является гамма-распределением с масштабным параметром λ и параметром формы k . Используя этот факт и соотношение (43), получаем

$$P[N(t) = k] = (\lambda t)^k \exp(-\lambda t) / k!, \quad (53)$$

т. е. $N(t)$ имеет распределение Пуассона со средним λt . Таким образом, для показательного распределения с интенсивностью отказов λ имеем

$$M(t) = \lambda t. \quad (54)$$

Результаты анализа процесса восстановления для случая нормального распределения не могут быть представлены в таких простых формулах, как (53) и (54), но поддаются численной обработке, поскольку k -кратная свертка нормального распределения со средним μ и дисперсией σ^2 является нормальным распределением со средним $k\mu$ и дисперсией $k\sigma^2$. Функция восстановления для распределения Вейбулла была исследована в работах [14, 15].

Если $N(t)$ означает число восстановлений случайным образом выходящего из строя устройства с функцией F , имеющей среднее μ и дисперсию σ^2 , то $N(t)$ при возрастании t асимптоти-

чески приближается к нормальному распределению со средним t/μ и дисперсией $\sigma^2 t/\mu^3$. Для $N(t)$ и $M(t)$ известны следующие асимптотические приближения:

$$\lim_{t \rightarrow \infty} N(t)/t \rightarrow 1/\mu \quad (\text{п. д.})^1, \quad (55)$$

Теорема Блекуэлла. Если распределение F является перемешанным распределением²⁾ со средним μ , то

$$\lim_{t \rightarrow \infty} [M(t+h) - M(t)] = h/\mu. \quad (56)$$

Узловая теорема восстановления. Если g имеет конечную вариацию на интервале $[0, \infty]$ и F — перемешанная функция распределения со средним μ , то

$$\lim_{t \rightarrow \infty} \int_0^t g(t-u) dM(u) = \frac{1}{\mu} \int_0^{\infty} g(u) du. \quad (57)$$

Элементарная теорема восстановления.

$$\lim_{t \rightarrow \infty} M(t)/t = 1/\mu. \quad (58)$$

Более подробное изложение теории восстановления дается в работах [9, 20] и в главе о случайных процессах, написанной Смитом (т. 1, гл. 3—4). Примеры применения теории восстановления к решению задач надежности и обслуживания можно найти в работах [2, 3] и [11].

4. НЕКОТОРЫЕ ОЦЕНКИ НАДЕЖНОСТИ

Рассмотрим некоторые оценки надежности $\bar{F}(t)$ для различных типов распределения времени безотказной работы, а также оценки, применяемые при анализе восстановления и надежности последовательных, параллельных и связанных систем.

4.1. Оценки функции надежности $\bar{F}(t)$

Приводимая ниже оценка позволяет сравнить вероятности безотказной работы в случае распределений с возрастающей и убывающей интенсивностью, имеющих предписанный квантиль, с вероятностью безотказной работы в случае показательного распределения, имеющего тот же квантиль.

¹⁾ п. д. — почти достоверное событие, т. е. событие, вероятность которого равна 1.

²⁾ Случайная величина X имеет решетчатое распределение, если существует такое $h > 0$, что $P[X = nh, n = 0, 1, 2, \dots] = 1$.

Если распределение F является распределением с возрастающей средней интенсивностью отказа (или с убывающей средней интенсивностью отказов) с p -м квантилем ξ_p (т. е. $F(\xi_p) = p$) и $\alpha = -\log(1-p)/\xi_p$, то функция надежности определяется как

$$\bar{F}(t) \begin{cases} \geq (\leq) \exp(-\alpha t), & 0 \leq t \leq \xi_p, \\ \leq (\geq) \exp(-\alpha t), & t \geq \xi_p. \end{cases} \quad (59)$$

Далее даются нижние и верхние границы оценки вероятностей безотказной работы для распределения с возрастающей средней интенсивностью отказов.

Если распределение F является распределением с возрастающей средней интенсивностью отказов и средним μ_1 , то для фиксированного $t > 0$ функция надежности определяется как

$$\bar{F}(t) \leq \begin{cases} 1, & t \leq \mu_1, \\ \exp(-wt), & t > \mu_1, \end{cases} \quad (60)$$

где $w > 0$ является функцией t , удовлетворяющей соотношению

$$1 - w\mu_1 = \exp(-wt). \quad (61)$$

Если распределение F является распределением с возрастающей средней интенсивностью отказов и r -м моментом μ_r

$$\left(\mu_r = \int_0^{\infty} t^r dF(t) = r \int_0^{\infty} t^{r-1} \bar{F}(t) dt \right),$$

то функция надежности определяется как

$$\bar{F}(t) \geq \begin{cases} \min[\exp(-b_1 t), \exp(-c t)], & t < \mu_r^{1/r}, \\ 0, & t > \mu_r^{1/r}. \end{cases} \quad (62)$$

где b_1 удовлетворяет уравнению

$$r [1 - \exp(-b_1 t)] + \int_t^{\infty} b_1 x^r \exp(-b_1 x) dx = \mu_r, \quad (63)$$

$$c = (\mu_r/r!)^{1/r}. \quad (64)$$

Если распределение F является непрерывным распределением с возрастающей интенсивностью отказов и известным r -м моментом μ_r и если $\lambda_r = \mu_r/r!$, $r \geq 1$, то функция надежности определяется как

$$\bar{F}(t) = \begin{cases} \exp[-t/\lambda_r^{1/r}], & t < \mu_r^{1/r}, \\ 0, & t \geq \mu_r^{1/r}. \end{cases} \quad (65)$$

Эта оценка, которая не может быть улучшена, даст нам возможность сравнить вероятность безотказной работы в случае распределения с возрастающей интенсивностью отказов с вероятностью безотказной работы в случае показательного распределения.

Если известен любой из моментов распределения F , то наиболее естественно предположить, что среднее μ_1 также известно. В этом случае в соотношении (65) $r = 1$, и получаем следующий важный результат:

Если распределение F является распределением с возрастающей интенсивностью отказов и средним μ_1 , то функция надежности определяется как

$$\bar{F}(t) \geq \begin{cases} \exp(-t/\mu_1), & t < \mu_1, \\ 0, & t \geq \mu_1. \end{cases} \quad (66)$$

Эта оценка не может быть улучшена.

В заключение определим верхнюю границу оценки вероятности выживания для распределения с убывающей интенсивностью отказов.

Если распределение F является распределением с убывающей интенсивностью отказов и средним μ_1 , то функция надежности определяется как

$$\bar{F}(t) \leq \begin{cases} \exp(-t/\mu_1), & t < \mu_1, \\ \mu_1/(te), & t \geq \mu_1. \end{cases} \quad (67)$$

Эта оценка не может быть улучшена.

4.2. Оценки надежности при анализе процесса восстановления системы

Если устройство имеет распределение продолжительности срока службы F и если в случае отказа оно мгновенно заменяется новым устройством, срок службы которого не зависит от срока службы предшествующего устройства и имеет то же распределение, то возникает проблема «запасных частей». Задача состоит в том, чтобы определить количество запасных частей (деталей, блоков и т. д.), необходимых для непрерывного функционирования устройства, при этом вероятность такого функционирования должна быть равной $1 - \alpha$.

Если распределение F есть показательное распределение со средним временем безотказной работы μ , то из выражения (53) следует, что для $n = 1, 2, \dots$ имеем

$$P[N(t) \geq n] = \sum_{j=n+1}^{\infty} (t/\mu)^j [\exp(-t/\mu)]/j!, \quad (68)$$

где $N(t)$ — число необходимых замен, P — функция, описывающая распределение Пуассона, значения которой можно найти, например, в работах [10, 18].

Вероятность нехватки запасных частей при условии, что в наличии имеется n частей, определяется с помощью соотношения (68). Таким образом, решением задачи «запасных частей» является наименьшее целое число n^* , удовлетворяющее условию $P[N(t) > n^*] \leq \alpha$.

Если ограничиться только предположением, что распределение F является распределением с возрастающей интенсивностью и средним μ , то из соотношения (66) следует, что

$$P[N(t) > n] \leq \sum_{j=n+1}^{\infty} (t/\mu)^j [\exp(-t/\mu)]/j!. \quad 0 \leq t < \mu. \quad (69)$$

Это означает, что распределение Пуассона дает оценку вероятности возникновения более чем n отказов в интервале $[0, t]$ для $t < \mu$, если распределение F является распределением с возрастающей интенсивностью отказов и средним μ .

Следующий результат [3] содержит оценки, справедливые для всех $t \geq 0$ и для других типов распределений.

Пусть F непрерывна, $F(0) = 0$, и пусть $R(t) = -\log \bar{F}(t)$. а) Если распределение F является распределением типа «Новое лучше Существующего» («Новое хуже Существующего»), то имеем

$$P[N(t) > n] \leq (\geq) \sum_{j=n+1}^{\infty} [R(t)]^j [\exp(-R(t))]/j!, \quad t \geq 0, n = 1, 2, \dots \quad (70)$$

б). Если распределение F является распределением с возрастающей (или убывающей) интенсивностью отказов, то имеем

$$P[N(t) > n] \leq (\geq) \sum_{j=n+1}^{\infty} [nR(t/n)]^j [\exp(-nR(t/n))]/j!, \quad t \geq 0, n = 1, 2, \dots \quad (71)$$

Соотношение (70) может быть применено к распределению Вейбулла (40) и при $\alpha \geq 1$. Для $\alpha \geq 1$ распределение Вейбулла является распределением с возрастающей интенсивностью и, следовательно, распределением типа «Новое лучше Существующего». Поскольку $R(t) = (\lambda t)^\alpha$, то из соотношения (70) следует

$$P[N(t) > n] \leq \sum_{j=n+1}^{\infty} [(\lambda t)^\alpha]^j [\exp(-(\lambda t)^\alpha)]/j!, \quad t \geq 0, n = 1, 2, \dots \quad (72)$$

Точное определение значений $P[N(t) > n]$ для распределения Вейбулла может быть выполнено только численно, и поэтому оценка в терминах распределения Пуассона с интенсивностью $(\lambda t)^\alpha$ может оказаться достаточно полезной.

Приведем несколько оценок функции восстановления $M(t) = E[N(t)]$.

Если распределение F является распределением типа «Новое лучше Существующего» («Новое хуже Существующего»), то

$$M(h) \leq (\geq) M(t+h) - M(t). \quad (73)$$

Если распределение F , которому подчинен процесс восстановления, имеет конечное среднее μ , то функция восстановления определяется как

$$M(t) \geq (t/\mu) - 1, \quad t \geq 0. \quad (74)$$

Если распределение F имеет конечное среднее μ и является распределением типа «Новое в среднем лучше Существующего» (или «Новое в среднем хуже Существующего»), то функция восстановления определяется как

$$M(t) \leq (\geq) t/\mu. \quad (75)$$

Таким образом, если распределение F является распределением типа «Новое в среднем лучше Существующего», то объединение выражений (74) и (75) дает

$$(t/\mu) - 1 \leq M(t) \leq t/\mu, \quad t \geq 0. \quad (76)$$

Из этого следует, что если распределение F является распределением типа «Новое в среднем лучше Существующего» со средним μ , то его функция восстановления может быть аппроксимирована величиной $(t/\mu) - (1/2)$ и ошибка аппроксимации не превысит $1/2$ для всех $t \geq 0$. Поскольку в общем случае довольно трудно получить точное значение величины $M(t)$, то оценки с помощью выражения (76) могут оказаться полезными.

4.3. Оценки надежности связанных систем

Рассмотрим связанную систему, имеющую функцию надежности h и состоящую из n независимых элементов с возрастающей интенсивностью $F_1, \dots, F_n$ отказов и известными средними $\mu_1, \dots, \mu_n$. Найдем нижнюю границу оценки функции надежности системы за время t , т. е. $h(\bar{F}_1(t), \dots, \bar{F}_n(t))$. Поскольку функция h является возрастающей по каждому из аргументов, то, используя нижнюю границу оценки (66), получаем

$$h(\bar{F}_1(t), \dots, \bar{F}_n(t)) \geq h(\exp(-t/\mu_1), \dots, \exp(-t/\mu_n)), \quad (77)$$

$$t < \min(\mu_1, \dots, \mu_n).$$

Так как нас в основном интересует доверительная нижняя граница оценки надежности, то, используя выражения (19) и (66) и обозначая структурную функцию системы через φ , имеем

$$\begin{aligned} h(\bar{F}_1(t), \dots, \bar{F}_n(t)) &= P[\varphi(X_1(t), \dots, X_n(t)) = 1] \geqslant \\ &\geqslant \prod_{i=1}^n P[X_i(t) = 1] \geqslant \prod_{i=1}^n \exp(-t/\mu_i) = \\ &= \exp\left(-t \sum_{i=1}^n (1/\mu_i)\right), \quad t < \min(\mu_1, \dots, \mu_n). \end{aligned} \quad (78)$$

Нижняя оценка величины $\exp\left(-t \sum_{i=1}^n (1/\mu_i)\right)$ также может быть получена с помощью процедуры оценки надежности, названной методом «счета деталей». Если система состоит из элементов K различных типов с k_i элементами каждого типа, то надежность системы за время t оценивается как $\exp\left(-t \sum_{i=1}^K (k_i/\mu_i)\right)$.

Эта оценка была бы точной, если бы все элементы были независимы, соединены последовательно и имели бы показательное распределение времени безотказной работы со средними μ_i , $i = 1, \dots, K$, для i -го типа элементов. Как было показано, метод «счета деталей» дает доверительную нижнюю границу оценки надежности системы для $t < \min(\mu_1, \dots, \mu_K)$, если система является связанной и ее элементы имеют распределение времени безотказной работы с возрастающей интенсивностью и средними μ_i , $i = 1, \dots, K$, для элементов i -го типа.

4.4. Среднее время безотказной работы систем последовательно и параллельно соединенных элементов ¹⁾

Рассматриваются распределения F_i и G_i , удовлетворяющие условию $F_i(0) = G_i(0) = 0$.

Если

$$\int_0^t \bar{F}_i(x) dx \geqslant \int_0^t \bar{G}_i(x) dx, \quad t \geqslant 0, \quad i = 1, 2, \dots, n, \quad (79)$$

то

$$\int_0^t \prod_{i=1}^n \bar{F}_i(x) dx \geqslant \int_0^t \prod_{i=1}^n \bar{G}_i(x) dx, \quad t \geqslant 0. \quad (80)$$

¹⁾ В данном разделе приводятся результаты, полученные в работах [3, 17].

Так как функционал $\int_0^{\infty} \prod_{i=1}^n F_i(x) dx$ определяет среднее время безотказной работы системы последовательно соединенных независимых элементов, в которой i -й элемент имеет распределение времени безотказной работы F_i , и то же самое справедливо для $\int_0^{\infty} \prod_{i=1}^n \bar{G}_i(x) dx$, то приведенный выше результат означает, что если равенство (79) удовлетворяется, то среднее время безотказной работы системы последовательно соединенных независимых элементов больше, когда элементы имеют распределения $F_1, \dots, F_n$, чем когда они имеют распределения $G_1, \dots, G_n$.

Для систем параллельно соединенных элементов получен следующий результат:

Если

$$\int_t^{\infty} \bar{F}_i(x) dx \leq \int_t^{\infty} \bar{G}_i(x) dx, \quad t \geq 0, \quad i = 1, 2, \dots, n, \quad (81)$$

то

$$\int_t^{\infty} \left[1 - \prod_{i=1}^n F_i(x) \right] dx \leq \int_t^{\infty} \left[1 - \prod_{i=1}^n G_i(x) \right] dx, \quad t \geq 0. \quad (82)$$

Так как функционал $\int_0^{\infty} \left[1 - \prod_{i=1}^n F_i(x) \right] dx$ определяет среднее время безотказной работы системы параллельно соединенных независимых элементов, в которой i -й элемент имеет распределение F_i , и то же самое справедливо для функционала $\int_0^{\infty} \left[1 - \prod_{i=1}^n G_i(x) \right] dx$, то приведенный выше результат означает, что если неравенство (81) удовлетворяется, то среднее время безотказной работы системы параллельно соединенных независимых элементов меньше, когда элементы имеют распределения $F_1, \dots, F_n$, чем когда они имеют распределения $G_1, \dots, G_n$.

Перечислим четыре условия, при удовлетворении любого из которых справедливы неравенства (79) и (81), а следовательно, и сформулированные выводы:

1. F_i является распределением с возрастающей средней интенсивностью отказов $\bar{G}_i(t) = \exp(-\lambda_i t)$;
2. $\bar{F}_i(t) = \exp(-\lambda_i t)$, G_i является распределением с убывающей средней интенсивностью отказов;

3. F_i является гамма-распределением с параметром формы α_i ;
 G_i является гамма-распределением с параметром формы $\beta_i \leq \alpha_i$;
 4. F_i является распределением Вейбулла с параметром формы α_i ;

G_i является распределением Вейбулла с параметром формы $\beta_i \leq \alpha_i$ (распределения F_i и G_i должны иметь одинаковые средние).

Используя вывод, сформулированный вслед за выражением (34), а также выражения (36) и (37) и условие 1 или 2, получим следующие результаты.

Если n независимых элементов имеют распределения с убывающей средней интенсивностью отказов и средними $1/\lambda_1, \dots, \dots, 1/\lambda_n$, то соответственно

$$\mu_s \geq 1 / \sum_{i=1}^n \lambda_i, \quad (83)$$

$$\begin{aligned} \mu_p \leq \sum_{i=1}^n (1/\lambda_i) = \sum_{i < j} [1/(\lambda_i + \lambda_j)] + \\ + \sum_{i < j < k} [1/(\lambda_i + \lambda_j + \lambda_k) - \dots \pm 1 / \sum_{i=1}^n \lambda_i. \end{aligned} \quad (84)$$

где μ_s (μ_p) — среднее время безотказной работы системы последовательно (параллельно) соединенных элементов. Если $\lambda_1 = \dots = \lambda_n = \lambda$, то

$$\mu_p \leq \frac{1}{\lambda} \left(1 + \frac{1}{2} + \dots + \frac{1}{n} \right). \quad (85)$$

Если n независимых элементов имеют распределения с возрастающей средней интенсивностью отказов и средними $1/\lambda_1, \dots, \dots, 1/\lambda_n$, то соответственно

$$\mu_s \leq 1 / \sum_{i=1}^n \lambda_i, \quad (86)$$

$$\begin{aligned} \mu_p \geq \sum_{i=1}^n (1/\lambda_i) - \sum_{i < j} [1/(\lambda_i + \lambda_j)] + \\ + \sum_{i < j < k} [1/(\lambda_i + \lambda_j + \lambda_k) - \dots \pm 1 / \sum_{i=1}^n \lambda_i. \end{aligned} \quad (87)$$

Если $\lambda_1 = \dots = \lambda_n = \lambda$, то

$$\lambda_p \geq \frac{1}{\lambda} \left(1 + \frac{1}{2} + \dots + \frac{1}{n} \right). \quad (88)$$

Результаты (83) — (88) фактически справедливы при более общих предположениях, чем то предположение, согласно которому элементы являются связанными и i -й элемент имеет безусловное распределение со средним $1/\lambda_i$. Независимость является более сильным предположением, чем необходимо на самом деле.

5. ИНТЕРВАЛЬНАЯ НАДЕЖНОСТЬ И ЭКСПЛУАТАЦИОННАЯ ГОТОВНОСТЬ

5.1. Интервальная надежность

Рассмотрим устройство с распределением продолжительности срока службы F и средним временем безотказной работы μ . Если устройство перестает работать (отказывает), оно немедленно заменяется устройством с аналогичным распределением продолжительности срока службы, при этом срок службы одного устройства не зависит от срока службы другого. Вероятность $p_t(u)$ того, что устройство, нормально работающее в течение времени t , будет продолжать также работать еще в течение некоторого времени u , определяется как

$$p_t(u) = \bar{F}(t+u) + \int_0^t \bar{F}(t-x+u) dM(x), \quad (89)$$

где $M(x)$ — функция восстановления, соответствующая распределению F [3, 11].

Барлоу и Прошан [3] показали, что если распределение F является распределением типа «Новое хуже Существующего», то вероятность $p_t(u)$ может быть записана как

$$p_t(u) \leq (\geq) \bar{F}(u). \quad (90)$$

Используя узловую теорему восстановления (57), можно показать, что по мере увеличения времени вероятность $p_t(u)$ практически перестает зависеть от величины t , т. е. имеем

$$p(u) = \lim_{t \rightarrow \infty} p_t(u) = \frac{1}{\mu} \int_u^{\infty} \bar{F}(x) dx. \quad (91)$$

Согласно классификации распределений, если распределение F является распределением типа «Новое в среднем лучше Существующего» (или «Новое в среднем хуже Существующего»), то имеем

$$p(u) \leq (\geq) \bar{F}(u). \quad (92)$$

Вероятность $p_t(u)$ безотказной работы устройства, используемого в момент времени t и оставшийся срок службы которого есть $\gamma(t)$, может быть записана как

$$P[\gamma(t) > u] = p_t(u), \quad (93)$$

где

$$\gamma(t) = S_{N(t)+1} - t, \quad (94)$$

$N(t)$ — число восстановлений в интервале $[0, t]$ и S_n — время, в течение которого происходит n -е восстановление [6].

5.2. Готовность системы

Рассмотрим устройство с распределением продолжительности срока службы F и средним временем безотказной работы μ .

Предположим, что, когда устройство отказывает, оно заменяется другим устройством с тем же самым распределением времени безотказной работы, причем замена осуществляется мгновенно и для этого требуется время, имеющее функцию распределения G со средним ν . *Готовность $A(t)$ системы к эксплуатации в момент времени t* определяется как вероятность того, что устройство нормально функционирует в момент времени t .

Пусть T_i — продолжительность i -го периода нормального функционирования и D_i — продолжительность i -го ремонта или периода замены. (D означает время простоя). При этом нет необходимости предполагать независимость T_i и D_i , а достаточно предположить, что последовательность $\{T_i + D_i\}$, $i = 1, 2, \dots$, есть последовательность взаимно независимых величин. Пусть H обозначает распределение $T_i + D_i$, $i = 1, 2, \dots$, и пусть M_H является функцией восстановления, соответствующей H . В этом случае [11]

$$A(t) = \bar{F}(t) + \int_0^t \bar{F}(t-u) dM_H(u). \quad (95)$$

Используя узловую теорему восстановления (57), можно показать, что с ростом t величина $A(t)$ практически перестает зависеть от t . Предельное значение величины $A(t)$, обозначаемое A и называемое *предельной готовностью*, равно

$$A = \lim_{t \rightarrow \infty} A(t) = \frac{E(x)}{E(x) + E(D)} = \frac{\mu}{\mu + \nu}. \quad (96)$$

Заметим, что величина A не зависит от распределений времени нормального функционирования или времени простоя, а зависит от средних значений этих величин.

Барлоу и Прошан [3] исследовали две модели готовности систем. В первой из них рассмотрена связанная система независимых элементов, имеющих n векторов состояний. Распределение времени безотказной работы устройства в i -м состоянии есть F_i со средним μ_i и распределение времени простоя для i -го состояния есть G_i со средним ν_i . Пусть h есть функция надежности структуры ϕ и $A(t)$, $A(A_i(t))$, A_i — готовность системы в момент времени t и предельная готовность системы (i -го элемента) соответственно. Предположим, что в течение ремонта или замены любого отказавшего элемента все другие элементы продолжают нормаль-

но функционировать. Тогда для готовности системы в момент времени t имеем

$$A(t) = h(A_1(t), \dots, A_n(t)), \quad (97)$$

$$A = h(A_1, \dots, A_n) = h\left(\frac{\mu_1}{\mu_1 + \nu_1}, \dots, \frac{\mu_n}{\mu_n + \nu_n}\right). \quad (98)$$

Во второй модели рассматривается система последовательно соединенных элементов, т. е. система, которая выходит из строя при отказе любого элемента. Кроме того, предполагается, что, когда отказавший элемент ремонтируется или заменяется, все другие элементы находятся в состоянии готовности к повторному включению. Они возобновляют функционирование, когда неисправный элемент восстановлен, и с этого момента являются в той же мере хорошими, что и к моменту отказа системы, т. е. не «все равно что новые». Следующее предположение состоит в том, что два или более элемента не могут отказать одновременно. Это справедливо тогда, когда все функции распределения отказов непрерывны.

Если $U(t)$ и $D(t)$ — соответственно время нормального функционирования и время простоя системы в течение времени t , причем $U(t) + D(t) = t$, то имеем

$$\lim_{t \rightarrow \infty} \frac{U(t)}{t} = \left(1 + \sum_{j=1}^n \nu_j / \mu_j\right)^{-1} \quad (\text{п. д.}), \quad (99)$$

$$A_{\text{ср}} = \lim_{t \rightarrow \infty} \frac{E[U(t)]}{t} = \left(1 + \sum_{j=1}^n \nu_j / \mu_j\right)^{-1}, \quad (100)$$

где $A_{\text{ср}}$ — средний коэффициент готовности.

Среднее время нормального функционирования системы стремится к

$$\mu = \left(\sum_{i=1}^n 1/\mu_i\right)^{-1} \quad (\text{п. д.}).$$

Среднее время вынужденного простоя системы стремится к величине

$$\nu = \mu \sum_{i=1}^n \nu_i / \mu_i \quad (\text{п. д.}).$$

Автор выражает глубокую признательность профессорам Барлоу и Прошапу за предоставленную возможность ознакомиться с работой [3] до ее публикации.

ЛИТЕРАТУРА

1. Barlow R. E., Bounds on Integrals with Applications to Reliability Problems, *Ann. Math. Stat.*, **36**, 565—574 (1965).
2. Barlow R. E., Proshan F., Mathematical Theory of Reliability, Wiley, N.Y., 1965. [Имеется перевод: Барлоу Р., Прошан Ф. Математическая теория надежности. — М.: Сов. радио, 1969.]
3. Barlow R. E., Statistical Theory of Reliability and Life Testing: Probability Models, Holt, Rinehart and Winston, N.Y., 1974.
4. Barlow R. E., Marshall A. W., Bounds for Distributions with Monotone Hazard Rate. I & II, *Ann. Math. Stat.*, **35**, 1234—1274 (1964).
5. Barlow R. E., Tables of Bounds for Distribution with Monotone Hazard Rate, *J. Amer. Stat. Assoc.*, **60**, 872—890 (1965).
6. Barlow R. E., Bounds on Interval Probabilities for Restricted Families of Distribution, in Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, III, 229—257 University of California Press, Berkeley, 1967.
7. Birnbaum Z. W., Esary J. D., Marshall A. W., Stochastic Characterizations of Wearout for Components and Systems, *Ann. Math. Stat.*, **37**, 816—825 (1966).
8. Bodin L., Approximations to System Reliability Using a Modular Decomposition, *Technometrics*, **12**, No. 2, 335—344 (1970).
9. Cox D. R., Renewal theory, Wiley, N.Y., 1962.
10. General Electric Company, Tables of the Individual and Cumulative Terms of the Poisson Distribution, Van Nostrand, Princeton, 1962.
11. Гнеденко Б. В., Беляев Ю. К., Соловьев А. Д. Математические методы в теории надежности. — М.: Наука, 1965.
12. Harter H. L., New Tables of the Incomplete Gamma Function Ratio and of Percentage Points of the Chi-Square and Beta Distribution, U.S. Government Printing Office, Washington, D.C., 1964.
13. Karlin S., Total positivity. Stanford University Press, Stanford, 1968.
14. Leadbetter M. R., Smith W. A. On the Renewal Function for the Weibull Distribution, *Technometrics*, **5**, 393—396 (1963).
15. Lomnicki Z. A., A Note on the Weibull Renewal Process, *Biometrika*, **53**, 375—381 (1961).
16. Mann N. R., Schafer R. E., Singpurwalla N., Methods for Statistical Analysis of Reliability and Life Data, Wiley, N.Y., 1974.
17. Marshall A. W., Proschan F., Mean Life of Series and Parallel Systems, *Appl. Prob.*, **7**, 165—174 (1970).
18. Molina C. E., Poisson's Exponential Binomial Limit, Van Nostrand, N.Y., 1942.
19. Pearson K., Tables of the Incomplete Gamma Function, Cambridge University Press, Cambridge, 1934.
20. Smith W. L., Renewal Theory and its Ramifications. *J. Roy. Stat. Soc.*, Series B, **20**, 243—302 (1958).

I-10. Техническое обслуживание оборудования

Э. Жардин¹⁾

ВВЕДЕНИЕ

Задача технического обслуживания — это, по существу, ряд проблем, связанных с контролем за состоянием оборудования. Для решения подобных задач могут быть использованы как детерминированный, так и стохастический подходы. Выбор подхода зависит от наличия или отсутствия в формулировке задачи неопределенностей относительно сроков и последовательности принятия решений. Следует иметь в виду, что наилучшее решение может быть получено только при условии четкой постановки задачи, т. е. необходимо точно определить конечную цель, которая должна быть достигнута. На-

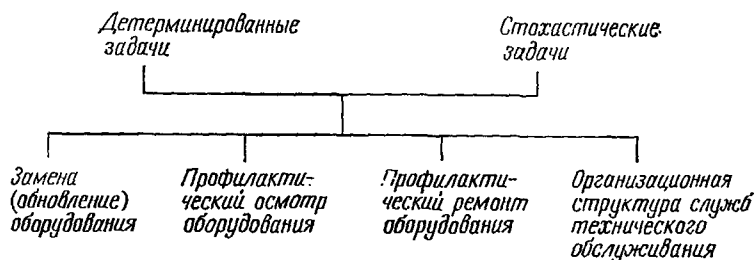


Рис. 1. Схема классификации задач технического обслуживания.

пример, целесообразно разрабатывать стратегию ремонта, в результате которой максимизируется коэффициент готовности оборудования, если данное предприятие интересуется только максимизацией прибыли. При некоторых обстоятельствах оптимальная стратегия ремонта позволяет достичь обеих целей, однако это не всегда удается. Обзор моделей систем управления техническим обслуживанием дан в работе [20].

Рассмотрим более подробно следующие основные типы задач контроля за состоянием оборудования: замена, профилактический осмотр, профилактический ремонт и восстановление, а также организация служб технического контроля (рис. 1).

1. ЗАМЕНА (ОБНОВЛЕНИЕ) ОБОРУДОВАНИЯ

Замена вышедшего из строя устройства, будь то отдельный элемент системы или вся система, строго говоря, означает обновление, а не замену оборудования. Детерминированный подход

¹⁾ Andrew K. S. Jardine, Ph. D., Associate Professor, Faculty of Business Administration at the University of Windsor in Windsor, Ontario, Canada.

к решению задач замены оборудования используется в тех случаях, когда возникает необходимость замены элементов, эксплуатационные характеристики которых изменяются (ухудшаются) во времени определенным образом. Например, ухудшение эксплуатационных характеристик некоторых элементов котельной установки, в частности воздухоподогревателя, приводит к увеличению затрат на эксплуатацию этой установки. Поскольку подобное увеличение может быть описано экспоненциальной функцией вида $A - Be^{-kt}$, то в данном случае нет какой-либо неопределенности относительно величины производственных затрат в момент времени t после осуществления замены.

Если замены производятся довольно часто, а период планирования, для которого определяется оптимальное решение, велик, то критерием оптимизации являются общие затраты (или прибыль, или другой соответствующий критерий) в единицу времени. В противном случае возникает необходимость нахождения оптимального решения на конечном интервале времени, и для анализа таких задач могут быть использованы методы динамического программирования.

1.1. Замена оборудования, изменение эксплуатационных характеристик которого носит детерминированный характер

Задача замены оборудования, осуществляемой через небольшие интервалы равной длительности, с бесконечным периодом планирования. Обычно затраты на эксплуатацию оборудования возрастают с течением времени. Чтобы уменьшить эти затраты, осуществляется замена оборудования. Поскольку замена оборудования также связана с определенными затратами, то задача сводится к нахождению оптимального интервала между заменами, для которого общие затраты в единицу времени оказываются минимальными. Обозначив общие затраты в единицу времени через $C(t_r)$, имеем

$$C(t_r) = \frac{\int_0^{t_r} c(t) dt + C_r}{t_r},$$

где t_r — срок эксплуатации данного оборудования, по истечении которого производится замена, $c(t)$ — эксплуатационные затраты в единицу времени в момент t (предполагается, что $c(t)$ — возрастающая функция времени) до замены и C_r — действительная (чистая) стоимость замены. Оптимальная величина t_r соответствует условию $C'(t_r) = 0$. Как показано в работе [7], оптимальное значение t_r достигается при $C(t_r) = c(t_r)$, т. е. когда общие затраты

на данный момент времени равны текущим эксплуатационным затратам в данный момент времени.

Задача замены, осуществляемой через значительные интервалы равной длительности, с бесконечным периодом планирования. Поскольку подобные замены производятся через интервалы, длительность которых измеряется годами, а не неделями или месяцами, как в предыдущем случае, то возникает необходимость рассматривать стоимость денег с учетом дохода следующего цикла. В результате задача сводится к разработке стратегии замены, позволяющей *максимизировать* суммарную приведенную чистую прибыль [14] или *минимизировать* общие приведенные затраты [3].

Рассмотрим задачу минимизации общих приведенных затрат (рис. 2). Введем следующие обозначения: A — первоначальная действительная стоимость нового оборудования; C_i — эксплуатационные затраты в i -й период времени, считая с момента установки нового оборудования (предполагаются монотонно возрастающими), которые приходятся на начало каждого периода; r — действующий коэффициент дисконтирования, равный $[1/(1+d)]$, где d — учетная ставка за рассматриваемый период; n — эксплуатационная долговечность (в периодах) оборудования, по истечении которой оборудование заменяется; $C(n)$ — общие приведенные затраты (при замене на идентичное оборудование) за бесконечный период времени при условии, что замена производится через каждые n периодов.

Поскольку конечной целью задачи является минимизация величины $C(n)$, следовательно, необходимо определить значение n , при котором эта величина достигает минимума, т. е. минимизирующее величину $C(n) = C_1(n) + C_2(n)r^n + C_3(n)r^{2n} + \dots$, где $C_1(n)$ — общие приведенные затраты за первый рабочий цикл, $C_2(n)$, $C_3(n)$ — соответственно за второй, третий и т. д. циклы (все затраты приводятся к началу соответствующего цикла).

Так как $C_1(n) = C_2(n) = \dots = C(n)$, то мы имеем дело с суммой бесконечной геометрической прогрессии, и поскольку $r < 1$, то получаем

$$C(n) = \frac{c(n)}{1-r^n},$$

где

$$c(n) = A + \sum_{i=1}^n c_i r^{i-1}$$

Оптимальное значение n находят с помощью методов дифференциального исчисления. Таким образом, полученное оптимальное решение поставленной задачи предусматривает замену оборудования на идентичное только в том случае, если эксплуатационные

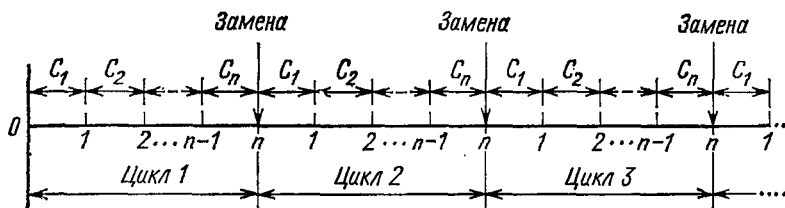


Рис. 2. Долгосрочная детерминированная задача замены.

затраты за следующий период оказываются больше, чем предшествующие приведенные средние затраты, в противном случае необходимо продолжать эксплуатацию имеющегося оборудования.

Предшествующие приведенные средние затраты с учетом периода n могут быть выражены следующим соотношением:

$$[A + \sum_{i=1}^n c_i r^{i-1}] / [1 + \sum_{i=1}^{n-1} r^i].$$

Следовательно, если величина c_{n+1} больше указанного выше приведенного среднего, то замену целесообразно производить через n периодов.

Решение задач замены, осуществляемой через длительные интервалы, с учетом льгот на налог и субсидий в случае приобретения основного оборудования рассматривается в работе [9]. Однако следует иметь в виду, что изменения оборотного капитала, обусловленные льготами на налог и субсидиями, учитываются только тогда, когда эти изменения имеют место в течение цикла эксплуатации оборудования.

Авторы работы [3] рассмотрели задачу замены существующего оборудования на технологически более современное и показали, что оптимальным решением является замена при условии, что затраты на эксплуатацию существующего оборудования за следующий период времени превысят приведенные средние затраты на установку и эксплуатацию нового оборудования, рассчитанные для оптимального срока его службы.

Поскольку речь идет о технологических усовершенствованиях, то основной проблемой является определение изменений в затратах, связанных с подобными усовершенствованиями. Подход к решению такой задачи с использованием метода динамического программирования при экспоненциальном изменении прибыли, эксплуатационных затрат и стоимости замены оборудования представлен в работе [2].

1.2. Замена оборудования, изменение эксплуатационных характеристик которого носит вероятностный характер

Аварийная замена оборудования обычно связана с большими единовременными затратами, чем профилактическая замена. Поэтому целесообразно принимать решения относительно необходимости профилактической замены и сроков ее осуществления. Интервалы между заменами (аварийными или профилактическими), как правило, достаточно короткие, и, следовательно, нет необходимости учитывать падение доходов. Критерием оптимизации служат ожидаемые общие затраты в единицу времени (при условии, что затраты являются подходящей мерой).

При анализе подобных задач замены оборудования мы будем использовать следующие обозначения [13]:

$f(t)$ — плотность распределения вероятностей времени безотказной работы оборудования;

$F(t) = \int_0^t f(x) dx$ — функция распределения или вероятность

отказа до наступления момента t ;

$R(t) = 1 - F(t)$ — функция надежности (функция долговечности), или вероятность того, что отказ случится не раньше, чем наступит момент t ;

$r(t) = f(t)/[1 - F(t)]$ — мгновенная интенсивность отказов как функция времени. Произведение $r(t) dt$ дает условную вероятность того, что отказ произойдет в интервале времени $(t, t + dt)$ при условии, что оборудование было исправно до момента времени t (см. гл. I-9).

Замены, осуществляемые через интервалы равной длительности. Как видно из рис. 3, профилактические замены производятся через постоянные интервалы времени длительностью t_p , а аварийные замены осуществляются по мере возникновения отказов. Если C_p — есть общие затраты при профилактической замене, C_f — общие затраты при замене в результате отказа и $C(t_p)$ — ожидаемые общие затраты в единицу времени при проведении профилактических замен с интервалом t_p , то получаем

$$C(t_p) = \frac{\text{Ожидаемые общие затраты/Экономический цикл}}{\text{Длительность цикла}} = \frac{H(t_p) C_f + C_p}{t_p},$$

где $H(t_p)$ — ожидаемое число отказов в течение интервала времени $(0, t_p)$, или функция восстановления.

Оптимальное значение величины t_p соответствует выполнению условия $C'(t_p) = 0$. Это дает

$$t_p h(t_p) = \frac{C_p}{C_f}.$$



Рис. 3. Профилактическая замена через интервалы равной длительности.

где $h(t_p)$, производная функции $H(t_p)$, есть плотность восстановления [1]. (Процедуры получения функций $H(t)$ и $h(t)$ описаны в работе [5]. См. также гл. 1—9.)

Для определения функции $H(t)$ с помощью методов теории восстановления может потребоваться использовались преобразования Лапласа. Если применение обратного преобразования вызывает определенные трудности, то может быть использована процедура дискретной аппроксимации. Один из возможных подходов сводится к решению рекуррентного уравнения вида [14]

$$H(t) = \sum_{i=0}^{T-1} [1 + H(T-i-1)] \int_i^{i+1} f(t) dt, \quad T \geq 1,$$

при $H(0) = 0$, где $H(T)$ — ожидаемое число отказов за T периодов. Как показано в работе [3], оптимальное решение, получаемое в этом случае, состоит в том, чтобы производить профилактическую замену в конце периода T , если затраты на замену неисправных элементов в течение этого периода больше, чем ожидаемые общие затраты к концу периода T .

Замена всех идентичных элементов или блоков. Иногда в целях уменьшения числа операций замен целесообразно осуществить одновременно замену всех идентичных элементов, а не одного элемента. Подобным образом выполняется замена ламп уличного освещения, т. е. все лампы заменяются одновременно через интервалы равной длительности t_p , при этом, однако, не исключается возможность выполнения аварийной замены. Такая замена почти аналогична замене через интервалы равной длительности с той лишь разницей, что в данном случае одновременно заменяются N элементов, т. е.

$$C(t_p) = \frac{NC_g + NH(t_p)C_f}{t_p},$$

где C_g — стоимость замены одного элемента при одновременной замене всех идентичных элементов.

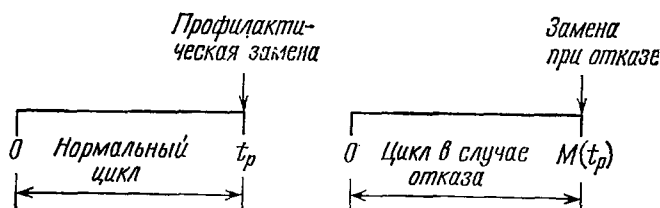


Рис. 4. Возможные рабочие циклы при замене по наработке.

Замена по наработке. Очевидный недостаток замены через интервалы равной длительности обусловлен существованием определенной вероятности того, что сразу же после профилактической замены может возникнуть необходимость в аварийной замене. Чтобы исключить такую возможность, выполняется замена по наработке, т. е. замена элемента осуществляется только тогда, когда длительность его эксплуатации достигает определенного предела t_p . В этом случае существуют два возможных варианта рабочего цикла (рис. 4). Если профилактическая замена производится после времени наработки t_p , то среднее время наработки $M(t_p)$ до выхода элемента из строя может быть представлено выражением вида

$$\frac{\int_0^{t_p} t f(t) dt}{1 - R(t_p)}.$$

Тогда ожидаемые общие затраты в единицу времени при проведении профилактической замены после того, как элемент проработал в течение времени t_p , будут равны

$$C(t_p) = \frac{C_p R(t_p) + C_f (1 - R(t_p))}{t_p R(t_p) + M(t_p) (1 - R(t_p))},$$

где C_p , C_f — общие затраты на профилактическую и аварийную замены соответственно.

Примечание. Справедливость использования отношения математических ожиданий доказана в работе [21], автор которой показал, что

$$\lim_{t \rightarrow \infty} \frac{K(t)}{t} = \frac{\text{Ожидаемые затраты на цикл}}{\text{Ожидаемая длительность цикла}},$$

где $K(t)$ — сумма ожидаемых затрат для последовательности циклов на интервале времени $(0, t)$ и $K(t)/t$ — ожидаемые затраты

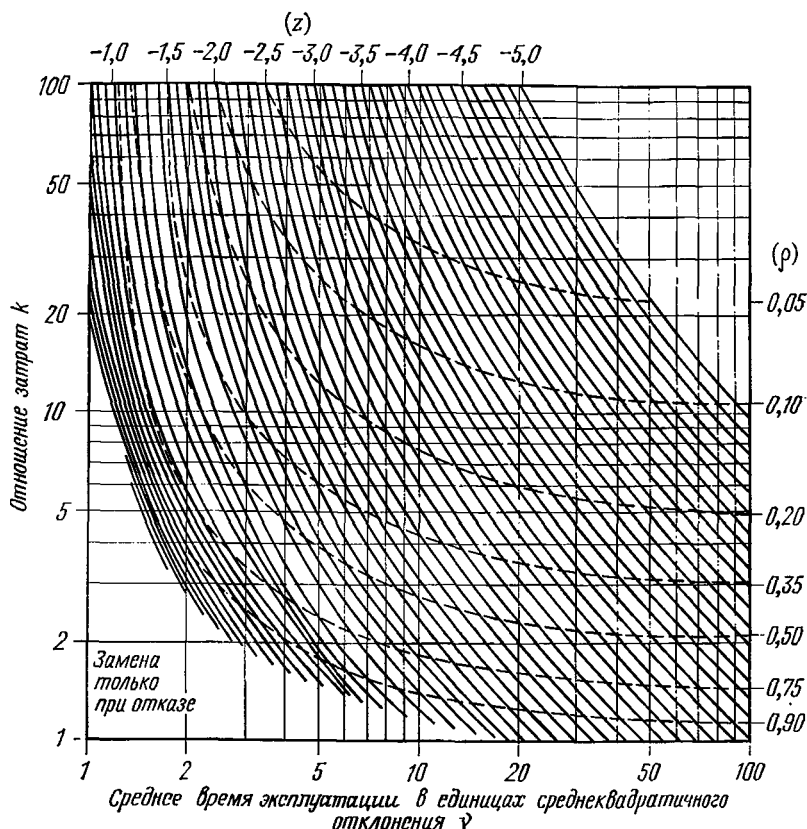


Рис. 5. Оптимальные стратегии при одновременной замене идентичных элементов или блоков (распределение Вейбулла).

в единицу времени. Оптимальное значение t_p определяется из условий $C'(t_p) = 0$ и

$$r(t_p) = \int_0^{t_p} R(t) dt - F(t_p) = \frac{C_p}{C_f - C_p}.$$

Необходимое и достаточное условие того, что замена по наработке более выгодна, чем аварийная замена, может быть выражено как [27]

$$\frac{C_p}{C_f} < 1 - \frac{M_R(a)}{M_R(0)} \text{ для некоторого } a > 0,$$

где C_p и C_f — общие затраты, связанные с профилактической и аварийной заменами соответственно. Величина $M_R(t)$ определяется как средняя остаточная долговечность элемента и представляет собой математическое ожидание времени отказа элемента, который уже использовался в течение времени t , т. е.

$$M_R(t) = \frac{\int_0^{\infty} x f(t+x) dx}{R(t)}.$$

Это условие означает, что, во-первых, общие затраты при аварийной замене должны быть больше, чем общие затраты, связанные с профилактической заменой, и, во-вторых, мгновенное значение интенсивности отказов $r(t)$ должно увеличиваться.

Если ограничиться рассмотрением элементов, для которых мгновенное значение отказов увеличивается, то необходимое и достаточное условие принимает вид

$$\frac{C_p}{C_f} < 1 - \frac{M_R(\infty)}{M_R(0)}.$$

В работе [10] показано, что если распределение частоты отказов с достаточной степенью точности является распределением Вейбулла и известны отношение затрат $k = (C_f/C_p)$ и среднее время эксплуатации элемента ν , измеренное по отношению к величине среднеквадратичного отклонения, то оптимальный интервал между профилактическими заменами может быть получен с помощью зависимостей, представленных на рис. 5 и 6. При заданных значениях k и ν определяется оптимальное значение t_p (на графике оно представлено в виде отклонения от среднего значения, выраженного через среднеквадратичное отклонение и обозначаемого через z). После того как найдено значение z , величина t_p определяется по формуле $t_p = E(t) + z\sigma(t)$, где $E(t)$ — средняя наработка элемента до его замены, а $\sigma(t)$ — среднеквадратичное отклонение времени безотказной работы. На рис. 5 и 6 приведены также кривые нормированного отношения ρ затрат [10]. Величина ρ , соответствующая заданным значениям k и ν , характеризует отношение затрат при оптимальной стратегии к затратам при аварийных заменах.

В работе [15] также приведены зависимости, подобные представленным на рис. 6, с помощью которых можно определить оптимальный период для профилактической замены при условии, что распределение отказов может быть описано распределением Вейбулла, нормальным, логарифмическим или гамма-распределением.

Следует отметить, что распределение Вейбулла нашло широкое применение в исследованиях задач технического обслуживания

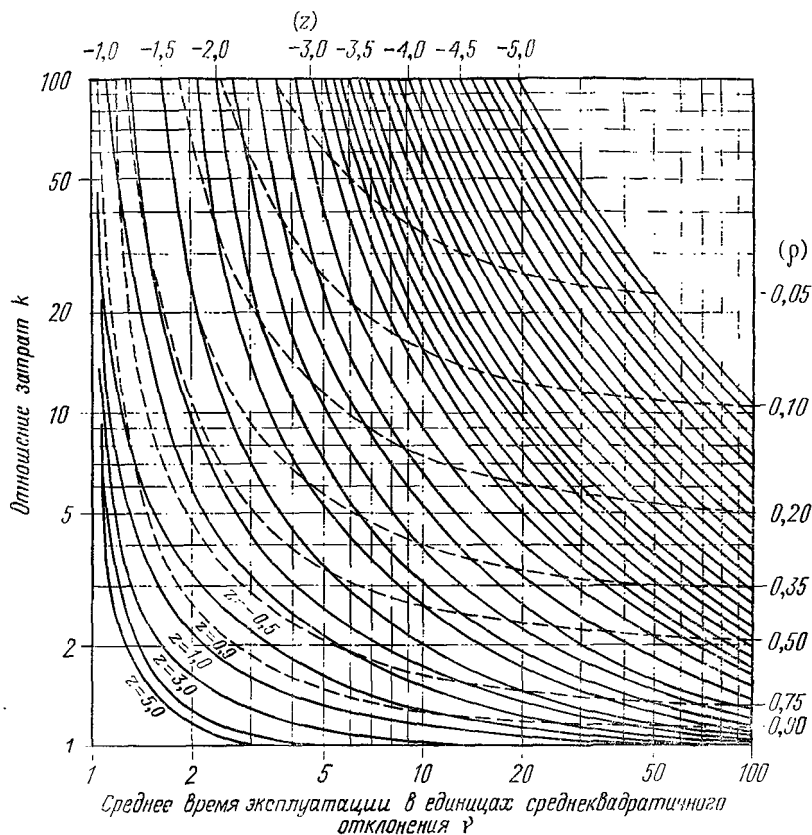


Рис. 6. Оптимальные стратегии при замене по наработке (распределение Вейбулла).

оборудования, поскольку, как это следует из работ Вейбулла по теории вероятности, оно хорошо соответствует экспериментальным данным [16]. Распределение Вейбулла может быть представлено в виде семейства кривых, форма которых определяется параметром β (рис. 7). Плотность вероятности для распределения Вейбулла имеет вид

$$f(t) = \frac{\beta}{\eta} \left(\frac{t-t_0}{\eta} \right)^{\beta-1} \exp - \left(\frac{t-t_0}{\eta} \right)^{\beta}, \text{ для } t \geq t_0, \eta > 0 \text{ и } \beta > 0,$$

где η — масштабный параметр (иногда пазываемый характеристической долговечностью), β — параметр, характеризующий форму

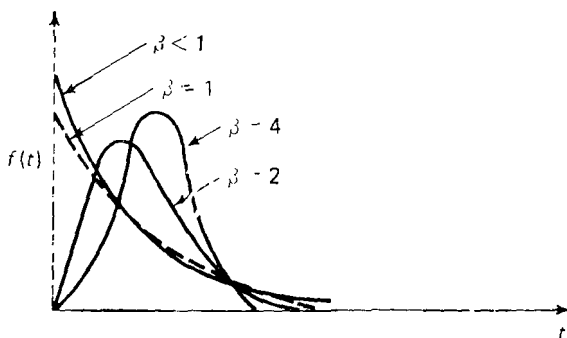


Рис. 7. Форма кривых распределения Вейбулла в зависимости от значений параметра β .

При значении $\beta=3,4$ распределение Вейбулла близко к нормальному, при $\beta=1$ оно совпадает с показательным распределением.

кривой, и t_0 — параметр сдвига (часто предполагается, что он равен нулю).

Замена с учетом времени простоя оборудования. В некоторых случаях оказывается возможным вышедшее из строя оборудование оставить в нерабочем состоянии (т. е. отключенным) до наступления момента очередной плановой замены. Поскольку, однако, вследствие простоя оборудования возникают потери, то необходимо определить оптимальное время t_i , в течение которого простоя оборудования не приводит к существенным издержкам. Значение оптимального времени получается путем решения относительно t_i следующего уравнения [25]:

$$\frac{C_f}{C_i} = \int_0^{t_i} R(t) dt,$$

где C_f — общие затраты на аварийную замену, а C_i — общие затраты в единицу времени, связанные с простоем отказавшего оборудования. Оптимальное решение состоит в том, чтобы заменять оборудование при наступлении плановых сроков замены, а отказавшее оборудование заменять, если только до следующей плановой замены осталось не меньше времени, чем t_i , в противном случае отложить замену до наступления очередного планового срока. Это соответствует выполнению условия $C_f < C_i M_R(0)$.

Замена при «удобном случае». В тех ситуациях, когда оборудование простаивает из-за отсутствия сырья или в результате аварии на каком-то участке производственного процесса, существует реальная возможность для проведения профилактической замены. В работе [26] рассмотрена эта задача и разработана модель, которую можно использовать для определения границ управления, позволяющих принимать решение относительно того, воспользо-

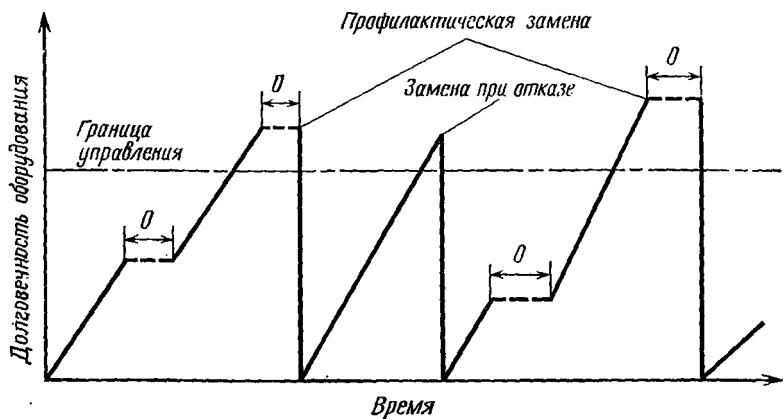


Рис. 8. Замена при «удобном случае».

зоваться ли удобным случаем для осуществления профилактической замены или нет (рис. 8). Приблизительный метод определения границы управления приведен в работе [27]. Обобщение этих результатов на случай возможности профилактических замен через интервалы одинаковой длительности выполнено в работе [8].

2. КОНТРОЛЬ ЗА СОСТОЯНИЕМ ОБОРУДОВАНИЯ

Основная цель контроля за состоянием оборудования состоит в своевременном выявлении и устранении дефектов. При этом важное значение имеют тщательность и частота контроля (или контрольного осмотра), а также выбор показателей (таких, как износ подшипников, градуировка, качество продукции), характеризующих состояние оборудования.

2.1. Контроль за состоянием оборудования, непрерывно находящегося в эксплуатации

Чтобы уменьшить количество поломок оборудования, непрерывно находящегося в эксплуатации, необходимо своевременно устранять небольшие неисправности, которые впоследствии могли бы привести к серьезным неполадкам. Для обнаружения подобных неисправностей проводят контрольные осмотры оборудования. Поскольку такие осмотры требуют определенных затрат (затраты на средства технического обслуживания и покрытие потерь в производстве вследствие запланированных простоев оборудования), то оптимальное решение должно обеспечивать сохранение определенного соотношения между общими затратами, связанными с проведением осмотра, и получаемыми выгодами.

В качестве конкретных целей могут выступать максимизация прибыли, минимизация затрат или минимизация простоев. В работе [24] рассмотрен случай максимизации прибыли, т. е. максимизации величины

$$P(n) = V \left(1 - \frac{\lambda(n)}{\mu} - \frac{n}{i} \right) - \frac{R\lambda(n)}{\mu} - \frac{In}{i},$$

где $P(n)$ — прибыль в единицу времени, получаемая при эксплуатации оборудования, если число осмотров (и малых ремонтов) в единицу времени равно n ; V — прибыль в единицу времени, получаемая при эксплуатации оборудования, если не возникает простоев; $\lambda(n)$ — среднее число поломок в единицу времени при частоте осмотров, равной n ; $1/\mu$ — средняя длительность ремонта (распределение предполагается экспоненциальным); $1/i$ — среднее время, необходимое для производства контрольного осмотра (распределение длительности контрольных осмотров предполагается экспоненциальным и не зависящим от частоты осмотров); R — стоимость ремонта в единицу времени и I — стоимость контрольного осмотра в единицу времени. Оптимальная частота контрольных осмотров в этом случае определяется из уравнения

$$\frac{d\lambda(n)}{dn} = -\frac{\mu}{i} \left(\frac{V+I}{V+R} \right).$$

2.2. Контроль за состоянием оборудования, используемого только в экстренных случаях

Поскольку некоторые виды оборудования и многие виды вооружения предназначены для использования исключительно в критических ситуациях, то, естественно, возникают проблемы, связанные с их длительным хранением. Чтобы исключить или по крайней мере уменьшить вероятность того, что после длительного хранения оборудование окажется не пригодным к эксплуатации, осуществляется практика контрольных осмотров. Если обнаружится, что оборудование неисправно, то производится его ремонт или замена, т. е. те виды технического обслуживания, в результате которого оборудование как бы «возвращается» к своему исходному состоянию. Поскольку контрольный осмотр, ремонт или замена требуют времени, задача сводится к максимизации времени, в течение которого оборудование пригодно к использованию, путем определения наилучшего интервала между осмотрами.

Исходя из предположения, согласно которому к концу контрольного осмотра или ремонта (если окажется, что он необходим) состояние оборудования эквивалентно исходному состоянию,

было показано [14], что коэффициент готовности оборудования в единицу времени $A(t_i)$ определяется формулой

$$A(t_i) = \frac{t_i R(t_i) + \int_0^{t_i} t f(t) dt}{t_i + T_i + T_r(1 - R(t_i))},$$

где t_i — интервал между контрольными осмотрами, а T_i и T_r — соответственно средние продолжительности контрольного осмотра и ремонта. Обзор работ в области исследования задач контроля за состоянием оборудования этого класса (называемых моделями подготовки) дан в работах [15, 17].

2.3. Контроль за состоянием оборудования, эксплуатационные характеристики которого могут меняться в процессе эксплуатации

Третья категория задач контроля состояния относится к оборудованию, у которого состояние отказа (т. е. неприемлемое состояние) связано с ухудшением характеристик.

В данном случае речь идет о контроле за оборудованием, состояние отказа которого проявляется в снижении показателей качества производства. Такой отказ может привести к выпуску продукции с выходящими за пределы допусков параметрами. Поэтому определить состояние оборудования можно не только с помощью контрольного осмотра, но и контроля качества выпускаемой продукции. Если обнаружится отказ, оборудование приводится в «исходное» состояние с помощью ремонта. Задача состоит в том, чтобы определить оптимальный график контроля, минимизирующий общие затраты в единицу времени, связанные с проведением контроля, ремонта и работой неисправного оборудования. Подобная задача рассмотрена авторами работы [1]. Более широкая постановка задачи, учитывающая возможность осуществления в процессе контроля (в зависимости от состояния оборудования) одного из видов технического обслуживания приведена в работе [22]. В работе [19] предложена процедура, позволяющая быстро решить задачу, сформулированную авторами работы [1].

3. ПРОФИЛАКТИЧЕСКИЙ ТЕКУЩИЙ РЕМОНТ И ВОССТАНОВЛЕНИЕ ОБОРУДОВАНИЯ

Профилактический текущий ремонт — это действия по обслуживанию оборудования, которые предпринимаются прежде, чем оборудование окончательно выйдет из строя. В последнем случае (т. е. в случае отказа) производится ремонт или восстановление оборудования. Ни один из этих видов обслуживания, бе-

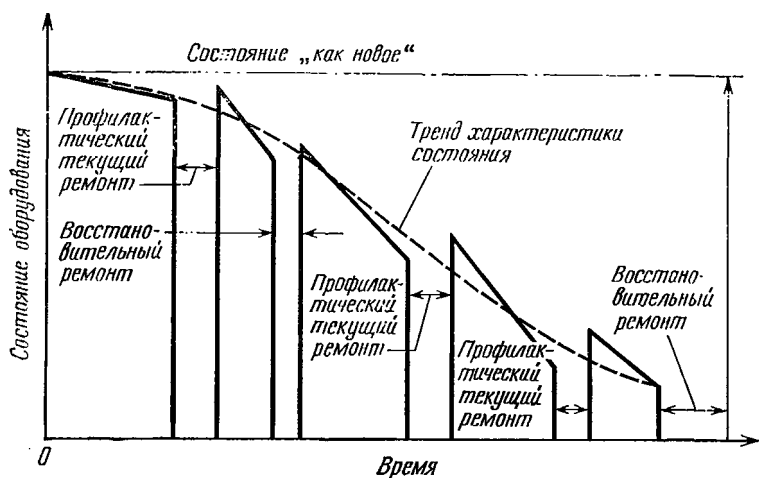


Рис. 9. Последствия проведения текущего и восстановительного ремонтов.

условно, не приводит оборудование к исходному состоянию. (Заметим, что состояние отказа не обязательно означает, что оборудование «сломалось», т. е. полностью вышло из строя. Такое состояние может быть связано, например, с производством деталей, не укладывающихся в пределы допуска, или окончанием срока службы оборудования, т. е. имеется определенный риск при продолжении его эксплуатации, например при производстве некоторых узлов самолетов.)

Основные проблемы, которые возникают при осуществлении профилактического текущего ремонта и восстановления оборудования, — это выбор межремонтного периода и определения степени влияния этих видов технического обслуживания на последующее состояние оборудования. При этом следует иметь в виду, что межремонтный период может быть бесконечным, т. е. один из видов обслуживания (текущий ремонт) не производится, а выполняется только ремонт в случае отказа. Кроме того, в предельном случае и текущий ремонт, и восстановление могут быть эквивалентны замене.

На рис. 9 показаны «обычные» последствия текущего ремонта и восстановления. Как видно из рисунка, и тот и другой вид обслуживания улучшают состояние оборудования, однако со временем оно изнашивается и полностью заменяется. Использование методов динамического программирования для решения задач текущего ремонта и восстановления рассмотрено в работе [14].

Задачи подобного типа, но с учетом установленных лимитов на ремонт оборудования были подробно проанализированы в ра-

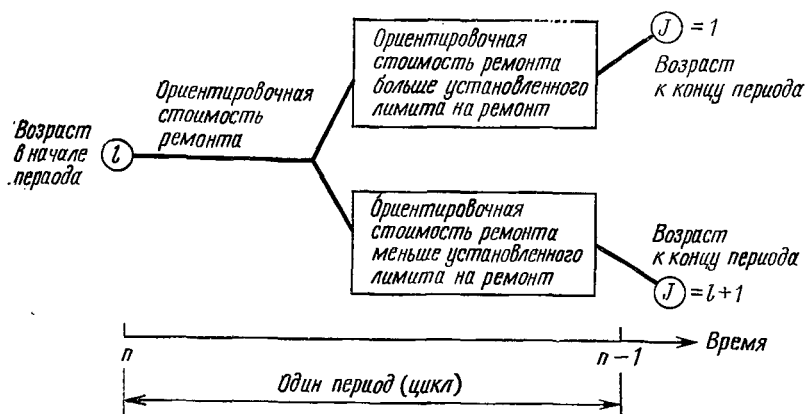


Рис. 10. Стратегия лимитированных затрат на ремонт.

боте [11] (рис. 10). Лимит на ремонт — это ограниченная сумма денег, которая одновременно может быть потрачена на ремонт оборудования. При отказе оборудования оценивается стоимость ремонтных работ по однократному восстановлению: если ориентировочные затраты на ремонт превышают установленный лимит, оборудование заменяется, в противном случае оно ремонтируется. Таким образом, в данном случае учитывается стоимость ремонта (если он производится) и предполагаемый срок службы оборудования (принимая во внимание ремонт) и проводится сравнение со стоимостью и характеристикой отказов оборудования, если оно заменяется. Это позволяет определить лимиты на ремонт оборудования разного срока службы, и решение относительно необходимости ремонта может быть принято на основе установленного лимита, т. е. максимального количества денег, которое целесообразно потратить на ремонт оборудования с данным сроком службы. О применении подобных лимитов при принятии решений о замене военных транспортных средств сообщается в работе [12].

4. ОРГАНИЗАЦИОННАЯ СТРУКТУРА СЛУЖБЫ ТЕХНИЧЕСКОГО ОБСЛУЖИВАНИЯ

Определение структуры организации, занимающейся техническим обслуживанием оборудования, предполагает решение двух основных взаимосвязанных проблем — это определение средств технического обслуживания (рабочей силы и оборудования), которыми должна располагать данная организация, и выбор спо-

соба реализации этих средств с учетом возможности использования внешних ресурсов (например, привлечения подрядчиков).

Обычно в распоряжении организации имеются некоторые собственные средства технического обслуживания, такие, как мастерские, склады и трудовые ресурсы. Кроме того, организация устанавливает определенные контакты с подрядными организациями, которые могут выполнять некоторые или все виды работ по техническому обслуживанию, необходимых для данной организации. Работы по обслуживанию могут выполняться либо сотрудниками организации, либо подрядчиками на территории организации или в мастерской подрядчика. Выбор одного из этих четырех возможных вариантов в каждом частном случае зависит от характера требуемой работы по обслуживанию, средств технического обслуживания, которыми располагает организация, наличия внешних ресурсов, объема работ по обслуживанию и их изменения во времени, а также от затрат, связанных с осуществлением каждого варианта.

Все эти факторы необходимо учитывать при определении наилучшего набора собственных средств технического обслуживания организации. Увеличение парка ремонтного оборудования, такого, как токарные станки, приводит к увеличению капиталовложений в предприятие и строения и увеличению потребности в рабочей силе. Если загрузка ремонтных подразделений меняется, может оказаться трудным сохранить состояние равновесия, при котором исключается как перегрузка средств технического обслуживания, так и простой производственных мощностей. Однако увеличение собственных средств уменьшает зависимость организации от внешних ресурсов, и, следовательно, необходимо найти такое решение, которое обеспечивало бы сохранение определенного соотношения между затратами, связанными с использованием собственных средств, и затратами, связанными с использованием внешних ресурсов. В этом случае возникает сложная задача оценки затрат, так как приходится учитывать не только затраты на внешние ресурсы, но и некоторые другие затраты, например связанные с контролем за выполнением работ или с простоем оборудования. Кроме того, возникает задача определения численности ремонтной бригады. При решении подобных задач следует иметь в виду, что увеличение численности бригады приводит к увеличению затрат на ее содержание, к сокращению времени простоя оборудования (имеется в виду время до начала ремонта) и уменьшению длительности самого ремонта (при условии, что работа выполняется всей бригадой).

Анализ задач организационной структуры служб технического обслуживания часто проводится с помощью методов теории массового обслуживания или математического моделирования [4, 18]. Авторы работы [6] детально рассмотрели различные задачи орга-

низации служб технического обслуживания, которые могут возникнуть, если ремонтная бригада отвечает за ремонт сразу нескольких машин. Одной из таких задач является определение числа обслуживаемых машин, закрепляемых за бригадой. При этом необходимо иметь в виду, что в течение ремонта одной из машин другие простаивают.

К числу работ, посвященных исследованию задач определения оптимального уровня средств технического обслуживания с учетом возможности использования подрядчиков, можно отнести работу [14]. Автор предложил модель, на основании которой можно получить оптимальное решение, минимизирующее общие затраты (постоянные и переменные) в единицу времени в предположении, что все возникающие в течение единичного отрезка времени работы будут завершены в пределах рассматриваемого интервала времени. Возможность переноса срока работ рассматривается в работе [23].

5. ПЕРСПЕКТИВЫ РАЗВИТИЯ ТЕХНИЧЕСКОГО ОБСЛУЖИВАНИЯ ОБОРУДОВАНИЯ

До сих пор существует мнение, согласно которому слишком длительное техническое обслуживание оборудования и связанные с ним замены следует рассматривать как «неизбежное зло», так как исключить подобного рода деятельность не представляется возможным, а вместе с тем с ней связаны определенные издержки. Однако в настоящее время есть все основания полагать, что управление техническим обслуживанием на научной основе окажется эффективным, поскольку такая деятельность может быть прибыльной для организации.

Правительство Великобритании придает огромное значение проблемам технического обслуживания оборудования и в 1970 г. организовало Комитет по теротехнологии (от греческого корня *tereo* — заботиться, ухаживать). Предполагалось, что комитет должен осуществлять «управление финансовой, технической и другими видами деятельности применительно к реальным активам с целью экономии затрат за срок их службы, составлять инструкции и проектировать эксплуатационную надежность и системы технического обслуживания заводов, машин, оборудования, зданий и сооружений при их монтаже, вводе в эксплуатацию, в процессе эксплуатации, модернизации и при замене с учетом поступающей от потребителей информации о качестве конструкций и затратах». Нам представляется, что такая постановка вопроса с необходимостью предполагает использование методов исследований операций для решения задач технического обслуживания оборудования.

ЛИТЕРАТУРА

1. Barlow R. E., Prochan F., *Mathematical Theory of Reliability*. Wiley, N.Y., 1965. [Имеется перевод: Барлоу Р. Е., Прошан Ф. Математическая теория надежности.— М.: Советское радио, 1969.]
2. Bellman R. E., Dreyfus S. E., *Applied Dynamic Programming*. Princeton Univ. Press, Princeton, N.J./Oxford Univ. Press, London, 1962. [Имеется перевод: Беллман Р., Дрейфус С. Прикладные задачи динамического программирования.— М.: Наука, 1965.]
3. Churchman C. W., Ackoff R. L., Arnoff E. L., *Introduction to Operations Research*. Wiley, N.Y., 1957. [Имеется перевод: Черчмен Р., Акрофф Р., Арнофф Э. Введение в исследование операций.— М.: Наука, 1968.]
4. Carruthers A. J., MacGow I., Hackmeier G. C., *A Study of the Optimum Size of Plant Maintenance Gangs, in Operational Research in Maintenance*, Jardine A. K. S. (Ed.), Manchester Univ. Press, Manchester/Barnes and Noble, N.Y., 1970.
5. Cox D. R., *Renewal theory*, Methuen, London/Wiley, N.Y., 1962. [Имеется перевод: в кн. Кокс Д., Смит В., Теория восстановления.— М.: Сов. радио, 1967.]
6. Cox D. R., Smith W. L., *Queues*, Methuen, London/Wiley, N.Y., 1961. [Имеется перевод: Кокс Д. Р., Смит У. Л., Теория очередей.— М.: Мир, 1966.]
7. Dean B. V., *Replacement Theory in Progress in Operation Research*, 1, R. L. Ackoff (Ed.), Wiley, N.Y., 1961.
8. Duncan J., Scholnick L. S., *Interrupt and Opportunistic Replacement Strategies for Systems of Deterioration Components*, *Operational Res. Quart.*, 24, No. 2, 271—283 (1973).
9. Eilon S., King J. R., Hutchinson D. E. *A Study in equipment Replacement*, *Operational Res. Quart.*, 17, No. 1, 59—71 (1966).
10. Glasser G. J., *Planned Replacement; Some Theory and Its Application*, *J. Quality Tech.*, 1, 110—119 (1969).
11. Hastings N. A. J., *The Repair Limit Replacement Method*, *Operational Res. Quart.*, 20, No. 3, 337—349 (1969).
12. Hastings N. A. J., *Equipment Replacement and the Repair Limit Method, in Operational Research in Maintenance*, A. K. S. Jardine (Ed.), Manchester University Press, Manchester/Barnes and Noble, N. Y., 1970.
13. Hastings N. A. J., Peacock J. B., *Statistical Distributions*, Butterworth, London/Halsted Press (Wiley), New York, 1974.
14. Jardine A. K. S., *Maintenance, Replacement and Reliability*, Pitman, London/Halsted Press (Wiley), N.Y., 1973.
15. Jorgenson D. W., McCall J. J., Radnor R., *Optimal Replacement Policy*, North-Holland, Amsterdam, 1967.
16. King J. R., *Probability Charts for Decision Making*, Industrial Press, N.Y., 1971.
17. McCall J. J., *Maintenance Policies for Stochastically Failing Equipment; a Survey*, *Management Sci.*, 11, No. 5, 493—524 (1963).
18. Morse P. M., *Queues, Inventories & Maintenance*, Wiley, N.Y., 1963, pp. 168—170.
19. Mumford A. G., Shahani A. K., *A Nearly Optimal inspection Policy*, *Operational Res. Quart.*, 23, No. 3, 373—379 (1972).
20. Pierskalla W. P., Voelker J. A., *A Survey of Maintenance Models: the Control and Surveillance of Deteriorating Systems*, *Nav. Res. Log. Quart.*, 23, No. 3, 353—388 (1976).
21. Smith W. L., *Regenerative Stochastic Processes*, *Proc. Roy. Stat. Soc.*, A, 232 (1955).
22. White D. J., *Setting Maintenance Inspection Intervals using Dynamic Programming*, *J. Ind. Eng.*, 18, No. 6, 376—381 (1967).

23. White D. J., An Example of Loosely Coupled Stages in Dynamic Programming, *Management Sci.*, 19, No. 7, 739—746 (1973).
24. White D. J., Donaldson W. A., Lawrie N. L., Operational Research Techniques, 1. Business Books, London, 1969, pp. 189—192.
25. Woodman R. C., Replacement Strategies, *Operational Res. Quart.*, 18, No. 2, 193—195 (1967).
26. Woodman R. C., Replacement Policies for Components that Deteriorate, *Operational Res., Quart.*, 18, No. 3 (1967).
27. Woodman R. C., Replacement Rules for Single and Multicomponent Equipment, *J. Roy. Stat. Soc., Series C*, 18, No. 1, 31—40 (1969).

Часть II

Применение методов исследования операций к социальным и производственным системам

II-1. Системы городской сферы обслуживания

И. Савас¹), Г. Грэммас²),

ВВЕДЕНИЕ

Затраты на содержание государственного аппарата в США составляют одну треть совокупного национального продукта, причем в государственных органах управления занята одна шестая часть всех служащих страны. Восемьдесят процентов служащих государственного аппарата работают в органах управления штатов и местного самоуправления. За период с 1955 по 1970 г. численность служащих государственного аппарата удвоилась, однако это не привело к существенному улучшению состояния дел в сфере обслуживания в государственном секторе. Несмотря на непрерывный рост затрат на образование, содержание полиции, социальное обеспечение, профессиональное обучение, жилищное строительство, транспорт, санитарную и здравоохранение, жители городов по-прежнему жалуются на преступность, нищету, безработицу, невыносимые жилищные условия, ухудшение состояния окружающей среды, перенаселенность, загрязненность улиц и астрономические счета за медицинское обслуживание. Они самым непосредственным образом ощущают недостатки работы соответствующих государственных служб, особенно неэффективность деятельности местных органов управления, что ставит местные власти в этом смысле в невыгодное положение. Несоответствие между осуществляемыми затратами и получаемыми результатами (т. е. несоответствие между входными и выходными потоками государственных служб как систем управления) принимает в глазах общества угрожающие размеры.

Для улучшения работы государственных органов используются методы политического давления. Программы повышения эффективности процессов обслуживания становятся неотъемлемой частью предвыборных обещаний кандидатов в местные органы власти, которые соответствующим образом реагируют на подобный рода нажим. Хотя и с запозданием, они вновь используют традиционные способы увеличения производительности в сфере услуг — капиталовложения в технические средства и оборудование, что, естественно, дает определенные положительные результаты. В настоящее время в рассматриваемой сфере деятельности стали использоваться методы системного анализа и научные методы управления, а также вычислительная техника.

¹) E. [S. Savas, Ph. D., Professor of Public Systems Management in the Graduate School of Business at Columbia University in New York, New York.

²) Gus W. Grammas, Ph. D., Assistant Professor of Management Science in the Graduate School of Business at Columbia University in New York, New York.

В ряде случаев это делается с излишней поспешностью и неоправданным энтузиазмом, но постепенно внедрение этих методов становится все более целенаправленным и квалифицированным. Использование таких методов местными органами управления началось приблизительно на пять — десять лет позже, чем в промышленности, однако разрыв постепенно сокращается, по мере того как общественный сектор начинает успешно конкурировать с частным сектором в привлечении способных работников, нехватка которых в прошлом мешала прогрессу в этой области.

Данная глава содержит обзор работ, посвященных наиболее важным приложениям исследования операций к процессам обслуживания в городе; особое внимание уделяется проблемам и трудностям, с которыми приходится сталкиваться исследователю систем.

1. ЭКСТРЕННЫЕ СЛУЖБЫ

Экстренные службы осуществляют один из самых важных и жизненно необходимых видов обслуживания городского населения. За последние годы в этой области достигнуты определенные успехи, чему немало способствовало систематическое использование методов исследования операций для решения функциональных задач таких служб [10, 21].

Городская экстренная служба может быть определена как система, функционирование которой характеризуется следующими показателями:

- Происшествия в городе, обуславливающие требования (заявки) на обслуживание, при этом время и место этих происшествий заранее точно не предсказуемы.

- Реакция экстренной службы на каждое требование, т. е. незамедлительная отправка на место происшествия одной или нескольких оперативных единиц ¹⁾. В некоторых случаях поступившие требования могут предварительно обрабатываться с целью установления приоритета (как, например, в диспетчерском пункте приема срочных вызовов по телефону в г. Нью-Йорке — 911).

- Время, за которое оперативная бригада прибывает на место происшествия. Это время является одной из основных характеристик качества обслуживания.

- Расположение бригад в момент отправления их по вызову, которое может быть фиксированным или непрерывно меняющимся.

- Способность службы определять срочность обслуживания поступающих требований.

- Возможность реализации ряда необходимых оперативных видов деятельности, не связанных с экстренным обслуживанием.

Последние три показателя являются особенно важными, так как связаны со способом организации служб и характеризуют сте-

¹⁾ В оригинале — emergency service unit. В дальнейшем в зависимости от контекста будут использоваться термины «бригада» («оперативная бригада») или принятые в отечественной литературе наименования оперативных единиц соответствующих экстренных служб. — *Прим. перев.*

пень организации их управления и, следовательно, эффективность их функционирования.

Рассмотрим эти показатели более подробно.

Исходное расположение некоторых бригад, например пожарных машин и бригад скорой медицинской помощи, как правило, фиксированно, в то время как для бригад других экстренных служб характерно непрерывное перемещение (например, курсирование полицейских патрульных автомобилей). Таким образом, если во втором случае имеется принципиальная возможность менять зону, контролируемую бригадой (например, полицейский патруль), ее размер и конфигурацию, то зона реакции такой бригады, как пожарная команда, зависит прежде всего от (постоянного) месторасположения пожарного депо. Различие между ¹⁾ курсирующими и станциопарно расположенными оперативными бригадами уменьшается в период большой нагрузки, когда их могут направлять по вызову либо непосредственно от одного места происшествия к другому, либо в момент их возвращения с места происшествия на постоянную базу. Кроме того, в такие периоды исходное расположение оперативных единиц почти не влияет на функционирование системы (в целом).

Что касается способности экстренных служб определять степень срочности обслуживания поступающих требований, то, как показывает практика, доля вызовов, не требующих экстренного обслуживания, для разных служб различна. Она составляет ~30% для службы скорой помощи, ~40% для пожарных депо (в г. Нью-Йорке) и 75% для полицейских участков. Способность установить приоритет вызова позволяет диспетчеру выбрать любое из следующих решений, которое в конкретной ситуации представляется наиболее рациональным: не посылать бригаду, держать заявку в очереди, ожидая появления свободной бригады в зоне вызова, поставить заявку в очередь (даже при наличии свободной бригады), если ожидается получение требования с более высоким приоритетом.

Однако, если приоритет заранее определить нельзя, то, чтобы установить его, диспетчеру придется послать по каждому требованию по крайней мере одну бригаду.

Ряд экстренных служб, исключая те, которые осуществляют незамедлительное обслуживание срочных вызовов, выполняет оперативные функции профилактического характера. Можно, например, предположить, что круглосуточное патрулирование полицейских автомобилей содействует предотвращению отдельных видов правонарушений. (Хотя существуют все основания для

¹⁾ Зона реакции определяется как часть обслуживаемой территории, постоянно или на определенный период времени закрепленная за данной оперативной бригадой в качестве первоочередного (или единственного) объекта обслуживания. — *Прим. перев.*

подобного утверждения, тем не менее оно подлежит проверке.) Поэтому, чтобы не загружать без необходимости патрулирующую машину, диспетчер полицейского участка в некоторых случаях может поставить поступившее требование в очередь.

Экстренные службы функционируют в сложной обстановке, нередко при наличии совокупности противоречивых целей, а иногда и жестких административных, правовых и политических ограничений. Обычно каждая система стремится обеспечить минимальное время ожидания начала обслуживания, одновременно стараясь установить обоснованную нагрузку персонала и не выходить за рамки финансовой сметы.

В общем случае основные задачи управления такими системами заключаются в нахождении оптимального распределения имеющегося персонала и оборудования и в определении состава персонала и количества оперативных единиц (бригад), необходимых для достижения некоторых слабо формализуемых целей. Несмотря на то что найти глобальный оптимум в решении функциональных задач таких систем, по-видимому, почти невозможно, использование методов исследования операций в этой области оказалось довольно успешным.

В данном разделе основное внимание уделяется имеющимся методам исследования, разработанным методам решения соответствующих задач и, что самое важное, тем факторам повышения качества обслуживания, которые могут быть реализованы при использовании количественных, аналитических методов исследования. Сначала будут рассмотрены три основных вида экстренных служб: полиция, пожарная охрана и скорая медицинская помощь. Следует заметить, что обсуждаемые методы исследования и соответствующие пояснения применимы и к оперативным бригадам других видов экстренной службы (например, специализированным грузовикам для буксировки, группам обезвреживания мин, аварийным машинам для ремонта газопроводов и т. п.).

Функционирование экстренных служб сопряжено со значительными затратами. Например, в 1971 г. только ассигнования на нужды департаментов полиции и противопожарной охраны составили около одного миллиарда долларов. Сфера экстренных услуг по сравнению с другими отраслями хозяйства является, по-видимому, одной из наиболее трудоемких и наименее обеспеченных оборотным капиталом. При этом основная часть выделенных средств (в некоторых случаях от 90 до 98%) для каждого вида экстренной службы идет на зарплату персоналу и дополнительные выплаты. Например, при существующем уровне зарплаты содержание круглосуточно патрулирующей полицейской машины с персоналом из двух человек обходится ежегодно в 100—200 тыс. долл., а ежегодные затраты на содержание одной полностью укомплектованной персоналом пожарной машины составляют около

500 тыс. долл. Таким образом, каждое управляющее решение, связанное с увеличением численности персонала, требует очень тщательного обоснования. Широкое применение методов исследования операций позволяет выявить существенные факторы, связанные с функционированием рассматриваемых служб. Во-первых, оказывается, что некоторые изменения в методах управления и/или распределения людей и оборудования нередко влияют на эффективность систем экстренного обслуживания совсем не так, как это подсказывает интуиция. Например, для минимизации усредненного по системе времени пробега пожарных машин к месту происшествия иногда необходимо направить не ту машину, которая находится ближе всех к этому пункту, а более удаленную бригаду. Во-вторых, существенно повысить эффективность систем можно путем использования более формализованных аналитических методов, а не обычных прикидок «по эмпирическим правилам». Например, применяя модели теории массового обслуживания вместо традиционных формул расчета по среднесуточной нагрузке, можно значительно точнее определить минимально необходимое число дежурных бригад. Наконец, простые изменения процесса управления нередко приводят к более заметному улучшению функционирования системы, чем установка нового дорогостоящего оборудования или увеличение численности персонала. Например, оптимальное распределение бригад скорой помощи по обслуживаемой территории может существенно снизить среднее время реакции.

Важнейшими задачами анализа систем управления экстренными службами являются: определение необходимого количества оперативных бригад в каждом районе, разбиение обслуживаемой территории на зоны¹⁾, размещение бригад и оборудования. Для улучшения функционирования систем могут быть рекомендованы следующие мероприятия: избирательное включение в очередь требований с низким приоритетом, изменение количества бригад во времени и по районам (зонам) территории, направление по вызову не обязательно тех бригад, которые находятся ближе всего к месту происшествия, пространственное перераспределение бригад в режимах перегрузки системы и, в случае полицейской службы, закрепление за машинами пересекающихся зон патрулирования. Во многих случаях реализация указанных рекомендаций позволяет уменьшить длину очереди и время реакции системы, более равномерно распределить нагрузку по оперативным бригадам и увеличить в нужных местах количество патрулирующих единиц, выполняющих профилактические функции.

Важной характеристикой, определяющей качество функционирования системы экстренного обслуживания, является число де-

¹⁾ Т. е. определение числа и границ зон реакции. — *Прим. перев.*

журных бригад в каждом районе города. Для определения этой величины, как правило, использовался либо метод, основанный на анализе структурных (географических) характеристик района, либо метод, предусматривающий использование формул для расчета среднесуточных нагрузок.

Первый из указанных методов обычно используется в тех случаях, когда необходимо соблюдение некоторых нормативов или требований инструктивного характера. Например, для того чтобы некоторый район имел «высокую оценку» по обеспечению противопожарной безопасности, пункты, расположенные в этом районе, не должны быть удалены более чем на 1 милю (1,6 км) от мест базирования бригад, оснащенных противопожарными средствами¹⁾, а также более чем на 1,25 мили — от мест базирования бригад, оснащенных средствами для спасательных работ в многоэтажных зданиях²⁾. Однако, как правило, методы, основанные на учете только географических факторов, не позволяют определить, достаточно ли на самом деле количество выделенных дежурных бригад. Это связано с тем обстоятельством, что время от момента получения вызова на обслуживание до прибытия бригады на место происшествия зависит не только от географических, но и от многих других факторов.

В методе, основанном на расчетах среднесуточной нагрузки (или потенциальной опасности), субъективно комбинируются все факторы, которые предположительно влияют на распределение оперативных бригад. В результате для каждого района определяется «показатель риска», и оперативные бригады распределяются по районам прямо пропорционально значениям этого показателя. В общем случае применительно к полицейской службе некоторого района такой показатель рассчитывается как средневзвешенная сумма ряда показателей (таких, как число задержаний, общая протяженность улиц и т. д.), причем весовые множители представляют собой субъективные оценки относительной значимости соответствующих показателей. Однако простота формул расчета среднесуточных нагрузок обычно не отражает тех сложных, сильно нелинейных и стохастических связей между параметрами системы, которые имеют место на самом деле. Кроме того, из-за отсутствия теоретической основы таких расчетов трудно дать рекомендации относительно повышения их точности.

1.1. Использование моделей массового обслуживания

В последние годы для определения необходимого числа дежурных бригад стали использоваться методы исследования опе-

¹⁾ В оригинале engine company.— *Прим. перев.*

²⁾ В оригинале ladder company.— *Прим. перев.*

раций. С точки зрения математического моделирования системы городских экстренных служб обладают двумя характерными особенностями: вероятностным характером распределения запросов и требований на обслуживание во времени и вероятностным характером распределения происшествий и оперативных бригад по районам города. Поскольку первое из указанных свойств присуще системам массового обслуживания, то естественно, что задачи управления экстренными службами, связанные с проявлением этого свойства, решаются с использованием моделей теории массового обслуживания. Вторая особенность, обусловленная пространственными характеристиками элементов систем и объектов обслуживания, требует использования моделей минимизации времени пробега. Для того чтобы учесть обе указанные особенности и получить обобщенный алгоритм распределения, можно прибегнуть к более совершенным методам исследования, например имитационному моделированию и динамическому программированию.

Основная цель аналитического исследования с использованием методов теории массового обслуживания состоит в определении минимально необходимого числа оперативных бригад, при котором для требований с высоким приоритетом гарантируется, что время ожидания в очереди или вероятность отсутствия свободной бригады в момент поступления вызова не будут превосходить установленных предельных значений. Поскольку время прибытия и время обслуживания определяются в условиях стохастических процессов, то ни одна система с конечным числом оперативных бригад не может обеспечить незамедлительной реакции на каждый вызов. В качестве простого примера функционирования одноканальной системы массового обслуживания рассмотрим случай, когда к некоторой зоне обслуживания приписана только одна оперативная бригада (канал обслуживания), причем ни одна из других бригад в данной зоне функционировать не может. С помощью стандартных формул расчета одноканальных систем массового обслуживания для каждого сектора этой зоны устанавливаются максимальные значения вероятности появления очереди или среднего ожидания в очереди. Затем секторы этой зоны уменьшают до тех пор, пока не будут удовлетворены установленные ограничения. Общее число секторов, полученных в результате такой процедуры, определяет требуемое количество оперативных бригад. Однако ни одна из реальных систем экстренного обслуживания не функционирует подобным образом, так как при фиксированных трудовых ресурсах перераспределение последних приведет к уменьшению времени реакции, а при фиксированном качестве обслуживания — к уменьшению требуемого количества бригад.

Как правило, при исследовании реальных систем приходится использовать модели многоканальных систем массового обслу-

живания. Так, например, в работе [5] рассматривается использование такой модели (система M/M/S)¹⁾ для определения необходимого числа операторов в каждый час суток в диспетчерском пункте приема срочных вызовов по телефону в Нью-Йорке — 911. Требования к качеству обслуживания предопределяют необходимость иметь достаточно большой штат операторов, с тем чтобы доля вызовов, обслуживаемых с задержкой в T или более единиц времени, была равна или меньше P . Время измеряется в единицах, равных периоду обслуживания; таким образом, если среднее время обслуживания одного вызова равно, например, 80 с, то интервал в 160 с будет эквивалентен двум периодам обслуживания. Для расчетов в качестве входной переменной должно быть также задано среднее ожидаемое число вызовов в каждый час работы системы (недостаточная точность прогнозирования этой величины ограничивает практическое использование рассматриваемой модели). Как пример ее применения (в рамках стандартных формул для расчета многоканальной системы массового обслуживания) рассмотрим случай, когда $T = 0,25$, $P = 0,05$, среднее время обслуживания равно 80 с, причем ожидается, что в данный час функционирования системы количество вызовов может достигать 840. Для рассматриваемого случая минимально необходимое число операторов S равно 24; при уменьшении P до нуля S будет равно 26. В модели многоканальной системы массового обслуживания M/M/S предполагается, что поток требований подчиняется закону Пуассона, время обслуживания распределяется по экспоненциальному закону, все S каналов обслуживания имеют одинаковые характеристики и обслуживание заявок осуществляется в порядке поступления (дисциплина обслуживания: «первым пришел — первым обслужен», т. е. прямой порядок обслуживания).

В работе [75] показано, что даже в том случае, когда второе и третье допущения несправедливы, для существенного уточнения функциональных характеристик системы совсем не обязательно ориентироваться на использование более сложных имитационных моделей. Несмотря на то что не все из допущений соответствуют условиям функционирования экстренной службы, рассмотренной в [75], модель многоканальной системы массового обслуживания была внедрена и ее практическое использование дало вполне удовлетворительные результаты.

Модели теории массового обслуживания нашли применение в исследованиях различных систем экстренного обслуживания. В работе [93] рассматривается использование модели многоканальной системы массового обслуживания применительно к службе полиции для решения задачи о распределении полицейских пат-

¹⁾ Общепринятые обозначения различных классов систем массового обслуживания, предложенное Д. Кендлом. — *Прим. перев.*

рульных машин в Сент-Луисе; в работе [76] модели обслуживания с приоритетом используются применительно к задаче об уменьшении времени ожидания при обслуживании наиболее важных требований. В работе [20] построена модель, позволяющая определить необходимое число дежурных бригад противопожарной службы в том случае, когда существует возможность направить к месту происшествия несколько бригад. Что касается службы скорой помощи (определения необходимого числа бригад в некотором районе), то для этого предлагается модель многоканальной системы массового обслуживания М/М/С в установившемся режиме [129], а для определения количества бригад скорой помощи, необходимых для незамедлительной реакции на 95 % (или 99 %) поступающих вызовов, может быть применена модель, рассматриваемая в работе [6]. В работе [51] проведено моделирование службы скорой помощи г. Детройта, которая рассматривается как многофункциональная стохастическая система массового обслуживания при допущениях, согласно которым время прибытия бригады на место вызова определяется в рамках квазимарковского процесса и время обслуживания зависит от текущего состояния системы.

1.2. Модели минимизации времени пробега

Несмотря на то что время пребывания требования в очереди составляет преобладающую часть общего времени отклика системы в период насыщения, в обычных режимах работы наибольшую долю этого времени составляет время пробега оперативной единицы к месту происшествия. Поэтому для описания систем необходимы модели, в которых время прибытия связывается с такими параметрами, как число дежурных бригад, географические характеристики района, интенсивность поступления требований и время их обслуживания на месте происшествия. Если оперативные бригады распределены по обслуживаемой территории, то для определения требуемого количества бригад в периоды небольших и умеренных нагрузок такие модели могут использоваться вместо обычных расчетов, основанных на использовании географических факторов. Если все оперативные единицы (бригады) имеют одну базу, как в случае службы скорой помощи [109], то время пробега от пункта базирования к месту происшествия определяется только географическими характеристиками территории и не зависит от числа дежурных бригад.

При определенных допущениях зависимость длины (или времени) пробега от характера распределения оперативных бригад по территории можно найти с помощью частных моделей, основанных на геометрических соотношениях, иногда называемых «законами обратной пропорциональности квадратному корню». В работе [79] показано, что если расположение оперативных бри-

гад обеспечивает минимизацию среднего времени пробега, то это время равно $0,47/\sqrt{n}$, где n — среднее число бригад, приходящихся на квадратную милю обслуживаемой территории. В работе [76] рассмотрено использование аналогичной модели для расчета длины пробега полицейских патрульных машин при случайном характере их распределения в контролируемой зоне. В работе [68] определена величина длины пробега каждой из n прибывших оперативных бригад пожарной службы при условии, что по одному вызову направляется несколько бригад. Наиболее существенным результатом этой работы является то, что в некоторых случаях оказалось возможным установить эмпирическую связь непосредственно между средними значениями времени пробега и плотностью распределения имеющихся в паличии бригад (при этом отпадает необходимость в использовании модели для расчета длины пробега). Используя эмпирические соотношения между временем и длиной пробега и характеристики заданного распределения бригад по территории, с помощью моделей для расчета длины пробега можно определить распределение времени пробега по бригадам, что в свою очередь позволяет найти вероятность того, что время пробега не превышает установленных предельных значений [67, 70].

Модели для расчета времени пробега в сочетании с моделями теории массового обслуживания можно использовать для определения требуемого количества бригад. Задавая общее число N бригад, с помощью моделей теории массового обслуживания определяется вероятность того, что n бригад будут свободны, а с помощью геометрической модели для этих условий определяется распределение времени пробега. При этом вводят два ограничения, одно из которых устанавливает некоторые предельные характеристики очереди (вероятность появления очереди, предельное время ожидания в очереди и т. д.), а другое — соответствующие предельные характеристики времени пробега (например, среднее время пробега). После этого определяется минимально допустимое значение N , при котором удовлетворяются оба ограничения.

Применение рассмотренных выше моделей обычно еще недостаточно для определения требуемого количества оперативных бригад. Это объясняется тем, что такие простые количественные характеристики, как вероятность появления очереди и среднее время пробега, не являются показателями конечного реального эффекта функционирования систем экстренного обслуживания. Так, например, трудно количественно определить фактическую выгоду от уменьшения времени реакции системы. Но, несмотря на это, такие показатели могут быть использованы с должным обоснованием и учетом реальных условий и во многих случаях могут рассматриваться как приемлемые эквиваленты более фундаментальных критериев эффективности (таких как уменьшение

числа смертных случаев, величина предотвращенного материального ущерба и т. д.). Этот аспект проблемы рассматривается в работе [17], где методами имитационного моделирования и теории массового обслуживания с использованием таких показателей, как время пробега и нагрузка пожарных бригад, определяется, в какой степени при увеличении числа бригад снижаются перегрузки, возникающие в отдельных районах города. Для того чтобы рациональным образом распределить бригады экстренных служб, необходимо учитывать сразу несколько показателей эффективности системы. Чтобы ввести в модели теории массового обслуживания и в расчетные схемы определения времени пробега такие показатели, как нагрузка бригад и количество бригад, осуществляющих профилактические функции, требуются более сложные методы исследования, например имитационное моделирование или динамическое программирование, которые позволяют построить алгоритмы размещения оперативных бригад для самого общего случая. В работе [17] применительно к службе пожарной охраны были использованы соответствующие имитационные модели, а в работе [76] для решения задачи, связанной с распределением полицейских патрульных машин по участкам, разработан алгоритм динамического программирования, в котором используется широкий набор показателей эффективности системы. В одном из ранних исследований, посвященном моделированию работы экстренных служб [109], была использована имитационная модель сравнительного анализа работы службы скорой помощи крупного города. Результаты исследования дали более полное представление о функционировании системы и содействовали разработке ряда радикальных мероприятий, реализация которых привела к значительному улучшению обслуживания. Эти мероприятия включали организационную централизацию службы, территориальную децентрализацию бригад скорой помощи за счет их распределения по вновь созданным периферийным базам, сортировку вызовов по приоритету для уменьшения потока заявок, перераспределение диспетчерских полномочий, а также установку новых технических средств, облегчающих выполнение работы диспетчера и органов оперативного управления.

1.3. Зоны реакции

Разбиение обслуживаемой территории на зоны является одним из важных аспектов проблемы эффективной организации системы экстренного обслуживания. Зона реакции конкретной оперативной единицы (полицейского патруля, пожарной машины или бригады скорой помощи) — это тот участок территории, который закреплен за данной единицей в качестве первоочередного объекта обслуживания. При определении площади и конфигурации

зоны реакции могут преследоваться несколько (и нередко противоречивых) целей, например минимизация времени реакции, выравнивание нагрузки, выдерживание демографической однородности участков, удобство управления. Очевидно, что с помощью только математического метода нельзя учесть всю совокупность факторов, относящихся к анализу данной проблемы, тем не менее в последнее время был разработан ряд моделей, обеспечивающих рациональные подходы к ее решению.

В целях минимизации времени, необходимого оперативной единице для прибытия на место происшествия, территория обычно разбивается на квадратные или круговые зоны. Поскольку, однако, скорость передвижения бригад может зависеть от направления движения, для каждого предполагаемого варианта планировки зон реакции используются количественные методы предварительных оценок параметров, определяющих время пробега. Например, в работе [76] было показано, что если каждый полицейский патруль действует только в своей зоне реакции, то среднее время пробега в пределах этой зоны минимизируется лишь при условии, что среднее время пробега в направлении восток — запад такое же, как и в направлении север — юг. В частности, для Манхэттена, где скорость движения в направлении север — юг почти в 4 раза выше, чем в направлении восток — запад, это означает, что размер зоны реакции в первом направлении должен быть в 4 раза больше, чем во втором. Можно ожидать, что при таком планировании зон среднее время пробега уменьшится приблизительно на 20% по сравнению со случаем обычной планировки квадратных или круговых зон реакции.

Рассмотрим другой пример. Администрация противопожарных служб обычно распределяет зоны реакции (участки) таким образом, чтобы к месту возникновения пожара направлялась бригада, наименее удаленная от данного пункта. В этом случае все пункты, расположенные на границе двух смежных участков территории, оказываются одинаково удаленными от обслуживающих их пожарных депо. Однако в работе [16] было показано, что даже в этом случае, когда под «ближайшей» бригадой понимается бригада, для которой «время пробега минимально», получаемое разграничение территории не оптимально и минимизация среднего времени пробега для системы в целом может быть достигнута при такой организации работы, когда к месту происшествия направляется не «ближайшая», а некоторая другая бригада. Следовательно, чтобы система в целом осталась в состоянии, позволяющем наилучшим образом обслужить ожидаемые требования, в данный момент может оказаться предпочтительным поступиться с теми потерями, которые обусловлены увеличением времени пробега. Авторы указанной работы показали также, что разграничение территории, при котором минимизируется среднее время

реакции для системы в целом, во многих случаях уменьшает и неравномерность нагрузки (т. е. различие в продолжительности активной работы наиболее и наименее загруженных оперативных единиц). Существенно упрощенный метод определения границ участков обслуживания с использованием эвристических методов приведен в работе [63]; принятое автором условие установления границ участков обслуживаемой территории состоит в том, что каждой базе оперативных бригад приписываются все те пункты, до которых либо расстояние, либо время пробега минимальны.

1.4. Размещение базовых пунктов и распределение оперативных единиц

Распределение оперативных единиц (бригад) и размещение баз (постов) экстренных служб по обслуживаемой территории являются наиболее важными факторами, которые характеризуют эффективность функционирования таких систем. Задачи размещения касаются выбора места для размещения новых баз, выбора нового места базирования при объединении двух или большего количества существующих баз, «исходного расположения» (т. е. расположения оперативных бригад перед отправкой их на линию) и порядка «перераспределения» бригад (т. е. как и при каких условиях корректировать размещение оперативных единиц при работе системы в режимах перегрузки). Как правило, задача о размещении средств обслуживания рассматривалась с учетом лишь экономических показателей, т. е. вероятностные аспекты функционирования системы не принимались во внимание. В работе [126] эта задача рассматривалась как задача о покрывающем множестве, причем каждое из рассматриваемых множеств включало все потенциально возможные места базирования, удовлетворяющие установленным относительно каждого обслуживаемого пункта ограничениям на время реакции или расстояние. Если затраты на размещение оборудования для всех вариантов расположения баз одинаковы, то для минимизации общего числа баз, удовлетворяющего нормативам (т. е. максимально допустимым значениям) времени реакции и удаленности от обслуживаемых пунктов, может быть использована модель линейного программирования.

Сфера использования вероятностных аналитических методов решения задачи размещения оперативных бригад экстренной службы обычно ограничивается случаями, когда число бригад или количество потенциальных мест их размещения невелико. Эти методы позволяют определить время пробега для всех возможных комбинаций мест базирования или использовать простые алгоритмы, помогающие установить оптимальную схему размещения. В работе [109] проанализированы сравнительные характеристики

работы службы скорой помощи для двух вариантов размещения средств обслуживания: исходного — при одной центральной базе и предлагаемого — при двух фиксированных пунктах базирования. При этом рассматривались все возможные варианты распределения n бригад между двумя пунктами базирования, а среднее время реакции определялось с помощью имитационной модели. В работе [79] также рассмотрен случай двух мест базирования и для одного из них допускалось произвольное расположение. Метод наискорейшего спуска позволил авторам этой работы найти такое расположение второй базы, при котором минимизируется среднее время пробега для системы в целом. Более существенным, однако, оказался результат, свидетельствующий о том, что оптимальное расположение новой базы почти не зависит от исходного распределения имеющихся в наличии бригад. В работе [123] при решении задачи о размещении бригад скорой помощи, минимизирующем среднее время реакции, с помощью процедур имитационного моделирования была получена информация о характеристиках системы при различных способах закрепления бригад и правилах диспетчеризации. Затем результаты имитационного моделирования были использованы для разработки аналитического метода оценки среднего времени реакции. После этого с помощью вероятностного алгоритма метода ветвей и границ была решена комбинаторная задача оптимизации, сформулированная в результате реализации предыдущих этапов исследования.

Были выполнены ряд работ, содержащих также аналитические исследования более широкого плана, включающие решение задач о перераспределении оперативных бригад во время патрулирования. В работе [76] показано, что локальное перераспределение полицейских патрулей (т. е. перераспределение их между соседними секторами обслуживаемой территории) обеспечивает уменьшение времени пробега только в тех случаях, когда патрули сосредоточены вблизи границ соответствующих участков. В работе [69] решение задачи перераспределения пожарных бригад базируется на введении понятия о «покрытии». Некоторый пункт называется «покрываемым», если по крайней мере для одной из существующих баз противопожарных или спасательных бригад время пробега до этого пункта не превышает T мин. Если ожидается, что один или несколько участков окажутся «непокрытыми» из-за слишком большого времени пробега, то (для постановки задачи в рамках математического программирования) используется эвристический алгоритм, позволяющий определить, какие незакрытые пожарные бригады следует послать в данный район и какие из имеющихся оперативных единиц нужно переместить на место посланных. Результаты экспериментальных расчетов свидетельствуют о том, что предложенный алгоритм хотя и не дает «оптимального» решения, все же намного лучше, чем существующие

методы, и, по-видимому, его использование дает весьма рациональные решения по перераспределению бригад. Его реализация на ЭВМ требует мало машинного времени и небольшого объема памяти. Поэтому при использовании ЭВМ данный метод может найти применение в рамках системы оперативного управления, функционирующей в реальном масштабе времени.

1.5. Имитационное моделирование с использованием ЭВМ

В настоящее время для количественной оценки предлагаемых мероприятий, связанных с размещением баз и бригад экстренных служб на территории, стали более широко применяться методы имитационного моделирования с использованием вычислительной техники. Обычно органам управления соответствующей экстренной службы приходится рассматривать одновременно несколько предложений по улучшению методов распределения технических и трудовых ресурсов. (Такие предложения могут касаться, например, установления дисциплины обслуживания, увеличения числа бригад в определенные часы суток, выбора новых мест размещения некоторых бригад, изменения зоны реакции или состава патрульного подразделения и совершенствования процедур перераспределения бригад.) Кроме того, для реализации некоторых мероприятий такого рода могут быть рекомендованы и определенные технологические нововведения (например, использование систем автоматического слежения за автомобилями). Поэтому, прежде чем сделать выбор одного из возможных вариантов, лица, принимающие решения, обычно считают необходимым провести объективный технико-экономический анализ каждого предложения по критерию «затраты — выгоды».

Как правило, для выполнения таких исследований необходимо применение имитационных моделей с использованием ЭВМ, что позволяет получить информацию о влиянии предлагаемых мероприятий на широкий круг параметров системы: время реакции на данный тип вызова, загрузку бригад, время ожидания в очереди, наличие (пезанятых) бригад и т. д. Ряд исследований, выполненных с использованием имитационных моделей [17, 109], был рассмотрен ранее. В работе [38] описана реализованная на ЭВМ модель, позволяющая предсказывать распределение полного времени реакции одной из реально функционирующих систем. Эта модель используется совместно с программой, осуществляющей непосредственный поиск такого распределения бригад скорой помощи по обслуживаемой территории, при котором среднее время реакции будет минимальным. Применение описанного подхода к анализу функционирования службы скорой помощи в Лос-Анджелесе показало, что с его помощью можно также существенно уменьшить и неравномерность загрузки бригад. В работе [76]

рассматривается имитационная модель для определения наиболее эффективных условий использования автоматической системы слежения за транспортными средствами полицейских участков. Как показали проведенные исследования, распространение сферы действия этой системы на патрульные машины, приписанные к непересекающимся секторам контролируемой территории, позволяет уменьшить среднее время пробега на 10—20%.

Практика использования моделей исследования операций для оценки функционирования систем экстренного обслуживания позволила выявить, на какие характеристики систем может оказать положительное или отрицательное влияние изменение отдельных процедур управления. Работы [55, 58, 77, 78] достаточно убедительно свидетельствуют о том, что сложные аналитические модели могут играть важную роль в исследованиях систем экстренного обслуживания. Следует ожидать, что в дальнейшем такие модели будут постоянно улучшаться как в отношении их применимости, так и по уровню разработки и, что самое важное, будут более широко использоваться работниками управления соответствующих служб в практической деятельности.

2. ОХРАНА ОКРУЖАЮЩЕЙ СРЕДЫ

Проблема охраны окружающей среды обычно подразделяется на три самостоятельные проблемы: охраны почвы, водного и воздушного бассейнов. Поскольку, однако, в природе все взаимосвязано, то естественно, что попытка решить проблему загрязнения почвы (в частности, сбор и уничтожение твердых отходов) может привести к проблеме загрязнения воздуха (сжигание отходов) или воды (сброс отходов). Кроме того, следует иметь в виду, что не всякий сброс отходов обязательно приводит к ухудшению состояния окружающей среды. Нежелательные последствия для пользователей земель, водой или воздушным бассейном возможны только в том случае, когда сброс отходов осуществляется в течение длительного периода времени и концентрация загрязняющих веществ, содержащихся в отходах, превышает установленные нормы. До недавнего времени считалось вполне естественным, что проблема охраны окружающей среды является прерогативой инженеров санитарных служб. Однако в настоящее время, когда охрана окружающей среды стала одной из актуальнейших проблем современности, для ее решения привлекаются различные специалисты и используются самые современные методы, причем главная роль отводится методам исследования операций и системному анализу.

Одна из самых ранних работ [5], содержащая всесторонний количественный анализ управления службой, ответственной за ликвидацию городских отходов, была выполнена по инициативе Regional Plan Association. Основные вопросы, которые рас-

смаатриваются в работе, — это перспективные оценки образования отходов, определение затрат и оценка факторов неопределенности, связанных с реализацией возможных вариантов управления, а также определение характера информации, необходимой для плановой организации работ по удалению отходов. Возможность использования методов исследования операций для решения соответствующих проблем обсуждается в работах [24, 26]. В частности, в работе [26] содержатся критические замечания в адрес исследователей, которые, не отрицая существование принципиальной возможности для улучшения контроля за загрязнением, тем не менее считают необходимым прежде всего осуществить реорганизацию соответствующих административных органов. Автор этой работы ратует за такой подход к решению рассматриваемых проблем, который возможен с учетом ограничений, накладываемых существующей системой управления. В работе [24] показана возрастающая роль исследователей систем в решении задач охраны окружающей среды, причем основными направлениями их работ, по мнению автора, являются следующие:

- определение условий загрязнения, разработка усовершенствованных методов регистрации и измерений загрязнения, стоимостная оценка последствий загрязнения;

- разработка методов контроля за загрязнением: саморегулирование и соблюдение требуемых стандартов качества;

- участие в исследованиях по оценке затрат и результатов для крупномасштабных систем контроля за качеством окружающей среды;

- разработка теоретических основ и комплексных моделей социальных и экономических процессов, связанных с охраной окружающей среды.

Существует два основных количественных подхода к решению проблемы управления сбросом отходов. Согласно первому из них, для источников загрязнения законодательным путем устанавливаются способ сброса и его максимальная величина. Такой подход обычно предусматривает минимизацию затрат при заданных ограничениях и приводит к построению оптимизационных моделей [99]. Другой подход основан на введении обязательных налогов, размер которых зависит от объема (количества) сброса загрязняющих веществ, содержащихся в отходах [40, 41]. Целью подобных экономических санкций является улучшение состава отходов, сбрасываемых источником загрязнения. Такой подход, как правило, приводит к построению математических моделей для определения (социальных) затрат, связанных с удалением загрязняющих веществ.

Модели исследования операций используются и для описания физических процессов в окружающей среде (например, модель диффузии использовалась применительно к процессам контроля

за качеством воздуха), а также для нахождения решений, обеспечивающих минимум затрат при соблюдении некоторого минимального уровня обслуживания (например, сетевые модели работ по уборке мусора).

2.1. Твердые отходы

Проблема сбора и удаления твердых отходов является одной из самых многоаспектных проблем охраны окружающей среды. Твердые отходы, как правило, содержат все нежелательные компоненты, которые к тому же очень трудно удалить из жидких или газообразных продуктов сброса. Управляющие решения в этой области связаны с планированием работы предприятий и использованием оборудования в рамках существующей социальной (и политической) системы. Состав и количество твердых отходов, образующихся в текущем периоде, или соответствующие прогнозные оценки, обычно рассматриваются как входной поток системы, и основное внимание уделяется сбору, транспортировке, накоплению, обработке и уничтожению этих отходов. Кроме того, в качестве одной из функций системы управления удалением твердых отходов рассматриваются утилизация полезных веществ и использование мусора в качестве топлива.

Теоретически процесс принятия решений включает две стадии. На первой стадии определяется желаемый характер обслуживания, а на второй — методы реализации такого обслуживания с минимальными затратами. Однако затраты, связанные с различным уровнем обслуживания, как правило, определяются лицом, принимающим решение и одновременно осуществляющим выбор «желаемого» уровня. Решения относительно уровня обслуживания касаются периодичности уборки, вида подлежащих уборке отходов, а также их сортировки в соответствии с требованиями переработки. Для оценки затрат, связанных с изменением качества обслуживания, используются разные дескриптивные и прогнозные модели. Модели, предусматривающие минимизацию затрат, позволяют определить число и виды сооружений (участков для свалки, мусоросжигательных станций, перевалочных пунктов), габариты и вид транспортных средств для сбора и перевозки отходов, составить маршруты движения этих средств, а также определить численность обслуживающих бригад.

С точки зрения моделирования система удаления твердых отходов включает следующие элементы: источники отходов, стационарные сооружения и оборудование для перемещения, обработки и уничтожения отходов, средства транспортировки отходов и персонал системы. Большинство математических моделей таких систем являются оптимизационными, за исключением моделей, связанных в основном с прогнозированием образования отходов.

Работа [8] содержит обзор оптимизационных моделей для решения различных задач в области организации работ по удалению твердых отходов.

Перспективные оценки образования твердых отходов требуются для решения многих задач управления процессами сбора и уничтожения мусора.

В работе [45] методы линейного регрессионного анализа используются для определения коэффициента пропорциональности в уравнении, связывающем среднюю плотность материала отходов с весовой долей содержащейся в них бумаги. Согласно условиям эксплуатации мусороперерабатывающих станций, необходимо, чтобы плотность мусора соответствовала установленным нормам. В работе [101] проводится сравнительный анализ данных о темпах накопления бытовых отходов при уборке мусора один и два раза в неделю. Оказалось, что при уборке мусора два раза в неделю наблюдалось заметное увеличение скорости его накопления, которое сохранялось в течение длительного времени. Установленная закономерность до сих пор не нашла должного объяснения.

Задачи выбора сооружений (предприятий) и мест их размещения достаточно просто решаются с помощью оптимизационных моделей, которые обычно являются моделями частично целочисленного программирования (см. работу [81], а также главу II-8 «Размещение и организация работы оборудования»). В этом случае задача может быть сформулирована следующим образом:

найти

$$\min \sum_i \sum_j (T_{ij} + D_j) X_{ij} + \sum_j F_j Y_j$$

при ограничениях

$$\sum_j X_{ij} = a_i \quad \text{для каждого } i\text{-го источника,}$$

$$b_j Y_j - \sum_i X_{ij} \geq 0 \quad \text{для каждого } j\text{-го пункта,}$$

$$Y_j \quad \text{— целые числа, равные 0 или 1,}$$

где Y_j равно 1, если сооружение строится в j -м пункте, и равно 0 в противном случае; F_j — постоянные затраты на строительство такого сооружения; D_j — затраты на ликвидацию 1 т отходов на мусоросжигательной станции в j -м пункте; T_{ij} — затраты на перевозку 1 т отходов от i -го источника (зоны сбора) до j -го пункта; a_i — количество отходов, образующееся в i -й зоне сбора; b_j — производственная мощность в j -м пункте; X_{ij} — количество отходов, которое должно быть перевезено из i -й зоны сбора в j -й пункт. Первая группа ограничений соответствует требованию, согласно которому отходы, накапливаемые в каждой из зон, должны быть собраны и вывезены полностью; вторая группа ограничений

выражает условие, согласно которому поток отходов, направляемый на соответствующее сооружение (предприятие), не превышает его мощности по переработке. Подобный класс моделей получил дальнейшее развитие в работах [92, 105], что позволило помимо указанных параметров определять оптимальное расположение перевалочных пунктов. С вопросами экономического анализа организации перевалочных станций (и пунктов перегрузки) читатель может ознакомиться в работе [46].

Для решения задач составления маршрутов транспортных средств, предназначенных для сбора и перевозки твердых отходов, необходимо разделить всю зону обслуживания на отдельные участки и составить маршруты движения транспортных средств от депо к соответствующему участку, затем — по территории этого участка и, наконец, назад — в депо. Целью такой задачи является обеспечение требуемого уровня обслуживания (периодичности сбора) при минимуме затрат. Первые попытки решения подобной задачи соответствовали в своих постановках классическим задачам «коммивояжера» и «китайского почтальона» [81].

В последнее время при постановке и решении задач маршрутизации стали использоваться другие модели исследования операций. Так, в работе [27] обсуждается модель организации работ по уборке мусора, включающая такие параметры, как характеристики планировки жилых застроек, время накопления отходов, численность бригад и расстояние от мест накопления отходов до пунктов их уничтожения. Авторами этой работы предложен математический метод распределения рабочих-мусорщиков по транспортным средствам и составления маршрутов этих средств таким образом, что совместная реализация этих процедур обеспечивает минимизацию общих затрат по уборке мусора. В работе [13] для нахождения наилучших маршрутов используются имитационные модели, реализация которых на ЭВМ позволяет осуществлять перебор большого числа возможных вариантов маршрутов и выбрать вариант, требующий минимальных затрат. В работе [7] для составления маршрутов, минимизирующих общее время пробега, и определения минимально необходимого количества транспортных средств, обеспечивающих заданный уровень обслуживания, были использованы сетевые модели и методы теории графов. Поскольку, однако, маршруты транспортных средств зависят от схемы размещения предприятий, то более совершенной является комплексная модель размещения и составления маршрутов. Тем не менее на практике решение задачи составления маршрутов является решением оперативного характера (краткосрочным), тогда как решение задачи о размещении предприятий — это решение стратегического плана (длительного срока действия).

Задачи календарного планирования работы персонала обычно сводятся к равномерному распределению нагрузки и использова-

нию соответствующего количества рабочих. Для определения необходимой численности персонала по уборке бытовых отходов в работе [3] была использована модель нелинейного программирования. Составление графика работы преследовало цель избежать чрезмерного накопления мусора у обочины тротуара, с тем чтобы минимизировать количество мусора, который накапливается в течение недели. При этом выдерживались правила профсоюзного устава рабочих-мусорщиков, а также ограничения на численность персонала и количество транспортных средств. Для решения аналогичной задачи в работе [12] были использованы эвристические алгоритмы. Такую же направленность имеет и работа [57], в которой для распределения бригад по сменам использована модель линейного программирования, причем основная цель состояла в оптимизации соотношения между фондом заработной платы и накоплением неубранного мусора при соблюдении соответствующих ограничений. Однако в работах [114, 116] показано, что принятые в пределах отдельных муниципальных районов организационные формы уборки мусора могут изменять затраты в три раза, что препятствует практической реализации тех преимуществ, которые можно было бы получить от мероприятий по маршрутизации и календарному планированию.

Экономический анализ во многих случаях содействует изучению возможности утилизации твердых отходов — как с целью извлечения полезных ресурсов (материалов), так и с целью получения топлива. В работе [89] обсуждается возможность введения системы стимулов (налогов) для минимизации количества образующихся твердых отходов; минимизации стоимости и максимизации эффективности сбора отходов, переработки и уничтожения отходов; максимизации доли таких твердых отходов, которые возможно экономичным способом отделить и повторно использовать; минимизации отрицательного воздействия процессов ликвидации твердых отходов на все экологические системы. В работе [1] показано, что в экономических и экологических аспектах утилизация обладает существенными потенциальными преимуществами по сравнению со способами удаления твердых отходов, которые используются муниципальными службами в настоящее время. Повторное извлечение ресурсов пока еще не самоокупается, однако, если его продукция будет обеспечена рынком сбыта, утилизация отходов может оказаться конкурентоспособной относительно тех возможных вариантов, которые имеются у городов для решения проблем ликвидации твердых отходов. В настоящее время проводятся исследования с целью использовать мусор, собираемый в г. Денвере, в качестве основного вида топлива для электростанции Coors Brewery в г. Годдене, шт. Колорадо; предполагается также в нижней части нью-йоркской бухты создать из твердых

отходов искусственный остров Ресар площадью в 225 акров, где из мусора будут извлекать полезные компоненты, а затем использовать его в качестве топлива.

2.2. Охрана и распределение водных ресурсов!

Методы исследования операций используются также при анализе систем, связанных с использованием водных ресурсов [48, 90, 91]. В общем случае основная проблема использования ограниченных водных ресурсов заключается в том, чтобы в достаточном количестве и с соблюдением требуемого качества распределить их по соответствующим направлениям использования, нередко конкурирующим между собой. Проблема качества воды непосредственно связана с определенной противоречивостью целей технического прогресса и охраны окружающей среды. Для выработки политики, препятствующей ухудшению качества воды, решающую роль приобретает тщательный анализ совокупности технических, экономических и социальных последствий, определяемых организацией управления сбросом отходов. Количественный аспект проблемы использования водных ресурсов связан с обеспечением рационального распределения воды среди различных потребителей. В отдельных случаях это обусловлено тем, что для конкретных нужд в некоторых местах воды или слишком много (например, на затопляемых сельскохозяйственных угодьях), или слишком мало.

Управление качеством воды требует решения таких основных проблем, как определение требуемого качества воды, определение путей обеспечения заданного качества при минимальных затратах и разработка плана реализации требуемых мероприятий. Для решения первой из них необходимо, чтобы требования к качеству воды были четко определены через физические параметры, которые можно измерить или по крайней мере достаточно надежно оценить. Наибольший вклад методы исследования операций призваны внести в решение второго и третьего вопросов. В последние годы для оценки возможных путей улучшения качества воды был разработан ряд математических моделей. Рассматриваемые варианты включают такие способы, как обработка сбросов у источника их образования, «закачка» отходов по трубопроводу в наименее чувствительный слой водного бассейна, интенсификация потока и струйная аэрация, введение таких экономических санкций, как налоги, с тем чтобы уменьшить потребление воды источниками чрезмерного загрязнения водных ресурсов.

При математическом моделировании систем, связанных с использованием водных ресурсов, обычно приходится применять сложные модели большой размерности. Это объясняется тем, что качество воды не может быть охарактеризовано одной переменной (например, содержанием растворенного кислорода), а определяется

совокупностью показателей, таких, как содержание и тип взвешенных и органических веществ, температура, биологическая норма кислорода, содержание хлоридов, фосфатов, нитратов, наличие водорослей и т. д. В работе [48], например, приводится описание обобщенной (нелинейной) математической модели большой размерности для определения оптимальной стратегии (экономических санкций), обеспечивающей требуемое качество воды при минимальных общих затратах на мероприятия по охране окружающей среды.

Существует два основных вида математических моделей для изучения процессов, определяющих качество воды. Первый тип моделей — это физические модели, описывающие причинно-следственные связи между величиной сброса отходов и качеством воды в конкретных природных условиях. Так, например, в работе [85] моделирование основного потока стохастического характера (который разбавляет продукты загрязнения) и последующая оценка различных вариантов сброса воды базируются на методе, в котором используются цепи Маркова. Второй тип моделей — это модели математического программирования для определения такого сочетания возможных вариантов, которое обеспечивает минимум затрат на достижение поставленной цели и может быть реализовано местными органами управления. Ряд моделей математического программирования возрастающей сложности (от линейных до нелинейных), предназначенных для решения подобной задачи, представлен в работе [90]. В качестве переменной x_j принято количество отходов, сбрасываемых j -м источником загрязнения; целевая функция модели — $\sum f_j(x_j)$ — общие затраты на удаление отходов; ограничение заключается в том, что исходное качество воды по крайней мере не ухудшится. Заметим, что в рассматриваемой модели качество воды определяется только некоторым множеством дискретных значений параметров и в постановке задачи отсутствует временной аспект (задача рассматривается в рамках установившегося режима, т. е. без учета стохастического характера потока воды и отходов, а также физических параметров). В работе [85] принято допущение о линейности функции $f_j x_j$ и соответствующих ограничений. Модели, построенные в данной работе, можно использовать не только для определения затрат на системы, обеспечивающие различные нормативы качества, но и для оценки этих затрат при изменении режима течения загрязненной воды и схемы размещения очистных сооружений. В работе [91] показано, что до сих пор еще не решены две важные проблемы, связанные с управлением качеством воды: размещение местных очистных сооружений, при котором затраты при заданном качестве воды будут минимальными, и нахождение критериев для распределения этих затрат.

Один из возможных подходов к решению проблемы контроля за качеством воды заключается в использовании экономических

рычагов для стимулирования требуемых изменений состава продуктов, выбрасываемых источником загрязнения. Авторы работы [65] отстаивают такой экономический подход, позволяющий руководству предприятия — источника загрязнений — принимать более эффективные решения, связанные с контролем за составом отходов. Однако интересно отметить, что комиссия по использованию бассейна р. Делавэр, по-видимому, будучи более уверенной в эффективности традиционных подходов, использовала для целей регламентации и соблюдения нормативов качества воды модели математического программирования, а не экономический подход [90].

Проблемы управления водоснабжением города связаны с влиянием урбанизации и развития города на запасы водных ресурсов. В работах [30, 31] рассмотрены вопросы проектирования гидротехнических сооружений, осуществляющих подачу воды и ее обработку для различных нужд. Применительно к проектированию таких сооружений в этих работах предлагается метод анализа по критерию затраты — эффективность; применение этого метода иллюстрируется на примере проектирования третьего водовода Нью-Йорка. В предлагаемом подходе сделана попытка преодолеть противоречивость целей двух заинтересованных сторон — органов управления и проектных организаций. По-видимому, главное достоинство предлагаемого метода состоит не столько в научном уровне разработки, как это обычно бывает, а в правильной постановке целей исследования. Существенную роль сыграли методы исследования операций и в решении вопросов, связанных с эксплуатацией систем водоснабжения. Водохранилища, например, имеют многоцелевое назначение: производство электроэнергии, защита от наводнения, водоснабжение, рекреация, причем реализация каждой из функций требует различных режимов эксплуатации. Для ознакомления с основными задачами в этой области — определением необходимых запасов воды и водоснабжением систем со стохастическим характером входных воздействий — читатель может обратиться к работам [103; 104], в которых расчет режимов эксплуатации водохранилища проводился с помощью методов линейного программирования с вероятностными ограничениями.

2.3. Контроль за загрязнением атмосферы

Воздух, как и вода, является природным богатством. В последние годы проблема ухудшения качества воздушного бассейна стала предметом многих научных исследований. Деятельность человека вызывает загрязнение атмосферы в основном за счет таких процессов, как сжигание топлива для получения энергии, сжигание отходов, выделение газообразных побочных продуктов при

работе различного рода предприятий. Хотя проблема загрязнения атмосферы — это проблема мирового масштаба, особенно актуальна она для городских агломераций, где источники загрязнения концентрируются на относительно небольшой площади с высокой плотностью населения. Обсуждение проблем, связанных с загрязнением атмосферы, и характеристику программы контроля за загрязнением атмосферы в г. Нью-Йорке читатель найдет в работе [33]. Существующие проблемы выбора между сокращением производства или загрязнением атмосферы в городских районах обсуждаются в работе [100]. Очевидно, что для сохранения нормального качества воздуха в крупных городах необходимы систематический анализ его состояния и плановая основа в управлении качеством воздушного бассейна.

Конечные цели программы контроля за загрязнением атмосферы обусловлены медицинскими, эстетическими и экономическими последствиями загрязнения. Однако современный уровень знаний в этой области фактически не позволяет установить надежные количественные связи между указанными конечными факторами и уровнем загрязнения атмосферы; нормативы для контроля трудно выразить в соответствующих количественных характеристиках. Но, несмотря на это, последние работы, выполненные в данном направлении, содержат обнадеживающие результаты. Так, в работе [102] изложены основные принципы последовательного подхода к стоимостной оценке ущерба здоровью или благосостоянию вследствие загрязнения атмосферы, причем принятый подход иллюстрируется на примере стоимостной оценки повышения смертности, вызванной загрязнением воздуха. Есть основания надеяться, что со временем нормативы качества атмосферы будут фиксированы в виде максимально допустимых годовых доз конкретных загрязняющих веществ. Между тем целевые установки разрабатываемых программ должны быть выражены через вторичные факторы, а именно через качество воздуха, которое само по себе поддается достаточно четкому количественному описанию и измерению (например, нормативы качества воздуха в отношении концентрации двуокиси серы можно выразить следующим образом: среднеарифметическое значение за год не должно превышать 0,03 части на миллион, а максимальное среднесуточное значение — 0,14 части на миллион)¹⁾. Однако решение этой технической проблемы связано преимущественно с правовыми и социально-политическими аспектами ее рассмотрения, как это и следует ожидать при решении любого вопроса, затрагивающего существенные стороны общественного развития. Как правило, программы борьбы с загрязнением атмосферы могут быть сформулированы в отношении допустимых видов топлива и устройств для его

¹⁾ Имеются в виду объемные доли. — *Прим. перев.*

сжигания, регулирования выбросов, а также мест размещения и режимов работы источников загрязнения.

Для планирования мероприятий по борьбе с загрязнением воздушного бассейна может быть использован количественный, аналитический подход, включающий следующую классическую последовательность этапов: количественное определение поставленных целей, составление исчерпывающего набора возможных вариантов стратегий, установление оценки критериев эффективности возможных вариантов (или их сочетания) и строгую количественную оценку возможных стратегий по критерию затраты -- выгода. В одном из наиболее полных научных исследований проблемы загрязнения атмосферы, которое было выполнено под эгидой American Association for the Advancement of Science [4], рассматриваются экономические последствия загрязнения атмосферы, государственная политика в области охраны атмосферы и вопросы формирования общественного мнения, которое в свою очередь влияет на выработку государственной политики в рассматриваемой области. В работе [107] изложены принципы построения системы имитационного моделирования, предназначенной для совершенствования планирования, осуществляемого соответствующими городскими службами контроля за загрязнением атмосферы, и дано описание элементов этой системы. Такая система может использоваться в процессах текущего и долгосрочного планирования; она обладает также определенными потенциальными возможностями предсказания качества воздуха в режиме реального масштаба времени. С точки зрения построения систем управления с обратной связью идентифицируемые входы, выходы, средства измерения и управляющие элементы могут быть синтезированы в рамках системы, реализующей некоторую совокупность стратегий управления качеством воздуха в городе.

Управляющие воздействия на источники загрязнения, выбросы которых контролируются системой, определяются степенью отклонения качества воздуха от установленного норматива. Кроме того, если можно предсказать влияние будущих погодных условий, то с помощью системы могут быть реализованы не только корректирующие, но и упреждающие воздействия.

Другой подход к решению проблем загрязнения атмосферы принят в работе [34], где обсуждается вопрос о том, каким образом методы анализа решений могут быть использованы государственными органами при составлении программы контроля за качеством атмосферы. В этой работе рассматривается совокупность целей и показателей эффективности, которые используются для описания последствий каждого из возможных вариантов решений. Наиболее важный аспект этой работы связан с идеей выработки политически жизнеспособных решений.

Существует два основных вида математических моделей для

исследования качества воздуха. Первый тип моделей описывает физические процессы, связанные с загрязнением атмосферы. Например, в работе [74] показано, что использование семейства кривых логарифмически-нормального (двухпараметрического) распределения вероятностей обеспечивает отличное совпадение с экспериментально полученными данными о кумулятивной плотности распределения суточных концентраций двуокиси серы. В работе [96] описано использование теории множеств и моделей диффузии для дистанционного контроля за источниками загрязнения. Второй вид моделей представлен моделями математического программирования, предназначенными для выбора возможных вариантов политики в области ограничения загрязнения атмосферы. В работе [66] представлена модель линейного программирования, базирующаяся на допущении о том, что для воздушного бассейна данного географического района существует максимально допустимая величина ежегодных выбросов определенного загрязняющего вещества. С помощью этой модели было получено решение применительно к воздушному бассейну Сент-Луиса с целью проверки утверждений о том, что введение ограничений на содержание серы в угле для уменьшения загрязнения атмосферы непригодно с экономической точки зрения. В этой работе показано, что данная мера оказалась оправданной лишь для определенных типов топок. Кроме того, был сделан вывод о том, что котлы с к. п. д. 75 % или ниже следует перевести на природный газ. Однако в целом возможности применения моделей математического программирования в области контроля за загрязнением атмосферы, по-видимому, ограничены. Большинство исследователей систем, использующих математические модели при решении проблем загрязнения атмосферы, основное внимание уделяли либо анализу и систематизации данных, либо моделированию процессов диффузии и атмосферных явлений.

Очевидно, что один из возможных способов управления качеством воздуха заключается в том, чтобы в целях предотвращения его загрязнения применить в отношении источников загрязнения такие санкции, как штрафы (или налоги). Такой подход использован, например, в работе [19], где рассмотрена проблема установления соответствующих обязательных налогов, взимаемых с каждого пользователя автомагистралей и пропорциональных стоимостной оценке ущерба, наносимого обществу.

3. САНИТАРНЫЕ СЛУЖБЫ. СЛУЖБЫ УЛИЧНОГО ДВИЖЕНИЯ

В отличие от рассмотренных ранее видов экстренных служб работа санитарных служб и служб уличного движения не столь заметна, тем не менее именно она в значительной степени определяет бытовые условия городского населения. Более того, недоста-

точный уровень развития таких служб может привести к серьезным политическим последствиям [113]. Основные функции рассматриваемых служб — это уборка снега, организация и обслуживание стоянок автомашин, уборка мусора, ремонт улиц и регулирование уличного движения.

9 февраля 1969 года в результате сильного бурана, разразившегося над Нью-Йорком, образовались снежные запасы, что привело к опасным для города осложнениям и политическому кризису. При анализе происшедшего события [113] основное внимание уделялось следующим вопросам: Какое количество снега обычно выпадает в Нью-Йорке? Какой объем работ должен быть выполнен для его уборки? Какие производственные мощности имеются для выполнения этих работ и какие мероприятия необходимы для ликвидации дисбаланса между необходимым объемом работ и имеющимися в наличии производственными мощностями? В результате был разработан план экстренной уборки снега, предусматривающий небольшие затраты и некоторое увеличение парка оборудования, его более рациональное распределение, а также мероприятия, обеспечивающие более быструю мобилизацию персонала и технических средств. В работе [25] показано, что уборка снега (льда) может быть значительно ускорена и за счет лучшей разработки маршрутов движения автомобилей, разбрасывающих соль.

Наличие и качество стоянок являются важными характеристиками деятельности служб уличного движения, тем более что проектирование стоянок автомашин может существенно улучшить состояние окружающей среды, в особенности качество воздуха, и понизить уровень шума в районах города с интенсивным движением. Основные вопросы, возникающие при организации такой службы, связаны с территориальным распределением стоянок, их вместимостью и типом, характером обслуживания (т. е. системой оплаты, временем работы и т. д.), строительством необходимых сооружений.

В работе [8] обсуждается имитационная модель для оценки одной из частных программ организации стоянок. Основу этой модели составляет задача линейного программирования, в которой минимизируются общий неудовлетворенный спрос всех пользователей (прибывающие автомашины распределяются по различным стоянкам) при ограничениях на вместимость каждой стоянки и условия удовлетворения всех требований. Подобная модель позволяет определить число пользователей каждой стоянки, для которых заданы конечный пункт назначения, число свободных мест на каждой стоянке и фактический доход государства.

Автомобили, паркующиеся у обочин тротуаров с нарушением существующих правил стоянки, препятствуют движению транспорта и затрудняют работу пожарных бригад, сбор мусора, расчистку улиц. Применительно к условиям Нью-Йорка был прове-

ден системный анализ соблюдения правил стоянки, их нарушений и использования соответствующих санкций в случае необходимости.

В работе [32] представлена простая аналитическая модель, с помощью которой оценивается ухудшение качества уборки мусора с помощью механических метел из-за нарушения правил стоянки как на размеченных, так и на не размеченных улицах. Степень соблюдения правил стоянки определяется величиной коэффициента $C = (N - I)/N$, где N — общее число мест для стоянки на данной стороне квартала, I — число машин, нарушающих правила стоянки. Показано, что подмести 70—75% поверхности обочины тротуара (минимальная градация «степени чистоты») можно только при условии высокого уровня соблюдения правил стоянки (85%). В работе [124] рассматривается модель, в основу которой положено понятие о конечной марковской цепи. Модель позволяет установить связь между системой штрафов за нарушение правил стоянки и ожидаемым уровнем соблюдения этих правил. Кроме того, в работе дается характеристика эффективности детерминированного и вероятностного подходов к осуществлению этих санкций и анализируется класс практически приемлемых методов, обеспечивающих удовлетворительные результаты. Полученные данные используются при составлении графиков дежурств и маршрутов движения полицейских инспекторов, что повышает результативность их работы.

В целях обеспечения безопасности на улицах города при движении транспорта, а также удовлетворения эстетических потребностей населения периодически осуществляется ремонт городских улиц. Можно выделить три основные категории таких ремонтных работ: общестроительные работы, обновление асфальтового покрытия мостовых и аварийный ремонт (ремонт выбоин и заделки отдельных участков). Проблема выбора между этими категориями работ по критерию затраты — эффективность рассмотрена в работах [97, 98]. Для нахождения оптимального распределения ремонтных средств во внимание принимаются ограничения по бюджету капиталовложений и сумме расходов, а также ограничения по производственным мощностям субподрядчиков и мощностям, которыми располагают соответствующие службы.

Поскольку основное назначение улиц города — обеспечение движения транспорта, естественно, что они используются для стоянок и передвижения транспортных средств. При этом возникают серьезные проблемы транспортной перегрузки городских улиц. Решение подобных проблем оказывается возможным с помощью методов исследования операций и путем создания систем регулирования, использующих ЭВМ и функционирующих в реальном масштабе времени. Регулирование на макроуровне осуществляется с помощью управляющих воздействий, которые распростра-

няются на относительно большой район города и характеризуются временными интервалами от минут до часов. Регулирование на микроуровне охватывает ограниченное число перекрестков и осуществляется при секундном отсчете времени.

Задачу транспортной перегрузки можно рассматривать как задачу о распределении: необходимо осуществить такое распределение фаз зеленого света светофоров, чтобы максимизировать некоторые показатели передвижения транспортных средств. Для расчета систем, реализующих такую последовательность фаз, используются модели частично целочисленного линейного программирования, которые, однако, базируются на допущении о том, что улицы сравнительно свободны. Поэтому уменьшение времени задержки у перекрестков, которого можно добиться, используя только управление светофорами, не превышает 10%. Регулирование транспортных потоков, при котором множество дорог и их пересечений рассматривается как ориентированная транспортная сеть с ограниченными пропускными способностями узлов, характеризуется более реальным подходом к решению проблемы и открывает возможности для дополнительных усовершенствований. Детальный обзор теоретических и прикладных исследований в области регулирования уличного движения представлен в работе [42].

4. ДРУГИЕ ВНУТРИГОРОДСКИЕ (МЕСТНЫЕ) СФЕРЫ ОБСЛУЖИВАНИЯ

Естественно, что помимо экстренных и санитарных служб, а также службы уличного движения в ведении органов местного самоуправления находятся и сферы социальных услуг, самой дорогостоящей из которых является образование. Проблеме образования посвящен ряд работ, в которых методы исследования операций используются для решения задачи планирования системы школьного обучения, прогнозирования контингента учащихся, выбора местоположения школ, календарного планирования, распределения учащихся, управления работой школьных столовых, составления маршрутов школьных автобусов и десегрегации (см. гл. II-3).

Система уголовного судопроизводства не является специфически городской и не ограничивается рамками полномочий органов местного самоуправления. Многие проблемы системы уголовного судопроизводства в США обусловлены ее фрагментарным характером. Федеральные, государственные и местные органы судопроизводства, ответственные за соблюдение законов, взаимодействуют с различными прокурорами, учитывают многообразие правил юрисдикции и наличие многочисленных социальных комиссий, которые должны способствовать пересмотру дел и реабилитации подсудимых, а также нести ответственность за досрочное

условное освобождение и условное освобождение на поруки несовершеннолетних. Полная операционная модель судебного процесса по уголовному делу построена в работе [41], а отдельные подсистемы судопроизводства моделируются в работах [59, 60, 125].

При решении проблем сферы социального обеспечения исследователю систем приходится сталкиваться с особыми сложностями, поскольку стоимостные критерии эффективности таких услуг трудно поддаются количественной оценке. В работе [120] обсуждаются различные аспекты функционирования служб по охране детства, составлено дерево целей этих служб и определены переменные, которые должны быть включены в модели соответствующей системы для их успешного применения. Особо подчеркивается, что главным препятствием при исследовании проблем городской сферы обслуживания является отсутствие необходимых данных.

Географическая миграция определенных слоев населения в большом Чикаго исследуется в работе [37]. Полученные результаты не подтверждают широко распространенной гипотезы о «миграции по расчету», согласно которой семьи с низким доходом переселяются в г. Чикаго с целью получить большую материальную помощь. В работе [97] также показано, что преобладающая часть мигрантов, получивших такую помощь, прожила на данном месте по крайней мере пять лет, прежде чем получила соответствующее пособие. В работе [47] показано, что осуществление предлагаемого плана помощи семьям при реализации программы поддержания доходов в г. Нью-Йорке потребует сокращения продолжительности рабочего дня для мужчин — глав семей.

Возможность определения границ политико-административных районов с использованием методов исследования операций и вычислительной техники впервые была рассмотрена в работе [54]; установление границ избирательного округа Бристольского района г. Нью-Йорка с помощью методов линейного программирования и ряда эвристических процедур описано в работе [94]. Применительно к условиям Нью-Йорка было показано, что более рациональная организация избирательных округов дает возможность снизить расходы, связанные с выборами, а реализация (вручную) простого метода расчета, основанного на машинном алгоритме, обеспечивает сокращение расходов более чем на один миллион долларов в год [110, 119].

Составление муниципального бюджета также стало объектом моделирования [28, 44], однако практически без внимания осталась такая важная функция органов местного самоуправления, как организация мест отдыха и развлечений. Работа [18], посвященная прогнозированию потоков туристов к местам отдыха, содержит оценку экономической эффективности организации мест отдыха на лоне природы, а в работе [36] рассматриваются вопросы

оптимальной эффективности использования средств для отдыха и развлечений. Общим недостатком этих работ является отсутствие выводов относительно планирования и организации городских парков. Несмотря на результаты работы [23], многое еще предстоит сделать в таких направлениях, как определение спроса различных слоев городского населения, касающегося сферы отдыха и развлечений, определение желаемого состава и мест расположения различных объектов этой сферы, установление связи между использованием таких объектов и затратами на их эксплуатацию и восстановление, т. е. по целому комплексу проблем планирования и управления, а также организации данной сферы городского обслуживания (в пространственном и временном аспектах).

Решения, связанные с муниципальным зонированием, обычно принимаются при участии заинтересованных политических сил, поэтому методы исследования операций вряд ли найдут здесь применение. Тем не менее в работе [29] предложена имитационная модель изменения предписания о разбиении на зоны г. Питтсбурга. Для проверки всех 125 изменений, которые были предложены в течение последних двух лет, проводились тестовые испытания модели, реализованной на ЭВМ. Рекомендации, которые выдала ЭВМ, в 90% случаев совпали с решениями, принятыми административными органами.

Рассмотрим кратко проблему моделирования развития города. Попытки разработать модели, выявляющие взаимосвязь различных факторов, влияющих на рост и развитие городов, предпринимались неоднократно [39, 52, 64, 87]. Однако, за исключением имитационной модели Форрестера, результаты выполненных работ до сих пор не нашли практического применения. По-видимому, это объясняется тем, что исследователи систем основное внимание уделяют не столько практическому решению проблемы, сколько разработке теоретических основ для такого решения, и в результате все усилия сводятся к обсуждениям, напоминающим схоластические дискуссии средневековья.

5. ОБРАТНАЯ СВЯЗЬ В СИСТЕМЕ УПРАВЛЕНИЯ ГОРОДСКОЙ СФЕРОЙ ОБСЛУЖИВАНИЯ

Более общий подход к исследованию городской сферы обслуживания и присущих ей свойств дается в работах [108, 110], где показано, что предоставление муниципальных услуг целесообразно рассматривать как процесс регулирования с обратной связью. В этом случае при анализе управления городом цели, административные воздействия, возмущения, динамические характеристики системы и информационная обратная связь рассматриваются с иных точек зрения. В указанных работах обращается внимание на то обстоятельство, что на основе претензий населения и объек-

тивных критериев можно установить показатели общественного развития и использовать их для определения степени рассогласования в муниципальной системе управления. В работах [71, 72] обсуждаются возможности осуществления автоматического управления и проведение оценки динамических характеристик систем с помощью анализа претензий, который проводится городскими органами управления.

Вопрос об обратной связи (от жителей города к органам управления) приобретает все большее значение, по мере того как осознается ее важная роль при решении задач социального развития [73, 121]. Так, например, с целью улучшения деятельности правительственных органов для губернатора шт. Пуэрто-Рико была разработана экспериментальная система обратной связи с населением [83, 84], отображающая существующие типы информационных потоков. Прямая связь с населением представляет собой информационный поток от правительства к населению, содержащий соответствующие прогнозные оценки и используемый для установления сопряженной обратной связи, т. е. для уточнения мнений и предложений населения относительно функционирования правительственных органов и степени участия населения в выработке управляющих решений. Сопряженная обратная связь усиливает ту компоненту обратной связи в цепи управления городской сферой обслуживания, которая характеризует претензии, требования и данные опроса населения. Компоненты цепи обратной связи и соответствующие системы исследовались при анализе условий жизни в шт. Массачусетс [82]. Для улучшения обратной связи с населением в различных условиях и на различной экспериментальной базе была разработана система «совместного управления», предусматривающая использование вычислительных машин и средств телесвязи [35, 61, 88]. В настоящее время все более широко осуществляются систематические обследования городского населения, результаты которых служат для оценки деятельности органов местного самоуправления [130], а также функционирования отдельных видов сферы обслуживания [9, 53]. Возможные формы участия населения города в процессе оказания социальных услуг обсуждаются в работе [131], причем привлечение населения рассматривается как средство пересмотра и улучшения реализации социальных программ, а не как социальная реформа.

6. ПРОБЛЕМЫ ПРИМЕНЕНИЯ МЕТОДОВ ИССЛЕДОВАНИЯ ОПЕРАЦИЙ В ГОРОДСКОЙ СФЕРЕ ОБСЛУЖИВАНИЯ

В предыдущих разделах данной главы рассматривалась возможность решения проблем, связанных с функционированием различных сфер городского обслуживания, с помощью методов

исследования операций. Для многих задач оперативного характера использование этих методов является весьма эффективным, однако в случае стратегических задач управления их применение оказывается менее успешным. Результаты, полученные в данной области применения, вызывают особенное разочарование, поскольку свидетельствуют о несостоятельности оптимистических высказываний некоторых исследователей-энтузиастов, считавших, что «городскую проблему» можно легко решить, взяв на вооружение современные методы управления, базирующиеся на методах исследования операций, системном анализе, системах «планирование — программирование — составление сметы» и т. п.

Чтобы понять причины ограниченных масштабов применения методов исследования операций в решении задач стратегического характера, необходимо проанализировать, какие факторы препятствуют внедрению этих методов, и рассмотреть природу систем, в рамках которых функционируют городские системы обслуживания [49, 56, 115]. Некоторые из трудностей, с которыми приходится сталкиваться исследователю систем при решении проблем городской сферы обслуживания, являются типичными и для других областей применения методов исследования операций, однако в условиях города они более очевидны. Существуют трудности, с которыми приходится иметь дело при решении всех задач управления в общественном секторе [120], и некоторые из них специфичны для местных органов власти, причем не только в Соединенных Штатах [129]. Обобщение этих проблем приводит к следующим выводам.

— Некоторая система является подсистемой другой системы. Сложность проблем городской сферы обслуживания затрудняет исследователю систем определение границ проблем, которые, с одной стороны, должны быть достаточно широкими, чтобы задача имела смысл, а с другой — достаточно узкими, чтобы можно было ее решить. Поэтому любое из возможных решений всегда уязвимо для критики, поскольку оно не учитывает интересов более крупной системы, в которую входит данная.

— Системный анализ некоторой проблемы является своего рода политическим актом, так как выбор конкретной области для исследования в какой-то мере свидетельствует о том, что в данной области имеются определенные упущения. Решение о проведении операционного анализа неизбежно содержит угрозу, как минимум, для одного бюрократического органа и доставляет удовольствие по крайней мере одному из них. Следовательно, намереваясь сделать что-либо существенное, исследователь систем сознательно подвергает себя перекрестным бюрократическим нападкам.

— Уверенно оперируя отвлеченными понятиями, связанными с распределением ресурсов и календарным планированием, исследователь систем становится беспомощным при оценке политиче-

ской обстановки исследуемых процессов, а именно: какая группа, чем, когда и где занимается, т. е. какую смысловую нагрузку несут понятия распределение и календарное планирование применительно к общественным ресурсам.

— Если решения относительно места посадки на Луне обосновываются исключительно техническими соображениями, то при решении таких вопросов, как размещение пункта уничтожения отходов или аэропорта, анализ технических аспектов, базирующийся на методах исследования операций, играет второстепенную роль. Это объясняется тем, что общественность не волнует вопрос, где совершат посадку космонавты — в Море Спокойствия или на Горах Волнения, однако населению совсем не безразлично, где будет размещен аэропорт.

Алгоритм, в соответствии с которым окажется, что «наилучшее» место для наркологической лечебницы находится совсем рядом с кварталом домов членов муниципального совета, на самом деле будет свидетельствовать лишь о том, что алгоритм неконструктивен, и рекомендации исследователя систем в лучшем случае будут оставлены без внимания.

— Обычно исследователи систем сетуют на то, что процесс принятия решений, относящихся к общественному сектору, носит излишне политический характер и недостаточно рационален. Это свидетельствует лишь о том, что такой специалист не представляет себе цели, которую преследует лицо, принимающее решение. Это лицо при составлении неявной целевой функции обычно учитывает показатели, о которых исследователь либо не подозревает, либо влияние которых он не может правильно учесть.

— Исследователь систем является своего рода носителем определенных изменений: его рекомендации никогда не предполагают сохранения *статус-кво*. Но изменение в сфере городского управления, как правило, означает изменение установленного порядка взаимоотношений, характера работы и полномочий, причем эти изменения легко рекомендовать, но трудно осуществить. Кроме того, роль носителя перемен в рассматриваемой сфере деятельности сопряжена с риском. Так, например, древнегреческому философу Сократу дорого обошлись попытки анализировать процессы человеческого мышления и сознания — он преследовался афинскими властями, был привлечен к суду и казнен.

— Традиционное представление о процессах принятия решений, сложившееся у исследователей систем, не соответствует тому, что на самом деле имеет место при решении проблем городской сферы обслуживания. Исследователь систем считает, что решение должно приниматься на основе тех выводов относительно последствий реализации возможных вариантов решения, которые он предоставляет в распоряжение лицу, принимающему решение. В действительности принятие административных решений редко

сводится к однозначному выбору одного из двух вариантов. В большинстве случаев это длительный и расплывчатый процесс с элементами неопределенности и плюрализма. Поэтому, если на основе результатов аналитического исследования и принято конкретное «решение», специалист (исследователь систем) должен стать активной силой, продвигающей резолюцию через невидимый фронт, а это для него весьма непривычная и малопривытная роль. В работе [62] красноречиво описывается процесс принятия решения в государственном секторе.

Важно помнить, что принятие решения . . . — это процесс. Такие выражения, как «процесс принятия решения» или «процесс выработки политики», не просто заклипания; они характеризуют непрерывную последовательность решений, важных и второстепенных, которые составляют в итоге неразрывную цепь выработки политики и административных действий . . . властей. Динамический характер политических процессов делает работу консультанта исключительно трудной. Это означает, что в данном случае не существует обычного порядка, когда консультант, изложив свою точку зрения или объяснив результаты исследования, может удалиться с арены действий, будучи уверенным, что его рекомендации не останутся без внимания, так как различные второстепенные проблемы, а также вопросы конкуренции требуют определенного внимания со стороны лица, принимающего решения. Поэтому, даже если какая-нибудь проблема считается уже решенной, не исключено, что спустя неделю она вновь будет включена в рассмотрение. Консультант может столкнуться с трудной задачей: сначала добиться, чтобы его точка зрения была достаточно внимательно выслушана, а затем в течение весьма долгого времени постоянно направлять свои усилия на то, чтобы сфокусировать внимание соответствующих лиц на выработанных рекомендациях, с тем чтобы обеспечить действия, необходимые для их реализации. Таким образом, последовательная смена этапов деятельности является существенной частью эффективного проведения в жизнь исследований, связанных с решением стратегических задач. Из этого следует, что для получения благоприятного решения относительно предлагаемых мер и их эффективной реализации совершенно недостаточно просто представить соответствующие рекомендации на верхний уровень управления.

Рассмотрим ситуацию, когда лицо, принимающее решение и занимающее высокий административный пост, благосклонно относится к рекомендациям группы консультантов и принимает решение «стратегического» характера, направленное на осуществление этих рекомендаций. Если подчиненные не проводят это решение в жизнь достаточно эффективно, то замысел в целом может оказаться неосуществленным. Понимание основного содержания решения более высоким уровнем руководства может служить важным фактором, обеспечивающим согласованность и энтузиазм тех звеньев, которые должны осуществлять это решение, принимая при этом несметное число других решений, которые могут определить успех или неудачу практической реализации исходного решения.

— Неуместность использования новейших технических средств в тех случаях, когда в этом нет необходимости, очевидна для политических лидеров и руководителей государственного аппарата, но отнюдь не для большого числа исследователей систем, которые, будучи вооружены математическими методами и программами для вычислительных машин, старательно ищут возможности для их

применения в городской сфере обслуживания. В этой связи уместно вспомнить следующее высказывание Рассела Аккоффа: «Я предпочел бы быть прооперированным квалифицированным хирургом, даже если он имеет в своем распоряжении всего лишь ржавый перочинный нож, чем коммуволяжером, который рекламирует набор прекрасных хирургических инструментов и, естественно, может воспользоваться любым из них».

— Оптимизация — классическая концепция исследования операций — при решении проблем городской сферы обслуживания приобретает различные смысловые оттенки. К тому же часто возникают ситуации, когда кажется, что допустимого решения вообще не существует. В таких случаях оптимизации обычно заключается в попытке определить, какие ограничения, наименее важные в политическом аспекте, можно волевым порядком отбросить, с тем чтобы получить некоторое осуществимое и, следовательно, оптимальное решение.

— Проблемы, не осознанные общественностью, являются проблемами, для решения которых нет должных стимулов. Понимание таких проблем интеллектуальной элитой, например исследователями систем, недостаточно для побуждения к действию. Так, общественность может мириться с тем, что не получает непосредственных благ от долгосрочных политических решений, которые, однако, предотвращают возникновение проблем, и не утруждает себя анализом возможных последствий таких решений, хотя цена такого «бездействия» — это недостатки, за которые придется расплачиваться в будущем.

— Модели для прогнозирования изменения процессов, происходящих в жизни города, разработаны явно недостаточно. Это объясняется тем, что трудно предсказать влияние возмущений, действующих на систему, а также определить вид и величину упреждающих корректирующих воздействий, которые следует осуществить для нейтрализации этих возмущений.

— На функционирование сложных городских систем скрытно и противоречиво влияют многие факторы, вследствие чего конечный результат некоторого воздействия оказывается прямо противоположным ожидаемому.

Пример. Рассмотрим сложную бюрократическую систему контроля и балансов, разработанную для обеспечения приемлемого уровня цен при закупках, осуществляемых для нужд города. В рамках такой системы разработаны методы, исключающие взятки и коррупцию при заключении контрактов на поставку сырья, материалов, оборудования, а также на обслуживание. Однако в некоторых городах следствием таких мер является чрезмерная задержка предложений о заключении контрактов и оплаты счетов, в результате чего деловые отношения с городом прерывают многие потенциальные поставщики, в том числе и те, кто действительно должен повысить цены, чтобы компенсировать дополнительные расходы, связанные с поставками для города. Таким образом, стратегия, рассчитанная на усиление конкурен-

ции и снижение цен на товары, приводит к обратному результату — уменьшению конкуренции и росту цен на товары [110].

— При решении городских проблем этап аналитических исследований в общем процессе занимает небольшую часть, в то время как основными этапами являются определение проблемы, сбор данных и реализация результатов работы. В связи с этим исследователям систем следует включаться в работу на более ранних стадиях, связанных с определением проблем и изучением тонкостей существующих ограничений и целей, и не уклоняться от участия в работе на более поздних стадиях — при внедрении. К сожалению, в настоящее время наблюдается прямо противоположное, что, по-видимому, объясняется непониманием важности участия исследователя систем в проведении работ подобного рода.

— Исследователю систем, заинтересованному в реализации своих предложений (как это и должно быть), следует знать следующие пять классических выражений, одно из которых непременно будет использовано работниками бюрократического аппарата в качестве «веского» аргумента нецелесообразности рекомендуемых изменений:

- Дело в том, что мы действовали аналогичным образом.
- Видите ли, в данном случае это неприемлемо.
- К сожалению, данные, полученные еще в 1948 г., свидетельствуют об обратном.
- Нам представляется, что нечто подобное мы уже делаем.
- Мне кажется, что этим можно было бы заняться, но даже в случае положительных результатов вряд ли что-нибудь изменится!

7 НОВЫЕ НАПРАВЛЕНИЯ В ИСПОЛЬЗОВАНИИ МЕТОДОВ ИССЛЕДОВАНИЯ ОПЕРАЦИЙ ДЛЯ РЕШЕНИЯ ПРОБЛЕМ ГОРОДСКОЙ СФЕРЫ ОБСЛУЖИВАНИЯ

При операционном анализе городской сферы обслуживания главное внимание уделяется проблемам и программам. Однако улучшение условий жизни в городе и совершенствование управления городской сферы услуг невозможно без структурных или социальных изменений [14, 43, 106]. Так, разработка оптимальных маршрутов для полицейских патрульных машин выглядит до некоторой степени неоправданной, поскольку из 31 тыс. полицейских департамента полиции г. Нью-Йорка одновременно патрулирование осуществляет только 1 тыс. полицейских [15]. Аналогичная ситуация возникает и при рассмотрении проблемы улучшения процесса сбора мусора в городе: стоит ли тратить столько усилий на разработку оптимизационных моделей календарного планирования, маршрутизации и размещения пунктов уничтожения отходов, если затраты частных фирм на аналогичные работы без использования таких методов в три раза мень-

ше, чем в муниципальных службах [114]. Монопольный характер некоторых видов муниципальных услуг уже сам по себе может служить объяснением их низкой эффективности [112]; очевидный, но чреватый последствиями политического характера выход из существующего положения — это создать конкуренцию в городской сфере обслуживания [114, 116].

Следует отметить, что не всегда методы исследования операций должным образом используются в системе городской сферы обслуживания. Если эффективность этих методов в области планирования семьи не вызывает сомнения, то совсем иначе обстоит дело в сфере сбыта. Результаты использования методов, принятые на вооружение современной системой сбыта, оказались удручающими: увлечение кормлением грудных детей из бутылочек наблюдается даже в слаборазвитых странах с высокими коэффициентами рождаемости. Если учесть, что кормление грудью подавляет овуляцию, а при кормлении из бутылочек этот естественный контрацептивный процесс нарушается, то вывод напрашивается сам собой: решение одной проблемы породило другую. Установлено, что продление периода кормления грудью в странах с высоким уровнем рождаемости могло бы снизить коэффициент рождаемости почти на 20% [128].

Методами исследования операций решены многие классы задач, в том числе задачи календарного планирования и распределения работ, однако когда стимулирование работников строго ограничено [118], то, по-видимому, бессмысленно уточнять параметры модели с целью отразить уменьшение производительности в этой сфере услуг.

Истинные характеристики эффективности государственной сферы обслуживания стали находить отражение в соответствующих аналитических исследованиях только в последнее время [53]. Между тем именно такие характеристики должны служить отправной точкой подобных исследований. Например, чтобы оценить эффективность полицейских служб, недостаточно пользоваться такими показателями, как время отклика системы, а следует проанализировать и динамику правонарушений. Аналогичная картина наблюдается и тогда, когда органы городского управления с гордостью сообщают о количестве собранного мусора и обычно не упоминают ни слова о том, сколько мусора осталось неубранным. Лишь совсем недавно эта проблема была сформулирована правильным образом, хотя, может быть, такая формулировка звучит несколько непривычно: насколько чист наш город? [9]

Государственные органы обычно указывают на потребности в услугах городской сферы обслуживания и запрашивают ресурсы для удовлетворения этих потребностей. Надо надеяться, что в целях эффективного использования этих ресурсов они все шире будут применять методы исследования операций. Поскольку, однако,

при таком подходе практически отсутствуют стимулы к снижению потребностей в услугах, то в данном случае открываются широкие возможности для исследований проблем использования рычагов ценообразования или рыночного механизма для изменения структуры и снижения уровня потребностей [95] независимо от того, относится ли это к общественному транспорту в часы пик или к потреблению воды в периоды пиковых нагрузок. Например, реализация предложения об использовании штатных медсестер для установления несрочных требований на обслуживание бригадами скорой помощи привела к сокращению рейсов бригад на 10% [108]; две медсестры оказывают такое же влияние на снижение среднего времени отклика, как и десять дополнительных бригад скорой помощи, но только при гораздо меньших затратах [80].

Одной из центральных проблем, связанных с качеством обслуживания населения городов, является проблема структуры органов местного самоуправления. Во многих городах естественные географические границы для организации обслуживания не соответствуют законодательно установленным границам полномочий органов, ответственных за это обслуживание, что, естественно, затрудняет четкое распределение соответствующих функций: какой уровень управления должен обеспечивать данный вид обслуживания, кого обслуживать, как должно оплачиваться это обслуживание. Сюда входят также вопросы распределения доходов, регионального деления, управления округами. Функции по социальному обеспечению лучше всего централизовать на уровне общегосударственного руководства, в то время как функции по транспортному обслуживанию, предотвращению загрязнений, водоснабжению и уничтожению отходов носят чисто региональный характер. Однако такие функции, как поддержание чистоты улиц, их ремонт, а также сбор мусора, имеют главным образом локальный характер и могут быть сосредоточены в руках органов местного самоуправления. Следовательно, чтобы дополнить или заменить устаревшие административные службы, сохранившиеся в некоторых городах США еще с колониальных времен, необходима новая иерархическая структура управления — региональной и местной.

В решении всех этих сложных проблем, включая социальные, методы исследования операций призваны сыграть важную роль. Помимо всего прочего, эти методы необходимы для разработки адаптивных систем, которые должны обеспечить сохранение наших городов, основной функцией которых является обеспечение благ человека. Чтобы решить поставленные задачи, необходим союз исследования операций с другими научными дисциплинами, который явился бы основой для развития науки века, или науки о системах [2].

ЛИТЕРАТУРА

1. Abert J. G., Alter H., Bernheisel J. F., The Economics of Resource Recovery from Municipal Solid Waste, *Science*, 183, 1052—1058 (March 15, 1974).
2. Ackoff R., Science in the Systems Age: Beyond IE, OR, and MS, *Operations Res.*, 21, No. 3, 661—671 (1973).
3. Altman S. M., Beltrami E. J., Rappaport S. S., Schoepfle G. K., Nonlinear Programming Model of Crew Assignments for Household Refuse Collection, *IEEE Trans. Systems, Man, and Cybernetics*, SMC-1, No. 3, 289—291 (1971).
4. American Association for the Advancement of Science, Air Conservation, AAAS, 1965.
5. Anderson R. T. (Ed.), Waste Management: Generation and Disposal of Solid, Liquid, and Gaseous Wastes in the New York Region, Regional Plan Association, New York, 1968.
6. Bell C. E., Allen D., Optimal Planning of an Emergency Ambulance Service, *Socio-Econ. Plan. Sci.*, 3, 95—101 (1969).
7. Beltrami E. J., Bodin L. D., Networks and Vehicle Routing for Municipal Waste Collection, *J. Networks*, 4, 65—94 (1973).
8. Bennett J. C., Ellis R. H., Rassam P. R., A Guide to Parking Systems Analysis, U.S. Department of Transportation/Federal Highway Administration No. 5001—00049, October 1972.
9. Blair L. H., Schwartz A. I., How Clean is Our City?, Urban Institute, Washington, D.C., 1972.
10. Blum E. H., Deployment Research of the New York City Fire Project, *Urban Anal.*, 1, 63—94 (1972).
11. Blumstein A., Larson R. Models of a Total Criminal Justice System, *Operations Res.*, 17, No. 2, 199—232 (1969).
12. Bodin L. D., Tucker A. G., Altman S. M., Beltrami E. J., Heuristics in Manpower Scheduling, Urban Science and Engineering Technical Report No. 71—7, State University of New York at Stony Brook, March 1971.
13. Bodner R. M., Cassel E. A., Andros P. J., Optimal Routing of Refuse Collection Vehicles, *J. Sanitary Eng. Div., ASCE*, 96, No. SA4, 893—904 (1970).
14. Brewer G. D., The Politician, the Bureaucrat, and the Consultant, Basic Books, N.Y., 1975.
15. Burnham D., 1,000 of the City's 31,000 Policemen Are Out in Radio Cars At Any One Time, *New York Times*, p. 1 (February 18, 1974).
16. Carter G. M., Chaiken J. M., Ignall E., Response Areas for Two Emergency Units, *Operations Res.*, 20, No. 30 571—594 (1972).
17. Carter G. M., Ignall E. J., A Simulation Model of Fire Department Operations: Design and Preliminary Results, *IEEE Trans. Systems Sci. and Cybernetics* SSC-6, No. 40, 282—293 (1970).
18. Cesario F. J., Operations Research in Outdoor Recreation, *J. Leisure Res.*, 1, 33—51 (1969).
19. Cesario F. J., Optimal Road Pricing of Air Pollution Control, *The Regional Sci. Assoc. Pap.*, 30, 97—110 (1973).
20. Chaiken J. M., Number of Emergency Units Busy at Alarms Which Require Multiple Servers, New York City — Rand Institute, R-531-NYC/HUD, March 1971.
21. Chaiken J. M., Larson R. C., Methods for Allocating Urban Emergency Units: A Survey, *Management Sci.*, 19, No. 4; Part II, 110—130 (1972).
22. Chan K. (Ed.), Urban Dynamics: Extensions and Reflections, *IEEE Trans. Systems, Man, and Cybernetics*, SMC-2, No. 2, 121—220 (1972).
23. Cicchetti C. J., Some Economic Issues in Planning Urban Recreation Facilities, *Land Econ.*, 47, 14—23 (1971).

24. Clogh D. J., Environmental Pollution, in: Operational Research 72, M. Ross (Ed.), North-Holland Publishing Company, Amsterdam, 1973.
25. Cook T. M., Alpin B. S., Snow and Ice Removal in a Urban Environment, *Management Science*, 23, No. 3, 227—234 (1976).
26. Coulter J. B., The Role of Government and Systems Analysis in Waste Management, in The Challenge to Systems Analysis: Public Policy and Social Change, G. J. Kelleher (Ed.), Wiley, N.Y., 1970.
27. Coyle R. G., Martin M. J. C. Case Study: The Cost Minimization of Refuse Collection Operations, *Operations Res. Quart.*, 20, Special Conference Issue, 43—56 (1969).
28. Crecine J. P., A Computer Simulation Model of Municipal Budgeting, *Management Sci.*, 13, No. 11, 786—815 (1967).
29. Davis O., Rueter F., A Simulation of Municipal Zoning Decisions, *Management Sci.*, 19, No. 4, P39—P78 (1972).
30. De Neufville R., Cost-Effectiveness Analysis of Civil Engineering Systems: New York City's Primery Water Supply, *Operations Res.*, 18, No. 5, 785—804 (1970).
31. De Neufville R., Schaake J., Stafford J. H., Systems Analysis of Water Distribution Networks, *J. Sanitary Eng. Div., ASCE*, 97, No. SA6; 835—842 (1971).
32. De Stefano L. T., Levis A. H., A Model for the Effect of Illegally Parked Cars on Street Cleaning, *J. Environ. Syst.*, 2, 109—129 (1972).
33. Eisenbud M., Environmental Protection in the City of New York, *Science*, 170, 706—712 (November 13, 1970).
34. Ellis H. M., Keeney R. L., A Rational Approach for Government Decisions Concerning Air Pollutin, in: Analysis Public Systems, A. Drake, R. Keeney, P. Morse (Eds.), M. I. T. Press, Cambridge, Massachusetts, 1972.
35. Etzioni A., Minerva and Electronic Town Hall, *Policy Sci*, 457—474 (1972).
36. Fisher A., Krutilin J. V., Determination of Optimal Capacity of Resource-Based Recreational Facilities, *Natural Resource J.*, 12, 417—444 (1972).
37. Fisher I. N., Purnell S. W., The Connection Between Migration and Welfare Dependency in the Chicago Metropolitan Area, Rand Corp. Report R-1388-IIISP, Santa Monica, California, 1973.
38. Fitzsimmons J., A Methodology for Emergency Ambulance Deployment, *Management Sci.*, 19, No. 6, 627—636 (1973).
39. Forrester J. W., Urban Dynamics, M. I. T. Press, Cambridge, Massachusetts, 1968.
40. Freeman A. M., Haveman R. H., Residual Charges for Pollution Control: A Policy Evaluation, *Science*, 177, 322—329 (July 28, 1972).
41. Freeman A. M., Haveman R. H., Kneese A. V., The Economics of Environmental Policy, Wiley, N.Y., 1973.
42. Gaxis D. C., Traffic Flow and Control: Theory and Applications, *Am. Scientist*, 60, 414—424 (1972).
43. Gordon D. R., City Limits: Barriers to Change in Urban Government, Charterhouse, N.Y., 1973.
44. Governmental Problem Solving: A Computer Simulation of Municipal Budgeting, Rand McNally, Chicago, 1967.
45. Green J. A., Predicting Future Quantities of Refuse, Local Government Operational Research Unit Report No. T. 20, August 1969.
46. Green J. A., Nice R. W., The Economics of Transfer Stations and Vehicle Relays as Methods of Refuse Collection, Local Government Operational Research Unit. Report No. T. 19, July 1969.
47. Greenberg D. H., The Potential Impact of the Family Assistance Plan on New York City, New York City — Rand Institute Rept. R-658-NYC, New York, 1971.
48. Haines Y. Y., Pollution and Ecology, in Systems Approach and the City,

- M. D. Mesarovic and A. Reisman (Eds.), North-Holland Publishing Company, Amsterdam, 1972.
49. Halbrecht H., Savas E. S., Hoffman G., Ayres H. F., Radnor M., Edelman F., Through a Glass Darkly, *Interfaces*, 2, No. 4, 1—17 (1972).
 50. Hall W. K., Management Science Approaches to the Determination of Urban Ambulance Requirements, *Socio-Econ. Plan. Sci.*, 5, 491—499 (1971).
 51. Hall W. K. The Application of Multifunction Stochastic Service Systems in Allocating Ambulances to an Urban Area, *Operations Res.*, 20, No. 3, 558—570 (1972).
 52. Harris B. (Ed.). Urban Development Models: New Tools for Planning, *J. Am. Inst. Planners*, 31, No. 2 (1965).
 53. Hatry H. P., Measuring the Quality of Public Services in: Improving the Quality of Urban Management, W. Hawley and D. Rogers (Eds.), Sage Publications, Inc., Beverly Hills, California, 1974.
 54. Hess S. W., Weaver J. B., Siegfeldt H. J., Whelan J. N., Zitlau P. A., Nonpartisan Political Redistricting by Computer, *Operations Res.*, 13, No. 6, 998—1006 (1965).
 55. Hoey J. M., Planning for an Effective Hospital Administered Emergency Ambulance Service in the City of Boston, M.I.T. Operations Research Center Report No. RR-01-73, September 1973.
 56. Hoos I. R., Systems Analysis in Public Policy: A Critique, University of California Press, Berkeley, Calif., 1972.
 57. Ignall E., Kolesar P., Walker W., Linear Programming Models of Grew Assignments for Refuse Collection, *IEEE Trans. Systems, Man, and Cybernetics*, SMC-2, No. 5, 664—666 (1972).
 58. Jarvis J. P., Optimal Dispatch Policies for Urban Server Systems, M.I.T. Operations Research Center Report No. TR-02-73, September 1973.
 59. Jennings J. B., Quantitative Models of Criminal Courts, Rand Corp. Report P-4641, Santa Monica, California, 1971.
 60. Jennings J. B., The Flow of Arrested Adult Defendants Through the Manhattan Criminal Court in 1968 and 1969, New York City-Rand Institute Report R-638-NYC, New York, 1971.
 61. Johnson N., Ward E., Citizen Information Systems: Using Technology to Extend the Dialogue Between Citizens and Their Government, *Management Sci.*, 19, No. 4, P21-P36 (1972).
 62. Kash D. E., Research and Development of the University, *Science*, 160, 1313—1318 (June 21, 1968).
 63. Keeney R. L., A Method for Districting Among Facilities. *Operations Res.*, 20, No. 4, 613—618 (1972).
 64. Kilbridge M. D., O'Block R. P., Teplitz P. V., A Conceptual Framework for Urban Planning Models, *Managements Sci.*, 15, No. 6, B246—B266 (1969).
 65. Kneese A. V., Bower B. T., Managing Water Quality: Economics, Technology Institutions, Johns Hopkins Press, Baltimore, Maryland, 1968.
 66. Kohn R. E., Application of Linear Programming to a Controversy on Air Pollution Control, *Management Sci.*, 17, No. 10, B609-B621 (1971).
 67. Kolesar P., A Model for Predicting Average Fire Engine Travel Times, *Operations Res.*, 23, No. 4, 603—613 (1975).
 68. Kolesar P., Blum E. H., Square Root Laws for Fire Engine Response Distances, *Management Sci.*, 19, No. 12, 1368—1378 (1973).
 69. Kolesar P., Walker W. E., An Algorithm for the Dynamic Relocation of Fire Companies, *Operations Res.*, 22, No. 2, 249—274 (1974).
 70. Kolesar P., Walker W. E., Hauser J., Determining the Relation Between Fire Engine Travel Times and Travel Distances in New York, *Operations Res.*, 23, No. 4, 614—627 (1975).

71. Krendel E. S., A Case Study of Citizen Complaints as Social Indicators, *IEEE, Trans. System. Sci., and Cybernetics*, SSC-6, No. 4, 265—272 (1970).
72. Krendel E. S., Social Indicators and Urban Dynamics, *Socio-Econ. Plan. Sci.*, 5, 387—393 (1971).
73. Krieger M. H., Social Reporting for a City: A Perspective and Some Problems, New York City-Rand Institute, P-4651 (1971).
74. Larsen R. I., A New Mathematical Model of Air Pollutant Concentration Averaging Time and Frequency, *J. Air. Pollution Control Assoc.*, 19, No. 1 (1969).
75. Larson R. C., Improving the Effectiveness of New York City's 911, in: Analysis of Public Systems, A. Drake, R. Keeney, and P. Morse (Eds.), M.I.T. Press, Cambridge, Massachusetts, 1972.
76. Larson R. C., Urban Police Petrol Analysis, M.I.T. Press, Cambridge, Massachusetts, 1972.
77. Larson R. C., Illustrative Police Sector Redesign in District 4 in Boston, M.I.T. Operations Research Center Report No. PP-01-73, November 1973.
78. Larson R. C., Approximating the Performance of Urban Emergency Service Systems, *Operations Res.*, 23, No. 5; 845—866 (1975).
79. Larson R. C., Stevenson K. A., On Insensitivities in Urban Redistricting and Facility Location, *Operations Res.*, 20, No. 3, 595—612 (1972).
80. Lee V., Nurses Screering 911 Calls to Save Ambulance Runs, New York Daily News, January 2, 1971.
81. Liebman J. C., Systems Approaches to Solid Waste Collection Problems, in: Systems Approach to Environmental Pollution, G. K. Chacko (Ed.), Operations Research Society of America, Arlington, Va., 1972.
82. Little J. D. C., Sheridan T. B., Stevens C. H., Tropp P., Citizen Feedback Components and Systems, M.I.T. Operations Research Center Report No. 76, June 1972.
83. Little J. D. C., Stevens C. H., Tropp P., Citizen Feedback Systems: The Puerto Rico Model, *Nat. Civic. Rev.*, 60, 191—203 (1971).
84. Little J. D. C., Gusdorf N., Puerto Rico's Citizen Feedback System, M.I.T. Operations Research Center Report No. 59, April 1971.
85. Loucks D. P., Lynn W. R., Probabilistic Models for Predicting Stream Quality, *Water Resources Res.*, 2, No. 3 (1966).
86. Loucks D. P., ReVelle C. S., Lynn W. R., Linear Programming Models for Programming Models for Water Pollution Control, *Management Sci.*, 14, No. 4, B166-B181 (1967).
87. Lowry I. S., A Model of Metropolis, Rand Corp. Report RM-4035-RC, Santa Monica, Calif., 1964.
88. Lucas W. A., Yin R. K., Serving Local Needs with Telecommunications, Rand Corp. Report R-1345-MF, Santa Monica, Calif., 1973.
89. Management and Behavioral Science Center, A Systems Approach to the Problems of Solid Waste and Litter, Wharton School of Finance and Commerce, University of Pennsylvania, September 1971.
90. Marks D. H., Operations Research and Water Quality Management, *Urban Anal.*, 1, No. 1, 29—44 (1972).
91. Marks D. H., Models in Water Resources, in: A Guide to Models in Governmental Planning and Operations, S. I. Gass and R. L. Sisson (Eds.), Sauger Books, Potomac, Maryland, 1975.
92. Marks D. H., Liebman J. C., Location Models: Solid Waste Collection Example, *J. Urban Plan and Dev., ASCE*, 97, No. UP1, 15—30 (1971).
93. McEwen T., Project Director, Allocation of Patrol Manpower Resources in the Saint Louis Police Department, I, II (1966).
94. Mills G., The Determination of Local Government Electoral Boundaries, *Operational Res. Quart.*, 18, No. 3, 243—255 (1967).
95. Mushkin S. (Ed), Public Prices for Public Products, Urban Institute, Washington, D.C., 1972.

96. Naden R. A., Leeds J. V., Pfeiffer P. E., Use of the Reciprocity Theorem and Mapping Theorems in Locating Anomalous Air Pollution Sources, *IEEE Trans. Systems, Man, and Cybernetics*, SMC-3, No. 2, 147—154 (1973).
97. Naphitali L. M., Maintenance Planning System for the Department of Highways, City of New York, Office of the Mayor, November 16, 1971.
98. Parker C., Renold J., Value for Money in Highway Maintenance: Report of a Feasibility Study, Local Government Operational Research Unit Report No. C98, July 1971.
99. Parvin M., Grammas G. W., Optimization Models for Environmental Pollution Control: A Synthesis, *Journal of Environmental Economics and Management*, 3, No. 2, 113—128 (1976).
100. Parvin M., Technological Abatement vs. Locational Adjustment: A Time-Space Dilemma, in: M. Chatterji and P. Van Rompuy (Eds.), *Energy, Regional Science and Public Policy*, Springer-Verlag, New York, 1976.
101. Quon J. E., Tanaka M., Charnes A., Refuse Quantities and the Frequency of Service, *J. Sanitary Eng. Div., ASCE*, 94, No. SA2, 403—420 (1968).
102. Rabow G., Cost of Pollution, *IEEE Trans. Systems, Man, and Cybernetics*, SMC-3, No. 3, 275—276 (1973).
103. ReVelle C. S., Joeres E., Kirby W., The Linear Decision Rule in Reservoir Management and Design: 1. Development of the Stochastic Model, *Water Resources Res.*, 5, No. 4, 767—777 (1969).
104. ReVelle C. S., Kirby W., The Linear Decision Rule in Reservoir Management and Design: 2. Performance Optimization, *Water Resources Res.*, 6, No. 4, 1033—1044 (1970).
105. ReVelle C. S., Marks D. H., Liebman J. C., A Analysis of Private and Public Sector Location Models, *Management Sci.*, 16, No. 11, 692—707 (1970).
106. Rosenloom R. S., Russel J. R., New Tools for Urban Management, Harvard University, Boston, 1971.
107. Savas E. S., Computers in Urban Air Pollution Control Systems, *Socio-Econ. Plan. Sci.*, 1, 157—183 (1967).
108. Savas E. S., City Hall and Cybernetics, in *Cybernetics and the Management of Large Systems*, E. M. Dewan (Ed.), Spartan Books, N.Y., 1969.
109. Savas E. S., Simulation and Cost-Effectiveness Analysis of New York's Emergency Ambulance Service, *Management Sci.*, 15, No. 12, B608-B627 (1969).
110. Savas E. S., Cybernetics in City Hall, *Science*, 168, 1066—1071 (May 29, 1970).
111. Savas E. S., A Computerized System for Forming Efficient Election Districts, *Operations Res.*, 19, No. 1, 135—155 (1971).
112. Savas E. S., Municipal Monopoly, *Harpers Magazine*, 55—60 (December 1971).
113. Savas E. S., The Political Properties of Crystalline H₂O: Planning for Snow Emergencies in New York, *Management Sci.*, 20, No. 2, 137—145 (1973).
114. Savas E. S., Municipal Monopolies vs. Completion in Delivering Urban Services, in *Improving the Quality of Urban Management*, W. Hawley and D. Rogers (Eds.), Sage Publications, Inc., Beverly Hills, California, 1974.
115. Savas E. S., New Directions for Urban Analysis, *Interfaces*, 6, No. 1 (1975).
116. Savas E. S., The Organization and Efficiency of Solid Waste Collection, Lexington Books, Lexington, Massachusetts, 1977.
117. Savas E. S., Berenyi J., Systems Analysis of Parking Regulations, Violations and Enforcement Activities, Office of the Mayor, New York, 1971.
118. Savas E. S., Ginsburg S., The Civil Service: A Meritless System? *The Public Interest*, No. 32, 70—85 (1973).

119. Savas E. S., Lipton H., Burkholz L., Implementation of an OR Approach for Forming Efficient Election Districts, *Operations Res.*, 20, No. 1, 46—48 (1972).
120. Splindler A., Social and Rehabilitation Services: A Challenge to Operations Research, *Operations Res.*, 18, No. 6, 1112—1124 (1970).
121. Stevens C. H., Science, Government and Citizen Feedback, *Operations Res.*, 18, No. 4, 577—591 (1970).
122. Stevenson K. A., Emergency Ambulance Transportation, in Analysis of Public Systems, A. Drake, R. Keeney, P. Morse (Eds.), M.I.T. Press, Cambridge, Massachusetts, 1972.
123. Swoveland C., Uyeno I., Vertinsky I., Vickson R., Ambulance Location: A Probabilistic Enumeration Approach, *Management Sci.*, 20, No. 4, Part II, 686—698 (1973).
124. Tartarone S., Levis A. H., Enforcement Policies for Parking Regulations, *IEEE Trans. Systems, Man, and Cybernetics*, SMC-3, No. 6, 604—610 (1973).
125. Taylor J. G., Navarro J. A., Cohen R. H., Simulation Applied to a Court System, *IEEE Trans. on Systems Sci., and Cybernetics*, SEC-4, No. 4, 376—379 (1968).
126. Toregas C., Swain R., ReVelle C., Bergman L., The Location of Emergency Service Facilities, *Operations Res.*, 19, No. 6, 1363—1373 (1971).
127. Urban G. L., A Model for Managing a Family Planning System, *Operations Res.*, 22, No. 2, 205—233 (1974).
128. Wade N., Bottle-Feeding: Adverse Effects of a Western Technology, *Science*, 184, 45—48 (1974).
129. Ward R. A., Operations Research in British Local Government, in: Proceedings of the Fifth International Conference on Operational Research, J. Lawrence (Ed.), Tavistock Publications, Ltd., London, 1970.
130. Webb K., Hatry H. P., Obtaining Citizen Feedback, The Urban Institute, Washington, D.C., 1973.
131. Jin R. K., Lucas W. A., Szanton P. L., Spindlen J. A., Citizen Organizations: Increasing Client Control Over Services, Rand Corp. Report R-1196-HEW, Santa Monica, Calif., 1973.

II-2. Система здравоохранения

*Р. Феттер*¹⁾

ВВЕДЕНИЕ

Методы исследований операций используются для составления как микро-, так и макромоделей системы здравоохранения. Первоначально эти методы применялись в основном для решения частных проблем, таких, как планирование амбулаторного медицинского обслуживания, выбор соответствующего метода управления запасами материалов для лечебных учреждений (белья, медикаментов и т. д.), управление запасами консервированной крови, использование ресурсов в стационарах, относительная эффективность различных способов оказания медицинской помощи.

В последнее время методы исследований операций и системный анализ стали использоваться в системе здравоохранения для решения более сложных проблем, например проектирования и эксплуатации больниц, планирования средств по оказанию медицинской помощи, распределения трудовых ресурсов, создания информационных систем, а также осуществления контроля качества медицинского обслуживания.

В настоящей главе основное внимание уделяется применению методов исследования операций для решения проблем управления процессом оказания медицинской помощи, однако сам процесс при этом не исследуется. Именно этим объясняется тот факт, что в рассмотрение не включены работы по биохимическим исследованиям, в том числе контроль за состоянием пациента, лабораторные радиологические исследования, проектирование оборудования и т. п., а также работы по эпидемиологии заболеваний, эффективности лечения и рациональности питания, очистке воды, автоматизации функционирования клиник, диагностике и т. д.

Следует иметь в виду, что обсуждаемые методы исследования операций сами по себе не рассматриваются, поскольку предполагается, что читатель знаком с основными подходами и методами исследования операций в объеме, достаточном для понимания изложенного материала.

1. КРАТКАЯ ХАРАКТЕРИСТИКА И ПРОБЛЕМЫ СИСТЕМЫ ЗДРАВООХРАНЕНИЯ

Для того чтобы был ясен общий принцип изложения, рассмотрим кратко основные элементы системы здравоохранения.

Процесс медицинского обслуживания включает следующие этапы: профилактический контроль за состоянием здоровья населения с целью своевременного выявления отклонений от нормального состояния, принятие решения относительно использования имеющихся средств для оказания медицинской помощи в случае установления отклонений от нормального состояния, применение

¹⁾ Robert B. Fetter, Ph. D., Professor in the Center for the Study of Health Services, Institution for Social and Policy Studies at Yale University, New Haven, Connecticut.

намеченных лечебных средств с целью перевода индивидуума в состояние (гомеостазис), настолько близкое к исходному (нормальному), насколько это позволяют сделать характер заболевания и имеющиеся лечебные средства. Таким образом, наиболее важными элементами процесса являются, во-первых, индивидуум (пациент) и, во-вторых, набор лечебных средств. Именно эти два фактора определяют количественные и качественные характеристики процесса медицинского обслуживания. Учет других факторов, связанных, например, с возникновением заболевания, его выявлением и диагностикой, наличием соответствующих лечебных средств, приводит к необходимости рассматривать систему медицинского обслуживания как сложную динамическую систему, которая, однако, сравнительно мало изучена, что объясняется большим разнообразием способов ее эксплуатации и реализации. Это обстоятельство, а также возможность существования в пределах рассматриваемой системы органов, выполняющих одни и те же функции, которые к тому же могут находиться в конфликтных отношениях или конкурировать между собой, свидетельствует о необходимости усовершенствования ее организации. Подобная необходимость становится еще более очевидной, если учесть, что для системы здравоохранения характерно отсутствие не только данных о потребностях в медицинском обслуживании, но и данных, необходимых для размещения основных служб этой системы.

Характер указанных выше проблем позволяет сделать вывод о целесообразности и перспективности применения методов исследования операций для их решения, хотя возникающие при этом задачи объемны, сложны и часто не до конца формализованы.

Ниже перечисляются проблемы службы здоровья, представляющие интерес с точки зрения анализа с помощью методов исследования операций.

Наиболее общей и распространенной проблемой является организация региональной службы здоровья, хотя в работах по исследованию операций именно этой проблеме уделяется мало внимания. Решение данной проблемы предполагает рассмотрение широкого круга вопросов, таких, как прогнозирование потребностей, определение объема и размещение средств обслуживания, выбор вариантов при проектировании и организации средств обслуживания, обеспеченность кадрами, установление факта необходимости оказания медицинской помощи, финансирование, оплату и т. д.

Особенно важной представляется проблема размещения ресурсов в системе службы здоровья с учетом требований и целей других систем. В этом случае в первую очередь необходимо решить задачу информационного обеспечения управления, включающего предсказания последствий различных вариантов решения относительно размещения в разных условиях.

Важные проблемы системного характера возникают при планировании различных программ службы здоровья, связанных, например, с очисткой воды, питанием, регулированием численности семьи, профилактическими прививками населения и т. д. Исследование динамики таких проблем представляет особый интерес, так как это дает возможность определить режим потребления ресурсов в системе и эффективность реализации соответствующего решения.

Существенный круг проблем связан с распределением трудовых ресурсов, планирование которых осуществляется для различных систем медицинского обслуживания, включая национальные и международные, локальные и региональные. Потребность в трудовых ресурсах зависит в основном от изменения объема оказываемых услуг, а обеспеченность трудовыми ресурсами — от возможности обучения, наличия стимулов для поступления на работу или увольнения, от взаимосвязи различных программ обучения и, несомненно, от требований, предъявляемых к медицинскому персоналу, а также от миграции населения. Исследование динамики взаимодействия всех этих факторов с целью предсказания уровня наличия и степени использования трудовых ресурсов представляет собой сложную аналитическую задачу.

Не менее сложной проблемой, настоятельно требующей разрешения, является контроль качества медицинского обслуживания, осуществляемого в различных условиях. В данном случае главная трудность заключается в отсутствии единого мнения о выборе метода оценки качества.

Проблема, которой до сих пор не уделялось должного внимания, хотя она является достаточно актуальной, — это функционирование учреждений системы здравоохранения.

Большое число новых проблем возникло в сфере амбулаторного обслуживания. Это объясняется существенными изменениями, которые произошли в характере амбулаторного лечения и способах оказания медицинской помощи. По мере того как все больше осознается настоящая и будущая роль амбулаторного лечения, возникает потребность в разработке новых и более качественных методов анализа эффективности соответствующих систем.

Исследователи систем большое внимание уделяют вопросам, связанным с организацией неотложной медицинской помощи, включая управление транспортом и службой оповещения. Тем не менее многие важные проблемы служб неотложной медицинской помощи до сих пор остаются вне поля зрения исследователей. К ним относится определение сфер и зон обслуживания, организация отдельных служб неотложной медицинской помощи и их взаимодействия со службами других подсистем здравоохранения, комплектование обслуживающего персонала, проведение обследований и множество других.

Важными областями исследования для разработки методов диагностики и лечения являются радиология, патология, лабораторный анализ. Повышение эффективности соответствующих служб здравоохранения связано с решением таких проблем, как распределение ресурсов, расстановка персонала, выбор технологии, использование вычислительной техники для обработки информации.

Круг проблем, относящихся к стационарному лечению, представляет наибольший интерес, поскольку их решение с помощью методов исследования операций оказывается экономически оправданным. Это объясняется тем, что затраты на стационарное лечение составляют большую часть общих затрат в системе здравоохранения. Таким образом, проблема рациональной организации стационарного обслуживания в различных условиях функционирования медицинского учреждения является в системе здравоохранения центральной. Кроме того, следует отметить, что немалые возможности для использования методов исследования операций в рамках решения данной проблемы предоставляет наличие в системе стационарного обслуживания разнообразных организационных структур, а также необходимость в рациональном размещении объектов данной системы.

При организации работы различных служб крупных больниц существует реальная возможность использования методов теории управления запасами. В этой связи можно указать несколько областей, таких, как организация поставок ресурсов различных типов, переработка белья, консервирование крови.

Не менее важными являются проблемы составления меню для лечебного и диетического питания. При решении этих проблем используются как методы математического программирования, так и эвристические методы.

Проектирование и внедрение в системе здравоохранения информационных систем представляют собой область, где применение методов исследования операций весьма перспективно как с точки зрения разработки экономических систем, так и с точки зрения обеспечения удобства пользователям. Следует заметить, что существовали серьезные сомнения относительно технической реализуемости и экономической целесообразности использования автоматизированных вычислительных систем управления крупными больницами на основе применения ЭВМ. Однако успехи в технологии производства и снижения стоимости ЭВМ позволяют прийти к заключению, что таким системам принадлежит будущее.

Перейдем к обзору работ по изучению системы здравоохранения, выполненных к настоящему времени. Следует подчеркнуть, что обзор не является исчерпывающим и основная часть изложенного материала относится к проблемам управления и применению методов исследования операций. Что касается теоретических

основ и исходной информации, то читатель может найти все интересующие его сведения в приведенной библиографии. Тем же, кто интересуется специальными вопросами, рекомендуем обратиться к Abstracts of Hospital Management Studies.

2. АНАЛИЗ ПРОБЛЕМ СИСТЕМЫ ЗДРАВООХРАНЕНИЯ С ПОМОЩЬЮ МЕТОДОВ ИССЛЕДОВАНИЯ ОПЕРАЦИЙ

Как уже отмечалось, методы исследования операций используются для создания макро- и микромоделей системы здравоохранения. В данном разделе основное внимание уделяется проблеме управления здравоохранением. Однако следует иметь в виду, что выделение подобной проблемы, так же как и классификация исследований на макро- и микроуровнях, являются весьма условными.

В настоящем обзоре рассматриваются только те работы, которые положили начало исследованиям системы здравоохранения с помощью методов исследования операций. Поскольку исторически раньше появились работы по составлению микромоделей, то начнем наш обзор именно с них.

2.1. Амбулаторное обслуживание

В числе первых применений методов исследования операций в области здравоохранения было использование теории массового обслуживания для решения задач проектирования и организации работ амбулаторий [304—306]. В этих работах определились основные направления исследований, затрагивающие такие проблемы, как составление расписания работы врачей и вспомогательного персонала, обслуживание пациентов, поступивших без предварительной записи, назначение времени приема больных. Поскольку амбулаторное лечение проводится в различных условиях, то для решения перечисленных проблем большую роль играет имитационное моделирование. Соответствующие имитационные модели представлены, например, в работах [96, 99, 150, 151, 253, 272, 308].

Следует отметить, что по ряду причин аналитическую теорию массового обслуживания не удастся непосредственно использовать для решения задач организации амбулаторного лечения. Объясняется это следующими причинами.

Поток заявок на амбулаторное обслуживание содержит как планируемую, так и непланируемую компоненты. Вместе с тем для организации служб амбулаторного лечения планирование заявок столь же важно, как и составление всевозможных расписаний обслуживания.

Поток пациентов является нестабильным и может существенно изменяться на протяжении коротких интервалов времени.

Многие системы здравоохранения имеют службы амбулаторного лечения, которые обеспечивают различный уровень обслуживания, включая консультации и повторное назначение с целью более эффективного лечения сложных форм заболеваний.

Службы амбулаторного лечения часто используют ресурсы, которые также необходимы для проведения стационарного лечения и другим службам системы здравоохранения. Это обстоятельство, видимо, должно приводить к формальным задачам, несколько отличным от тех, что рассматриваются в классической теории массового обслуживания.

Наряду с проблемами планирования и организации служб амбулаторного лечения важную роль играют вопросы обеспечения амбулаторных учреждений кадрами. В работе [127] рассматривается потенциальная возможность использования в амбулаториях неквалифицированного персонала. Планирование и управление персоналом в системе амбулаторного лечения являются предметом исследования многих проводимых в настоящее время работ по использованию исследований операций для анализа системы здравоохранения, включая разработку имитационных моделей.

2.2. Стационарное медицинское обслуживание

Большое число прикладных работ посвящено как весьма специальным, так и достаточно общим вопросам стационарного медицинского обслуживания. Среди первых наиболее важных работ были исследования, проведенные в Университете имени Джона Хопкинса в период с 1955 по 1965 г. с целью изучения проблем прогнозирования и контроля за состоянием больных [57, 323], разработки методов обучения обслуживающего персонала [103, 313, 314], проблемы профилактических осмотров [106], реализации прогрессивных методов лечения [112, 114]. В это же самое время в Йельском университете проводилась работа по изучению таких вопросов, как прогнозирование потребностей в службе охраны материнства [286], контроль за использованием средств хирургии [98, 179, 180], экономическая оправданность многопрофильных медицинских учреждений [289], анализ эффективности прогрессивных методов лечения [100, 101]. Целью всех перечисленных работ и большинства работ, указанных в библиографии, являлось повышение эффективности медицинского обслуживания в существующих лечебных учреждениях. В этих работах рассматриваются главным образом проблемы оказания неотложной медицинской помощи и лишь отдельные элементы или аспекты системы стационарного лечения. Данное обстоятельство однако не умаляет важности проведенных исследований.

Подразделения любой больницы могут быть организованы точно так же, как службы промышленных фирм, поэтому при анализе вопросов организации работы больниц используются традиционные инженерные подходы, а не общие методы исследования операций. Если, например, просмотреть литературу за 1973 г., то можно установить, какие из больничных служб являлись предметом исследований: это, во-первых, проблема организации аптек, прачечных, ремонтных мастерских, а также диетического питания [14, 16, 136, 81]. Исчерпывающий обзор прикладных работ в данной области был составлен исследовательской группой Технологического института в Джорджии [267].

Сравнительно мало работ проведено в области медицинского обслуживания пациентов с хроническими заболеваниями, хотя в этом виде обслуживания, весьма отличном от обслуживания в неотложных случаях, нуждаются многочисленные группы пациентов.

В заключение следует отметить, что исследователям систем при решении проблемы организации стационарного медицинского обслуживания следует принимать во внимание важные вопросы приемлемости предлагаемых изменений существующей системы стационарного лечения.

2.3. Информационные системы в здравоохранении

По своему назначению информационные системы в здравоохранении можно разделить, с одной стороны, на медицинские и административные, а с другой стороны — на исследовательские и функциональные. Схематически такая классификация может быть представлена следующим образом:

Информационные системы	Области применения	
	медицинские	административные
Исследовательские	1	2
Функциональные	3	4

Информационную систему, которая включает все области, выделенные на схеме, следовало бы назвать глобальной системой. Однако возможность построения такой системы вызывает серьезные сомнения. Наибольшие усилия предпринимались для создания информационных систем, относящихся к области 4. Эти информа-

ционные системы разрабатывались для типичных сфер бизнеса: формирование платежных ведомостей, оформление счетов, управление запасами и т. д. Определенные усилия предпринимались для создания и исследования информационных систем, относящихся к области 2. Однако большого успеха соответствующие разработки пока что не имеют.

Что касается области 3, то здесь было разработано и внедрено немало систем. Эти системы привязывались к таким объектам медицинского обслуживания, как клинические лаборатории, учреждения амбулаторного лечения, аптеки и т. п. Однако в результате всех этих разработок пока еще не удалось сформировать общую методику проектирования соответствующих информационных систем.

Возвращаясь к области 4, следует указать на разработки, проведенные фирмами Control Data [60], IBM [166], Lockheed [207]. Необходимо отметить, что разработка комплексных информационных систем, покрывающих области 1 и 4, представляется весьма перспективной, однако создание таких систем наталкивается на серьезные проблемы технологического и экономического характера [17, 169, 194—196, 258].

2.4. Неотложная медицинская помощь

В области оказания неотложной помощи исследователями систем была выполнена значительная работа, имевшая определенный успех. Работа осуществлялась по двум основным направлениям: неотложная помощь при коронарной недостаточности [176, 283] и при несчастных случаях [140, 228].

При рассмотрении вопросов организации и функционирования системы неотложной помощи важную роль играет ее взаимодействие с более устойчивейшей системой амбулаторного лечения. Дело в том, что на практике такое взаимодействие часто не обеспечивается, что снижает эффективность обеих систем.

Еще одна важная проблема, которой уделяется большое внимание, — это повышение эффективности служб оказания травматологической помощи. Очевидно, что эффективное решение этой проблемы в значительной степени зависит от уровня средств перевозки. При этом надо иметь в виду, что требуемые средства весьма разнообразны и дорогостоящи, что решающим фактором при оказании неотложной медицинской помощи является время и что места получения травм заранее не известны, хотя часто локализация интенсивных источников травм вполне предсказуема (например, таким местом может быть крупная автострада).

Наличие этих факторов показывает, что для повышения эффективности использования средств перевозки в случае скорой помощи необходимо проводить исследования с использованием методов моделирования и системного анализа.

3. ПЛАНИРОВАНИЕ НАЦИОНАЛЬНЫХ И РЕГИОНАЛЬНЫХ СИСТЕМ МЕДИЦИНСКОГО ОБСЛУЖИВАНИЯ

Необходимость использования формальных методов для планирования национальных систем медицинского обслуживания общепризнана [146]. Тем не менее примеров практического применения методов исследования операций в данной области не так уж много. Это объясняется недостатком информации, а также наличием целого ряда ограничений политического, административного и экономического характера. Среди работ по планированию медицинского обслуживания в государственном масштабе следует отметить исследования, проведенные в СССР [244] и Швеции [146]. Заслуживает внимания также разработка подхода к планированию, опирающегося на классические модели математического программирования [2]. Этот подход, в частности, был применен к некоторым странам Латинской Америки. В термпланах математического программирования была сформулирована одна из задач планирования службы здоровья Южной Кореи, осуществляющей контроль за заболеваниями туберкулезом [94].

Полученная модель неоднократно использовалась Всемирной организацией здравоохранения (ВОЗ). Применение этой модели в различных условиях позволяет вести непрерывную работу по ее уточнению.

Определенное внимание исследователей привлекает проблема интеграции служб лечения и профилактики в пределах региональной и национальной систем медицинского обслуживания [315, 320, 321]. Эта проблема рассматривается главным образом применительно к отдельным видам заболеваний и соответствующим группам населения.

3.1. Планирование потребностей в медицинском персонале

Для предсказания потребностей в персонале использовались сравнительно простые модели, в которых учитывались лишь наблюдаемые скорости изменения использования трудовых ресурсов [186, 319]. Более достоверные модели прогноза предложения и спроса персонала основных лечебных учреждений, входящих в систему медицинского обслуживания, пока что не построены. Тем не менее предпринимаются определенные попытки разработать необходимые планирующим органам аналитические методы исследования влияния изменений важных параметров системы

на степень использования медицинского персонала [97]. В частности, были выполнены исследования по взаимозаменяемости различных видов трудовых ресурсов, при этом использовались методы математического программирования [127].

3.2. Управление системой медицинского обслуживания

В системе медицинского обслуживания можно выделить два типа органов управления: административные и медицинские. При этом возросшая роль первых объясняется резким удорожанием медицинской помощи за последнее время, а вторых — главным образом правильной оценкой роли населения в организации медицинского обслуживания. В области административного управления разрабатываются бюджетные системы и системы экономического анализа [64, 186, 238, 310]. Что касается медицинских органов управления, то они ориентируются на установленное при организации системы медицинского обслуживания требование, согласно которому необходим контроль за службами стационарного лечения с целью обеспечения надлежащего уровня медицинского обслуживания группы населения, обслуживаемого за счет государственных субсидий. Такой контроль предусматривает выбор историй болезни для ретроспективного анализа, причем выбор обычно осуществляется с учетом длительности госпитализации (например, могут рассматриваться только те истории болезни, в которых длительность госпитализации превышает заданное число дней). Выбранные истории болезни представляются на рассмотрение государственным или региональным экспертным комиссиям.

Следует подчеркнуть, что подобный контроль осуществляется для того, чтобы путем анализа и оценок ряда характерных случаев можно было повысить эффективность управления качеством лечебного процесса. Для реализации такого контроля важную роль имели исследования по использованию методов принятия решений как в вопросах выбора типичных случаев [220, 249, 311], так и в вопросах о назначении соответствующего лечения [74—76, 144, 248]. Большинство упомянутых исследований стимулировали разработку управляющей системы здравоохранения, в основе построения которой лежит положение о существовании зависимости между состоянием здоровья обслуживаемого населения и потребляемыми в системе медицинского обслуживания ресурсами.

3.3. Профилактический осмотр пациентов

В области профилактических осмотров изучалась возможность применения теории принятия решений и использования вычислительной техники, а также исследовались проблемы выявления

отдельных заболеваний [106, 202] и многопрофильных профилактических осмотров [55]. Следует заметить, что проведенные исследования, целью которых являлось улучшение медицинского обслуживания, концентрировались на проблемах предотвращения заболеваний и поддержания нормального состояния здоровья, а не на проблемах оказания неотложной помощи.

4. ПЕРСПЕКТИВНОСТЬ ИСПОЛЬЗОВАНИЯ МЕТОДОВ ИССЛЕДОВАНИЯ ОПЕРАЦИЙ ДЛЯ РЕШЕНИЯ ПРОБЛЕМ СИСТЕМЫ ЗДРАВООХРАНЕНИЯ

Проектирование и разработка технически реализуемых и экономически эффективных систем медицинского обслуживания, составляющих основу здравоохранения, представляют наиболее благодатную почву для приложения методов исследования операций. Это связано с тем, что громадное число возникающих при этом задач может быть решено с помощью аналитических методов, эффективность использования которых зависит главным образом от способности исследователей применить свои знания и опыт для решения проблем медицинского обслуживания, характеризующихся необычайно большим числом ограничений. В развитых странах, таких как, например, США и Англия, существующие традиции и нормы жизни таковы, что любые изменения в системе медицинского обслуживания наталкиваются на серьезные ограничения. В менее развитых странах многие разумные предложения по изменению системы медицинского обслуживания также могут оказаться недостаточно эффективными из-за низкого уровня жизни или политики, которую проводит правительство. Тем не менее не стоит отчаиваться, следует лишь иметь в виду, что традиционные методы исследования операций могут и не дать существенных результатов в области здравоохранения. В частности, в рамках традиционных подходов, хотя они в некотором смысле и полезны, пока не удастся разрешить сложные проблемы планирования, возникающие в системе здравоохранения.

В следующем разделе приводится математическая модель, позволяющая формализовать широкий класс важных задач планирования в системе медицинского обслуживания. При этом рассматриваемая модель имеет достаточное число модификаций, что дает возможность анализировать различные частные задачи планирования.

5. МОДЕЛЬ РАСПРЕДЕЛЕНИЯ РЕСУРСОВ В СИСТЕМЕ ЗДРАВООХРАНЕНИЯ

При анализе подсистем здравоохранения с точки зрения их проектирования или реорганизации главной проблемой является распределение ограниченных ресурсов, необходимых для реали-

зации различных методов лечения. С этой проблемой связан широкий круг вопросов: от достаточно простых, например сравнение различных методов лечения данного заболевания, до весьма сложных, например исчерпывающий анализ состояния здоровья различных групп населения при использовании тех или иных лечебных средств.

В таком анализе было бы необходимо оценить и предсказать процесс распространения заболевания, наметить основные методы лечения и определить стоимостные оценки для необходимого ресурса, в некоторой мере оценить эффект от проведенного лечения и, наконец, описать динамику протекания заболевания и проводимого курса лечения. Данный анализ может быть проведен в рамках описываемой ниже модели. Одна из модификаций предлагаемой модели имеет следующий вид:

$$e_{prt}P_{prt} - [Y_{prt} + \sum_{h=0}^t g_{prh}X_{prmt-t-h}] = 0. \quad (1)$$

где e_{prt} — ожидаемая степень распространения данного заболевания в группе населения p -го типа в r -й области в период времени t , P_{prt} — численность группы населения p -го типа в r -й области в период времени t , Y_{prt} — численность убывающей из r -й области в период времени t группы населения p -го типа, X_{prmt} — число случаев, когда в рассматриваемой группе населения в соответствующей области и в соответствующий период времени был применен m -й метод лечения, g_{prh} — величина, определяющая влияние m -го метода лечения на степень распространения заболевания в соответствующей группе населения спустя h периодов времени.

Из соотношения (1) вытекает зависимость степени распространения заболевания от предполагаемой численности группы населения. Кроме того, соотношение (1) характеризует динамическую связь между степенью распространения заболевания и принятыми решениями по выбору метода лечения.

Ограничение модели имеет вид

$$\sum_{m,p} a_{km}X_{m,prt} \leq A_{krt}, \quad (2)$$

где a_{km} — объем k -го ресурса, необходимого для реализации m -го метода лечения, и A_{krt} — общее количество k -го ресурса в r -м районе в период времени t .

Соотношение (2) представляет собой ограничение на использование ресурсов при различных методах лечения.

Другая модификация модели может быть представлена в виде

$$\sum d_{bmprt}X_{m,prt} - B_{bprt}^+ - B_{bprt}^- = B_{bprt}, \quad (3)$$

где d_{bpr} — величина, определяющая вклад в достижение цели лечения b для группы населения p -го типа p в r -й области в период времени t при использовании m -го метода лечения (вклад относится к одному случаю лечения); B_{bpr} — желаемая степень достижения цели b , B_{bpr}^- и B_{bpr}^+ — отклонения соответствующего знака от этой степени.

Соотношение (3) определяет степень достижения желаемого уровня состояния здоровья населения в зависимости от проводимых мероприятий по лечению заболевания. Например, с помощью такого соотношения можно было бы выявить уменьшение уровня смертности среди различных групп населения в отдельных районах.

Еще одна модификация модели имеет вид

$$Z = \min \sum_{bpr} w_{bpr}^+ B_{bpr}^- + w_{bpr}^- B_{bpr}^+.$$

Следует отметить, что с помощью одной из модификаций типа (3) можно непосредственно определять стоимость проводимых мероприятий по лечению населения, и тогда задача свелась бы к минимизации единственного критерия. Однако в действительности возникает необходимость в использовании многоцелевой структуры модели вместе с предположениями о наличии оценок относительной важности для различных целей.

Расширение сформулированной модели путем включения в нее не одного, а некоторого множества заболеваний потребовало бы добавления еще одного индекса (индекса характера заболевания). Кроме того, весьма вероятно, что может возникнуть необходимость включить в модель имеющиеся или планируемые к использованию средства лечения в виде ресурсов определенного типа, с тем чтобы осуществлять непосредственное их планирование. Особенно это относится к случаям, в которых планирование пацелено на небольшие изменения имеющихся средств медицинского обслуживания.

Важный вывод, который может быть сделан в результате анализа приведенной модели, состоит в том, что математическое моделирование в области медицинского обслуживания должно быть построено на всестороннем учете различных факторов, оставаясь при этом логически ясным и стройным. Руководящие и плановые работники системы здравоохранения должны располагать определенной информацией для выработки правильных решений. Моделирование как раз и служит для решения проблемы информационного обеспечения.

6. ПУТИ ДАЛЬНЕЙШЕГО УСОВЕРШЕНСТВОВАНИЯ СИСТЕМЫ МЕДИЦИНСКОГО ОБСЛУЖИВАНИЯ

Есть все основания считать, что система медицинского обслуживания, по крайней мере в США, в недалеком будущем должна претерпеть серьезные изменения. Быстрый рост потребностей в медицинской помощи, значительное расширение наших знаний о болезнях и человеческом организме, бесконтрольный рост цен в данной области — все это свидетельствует о необходимости изменения системы медицинского обслуживания и создает для этого благоприятную почву. Важную роль при этом должны сыграть работы в области исследования операций, которые, конечно, будут продолжаться во всех тех направлениях, о которых говорилось выше. В частности, более широко будут проводиться работы по исследованию макромоделей, причем важное место в прикладных разработках должно отводиться проблемам планирования и оценке влияния возмущений, которые вносятся в систему медицинского обслуживания в результате ее частичной реорганизации.

Служба здоровья, безусловно, представляет собой такую область, где применение методов исследования операций может дать существенный вклад, величина и эффективность которого в большой степени зависит от профессионального уровня исследователей систем.

ЛИТЕРАТУРА

1. Abel-Smith B., An International Study of Health Expenditures and its Relevance for Health Planning, WHO Public Health Paper No. 32, Geneva. 1967.
2. Ahumada J., et al., Health Planning: Problems of Concept and Method, Pan American Health Organization, Washington, D.C., 1965.
3. Alper P., Armitage P. H., Smith C. S., Educational Models, Manpower Planning and Control, *Operational Res. Quart.*, 18, No. 2, 93—103 (1967).
4. Arrow K. J., Uncertainty and the Welfare Economics of Medical Care, *Am. Econ. Rev.*, 53, 941—973 (1963).
5. Badger K. H., Queuing Theory: Predicting Optional Obstetric Service Utilization on Medical Hill, University of California, Graduate Program in Hospital Administration, 1967.
6. Bailey N. T. J., Operational Research in Medicine, *Operational Res. Quart.*, 3, 24—29 (June 1952).
7. Bailey N. T. J., A Study of Queues and Appointment System in Hospital Out-patient Departments, With Special Reference to Waiting-times, *J. Roy. Stat. Soc., Series B.*, 14, 185—199 (1958).
8. Bailey N. T. J., Queuing for Medical Care, *Appl. Stat.*, 3, 137—145 (November 1954).
9. Bailey N. T. J., Statistics in Hospital Planning and Design, *Appl. Stat.*, 5, 146—157 (November 1956).
10. Bailey N. T. J., Operational Research in Hospital Planning and Design, *Operational Res. Quart.*, 8, 149—157 (September 1957).

11. Bailey N. T. J., The Mathematical Approach to Biology and Medicine, Wiley, N.Y., 1967.
12. Baligh H. H., Laughunn D. J., An Economic and Linear and Linear Model of the Hospital, *Health Services Res.*, 4, 293—303 (1969).
13. Balintfy J. L., A Stochastic Model for the Analysis and Prediction of Admissions and Discharges in Hospitals. in C. W. Churchman and M. Verhulst (Eds.), Management Sciences: Models and Techniques, 2, Pergamon Press, N.Y., 1960, pp. 288—299.
14. Balintfy J. L., Linear Programming Models for Menu Planning, in H. E. Smalley and J. R. Freeman (Eds.), Hospital Industrial Engineering, Reinhold, N.Y., 1966, pp. 402—415.
15. Balintfy J. L., A Hospital Census Predictor Model. in H. E. Smalley and J. R. Freeman (Eds.), Hospital Industrial Engineering, Reinhold, N.Y., 1966, pp. 312—316.
16. Balintfy J. L., Nebel E. C., Experiments with Computer-Assisted Menu Planning, *JAHHA*, 40, 88—96 (June 1966).
17. Ball M. J., An Overview of Total Medical Information Systems, *Meth. Info. in Med.*, 10, 73—82 (April 1971).
18. Barker J., Computers in Hospitals, *The Hospital*, 67, 279—282 (August 1971).
19. Barnett G. O., Computers in Patient Care, *New England J. Med.*, 279, 1321—1327 (December 1968).
20. Barnett G. O., The Use of Computers in Clinical Data Management: The Ten Commandments, in Committee on Computer System in Medicine, American Medical Association (Ed.), Symposium on Computer Systems in Medicine, Chicago, American Medical Association, 1971, p. 85—89.
21. Barnett G. O., Castleman P. A., A Time-sharing Computer System for Patient-care Activities, *Computers and Biomed. Res.*, 1, 41—51 (March 1967).
22. Barnett G. O., Greenes R. A., Interface Aspects of a Hospital Information System, *Ann. N.Y. Acad. Sci.*, 161, 756—768 (September 1969).
23. Barnett G. O., Hofmann P. B., Computer Technology and Patient Care: Experiences of a Hospital Research Effort, *Inquiry*, 5, 51—57 (September 1968).
24. Barnoon S., Wolfe H., Scheduling a Multiple Operating Room System: A Simulation Approach, *Health Service Res.*, 3, 272—285 (Winter 1968).
25. Barr A., Measuring Nursing Care, in G. McLachlan (Ed.), Problems and Progress in Medical Care, Oxford University Press, London, 1964.
26. Barlett M. S., Monte Carlo Studies in Ecology and Epidemiology. IV, Proceeding of the 4th Berkeley Symposium on Mathematics. *Stat. and Prob.*, IV, 39—55 (1961).
27. Bartscht K. G., Hospital Staffing Methodologies — Concepts. Development, and Uses, *J. Ind. Eng.*, 18, 708—717 (December 1967).
28. Bartscht K. G., Rothenbuhler E. F., Man-minute Computations from the Basis for New Staffing Methodology, *Hospitals, JAHHA*, 40, 62—66 (March 1966).
29. Bartscht K. C., Bayha F. H., Time Analyzed for Alternative Food Flow Systems, *Hospitals, JAHHA*, 40, 88—92 (March 1966).
30. Bartscht K. C., Estrella M. A., Range of Required Skills Detailed in Study of the Pharmacy Function, *Hospitals, JAHHA*, 40, 94—98 (March 1966).
31. Bartscht K. C., Grimes R. M., Linen Production Methods Analyzed in the Laundry, *Hospitals, JAHHA*, 40, 107—110 (March 1966).
32. Baruch J. J., Hospital Research and Administration with a Digital Computer, *Circulation Res.*, 11, 629—636 (September 1962).
33. Battelle-Northwest, Systems Programs for Hospitals, Battelle Northwest, Richland, Washington, 1969.

34. Battistella R. M., Chester T. E., Role of Management in Health Services in Britain and United States, *Lancet*, 626—630 (March 1972).
35. Bisphram K., Holland W. W., Stringer J., Planning for Health. *The Hospital* (March 1974).
36. Bithell J. F., A Class of Discrete-time Models for Study of Hospital Admission Systems, *Operations Res.*, 17, 48—69 (January — February 1969).
37. Bithell J. P., Devlin H. B., Prediction of Discharge of Inpatients, *Health Services Res.*, 3, 174—184 (Fall 1968).
38. Blanco-White M. J., Pike M. C., Appointment Systems in Outpatients Clinics and the Effect of Patients' Unpunctuality, *Med. Care*, 2, 113—145 (1964).
39. Blumberg M. S., Systems Analysis and Health Manpower: Medical Manpower Continuing Crisis, *JAMA*, 201, 856—860 (1967).
40. Bodenham K. E., Wellman F., Foundations for Health Services Management: A Seicon Report for the Scottish Home and Health Department on the Requirements for a Health Service Information System. Oxford University Press, London (1972; for the Nuffield Provincial Hospital Trust).
41. Bogatyrev J. D., O Kriterijah Effektivnosti Dlia Matematicheskogo Modelirovaniia Razlichnykh Vidov Medicinskoj Pomoschi Naseleniju. (Evaluation Criteria for the Mathematical Modelling of Various Form of Medical Care). *Sovetsk. Zdravochr* 5, 8—13 (1969).
42. Boodman D. M., Scientific Inventory Control, *Hosp. Progress*, 48, 78 (November 1967).
43. Bove J. R., McKay D. K., Computer Approach to Hospital Blood Bank Inventory Control, *Transfusion*, 9, 143—150 (May — June 1969).
44. Brown P. T. S., Computers in Hospitals, in J. Rose (Ed.), *Computers in Medicine*, J. & A. Churchill, London, 1969, pp. 108—119.
45. Bruce R. A., Yarnall S. R., Computer-Aided Diagnosis and Diagnosis of Cardiovascular Disorder, *J. Chronic Dis.*, 19, 473—484 (1966).
46. Buchanan J. M., Automated Hospital Information Systems, *Military Med.*, 131, 1510—1512 (December 1966).
47. Budd P. J., AHIS — Automated Hospital Information System. Information Bulletin IB 30-70-2, Veterans Administration, Department of Data Management, Washington, D.C., 1970.
48. Caceres C. A., Rilke A. E., The Digital Computer as an Aid in the Diagnosis of Cardiovascular Disease, *Trans. N.Y. Acad. Sci.*, 23, 240—244 (January 1961).
49. Carpenter R. J., A Review of Developed and Developing Routine Computer Applications in Medicine. WHO Documents. HS/ADP/66.1 and HS/ADP/66.1 Add. 1. Geneva, 1966.
50. Calassi C. A., Peterson E. L., The Blood Inventory Control System — Helping Blood Bank Management Through Computerized Inventory Control, *Transfusion*, 7, 60—69 (January — February 1967).
51. Catherer A., Warren M. D. (Eds.), *Management and the Health Services*. Pergamon Press, Oxford, 1971.
52. Catliff G. C., Critical Path Analysis as an Aid to Hospital Planning, *Brit. Hosp. Soc. Serv. J.*, 1204—1206 (1964).
53. Chacko G. K. (Ed.), *The Recognition of Systems in Health Services*. Operations Research Society of America, Arlington, Virginia, 1969.
54. Clark E. L., Diggs W. W., Quantifying Patient Care Needs, *Hospitals, JAMA*, 45, 96—100 (September 1971).
55. Collen M. F., General Requirements for a Medical Information System (MIS) in M. F. Collen (Ed.), *Medical Information Systems*, U.S. Department of Health Education, and Welfare, Health Services and Mental Health Administration, Washington, D.C., 1970, pp. 1—16.
56. Collen M. F., et al., Automated Multiphasic Screening and Diagnosis, *Am. J. Public Health*, 54, 741—750 (May 1964).

57. Connor R. J., A Hospital Inpatient Classification System, Doctoral Dissertation, Department of Industrial Engineering, The Johns Hopkins University, 1960.
58. Connor R. J., A Work Sampling Study of Variations in Nursing Work Load, *Hospitals, JAHA*, 35, 40 (May 1961).
59. Connors M. M., A Stochastic Elective Admissions Scheduling Algorithm, *Health Services Res.*, 5, 308—319 (Winter 1970).
60. Control Data Corporation, Integrated Medical Systems, Pub. No. X0010103, CDC Corp., La Jolla, California, 1970.
61. Cooperative Information Center for Hospital Management Studies, Abstracts of Hospital Management Studies Quarterly and Annual Cumulative, (various dates), University of Michigan, Ann. Arbor, Michigan.
62. Cross K. W., Roberts J. L., Management Control in Medical Care, *The Hospital*, 66, 45—47 (February 1970).
63. Cross K. W., Roberts J. L., Droar J., Electronic Processing of Hospital Records, G. McLachlan and R. A. Schegog. (Eds.), *Computers in the Service of Medicine*, Vol. 1, Oxford University Press, London, 1969, pp. 23—39.
64. Crystal R. A., Brewster A. W., Cost Benefit and Cost Effectiveness Analysis in the Health Field, *Inquiry*, 3, 3—13 (1966).
65. Crystal R. A., Brewster A. W., Improving the Use of Hospital Maternity Beds, *Operational Res. Quart.*, 16, 65—76 (1965).
66. Cvetanovic B., Grab B., Uemura K., Epidemiological Model of Typhoid Fever and its Use in the Planning and Evaluation of an Anti-Typhoid Immunization and Sanitation Programmes, *Bulletin, World Health Organization*, 45, 53—75 (1971).
67. Davies J. O. F., et al., Towards a Measure of Medical Care: Operational Research in the Health Services, Oxford University Press, London, 1962.
68. Davis L. S., et al., Computer-stored Medical Record, *Computers and Biomed. Res.*, 1, 452—469 (May 1968).
69. Deboeck G., User Oriented Modelling for Health Planning and Programming, IIASA Conference on Systems Aspects of Health Planning, Baden, Austria, August 1974.
70. DeLand E. C., Waxman B. D., Review of Hospital Information Systems, Report No. P-4337, The Rand Corporation, Santa Monica, California, 1970.
71. Denison R. A., Wild R., Martin M. J. C., A Bibliography of Operational Research in Hospitals and the Health Services, University of Bradford, Management Center, Bradford, England, 1969.
72. Dennis L. C., II, Balintfy J. L., A Linear Programming Analysis of Institutional Management Policies, Paper presented at Joint Meeting of the Operations Research Society of America and the Institute of Management Sciences, San Francisco, May 1968.
73. Dixon R. A., A Model of a Hospital's Patient Medical Information System, *Methods of Info. in Med.*, 9, 88—97 (April 1970).
74. Donabedian A., Evaluating the Quality of Medical Care, *Milbank Memorial Fund Quart.*, Part 2, 44, 166—206 July (1966).
75. Donabedian A., Promoting Quality Through Evaluating the Process of Patient Care, *Medical Care*, 6, 181—202 (May—June 1968).
76. Donabedian A., Medical Care Appraisal—Quality and Utilization, in: A Guide to Medical Care Administration Vol. 2, American Public Health Association, 1969.
77. Douville G., La Technique de la Simulation Appliquée au Block Operatoire, *Administration at Santé*, 3, 5—15 (Iliver) (1968).
78. Dowling W. L., The Application of Linear Programming to Decision-Making in Hospitals, *Hosp. Admin.*, 16, 66—75 (Summer 1971).
79. Dunn H. D., Hindle A., Decision Making in the Hospital Environment, Paper presented at the 31st Conference of the Operations Research Society of America, N.Y., May 1967.

80. Dunn R. G., Scheduling Elective Admissions, *Health Services Res.*, **21**, 181—215 (Summer 1967).
81. Eckstein E., Is the 'Diet Problem' Identical to the 'Menu Planning Problem'? *Management Sci.*, **16**, 527—528 (May 1970).
82. Edgecumbe R. H., The CASH Approach to Hospital Management Engineering, *Hospitals. JAH*, **39**, 70—74 (March 1965).
83. Elston R. C., Pickrel J. C., A Statistical Approach to Ordering and Usage Policies for a Hospital Blood Bank, *Transfusion*, **3**, 41—47 (January — February 1963).
84. Elston R. C., Pickrel J. C., Guides to Inventory Levels for a Hospital Blood Bank Determined by Electronic Computer Simulation, *Transfusion*, **5**, 465—470 (September — October 1965).
85. Elveback L., Varma A., Simulation of Mathematical Models for Public Health Problems, Public Health Repts., **80**, No. 12, 1067—1076 (1965).
86. Esogbue A. O., Experiments on Scheduling Disciplines in Surgery: A Simulated Queuing Approach, Technical Memorandum No. 168, Case Western Reserve University, Department of Operations Research, 1969.
87. Fagnani F., Introduction a la Recherche Operationelle en Santé Publique, *Bulletin du l'Institut National de la Santé et de la Recherche Médicale*, **23**, 837—856 (1968).
88. Fairry M. J., Information Systems in Hospital Administration, in M. E. Abrams (Ed.), Medical Computing, American Elsevier, N.Y., 1970, pp. 384—389.
89. Fanshel S., Bush J. F., A Health Status Index, *Operations Res.*, December 1970.
90. Feldstein M. S., Economic Analysis for Health Service Efficiency, North Holland Publishing, Amsterdam, 1967.
91. Feldstein M. S., Economic Analysis, Operational Research, and the National Health Service, *Oxford Econ. Pap.*, **15**, 19—31 (March 1963).
92. Feldstein M. S., Effects of Differences in Hospital Bed Scarcity on Type of Use, *Brit. Med. J.*, **2**, 561—564 (August 1964).
93. Feldstein M. S., An Aggregate Planning Model of the Health Care Sector, *Med. Care*, **5**, 369—381 (November — December 1967).
94. Feldstein M. S., Piot M. A., Sundaresan T. K., A Resource Allocation Model for Public Health Planning, *WHO Bull. Supplement*, 1973.
95. Feldstein P. J., German J. J., Predicting Hospital Utilization: An Evaluation of Three Approaches, *Inquiry*, **2**, 13—36 (June 1965).
96. Fetter R. B., Mills R., HOSPSIM: A Simulation Modeling Language for Health Care Systems, WP25, Simulation, March 1973, pp. 73—80.
97. Fetter R. B., Mills R., A Language to Aid in Manpower Planning, Puter Associates, New Haven, Connecticut, 1972.
98. Fetter R. B., Thompson J. D., The Simulation of Hospital Systems, *Operations Res.*, **13**, 689—711 (September — October 1965).
99. Fetter R. B., Thompson J. D., Patient's Waiting Time and Doctor's Idle Time in the Outpatient Setting, *Health Services Res.*, **1**, 66—90 (Summer 1966).
100. Fetter R. B., Thompson J. D., A Decision Model for the Design and Operation of a Progressive Patient Care Hospital, *Med. Care*, **7**, 450—462 (November — December 1969).
101. Fetter R. B., Thompson J. D., A Planning Model for the Delivery of Health Services, in A. Reisman and M. Kiley (Eds.), Health Care Delivery Planning, Gordon & Breach, N.Y., 1973.
102. Fetter R. B., Interactive Modelling of Health Care Delivery Systems, Proceedings of the Colloquium on Systems Theory and Scientific Management of Public Utilities, University of Montreal, January 1975.
103. Flagle C. D., Operations Research in a Hospital, in C. D. Flagle, W. H. Hug-

- gins. and R. H. Roy (Eds.), *Operations Research and Systems Engineering*, John Hopkins Press, Baltimore, 1960.
104. Flagle C. D., *Operations Research in the Health Services*, *Operations Res.*, **10**, 591-603 (September - October 1962).
105. Flagle C. D., *Operations Research in Community Services*, in D. B. Hertz and R. T. Eddison (Eds.), *Progress in Operations Research*, Wiley, N.Y., 1964.
106. Flagle C. D., A Decision Theoretical Comparison of Three Methods of Screening for a Single Disease. *Proceedings of the 5th Berkeley Symposium on Mathematics, Statistics and Probability*, Vol. 5, 1965, pp. 887-898.
107. Flagle C. D., A Decade of Operations Research in Health, in F. Swicky and A. G. Wilson (Eds.), *New Methods of Thought and Procedure*, Springer-Verlag, N.Y., 1967, pp. 33-41.
108. Flagle C. D., Technological Development in the Health Services, *Health Services Res.*, **4**, 6-13 (Spring 1969).
109. Flagle C. D., Communication and Control in Comprehensive Patient Care and Health Planning, *Ann. N.Y. Acad. Sci.*, **161**, 714-729 (September 1969b).
110. Flagle C. D., The Role of Simulation in the Health Services, *Am. J. Public Health*, **60**, 2386-2394 (December 1970).
111. Flagle C. D., Young J. P., Application of Operations Research and Industrial Engineering to Problems of Health Services, Hospitals, and Public Health, *J. Ind. Eng.*, **17**, 609-614 (November 1966).
112. Flagle C. D., et al., Optimal Organization and Facility for a Nursing Unit, Progress Report, Dec. 1957 - Dec. 1959, Johns Hopkins Hospital, Operations Research Division, Baltimore, 1960.
113. Flagle C. D., et al., The Problems of Organization for Hospital Inpatient Care, in C. W. Cruchman and M. Verhulst (Eds.), *Management Sciences: Models and Techniques*, Vol. 2, Pergamon Press, N.Y., 1960, pp. 275-287.
114. Flagle C. D., et al., The Progressive Patient Care Hospital Estimating Bed Needs, Public Health Service Publication No. 930-C-2, U. S. Government Printing Office, Washington, D.C., 1963.
115. Frankfurter G. M., et al., A Regional Computer-Based Blood Inventory Information System, *Transfusion*, **10**, 203-214 (July - August 1970).
116. Freibron R. B., Two Years Invested in Planning New Loyola Information System, *Modern Hospital*, Vol. 112, February, 1969.
117. Fretheim H. E., Nussbaum E. E., A Computer Based Medical Information System, Control Data Corporation, Medical System Research Laboratory Report, Burlington, Vermont, 1971.
118. Friend J. K., Jessop W. N., *Local Government and Strategic Choice: An Operational Research Approach to the Processes of Public Planning*, Tavistock Publications, London, 1969.
119. Fries B. E., The Application of Operations Research to the Hospital Complex, Technical Report No. 92, Cornell University, Department of Operations Research, Ithaca New York, 1969.
120. Fuchs V. R., The Basic Forces Influencing Costs of Medical Care, In U.S. Department of Health, Education, and Welfare, Report of the National Conference on Medical Costs, June 27-28, U.S. Government Printing Office, Washington, D.C., 1967, pp. 16-34.
121. Gargov K., Colakov V., Borisov D., Edraveopazvane i Novijat Nacin na Planirane i Rokovodstvo na Narodnoto Stopanstvo (The Health Services and New Methods of Planning and Managing the National Economy), *Hig. Zdraveop (Sofija)*, **4**, 309-315 (1969).
122. Geisler R., The THOMAS Medical Information System, *Datamation*, **16**, 133-136 (June 1970).
123. General Electric Company, MEDINET Nursing Staff Allocation, General Electric, Watertown, Mass., 1969.

124. Gillette P. J., Rathbun P. W., Wolfe H. B., Hospital Information Systems, Parts 1 and 2. *Hospitals. JAHHA*, 44, 76-78 (August 1970), 44, 45-48 (September 1970).
125. Gocka E. F., A Survey of Medical Information Systems, paper presented at the 11th American Meeting of the Institute of Management Sciences, Los Angeles October 1970.
126. Goldman J., Knappenberger H. A., How to Determine the Optimum Number of Operating Rooms. *Mod. Hosp.*, 11, 114 (September 1968).
127. Golladay F. L., Smith K. R., Miller M., Allied Health Manpower Strategies Estimates of the Potential Gains From Efficient Task Delegation, Research Report No. 5, University of Wisconsin, Health Economics Research Center, Madison, Wisconsin, October 1971.
128. Gordon B. L., Medical Data Repository: Resources; Development: Utilization, in Committee on Computer Systems in Medicine (Ed.), Symposium on Computer Systems in Medicine, American Medical Association, Chicago, 1971, pp. 12-21.
129. Corry G. A., Problem Solving Strategies in a System for Computer-Aid Diagnosis, Working Paper No. 267-68, Massachusetts Institute of Technology, Sloan School, Cambridge, Massachusetts, 1967.
130. Gouveia W. A., Computer Applications in the Hospital Pharmacy. *Hospitals. JAHHA*, 45, 80-83 (January 1971).
131. Greenes R. A., Sidel V. W., The Use of Computer Mapping in Health Research. *Health Services Res.*, 2, 243-258 (1967).
132. Grooms H. R., Au T., The Simulation of Hospital Patient Treatment Systems, in Proceedings of the Third Conference on Applications of Simulation, Los Angeles, December 8-10, 1969.
133. Grundy F., Reinke W., Health Practice Research and Formalized Managerial Methods. *WHO Publications*, Geneva, 1972.
134. Gue R. L., Operations Research in Health and Hospital Administration, *Hosp. Admin.*, 10, 6-25 Fall (1969).
135. Gue R. L., An Introduction to the Systems Approach in the Dietary Department. *Hospitals. JAHHA*, 43, 100-101 (September 1969).
136. Gue R. L., Liggett J. C., Mathematical Programming Models for Hospital Menu Planning. *J. Ind. Eng.*, 17, 395-400 (August 1966).
137. Gupta I., Zoreda J., Kramer N., Hospital Manpower Planning by Use of Queuing Theory, *Health Services Res.*, 6, 76-82 (Spring 1971).
138. Halpert H. P., Horvath W. J., Young J. P., An Administrator's Handbook on the Application of Operations Research to the Management of Mental Health Systems, Pub. No. 1003, National Clearinghouse for Mental Health Information, U.S. Government Printing Office, Washington, D.C., 1970.
139. Hammon G. L., Jacobs S. E., Shared Computer Systems, Parts 1 and 2, *Hospitals. JAHHA*, 44, 50-53; (May 1970); 44, 46 (1970).
140. Handside A. J., Morris D., Simulation of Emergency Bed Occupancy, *Health Services Res.*, 2, 287-297 (Fall-Winter 1967).
141. Harrington J. H., Operations Research - A Relatively New Approach to Managing Man's Environment, *New England J. Med.*, 275, 1342-1350 (December 1966).
142. Harvey N. A., Cybernetic Application in Medicine, *N.Y. State J. Med.*, 65, 765-772, 871-875, 995-1002 (1965).
143. Hearn C. R., Bishop J. M., Computer Model Simulating Medical Care in Hospital. *Brit. Med. J.*, 3, 396-399 (August 1970).
144. Herring C., Ohio Hospital's Quality Control and Staff Utilization Program, *Hosp. Progress*, 51, 38-44 (July 1970).
145. Hershey J. C., Abernathy W. J., Baloff N., A Variable Nurse Staffing Model, Research Paper No. 45, Stanford University, Graduate School of Business, Stanford, California, 1971.

146. Hilleboe H. E., Berkhuus A., Thomas W. C., Approaches to National Health Planning, WHO, Public Health Paper No. 46, 1972.
147. Hills P. M., Towards a Hospital Information System, in M. E. Abrams (Ed.), Medical Computing, American Elsevier, N.Y., 1970.
148. Hirsch R. L., Brodheim E., Ginsberg F. E., A Computer-Based Blood Inventory and Information System for Hospital Blood Banks as Part of a Regional Blood-Management Program, *Transfusion*, 10, 194—202 (July — August 1970).
149. Hofmann P. B., Gouveia W. A., Computers: Great Future, Perilous Present, *Mod. Hosp.*, 111, 98 (July 1968).
150. Hofmann P. B., Rockart J. F., Implications of the Noshov Rate for Scheduling OPD Appointments, *Hosp. Prog.*, 50, 35—40 (August 1969).
151. Hofman P. B., Barnett G. O., Planning for an Automated Clinic Appointment System, *Hosp. Topics*, 48, 37—42 (October 1970).
152. Högman C. F., Ramgren O., Computer System for Blood Transfusion Service, *Transfusion*, 10, 121—132 (May — June 1970).
153. Horvath W. J., The Systems Approach to the National Health Problem, *Management Sci.*, 12, B-391—B-392 (1966).
154. Horvath W. J., British Experience with Operations Research in the Health Services, in K. L. White (Ed.), Medical Care Research, Pergamon Press, New York, 1965, pp. 55—64.
155. Horvath W. J., Operations Research in Medical and Hospital Practice, in P. M. Morse (Ed.), Operations Research for Public Systems, The MIT Press, Cambridge, Massachusetts, 1967, pp. 127—157.
156. Horvath W. J., Organizational and Management Problems in the Delivery of Medical Care, *Management Sci.*, 14, 275—279 (February 1968).
157. Howell J. T., Systems Opportunities in Modern Medical Organizations, *World Hosp.*, 5, 153—156 (July 1969).
158. Howland D., Approaches to the Systems Problem, *Nursing Res.*, 12, 172—174 (1963).
159. Howland D., A Hospital System Model, *Nursing Res.*, 12, 232—236 (1963).
160. Howland D., A Model for Hospital System Planning, in G. Kreweras and G. Morlat (Eds.), Proceeding of the Third International Conference on Operational Research, English Universities Press, London, 1964, pp. 204—211.
161. Howland D., Approach to Nurse-Monitor Research, *Am. J. Nursing*, 66, 556—558 (March 1966).
162. Howland D., Toward a Community Health-System Model, *Bio-Med. Computing*, 1, 11—29 (1970).
163. Ilurburt E. L., Jones A. R., Blood Band Inventory Control, *Transfusion*, 4, 126—133 (March — April 1964).
164. Ingbar M. L., Organization of Health Services: Challenge to Operations Research, in G. K. Chacko (Ed.), The Recognition of Systems in Health Services, Operations Research Society of America, Arlington, Virginia, 1969, pp. 5—27.
165. Inquiry, Special Issue — The Computer and Hospital Information Systems, 5 (September 1968).
166. International Business Machines Corporation, Medical Information Systems Program (MISP), Version 2, Application Description Manual, IBM Corp., White Plains, New York, 1970.
167. International Business Machines Corporation, HIS-MISP: Hospital Information System — Medical Information System Programs, Mimeographed, IBM Corp., White Plains, New York (undated).
168. Jackson R. R. P., Welch J. D., Fry J., Appointment Systems in Hospitals and General Practice, *Operations Res. Quart.*, 15, 219—237 (1964).
169. James F. D., The Development and Implementation of a Computer-Based Hospital Information System, in G. McLachlan and R. A. Shegog (Eds.),

- Computers in the Service of Medicine, Vol. 2, Oxford University Press, London, 1969, pp. 155—167.
170. Jelinek R. C., A New Approach to the Analysis of Nursing Activities, *Hospitals, JAMA*, 40, 89—91 (October 1966).
171. Jelinek R. C., A Structural Model for the Patient Care Operation, *Health Services Res.*, 2, 226—242 (Fall — Winter 1967).
172. Jelinek R. C., An Operational Analysis of the Patient Care Function, *Inquiry*, 6, 53—58 (June 1969).
173. Jennings J. B., An Analysis of Hospital Blood Bank Whole Blood Inventory Control Policies, *Transfusion*, 8, 335—342 (November — December 1968).
174. Jennings J. B., Information Needs for Hospital Blood Bank Inventory Control, *Transfusion*, 9, 214—216 (July — August 1969).
175. Kant V., *Mathematiceskie i ekonomicheskie metody v zdravoohraneni.* (The Use of Mathematical and Economic Methods in the Health Services) Kisinev.
176. Kao E. P. C., A Semi-Markov Model to Predict Recovery Progress of Coronary Patients, *Health Services Res.*, 7, No. 3, 191—208 (1972).
177. Kao E. P. C., A Semi-Markovian Population Model with Application to Hospital Planning, *IEEE Trans. on Syst., Man. and Cybernetics*, SMC-3, No. 4, 327—336 (1972).
178. Kast F. E., Rosenzweig J. E., Hospital Administration and Systems Concepts, *Hosp. Admin.*, 11, 17—33 (Fall 1966).
179. Kavet J., The Application of Computer Simulation Techniques to the Surgical Subsystem, Unpublished Master's Essay, Yale University. School of Medicine, New Haven, Connecticut, 1967.
180. Kavet J., Thompson J. D., Computers Can Tell You What Will Happen Before it Happens, *Mod. Hosp.*, 109, 102—105 (December 1967).
181. Kendall D. G., Mathematical Models of the Spread of Infection. Mathematics Research Council Mathematics and Computer Science in Biology and Medicine, London, 1945, pp. 213—225.
182. Kennedy F. D., Development of a Community Health Service System Simulation Model, in A. Reisman and M. Kiley (Eds.), *Health Care Delivery, Gordon and Breach*, N.Y., 1973.
183. Kiley J. M., et al., A Computer-based Medical Record, *J. JAMA*, 205, 571—576 (August 1968).
184. Kilpatrick K. E., Freund L. E., A Simulation of Tank Oxygen Inventory at a Community General Hospital, *Health Services Res.*, 2, 298—305 (Fall-Winter 1967).
185. Kirk J. F., Murray A., Neilson J. M., Automation of Hospital Medical Records — Data Processing, *Health Bull.*, 28, 49—52 (January 1970).
186. Kissick W. L., Forecasting Health Manpower Needs, *Hospitals, JAMA*, 41, 47—51 (September 1967).
187. Kissick W. L., How Imagination and Innovation Can Help Bridge Manpower Gaps, *Hospitals, JAMA*, 41, 76 (October 1967).
188. Kissick W. L., Planning, Programming and Budgeting in Health, *Med. Care.*, V, 201—207 (1967).
189. Klarman H. E., Increase in the Cost of Physician and Hospital Services, *Inquiry*, 7, 22—36 (March 1970).
190. Klarman H. E., Some Technical Problem in Areawide Planning for Hospital Care, *J. Chronic Dis.*, 17, 735—747 (1964).
191. Kleczkowski B. M., Planning and Evaluation of the Community Health Care Programmes: Model of a General Optimizing Method Sante Publ., Bucharest, Vol. 3, 1971, pp. 259—273.
192. Kolesar P., A Markovian Model for Hospital Admission Scheduling, *Management Sci.*, 16, 384—396 (February 1970).
193. Krismer J. R., Comprehensive Health Operating and Information Systems,

- Trans. Am. Acad. Ophthalmology and Otolaryngology*. 75, 1126—1131 (September — October 1971).
194. Lair B. M., Computerizing Total Hospital Activities — A Pioneer Report, *Comput. and Autom.*, 18, 30—32 (July 1969).
 195. Lamsun B. G., et al., The First Decade of Effort: Progress Toward a Hospital Information System at the UCLA Hospital, Los Angeles, California, *Methods Info. in Med.*, 9, 73—80 (April 1970).
 196. Lawton M. D., Systems Design for a Management-oriented Hospital Information System, in M. E. Abrams (Ed.), *Medical Computing*, American Elsevier, N.Y., 1970, pp. 358—373.
 197. Lechat M. F., Flagle C. D., Statistical Decision Theory and the Selection of Diagnostic and Therapeutic Strategies in Public Health, *Proceeding of the Third International Conference on Operational Research*, Oslo, 1963.
 198. Lechat M. F., Flagle C. D., Allocation of Medical and Associated Resources to the Control of Leprosy, in N. N. Barish and M. Verhulst (Eds.), *Management Sciences in the Emerging Countries*, Pergamon Press, Oxford, Great Britain, Chapter 9, 1965.
 199. Leighon E., CHD: California's Computerized Information Information System for Hospitals, *Hosp. Forum*, 13, 6 (May 1970).
 200. Levy R. R., Cammarn M. R., Information Systems Applications in Medicine, in C. A. Cuadra (Ed.), *Annual Review of Information Science and Technology*, V. 3, Encyclopedia Britannica, Chicago, 1968, pp. 397—428.
 201. Levy R. R., Cammarn M. R., Smith M. J., Computer Handling of Ambulatory Clinic Records, *JAMA*, 190, 1033—1037 (December 1968).
 202. Lincoln T. L., Parker R. D., Medical Diagnosis Using Baye's Theorem, *Health Services Res.*, 2, 34—45 (1967).
 203. Lipkin B., Engle R. L., Jr., et al., Digital Computer as an Aid to Differential Diagnosis. Use in Hematological Diagnosis, *Arch. Internal Med.*, 108, 57—72 (1961).
 204. Lockheed Missiles & Space Company, Proposal for a Mayo Clinic Medical Information System: Systems Analysis and Design Phase. LMSC, Sunnyvale, California, 1966.
 205. Lockheed Missiles & Space Company, Analysis of a Clinical Information System: University of Saskatchewan Hospital, LMSC, Hospital Systems Study Group, Sunnyvale, California, 1967.
 206. Lockheed Missiles & Space Company, Resumé of Capabilities and Interests in Medical Information Systems. Mimeographed, LMSC, Sunnyvale, California, 1968.
 207. Lockheed Missiles & Space Company, Analysis of Information Needs of Nursing Stations. Technical Report No. LMSC-682684. LMSC, Sunnyvale, California, 1969.
 208. Long M., Feldstein P. J., The Economics of Hospital Systems: Peak Loads and Regional Coordination, *Am. Econ. Rev.*, B-2, 119—124 (1967).
 209. Lubin J. W., Drossness D. L., Wylie L. G., Highway Network Minimum Path Selection Applied to Health Facility Planning, *Public Health Rep.*, 80, 771—777 (1965).
 210. Lubin J. W., Reed I. M., Worstell G. L., Drossness D. L., How Distance Affects Physician Activity, *Mod. Hosp.*, 107, 80—82ff (1966).
 211. Luck G. M., Hospital Operation Policies at the Commissioning Stage, Institute for Operations Research, Health Report, No. 3, 1969.
 212. Luck G. M., Luckman J., Smith B. W., Stringer J., Patients, Hospitals and Operational Research, Tavistock, London, 1971.
 213. Luckman J., Management Policies for Large Ward Units, Institute for Operations Research, Health Report No. 1, 1969.
 214. Luckman J., Planning Surgery — A Computer Simulation. Institute for Operations Research, Health Report No. 2, 1972.
 215. Mahler H. T., Piot M. A., Essais d'Application de la Recherche Opera-

- tionnelle Dans la Lutte Anti-Tuberculeuse, *Bull. de l'Institut National de la Santé et de la Recherche Médicale*, 21, 855—882 (1966).
216. Masouredis S. P., et al., Development of an Automated Blood Inventory and Information System for a Regional Transfusion Service, *Transfusion*, 10, 182—193 (July — August 1970).
 217. Michael J. M., Spatfore J., Williams E. R., A Basic Information System for Health Planning, *Public Health Rep.*, 83, No. 1, 21—28 (1968).
 218. Mills R. E., Fetter R. B., CML: A Conversational Modeling Language, Technical Report No. 53, Yale University, Department of Administrative Sciences, 1972.
 219. McCarn D. B., Moriarty D. G., Computers in Medicine, *Hospitals, JAMA*, 45, 37—39 (January 1971).
 220. McClain J. O., Riedel D. C., Screening for Utilization Review: On the of Explicit Criteria and Non-Physicians in Case Selection, *Am. J. Public Health*, 63, No. 3, 247—251 (1973).
 221. McLachlan G. (Ed.), Operational Research in the Health Services, a Symposium, Oxford University Press, London, 1962. (For the Nuffield Provincial Hospital Trust.)
 222. McLachlan G., Shegog R. A. (Eds.), Computers in the Service of Medicine. Essays on Current Research and Applications, Vol. 1, 2, Oxford University Press, London, 1968 (For the Nuffield Provincial Hospital Trust).
 223. McLachlan G. (Ed.), Health Operations Research and Systems Analysis Literature, in A. Sheldon, F. Baker and C. P. McLaughlin (Eds.), Systems and Medical Care, MIT Press, Cambridge, Massachusetts, 1970, pp. 27—48.
 224. McNerney W. J., Why Does Medical Care Cost So Much? *New England J. Med.*, 282, 1458—1465 (June 1970).
 225. Nasta M. D., Beddow J. K., Shapiro R. A., A Deterministic Input-Output Model to Facilitate Management of a Hospital System, *J. Operations Res. Soc. Japan*, 15, 19—33 (March 1972).
 226. Navarro V., Systems Analysis in the Health Field, *Socio-Econ. Plan. Sci.*, 3, 179—189 (August 1969).
 227. Navarro W., Parker R., White K., A Stochastic and Deterministic Model of Medical Care Utilization, *Health Services Res.*, 5, No. 4, 342—357 (1970).
 228. Newell D. J., Provision of Emergency Beds in Hospitals, *Brit. J. Preventive Social Med.*, 8, 77—80 (April 1954).
 229. Newell D. J., Immediate Admissions to Hospital, in G. Kreweras and G. Morlat (Eds.), Proceeding of the Third International Conference on Operational Research, English Universities Press, London, 1964, pp. 224—233.
 230. Newell D. J., Problems in Estimating the Demand for Hospital Beds, *J. Chronic Dis.*, 17, 749—759 (September 1964).
 231. Newell D. J., Hospital Bed Usage, *Brit. J. Hosp. Med.*, 3, 915—917 (June 1970).
 232. Newhouse J. P., Toward a Theory of Nonprofit Institutions: An Economic Model of a Hospital, *Am. Econ. Rev.*, 50, 64—74 (March 1970).
 233. Ockenden J. M., Bodenham K. E., Focus on Medical Computer Development, Oxford University Press, London, 1970.
 234. O'Connel B. P., McFarlane A. H., A Medical Care Information System: An Approach to the Computerization of Ambulatory Patient Records, *Med. Care*, 8, 82—87 (January — February 1970).
 235. O'Hara M., Josephson A. M., Regional Automated Data-Processing System for Blood Banks, *Transfusion*, 10, 215—220 (July — August 1970).
 236. O'Malley C. D., Application of Systems Engineering in Nursing, *Am. J. Nursing*, 69, 2155—2160 (October 1969).
 237. Overall J. E., Williams E. M., A Computer Procedure for Diagnosis of Phryoid Function, in K. Enslein (Ed.), Data Acquisition and Processing in Biology and Medicine, Vol. 2, Pergamon Press, London, 1962.

238. Packer A. H., Applying Cost-Effectiveness Concepts to the Community Health System. *ORSA J.*, 16, 343—350 (1968).
239. Pan American Health Organization, Regional Advisory Committee on Computers in Health First Meeting, Scientific Publication No. 211, Washington. 1970.
240. Park K. S., Freeman J. R., Community Health Resources Allocation with Linear Programming Methods, Technical Report No. 2, University of Florida. Health Systems Research Division, 1969.
241. Pegels C. C., Jelmer A. E., An Evaluation of Blood-Inventory Policies: A Markov Chain Application, *Operations Res.*, 18, 1087—1098 (November — December 1970).
242. Piot M., Sundaresan T. K., A Linear Programme Decision Model for Tuberculosis Control. Progress Report on the First Test-Runs. WHO/TB/Techn. Information/67—55, WHO Geneva, 1967.
243. Палкин Б. Н., Андерсон В. Я., Кононов А. Г. Оптимальное распределение прироста числа коек в лечебных учреждениях по специальностям и территории с применением электронно-вычислительных машин. — Советское здравоохранение, 1969. 10, с. 29—34.
244. Popov G. A., Principles of Health Planning in the USSR, WHO Public Health Papers No. 43, Geneva, 1970.
245. Rath G. J., Kennedy O. G., Lapp R. E., Developing a Composite Hospital Simulation Model, Report DDC-69-81, Northwestern University, The Technological Institute, Design and Development Center, Evanston, Ill., 1969.
246. Reed R., Jr., Stanley W. E., Optimizing Control of Hospital Inventories, *J. Ind. Eng.*, 16, 48—51 (January — February 1965).
247. Rice D. P., Measurement and Application of Illness Costs, *Public Health Rep.*, 84, No. 2 (1969).
248. Riedel D. C., Brenner M. H., et al., Psychiatric Utilization Review as Patient Care Evaluation, *Am. J. Public Health*, 62, No. 9, 1222—1228 (1972).
249. Riedel D. C., Fetter R. B., et al., Basic Utilization Review Program (BURP), Technical Report WP21, Yale University. Health Services Research, 1972.
250. Riedel D. C., Fitzpatrick T. B., Patterns of Patient Care, University of Michigan Press, Ann. Arbor, Michigan, 1964.
251. Riedel D. C., Fitzpatrick T. B., Leifer R. P., Prediction of Hospital Length of Stay, *Health Services Res. J.*, 1, 287—300 (Winter 1966).
252. Riedel D. C., Fitzpatrick T. B., Wing P., Computer Simulation of Hospital Patient Scheduling Systems, *Health Services Res.*, 3, 130—141 (Summer 1968).
253. Rockart J. F., Hofmann P. B., Physician and Patient Behavior Under Different Scheduling Systems in a Hospital Outpatient Department, *Med. Care*, 7, 463—470 (November — December 1969).
254. Rockwell T. H., Barnum R. A., Giffin W. C., Inventory Analysis as Applied to Hospital Whole Blood Supply and Demand, *J. Ind. Eng.*, 13, 109—114 (March — April 1962).
255. Romania, Ministry of Health, Studies and Methods of Operational Research in Public Health, Bucharest, 1969.
256. Rösch G., Operational Research in the Medical Field in France, WHO Document No. EURO 0408/11, Copenhagen, 1969.
257. Rosen D., Medical Care Information System of the Veterans Administration, *Public Health Rep.*, 83, 363—371 (May 1968).
258. Saunders M. G., The Computer and Total Medical Information Processing — A Present Fiction, *Canadian Hosp.*, 46, 64—67 (October 1969).
259. Schwartz M. D., Status of Hospital Information Systems, *Hosp. Progress*, 51, 53—60 (June 1970).

260. Scheff T. J., Decision Rules, Types of Errors and Their Consequences for Medical Diagnosis. *Behavioral Sci.*, 8, 97—107 (1963).
261. Shegog R. F. A., Reviewing Some Applications for Computers to Medicine, in G. McLachlan (Ed.), Problems and Progress in Medical Care, Third Series, Oxford University Press, London, 1968, pp. 146—170.
262. Sheldon A., Baker F., McLaughlin C. P. (Eds.). Systems and Medical Care, MIT Press, Cambridge, Massachusetts, 1970.
263. Siegel S. J., Developing an Information System for a Hospital. *Public Health Rep.*, 83, 359—362 (May 1968).
264. Silver A., Silver A. M., An Empirical Inventory Control System for Hospital Blood Banks, *Hospitals, JAMA*, 38, 56 (August 1964).
265. Singman D., et al., Computerized Blood Bank Control, *J. JAMA*, 194, 113—116 (November 1965).
266. Smalley H. E., Hospital Systems Research, *Hosp. Admin.*, 11, 42—50 (Winter 1966).
267. Smalley H. E., Freeman J. R., Hospital Industrial Engineering, Reinhold, N.Y., 1966.
268. Smalley H. E., et al., Inventory Policies, *Hosp. Management*, 97, 92—97 (March 1964).
269. Smallwood R. D., et al., A Medical Service Requirement Model for Health System Design, Proceedings of the IEEE (Special Issue on Automated Health Services), 1969.
270. Smallwood R. D., Sondik E. J., Offensend F. L., Toward An Integrated Methodology for the Analysis of Health Care Systems, *Operations Res.*, 19, 1300—1322 (October 1971).
271. Smith B. W., Luck G. M., Networks for the Commissioning of Walsgrave Hospital, Coventry, Institute for Operations Research, Health Report No. 4, 1970.
272. Soriano A., Comparison of Two Scheduling Systems. *Operations Res.*, 14, 388—397 (May — June 1966).
273. Souder J. J., Computers Can Bring a New Rationality Into Hospital Design, *Mod. Hosp.*, 110, 80—86 (March 1968).
274. Souder J. J., et al., Planning for Hospitals, American Hospital Association, Chicago, 1964.
275. Stimson D. H., Utility Measurement in Public Health Decision Making, *Management Sci.*, 16, 17—30 (October 1969).
276. Stimson D. H., Health Agency Decision-Making: An Operations Research Perspective, in M. F. Arnold, L. V. Binkenship and J. M. Hess (Eds.), Administering Health Systems: Issues and Perspectives, Aldine-Atherton N.Y., 398—433 (1971).
277. Stimson D. H., Stimson R. H., Operations Research in Hospitals. Hospital Research & Educational Trust, Chicago, Ill., 1972.
278. Stringer J., Operations Research for Multi-Organizations, *Operational Res. Quart.*, 18, No. 2, 105—120 (1967).
279. Stringer J., Operational Research in Planning and Designing Medical Facilities in the United Kingdom. WHO Document EURO 0408/12, Copenhagen.. 1969.
280. Strand L., Moznisti Aplikace Metody Monte Carlo ve Zdravotnických Zariadení (The Use of the Monte Carlo Method in Planning the Work of Medical Institutions), *Cs. Zdrav.*, 4—5, 193—198 (1969).
281. Teitz M. B., assisted by Polly Bart, Models of Patient Flows in a Hospital System: A Preliminary Analysis, Internal Working Paper. University of California, Department of City and Regional Planning, Berkeley, California, 1968.
282. Thomas W. H., A Model for the Prediction of Demand for Nursing Services, Paper presented at the 27th National Meeting of the Operations Research Society of America, Boston, May 1965.

283. Thomas W. H., A Model for Predicting Recovery Progress of Coronary Patients, *Health Services Res.*, 3 185—213 (1968).
284. Thomas W. H., et al., A Model of a Hospital Information System, Research Report No. 4, State University of New York at Buffalo, Department of Industrial Engineering, 1965.
285. Thompson J. D., On Reasonable Costs of Hospital Services, *Milbank Memorial Fund Quart.*, Part 2, 46, 33—41 (January 1968).
286. Thompson J. D., Fetter R. B., The Economics of the Maternity Service, *Yale J. Biol. and Med.*, 36, 91—103 (August 1963).
287. Thompson J. D., Fetter R. B., Research Helps Calculate OB Bed Needs, *Mod. Hosp.*, 102, 98—101 (January 1964).
288. Thompson J. D., Fetter R. B., The Application of Simulation to Hospital Planning, in: *Proceedings on Simulation in Business and Public Health*, First Annual Conference of American Statistical Association (New York Chapter Area) and Public Health Association of New York City, 1966.
289. Thompson J. D., Fetter R. B., Economics of Occupancy with Varying Mixes of Private and Other Patient Accommodations: A Simulation, *Health Services Res.*, 42—52 (Spring 1969).
290. Thompson J. D., Fetter R. B., Riedel D. C., CUISS, A Multipurpose Regional, Medical Information System Based on Three Inputs for Each Patient, Mimeographed, Yale University, Department of EPH, New Haven, Connecticut, 1969.
291. Thompson J. D., Avant O. W., Spiker E. D., How Queuing Theory Works for the Hospital, *Mod. Hosp.*, 94, 75—78 (March 1960).
292. Thompson J. D., et al., Predicting Requirements for Maternity Facilities, *Hospitals, JAMA*, 37, 45 (February 1963).
293. Thompson J. D., The Use of Computers in Hospitals, Report No. PB 195 629, National Technical Information Service, National Center for Health Services and Development, Springfield, Virginia, 1969.
294. Tsao R. F., A Second Order Exponential Model for Multidimensional Dichotomous Contingency Tables, with Applications in Medical Diagnosis, Report No. 320-2014, IBM Corporation, Scientific Center, Cambridge, Massachusetts, 1967.
295. Vallbona C., Applications of Computers for Hospital Usage, *J. Chronic. Dis.*, 19, 461—472 (April 1966).
296. VanBrunt E. E., The Kaiser-Permanente Medical Information System, in M. F. Collen (Ed.), *Medical Information Systems*, U.S. Department of Health, Education and Welfare, Health Services and Mental Health Administration, Washington, D.C., 1970, pp. 87—101.
297. Венедиктов Д. Д. Проблемы расстановки и использования медицинских кадров. — *Советское здравоохранение*, 1967, 8, с. 3—9.
298. Waaler H., Gesser A., Anderson S., The Use of Mathematical Models in the Study of the Epidemiology of Tuberculosis, *Am. J. Public Health*, 52, 1002—1013 (1962).
299. Warner D. M., Prawda J., A Mathematical Programming Model for Scheduling Nursing Personnel in a Hospital, Mimeographed, University of Michigan, Bureau of Hospital Administration, Ann. Arbor, Michigan, 1971.
300. Warner H. R., Toronto A. F., et al., A Mathematical Approach to Medical Diagnosis, *JAMA*, 177, 75—81 (1961).
301. Weinerman E. R., Research on Comparative Health Services Systems, *Med. Care*, 19, No. 3 (1972).
302. Weir R. D., Fowler G. B., Dingwall-Fordice I., The Prediction and Simulation of Surgical Admission in G. McLachlan and R. A. Schegog (Eds.), *Computers in the Service of Medicine*, Vol. 2, Oxford University Press, London, 1969, pp. 141—154.

303. Welch J. D., Hospital Applications, *Operation Res. Quart.*, 3, 8—10 (March 1952a).
304. Welch J. D., Some Research Into the Organization and Design of Hospital Outpatient Deapartments, *J. Roy. Sanitary Inst.*, 72, 298—311 (July 1952).
305. Welch J. D., Appointment Systems in Hospital Outpatient Departments, *Operational Res. Quart.*, 15, 224—231 (September 1964).
306. Welch J. D., Bailey N. T., Appointment Systems in Hospital Outpatient Department, *Lancet*, 1, 1105—1108 (May 1962).
307. Whitston C. W., An Analysis of the Problem of Scheduling Surgery, Parts 1 and 2, *Hosp. Management*, 99, 58 (April 1965); 99, 45—49 (May 1965).
308. Williams W. J., Covert R. P., Steele J. D., Simulation Modeling of a Teaching Hospital Outpatient Clinic, *Hospitals, JAHIA*, 41, 71—75 (November 1967).
309. Wilson H. H., Schulmeister M. R., Medical Systems Engineering, *Inquiry*, 5, 35—41 (September 1968).
310. Wiseman J., Cost-Benefit Analysis and Health Service Policy, in A. T. Peacock and D. J. Robertson (Eds.), *Public Expenditure: Appraisal and Control*, Oliver & Boyd, Edinburgh, 1963, pp. 128—145.
311. Wolfe H., Computerized Screening Device for Selecting Cases for Utilization Review, *Med. Care*, 5, 44—51 (1967).
312. Wolfe H., Iskander M., Raffin T., A Study of Obstetrical Facilities, in C.K. Chacko (Ed.), *The Recognition of Systems in Health Services*, Operations Research Society of America, Arlington, Virginia, 1969, p. 369—392.
313. Wolfe H., Young J. P., Staffing the Nursing Unit: Controlled Variable Staffing, *Nursing Res.*, 14, 236—245 (1965).
314. Wolfe H., Young J. P., Staffing the Nursing Unit: the Multiple Assignment Technique, *Nursing Res.*, 14, 299—303 (1965).
315. World Health Organization, Regional Office for Europe, Methods of Evaluating Public Health Programmes, Report on a Symposium, EURO 0375, Report on a Symposium, Kiel (November 1967). WHO Copenhagen, 1968
316. World Health Organization, Regional Office for Europe, Health Operational Research, EURO 0408, Report on a Seminar, Bucharest, September 1963, WHO, Copenhagen, 1969.
317. World Health Organization, Regional Office for Europe, The Regional Office's Activities in the Field of Medical Computing, EURO 3093, Report on a Working Group, Bratislava, February 1970; WHO, Copenhagen, 1970.
318. World Health Organization, Regional Office for Europe, The Use of Operational Research in Health Services, EURO 0408 (1), Report on a Meeting, Copenhagen, February 1970; WHO, Copenhagen, 1970.
319. World Health Organization, Development of a Methodology for Economic Assessment of Parasitic Disease Control Programmes, Report of Working Group PD/71.8, WHO Geneva, 1971.
320. World Health Organization, A Manual of Concepts and Procedures, Health Projection Formulation, WHO Document PSA/71.1, Rev. 1, Project System Analysis, WHO Geneva, 1971.
321. World Health Organization, Statistical Indicators for the Planning and Evaluation of Public Health Programmes, WHO Technical Report Series No. 472, Report of a WHO Scientific Group, WHO, Geneva, 1971.
322. World Health Organization, The Development of Studies in Health Manpower, WHO Technical Report Series No. 481, Report of a WHO Scientific Group, WHO, Geneva, 1971.
323. Young J. P., Stabilization of Inpatient Bed Occupancy Through Control of Admissions, *Hospitals, JAHIA*, 39, 41—48 (October 1965).

- 324. Young J. P., Administrative Control of Multiple-Channel Queuing Systems with Parallel Input Streams, *Operations Res.*, 14, 145—156 (January — February 1966).
- 325. Young J. P., Information Nexus Guides Information Systems, *Mod. Hosp.*, 106, 101—105 (February 1966).
- 326. Young J. P., A Conceptual Framework for Hospital Administrative Decisions Systems, *Health Services Res.*, 3, 79—95 (1968).
- 327. Young J. P., No Easy Solutions, in G. K. Chacko (Ed.), *The Recognition of Systems in Health Services*, Operations Research Society of America, Arlington, Virginia, 1969, pp. 395—398.

II-3. Процессы обучения

А. Муд¹⁾

ВВЕДЕНИЕ

В данной главе основное внимание уделяется анализу процесса обучения с помощью методов исследования операций. Однако с таким же успехом эти методы могут быть использованы и для решения других проблем, связанных с различного рода деятельностью учебных заведений. Это может быть составление расписаний и программ обучения, обеспечивающих более эффективное использование преподавателей и средств обучения; организация в школах соответствующих систем информации с учетом данных о каждом учащемся и индивидуальных планов обучения; решение организационных вопросов, связанных, например, с наличием длинных очередей у столов регистрации и в кафетериях школ или насыщенностью школьной жизни спортивными событиями; составление маршрутов движения школьных автобусов. Некоторые из этих проблем рассматриваются в других главах настоящего тома, а краткий перечень работ, посвященных этим проблемам, приводится в разд. 2 данной главы.

1. АНАЛИЗ СИСТЕМЫ ОБРАЗОВАНИЯ

Анализ системы обучения позволяет получить количественное описание этой системы и разработать наиболее эффективные методы управления. Возможность проведения подобного анализа обусловлена аналогией, существующей между процессами обучения и производственными процессами. Действительно, систему образования можно рассматривать как некоторую производственную операцию, для осуществления которой требуются капитальное оборудование (школьные здания, лаборатории, автобусы, библиотеки, спортивные площадки и т. д.), персонал (администрация, преподаватели, воспитатели, репетиторы и т. д.), а также соответствующие программы (учебные и факультативные). В этом случае для количественного описания системы образования ее можно представить в виде системы с входными (лица, прошедшие часть курса обучения) и выходными (лица, перешедшие на следующую ступень обучения) переменными. Разработка адекватной количественной модели, с помощью которой можно было бы установить точную взаимосвязь между этими переменными, имеет

¹⁾ Alex M. Mood, Ph. D., Professor Emeritus of Administration, Public Policy Research Organization, at the University of California in Irvine, California.

первостепенное значение для администрации учебных заведений и лиц, принимающих решения относительно управления системой образования, распределения ресурсов внутри системы и определения доли национального дохода на удовлетворение нужд образования. Такая модель, несомненно, позволила бы увязать между собой большое количество экспериментальных данных и, таким образом, разработать общие рекомендации для всей системы образования в целом. Кроме того, с помощью подобной модели можно было бы легко выявить все доводы за и против осуществления той или иной политики в области развития системы образования. Наконец, использование такой модели позволило бы подойти с научных позиций к обоснованию методики обучения, что в свою очередь дало бы возможности администрации учебных заведений использовать научно обоснованные методы управления.

1.1. Входные и выходные переменные системы образования

Первые попытки разработать глобальную модель системы образования были предприняты по инициативе министерства просвещения США [15]. Несмотря на то что ввиду своей чрезвычайной сложности эта модель не может быть реализована в настоящее время, она представляет определенный интерес, так как является начальным этапом дальнейших работ в этом направлении. Основными параметрами модели являются следующие шесть показателей:

- личные способности учащегося и его отношение к учебе,
- моральная поддержка учащегося со стороны членов семьи,
- моральная поддержка учащегося со стороны одноклассников,
- моральная и материальная поддержка системы образования со стороны совета попечителей,
- функциональные характеристики системы обучения,
- влияние государства на образование.

Приведенные показатели дают приближенное представление о взаимосвязях, существующих в системе образования (рис. 1). Эти показатели являются агрегированными, поэтому, прежде чем перейти к количественной оценке этих показателей, представим их в виде совокупности компонент.

Личные способности учащегося и его отношение к учебе

1. Успехи в приобретении соответствующих навыков и знаний:

- овладение средствами общения,
- изучение математики и алгоритмических языков,
- изучение естественных наук,

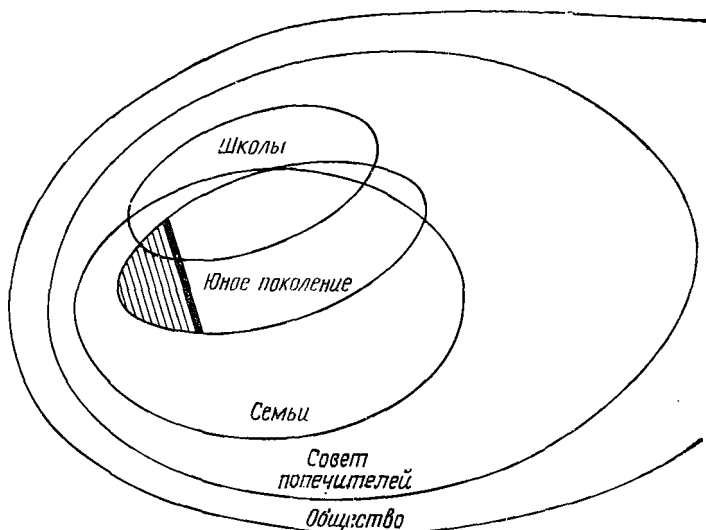


Рис. 1. Схематическое изображение взаимосвязей, существующих в системе образования.

изучение общественных наук,
изучение классических языков и литературы,
изучение гуманитарных наук.

2. Индивидуальное развитие учащегося:

компетентность,
чувство ответственности,
уверенность в себе,
способность к творчеству,
этичность,
честолобие,
умение сосредотачиваться,
рассудительность.

Моральная поддержка учащегося со стороны членов семьи (в физическом и физиологическом смысле)

создание всех условий для занятий дома,
воздействие на учащегося с целью воспитания у него сознательного отношения к учебе,
воздействие на школу в целях обеспечения хорошего уровня обучения,
выработка у учащегося соответствующих навыков обучения.

Моральная поддержка учащегося со стороны одноклассников:

позитивное отношение к успехам в учебе,
стремление достичь больших успехов в учебе, помощь в учебе.

Моральная и материальная поддержка системы образования со стороны совета попечителей:

социальное влияние,
уверенность в важности роли образования,
участие в школьных мероприятиях,
финансовая поддержка,
наличие средств для повышения культурного уровня
и отдыха.

Функциональные характеристики собственно системы образования

1. Преподаватели:

вера в «обучаемость» детей,
компетентность в вопросах организации учебной работы
в аудиториях,
знание предмета преподавания,
владение методикой преподавания,
умение контактировать с учащимися,
умение поощрять учащихся.

2. Администрация:

формирование целей обучения с учетом требований общества и стремлений учащегося,
компетентность в вопросах эффективной организации труда преподавателей и учащихся с точки зрения достижения намеченных целей,
компетентность в контроле учебных программ и в решении вопросов их корректировки в случае, когда достижение целей обучения находится под угрозой, умение поддерживать непосредственную связь между школой и обществом.

3. Учебный план и другие аспекты учебного процесса:

связь с направлениями в социологии и экономике, приемлемость с точки зрения способностей и желаний учащихся, возможность адаптации к индивидуальным запросам учащихся,
организация питания и оздоровительных мероприятий, наличие дополнительных программ учебного плана, а также программ отдыха и развлечений.

4. Средства обучения:

количество и качество аудиторий, лабораторий, библиотек и т. д.;
наличие учебных и технических средств обучения и организация отдыха и развлечений (книги, магнитные записи, проекторы, компьютеры, телевизионные установки, музыкальные инструменты, произведения искусства, спортивный инвентарь и т. д.);
пригодность окружающего пространства, прилегающей тер-

ритории и спортивных полей для нормального функционирования школы;
разнообразие средств, которыми располагает совет попечителей и которые используются школой в рамках культурно-развлекательной программы.

Участие государства в области образования:

возможность приобретения знаний,
государственные законы и политика, относящиеся к управлению школой, организации школьной жизни, учебным планам, учащимся и режиму посещаемости занятий,
предоставление ресурсов системе образования.

Первые четырнадцать компонент, характеризующие образовательный уровень учащихся, являются одновременно входными в начале учебного года и выходными (в конце года) переменными системы. Все остальные компоненты можно рассматривать как факторы, влияющие на процесс перехода учащегося в течение учебного года из начального состояния в конечное.

1.2. Способ оценки показателей процесса обучения

Если сформировано некоторое множество компонент процесса обучения, то можно переходить к составлению набора критериев оценки для каждой из компонент. Такие критерии в сочетании с весовыми множителями, приписываемыми каждой компоненте, могут быть использованы как мера «полезности» (или значимости) данной компоненты. Например, для компоненты «создание всех условий для занятий дома» к качеству критериев оценки могут быть использованы следующие: качество пищи, качество одежды, качество жилища, обеспеченность площадью, с кем из родителей живет ребенок, степень близости детей с родителями, степень откровенности родителей с детьми, возможность спокойно заниматься, наличие (или отсутствие) сильных эмоциональных потрясений, которые могли бы отвлекать учащегося от школьных занятий, необходимость совмещать учебу с работой.

При качественной оценке компонент, а следовательно, и показателей не следует ориентироваться на максимально хорошие условия жизни. Это означает, что при рассмотрении критерия «качество пищи» необходимо исходить из условия существования системы сбалансированного питания, а при обсуждении критерия «качество одежды» максимальная оценка по соответствующей шкале должна учитывать степень чистоты и удобства одежды, способность сохранять тепло, а также степень эстетического наслаждения, которое получает ребенок от той или иной формы одежды.

Если рассмотреть другую компоненту, а именно «воздействие членов семьи на учащегося с целью воспитания сознательного отношения к учебе», то в этом случае критериями оценки могут быть: поощрение за успехи в учебе; меры воздействия при низкой успеваемости; осведомленность родителей о требованиях, предъявляемых к выполнению домашних заданий, и контроль за их выполнением; обеспечение родителями регулярной посещаемости учащегося школы; участие родителей в формировании желания освоить перспективные учебные программы; участие родителей в формировании желания сделать карьеру, требующую получения всестороннего образования.

Любой критерий компоненты типа «меры воздействия при низкой успеваемости» предполагает ее качественную оценку с помощью соответствующей шкалы, которая может иметь следующие градации: строгий режим в случае низкой успеваемости, суровые меры наказания, лишение важных привилегий и прав; незначительные наказания, разумные, обоснованные меры воздействия; незначительное беспокойство; полная удовлетворенность при беглом просмотре оценок.

Ни один из шести указанных показателей системы образования, так же как и ни одна из перечисленных компонент, не может быть оценен непосредственно. Поэтому для их количественного описания используются соответствующие числовые индексы или другие критерии оценки. Например, в системе образования для выявления уровня знаний учащихся широко применяются всевозможные тесты. Кроме того, социологи разработали социально-экономический критерий. Хотя некоторые компоненты могут быть достаточно просто оценены на основании имеющихся данных, тем не менее для большинства компонент способы оценки до сих пор не разработаны. Поэтому для каждой такой компоненты составляется набор критериев, на основе которого разрабатываются тесты или анкеты, позволяющие получить удовлетворительную оценку данной компоненты. В большинстве случаев подобные оценки являются весьма приближенными. Это объясняется тем, что ввиду сложности и абстрактности компонент не представляется возможным составить исчерпывающий набор критериев. По-видимому, в ряде случаев даже такая оценка является вполне приемлемой. Проиллюстрируем это на следующем примере

Пример. Для формирования критерия «стоимость жизни» статистическое бюро США составляет довольно длинный перечень цен на товары и услуги с целью дать более полное представление о том, чем может располагать американская семья. Однако вместо такого перечня, очевидно, целесообразнее было бы перечислить цены, например, только на хлеб и автомобили и указать плату за квартиру и лечение в стационаре. Если при этом подобрать соответствующие весовые множители, то полученная приближенная оценка будет хорошо коррелировать с полной оценкой уровня жизни. При этом

надо иметь в виду, что наличие указанной корреляции во многих случаях недоказуемо и обосновывается лишь здравым смыслом.

Отметим следующее важное обстоятельство. Широкое использование количественных методов при исследовании системы образования с необходимостью предполагает, что между исследователями достигнуто соглашение относительно набора критериев оценки каждого показателя изучаемой системы. Более того, выбранные для области образования критерии должны быть такими же существенными, как, например, «уровень стоимости жизни», «общий национальный продукт», «индекс оптовых цен» и другие показатели в области экономики. Этого необходимо добиваться, несмотря на то что критерии оценки системы образования являются менее определенными и соответственно менее обоснованными, чем критерии оценки экономики.

При наличии общепризнанных способов оценки основных показателей процесса обучения исследователи смогут выбрать критерии для основных объектов исследования (учащиеся, учителя, родители, школьные администраторы, учебные планы и т. д.) и осуществить проверку их адекватности. В результате этого может быть построена некоторая система количественных оценок основных показателей процесса обучения. Следует иметь в виду, что строгое обоснование выбранной системы количественных оценок не означает, что такие оценки неизменны. Они могут периодически пересматриваться (так же, кстати, как и экономические критерии), с тем чтобы отражать эволюционные изменения, происходящие в жизни, и в частности в целях и направленности процесса обучения.

Модель системы образования должна быть полной, т. е. по возможности в ней должны быть учтены все существенные входные или выходные параметры. В противном случае исказится зависимость между параметрами модели, что, естественно, приводит к совершенно неправильному представлению о влиянии входов модели на ее выходы. Особенно сложная ситуация складывается при рассмотрении выходных переменных исследуемой системы. Дело в том, что в области образования разработаны способы оценки только познавательных способностей учащегося и достигнутого им уровня знаний (т. е. для первых шести компонент образовательного состояния учащегося в приведенном выше списке). Что же касается компонент показателя «индивидуальное развитие учащегося», то для них пока что аналогичных способов оценки не существует. Это обстоятельство приводит к тому, что в области образования при анализе проблем стратегического и оперативного управления наблюдается тенденция уделять наибольшее внимание тем компонентам, для которых разработаны способы оценки, и пренебрегать теми, для которых такие способы

еще не найдены. Интересно отметить, что указанная тенденция существует, несмотря на широко распространенное мнение о том, что компоненты индивидуального развития имеют важное практическое значение.

1.3. Регрессионные модели

Можно надеяться, что будет разработана довольно полная модель процесса обучения, математическое описание которой должно включать систему уравнений, характеризующих связь между входными и выходными переменными, а также связи только между входными и только между выходными переменными. По-видимому, такая модель будет аналогична модели экономики США, состоящей из 100 уравнений, разработанных Институтом Брукинса [4]. Однако в настоящее время разработка подобной модели затруднительна ввиду отсутствия единого мнения относительно выбора переменных, недостаточного понимания связывающих их закономерностей и отсутствия информации, необходимой для выявления этих закономерностей. В результате были созданы самые простые модели, для описания которых обычно используется одно уравнение регрессии. Это уравнение должно воспроизводить зависимость между оценками, получаемыми с помощью тестов по проверке успеваемости, и несколькими независимыми переменными, характеризующими степень влияния на учащегося семьи и школы.

Используя приведенные выше входные и выходные показатели процесса обучения, можно представить общую регрессионную модель этого процесса в виде $S_1 = f(S_0, P, F, L, C, Y)$, где S_0 — вектор (14 компонент), характеризующий начальный уровень знаний учащегося, S_1 — вектор (14 компонент), характеризующий достигнутый уровень знаний учащегося, P, F, L, C и Y — критерии оценок основных показателей процесса обучения (ученики, семья, школа, совет попечителей, общество) и функция f — линейна по коэффициентам регрессии. Каждую из пяти оценок обычно представляют в виде линейной функции регрессии от компонент, которые в свою очередь записываются в виде линейных функций регрессии от соответствующих показателей. Функция f , как правило, представляет собой взвешенную сумму степеней и произведений переменных, весовые множители которых совпадают с коэффициентами регрессии. Отметим, что в тех случаях, когда S состоит из одной компоненты, часто удобно ввести шкалу критерия оценки и таким образом автоматически обеспечить линейность функции f (2, 15).

При разработке упомянутых регрессионных моделей обычно используются независимые переменные, которые соответствуют основным показателям процесса обучения, и не используются

ортогональные переменные, получаемые с помощью факторного анализа. Это объясняется тем, что интерпретация ортогональных переменных вызывает определенные трудности, так как такие переменные, как правило, представляют собой комбинацию различных компонент основных показателей. Регрессионная модель процесса обучения предназначается не только для исследователей систем, но и для административных работников системы образования, членов советов попечителей и представителей законодательных органов. Поэтому при описании модели и интерпретации результатов, получаемых с ее помощью, необходимо пользоваться терминологией, доступной для понимания такого рода специалистами. В противном случае практическая ценность модели будет сомнительна.

1.4. Разделение дисперсии

Между основными показателями процесса обучения, по-видимому, существует сильная взаимосвязь, так как все усилия по разделению их влияния на процесс пока что не увенчались успехом. Вполне вероятно, что такое разделение вообще невозможно. Действительно, в процессе обучения возникает такая тесная взаимосвязь показателей семьи и школы, что едва ли можно надеяться на полное выделение влияния каждого из них.

Предположим, что для основных показателей процесса обучения выбраны пять критериев оценки. Исследуем с помощью линейной регрессии по этим критериям множество A данных об успеваемости по арифметике в пятом классе. Уравнение регрессии будет иметь вид

$$A = aP + bF + cL + dC + eY,$$

где a , b , c , d и e — коэффициенты регрессии.

Величина дисперсии D оценки успеваемости A будет равна $(1 - R^2) D$, где R^2 — коэффициент множественной корреляции.

Как уже отмечалось, между выбранными критериями существует сильная взаимная корреляция, поэтому любые четыре из них будут уменьшать дисперсию почти на ту же величину, что и все пять, а использование любых трех не намного хуже, чем использование любых четырех.

Один из способов оценки влияния на успеваемость одной компоненты или совокупности компонент состоит в разделении дисперсии [3, 15, 17]. Идею такого разделения лучше всего продемонстрировать на двух независимых переменных.

Пусть D есть дисперсия оценки успеваемости A . Предположим, что регрессия только по переменной P приводит к уменьшению дисперсии на величину R_P^2 , а регрессия только по переменной F приводит к уменьшению дисперсии на величину R_F^2 .

регрессия по обоим переменным P и F , т. е. представление $A = aP + bF$, приводит к уменьшению дисперсии на величину R_{PF}^2 . Тогда величина R_{PF}^2 может быть представлена в виде трех составляющих: $(R_{PF}^2 - R_P^2)$ — относительная величина дисперсии D , обусловленная только переменной F ; $(R_{PF}^2 - R_F^2)$ — относительная величина дисперсии D , обусловленная только переменной P ; оставшаяся доля дисперсии, обусловленная как переменной P , так и переменной F , определяется как $R_{PF}^2 - (R_{PF}^2 - R_P^2) - (R_{PF}^2 - R_F^2) = R_P^2 + R_F^2 - R_{PF}^2$ (эта величина характеризует взаимосвязь факторов P и F).

В общем случае, когда имеется n независимых переменных, можно провести аналогичное разбиение дисперсии на $(2^n - 1)$ частей. Если при этом для обозначения переменных вместо P , F , L и т. д. использовать целые числа $1, 2, \dots, n$, то можно довольно просто получить соответствующие формулы. Проиллюстрируем это на следующем примере.

Пусть требуется определить относительную величину дисперсии, обусловленную совместным действием переменных 2, 4 и 7 при общем числе переменных, равном 8. Для этого необходимо раскрыть скобки в выражении

$$-(1 - \chi_2)(1 - \chi_4)(1 - \chi_7)\chi_1\chi_3\chi_5\chi_6\chi_8$$

и заметить в полученной сумме каждое слагаемое символом R^2 с индексами, совпадающими с индексами входящих в данное слагаемое (сомножителей) переменных χ . Например, при четырех переменных доля дисперсии, обусловленная только переменной 2, определяется как

$$U_2 \approx -(1 - \chi_2)\chi_1\chi_3\chi_4 = R_{1234}^2 - R_{134}^2,$$

а относительная величина дисперсии, которая может быть обусловлена либо переменной 1, либо переменной 3, определяется как

$$E_{13} \approx -(1 - \chi_1)(1 - \chi_3)\chi_2\chi_4 = -R_{1234}^2 + R_{124}^2 + R_{134}^2 - R_{24}^2.$$

Используя рассмотренный прием при получении формулы для доли дисперсии, обусловленной совместным действием всех n независимых переменных, необходимо после раскрытия скобок в выражении

$$-(1 - \chi_1)(1 - \chi_2)(1 - \chi_3) \dots (1 - \chi_n)$$

вычесть из него -1 .

1.5. Примеры, иллюстрирующие эффективность использования регрессионной модели

Возможность использования в широких масштабах регрессионной модели процесса обучения представилась в связи с проведением в 1964 г. через конгресс США Закона о гражданских правах. В разделе закона под номером 402 содержалось следующее требование, обращенное к специальному советнику по вопросам образования: изучить «проблему недостаточного обеспечения равных прав на образование для лиц с различным цветом кожи, различной расовой, религиозной и национальной принадлежности». Соответствующее исследование было проведено в 1965 г., а спустя год был опубликован итоговый отчет [3]. В ходе проведенного исследования была собрана громадная и исчерпывающая информация, относящаяся к школам, учащимся, семьям учащихся и школьным советам попечителей. Эта информация была сформирована из данных обследования 600 000 учащихся и 4000 государственных школ. При этом были опрошены не только учащиеся, но также преподаватели, старшие преподаватели и директора школ. Кроме того, преподаватели и учащиеся участвовали в тестировании. Приведенные ниже результаты выбраны из тщательных исследований [40, 41], выполненных уже после того, как был опубликован отчет [3]. В этом исследовании четыре оценки успеваемости (относящиеся к третьей, шестой, девятой и двенадцатой ступеням обучения) объединялись в единый критерий, а весовые множители учитываемых факторов определялись на основе данных анкет по максимальному коэффициенту корреляции между показателями и критерием. Для того чтобы функция регрессии была линейна, показатели, фигурирующие в анкетах, располагались по их относительной значимости.

При исследовании дисперсии оценки успеваемости в национальном масштабе оказалось, что около одной трети величины этой дисперсии обуславливается различиями между школами и около двух третей — различиями между учащимися в рамках отдельных школ. Экспериментальные данные, приведенные ниже, относятся только к черной составляющей дисперсии. В этих данных соотносятся результаты регрессии средних сведений об успеваемости со средними значениями критериев основных показателей процесса обучения в школе.

Таблица 1

Относительная величина дисперсии оценки успеваемости, обусловленная регрессией по показателям P и L

Составляющие дисперсии	Третьей класс	Шестой класс	Девятый класс	Двенадцатый класс
R_P^2	0,5254	0,7884	0,8207	0,8221
R_L^2	0,4901	0,7322	0,7601	0,7865
R_{PL}^2	0,5646	0,8296	0,8662	0,8617
Только P	0,0745	0,0974	0,1061	0,0752
Только L	0,0392	0,0412	0,0455	0,0366
Как P , так и L	0,4509	0,6910	0,7146	0,7469

В табл. 1 приводятся данные об уменьшении дисперсии в результате регрессии отдельно (первые две строки таблицы) или совместно (третья строка таблицы) по двум основным показателям

учебного процесса: P (сокурсники) и L (школа). В последних трех строках табл. 1 приводятся данные об относительной величине дисперсии, обусловленной соответственно только переменной P , только переменной L и совместным влиянием переменных P и L .

Из анализа данных последнего столбца табл. 1 следует, что за счет регрессии только по показателю P величина дисперсии оценки уменьшается на 82%, за счет регрессии только по показателю L — на 79%. Регрессия по обоим показателям дает уменьшение величины дисперсии на 86%. Следовательно, взаимосвязь показателей P и L очень мала. Поэтому основное уменьшение дисперсии может быть связано либо с изменением характеристик окружения учащегося, либо характеристик школы.

Таблица 2

Относительная величина дисперсии оценки успеваемости, обусловленная регрессией по показателям P , T и O

Составляющие дисперсии	Третий класс	Шестой класс	Девятый класс	Двенадцатый класс
R_P^2	0,5254	0,7884	0,8257	0,8221
R_T^2	0,4751	0,7145	0,7195	0,7441
R_O^2	0,1129	0,1588	0,2543	0,3178
R_{PT}^2	0,5569	0,8197	0,8449	0,8399
R_{PO}^2	0,5336	0,8028	0,8495	0,8444
R_{TO}^2	0,4901	0,7322	0,7601	0,7865
R_{PTO}^2	0,5646	0,8296	0,8662	0,8617
Только P	0,0745	0,0974	0,1061	0,0752
Только T	0,0310	0,0268	0,0167	0,0173
Только O	0,0077	0,0099	0,0213	0,0218
Как P , так и T	0,3462	0,5466	0,4891	0,4514
Как P , так и O	0,0073	0,0058	0,0193	0,0206
Как T , так и O	0,0005	0,0045	0,0075	0,0005
Как P , так и T , так и O	0,0974	0,1386	0,2062	0,2749

Данные табл. 2 характеризуют относительную важность показателя уровня преподавания по сравнению с другими показателями процесса обучения в школе, причем показатель L представлен в виде двух компонент T и O , где T — качество преподавания, а O включает в себя все остальные показатели качества, такие, как размеры классов, время постройки школьного здания, наличие лабораторий и библиотек, обеспеченность учебниками, укомплектованность радиотелевизионной аппаратурой и другим

оборудованием, содержание учебных планов, состав администрации, пригодность прилегающей территории для спортивных занятий, факультативные занятия и т. д. Первые семь строк табл. 2 содержат квадраты коэффициентов множественной корреляции, обусловленных регрессией по P , T и O , затем по каждой паре этих показателей и наконец сразу по всем трем показателям.

Сравнение данных, приведенных во второй строке табл. 1 и во второй и третьей строках табл. 2, позволяет сделать следующий важный вывод: при анализе влияния фактора «школа» на процесс обучения можно ограничиться рассмотрением показателя качества преподавания, поскольку другие показатели школы влияют на величину дисперсии в значительно меньшей степени. Действительно, сравнивая данные третьей и последней строк табл. 2, можно видеть, что влияние показателя O значительно сильнее, чем показателей T и P .

1.6. Приращения дисперсии в сравнении с коэффициентами регрессии

Использование для оценки значимости показателей (или их компонент) относительной величины дисперсии, обусловленной регрессией по соответствующим переменным, а не коэффициентов регрессии, свидетельствует о несовершенстве используемой модели. Исследователи могли бы вычислить коэффициент регрессии для любой основной переменной, например соответствующей показателю P , а затем вычислить коэффициенты регрессии, относящиеся к отдельным компонентам этого показателя. Однако такой способ не является оправданным, поскольку учитываемые компоненты не полностью характеризуют показатели. Что касается приращений дисперсий для показателя, то оно оказывается менее чувствительным к изменению состава его компонент.

Трудность определения коэффициентов регрессии объясняется еще и существованием сильной корреляции между переменными, которая приводит к нестабильности коэффициентов регрессии. Согласно данным еще не опубликованной работы Битена, для соответствующей совокупности переменных нестабильность коэффициента множественной корреляции меньше, чем нестабильность коэффициентов регрессии.

Следует иметь в виду, что характеристики взаимосвязи некоторых совокупностей переменных могут принимать отрицательные значения (хотя в табл. 1 и 2 все величины положительны). Это происходит в тех случаях, когда две независимые переменные отрицательно коррелируют между собой, а каждая из них положительно коррелирует с некоторой зависимой от них переменной.

В связи с этим можно привести следующий пример, рассмотренный Битеном. Мастерство игры в футбол положительно коррелирует как со скоростью, так и с весом игрока, но скорость и вес коррелируют между собой отрицательно. Поэтому при вычислении регрессии переменной «мастерство» по независимым переменным «скорость» и «вес» величина, характеризующая их взаимосвязь, имела бы отрицательное значение. Это объясняется тем, что влияние одной из переменных значительно увеличивается при включении в регрессию другой переменной. Так, если для оценки мастерства игры в футбол проводится регрессия по группе переменных, которая в одном случае содержит, а в другом не содержит переменную «вес» и в обоих случаях содержит переменную «скорость», то уменьшение величины дисперсии ошибки в первом случае будет существенно большим, чем во втором.

1.7. Анализ структурных моделей процесса обучения

Главной целью многочисленных работ по изучению возможностей моделирования процесса обучения является создание такой модели, которая достаточно полно учитывала бы причинно-следственные связи, существующие между входными и выходными показателями процесса. Такая модель должна была бы состоять из некоторого множества уравнений, адекватно описывающих основные причинно-следственные связи и включающих четко определенные и точно измеренные переменные, т. е. предполагаемая модель должна быть аналогична моделям, используемым в экономике (например, модели, описываемые в работах [5, 6]). Достаточно общее структурное описание модели процесса обучения дано в работе [7].

Первые шаги в этом направлении были предприняты в работах [1, 9, 13]. Так, например, в работе [13] для частичного анализа информации, относящейся к проблеме реализации равных прав на получение образования, используется следующая система уравнений [9]:

$$A = b_1 S + c_1 G + \sum_i d_{1i} X_i,$$

$$S = a_2 A + \sum_i d_{2i} X_i,$$

$$G = a_3 A + \sum_i d_{3i} X_i,$$

где A — оценка успеваемости, S — оценка отношения учащегося к собственным способностям и шансам на успех, G — оценка желания учащегося получать в школе высокие баллы. В обосновании данной модели лежат следующие соображения. Прежде всего, включение в нее только одного, первого, уравнения было бы недостаточным, поскольку показатель S в определенной степени зависит от успешного обучения, т. е. от показателя A . Следовательно, в модель необходимо включить второе уравнение. Точно так же наряду с наличием влияния на показатель A по-

казателя G имеет место и обратное влияние. Поэтому третье уравнение также необходимо включить в модель.

Если пользоваться терминологией экономики, переменные, соответствующие факторам A , S и G , следует называть эндогенными (внутренними). В той же терминологии переменные X , соответствующие показателям образовательной подготовки учащегося и характеристикам школы, следует называть экзогенными (внешними). Буквы a , b , c и d обозначают коэффициенты, которые

Таблица 3

Весовые множители переменных в моделях, описывающих процесс обучения

Независимые (экзогенные) переменные	Коэффициент значимости	
	простая регрессионная модель	модель, не учитывающая переменные S и G
Пол учащегося	0,8	0,8
Возраст учащегося	-7,5	-6,8
Пробывание ребенка в детском саду до поступления в школу	2,1	2,1
Размеры семьи	-0,5	-0,6
Материальное положение семьи	1,4	1,3
Наличие у учащегося матери или лица, выполняющего ее функции		-0,4
Занятость матери работой		-0,3
Наличие у учащегося отца или лица, выполняющего его функции		-0,5
Образовательный уровень отца	0,3	0,4
Оценочный тест преподавателя		0,3
Характеристика учебного заведения, которое закончил преподаватель		7,7
Опыт преподавателя	0,4	0,8
Желание преподавателя перейти работать в другую школу		1,0
Расовое различие между преподавателем и учащимися	-2,7	
Текучесть кадров (преподавателей)		-0,2
Объем фондов школьной библиотеки	0,3	0,5
Наличие в школе актового и спортивного залов и кафетерия	0,7	

Из всех возможных экзогенных переменных, характеризующих показатель A , в эту таблицу включены только те, которые дают наиболее существенный вклад, если показатель оценивается по величине остаточной суммы квадратов.

могут быть оценены по исходной информации. Описанная модель позволяет оценить влияние фактора A на факторы S и G . Кроме того, если подставить второе и третье уравнения модели в первое, получим выражение, описывающее зависимость фактора A только от переменных X , т. е.

$$A = b_0 + \sum_i b_i X_i.$$

Это обеспечивает более реальную оценку зависимости показателя A от экзогенных переменных, чем аналогичное представление, полученное с помощью обычной регрессии.

В приводимой ниже табл. 3 сравниваются два способа представления показателя A .

ЛИТЕРАТУРА

1. Adelman I., Parti M., The Determinants of Student Achievement, Research Bulletin, Rutgers University, New Brunswick, New Jersey, 1972.
2. Beaton A. E., Criterion Scaling of Questionnaire Items, *Socio-Econ. Plan.*, 2, 335—362 (1969).
3. Coleman J. S., et al., Equality of Educational Opportunity, U.S. Government Printing Office (FS 5.238:38004) Washington, D. C., 1966.
4. Dusenberry J. S., et al., The Brookings Econometric Model of the United States, Rand McNally, Chicago, 1965.
5. Goldberger A. S., Econometric Theory, Wiley, N.Y., 1964.
6. Johnson J., Econometric Methods, McGraw-Hill, N.Y., 1963.
7. Jöreskog K. C., et al., A General Program for Analysis of Covariance Structures Including Generalized MANOVA, Educational Testing Service, Princeton, N.J., 1971.
8. Kershaw J. A., McKean R. N., Systems Analysis and Education, The Rand Corporation, Santa Monica, Calif., 1959.
9. Levin H. M., A New Model of School Effectiveness, Do Teachers Make A Difference?, U.S. Government Printing Office (HE 5.258 : 58042) Washington, D.C., 1970, pp. 55—78.
10. Mayeske G. W., et al., A Study of Our Nation's Schools, U.S. Government Printing Office (HE5.210 : 10085), Washington, D.C., 1972.
11. Mayeske G. W., et al., A Study of the Achievement of Our Nation's Students, U.S. Government Printing Office (1780—1055) Washington, D.C., 1973.
12. Mayeske G. W., et al., A Study of the Attitude Toward Life of Our Nation's Students, U.S. Government Printing Office (1780—01146) Washington, D.C., 1973.
13. Michelson S., The Association of Teacher Resources with Children's Characteristics, Do Teachers Make a Difference?, U.S. Government Printing Office (HE 5.258 : 58042) Washington, D.C., 1970, pp. 120—168.
14. Michelson S., Equal School Resource Allocation, *J. Human Resources*, 7, 283—306 (1972).
15. Mood A. M., Macro-Analysis of the American Educational System, *Operations Res.*, 17, 770—784 (1969).

16. Michelson S., Partitioning Variance in Multiple Regression Analyses as a Tool for Developing Learning Models, *Am. Educ. Res. J.*, 8, 191—202 (1971).
17. Newton R. G., Spurrell D. J., A Development of Multiple Regression for Analysis of Routine Data, *Applied Statistics*, 16, 51—64 (1967).

2. АННОТИРОВАННАЯ БИБЛИОГРАФИЯ

Настоящий раздел, по существу, носит библиографический характер, так как в нем приводится перечень работ, посвященных применению методов исследования операций для решения проблем, которые не рассматривались в предыдущем разделе.

2.1. Образование как сектор экономики

Число опубликованных работ в области экономики, в которых система образования рассматривается как сектор экономики (и в обсуждаемых экономических моделях ей отводится существенная роль), столь велико, что не представляется возможным перечислить все работы. Поэтому в настоящем разделе приводятся как классические работы, так и ряд работ, выполненных совсем недавно, по которым, кстати, можно составить полное представление о современном уровне исследований в рассматриваемой области. Кроме того, в библиографию включен достаточно полный обзор, составленный Хофнером [10].

Теодор В. Шульц, основоположник работ, считает, что с экономической точки зрения образование есть не что иное, как «капитал». Это означает, что затраты на обучение следует рассматривать как инвестиции в производительный капитал [16]. В ряде исследований [11] эффект от такого инвестирования оценивался путем определения дополнительной заработной платы рабочих и служащих, прошедших дополнительную подготовку. В то же время классическое исследование [2] показало, что дополнительный заработок, часто приписываемый более высокому образовательному уровню, на самом деле должен быть отнесен за счет более высоких способностей. Действительно, более способные люди хорошо учатся в школе и, как правило, имеют склонность к продолжению образования. Поэтому весьма вероятно, что они зарабатывали бы больше, чем менее способные люди, и без получения дополнительного образования.

В США ведутся интенсивные исследования экономических аспектов проблем образования для оценки Федеральных программ, предназначенных для оказания помощи различным группам населения в повышении их образовательного и профессионального уровня. В библиографию включен ряд работ, связанных с такими оценками.

1. Ackoff R. L., Toward Strategic Planning of Education, Efficiency in Resource Utilization in Education, OECD, Paris, 1969.

Описывается сложная модель для анализа роли образования. Модель является углубленным вариантом ранее разработанной модели [14], которая, однако, не была в достаточной степени детализирована и опробована.

2. Becker G. S., Human Capital, Columbia Univ. Press, N.Y., 1964.
3. Bellante D. M., A Multivariate Analysis of a Vocational Rehabilitation Program, *J. Human Resources*, 7, 226—241 (1972).

Приводятся соотношения между затратами и выгодами для различных групп лиц, участвующих в процессе обучения.

4. Blaug M., The Private and Social Returns of Investment in Education: Some Results for Great Britain, *J. Human Resources*, 3, 330—346 (1967).
- Проводится анализ социального эффекта образования.

5. Danieri A., Mechling J., Direct Marginal Productivity of College Education in Relation to College Aptitude of Students and Production Costs of Institutions, *J. Human Resources*, 5, 51—70 (1970).

Оцениваются заработки студентов с разными способностями, посещающих институты различного профиля.

6. Denison E. E., Measuring the Contribution of Education to Economic Growth, The Residual Factor and Economic Growth, OECD, Paris, 1964, pp. 13—55.

Вклад образования в экономику оценивается величиной «остатка», который получается после исключения из общего вклада той его части, которая связана с обычными экономическими показателями.

7. Eckaus R. S., Estimating the Returns to Education, Carnegie Commission on Higher Education, Berkeley, California, 1973.

Используется показатель повременной оплаты, а не показатель годового дохода.

8. Hanoch G., An Economic Analysis of Earnings and Schooling, *J. Human Resources*, 3, 310—329 (1967).

Оценивается предельная норма дохода в течение одного учебного года. При этом используются данные переписи 1960 г.

9. Hansen W. L., Weisbrod B. A., The Distribution of Costs and Direct Benefits of Public Higher Education: The Case of California, *J. Human Resources*, 4, 176—191 (1969).

Исследуется распределение общественных фондов в системе высшего образования для различных статей поступлений.

10. Hüfner K., Economics of Higher-Education and Educational Planning: A Bibliography, *Socio-Eco. Plan. Sci.*, 2, 25—101 (1969).

Содержит 1333 работы, которые распределены по следующим разделам: Сборники библиографии; Общие обзоры; Обзоры, относящиеся к отдельным странам или институтам; Организационные вопросы; Университет как производственная ячейка; Анализ издержек; Финансирование; Государственные ассигнования; Частные ассигнования; Социально-экономические факторы; Концепция «человеческий капитал»; Экономическая эффективность затрат в системе образования; Образование и экономический рост; Государственное планирование и использование кадров; Потребность в кадрах определенной квалификации и др.

11. Miller H. P., Annual and Lifetime Income in Relation to Education: 1937—1959, *Amer. Eco. Rev.*, 50, 962—986 (1960).
12. Mincer J., On-The-Job Training: Costs, Returns and Implications, *J. Political Economy*, 7 (Supplement), 50—59 (1962).
13. Mincer J., Investment in Human Capital and Personal Income Distribution, *J. Political Economy*, 66, 281—302 (1958).
14. Platt W. J., Educational Policy for Economic Growth, Stanford Research Institute, Menlo Park, Calif., 1961.
15. Ribich T. I., Education and Poverty, Brookings Institution, Washington, D.C., 1968.

Для оценки некоторых программ профессиональной переподготовки, предпринятых федеральным правительством в качестве меры повышения уровня жизни, проводится анализ затрат и результатов.

16. Schultz T. W., Capital Formation by Education, *J. Political Economy*, 68, 571—583 (1960).
17. Schultz T. W., Rise in the Capital Stock Represented by Education in the U.S., 1900—1957, Economics of Higher Education, S. J., Mushkin (Ed.), Dept HEW, U.S. Government Printing Office, Washington, 1962, pp. 93—101.
18. Schultz T. W., The Economic Value of Education, Columbia Univ. Press, N.Y., 1962.

19. Sen A. K., Education, Vintage, and Learning by Doing, *J. Human Resources*, 1, 3—21 (1966).

Оценивается экономическая эффективность обучения непосредственно на рабочем месте.

20. Somers G. G., Stomdorfer E. W., A Cost-Effectiveness Analysis of the In-School and Summer Neighborhood Youth Corps Program, *J. Human Resources*, 7, 446—459 (1972).

Дается оценка национальной программы обучения.

21. Spiegelman R. G., A Benefit-Cost Model to Evaluate Educational Programs, *Socio-Eco. Plan. Sci.*, 1, 443—460 (1968).

Приводится модель для оценки учебных программ, состоящая из 13 уравнений, включающих в качестве параметров общественные и частные затраты, а также общественные и личные выгоды. Модель применяется к программе раздела 1 Закона о начальном и среднем образовании с использованием данных, относящихся к Сан-Франциско.

22. Weisbrod B. A., Education and Investment in Human Capital, *J. Political Economy*, 7 (Supplement), 106—123 (1962).

2.2. Стохастические модели процесса обучения

Данное направление зародилось около 20 лет назад, когда была принята попытка разработать простые вероятностные модели, с помощью которых можно было бы определить, сколько раз необходимо повторить один и тот же материал, для того чтобы его запомнить. Работа [2] явилась первым систематизированным описанием моделей такого типа. Следует отметить, что развитию данного направления исследований в значительной мере способствовал тот факт, что на протяжении почти десяти лет оно имело свой собственный журнал (журнал математической физиологии). Кстати, тем, кто желает ознакомиться с современным состоянием дел в области разработки стохастических моделей процесса обучения, лучше всего начать изучение вопроса с доскональной проработки последних томов этого журнала.

1. Bush R. R., Estes W. K., *Studies in Mathematical Learning Theory*, Stanford Univ. Press, Stanford, Calif., 1959.
2. Bush R. R., Mosteller F., *Stochastic Models of Learning*, Wiley, N.Y., 1955.
3. Luce R. D., Bush R. R., Galanter E. (Eds.), *Handbook of Mathematical Psychology*, Vols. I, II, and III, Wiley, 1963—1965.
4. *Journal of Mathematical Psychology*, Vols. 1—10, 1964—1973.

2.3. Анализ процесса формирования учебных групп

Планирование учебного процесса, выполнение его элементов, а также состав администрации учебных заведений существенным образом зависят от количественного и качественного состава учащихся. В связи с этим большое внимание уделяется разработке моделей формирования учебных групп и прогнозированию числа учащихся, переходящих из одной образовательной категории в другую. Наиболее широко используются регрессионные модели, и в особенности авторегрессионные модели. Примером последней является модель, с помощью которой можно определить число учащихся, зачисляемых на предпоследний курс следующего года, в зависимости от числа учащихся, зачисленных на второй курс текущего года. Помимо моделей указанных типов применяются также потоковые модели. Так как при разработке потоковых моделей переход учащегося из одной категории в другую рассматривается как некоторые входной или выходной поток системы обучения, то для математического описания этих моделей используется система уравнений сохра-

нения потока. Важность моделей данного типа объясняется наличием большого числа потоков как внутри системы образования, так и вне ее. Например, выпускники высшей школы образуют входные потоки для колледжей, школ профессионального обучения, армии и для сфер занятости и незапнятости.

Поскольку процесс обучения имеет годовой цикл, то наряду с упомянутыми моделями широкое распространение получили различные варианты марковской модели. При этом предполагается, что учащиеся распределены по множеству состояний и что изменение этих состояний происходит мгновенно, т. е. в момент окончания учебного года. Указанное изменение состояний описывается с помощью некоторой матрицы переходов, аналогичной вероятностной матрице стохастических марковских процессов. Однако в данном случае элементы матрицы переходов имеют смысл не вероятностей перехода, а представляют собой относительное число учащихся, переходящих из одного определенного состояния в другое. Поэтому такие модели следовало бы называть моделями типа марковских, а не марковскими.

1. Armitage P. H., Smith C. S., Alper P., *Decision Models for Educational Planning*, Allen Lane, London, 1969.
2. Armitage P. H., Phillips C. M., Davies J., *Towards a Model of the Upper School System*, *J. Roy. Stat. Soc.*, 133, 166—205 (1970).
3. Baisuck A., Wallace W. A., *A Computer Simulation Approach to Enrollment in Higher Education*, *Socio-Eco. Plan. Sci.*, 4, 365—381 (1970).

Для анализа данных, относящихся к Политехническому институту Пенселера, применяются модели марковского типа.

4. Balachandran K. R., Gerwin D., *Variable-Work Models for Predicting Course Enrollment*, *Operations Res.*, 21, 823—834 (1973).

Приводятся некоторые модели марковского типа и результаты их применения.

5. Bell C., *Can Mathematical Models Contribute to Efficiency in Higher Education?*, *Papers on Efficiency in the Management of Higher Education*, Carnegie Commission on Higher Education, Berkeley, Calif., 1972, pp. 43—60.
6. Cartter A. M., *The Supply of and Demand for College Teachers*, *J. Human Resources*, 1, 22—38 (1966).

Впервые дается прогноз относительной избыточности числа людей, имеющих степень доктора философии.

7. Clough D. J., McReynolds W. P., *State Transition Model of an Educational System Incorporating a Constraint Theory of Supply and Demand*, *Ontario J. Educ. Res.*, 9, 53—65 (1966).
8. Conlisk J., *The Determinants of School Enrollment and School Performance*, *J. Human Resources*, 4, 140—157 (1969).

На основе данных переписи 1960 г. осуществляется регрессионный анализ зависимости между уровнем образования, получаемого в общественных школах, и такими характеристиками учащихся, как пол, расовая принадлежность, места проживания (город или сельская местность), доход семьи, образование родителей.

9. Corazzini A. J., Dugan D. J., Grabowski H. H., *Determinants and Distributional Aspects of Enrollment in U.S. Higher Education*, *J. Human Resources*, 7, 39—59 (1972).

Проводится регрессионный анализ степени зависимости посещаемости колледжей от доходов семей.

10. Correa H., *Quantitative Methodologies of Educational Planning*, International Textbook Co., N.Y., 1969.

Дается детальное описание потоковых моделей планирования программ и средств профессионального обучения. Модели предназначены

для анализа проблемы удовлетворения потребностей в рабочей силе для обеспечения экономического роста.

11. Correa H., Tinbergen J., Quantitative Adoption of Education to Accelerated Growth, *Kyklos*, 15 (1962).

Представлен один из первых вариантов потоковой модели.

12. Dyer J. S., Assessing Effects of Changes in the Cost of Higher Education on Student Enrollment, *Socio-Eco. Plan. Sci.*, 5, 307—316 (1971).

13. Fox K. A., Sengupta J. K., The Specification of Econometric Models for Planning Educational Systems, *Kyklos*, 21, 665—694 (1968).

Проводится общее рассмотрение моделей типа «вход — выход», моделей марковского типа и моделей линейного программирования.

14. Gani J., Formulae for Projecting Enrollments and Degrees Awarded in Universities, *J. Royal Statistical Soc.*, A126, 400—409 (1963).

Представлена модель марковского типа.

15. Hopkins D. S. P., On the Use of Large-Scale Simulation Models for University Planning, *Rev. of Educ. Res.*, 41, 467—478 (1971).

Рассматриваются вопросы целесообразности использования экспоненциальных методов оценки быстроменяющихся параметров.

16. Marshall K., Oliver R. M., A Constant-Work Model for Student Attendance and Enrollment, *Operations Res.*, 18, 193—206 (1970).

Дается модель, которая позволяет интерпретировать данные о посещаемости колледжей. При этом используется идея о существовании определенного количества курсов, которые необходимо пройти для получения степени, и некоторой вероятности успешного прохождения отдельного курса.

17. Marshall K. T., A Comparison of Two Personnel Prediction Models, *Operations Res.*, 21, 810—822 (1973).

Проводится сравнение модели марковского типа с моделями процесса формирования групп на основе данных, относящихся к Университетскому колледжу в г. Беркли.

18. Mathematical Models in Educational Planning, OECD, Paris, 1967.

19. Masters S., The Effect of Family Income on Children's Education, *J. Human Resources*, 4, 158—176 (1969).

Проводится регрессионный анализ данных переписи 1960 г.

20. Moser C. A., Redfern P., A Computable Model of the Educational System in England and Wales, Proceedings of the 35th International Statistical Institute, Belgrade, 1966.

21. Nam C. B., Rhodes A. L., Herriot R. E., School Retention by Race, Religion and Socio-economic Status, *J. Human Resources*, 3, 171—190 (1968).

Проводится регрессионный анализ данных обзора о населении США 1965 г.

22. Oliver R. M., Hopkins D. S. P., An Equilibrium Flow Model of a University Campus, *Operations Res.*, 20, 249—264 (1972).

Дается простая потоковая модель процесса формирования групп (число групп равно восьми), устанавливающая зависимость между количеством зачисляемых студентов и конечной потребностью в специалистах, прошедших соответствующий курс обучения. Модель используется для анализа данных, относящихся к Университету г. Беркли, шт. Калифорния.

23. Orr L., The Dependence of Transition Proportions in the Educational System on Observed Social Factors and School Characteristics, *J. Roy. Stat. Soc.*, 135, 74—95 (1972).

24. Simon K., Projections of Educational Statistics, U.S. Office of Education, Dep't of HEW, an annual publication.

Представляет собой ежегодные выпуски, в которых приводятся детальные десятилетние планы набора различных категорий учащихся (от набора в детские сады до формирования групп докторов философии)

- и соискателей степеней, присуждаемых в различных областях знаний. Подобные планы составляются с помощью тщательно разработанной регрессионной модели, которая в деталях описывается в приложении.
25. Stone R., A Model of the Educational System, *Minerva*, 3, 172—186 (1965).
Описывается общий вариант модели марковского типа.
 26. Stone R., Input-Output and Demographic Accounting: A Tool for Educational Planning, *Minerva*, 4, 363—380 (1966).
Описывается динамическая модель планирования марковского типа.
 27. Stone R., Stone G., Gunton J., An Example of Demographic Accounting: The School Ages, *Minerva*, 6, 185—212 (1968).
Приводятся результаты применения модели, предложенной одним из авторов работы, для анализа данных, относящихся к Великобритании.
 28. Zabrowski E. K., The Dynamod Model of Student and Teacher Population Growth, *Socio-Eco. Plan. Sci.*, 2, 455—464 (1968).
Приводится описание имитационной модели марковского типа с использованием ЭВМ для определения 140 студенческих и преподавательских групп для заданных интервалов времени.

2.4. Модели распределения трудовых ресурсов

Трудно провести четкое разграничение между работами, связанными с распределением профессионально подготовленной рабочей силы, и некоторыми работами, включенными в предыдущий раздел. Дело в том, что многие модели формирования групп имеют отношение к комплектованию штата различного рода учебных заведений. Однако работы, которые, по нашему мнению, представляют интерес, в основном посвящены вопросам обеспеченности трудовыми ресурсами.

1. Adelman I., A Linear Programming Model of Educational Planning, The Theory and Design of Economic Development, I. Adelman and E. Thorbecke (Eds.), Johns Hopkins Univ. Press, Baltimore, 1966, pp. 385—411.
Рассматриваются вопросы планирования программ обучения, ориентированных на максимальное улучшение показателей экономического роста. Предлагаемая модель применяется к данным, относящимся к Аргентине.

2. Alper P., Armitage P. H., Smith C. S., Educational Models, Manpower, Planning, and Control, *Operational Res. Quart.*, 18, 93—103 (1967).

Рассматривается модель марковского типа, предназначенная для анализа проблемы обеспечения на национальном уровне необходимых объемов трудовых ресурсов различных видов.

3. Balinsky W., Reisman A., Some Manpower Planning Models Based on Levels of Educational Attainment, *Management Sci.*, 18, B691—B705 (1972).

Предлагается потоковая модель распределения кадров в сфере обучения с учетом степени занятости. В модели минимизируются затраты, возникающие в результате недостатка или избытка различных специалистов требуемой квалификации.

4. Bolt K. H., Koltun W. L., Levine O. H., Doctoral Feedback into Higher Education, *Science*, 148, 918 (1965).
5. Bowles S., Planning Educational Systems for Economic Growth, Harvard Univ. Press, Cambridge, 1969.

Приводится тщательно разработанная модель линейного программирования, которая была использована правительством Нигерии.

6. LeVasseur P. M., A Study of Inter-Relationships between Education, Manpower, and Economy, *Socio-Eco. Plan. Sci.*, 2, 269—295 (1969).

Описывается потоковая модель, реализованная с использованием ЭВМ, предназначенная для планирования на национальном уровне в области образования, распределения кадров и экономики.

7. Levine S. N., Economic Growth and the Development of Educational Facilities, *Socio-Eco. Plan. Sci.*, 1, 27—32 (1967).

Описывается имитационная модель, с помощью которой устанавливается взаимосвязь удовлетворения потребностей в трудовых ресурсах с общим развитием системы образования.

8. Reisman A., The Generation of Doctorates and Their Feedback into Higher Education, *Socio-Eco. Plan. Sci.*, 2, 473—486 (1969).

Описывается достаточно полная имитационная модель, включающая 200 уравнений. Модель предназначена для исследования различных стратегий распределения кадров.

9. Skorov G., Manpower Approach to Educational Planning, Economic and Social Aspects of Educational Planning, UNESCO, Paris, 1964.

Обсуждается планирование использования кадров в СССР.

10. Tinbergen J., Bos H. C., A Planning Model for Educational Requirements of Economic Growth, The Residual Factor and Economic Growth, OECD, Paris, 1964. pp. 147—169.

Проводится анализ формирования групп с учетом потребностей в кадрах различной квалификации.

2.5. Эффективное использование ресурсов в системе образования

Большая часть работ, включенных в данный раздел, относится к проблемам эффективного распределения ресурсов в пределах системы образования. Несколько работ относится к важной социальной проблеме, связанной с определением количества ресурсов, которое должно быть выделено системе образования в целом. Кроме того, в настоящий раздел включено несколько работ, относящихся к проблеме оценки эффективности отдельных элементов учебного процесса.

1. Abt C. C., Design for an Educational System Cost-Effectiveness Model, *Socio-Eco. Plan. Sci.*, 2, 399—415 (1968).

Описывается класс имитационных моделей, пригодных для агрегированной оценки программ раздела 1 Закона 1964 г. о начальном и среднем образовании.

2. Besel R., A Linear Model for the Allocation of Instructional Resources, *Socio-Eco. Plan. Sci.*, 6, 501—506 (1972).

Описывается модель линейного программирования, предназначенная для анализа проблемы распределения ресурсов, проводимого с точки зрения достижения требуемых уровней успеваемости. Модель строится в предположении, что учащиеся в зависимости от способностей требуются различные объемы ресурсов.

3. Bruno J. E., A Mathematical Programming Approach to School Finance, *Socio-Eco. Plan. Sci.*, 3, 1—12 (1969).

Описывается модель линейного программирования, предназначенная для анализа проблемы размещения капиталов в системе образования. В модели учитываются соображения политического, социального и экономического характера.

4. Bruno J. E., Using LP Salary Evaluation Models in Collective Bargaining Negotiations with Teacher Unions, *Socio-Eco. Plan. Sci.*, 3, 103—117 (1969).

Описывается модель, предназначенная для решения проблемы удовлетворения потребностей различных объединений преподавателей при наименьших затратах.

5. Bruno J. E., The Use of Monte Carlo Techniques for Determining Optimal Size of Substitute Teacher Pools in Large Urban School Districts, *Socio-Eco. Plan. Sci.*, 4, 415—428 (1970).

Исследуются возможности замены временных преподавателей постоянными.

6. Burkhead J., Fox T. G., Holland J. W., Input and Output in a Large City High School, Syracuse Univ. Press, Syracuse, N.Y., 1967.

Описывается модель линейного программирования, в которой успеваемость рассматривается как функция от бюджета соответствующего учебного заведения. Модель рассматривается применительно к данным, относящимся к высшим учебным заведениям Чикаго и Атланты.

7. Durstine R. M., In Quest of Useful Models for Educational Systems, *Socio-Eco. Plan. Sci.*, 2, 417—437 (1968).

Исследуются потоковые модели, описывающие определенные аспекты системы образования.

8. Dyer J. S., A Procedure for Selecting Educational Goal Areas for Emphasis, *Operations Res.*, 21, 835—845 (1973).

Проводится разработка функции практической полезности.

9. Gerwin D., A Process Model of Budgeting in a Public School Systems, *Manag. Sci.*, 15, 338—361 (1969).

Описывается достаточно полная имитационная модель процесса финансирования школ некоторого округа. Модель рассматривается применительно к системе средних школ Питтсбурга.

10. Judy R. W., Levine J. B., A New Tool for Educational Administration, Univ. of Toronto Press, 1965.

Детально описывается модель типа «вход — выход», которая весьма эффективна для анализа проблем планирования на уровне отдельных институтов системы высшего образования.

11. Kaufman R. A., Corrigan R. E., Johnson D. W., Towards Educational Responsiveness to Society's Needs, *Socio-Eco. Plan. Sci.*, 3, 151—157 (1969).

Описывается модель, в которой использование индивидуальных функций полезности для общества служит основой планирования в системе образования.

12. Kodoma F., An Approach to the Analysis of Vocational Education and Training Requirements, *Management Sci.*, 1, № 17, B178—B191 (1970).

Обсуждается вопрос распределения времени между элементами программы профессионального обучения с целью минимизации ожидаемого числа несчастных случаев и ошибок в последующей работе.

13. Koenig H. E., Keeney M. G., A Portotype Planning and Resource Allocation Program for Higher Education, *Socio-Eco. Plan. Sci.*, 2, 201—215 (1968).

Дается описание в терминах пространства состояний проблемы использования ресурсов в университетах при реализации различных учебных программ.

14. Lee S. M., Clayton E. R., A Goal Programming Model for Academic Resource Allocation, *Management Sci.*, 18, B395, B408 (1972).

Описывается модель линейного программирования распределении ресурсов. При этом критерием является относительная значимость свободных (слабых) переменных. Модель применяется для анализа данных, относящихся к Политтехническому институту в Вирджинии.

15. Lerman R. I., Some Determinants of Youth School Activity, *J. Human Resources*, 7, 366—379 (1972).

Проводится регрессионный анализ данных переписи с целью выявле-

ния влияния факторов семьи на образовательный уровень и занятость различных слоев молодежи.

16. Levin H. M., A Cost-Effectiveness Analysis of Teacher Selection, *J. Human Resources*, 5, 24—33 (1970).

Для изучения различных критериев выбора преподавателей используется функция простого воспроизводства.

17. McNamara J. F., A Regional Planning Model for Occupational Education, *Socio-Eco. Plan. Sci.*, 5, 317—339 (1971).

Для изучения распределения фондов профессионального обучения по различным программам, осуществляемым в Пенсильвании, используется модель линейного программирования.

18. Newton R. M., Thomas W. H., Design of School Bus Routes by Computer, *Socio-Eco. Plan. Sci.*, 3, 75—85 (1969).

Предлагается процедура определения оптимального маршрута движения школьного автобуса, состоящая в разбиении оптимального маршрута в соответствующей задаче коммивояжера.

19. Owen J. D., The Distribution of Educational Resources, in Large American Cities, *J. Human Resources*, 7, 26—38 (1972).

Для исследования механизма распределения средств обучения в среде бедного и цветного населения используется регрессионный анализ.

20. Ploughman T., Darnton W., Heuser W., An Assignment Program to Establish School Attendance Area Boundaries and Forecast Construction Need, *Socio-Eco. Plan. Sci.*, 1, 243—258 (1967).

Предлагается имитационная модель для планирования роста системы школ на период, охватывающий несколько лет. Модель разработана применительно к данным, относящимся к округу Окленд, шт. Мичиган.

21. Reisman A., Taft M. I., A Systems Approach to the Evaluation and Budgeting of Educational Programs, *Socio-Eco. Plan. Sci.*, 3, 245—277 (1969).

Описывается механизм распределения ресурсов, основанный на использовании агрегированной функции полезности и отражающий полезность отдельных программ обучения.

22. Sinha B. K., Gupta S. K., Sisson R. L., Towards Aggregate Models of Educational Systems, *Socio-Eco. Plan. Sci.*, 3, 25—36 (1969).

Описываются потоковая модель распределения учащихся и модель определения уровня успеваемости учащихся на основе данных об исходных ресурсах.

23. Sisson R. L., Can We Model the Education Process? *Socio-Eco. Plan. Sci.*, 2, 109—119 (1968).

Проводится общее обсуждение проблемы выявления зависимости между выходными переменными, характеризующими образовательный уровень, и входными переменными, характеризующими предоставленные ресурсы.

24. Szekely B., Stankard M., Sission R. L., Design of a Planning Model for an Urban School District, *Socio-Eco. Plan. Sci.*, 1, 231—242 (1967).

Описывается имитационная модель планирования учебной работы в рамках некоторого округа.

2.6. Комплексное проектирование учебных заведений

Работы, вошедшие в данный раздел, могли быть включены и в предыдущий. Поскольку, однако, они составляют довольно большую группу, то представляется целесообразным выделить их особо.

1. Crandall R. H., A Constrained Choice Model for Student Housing, *Management Sci.*, 16, 112—120 (1969).

Для определения оптимального размещения учащихся в общежитиях, включая общежития, расположенные на территории колледжа, ис-

пользуется модель линейного программирования. Критерием является минимум затрат.

2. Eastman C. M., Kortanek K. D., Modeling School Facility Requirements in New Communities, *Management Sci.*, 16, B784—B799 (1969).

Для выявления связи между возрастающим запасом мест в общежитиях, а также возрастающей степенью перемещений внутри общежитий и требованиями к обеспечению условий жизни учащихся используется модель линейного программирования.

3. Graves R. J., Thomas W. H., A Classroom—Allocation Model for Campus Planning, *Socio-Eco. Plan. Sci.*, 5, 191—204 (1971).

Для решения проблемы проектирования общежитий при колледжах (Буффало), удовлетворяющих некоторым условиям нормального проживания при минимальных затратах, используется модель линейного программирования.

4. Maver T. W., Spatial Environment in Comprehensive Schools, *Operational Res., Quart.*, 23, 205—322 (1972).

Описываются компьютерные программы, предназначенные в помощь архитекторам при проектировании школ, когда целевой функцией является достижение максимальной гибкости использования помещений для различных целей.

5. O'Brien R. J., Model for Planning the Location and Size of Urban Schools, *Socio-Eco. Plan. Sci.*, 2, 141—153 (1969).

Дается некоторое множество моделей, с помощью которых может быть составлена функция полезности.

6. Smith R. L., Accommodating Student Demand for Courses by Varying the Classroom-Size Mix, *Operations Res.*, 19, 862—874 (1971).

Описывается процесс распределения курсов по учебным аудиториям. Критерием является минимум учащихся, не полавших на данный курс ввиду ограниченности аудиторий.

2.7. Проблема интеграции

Более эффективное решение проблемы интеграции школ оказалось возможным благодаря изданию количественных методов анализа.

1. Belford P., Ratliff H. D., A Network-Flow Model of Racially Balancing Schools, *Operations Res.*, 20, 619—628 (1972).

Описывается потоковая модель, предназначенная для анализа проблемы интеграции. В модели используется критерий минимума суммарного расстояния от места жительства учащихся до школы. Модель рассматривается применительно к высшим учебным заведениям округа Гейнсвилл, шт. Флорида.

2. Clarke S., Surkis S., An Approach to Racial Desegregation in School System, *Socio-Eco. Plan. Sci.*, 1, 259—272 (1968).

Для решения проблемы формирования состава учащихся школы (округ Бруклин, шт. Нью-Йорк), удовлетворяющего определенным условиям этнического характера, используется модель линейного программирования. Критерием является минимизация суммарного времени, затрачиваемого учащимися на дорогу в школу.

3. Heckman L. B., Taylor H. M., School Rezoning to Achieve Racial Balance, *Socio-Eco. Plan. Sci.*, 3, 127—133 (1969).

Для определения границ районов, обслуживаемых отдельными школами, при условии решения проблемы интеграции и при критерии минимума затрат на поездки в школу используется модель линейного программирования.

4. Koenigsberg E., Mathematical Analysis Applied to School Attendance Areas, *Socio-Eco. Plan. Sci.*, 1, 465—475 (1968).

Описывается модель линейного программирования, предназначенная для анализа задачи прикрепления учащихся к школам. Задача состоит в решении проблемы интеграции при различных критериях.

5. Levine S. N., A Model of Racially Changing Neighborhoods, *Socio-Eco. Plan. Sci.*, 1, 477—479 (1968).

Приводится модель двух параметров, пригодная для представления изменяющегося этнического состава определенной части живущего по соседству населения.

II-4. Транспорт

Б. Нийл¹⁾, Э. Кэтенис²⁾, Д. Хайндс³⁾

ВВЕДЕНИЕ

Транспорт — это система средств, предназначенных для организации и осуществления перевозки людей и грузов с определенными целями. Из приведенного определения следует, что проблема транспорта включает большой круг вопросов. Несмотря на то что для решения этих вопросов могут быть использованы самые различные методы, есть все основания полагать, что в данном случае наиболее эффективными окажутся методы исследования операций. Транспортная сеть предоставляет исследователям систем широкое поле деятельности, где они могут продемонстрировать возможности методов исследования операций при изучении как отдельных аспектов указанной проблемы, так и проблем, возникающих при рассмотрении транспортной сети как подсистемы общей городской системы.

Иногда существование эффективной транспортной сети воспринимается как нечто само собой разумеющееся, причем даже не до конца осознается вся ее значимость. Только неопровержимые факты, например сведения о том, что 20% валового национального продукта США связано с функционированием транспортной сети [22] или 35—40% всех застрахованных территорий отводится на развитие транспорта [41], могут заставить людей трезво оценить реальное положение дел. Вообще людям свойственно осознавать наличие транспортных проблем только тогда, когда они сами становятся жертвами заторов, несогласованности в расписании движения самолетов или обнаруживают промажу принадлежащего им груза. В этой связи следует заметить, что меры по улучшению обслуживания и ликвидации недостатков, как правило, принимаются лишь в тех случаях, когда подобные инциденты неоднократно имели место и критика в адрес соответствующих организаций достигла достаточно высокого уровня.

По мере повышения сложности транспортных сетей, которое сопровождается возрастанием роли проблем обеспечения их надежности, во многих странах растет и интерес к транспорту как к объекту исследования. Например, люди, занимающиеся проблемой перевозок грузов, не только прекрасно понимают необходимость создания хороших транзитных систем и связанные с этим выгоды (более эффективное использование энергии, оживление и восстановление деловой части города и т. п.), но и невозможность их реализации без специальных исследований. В настоящее время на многих промышленных предприятиях распределение потоков продукции анализируется с использованием системного подхода и при этом используются существующие

¹⁾ Brian F.O'Neil, Ph. D., Associate Professor of Management and Logistics in the School of Business Administration at the University of Miami, Coral Gables, Florida.

²⁾ Anthony J. Cantanese, Ph. D., Dean of the School of Architecture and Urban Planning at the University of Wisconsin-Milwaukee.

³⁾ David H. Hinds, Chief of Programming and Planning Support in the Office of Transportation Administration of Metropolitan Dade County, Miami, Florida.

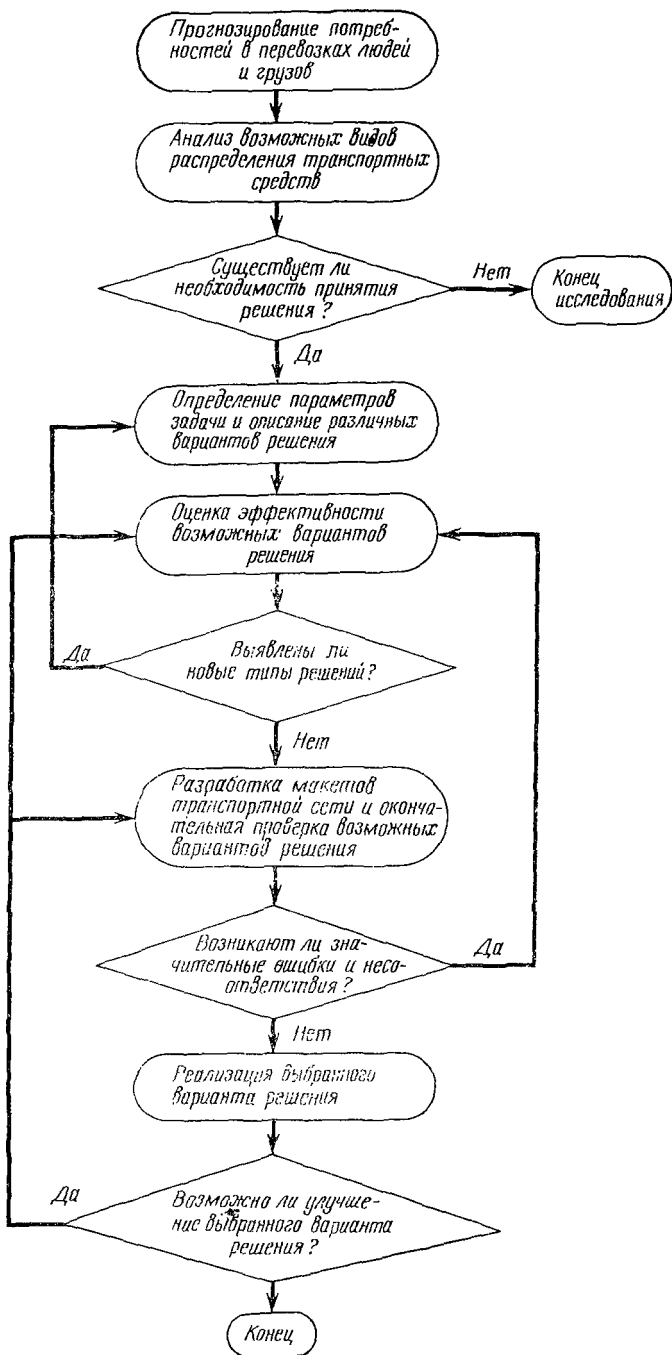


Рис. 1. Общая схема анализа транспортных средств

зависимости между отдельными элементами системы распределения продукции. Результаты такого анализа часто оказываются совершенно поразительными с точки зрения громадной экономии, к которой они приводят.

При изучении проблем транспорта необходимо помнить, что он не представляет собой изолированную систему, а тесно связан с другими системами. В общем случае транспорт можно рассматривать как средство достижения некоторой цели, которая формируется в рамках другой системы. Например, перевозка грузов внутри предприятия и между предприятиями является лишь частью всей системы распределения продукции и в то же время тесно взаимодействует с такими подсистемами, как управление производственными запасами, обработка заказов, производство, обработка информации и общее управление предприятием. Если обратиться к городской транспортной системе, включающей сложную сеть улиц и магистралей, системы транзитной перевозки грузов, скоростного проезда автомобилей и т. д., то недостаточно отметить, что эта система выполняет функции обеспечения других городских систем (экономической, социальной, политической и культурной). Необходимо также со всей определенностью подчеркнуть, что транспортная подсистема является неотъемлемой органической частью общегородской системы. При этом важно иметь в виду, что планирование работы транспорта должно обязательно включать исчерпывающий анализ взаимодействия транспортной подсистемы города с другими его подсистемами.

В данной главе широко используются термины «государственный сектор» и «частный сектор». Государственный сектор транспорта включает крупные общественные магистрали, общественные системы для транзитных перевозок, а также внутригородские или междугородные системы для личного транспорта. Частный сектор включает транспортные системы распределения продукции и другие находящиеся в частной собственности, хотя и регулируемые правительством, системы перевозки грузов. Для анализа и проектирования большинства транспортных подсистем как общественных, так и частных, может быть применена хорошо известная общая схема исследования (рис. 1), в которой весьма точно реализуются общие принципы системного анализа. Основными этапами такого исследования являются:

- прогнозирование потребностей в перевозках людей и грузов,
- анализ возможных видов распределения транспортных средств,
- определение параметров задачи и описание различных вариантов решения,
- оценка эффективности возможных вариантов решения,
- разработка макета транспортной сети и ее окончательная проверка,
- реализация выбранного варианта решения.

Рассмотрим более подробно каждый из перечисленных этапов исследования.

1. ПРОГНОЗИРОВАНИЕ ПОТРЕБНОСТЕЙ В ПЕРЕВОЗКАХ ЛЮДЕЙ И ГРУЗОВ

Прогнозирование потребностей в каких-либо перевозках должно быть основано на исследовании, включающем анализ следующих восьми элементов транспортной системы: пункт отправления, пункт назначения, объем перевозок, вид транспорта, транспортная линия, объект перевозки, расписание перевозок, характеристика времени осуществления перевозок. Объектами перевозок, как уже указывалось, могут быть люди и грузы (табл. 1).

В производственной сфере основу исследования составляют потребности в перевозках, характеризуемые грузопотоками, в то

Таблица 1

Элементы транспортной системы	Характер информации, используемой при анализе	
	государственный сектор	частный сектор
Пункт отправления (откуда)	Место жительства	Оптовые базы или магазины розничной торговли
Пункт назначения (куда)	Общественные предприятия, биржи труда, торговые центры, центры развлечений и отдыха, правительственные учреждения	Заводы, склады, центры распределения продукции, сырьевые базы
Транспортная линия (каким путем)	Скоростные магистрали, дороги и улицы, транзитные магистрали, железные дороги	Магистрали, связывающие штаты; внутренние дороги штатов; воздушные, речные и морские линии; железные дороги
Объект перевозки (кто, что)	Рабочие, служащие и школьники; покупатели; престарелые и инвалиды	Грузы всевозможных видов и габаритов
Вид транспорта (на чем)	Автомобиль, автобус; велосипед, движение пешком	Собственный или арендованный грузовой транспорт; общественный транспорт; самолет, корабль, поезд
Объем перевозок (сколько)	Средний ежедневный объем перевозок по воздушным линиям, мощности пиковых потоков грузов	Вес или другая количественная характеристика груза; полностью или не полностью загруженное транспортное средство (железнодорожная платформа, автомобиль и т. п.)
Расписание перевозок (когда)	Расписание работы рабочих и служащих, характерные время закупок для домохозяйек, школьное расписание, время наступления отдельных событий	Режим работы транспортного средства, определяемый диспетчером или заранее составленным расписанием
Характеристика времени осуществления перевозок	Часы пик, сезон или период межсезонья	Период отправления грузов, период межсезонья

время как в непроизводственной — потребности в перевозках пассажиров, характеризуемые пассажиропотоками. Прогноз может разрабатываться как для периода времени, равного всего лишь одному дню (например, при составлении производственных расписаний), так и для периода времени до 20 лет (например, при проектировании скоростной системы транзитных перевозок). Выбор наиболее желательного варианта решения при анализе

транспортных систем осуществляется с учетом ожидаемого поведения больших групп людей, имеющих общие интересы.

В общем виде характерные особенности организации перевозок людей и грузов состоят в следующем. Рост потребностей в перевозках грузов стимулируется развитием промышленного производства, которое определяется совокупной потребностью общества в товарах. Поэтому выбором пункта погрузки, способа перевозки, маршрута и составлением расписания перевозок занимается соответствующая транспортная организация, связанная с промышленностью. С другой стороны, перевозки общественным транспортом обуславливаются потребностью в перемещении населения из одной точки города в другую, причем выбор одного варианта передвижения из нескольких возможных осуществляется самим пассажиром. Органы городского управления, по существу, не оказывают влияния на работу транспортной сети, если не считать, например, таких действий, как запреты на движение определенных транспортных средств в некоторых районах города. В то же время эти органы непрямым образом контролируют проектирование и составление расписания работы общественных транспортных систем, а также формируют принципы распределения земельных участков для развития транспорта. Хотя решение обоих указанных вопросов может оказать существенное влияние на организацию перевозок, важно подчеркнуть, что органы городского управления не ставят целью непосредственно контролировать, где, когда и как перемещается пассажир. Потребность в таком перемещении носит случайный характер, что существенным образом затрудняет прогнозирование развития городских транспортных сетей.

Мало того, что потребности в перевозках носят стохастический характер, для отдельных моментов времени их просто трудно определить, так как для этого необходимо было бы провести исследование перемещений больших групп населения. Целесообразность получения такой информации должна подвергаться всесторонней оценке, так как затраты на ее получение могут достигать сотен тысяч долларов [25]. Кроме того, в столь обширных исследованиях качество информации может быть весьма невысоким из-за отсутствия непосредственного контроля за сборщиками информации.

При прогнозировании потребностей в перевозках часто используется метод множественной регрессии. Так, на основе уравнения множественной регрессии, учитывающего такие показатели, как валовый национальный продукт, национальный научно-технический уровень, объемы грузопотоков, коэффициент промышленного роста и ряд других, были предсказаны объемы воздушных грузовых перевозок и, как следствие, необходимые для этого объемы производства транспортных самолетов [3, 32]. Коэффициен-

ты регрессии в этом случае определялись на основе исходных данных стандартным методом наименьших квадратов. При исследовании транспортных перевозок чикагского региона [8] применение регрессии, учитывающей такие показатели, как количество владельцев автомобилей и плотность населения, позволило довольно точно оценить количество поездок, приходящихся на одну семью. В этом исследовании использовались методы нелинейной интерполяции, в частности полиномы и функции Гаусса для интерполяции нелинейных временных зависимостей числа автобусных и автомобильных поездок, приходящихся на одного жителя.

Более сложным и потому реже используемым методом прогнозирования является факторный анализ. Этот метод состоит в комбинировании большого числа входных переменных в существенно меньшее число групп, включающих сильно коррелирующие переменные. Иногда данный метод используется перед применением регрессионного анализа, благодаря чему последний становится более эффективным. Факторный анализ был использован в работе [17] для выявления характерных участков транспортных сетей и основных направлений потоков грузов между различными странами, а в работе [25] — для определения эффективных потребностей в различных видах городского транспорта.

При прогнозировании потребностей в перевозках людей и грузов возможно также применение имитационных моделей, реализованных на ЭВМ. Обсуждение некоторых наиболее значительных имитационных моделей будет дано при рассмотрении одного из следующих этапов общей схемы исследования. Однако имитационную модель [21] целесообразно рассмотреть в настоящем разделе. Основное назначение данной модели состоит в предсказании требований к оборудованию летательных аппаратов (прогноз выполняется на срок до 10 лет путем обработки данных о функционировании авиалиний США за предыдущий десятилетний период). Хотя первоначально эта модель предназначалась для предсказания конъюнктуры рынка, тем не менее оказалось возможным прогнозировать тенденции изменения характеристик воздушного транспорта (например, необходимость в новых системах авиалиний, рост объемов перевозок и их стоимости, развитие средств обслуживания пассажиров, улучшение перевозки багажа и грузов, технического обслуживания самолетов, а также общей стоимости системы воздушных сообщений). Метод имитационного моделирования оказался наиболее подходящим для решения задачи, поставленной фирмой Lockheed и имевшей целью формирование требований к развитию большого числа различных авиалиний США.

Для руководителей какой-то определенной авиакомпании вполне естественным является желание предсказать те требова-

ния, которым должна отвечать эта авиакомпания, чтобы в будущем выстоять в конкурентной борьбе (возможно, при таком анализе некоторые конкуренты будут выделены особо). С помощью упомянутой модели можно решить эту задачу практически для любой авиакомпании США. Решение осуществляется путем сведения некоторого числа частных требований, составленных с точки зрения отдельной авиакомпании в общую систему требований с последующей конкретизацией оценок, полученных в такой укрупненной модели. Другими словами, данный подход позволяет получать детальный обзор взаимосвязанных характеристик отдельных авиакомпаний. Подобные возможности рассматриваемой модели являются уникальными, и, видимо, имеет смысл попытаться разработать аналогичные модели и для других видов транспорта, таких, как грузовой автомобильный, железнодорожный и морской транспорт.

Рассмотрим теперь более подробно методы прогнозирования применительно к городскому транспорту. Используемые в данном случае так называемые модели распределения поездок включают в себя модели развития, конфликтующих возможностей, равных возможностей, предпочтения и притяжения [11]. Эти модели построены на различных теоретических предположениях относительно того, каким образом локализация пунктов отправления и назначения, объем перевозок и другие элементы транспортной системы взаимосвязаны с остальными выходными переменными. В качестве последних выступают такие параметры, как планируемый рост тарифов, коэффициенты реальной занятости, расположение торговых зон или зон развлечений и отдыха, пространственное и временное разделение различных городских территорий, факторы привлекательности тех или иных районов города.

Локализация пунктов отправления и назначения может быть определена после разбиения всей городской территории на зоны. Объем перевозок обычно выражается числом поездок, приходящихся либо на одного человека, либо на одно транспортное средство. Общепринятой классификацией поездок в зависимости от пункта назначения является следующая: поездки домой, на работу, в школу, в места отдыха, развлечений, проведения общественных мероприятий, а также регулярные и нерегулярные поездки, связанные с покупкой товаров.

Модели развития позволяют определить будущее распределение поездок между различными парами выделенных зон на основе исходных данных о количестве выездов и въездов для отдельных зон и количестве поездок между различными парами зон. Прогнозируемые значения выходных переменных получают из известных (на момент составления прогноза) значений с помощью коэффициента развития, который представляет собой оценку ожидаемых изменений в плотности населения и в степени

использования территорий рассматриваемых зон. На основе модели развития разработано несколько методов анализа городского транспорта, среди них — методы постоянного коэффициента развития, среднего коэффициента развития, здравого смысла и так называемый детройтский метод. Однако наиболее тщательно разработанным и потому широко распространенным является метод Фратера [17].

Основное предположение, используемое в методе Фратера, состоит в том, что количество поездок из i -й зоны в j -ю для некоторого будущего момента времени пропорционально исходному числу всех выездов из i -й зоны, умноженному на коэффициент развития j -й зоны. Общее соотношение, используемое в методе Фратера, имеет следующий вид:

$$T_{ij(i)} = t_{ij} F_i F_j \sum_x t_{ix} \left(\sum_x t_{ix} F_x \right).$$

В этом и следующем выражениях приняты такие обозначения: $T_{ij(i)}$ — прогнозируемое количество поездок между i -й и j -й зонами, обусловливаемое развитием i -й зоны; $T_{ij(j)}$ — прогнозируемое количество поездок между i -й и j -й зонами, обусловливаемое развитием j -й зоны (величину $T_{ij(j)}$ можно получить из формулы для $T_{ij(i)}$, если в правой части этой формулы индексы i и j поменять местами); t_{ij} — исходное число всех поездок между i -й и j -й зонами; F_x — коэффициент развития для зоны с номером x , где x принимает множество значений, соответствующих всем рассматриваемым зонам.

Если через T_{ij} обозначить прогнозируемое количество всех поездок между i -й и j -й зонами, то имеем

$$T_{ij} = \frac{T_{ij(i)} + T_{ij(j)}}{2}.$$

Поскольку величины, определяемые из двух приведенных выше соотношений, как правило, не удовлетворяют условию

$$\sum_j T_{ij} = F_i \sum_j t_{ij},$$

то для его выполнения необходимо использовать метод последовательных приближений.

Модель конфликтующих возможностей предложена в работе [29] и использована при исследовании транспортных проблем Чикагского региона [8]. Затем эта модель рассматривалась также в работе [9]. Модель основана на предположении о том, что число людей, выезжающих из некоторого пункта, прямо пропорционально числу «возможностей», имеющихся на маршруте, и об-

ратно пропорционально числу «реализованных возможностей». В качестве упомянутых «возможностей» можно рассматривать, например, места, в которых может быть предложена работа (для людей, ищущих работу), городские и загородные парки (для людей, совершающих поездку с целью отдыха или развлечений), наконец, торговые центры (для людей, выезжающих за покупками).

Модель конфликтующих возможностей математически может быть представлена с помощью выражения

$$T_{ij} = O_i [e^{-LD} - e^{-L(D+D_j)}],$$

где T_{ij} — количество поездок из i -й зоны в j -ю; O_i — общее количество поездок, начинающихся в i -й зоне; D — количество имеющихся (или отсутствующих) целей на маршрутах, ведущих в j -ю зону; D_j — количество поездок, заканчивающихся в j -й зоне; L — некоторая оценка вероятности того, что в случайно выбранном пункте назначения может быть достигнута цель отдельной поездки (L представляет собой величину, которую необходимо определить на основе исходных данных; она характеризует степень убывания величины T_{ij} с увеличением числа целей и длины маршрута), e — основание натурального логарифма.

Основой модели равных возможностей [30] является распределение поездок, характеризующихся одинаковыми временем, расстоянием и стоимостью, по группам однотипных поездок, т. е. поездок с одинаковыми целями. Выделенные таким способом поездки равновероятны в пределах каждой группы.

Модель предпочтения [21] представляет собой модель распределения поездок несколько иного типа, чем рассмотренные выше. Основное предположение, используемое при построении этой модели, состоит в том, что для каждого пункта отправления можно оценить предпочтительность всех возможных пунктов назначения, а для каждого пункта назначения можно оценить предпочтительность всех возможных пунктов отправления. В данной модели сочетание двух определенных пунктов отправления и назначения может быть оптимальным только для одного из пунктов, причем такое сочетание должно удовлетворять некоторым условиям устойчивости.

В модели притяжения [33], являющейся наиболее широко распространенной моделью распределения поездок, делается оценка количества поездок из i -й зоны в зону j -ю на основе предположения о том, что рассматриваемая величина прямо пропорциональна некоторому коэффициенту «привлекательности» j -й зоны и обратно пропорциональна коэффициенту, характеризующему удаленность i -й и j -й зон друг от друга (в единицах времени или расстояния).

Модель притяжения можно представить математически с помощью следующего выражения:

$$T_{ij} = \frac{P_i A_j F_{ij} K_{ij}}{\sum_i A_j F_{ij} K_{ij}},$$

где T_{ij} — количество поездок из 1-й зоны, совершаемых благодаря «привлекательности» j -й зоны; P_i — общее количество поездок, начинающихся в i -й зоне; A_j — общее количество поездок, «привлеченных» j -й зоной; F_{ij} — эмпирически определяемый коэффициент удаленности (F_{ij} представляется в виде d_{ij}^{-b} или, точнее, $d_{ij}^{b_{ij}}$, где d_{ij} — расстояние между i -й и j -й зонами, а b_{ij} — некоторый показатель степени, зависящий от величины d_{ij} и обычно определяемый с помощью линейной регрессии); K_{ij} — коэффициент, с помощью которого учитываются эффекты социального и экономического характера (величина K_{ij} определяется по исходным данным о распределении поездок). Таким образом, в модели притяжения должны быть предварительно определены (по исходным данным) два параметра — F и K .

Успех этой модели объясняется, главным образом, ее простотой, а также тем, что на агрегированном уровне рассмотрения имеется небольшое число параметров, которые необходимо определять предварительно с требуемой для прогноза степенью точности. Однако вопрос о возможности использования модели притяжения при более низком уровне агрегации [20] вызвал оживленную дискуссию. Интересно отметить, что в качестве альтернативы рассмотренной модели в работе [20] предлагается качественный метод, связанный с анализом данных об объемах перевозок, личных доходах, подвижности населения и стоимости поездок. В этой работе приводятся три примера практического применения указанного подхода, в которых была достигнута достаточно хорошая точность.

Модели притяжения, конфликтующих и равных возможностей, а также метод Фратера были подвергнуты сравнительному анализу путем их одновременного использования для изучения распределения поездок в г. Вашингтоне в 1948 и 1955 гг. [19]. Оказалось, что прогноз на основе модели притяжения является несколько более точным и полным, чем на основе модели конфликтующих возможностей. Однако исключение из первой модели коэффициента K_{ij} , учитывающего социально-экономические факторы, может привести к обратному результату. Это обстоятельство указывает, в частности, на то, что степень точности предварительного определения параметра K_{ij} влияет на точность прогноза, составленного с использованием этой модели. С точки зрения надежности и полезности модели притяжения и конфликтующих возможностей оказались примерно равноценными, хотя для

последней, по-видимому, несколько проще проводить предварительное определение соответствующих параметров. В отличие от этих моделей эффективность метода Фратера проявилась лишь при анализе стабильных ситуаций. Если же за период прогноза происходили какие-либо изменения, например в характере использования городской территории, то метод Фратера оказывался совершенно непригодным. Что касается модели равных возможностей, то она по-настоящему не смогла конкурировать с тремя другими моделями, поскольку в рамках рассмотренной зональной структуры г. Вашингтона для нее невозможно было провести определение соответствующих параметров. Кроме того, значительное число небольших зон крайне затруднило установление оптимальных тарифов на проезд, что является серьезным недостатком модели равных возможностей, так как во многих исследованиях, связанных с использованием городской территории, необходимо рассматривать как раз весьма детальное разбиение на зоны.

Самое последнее достижение в области моделей распределения поездок представляет модификация модели притяжения, названная моделью максимизации энтропии. Более подробное обсуждение моделей распределения читатель может найти в работе [27].

2. АНАЛИЗ РАСПРЕДЕЛЕНИЯ ВОЗМОЖНЫХ ВИДОВ ТРАНСПОРТНЫХ СРЕДСТВ

Данный этап исследования включает тщательный анализ состояния существующей транспортной сети. Основной целью такого анализа является сопоставление прогнозируемых потребностей в транспортном обслуживании с имеющимися в момент составления прогноза возможностями. По существу, на этом этапе решается вопрос, позволят ли изменения в организации сети и увеличение ее пропускной способности удовлетворить будущие потребности в перевозках людей и грузов.

Необходимо сразу же обратить внимание на то, что рассмотрение названных вопросов применительно к государственному и частному секторам проводится по-разному. Поскольку в частном секторе выбором транспортных средств и маршрутов движения, а также выбором поставщиков занимается какое-то предприятие, то распределение транспортных средств с целью удовлетворения будущих потребностей осуществляется в соответствии с решениями руководства данного предприятия. В государственном же секторе транспорта основные элементы транспортной системы связаны со случайными факторами, а удовлетворение потребностей является заботой самих людей, нуждающихся в транспортном обслуживании. Поэтому в государственном сек-

торе распределение транспортных средств с целью удовлетворения будущих потребностей представляет собой проблему, которая может быть решена лишь методами прогнозирования.

С помощью методов исследования операций был разработан ряд моделей для обеспечения возможности наиболее эффективного распределения транспортных средств (в частном секторе) или для получения наиболее правильного варианта при прогнозировании распределения транспортных средств (в государственном секторе).

Применительно к частному сектору разработаны модели, предназначенные для исследования вопросов управления запасами, составления производственных расписаний, выбора транспортных средств, составления расписаний их движения, локализации средств транспортного обслуживания и т. п., т. е. для рассмотрения практически всего комплекса вопросов, связанных с организацией системы распределения транспортных средств. Многие из этих вопросов рассматриваются в других главах данного тома. Важно отметить, что при применении упомянутых моделей необходимо делать определение предположения относительно всей транспортной системы, а решение, полученное на основе любой из моделей, является лишь частично оптимальным, так как позволяет оптимизировать только часть всей системы. Чтобы обойти эти трудности, были разработаны имитационные модели большой размерности, реализованные на ЭВМ, в которых удалось объединить все элементы системы распределения транспортных средств. Такие модели рассматриваются в одном из следующих разделов, который посвящен этапу «Оценка...» общей схемы исследования.

Особой областью приложения методов исследования операций в частном секторе транспорта является выбор конкретных транспортных средств. Модели выбора транспортных средств представляют собой описание некоторой процедуры выработки решения на основе подсчета полных затрат на реализацию проекта. В рамках этих моделей проводится сопоставление затрат с уровнем обслуживания, которое обеспечивается выбираемым транспортным средством. Затраты включают расходы на приобретение транспортного средства и эксплуатационные расходы. Уровень обслуживания характеризуется временем осуществления перевозки, грузоподъемностью и надежностью функционирования соответствующего транспортного средства.

На транспорте часто бывает очень сложно определить разницу в затратах, связанных с обеспечением различных уровней обслуживания. В качестве иллюстрации этого положения приведем следующий пример. Меньшее время перевозки грузов, например в случае их доставки по воздуху, а не по железной дороге, позволяет снизить уровень запасов при сохранении уровня надежности (под надежностью в данном случае можно понимать выражен-

ное в процентах количество случаев, в которых заявка потребителя не удовлетворяется за счет имеющихся запасов; при этом предполагается, что в рассматриваемой системе потребности формируются случайным образом). Выигрыш, связанный с более низким уровнем производственных запасов, определяется уменьшением потерь, обусловленных старением продукции и возможными перебоями в обслуживании. В таком случае конкретные значения потерь и сопоставление осуществляемых затрат часто определяются все же не на основе модельных представлений, а из соображений здравого смысла.

Интересная модель выбора транспортных средств была предложена в работе [24]. В этой модели представлены взаимосвязи между 28 известными и впервые введенными показателями эффективности использования грузовых транспортных средств. В отличие от многих моделей выбора, реализованных на ЭВМ, рассматриваемая модель представляет собой ряд уравнений, эмпирически выбираемых с таким расчетом, чтобы искомое решение могло быть найдено без применения вычислительной техники. Другая, также интересная модель выбора транспортных средств в частном секторе была предложена в работе [14].

Как уже отмечалось, в государственном секторе транспорта модели распределения средств, удовлетворяющих будущие потребности в перевозках, являются, по существу, моделями для разработки прогнозов. Некоторые из таких моделей, называемые транспортными моделями распределения маршрутов, позволяют определить, как в обеспечении перевозок могут использоваться существующие и проектируемые транспортные сети. Среди других моделей следует выделить модели кратчайшего пути (основанные на алгоритмах построения дерева решений и алгоритмах так называемого типа «все или ничего») и модели сетей минимальной стоимости (основанные на алгоритмах типа «беспорядок»). Другие модели распределения, известные под названием моделей разделения транспортных средств, предназначены для определения доли пассажиров, выбирающих то или иное конкретное транспортное средство.

При решении вопросов распределения транспортных средств используются два различных принципа [35]. Первый принцип, который можно назвать «оптимизацией для потребителя», основан на предположении о том, что в системе возможно установление некоторого равновесного состояния. Последнее характеризуется тем, что ни одно транспортное средство не имеет возможности сократить время пробега за счет изменения маршрута, так как маршрут выбран исходя из требования минимизации пути. Вторым принципом, который можно назвать «оптимизацией для системы», основан на минимизации среднего времени пробега.

Использование алгоритмов определения наиболее выгодного мар-

шрута (или кратчайшего пути) предполагает, что люди, совершающие поездку, выбирают наиболее выгодный, с их точки зрения, или близкий к нему маршрут («оптимизация для потребителя»).

Алгоритмы построения дерева решения позволяют решать проблему поиска оптимальных маршрутов путем формирования дерева транспортных линий, соединяющих определенный пункт отправления с различными возможными пунктами назначения, причем одновременно может быть учтена и стоимость проезда по соответствующим маршрутам. Дрейфус в работе [14], критически проанализировав модели кратчайшего пути, указал трудности, которые могут встретиться при практическом использовании результатов прогноза. По мнению автора, для устранения этих трудностей в такие модели следует вводить «запрет на разворот», с помощью которого могут быть учтены реальные трудности изменения направления движения транспорта на обратное и исключена возможность организации маршрутов типа «звезда». Заметим, что предположение, лежащее в основе «оптимизации для потребителя» (выбор наиболее выгодного маршрута), может не выполняться. Например, в работе [34] приводится много не только объективных, но и субъективных факторов, на основании которых люди, совершающие поездку, выбирают тот или иной маршрут движения. Поэтому затраты на эксплуатацию транспортной линии следует рассматривать как случайную величину, связанную с труднопрогнозируемым отношением людей к выбору маршрутов движения.

Другая проблема, относящаяся к определению стоимости проезда по транспортным линиям, состоит в том, что практически она должна учитывать степень интенсивности транспортного потока. Например, время проезда по главным транспортным линиям существенно увеличивается в часы пик, и тогда более приемлемым становится использование побочных транспортных линий. Проблема определения стоимости проезда в зависимости от степени интенсивности транспортного потока может быть решена с использованием метода ограниченной пропускной способности. Метод представляет собой повторяющийся (итеративный) процесс, в котором стоимость проезда по соответствующей транспортной линии увеличивается каждый раз, когда интенсивность транспортного потока превышает пропускную способность этой линии (описание этого метода можно найти в работе [8]).

Среди методов распределения маршрутов, относящихся к методам определения наиболее выгодного маршрута, наиболее распространенным является метод «все или ничего». Суть метода состоит в том, что весь транспорт, имеющий одни и те же пункты отправления и назначения, пропускается по наиболее выгодному маршруту, а все другие транспортные средства этим маршрутом не пользуются. Данный метод обеспечивает «оптимизацию для системы».

Как уже отмечалось выше, «оптимизация для потребителя» основана на предположениях, выполняющихся далеко не во всех случаях. Однако надо иметь в виду, что имеются пакеты программ, которые позволяют осуществлять «рассредоточение маршрутов» [5]. В частности, весь транспорт, перемещающийся между двумя пунктами, может распределяться не по одному, а по двум наиболее выгодным маршрутам, соединяющим эти пункты. Имеются возможности и для распределения транспорта по многим маршрутам [6].

«Оптимизация для системы» может осуществляться также с помощью алгоритмов построения сетей минимальной стоимости. Соответствующая проблема при условии задания пунктов отправления и назначения, а также в отсутствие ограничений на пропускную способность и ориентацию линий может быть сформулирована как задача линейного программирования, которая эффективно решается методами линейного программирования.

3. ОПРЕДЕЛЕНИЕ ПАРАМЕТРОВ ЗАДАЧИ И ОПИСАНИЕ РАЗЛИЧНЫХ ВАРИАНТОВ РЕШЕНИЯ

К настоящему этапу исследования становится ясно, должно ли оно завершиться принятием решения. Если должно, то для рассматриваемой задачи необходимо задать значения параметров и определить входные переменные. Так как описать процесс выполнения этой работы в общем виде довольно сложно, ниже приводится лишь перечень наиболее типичных входных переменных, с которыми приходится иметь дело при решении задач, возникающих как в частном, так и в государственном секторе транспорта (табл. 2).

Следует отметить, что правильность выбора каких-то показателей в качестве входных переменных системы зависит от уровня рассмотрения задачи и имеющихся в ней ограничений. Например, при рассмотрении вопросов планирования городского хозяйства постройка подземной скоростной транспортной магистрали (входная переменная) может оказаться совершенно неприемлемой из-за ограниченности фондов, выделяемых на соответствующие цели.

4. ОЦЕНКА ЭФФЕКТИВНОСТИ ВОЗМОЖНЫХ ВАРИАНТОВ РЕШЕНИЯ

После того как сформировано множество возможных вариантов решений, должна быть осуществлена оценка каждого из этих вариантов. При рассмотрении вопросов, относящихся к частному сектору транспорта, для проведения указанной оценки используются такие же методы, что и на этапе «Анализ возможных видов распределения» общей схемы исследования. Среди них сле-

Таблица 2

Перечень типичных входных переменных при решении задач, возникающих в частном и государственном секторах транспорта

Частный сектор	Государственный сектор
«Явные» переменные	
Пункты отправления и назначения Местоположение центров распределения продукции, складов, предприятий и других элементов транспортной системы Транспортные средства Частные, государственные или арендованные; автомобиль, железнодорожный состав или самолет и т. д. Маршруты и расписание движения транспорта	Транспортные потоки Ограничения на вместимость транспортных средств

«Неявные» переменные

Реклама Цены	Государственный транспорт Маршруты транспортных средств, пропускная способность линий, расписание движения, местоположение остановок городского транспорта Законодательство Законы образования государственных зон, принципы организации транспортных потоков
----------------------------	---

дует выделить методы составления расписания движения по заданным маршрутам, выбора транспортных средств, а также определения оптимального размещения складов, центров распределения продукции и других элементов транспортной системы. Что касается государственного сектора транспорта, то в этом случае при предсказании реакции общества на предполагаемые изменения в транспортной системе определенную роль должны сыграть модели распределения транспортных средств и возможных поездов.

4.1. Критерии, используемые при оценке возможных вариантов решения

Поскольку многие элементы, учитываемые в каждом отдельном возможном решении, тесно связаны друг с другом, они должны оцениваться одновременно. Это требует проведения анализа типа затраты—выгоды или затраты—полезность или, наконец, имитационного моделирования с использованием ЭВМ.

При оценке возможных вариантов решения для исследуемой транспортной системы обычно устанавливается соответствие между уровнем обслуживания и его стоимостью. В частном секторе транспорта разные уровни обслуживания характеризуются различием во времени доставки грузов и их состоянием, а также во времени удовлетворения специальных заявок. В государственном секторе транспорта разные уровни обслуживания характеризуются различной степенью загрузки транспортных сетей, например пропускной способностью воздушных линий, временем перевозок в городе на частном или на общественном транспорте, степенью доступности любого района города даже для небольшой части населения, уровнем травматизма, уровнем шума и другими показателями. Доступность рассматривается в основном с точки зрения времени проезда в тот или иной район города и равномерности распределения маршрутов государственного транспорта.

Для транспортных проблем большого масштаба оценку возможных вариантов решения редко удается проводить с использованием только количественных методов. Масштабность проблемы, как правило, вызывает присутствие в ее описании неформализуемых качественных показателей. Поэтому используемые критерии оценки возможных вариантов решения в ряде случаев могут выявлять лишь отдельные аспекты общей проблемы.

4.2. Методы анализа возможных вариантов решения

Имитационное моделирование с использованием ЭВМ представляет собой один из наиболее важных методов анализа. Самая существенная особенность такого моделирования состоит в том, что оно позволяет учесть многие качественные показатели транспортных систем. Иными словами, при некоторых предположениях относительно выявленных качественных показателей можно сопоставить возможные варианты решения, определив для каждого уровня стоимости обслуживания набор соответствующих ему выходных переменных.

Одной из лучших имитационных моделей, предназначенных для анализа транспортных систем, является модель LREPS, которая носит название «Крупномасштабный имитатор планирования в различных внешних условиях». С помощью данной

модели можно имитировать динамику процесса распределения упакованных грузов и на этой основе проводить сравнение сложившихся и возможных структур систем распределения продукции. Модель разработана в Университете шт. Мичиган [4].

Основными критериями, используемыми при анализе различных систем распределения продукции, являются пропускная способность линий обслуживания заказчиков (имеются в виду скорость и надежность) и соответствующая ей полная стоимость обслуживания. В общем случае с увеличением указанной пропускной способности повышается полная стоимость перевозок, что приводит к необходимости принятия компромиссного решения. Ниже приводятся некоторые особенности модели, позволяющие оценить степень ее эффективности.

1. Модель является динамической, так как позволяет осуществлять планирование во времени и учитывать последствия принимаемых решений.

2. Модель является стохастической, поскольку предоставляет исследователю возможность рассматривать объем потребностей, промежуток времени между поступлениями заказов, время передачи сообщений, длительность обработки заявок и время перевозки как случайные величины. Кроме того, имеется также возможность учитывать детерминированные связи элементов транспортной сети.

3. В рамках модели допускается комплексное рассмотрение проблем распределения производственных запасов и размещения предприятий, т. е. допускается одновременное исследование проблем временного и пространственного разделения элементов транспортной сети.

4. Модель является многоэтапной. На первом этапе осуществляется поочередный ввод данных о производственных предприятиях (их число может достигать 100), а также о складах, принадлежащих этим предприятиям. На втором этапе в модель вводятся сведения о центрах распределения разнообразной продукции (их число также может достигать 100). На третьем, последнем, этапе учитываются до 20 тыс. сведений, характеризующих потребности как отдельных заказчиков, так и групп заказчиков.

На каждом этапе и при переходе от этапа к этапу связь между отдельными элементами транспортной сети устанавливается с помощью величин потоков либо продукции, либо информации.

Модель была успешно использована при проектировании и анализе крупных систем распределения продукции для фирм различного профиля, производящих кожевенные, химические и агрохимические товары, замороженные продукты питания, оборудование для систем водо- и теплоснабжения.

Однако следует отметить, что эта модель имеет и некоторые недостатки. Так, для использования модели требуется обширная

информация, на получение которой не каждая фирма может выделить необходимые средства и время. Кроме того, модель мало пригодна для исследования вопросов, не имеющих прямого отношения к транспорту.

Наряду с рассмотренной моделью существуют и другие имитационные модели для планирования и анализа систем распределения продукции. Наиболее полезной из них оказалась модель, имеющая название «Имитатор систем распределения» (DSS) [10]. Характерной особенностью этой модели является целенаправленная организация вопросов, заключающаяся в том, что переменные параметры имитационной модели определяются по ответам типа да — нет на 450 вопросов. Информация, получаемая с помощью такого вопросника, включает характеристики спроса и покупательной способности, стратегии формирования портфеля заказов и пополнения производственных запасов, размещения предприятий, их производственных возможностей, а также характеристики других важных элементов системы. Существенно, что все эти элементы в большинстве случаев могут быть легко проанализированы. В зависимости от ответов на поставленные вопросы могут быть сформированы 10^{12} различных вариантов рассматриваемой имитационной модели. При формировании этих вариантов используется специальная библиотека программ.

После того как определена структура имитационной модели, с помощью программы-редактора составляется специальный перечень информации, необходимой для проведения анализа. Благодаря такому перечню затраты на получение необходимой информации могут оказаться существенно меньше выгод от проведения данного исследования, поскольку для использования модели надо получать только ту информацию, которая действительно необходима.

Размерность модели может быть достаточно просто уменьшена путем соответствующей корректировки используемого вопросника. Такая возможность делает модель весьма ценной для средних по величине и небольших фирм. Другой полезной для небольших фирм особенностью рассматриваемой модели является то, что для ее применения нет необходимости проводить большое число отладок программы. Однако, видимо, самым главным ее достоинством является возможность установления взаимосвязей между различными элементами системы. Необходимо также иметь в виду, что при использовании данной модели исследователь не только участвует в процессе моделирования, но и существенно влияет на него через вопросник. При этом исследователь лучше понимает модель и больше ей доверяет.

Хотя подобная программа может показаться излишне автоматизированной и потому не слишком полезной, однако благодаря ей обеспечивается адекватность модели для весьма сложных

исследуемых систем. С помощью модели DSS можно анализировать сведения относительно трех групп элементов системы: сведения о потребностях, складах продукции и производственных ресурсах. Связи между этими элементами могут носить произвольный характер, каждый канал распределения продукции может характеризоваться двумя направлениями ее перевозок, все склады продукции могут быть связаны друг с другом, потребности могут возникать в любой точке системы, и, наконец, на каждом складе могут быть созданы запасы любого вида. При использовании этой модели оценки характеристик системы в основном определяются с помощью оценок подсистемы регионального распределения складов. К сожалению, в этой модели отсутствует возможность одновременного рассмотрения временных и пространственных параметров системы, а также возможность исследования ее динамических характеристик. Такие возможности, как уже указывалось, предоставляет рассмотренная модель.

В настоящее время многие фирмы хорошо знают, что правильное планирование систем распределения, проводимое на агрегированном уровне, может привести к весьма значительной экономии средств. Рассмотренные выше имитационные модели представляют собой как раз инструмент такого планирования. Они позволяют обеспечивать руководителей фирм информацией, на основе которой можно организовать или реорганизовать транспортную сеть таким образом, чтобы удовлетворить текущие или будущие потребности в перевозках грузов. Следует отметить, что для многих фирм создание совершенно новой системы распределения продукции может быть непринемлемо. Поэтому очень важно, что такое ограничение, как и любое другое, может быть учтено в имитационных моделях рассмотренных типов.

Имитационные модели, основанные на использовании ЭВМ, применяются и при анализе систем государственного сектора транспорта, но реже, чем в случае частного сектора. Одна из таких моделей была применена при анализе различных вариантов организации транспортных сетей городов Бостона, Хьюстона, Нью-Хейвена и Тусона [18].

Однако при всех достоинствах рассматриваемых имитационных моделей они все же обладают серьезным недостатком, связанным с невозможностью включения в модель таких переменных, которые не могут быть выражены количественно. К моделям, в которых делается попытка рассмотрения и количественных, и «качественных» переменных, относятся модели типа затраты — выгоды и затраты — эффективность. В моделях первого типа различным уровням общих затрат ставится в соответствие совокупность количественных и качественных оценок различных уровней обслуживания. Во многих случаях такое сопоставление осуществляется

с использованием определенного типа показателей относительной стоимости. Однако при рассмотрении ряда важных показателей, таких, например, как загрязнение окружающей среды, влияние заторов на движение транспорта, уровень обслуживания небольших групп пассажиров, степень безопасности пользования транспортом, становится ясно, что сведение их к показателям стоимости является практически невозможным.

Данное обстоятельство привело к тому, что была разработана модель типа затраты—эффективность. Главная цель разработки этой модели состояла в том, чтобы для каждого возможного значения полных затрат представить оценки показателей уровня обслуживания в наиболее ясном и естественном для этих показателей виде. В качестве примеров показателей, о которых здесь могла бы идти речь, можно указать такие показатели, как время в пути, количество транспортных происшествий, уровень загрязнения окружающей среды, уровень комфортабельности. Возможные варианты решения для транспортной системы, представленные в естественном виде через качественные показатели рассмотренного выше типа, могут быть оценены в некоторой балльной шкале группой экспертов, аналогично тому как преподаватель оценивает школьные работы.

В исследовании [12] возможные варианты решения были разбиты на группы в соответствии со следующей классификацией:

- транспортная сеть без изменений;
- улучшение транспортной сети путем ее модернизации,
- улучшение транспортной сети за счет существенного изменения некоторых элементов системы, при котором система в целом не претерпевает глобальной перестройки;
- улучшение транспортной сети за счет существенных изменений во всей системе, вызванных новыми научными открытиями или новыми техническими решениями, например за счет создания надземной скоростной транспортной сети.

Экспертам было предложено оценить эти группы различными способами, используя также показатели, как уровень комфортабельности, выгоды, надежность и безопасность. Другие показатели определялись на основе мнений экспертов о предполагаемой реакции общественности на реализацию возможных вариантов решения из числа входящих в четыре названные выше группы. Сделанные оценки использовались затем для выработки экспертами общей точки зрения.

Модели типа затраты—выгоды и типа затраты—эффективность являются наиболее полными среди существующих моделей анализа транспортных систем. Эти модели, по существу, объединяют отдельные частные модели (главным образом основанные на использовании ЭВМ имитационные модели) с методами учета качественных переменных. При этом формируются возможные варианты

решения, для которых проще, чем в других моделях, устанавливается связь ожидаемых результатов с общими затратами. Однако необходимо иметь в виду, что выбор окончательного варианта решения требует со стороны лиц, принимающих решения, большой работы для правильной оценки рассматриваемых вариантов.

5. РАЗРАБОТКА МАКЕТОВ ТРАНСПОРТНОЙ СЕТИ И ОЦЕНКА ВОЗМОЖНЫХ ВАРИАНТОВ РЕШЕНИЯ

Действующие макеты довольно широко используются на практике для оценки проектируемых предприятий промышленности или производственных процессов. Аналогичные способы проверки различных вариантов решения используют и организации, контролирующие состояние шоссежных дорог. При этом применяются имитационные макеты, включающие в рассмотрение не только основные транспортные линии, но и тротуары, дорожки, проезды, а также систему сигнализации.

В последнее время все большее внимание уделяется макетам и имитационным моделям как средствам окончательной проверки возможных вариантов решения в области транспорта. Замечательные успехи в авиационной промышленности привели к широкому распространению хорошо зарекомендовавших себя имитационных моделей-макетов при проектировании транспортных сетей и управлении ими как в государственном, так и частном секторе.

Например, с использованием такого подхода для решения задач государственного сектора был разработан маршрут автобуса-экспресса, когда требовалось обеспечить проезд по 25 основным городским магистралям. Применение же имитационных моделей-макетов в частном секторе показало, что планирование прибыли от сдачи в аренду вагонеток для однопколейных дорог должно осуществляться Национальной компанией, предоставляющей аренду грузового транспорта, не в масштабах всей страны, а в масштабах отдельных городов.

Наряду с существенным уменьшением затрат на перестройку исследуемой транспортной сети и уменьшением вероятности возникновения непредвиденных последствий принимаемых решений использование макетов имеет еще одно важное достоинство, заключающееся в возможности ведения «обратной связи». Эта обратная связь либо может быть организована в обычном для технических систем виде, либо может представлять реакцию общественного мнения на тот или иной вариант решения. Отметим, кстати, что в последнее время общественному мнению отводится все большая роль при выборе наилучшего варианта решения.

6. РЕАЛИЗАЦИЯ ВЫБРАННОГО ВАРИАНТА РЕШЕНИЯ

В данном разделе дается краткий обзор проблем реализации выбранных вариантов решений и причин, задерживающих эту реализацию. Такие проблемы и причины должны быть выявлены для того, чтобы исследователи могли оказывать непосредственное влияние на процесс принятия решения.

6.1. Особенности процесса принятия решения

К настоящему времени накоплено немало экспериментальных и теоретических данных, показывающих сходство процессов принятия решения в системах государственного и частного транспорта. В обоих случаях трудности, которые возникают в процессе принятия решения, приводят к задержкам в реализации рекомендаций, вырабатываемых с применением общих моделей.

Одна из проблем реализации таких рекомендаций связана с наличием разных интересов у различных предприятий. Существование разных интересов приводит к тому, что в процессе принятия решения определенную роль играют субъективный фактор и фактор «осторожности» (минимального риска). Выделенные факторы проявляются, например, в тех случаях, когда оптимальный вариант решения, предлагаемый для некоторого предприятия, является для него новым и неопробованным. Тогда фактор «осторожности» может привести даже к полному неприятию предложенного варианта. Таким образом, наличие целого ряда особых обстоятельств приводит к формированию либо положительного, либо отрицательного отношения к решениям определенного типа.

Другой проблемой, весьма близкой к только что рассмотренной, является проблема, связанная с наличием групп лиц, имеющих разные интересы. Каждая из таких групп поддерживает и отстаивает варианты решения определенного типа. Если учитывать динамику развития отдельных групп, характер поддерживаемых ими решений вполне можно предсказывать заранее. Аналитические транспортные модели строятся, как правило, без учета соображений об интересах различных групп. Поэтому вряд ли можно считать, что полученный с помощью модели тот или иной вариант решения будет удовлетворительным с точки зрения всех заинтересованных сторон. Принятое решение по организации транспортных сетей почти всегда вызывает оппозиционное отношение одних групп, поддержку других и равнодушное отношение со стороны остальных групп. Это обстоятельство приводит к тому, что принятое решение должно затем согласовываться аналогично процедурам политического характера (это справедливо как для государственного, так и для частного сектора транспорта).

6.2. Необходимость учета состояния окружающей среды и мнения людей в процессе принятия решений

Другие факторы, задерживающие практическое применение результатов анализа транспортных систем, связаны с необходимостью учета состояния окружающей среды и мнения людей при выборе наиболее желательного варианта решения. Эти факторы приходится учитывать при организации многих транспортных сетей государственного и частного секторов.

В последние годы в США, в отдельных штатах и округах, появились законы, требующие анализа всех предложений по организации транспортных сетей с точки зрения обеспечения условий, гарантирующих отсутствие отрицательного влияния транспорта на человека и окружающую среду. В связи с этим требованием моделирование транспортных систем должно включать дополнительный этап тщательной доработки модели, на котором делается попытка максимально возможной компенсации отрицательных воздействий транспорта на человека и окружающую среду. С помощью соответствующих переменных, учитывающих указанные воздействия и включаемых в общие модели анализа, удалось сделать некоторые оценки в рассматриваемой области, однако полученные данные оказались пока малоубедительными и дорогостоящими [28].

Таким образом, наличие воздействия транспорта на окружающую среду и человека требует некоторой модификации оптимальных решений, вырабатываемых без учета этих воздействий. Поэтому в будущих исследованиях нужно осуществлять такой учет и включать в количественном или хотя бы в качественном виде в анализ все необходимые для этого параметры, переменные и ограничения.

7. КЛАССИЧЕСКИЙ ПРИМЕР ПРАКТИЧЕСКОГО ИСПОЛЬЗОВАНИЯ ОБЩЕЙ СХЕМЫ ИССЛЕДОВАНИЯ ТРАНСПОРТНОЙ СИСТЕМЫ

В качестве примера, демонстрирующего использование общей схемы анализа транспортных систем (рис. 1), рассмотрим классическое исследование Лесли С. Эди [16], которое было выполнено для Управления порта Нью-Йорка и посвящено проблеме простоев автомобильного транспорта во время оплаты пошлин.

Работа [16] представляет собой первую попытку применения методов исследования операций к решению проблемы, относящейся к сфере деятельности Управления порта. Хотя после 1952 г., когда началось исследование Л. Эди, было выполнено немало других прикладных работ подобного рода, именно это исследование является наиболее известным и получившим наиболее широкое признание (в частности, в 1955 г. оно было удостоено Первой Манчестерской премии).

Рассмотрим сначала имевшиеся исходные данные. Управление порта осуществляло сбор пошлин за пользование находившихся в его ведении тоннелей и мостов. Необходимость обеспечения этих работ соответствующим персоналом привела к существенному росту затрат, осуществляемых Управлением порта. Поэтому оно заинтересовалось возможностью уменьшения числа сборщиков пошлин при условии сохранения постоянно высокого уровня обслуживания и разумного распределения периодов отдыха данной категории работников. Проблема сводилась к нахождению наилучшего компромиссного решения, позволяющего свести к минимуму простой транспорта при минимальных затратах на организацию сбора пошлины.

До проведения исследования Л. Эди для определения необходимого числа сборщиков пошлин использовался некий произвольный норматив, непосредственно не связанный с уровнем обслуживания. Произвольные нормативы использовались также при распределении сборщиков по рабочим местам и при определении количества кабин для сбора пошлин, которые должны быть открыты в тот или иной момент времени. При указанном методе сбора пошлин время простоя одного автомобиля составляло от 2 до 5 с. Однако при более формализованном подходе к решению проблем организации (на основе методов теории вероятностей) предполагалось возможным сократить среднее время простоя за счет выявления реальной зависимости между интенсивностью обслуживаемого транспортного потока, количеством каналов обслуживания и уровнем обслуживания.

Для разработки указанного подхода были сформулированы следующие конкретные цели исследования:

- оценка уровня обслуживания и определение характера его изменения в зависимости от интенсивности потока транспорта и количества каналов обслуживания;
- определение оптимальных уровней обслуживания;
- разработка более точного по сравнению с существующим метода контроля за затратами и объемами обслуживания, обеспечивающего также сборщикам пошлин возможности для отдыха.

По существу, достижение этих трех целей возможно с помощью методов теории массового обслуживания, оптимизации и составления расписания.

7.1. Прогнозирование потребностей в перевозках людей и грузов

Первый этап общей схемы (рис. 1) состоит в прогнозировании потребностей в транспортных перевозках. Для составления прогноза в рассматриваемом исследовании было подсчитано количество транспортных средств, прибывающих на отдельную площадку каждые 30 с, и количество транспортных средств, составляющих

очереди у открытых кабин для сбора пошлин. Полученные результаты были сгруппированы и представлены для интервалов времени длительностью по 20 мин. После этого для каждого интервала определяется степень занятости сборщиков пошлин и среднее время обслуживания автомобилей или время их простоя в очереди. Эти величины использовались для установления эмпирических зависимостей, которые затем сопоставлялись с теоретическими зависимостями, полученными на основе классической теории массового обслуживания для одного и нескольких одновременно работающих пунктов обслуживания. В дальнейшем эмпирические кривые оказались весьма полезными для прогнозирования интенсивности потоков прибывающего транспорта и времени его простоя.

7.2. Анализ возможных видов распределения транспортных средств

В результате проведенного исследования было установлено, что среднее время простоя одного автомобиля уменьшается, если увеличивается интенсивность потока автомобилей, приходящихся на одну пропускную линию.

Кроме того, был получен весьма важный побочный результат, который состоял в выявлении отрицательного влияния на уровень обслуживания правостороннего расположения кабин для сбора пошлин. При правостороннем расположении кабины автомобиль может подъехать к ней лишь со стороны правой дверцы, в то время как водитель занимает место у левой дверцы. Такое расположение кабин приводит к существенным задержкам транспорта. Поэтому Управление порта реконструировало все основные площадки таким образом, чтобы было обеспечено только левостороннее расположение кабин для сбора пошлины.

7.3. Определение управляющих переменных и описание различных вариантов решения

В рассматриваемом случае было совершенно очевидно, что проблема принятия решения действительно существовала. Для выработки решения были установлены такие зависимости, которые являются типичными для задач массового обслуживания. Так, например, была выявлена зависимость времени нахождения транспорта на платной площадке от числа работающих кабин и степени заполнения данной площадки, а также зависимость среднего времени обслуживания одного автомобиля от интенсивности потока транспорта, попадающего в течение часа на отдельную пропускную линию, и от числа пропускных линий.

Как только были выявлены эти два вида зависимостей, соответствующая задача массового обслуживания получила свое реше-

ние, а параметры системы обслуживания оказались определенным явным образом.

Что касается оптимизационной задачи, то ее решение сводилось к определению правила, регулирующего открытие и закрытие пропускных линий и обеспечивающего удовлетворение имеющейся потребности в обслуживании. После оценки возможных вариантов решения был выбран вариант, который сводился к следующему: поддерживать открытыми пропускные линии в таком количестве, какое необходимо для обеспечения задержки машины длительностью в среднем не более 11 с, ограничить максимальную длину очереди величиной, которая не более чем на три единицы превышает величину, определяемую по кривой пуассоновского распределения.

Задача составления расписания, в основном касающаяся распределения сборщиков пошлин по рабочим местам, оказалась наиболее сложной из трех задач, сформулированных в предыдущем разделе. Вначале для решения задачи был использован метод проб и ошибок, в основе которого лежит построение диаграммы Ганта. Однако получаемые при этом результаты оказались неудовлетворительными. Тогда для решения задачи было использовано линейное программирование. Но и этот подход также дал неудовлетворительные результаты, в частности потому, что нецелочисленные значения переменных в рассматриваемой задаче не имели смысла.

Тем не менее удовлетворительный метод решения задачи составления расписания был все же найден. Этот метод представлял собой некоторую последовательность повторяющихся расчетов, в результате которых определяются правила назначения основного и запасных сборщиков пошлин. С помощью такого метода оказалось возможным составлять различные расписания.

7.4. Оценка возможных вариантов решения

После того как были четко описаны задачи массового обслуживания, оптимизации, составления расписания и определены возможные варианты решения этих задач, необходимо было перейти к выбору такого варианта, который удобен для реализации.

На данном этапе исследования Л. Эди не стал строить модель, объединяющую три частные модели, а каждую из последних проанализировал отдельно, сопоставив затем полученные результаты. В частности, используя два вида зависимостей, найденных для задачи массового обслуживания, Л. Эди сумел получить зависимость времени простоя от интенсивности входящего транспортного потока и от числа пропускных линий. Далее, в результате экспериментальной проверки ряда возможных вариантов решения исследователь выбрал один из них в качестве оптимального варианта

решения задачи составления расписания. Этот вариант был затем использован при формировании исходных данных для оптимизационной задачи.

7.5. Разработка макетов транспортной сети и окончательная проверка возможных вариантов решения

Для оценки предложенных вариантов решения задач массового обслуживания и оптимизации была проведена их широкая проверка. Кроме того, было составлено ориентировочное расписание работы пропускных пунктов, в котором сборщикам пошлины выделялось время на отдых и прием пищи.

7.6. Реализация выбранного варианта решения

Рекомендованные решения были реализованы Управлением порта.

7.7. Обратная связь и последующие улучшения

Результаты работы [13] позволили осуществить следующие мероприятия в области организации сбора пошлины:

- улучшение контроля за уровнем обслуживания и уровнем соответствующих затрат на основе четких представлений о зависимости между ними;
- реконструкция пропускных линий, обеспечивающих левостороннее расположение кабин сбора пошлины;
- разработка устройства для автоматической печати транспортных счетов и их регистрации на перфокартах;
- разработка специальных камер для регистрации транспорта, способствующих более быстрой реакции системы на изменяющиеся условия функционирования;
- формирование исходных данных для проектирования автоматизированной системы сбора пошлины.

Исследование, выполненное Л. Эди, послужило стимулом для использования методов исследования операций при решении других проблем, связанных с организацией сбора пошлины, например проблемы пешего и моторизованного патрулирования.

8. ТИПИЧНЫЕ ПРОБЛЕМЫ, ВОЗНИКАЮЩИЕ ПРИ ИССЛЕДОВАНИИ ОБЩЕЙ СХЕМЫ АНАЛИЗА ТРАНСПОРТНОЙ СИСТЕМЫ

8.1. Проблема сбора необходимой информации и формирование исходных данных

Сведения, доступные исследователям транспортных систем, как правило, являются недостаточными для проведения анализа. Обычно отсутствие необходимых данных объясняется тем, что транспортные перевозки требуют больших затрат. Действитель-

но, получение качественной информации, необходимой для анализа и моделирования транспортных систем, является дорогостоящей процедурой. Ряд примеров анализа систем государственного сектора транспорта показал, каких громадных затрат требует получение соответствующей информации [25]. Аналогичное положение дел характерно и для частного сектора транспорта. Однако все это не может служить основанием для того, чтобы вообще не заниматься сбором необходимых данных.

Общая причина возникновения проблем при сборе информации и формирования на ее основе исходных данных для анализа состоит в необходимости рассмотрения и измерения большого числа качественных и неявных переменных. Что касается явных и количественных переменных, то им часто приходится присваивать какие-то значения в условиях полного или почти полного отсутствия соответствующих исходных данных. Данное обстоятельство вносит в транспортные модели некоторый элемент неоднозначности. Более того, поскольку во многих случаях имеет место субъективизм при определении конкретных значений для ряда переменных, то этот субъективизм сохраняется и во всей процедуре анализа.

Когда наряду с проблемой сбора информации и получения на ее основе исходных данных возникают проблемы измерения хотя и явных, но являющихся качественными переменными, то появляется новая трудность в успешном осуществлении анализа. Эта трудность связана с проблемой увеличения ошибок исследования. В ряде работ показано, что наличие в транспортных моделях достаточно сложных математических зависимостей в совокупности с трудностями сбора информации и формирования данных, относящихся к неявным и качественным переменным, может приводить к увеличению ошибки в процессе проведения анализа до 50% [13].

8.2. Традиционность в оптимальных решениях

То обстоятельство, что оптимальные варианты решения транспортных задач обладают определенными традиционными тенденциями, послужило основанием для критических замечаний по поводу целесообразности применения методов исследования операций при анализе транспортных систем. Указанная особенность оптимального решения обусловлена тем, что при анализе транспортных систем необходимо учитывать ограничения, которые являются довольно стабильными. Кроме того, если наряду со стабильностью ограничений имеет место субъективизм при оценке переменных или при исключении из анализа тех переменных, которые с трудом поддаются измерению, то получаемые решения становятся ординарными, а в некоторых случаях просто тривиальными. Например, при планировании городского транспорта главные вы-

годы той или иной организации транспортной сети часто связывают лишь с экономией времени пассажиров, с минимизацией общей протяженности маршрутов и с уменьшением эксплуатационных расходов. Однако совершенно очевидно, что действительные выгоды для людей, пользующихся городским транспортом, связаны не только с экономией времени и что решение, обеспечивающее минимизацию общих затрат, может быть вовсе не тем решением, которое в действительности требуется. Еще один довод против применения на транспорте традиционных решений связан с тем, что в настоящее время действующие транспортные сети требуют коренных изменений в целом, а такие изменения не всегда соответствуют оптимальным вариантам решений, получаемых с помощью количественных моделей.

Подход, который может обеспечить большую гибкость применения моделей для анализа транспортных систем, состоит в нахождении не только оптимальных вариантов решения, но и вариантов, которые могут быть приняты на основании детализации, обобщения, модификации, экстраполяции или корректировки оптимального варианта. По аналогии с подходом, предложенным в работе [12], можно было бы выделить следующие четыре группы вариантов решений [7].

Первая группа включает оптимальный вариант решения исходной задачи и является как бы отправной для формирования остальных групп.

Вторая группа включает варианты так называемых минимальных решений, которые являются модификациями оптимального варианта и выбираются по принципу наименьших затрат времени, усилий и ресурсов на реализацию решения. Можно считать, что в минимальных решениях экстраполируются текущие потребности в транспортных перевозках.

Третью группу составляют варианты решения, называемые максимальными, для осуществления которых требуется использование максимально возможного количества ресурсов и реализация которых позволяет максимально улучшить транспортную сеть. Хотя, как правило, ресурсные ограничения таковы, что оптимальное решение входит в число максимальных, тем не менее в некоторых случаях это не так. Например, при организации строительства некоторой автомагистрали рекомендуемое решение не обязательно требует использования всех средств, выделенных из фондов Департамента автомобильных дорог.

Имеющийся практический опыт позволил установить, что наиболее приемлемые решения находятся в области между максимальными и минимальными решениями. Выделенная в этой области группа нормативных решений представляет собой последнюю из рассматриваемых групп. Нормативные решения могут быть как непосредственно связаны, так и не связаны с оптималь-

ным решением. При формировании данной группы используется принцип «необходимого действия», а не принципа оптимальности, заложенный в аналитической модели.

Таким образом, формирование рассмотренных групп решений или подучающихся из них смешанных групп позволяет использовать оптимальные решения в качестве основы для формирования возможных вариантов решения. Это множество вариантов решений затем оценивается с точки зрения рассмотренных выше проблем и с точки зрения возможности их реализации.

ЗАКЛЮЧЕНИЕ

На основе материала, изложенного в данной главе, могут быть сделаны следующие основные выводы относительно использования методов исследования операций при анализе систем государственного и частного секторов транспорта.

При анализе транспортной системы частного сектора главную роль играют методы прогнозирования (о чем свидетельствуют результаты применения многочисленных моделей распределения и целей поездок, специально предназначенных для предсказания потребностей населения в транспортных средствах). Это объясняется двумя причинами. Во-первых, в частном секторе контроль за удовлетворением большей части потребностей в транспортных средствах осуществляется централизованно соответствующими органами управления определенного предприятия; в государственном секторе аналогичный контроль, по существу, распределен по всей системе и осуществляется миллионами людей.

Во-вторых, целью планирования в частном секторе, как правило, является максимизация прибыли, в то время как в государственном секторе цель планирования -- это удовлетворение потребностей миллионов людей в транспортных средствах. Поэтому часто довольно трудно определить критерий оценки возможных вариантов решений при планировании городской транспортной сети, и возникает необходимость выработки стратегии, позволяющей устанавливать правильность и важность выбираемых целей.

Для планирования различных и сходства в моделях, используемых в частном и государственном секторах транспорта, рассмотрим модели, позволяющие сделать выбор транспортных средств. С помощью таких моделей проблема выбора транспортных средств как в частном, так и в государственном секторе решается путем определения соответствующих затрат и выгод. Однако в государственном секторе выбор транспортных средств осуществляется, по существу, населением, и используемые модели распределения предназначены для предсказания интенсивности потока пассажиров, выбирающих каждое из имеющихся транспортных средств. В частном секторе управление процессом выбора транспортных средств является функцией соответствующего планирующего органа, и поэтому модели для выбора транспортных средств направлены на выявление одного или нескольких транспортных средств, наилучших в имеющейся парке. Следовательно, модели, используемые в частном секторе, являющиеся оптимизационно-ориентированными, а модели, используемые в государственном секторе, -- прогностически-ориентированными. Это положение справедливо для большинства моделей, применяемых на различных этапах общей схемы планирования транспортных сетей.

Что касается перспективы использования методов исследования операций для решения проблем транспортной сети, то, по-видимому, сложные и дорогостоящие имитационные модели должны уступить место менее сложным и менее общим моделям. Несомненно также, что более широкое применение найдут имитационные и операционные игры.

ЛИТЕРАТУРА

1. Alonso W., Predicting Best with Imperfect Data, *J. Am. Inst. Planners*, 34, № 4, 248—255 (July 1968).
2. Appa G. M., The Transportation Problem and its Variants, *Operational Res. Quart.*, 24, № 1, 79—99 (March 1973).
3. Bluestone D. W., U. S. Airline Passenger Traffic Will Scar in the Seventies, *Airline Management and Marketing Including American Aviation*, 2, № 1, 24—32 (January 1970).
4. Bowersox D. J., Coping: Dynamic Simulation of PD, *Distribution World Wide*, December 1972, pp. 24—31.
5. Bowersox D. J., Planning Distribution Operations with Dynamic Simulation, *J. Marketing*, January 1972, pp. 17—25.
6. Brokke G. E., Urban Transportation Planning Computer System, American Association State Highway Officials Conference, Minnesota 1967.
7. Burrell J. E., Multiple Route Assignment and its Application to Capacity Restraint, Fourth International Symposium on Theory of Traffic Flow, Karlsruhe, 1968.
8. Catanese A. J., Steiss A. W., Systemic Planning: Theory and Application, D. C. Heath, Boston, 1970.
9. Chicago Area Transportation Study, Illinois Department of Public Works and Buildings in Cooperation with the Bureau of Public Roads, U.S. Department of Commerce, Vols. I—III, 1959, 1960, 1962.
10. Clark C., Peters G. H., The Intervening Opportunities' Method of Traffic Analysis, *Traffic Quart.*, 19, № 1, 101—119 (January 1965).
11. Connors M. M., Coray C., Cucarro C. J., Green W. K., Low D. W., Markowitz H. M., The Distribution System Simulator, *Management Sci.*, 18, № 8, 425—453 (April 1972).
12. Creighton R. L., Urban Transportation Planning, University of Illinois Press, Chicago, 1970.
13. Dodson E. N., Cost-Effectiveness in Urban Transportation, *Operations Res.*, 17, № 3, 373—394 (May-June 1969).
14. Dreyfus S. E., An Appraisal of Some Shortest-Path Algorithms, *Operations Res.*, 17, № 3, 395—412 (May-June 1969).
15. Eatherly B. J., Summary of a Simple Model of Mode Choice, *Management Sci.*, 19, № 2, 201—204 (October 1972).
16. Edie L. C., Traffic Delays at toll Booths, *Operations Res.*, 11, № 1, 107—138 (May — June 1954).
17. Fratar T. J., Forecasting Distribution of Interzonal Vehicular Trips by Successive Approximation, Highway Research Board Proceedings, 33th Annual Meeting, 1954, pp. 376—384.
18. Garrison W. L., Marble D. F., A Prolegomenon to the Forecasting of Transportation Development, Transportation Center Report, Northwestern University, August 1969.
19. General Research Corporation, Systems Analysis of Urban Transportation, Vols. 1—4, January 1968.
20. Heanue K. E., Pyers C. E., A Comparative Evaluation of Trip Distribution Procedures, Public Roads, June 1966, pp. 43—51.
21. Heggie I. G., Are Gravity and Interaction Models a Valid Technique for Planning Regional Transport Facilities?, *Operational Res. Quart.*, 20, № 1, 93—110 (March 1969).
22. Kirby R. F., A Preferencing Model for Trip Distribution, *Transport Sci.*, 4, № 1, 1—35 (February 1970).
23. Kriebel F. E., Transportation in the 70's, *Distribution World Wide*, July 1970, pp. 39—42.
24. Lee Douglas B., Requiem for Large Scale Models, *J. Am. Inst. Planners*, 39, № 3, 163—178 (May 1973).

25. McQuei R., Cargo Vehicle Productivity, *Management Sci.*, 18, № 2, B36—B51 (October 1971).
26. Mouchahoir G. E., Economy in Data Collection for Planning Studies, *Traffic Quart.*, 26, № 2, 289—298 (April 1972).
27. Parrish R. L., Lockheed's «Paper Airline», *Management and Marketing Including American Aviation*, January 1970, pp. 64—65.
28. Potts R. B., Oliver R. M., *Flows in Transportation Networks*, Academic Press, Inc., N. Y., 1972.
29. Southeast Wisconsin Regional Planning Commission, *Recommended Transportation and Land Use Plan: 1990*, Milwaukee, 1966.
30. Stauffer S. A., Intervening Opportunities: A Theory Relating Mobility and Distance, *Amer. Soc. Rev.*, 5, 845—867 (1940).
31. Tomazinis A. R., A New Method of Trip Distribution in an Urban Area, *Highway Research Board Bulletin*, 347, 77—99 (1962).
32. Tomlin J. A., A Mathematical Programming Model for the Combined Distribution-Assignment of Traffic, *Transport, Sci.*, 5, № 2, 122—140 (1971).
33. Verleger P. K., Jr., Models for the Demand for Air Transportation, *The Bell Journal of Economics Management Science*, 3, № 2, 437—457 (1972).
34. Voorhees A. M., A General Theory of Traffic Movement, *Proc. Inst. Traffic Engineers*, 1955, pp. 46—56.
35. Wachs M., Relationships Between Drivers' Attitudes Toward Alternate Routes and Driver and Route Characteristics, *Highway Res. Board Record*, 197, 70—87 (1967).
36. Wardrop J. G., Some Theoretical Aspects of Road Traffic Research, *Proc. Inst. Civil Engineers*, 1, Part II, 325—378 (1952).

II-5. Военные системы

*Э. Квейд*¹⁾

ВВЕДЕНИЕ

Примеры принятия решений военным командованием на основе анализа, который, не используя современную терминологию, следовало бы назвать «системным», можно найти даже в истории древних веков. Так, правитель Сиракуз, несмотря на глубокую убежденность, что все исходит от бога, в том числе и решения относительно ведения военных операций, все же обратился к Архимеду с просьбой помочь осажденному городу прорвать осаду римлян.

Тем не менее лишь с появлением науки управления (конец XIX — начало XX в.) анализу, напоминающему современную дисциплину «исследование операций», стали уделять серьезное внимание и разрабатывать соответствующие методы.

Применительно к военным системам (операциям) подобные методы иногда мало чем отличались от традиционных (кстати, недостаточно подробно освещенных даже в военных документах), которые просто усложнились математически. Например, схема составления оперативной сводки для «оценки обстановки», предусматривающая использование данных анализа, почти не отличалась от обычной схемы, принятой в военных штабах, за исключением того, что в первом случае больше внимания уделялось количественным оценкам.

1. ИСТОРИЧЕСКИЕ АСПЕКТЫ ИСПОЛЬЗОВАНИЯ МЕТОДОВ ИССЛЕДОВАНИЯ ОПЕРАЦИЙ В ВОЕННЫХ СИСТЕМАХ

Хотя термин «операционные исследования» или, в американском варианте, «исследование операций» впервые был введен в 1938 г. для характеристики рода деятельности небольшой исследовательской группы, созданной по инициативе Air Ministry Research Station, по существу, для военных организаций он не явился чем-то новым.

1.1. Первые исследования

Согласно документу, датированному 1886 г., военное командование выпущдено было провести системный анализ, чтобы принять решение относительно производства 12-дюймовых орудий, заряжающихся с казенной части и предназначенных для исполь-

¹⁾ Edward S. Quade, Ph. D., Head (Retired) of the Mathematics Department at the Rand Corporation in Los Angeles, California.

зования в береговой артиллерии. Необходимо было сделать выбор между орудием, принятым на вооружение и выпускаемым фирмой Кигурр, и орудием нового образца американского производства. С этой целью был проведен анализ соответствующих характеристик орудий данных образцов и затрат на их производство. Оказалось, что отношение эффективности к стоимости для орудия, принятого на вооружение, составляло всего лишь 0,8 аналогичного соотношения для орудия нового образца, при этом мерой эффективности служила «мощность» орудия (энергия снаряда на дульном срезе), выраженная в футо-тоннах. Так как военное командование не сочло возможным принять решение только на основе этих данных, то последовало распоряжение продолжать исследования.

Ниже приводятся выводы относительно второго этапа исследований [97].

С помощью хорошо известных методов анализа соответствующих характеристик артиллерийских орудий со сходной системой заряжения можно легко определить калибр и вес орудия каждого образца при данной мощности орудия или при условии, что для однотипных орудий задана одна и та же мощность. Таким образом все данные можно привести к абсолютному стандарту и тем самым придать анализу строго количественный характер.

Результаты второго этапа исследований показали, что стоимость орудия нового образца с той же мощностью, что и орудия, принятого на вооружение, составляла лишь 0,8 стоимости последнего. Более того, после того как подсчитали стоимость уже существующего орудия с той же мощностью, что и орудие нового образца, оказалось, что она почти на 20% превышает стоимость последнего. Тем не менее было принято однозначное решение рекомендовать к производству орудие, принятое на вооружение. В этой связи интересно заметить, что доводы, которые приводились в пользу такого выбора, можно нередко услышать еще и сегодня, например:

...К сожалению, мы не располагаем данными относительно прочности орудий нового образца...

...Вы же видите, разница в стоимости производства небольшая, и если учесть, что дешевизна, как утверждают, является чуть ли не главным достоинством орудия нового образца, то...

...дело в том, что производство орудий в таком количестве, несомненно, пагубно отразится на развитии сталелитейной промышленности страны...

...После стольких лет ожиданий возможности использования опыта ведущих в области производства тяжелой артиллерии стран, когда наконец такая возможность стала реальностью и мы можем начать производство средств защиты страны (защиты, которой длительное время пренебрегали), было бы крайне неразумно с нашей стороны выбрать для производства пушек материал, который европейские державы отвергли после многократного опробования...

Если обратиться к истории 1-й мировой войны, то можно найти немало примеров использования методов исследования операций при решении проблем, в частности, связанных с действиями военно-морских и сухопутных сил США. Рассмотрим два примера. Один из них относится к области ведения борьбы против подводных лодок.

Проведение исследований в этой области было поручено Т. А. Эдисону, которого, по общему мнению, привлекли к работе не столько в качестве исполнителя, сколько в качестве эксперта, учитывая его талант изобретателя и интеллект ученого. Эдисон начал свою деятельность с проведения экспериментов с подводными акустическими устройствами с целью их усовершенствования. Однако вскоре он изменил направление исследований, придя к выводу, что усовершенствование акустических устройств, предназначенных для обнаружения подводных лодок, являются лишь одним из направлений решения главной задачи — спасти корабли от нападения.

В связи с этим летом 1917 г. Эдисон с тремя помощниками выехал в Вашингтон, где намеревался тщательно изучить статистические данные о деятельности подводных лодок и использовать их для разработки стратегических планов борьбы с подводными лодками. Эдисон рассчитывал найти в правительственных документах ясные, подробные и привязанные к картам сведения о случаях потопления судов. Не обнаружив ничего подобного, он был вынужден сам заняться подготовкой соответствующих сведений и составлением карт.

По этим данным было установлено, что пароходные компании составляют для своих судов те же маршруты, что и до войны. Согласно статистическим данным, только 6% судов было потоплено в ночное время... Очевидным стал и тот факт, что пароходные компании не усвоили уроков плаванья в ночное время в опасной зоне. Кроме того, результаты изучения данных показали, что только около 4% первоклассных торговых судов Великобритании имели на своем борту современную акустическую аппаратуру. Что касается радиоприемной аппаратуры, то на британских торговых судах она практически отсутствовала.

Располагая достаточными данными и фактами, Эдисон приступил к разработке плана, который в общих чертах заключался в следующем:

- в опасной зоне суда могут находиться только в ночное время;
- необходимо, насколько это возможно, отказаться от стандартных маршрутов;
- днем суда должны становиться на якорь в мелких водах и гаванях, с тем чтобы подводные лодки не могли к ним приблизиться;
- суда могут заходить в глубоководные порты согласно маршруту только в ночное время.

К сожалению, работа Эдисона почти не имела практического выхода. Вполне вероятно, что его отчеты могли затеряться среди других бумаг, и, по-видимому, только этим можно объяснить тот факт, что в начале 2-й мировой войны ВМС США вынуждены были заново открывать то, к чему пришел Эдисон в результате своих статистических исследований.

Следующий пример иллюстрирует применение методов исследования операций в сфере сухопутных войск США.

Ответственность за проведение подобных исследований была возложена на полковника Леопарда Айреса, усилиями которого был создан статистический отдел при генеральном штабе. Кроме того, Айресу удалось добиться разрешения докладывать о результатах работы непосредственно начальнику штаба, минуя промежуточные инстанции. Все это, несомненно, способствовало успешной разработке методов исследования операций и их практическому использованию. Вот что писал по этому поводу Айрес [3]:

...Получив назначение в одну из штаб-квартир при военном ведомстве, я не мог себе представить, что чуть ли не в первый же день мне будет дано распоряжение ознакомиться с работой Транспортного отдела, который занимался подсчетом грузооборота транспортных судов, предназначенных для перевозки войск и груза, и планированием ежемесячного объема грузов и численности войск, которые должны быть перебронированы на другой континент в соответствии с планом совместных действий военно-морских и сухопутных сил. Мне не составило большого труда установить, что все вычисления проводились неправильно и, естественно, не имели никакого практического выхода (т. е. грузооборот не увеличивался). Я предложил изменить методы расчета с учетом того, что статистические данные о скорости движения судов и об объеме перевозимого груза не соответствуют данным, используемым в расчетах. Для большей убедительности я привел следующий пример. Предположим, что грузовая флотилия состоит из трех судов, одно из которых тихоходное с грузооборотом в 50 дней, а два других — быстроходные с грузооборотом в 25 дней каждое. Тогда среднее время грузооборота для всех трех судов будет составлять 33,3 дня. Это означает 3 грузооборота для каждого судна за 100 дней. Однако на самом деле каждое быстроходное судно может совершить за 100 дней 4 грузооборота, а тихоходное — только 2, что в сумме составит 10 грузооборотов. Таким образом, в первом случае мы имеем 9 грузооборотов, во втором — 10, причем оба подсчета основаны, казалось бы, на совершенно правильных арифметических вычислениях.

На следующий день ко мне зашел дежурный офицер и сообщил, что начальник штаба одобряет мои действия и просит меня к нему зайти. Нет необходимости говорить о том, что означает такая просьба. Бросив все дела, я явился в штаб-квартиру.

В комнате для секретных совещаний, куда я был введен, помимо начальника штаба и его помощников находился начальник военно-морских операций со своей свитой. Между мной и генералом Маричем состоялся следующий диалог:

Генерал. Нам доложили, что вчера, просматривая в Транспортном отделе данные о грузообороте, вы высказались за необходимость отказаться от существующего способа вычислений и предложили заняться, как вы изволили выразиться, определением «эффективного среднего». Не могли бы вы уточнить, что следует понимать под «эффективным средним».

Я. Сэр, согласно математической терминологии, эффективное среднее означает «наилучшее усредненное значение».

Генерал. Каким образом вы предполагаете использовать этот показатель в данном случае?

Я. В данном случае этот показатель может быть использован при определении уровня загрузки и скорости движения судов.

Генерал. Дайте, пожалуйста, определение наилучшего усредненного значения.

Я. Сэр, наилучшее усредненное значение для ряда заданных переменных есть величина, обратная среднему значению обратных величин этих переменных.

Генерал. Совершенно правильно. Я считаю, что вычисления действительно следует проводить по методу, который был предложен вами.

1.2. Разработка методов исследования операций

Вскоре после того, как методы исследований операций стали применяться военными специалистами Великобритании, они были взяты «на вооружение» и военными специалистами других стран. В США и Канаде были созданы специальные подразделения, занимающиеся анализом военных операций. Такое быстрое распространение методов исследований операций в военной области объяснялось необходимостью освоить новые виды вооружения. Новые виды и системы оружия были настолько необычными по характеру, конструкции и технологии производства, что при их внедрении нельзя было полагаться только на опыт, накопленный военными специалистами, требовались научные знания и опыт ученых. Участие ученых в анализе военных операций сразу дало положительные результаты, которые, кстати, свидетельствовали об эффективности использования этих методов. В качестве классического примера успешного применения методов исследования операций можно привести создание глубоководной бомбы: возможность создания такой бомбы была установлена на основе научного анализа данных о ее эффективности с использованием методов теории вероятностей и статистики.

Во время 2-й мировой войны методы исследования операций применялись в основном в области тактики. В частности, рассматривались следующие вопросы: что применять в первую очередь в качестве радиолокационных помех — пассивные или активные, как определить наиболее эффективные цели для бомбометания, какие из способов обнаружения подводных лодок являются лучшими, какой способ расстановки эсминцев для конвоя кораблей является наиболее целесообразным. Исследования в основном касались изучения наиболее эффективного использования существующих технических средств с известными характеристиками. Таким образом, для этих исследований, которые существенно отличаются от более поздних, были известны исходные данные и определены конечные цели при низком уровне неопределенности.

После войны центр тяжести исследований сразу же переместился с тактических задач на проблемы планирования, т. е. возникла необходимость создания новых видов оружия. Постепенно сфера применения методов исследования операций расширялась и уже охватывала не только проблемы выбора оружия и его содержания в мирное время, но также и военно-политические доктрины государства.

Например, одним из первых приложений методов исследования операций было определение оптимальной процедуры поиска эсминцем местоположения вражеской подводной лодки, угрожающей конвою. В настоящее время при использовании методов исследования операций для решения аналогичной задачи рассматрива-

ется не только угроза, которую представляют вражеские подводные лодки для кораблей, но также учитывается возможная угроза городам и военным базам от баллистических ракет, находящихся на таких лодках. Необходимо найти способ борьбы с этими баллистическими ракетами в виде новых комплексов по обнаружению и перехвату, некоторые из которых, может быть, еще находятся в стадии производства или даже в стадии разработки. Кроме того, следует разработать подходящие тактические приемы использования этих комплексов. Короче говоря, проблемы, связанные с выбором новых видов вооружения и разработкой стратегии будущих войн, требуют использования методов исследования операций в масштабах, значительно отличающихся от тех, которые были характерны для 2-й мировой войны.

Методы исследования операций, которые применялись при составлении программ оборонных комплексов, обычно называют методами системного анализа. Основная особенность этих методов, являющихся достаточно общими методами исследования операций, состоит в том, что они могут быть доступными и для специалистов по исследованию операций. Обычно методы системного анализа используются при рассмотрении проблемы обороны создания соответствующих средств противодействия.

Термин «системный анализ» был введен для характеристики нового метода исследований, проводимых в военной области после 2-й мировой войны, целью которых являлся анализ проблем выбора и оценки эффективности новых систем вооружений, и в частности новых систем вооружения бомбардировщиков. Естественно, что в процессе разработки новых систем и видов вооружения, для осуществления которой требовалось несколько лет, могли изменяться как цели, так и исходные характеристики средств вооружения. Поэтому анализ проводился в условиях достаточно высокой степени неопределенности. Поскольку в мирное время невозможно было с приемлемой достоверностью определить эффективность того или иного вида оружия, основное внимание было сосредоточено на производственных затратах. В связи с этим более важную роль стали играть соображения экономического характера. Появился особый вид анализа, использующий критерий «стоимость — эффективность». Типичным примером применения анализа указанного типа является выбор наименее дорогостоящей программы достижения заданной цели. С помощью такого анализа можно из пяти проектов военного истребителя выбрать наиболее эффективный проект при ограниченном бюджете. Однако данная процедура не могла бы быть использована для решения более сложной проблемы: целесообразно ли, вообще, разрабатывать новой истребитель. Кстати, проблемы такого типа не возникали до тех пор, пока не прекратился рост военных бюджетов на авиацию.

Хотя уяснение различия между системным анализом и исследованием операций может быть полезным [78], однако с точки зрения методологии применения исследования операций это не играет существенной роли. Наблюдается тенденция даже все виды «совещательного» (экспертного) анализа относить к исследованию операций.

1.3. Возможные области применения методов исследования операций

Общие методы исследований операций и методы, относящиеся к таким его подразделам, как системный анализ, научные методы управления и экономическая эффективность, могут быть применены для решения многих проблем в военной области, начиная с организации ежедневной административно-хозяйственной работы и кончая ведением военных действий или выработкой определяющих решений относительно обеспечения государственной безопасности. Учитывая специфику этих проблем, их можно разбить на следующие четыре класса:

I. Методы управления людьми, вооружением и техникой в мирное время, обеспечивающие достаточно высокий уровень их экономичности и боеготовности.

Примеры. Какие методы управления следует использовать главному центру снабжения [88]? Каким образом можно предусмотреть необходимую численность военного персонала с учетом различных военных специалистов [102]?

II. Методы использования имеющихся в распоряжении людей, оружия, техники, обеспечивающие выполнение поставленной военной задачи.

Примеры. Сколько видов оборонительного вооружения должно приходиться на один бомбардировщик, чтобы обеспечить прорыв системы защиты «воздух — воздух» [34]? Какое распределение ракетного оружия по целям является оптимальным при существующем запасе ракет и заданных целях [64]?

III. Определение видов оружия, техники, контингента войск, материалов, которые необходимо разработать, создать, приобрести и развернуть в рамках общей оборонительной стратегии.

Примеры. Сколько и какого типа требуется стратегического ракетного вооружения [66]? Каков наилучший план разработки военной техники, например самолетов или ракет [55]?

IV. Разработка оборонительной стратегии США

Примеры. Какой стратегии следует придерживаться США, исходя из имеющихся разведывательных данных и прогнозов [11]? Какую политику следует проводить США во время международного кризиса [1]?

1.4. Характеристика проблем

Приведенные выше проблемы расположены ориентировочно в порядке возрастания их важности при принятии соответствующего решения и одновременно в порядке уменьшения возможностей проведения системного анализа для выработки правильных рекомендаций. Естественно, что не существует четких разграничений между этими проблемами, и, как правило, решение одной из них должно быть допустимым для решения проблем «следующего» класса.

Применение методов исследования операций для решения проблем, связанных с поддержанием надлежащего уровня вооруженных сил и выработки стратегии обороны, предполагает рассмотрение не только инженерных, экономических, научных и управленческих аспектов соответствующих проблем, но также политических и государственных соглашений. По этой причине фактически не существует в строгом смысле решений подобных проблем, и в таких случаях может быть выработана лишь некоторая рекомендация. Как и следует ожидать, чем меньше данных для анализа, тем он менее точен и, чем меньше он опирается на эксперимент, тем менее достоверными получаются результаты, какими бы совершенными методами ни проводился анализ.

В решении проблем первого класса используются наиболее формализованные методы, являющиеся в этом смысле наиболее продуктивными. Если отвлечься от существа проблем, то соответствующие методы почти не отличаются от методов исследования операций, применяемых при принятии решений и распределении ресурсов в области торговли и промышленности (имеются в виду проблемы управления запасами, назначения персонала, проверки надежности, составления маршрутов и т. д.). Все эти методы исследования операций, или, более конкретно, научные методы управления, предназначены для того, чтобы повысить эффективность человеко-машинных процедур управления, где под «более эффективным» подразумевается некоторый «военный эквивалент» понятия «максимально выгодный». Характерная особенность проблем первого класса состоит в том, что применяемые для их решения методы исследования операций представляют собой систематизированные вычислительные процедуры, которые «привязаны» к достаточно общей модели. Изменяя параметры модели, можно использовать ее для решения широкого круга проблем.

Следует отметить, что экономический эффект от использования методов исследования операций в решении задач управления военными операциями в мирное время огромен. Например, методы, разработанные исследователями систем для решения проблем управления запасами только для военно-воздушных сил США, дали ежегодную экономию в размере 250 млн. долл.

Характер проблем остальных классов всегда предполагает наличие конфликтной ситуации. Это обстоятельство требует включения в модель элемента конфликтности. Вероятно, главную роль в чисто военных вопросах играет действительное или потенциально возможное взаимодействие с врагом, а не взаимосвязь между различными решениями и затратами. Поэтому проведение обычного анализа в области торговли или промышленности проще, так как в данных случаях конфликтная ситуация не играет значительной роли. По этому поводу Альберт Вуолстетер писал [103]:

«Компания Bell Telephone System в основном беспокоится о том, чтобы в телефонные автоматы не попадали жетоны и фальшивые монеты. В то же время компании совершенно не надо опасаться того, что кто-то будет создавать помехи на высокочастотных релейных линиях с единственной целью — сорвать нормальные для мирного времени доходы компании.»

Однако нет таких методов исследования операций, которые применялись бы исключительно в военной области. Даже классическая военная игра с участием трех военных подразделений нашла применение в гражданской сфере, хотя в основном для тренировки и обучения.

Парадокс использования методов исследования операций в военной области заключается в том, что из всех изучаемых этими методами боевых операций осуществляются только некоторые, остальные же, на наше счастье, возможно, никогда не придется осуществлять. Очень часто какая-нибудь изучаемая система оружия или боевая операция оказывается устаревшей, даже не пройдя испытаний в бою, и мы так и остаемся в неведении относительно правильности проведенного анализа. Эффективность тех видов оружия, которые в мирное время используются в качестве средства обороны, определяется тем, как они повели бы себя во время войны. Таким образом, основная трудность в применении методов исследования операций в военной области связана с получением значительной по объему количественной информации, на основе которой можно строить прогнозы. Даже во время военных действий существует мало возможностей для получения такой экспериментальной информации. Поэтому ее источниками в основном являются учения, полевые испытания (которые зачастую организуются так, что не дают ощутимых результатов), учения по карте и боевой опыт (который вряд ли окажется полезным в будущем, когда военные действия будут вестись совершенно другими видами оружия).

Прежде чем планировать приобретение будущих систем вооружения, необходимо теоретически оценить эффективность этих систем в различных боевых ситуациях. Такая оценка, конечно, основывается на имеющихся данных о различных частях системы. Совершенно ясно, что в мирное время специалисты по исследованию

операций занимаются не столько статистическим анализом имеющихся данных, сколько теоретическими подсчетами на основании весьма неполных данных. Это привело к появлению таких новых методов анализа, как стратегические (операционные) игры и имитационное моделирование, а также к дальнейшему развитию методов планирования в условиях неопределенности, в частности методов проверки чувствительности соответствующих моделей к изменению ее параметров.

1.5. Организация исследований

После 2-й мировой войны одни страны продолжали организацию учреждений, занимающихся исследованиями военных операций, а другие только приступали к их созданию. В большинстве стран, за исключением Соединенных Штатов, такие учреждения, как правило, относились к различным родам войск и руководство ими осуществлялось в основном военными специалистами. В Соединенных Штатах каждый род войск и министерство обороны были связаны с более или менее независимыми организациями, ведущими соответствующие исследования и укомплектованными главным образом гражданскими лицами. Наиболее известной из таких организаций является фирма Rand Corporation [83]. Другими известными организациями, проводящими исследования в военной области, являются Center for Naval Analysis, Institute of Defense Analysis, Stanford Research Institute, Research Analysis Corporation (сейчас входит в состав General Analysis Corporation). Кроме того, исследованием операций в военной области занимаются почти все предприятия авиационной промышленности, а также многие другие промышленные фирмы.

Фактически в Соединенных Штатах до конца 60-х годов работы в области исследования операций велись в основном благодаря финансовой поддержке министерства обороны. Однако затем конгресс обязал министерство обороны сократить расходы на фундаментальные исследования. Тем не менее многие разделы исследования операций — теория игр, линейное программирование, динамическое программирование, теория потоков в сетях, имитационное моделирование и другие — были разработаны при финансовой поддержке таких организаций, как Office of Naval Research, Army Operations Research Office, Air Force Office of Scientific Research, Air Force Project Rand. В итоге на применение методов исследования операций в военной области были затрачены огромные средства. Согласно данным Главного статистического управления США [96], затраты на разработку всех «активных» моделей, имитационных моделей и игр для министерства обороны составили 175,5 млн. долл. Причем это, конечно, лишь часть действительных расходов.

1.6. Пример использования методов исследования операций в военной области

Чтобы проиллюстрировать некоторые детали анализа операций в военной области, рассмотрим гипотетический случай, который, однако, вполне мог бы иметь место и в действительности. Речь пойдет об анализе варианта проблемы, вопрос о решении которой был поставлен перед конгрессом США: следует ли для стратегических ракет использовать систему противоракетной защиты.

Стратегические ракеты наземного базирования являются важной составной частью оборонного потенциала США и находятся в состоянии готовности нанести ответный удар в случае атаки противника. Для реализации такой готовности необходимо, чтобы после атаки противника уцелело определенное число боеспособных стратегических ракет.

Существуют различные способы повышения неуязвимости ракет. Можно замаскировать ракеты, сделать их мобильными, создать для них прикрытие, произвести такое количество ракет, что ни один противник не сможет уничтожить сразу все. Наконец, можно организовать активную защиту ракет.

Предположим, что система противоракетной активной защиты возможна и даже частично разработана. Проанализируем, какие методы исследования операций могут оказаться наиболее эффективными при принятии решения в данной ситуации.

Следует иметь в виду, что без проведения анализа необходимость приобретения данной системы защиты не вызывает сомнения только в том случае, когда предлагаемая система защиты дешева и эффективна или существует реальная опасность нападения. Однако ответы на вопросы относительно стоимости системы и ее эффективности, реальности опасности, типа необходимой защиты зависят от многих показателей, взаимозависимых факторов и параметров, некоторые из которых известны, другие могут быть определены, а относительно третьих приходится делать соответствующие предположения. Таким образом, для того чтобы принять рациональное решение, необходимо иметь схему, которая увязала бы в единое целое различные аспекты проблемы. Методология исследования операций как раз и предлагает такую схему, которая сводится к следующему.

Определение целей и задач рассматриваемой системы защиты. Для того чтобы определить цели и задачи системы защиты стратегических ракет, рассматриваемой с точки зрения дальнейшей разработки и приобретения, необходим анализ более крупной системы, включающей как систему защиты, так и защищаемые ракеты и все то, что имеет значение для приведения системы защиты в действие: информация о возможности атаки про-

тивника, РЛС дальнего обнаружения, политическая обстановка и т. п. Дать оценку системе защиты можно, лишь выяснив насколько она, расходуя определенный ресурс, способствует или препятствует повышению эффективности защищаемых ракетных сил. Таким образом, когда речь идет о целях системы, то необходимо рассматривать общие цели, стоящие как перед стратегической системой, так и перед системой ее защиты. Решение вопроса о том, следует ли планировать защиту ракет, зависит от того, каким образом они будут использоваться. Если они предназначены для нанесения неожиданного удара, то нет смысла их защищать; если стратегические ракеты и система их защиты должны быть чем-то вроде средства сдерживания атаки противника, то такая защита необходима. Мы говорим при этом «чем-то вроде», так как нам бы хотелось, чтобы стратегические ракеты могли служить не только средством сдерживания атак противника.

Выбор способа оценки «сдерживающего эффекта». Допустим, что цель ракетных сил — сдерживание. Поскольку различные системы обеспечивают разный «сдерживающий эффект», необходимо найти некоторый способ оценки этого эффекта, т. е. выявить, как использование той или иной системы защиты влияет на эффективность всей системы как средства обороны.

Наиболее разумным было бы измерить «сдерживающий эффект» количеством стратегических ракет, уцелевших после атаки противника и готовых к запуску. Такая оценка хотя не учитывает всех аспектов обороны (сдерживания), тем не менее имеет правильную направленность. Разработка оценок такого типа имеет решающее значение во многих исследованиях. Если бы удалось найти подходящую меру эффективности, то, очевидно, результаты анализа не вызвали бы сомнения. (В противном случае результаты анализа всегда будут сомнительными.) Однако нет полной уверенности в том, что такую меру удастся найти. Дело в том, что сдерживание — это действие, которое диктуется разумными соображениями. Если бы имелись две системы, в одной из которых после первой атаки противника уцелело 500 стратегических ракет, а в другой только 250, то трудно определить, насколько большим «сдерживающим эффектом» обладает система с вдвое большим числом уцелевших ракет.

Определение факторов, которые могут быть причиной начала военных действий. Рассмотрим данную систему защиты с учетом политической и военной обстановки, в которой должна действовать эта система. В этой связи особое значение приобретают факторы, которые могут явиться причиной начала военных действий, например разрастание кризиса, преднамеренная эскалация военных действий или, что наиболее часто допускается, внезапное нападение противника. Другими словами, необходимо предусматривать все

возможные направления событий и их характер, так как в разных ситуациях требуются различные состояния боевой готовности ракетных сил и различные расходы на поддержание их в этом состоянии. Следовательно, определение этих факторов может оказать значительное влияние на сравнительный анализ.

Определение возможных вариантов систем. Для этого необходимо изучить степень реализуемости различных вариантов и сократить их число до разумной величины, оставив только те варианты, которые отличаются друг от друга по наиболее важным параметрам. Таких вариантов должно быть по крайней мере два. Наиболее очевидный вариант состоит в том, чтобы использовать имеющиеся фонды не на систему защиты, а на покупку дополнительных ракет.

Уточнение постановки задачи. Поскольку трудно определить, какое число уцелевших ракет понадобится для обеспечения сдерживания, но известны ассигнования на приобретение системы защиты, то представляется разумным в постановке задачи зафиксировать сумму выделенных средств и при этом ограничении максимизировать эффективность системы защиты. Можно сформулировать критерий, или правило, выбора системы следующим образом: приемлемым считается такой вариант системы защиты, который при ограниченной стоимости обеспечивает максимальное число непораженных ракет.

Составление схемы (или модели) определения расходов. Чтобы такая схема была полезной, она должна учитывать изменения в методике применения выбранной системы, происходящие в период ее разработки. В первом приближении подсчитываемые расходы включают в себя сумму (выражаемую в долларах или его эквивалентах), затрачиваемую на поддержание имеющейся в настоящее время системы ракет, плюс ассигнования на ее усиление либо с помощью какой-нибудь системы защиты, либо за счет разработки дополнительных ракет. Далее, следует определить, какое вооружение можно приобрести на данную сумму ассигнований. Эта задача оказывается довольно сложной, так как необходимо учитывать жесткие ограничения на ресурсы, а также нематериальные затраты. Тут возможны нежелательные побочные эффекты, которые могут помешать осуществлению программы и поэтому должны быть изучены (нежелательные эффекты могут быть связаны, например, с неудачным выбором местоположения пусковых установок).

Разработка модели оценки возможных вариантов системы защиты. Другими словами, требуется разработать вычислительную схему или вычислительный процесс, например систему уравнений или машинную программу, которые бы для каждого варианта рассчитывали число непораженных ракет. Эти предположения должны учитывать объемы выделенных средств,

характеристики ракет противника (точность, надежность, разрушительную силу, число и т. п.); характеристики системы защиты (надежность, вероятность поражения атакующей ракетой и т. п.), а также характеристики ракетной системы, подвергшейся атаке (прочность, местонахождение, число ракет и т. п.)

Рассмотрим простейшую схему расчета затрат. Введем следующие обозначения: M — число ракет, которые можно приобрести при заданном бюджете, не тратя средств на систему защиты; s_M — стоимость одной незащищенной ракеты; s_D — стоимость защитных средств, приходящихся на одну ракету; Ms_M — заданная сумма ассигнований, а $(s_M + s_D)$ — общая стоимость ракеты и средств ее защиты.

Описываемая здесь модель определения эффективности [81] построена по аналогии с моделью, представленной в работе [16] и предназначенной для описания системы защиты, жестко связанной с пусковыми установками.

В любой модели определения эффективности используются следующие показатели: A — число ударов (выстрелов), нанесенных противником по ракете; p — вероятность уничтожения одним выстрелом незащищенной ракеты; p_D — вероятность уничтожения одним выстрелом защищенной ракеты.

Сначала рассмотрим случай атаки на незащищенные ракеты. Если A — число ударов по ракетам, а M — число ракет (мишеней), то A/M — число ударов по одной мишени. Для того чтобы упростить математические выкладки, допустим, что в данном случае A/M — целое число. Вероятность того, что ракета уцелеет после нанесения ударов, равна $(1 - p)^{A/M}$. Тогда ожидаемое число уцелевших ракет составит $M(1 - p)^{A/M}$.

Приведенная модель не позволяет непосредственно получить численные результаты, поскольку для определения некоторых ее параметров требуются отдельные частные модели. Например, для того чтобы определить вероятность p того, что один удар поразит незащищенную цель, необходимо сделать некоторые предположения относительно характера цели, о достижимой точности и характере атакующего удара.

Теперь рассмотрим защищаемую систему. Поскольку известна общая сумма ассигнований Ms_M и общая стоимость ракеты и ее защиты $s_M + s_D$, то число защищенных ракет, которые могут быть приобретены на данную сумму, составит $Ms_M/(s_M + s_D)$. В условиях установки системы защиты необходимо предполагать, что при том же числе атакующих ударов A вероятность поражения цели уменьшается до величины p_D . Число атакующих ракет или ударов на одну защищенную ракету составит

$$A \approx \frac{Ms_M}{s_M + s_D} = \frac{A(s_M + s_D)}{Ms_M}.$$

Сделаем упрощающее предположение (которое относится лишь к данному случаю, а не к анализу вообще) о целочисленности величины, определяемой отношением числа атакующих ударов числу ракет. Если p_D — вероятность уничтожения противником защищенной ракеты, то вероятность того, что ракета уцелеет под ударами всех атакующих ракет, определяется по формуле

$$(1 - p_D)^A (s_M + s_D)^{M s_M}.$$

Следовательно, ожидаемое число защищенных ракет, уцелевших после атаки, равняется указанной вероятности, умноженной на исходное число этих ракет.

Если учесть сформулированный выше критерий оценки возможных вариантов системы защиты и сравнить ожидаемые значения соответствующих величин, то получим следующее условие, при выполнении которого приобретение системы защиты становится целесообразным:

$$\frac{M s_M}{s_M + s_D} (1 - p_D)^A (s_M + s_D)^{M s_M} > M (1 - p)^{A/M}.$$

При наличии нескольких вариантов системы защиты следует выбрать тот, для которого левая часть этого неравенства будет наибольшей. По ряду причин также может возникнуть необходимость в исследовании полного распределения вероятностей числа пораженных ракет, однако здесь это не будет рассмотрено. Следует иметь в виду, что не были учтены такие важные факторы, как возможность обеспечения защиты только части ракетного арсенала, пути уменьшения стоимости системы защиты (например, путем группировки ракет или с помощью маскировки). Конечно, не требуется слишком тщательно маскировать ракеты (т. е. если противник не будет знать, что у нас есть наступательные ракеты, нам не будет нужна система противоракетной обороны) или располагать их слишком близко друг к другу, так что один удар смог бы уничтожить не одну ракету. Не принимались во внимание и другие возможные действия противника, например проведение рекогносцировки между атаками, с тем чтобы избежать ударов по уже уничтоженным целям и не производить избыточных ударов, или нанесение удара в первую очередь по самой системе защиты.

Все эти факторы, а также наличие предупреждения об атаке, установка ложных стартовых площадок, создание электронных помех могут быть учтены при разработке более сложных моделей и проведении более дорогостоящих исследований.

Далее необходимо проследить за тем, чтобы все неопределенности были тщательно изучены. Следует выяснить, оказывает ли принципиально неустранимая неопределенность относительно будущей обстановки в мире или развития техники существенное влия-

ние на принимаемые решения. Например, необходимо выяснить роль степени риска или временной задержки осуществления разработок. Следует также провести анализ на чувствительность, чтобы обнаружить, как отсутствие определенности в значениях различных параметров влияет на результаты. В частности, хорошо было бы выявить зависимость вероятности уничтожения ракет от точности удара и прочности пусковых ракетных установок. Для этого пришлось бы анализировать более специализированную модель. В целом необходимо иметь возможность варьировать один за другим или одновременно несколько основных параметров, используя для задания области изменения параметров, например метод Монте-Карло.

Рассмотрим факторы, которые до сих пор не принимались во внимание. К ним можно отнести неясные факторы военного и политического характера, связанные, например, с тем, что в арсенале имеется большее число ракет, чем число ракет, уцелевших после возможной атаки, или с наличием некоторой системы защиты, пусть неэффективной, но способной поднять моральный дух населения. Другим важным следствием наличия системы защиты является ее влияние на переговоры о сокращении вооружений.

Теперь можно было бы на основе анализа дать некоторые рекомендации. Оказывается, что это не так-то просто сделать, так как выбор решения всегда зависит от большего количества факторов, чем включено в анализ. Так, например, мы должны учитывать распределение доходов и расходов между заинтересованными группами лиц, на которые повлияет решение, и также чистую прибыль этих лиц. Для установки новой системы защиты может потребоваться земля, находящаяся в распоряжении определенной группы лиц. Кроме того, люди, живущие по соседству, могут возражать против установки новой системы защиты из опасения оказаться мишенью. Следовательно, люди должны получить компенсацию не только за дополнительные расходы, но и за моральные издержки, связанные со страхом за свою жизнь. Поскольку решение всех этих вопросов может оказаться не простым делом, политическая жизнеспособность предпринятой акции может оказаться под угрозой.

2. АНАЛИЗ ВОЕННЫХ НЕБОЕВЫХ ОПЕРАЦИЙ

Многие проблемы эксплуатации и управления, с которыми приходится сталкиваться военным специалистам, аналогичны проблемам, существующим в области торговли и промышленности, что позволяет для их решения использовать хорошо известные разработанные методы. Например, для определения оптимального объема запасных частей используются теоретические основы технического обслуживания и теория запасов, для определения опти-

мального фонда на ремонтные работы — теория массового обслуживания и теория управления запасами, для определения максимальной загрузки наземных и воздушных транспортных средств — линейное программирование, при разработке и производстве новых видов оружия и военной техники — метод критического пути. Военные специалисты фактически первыми стали использовать методы исследования операций при анализе многих общих функциональных процессов, имеющих место как в военной, так и гражданской сфере.

Отличие исследования небоевых военных операций от гражданских состоит в необходимости учитывать вероятность неожиданного возникновения конфликтной ситуации, которая может потребовать принятия самых различных решений — от участия в «демонстрации силы» до непосредственных военных действий. Степень влияния возможного возникновения конфликтной ситуации на анализ зависит от характера анализируемой деятельности. Так, например, наличие такой возможности будет иметь решающее значение при определении технических характеристик и количества транспортных самолетов и играть менее существенную роль при составлении расписания транспортировки домашнего имущества и родственников военнослужащих, находящихся за границей.

2.1. Организация перевозок и снабжения

В военном деле организация перевозок и снабжения состоит в планировании и осуществлении операций, связанных с разработкой, приобретением, хранением, перевозкой и сохранностью груза (материальной части и имущества), в обеспечении всех видов перевозок и в организации их функционирования. Данная область благоприятна для разработки и применения методов исследования операций, что в свою очередь способствовало созданию единой системы перевозок и снабжения... В рамках этой системы удастся координировать функционирование таких ее элементов, как снабжение, транспортировка, резервы, эксплуатация и контроль информации, обеспечивая тем самым высокий уровень боеготовности техники при небольших расходах на ее содержание [2, 10]. Следует отметить, что применение в военной области методов прогнозирования и теории управления запасами дало большой экономический эффект. В этой связи заслуживает внимания работа [52], посвященная выработке методики принятия решений относительно того, в каких случаях необходимо уменьшить или увеличить нормальный уровень запасов.

В работе [25] рассматривается стратегия обслуживания технических средств, которые ввиду их дефицитности и длительности установки редко или вообще не замечаются, но требуют

определенных запасных частей; анализ уровней запаса с помощью статистического метода приводится в работе [77]; модели управления запасами, используемые в ВМС США, рассматриваются в работе [79], а математическое обоснование правил замены частей самолетов и ракет дается в работе [50].

Решение о целесообразности ремонта какого-нибудь оборудования обычно принимается с учетом экономических показателей, таких, как эксплуатационные расходы, ликвидационная стоимость, средний срок службы, устаревание, сроки ремонта и т. п. Полная логическая схема принятия решений по вопросам ремонта или списания оборудования, в которой используются экономически обоснованные ограничения расходов на ремонт, представлена в работах [5, 33, 60].

Для исследования проблем организации перевозок и снабжения в военной области широко применяются имитационные модели, ориентированные на расчет вручную и с помощью ЭВМ. Одной из совершенных моделей этого типа является модель SAMSON (Support-Availability Multi-System Operations Model). Это машинная имитационная модель, предназначенная для изучения влияния колебаний уровня запасов и изменений в политике на боеспособность авиации, т. е. изучения зависимости между уровнем организации перевозок и снабжения и возможностью проведения боевой операции. В модели учитываются характеристики надежности самолета и обеспечение полета различными его подсистемами, планы и сроки проведения операций и осуществления перевозок и снабжения, а также проблемы обеспечения ресурсами и средствами противовоздушной обороны, необходимыми для реализации этих планов и сроков.

Имитационные модели также использовались для разработки и оценки возможных стратегий осуществления перевозок и снабжения в системе межконтинентальных баллистических ракет, построенной на новых принципах организации.

2.2. Авиационный и морской военный транспорт

Одним из первых опытов применения методов решения транспортной задачи было составление расписания для флотилии военных танкеров [37]. Что касается авиации, то при определении технических требований к оперативно-тактической авиации важно учитывать пригодность посадочной площади (или летного поля) для приема самолетов. Значение степени пригодности военного аэродрома с точки зрения требований к конструкции самолета изучалось в работе [51].

Аналитический подход к исследованию авиационного военного транспорта включал использование линейного программирования. В работе [62] приводится пример, в котором первоначально про-

стой анализ становится более сложным и достоверным по мере более глубокого изучения проблемы, т. е. от использования модели линейного программирования переходят к использованию динамической модели, а затем к комбинации динамической и имитационной моделей. В работе [91] описывается модель, использующая процедуру Монте-Карло, предназначенную для исследования военно-воздушного транспорта, включая условия перехода от мирных к военным действиям. С точки зрения организации воздушных перевозок определенным интерес представляет работа [41], где рассматриваются технические характеристики самолетов, предназначенных для перевозки груза.

3. АНАЛИЗ БОЕВЫХ ОПЕРАЦИЙ

Первоначальной целью применения методов исследования операций в военной области было изучение эксплуатационных характеристик новой техники или вооружений, с тем чтобы либо улучшить их конструкцию, либо повысить эффективность их использования. Подробное описание методов, которые использовались во время 2-й мировой войны, можно найти в работах [9, 68, 69].

Методы исследования операций, разрабатываемые соответствующими подразделениями военных штабов, в настоящее время в основном ориентированы на достижение вышеуказанной цели. В области тактики результаты анализа боевых качеств оружия и оборудования, проводимого с помощью методов исследования операций, могут быть соотнесены с данными экспериментов и полетных испытаний, полученных в ходе, например, учебных воздушных боев, а иногда и в ходе действительных боевых действий. (При этом иногда приходилось непосредственно использовать разработанные методы.)

Практически ни в одной другой области применения методов исследования операций нет таких благоприятных условий для создания и развития новых методов, как в военной области. Почти каждый аспект боевых операций был объектом анализа, однако различных аспектов настолько много, что мы можем лишь кратко упомянуть некоторые из наиболее важных.

Проблемы, рассматриваемые в данном разделе, являются, как правило, очень сложными и требуют использования достаточно совершенных и сложных методов. Согласно работе [80], можно для решения этих проблем использовать аналитический подход, имитационное моделирование на ЭВМ и методы теории игр. Поскольку прикладные задачи, возникающие в военной области, весьма разнообразны, трудно провести классификацию с учетом их характера. Поэтому последующий материал рассматривается в соответствии с указанными подходами.

3.1. Аналитические модели

В силу абстрактности аналитическая модель может учитывать только часть переменных задачи и, описывая последнюю, зачастую может приводить к упрощению. По этой причине аналитические модели обычно не применяются для непосредственного решения проблем, связанных с выполнением реальных операций. Однако очень часто такие модели, так же, кстати, как и игровые модели, позволяют проникнуть в суть проблемы.

Примеры применения математического моделирования к проблемам бомбардировки, противовоздушной обороны, артиллерийской стрельбы и запуска ракет весьма многочисленны. В соответствующих исследованиях во внимание принимаются два или более показателя (фактора), влияющие на выбор тактики (например, геометрия мишени, особенности цели, количество бомб или снарядов, вид стрельбы, ошибки наводки, характер повреждений, уязвимость мишени и т. п.). Примеры применения методов исследования операций к решению проблем указанного типа в период 2-й мировой войны можно найти в работах [13, 40, 56, 57].

С появлением ядерного оружия, обычных и баллистических ракет, акцент в рамках рассматриваемых проблем несколько сместился, однако аналитические методы исследований по-прежнему находят применение [28]. В некоторых случаях эти методы совпадают с ранее разработанными. Работы советских ученых по данной проблематике можно найти в обзоре Киприлова [54].

Многие военные проблемы связаны с пространственным и временным размещением вооружений для ведения боевых действий. Другие проблемы относятся к дуэльным сражениям (например, между 2 танками, 2 самолетами или танком и самолетом [27]), а также к преследованию и поиску (например, поиск самолетом подлодки [49]). Следует отметить, что ко многим таким проблемам применяется теория игр, что не только способствует лучшему их пониманию, но иногда помогает выработать практические рекомендации.

Определенные аспекты ведения войны были представлены в игровых математических моделях, в которых подробно рассматриваются двусторонние (а иногда и с большим числом сторон) боевые действия в рамках теории игр с нулевой суммой. Однако реальные боевые ситуации оказываются намного сложнее, чем простые тактические столкновения, и часто не могут быть сведены к игре с нулевой суммой (например, в случае ядерной войны действия не могут сводиться к чистому противодействию двух сторон). Кроме того, некоторые факторы, не включенные в анализ в целях упрощения, например психофизиологические факторы, могут играть решающую роль в понимании происходящего. Тем не менее подобные модели успешно применялись для выработки воен-

ной доктрины [44], для разработки тактики воздушной войны [8 38], боевых действий истребителя против бомбардировщика [19].

В целом применение теории игр в военной области было более полезным для разработки представления о конфликтных ситуациях, чем для практического использования. Клейтон Томас, сотрудник отдела исследования операций ВВС США, по этому поводу писал следующее [92]:

Неверно было бы утверждать, что теория игр совсем не нашла практического применения, так же как утверждать, что она используется слишком интенсивно и даже в ущерб разумному планированию способов и средств обороны. Правильнее было бы сказать, что применение методов теории игр для решения задач тактического плана оказалось более эффективным, чем для разработки главной стратегии.

При организации оборонных мероприятий широко применяются определенные принципы, такие, например, как принцип «ни одного слабого звена» [27], согласно которому оборона объектов должна быть такой, чтобы атакующий самолет встретил одинаковый отпор при нападении на любой из объектов.

3.2. Имитационное моделирование

Имитационное моделирование военных задач по сравнению с математическим моделированием охватывает намного большее число переменных и ставит основной целью максимальное приближение к действительности. Несмотря на высокую стоимость многомерных имитационных моделей, реализованных на ЭВМ, они довольно широко применяются, и буквально сотни таких имитационных моделей находятся в распоряжении министерства обороны. Тем не менее в работе [80] делается следующее заключение относительно их использования:

Есть определенные указания на то, что крупномасштабные модели не находят широкого применения. Большой объем и длительный срок разработки, а также общий характер этих моделей создают трудности для проверки их адекватности. В случае использования крупномасштабных моделей вырастает вероятность нежелательных последствий, смены обслуживающего персонала, плохо составленной документации, слабой концептуальной проработки и недостаточно высокого уровня профессиональных связей.

Крупномасштабные цифровые имитационные модели, как правило, используются для описания боевых операций стратегических бомбардировщиков. Первая попытка проимитировать с помощью быстродействующей ЭВМ большинство основных элементов стратегической воздушной войны была предпринята с помощью модели стратегических операций, разработанной фирмой Rand Corp. В дальнейшем модель была дополнена другими детально разработанными элементами, такими, как диспетчеризация полетов,

проводка самолетов по определенным маршрутам, дозаправка, наличие атакующих истребителей и системы локальной защиты, бомбовые разрушения наступательной техники, ограничения на ход операций в виде наличия радиоактивных осадков. Причем все перечисленные элементы имитировались для конкретных географических условий при наличии определенных вооруженных сил, заданных характеристиках самолетов, известной эффективности защиты и вооружения, а также при ограниченных резервах военных баз.

Эта модель, как и другие подобные имитационные модели, описывает обычные военные действия, которые рассматриваются в виде двустороннего конфликта, когда можно проследить все подробности происходящего. Моделирование начинается с задания начальных условий для каждой стороны: количество и типы самолетов (бомбардировщиков, заправщиков топлива, транспортных), местонахождение баз бомбардировщиков, размещение средств защиты (РЛС, систем локальной защиты, баз истребителей) и состояние всей этой техники. Кроме того, должен быть разработан ряд стратегий для каждой из сторон, предусматривающих не только составление расписания взлетов и контроль за полетами, но и учитывающих непредвиденные обстоятельства, которые могут возникнуть в ходе военных действий, и возможность ведения военных действий в различной обстановке.

Другая имитационная модель [59] позволяет определять степень влияния поддержки ближней авиации, минометов и небольших орудий на исход сражения, в котором участвуют пехотные подразделения численностью не более полка. Эта модель разработана в деталях в предположении, что рассматриваемая ситуация может быть промоделирована непосредственно [87].

В ряде работ рассматриваются сложные взаимодействия военных, технических, геополитических и социально-экономических факторов, с наличием которых связаны основные трудности разработки программ обороны по созданию средств отражения попыток противника нарушить установленные границы. В работе [77] приводятся два варианта соответствующих программ, работающих в реальном масштабе времени и созданных на основе модели пограничного контроля [76]. С помощью этих программ можно анализировать конфликтную ситуацию без проведения вручную математических вычислений. Варианты программ соответствующей модели дают готовый анализ специфических ситуаций и позволяют быстро проверять новые решения с точки зрения их возможной эффективности в различных непредвиденных обстоятельствах. Программы позволяют также провести тщательный количественный анализ предлагаемых систем по обеспечению безопасности границ. С помощью данных программ можно про-

водить детальное исследование и предсказывать развитие во времени партизанских войн в интересующем районе. При этом расчеты могут проводиться для условий инфильтрации, наличия огневого заграждения, наличия подкреплений и истощения сил у воюющей стороны. Для любого будущего момента получаемые с помощью программы данные включают значения некоторых показателей, плотности их распределения и темпы изменения.

3.3. Стратегические игры

Стратегические игры (или операционные, как их обычно называют в гражданской сфере) тесно переплетаются как с теорией игр, так и с имитационным моделированием и чаще всего применяются для исследования конфликтной ситуации с помощью какой-либо аналитической или имитационной модели.

В военных играх в помощь «игрокам» могут быть выделены вычислительные машины, но обычно всего лишь для сбора и обработки информации. Использование методов исследования операций способствовало тому, что традиционная классическая стратегическая игра с участием трех военных подразделений [106], которая первоначально использовалась в основном для тактических учений [47] и анализа запланированных операций, стала современным инструментом исследований. Военные игры хорошо описаны в работе [92], а соответствующие методы — в работах [65, 98].

Примеры игр с фиксированной структурой можно найти в работах [72, 98]. Целью этих исследований было выявление возможных недостатков в боеспособности вооруженных сил, в используемой военной доктрине и в системах оружия при ведении локальных военных действий, а также выработка действенных способов их устранения. Это были открытые игры, включающие большое количество показателей планирования. Следует отметить, что решения в этих играх, как правило, принимались не в соответствии с какими-то установленными определенными правилами, а некоторым лицом, исполняющим роль арбитра. Игры данного типа были использованы для сравнения различных систем оружия, изучения требований, предъявляемых к вооруженным силам, и для сопоставления программ оказания военной помощи [101]. Среди используемых игр были как простые и неформальные игры (например, обнаружение подлодок с воздуха) [67], так и более сложные игры с использованием вычислительных машин, например метод оценки структуры вооруженных сил США [15].

Игры могут представлять собой расширенные учения с участием войск и техники, например учения Sagebrush. Такие игры дают ценные данные. В частности, доклад о результатах операции

Sagebrush содержал не менее 78 рекомендаций, и некоторые из них были использованы министерством обороны и департаментом военно-воздушных сил [26]. Центр разработки боевых операций в Форт Орде, шт. Калифорния, провел военные игры с участием реальных войск и техники с использованием метода Монте-Карло.

При использовании методов теории игр можно как привлечь людей, которые должны будут принять решения в смоделированной (игровой) ситуации, так и использовать цифровые и аналоговые вычислительные машины. Необходимость участия людей вызвана тем, что психофизиологические показатели (оценка ситуации и интуиция) играют большое значение, но не могут быть учтены в модели. Кроме того, иногда участие в игре людей обходится дешевле, чем разработка соответствующего математического обеспечения.

Первые попытки использовать имитационные модели с ЭВМ были предприняты в период 1952—1954 гг. по инициативе Rand Corp. с целью разработки мер по организации эффективной противовоздушной обороны. На основе результатов, полученных при использовании этой модели, был проведен ряд учений личного состава войск противовоздушной обороны США.

Первая разработка игровой модели с использованием аналоговых вычислительных машин являлась частью программы исследований по тактике боевых действий типа «воздух—земля» [12, 82]. В этой игре основное внимание уделяется проблемам распределения средств, стоящим перед командным составом тактической авиации в театре основных военных действий. Эти проблемы обусловлены необходимостью планировать ежедневно число оставляемых в резерве истребителей, находящихся в боевой готовности, истребителей-бомбардировщиков и бомбардировщиков, обеспечивающих противовоздушную оборону аэродромов, а также число бомбовых заграждений и поддержку с воздуха наземных войск. Позднее авторам работы [8] удалось аналитически описать и решить в рамках теории игр несколько упрощенных вариантов данной игры.

В настоящее время игровые модели, реализованные на ЭВМ, в военной области не применяются, за исключением высших учебных военных заведений.

3.4. Методы оценки

Впервые в военной области метод «Дельфи» (или метод прогнозирования на основе оценок экспертов) был использован в 1951 г., когда возникла необходимость определить требования, предъявляемые к бомбардировке промышленных объектов [23]. В последнее время метод «Дельфи» широко используется в военной области, в основном для выполнения технических прогнозов.

Многообразие всегда присутствующих политических аспектов последовательно принимаемых сложных решений, очевидно, требует осмысленного подхода к исследованию (см. [29]). Этот подход в совокупности с игровым подходом позволяет выявлять динамику взаимодействия политических и физических факторов.

4. РАЗЛИЧНЫЕ ТИПЫ АНАЛИЗА СОСТОЯНИЯ ВООРУЖЕНИЯ

Все военные проблемы в какой-то степени являются экономическими проблемами эффективного размещения ресурсов [46].

Как правило, существует множество способов выполнения какой-либо военной задачи. Эти способы связаны с различными стратегией и тактикой, с различной обеспеченностью живой силой и оружием, а также с различными приемами по организации, доставке и поддержанию в боеготовности войск и техники. Традиционные методы исследования операций могут быть использованы для решения проблем рационального выбора. Что же касается проблем распределения ресурсов, то традиционная для раннего этапа исследования постановка задачи в виде указания целей и средств их достижения должна быть расширена с учетом того, что цели и средства являются взаимосвязанными.

Эта взаимосвязь означает следующее: то, что является целью в одном случае, служит средством в другом, возможность осуществления цели зависит от эффективности средства для ее достижения. Эта взаимосвязь означает также, что любая заданная цель представляет собой один из путей достижения более сложной цели [31].

Методы исследования операций, применяемые для оценки состояния вооруженных сил, т. е. для отбора будущих систем оружия и для управления процессом разработки и приобретения этих систем, обычно называют методами системного анализа. Решаемые с помощью этих методов проблемы сложны и многообразны [30, 32, 72, 75, 94].

Использование методов исследования операций на государственном уровне подвергалось серьезной критике, которая, в частности, относилась к двум подходам к анализу проблем организации обороны, один из которых связан с анализом критерия «эффективность—стоимость», а другой опирается на систему «Планирования, программирования и финансирования» (систему PPBS).

4.1. Анализ типа стоимость — эффективность

Критерий стоимость—эффективность используется в военно-экономических исследованиях при сравнении различных путей решения задач государственной безопасности. При этом ищется либо путь, наиболее эффективный при заданных затратах, либо

путь, обеспечивающий достижение заданной цели при наименьших затратах.

Большие затраты на вооружение привели к необходимости применения методов исследования операций при выборе между количеством и эксплуатационными качествами военной техники. В одной из работ обзорного характера [7] отмечается, что поскольку затраты на производство самолетов типа F-14 превышают затраты на производство самолетов типа F-8 и A-4M соответственно в 5,8 и 10 раз, то, согласно закону квадратов Манчестера, самолет F-14 должен быть соответственно в 25, 64 и 100 раз эффективнее самолетов указанных типов, чтобы оправдать приобретение меньшего количества самолетов этого типа при заданном бюджете.

Пример, иллюстрирующий использование анализа типа стоимость — эффективность. Предположим, что требуется создать надежную систему военных средств связи в рамках всего мира. Рассмотрим следующие варианты систем (возможны, конечно, и другие варианты).

А. Усовершенствованная, широкодиапазонная и высокочастотная система, использующая обычные средства связи: кабель, наземные линии, спутники связи и т. п.

Б. Синхронизированная система спутников, предназначенная для военных целей. Система состоит из ряда спутников, вращающихся по экваториальной орбите и имеющих связь с наземными станциями, которые позволяют спутникам «покрывать» всю территорию Земли.

Анализ следовало бы начать с того, чтобы выявить все аспекты проблем, связанных с рассматриваемыми системами, так как если эти системы сравнивать только по пропускной способности для каналов передачи информации, то они практически эквивалентны. Причем ситуация мало меняется, если к пропускной способности добавить любую комбинацию из таких показателей (факторов), как уязвимость в условиях ведения военных действий, избыточность, пригодность. После проведения указанного начального этапа анализа можно переходить к определению общей стоимости систем [36]. Поскольку затраты на изготовление разных элементов возможных систем могут относиться к различным моментам времени, они должны быть дисконтированы к одному и тому же моменту.

К сожалению, некоторые элементы и параметры систем, видимо, вносят достаточно большую неопределенность в общую стоимость системы. В частности, к таким параметрам относятся коэффициенты дисконтирования ввиду возможных дополнительных инвестиций и колебаний в широких пределах эксплуатационных затрат, или общее среднее время отказа системы спутников, или вероятность P успешного запуска ракетопослителей исходных спутников и спутников пополнения.

Предположим, что общее среднее время работы системы до отказа и вероятность P вносят основную неопределенность. Тогда следует рассмотреть интервал возможных изменений в общей стоимости. Такое рассмотрение может быть проведено графически (рис. 1).

Если при фиксированной вероятности P успешного запуска среднее время работы до отказа превышает значение, определяемое точкой пересечения данных кривой и линии стоимости для системы А, то соответствующая кривая для системы В будет располагаться выше. На основе этих данных может быть построена зависимость P от среднего времени работы до отказа, позволяющая определить область предпочтительности систем А и В.

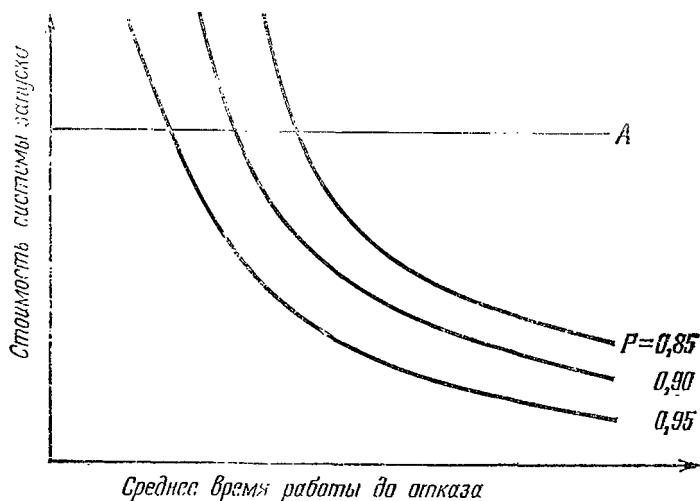


Рис. 1. Эффективность систем А и В при фиксированной вероятности успешного запуска.

Поскольку подобные зависимости наряду со значениями вероятности P и общего среднего времени работы до отказа учитывают также значения других параметров, при проведении анализа необходимо оценивать чувствительность общей «неопределенности» по отношению к этим параметрам. Это особенно важно, например, тогда, когда определение общей стоимости системы проводится на ранней стадии ее разработки [36].

Оценка чувствительности чаще всего проводится путем рассмотрения влияния изменения отдельной переменной в некоторый момент времени. В том случае, когда неопределенность вносится многими переменными, наилучшим методом исследования, видимо, должен быть метод Монте-Карло.

Определение стоимости будущих систем в долларах или других денежных единицах представляет собой важнейшую задачу анализа типа стоимость—эффективность. Хотя методы анализа затрат и использование ресурсов наиболее детально и глубоко разработаны в военной области, использование их не позволяет заметно уменьшить военные расходы, что является главным источником неудовлетворенности военными программами. Более подробное обсуждение рассмотренных методов можно найти в работе [36].

4.2. Управление системами

Решения, касающиеся управления какой-либо системой оружия, должны приниматься на протяжении всех стадий существования этой системы, от идеи и выработки определенного решения относительно приобретения, включая период разработки, производства и внедрения, до окончательного списания системы.

Для решения всех этих проблем довольно часто применяются методы исследования операций. Так, решения относительно приобретения отдельных систем оружия принимаются на основе построения либо графика Ганта, либо веховых диаграмм, а также на основе сбалансированной последовательности операций. Это позволяет повысить эффективность процесса составления расписаний и отчетов. Для анализа сложных итеративных операций [48] используются метод критического пути и метод ПЕРТ (метод оценки и пересмотра плана).

При реализации наиболее эффективного размещения ресурсов, имеющихся в распоряжении министерства обороны, в качестве главного инструмента используется Система планирования, программирования и финансирования, пригодная для управления весьма разнообразными и широко распространенными операциями. При создании этой системы преследовались следующие цели:

— Разработка нескольких вариантов данной программы таким образом, чтобы каждый из них подчеркивал ту или иную цель программы. При этом целесообразность реализации этих вариантов для каждого возможного решения определяется эффективностью затрат в течение длительного периода времени, например на организацию ответного стратегического удара, обороны континента, воздушного или морского моста по сравнению с такими же затратами на содержание войск, строительство или снабжение.

— Осуществление в процессе аналитического сопоставления различных программ, проектной разработки и оценки их стоимости и эффективности, обоснованной их классификации и оценки.

— Предоставление лицу, принимающему решения, информации о ходе выполнения программ и средствах контроля за их выполнением.

Данная характеристика рассматриваемой системы определяет ее значение для анализа и должна способствовать более широкому применению этой системы в министерстве обороны и непосредственно в различных родах войск. Система существенно изменилась и продолжает меняться в основном под воздействием критических замечаний со стороны конгресса и различных организаций, однако роль ее в анализе не снижается.

4.3. Примеры, иллюстрирующие неопределенность результатов анализа, проведенного методами исследования операций

В докладе, составленном Комитетом по национальной безопасности, сообщалось следующее:

Система планирования, программирования и финансирования (PPBS) была использована для обоснования закупки работающего на масле авиа-

носца стоимостью 277 млн. долл., который устарел еще до того, как был снесен на воду. Аналогичные просчеты были допущены при использовании результатов анализа типа стоимость — эффективность при заключении самого крупного в истории контракта на покупку военного самолета, стоимость которого была велика даже для его технических данных...

Подобная ситуация имела место, когда было принято решение относительно приобретения атомного авианосца CVN 67. В ответ на соответствующий запрос министр обороны Меламара дал подробные указания по поводу того, какие виды анализа необходимо провести. Исследования были проведены, однако министр не нашел их достаточно убедительными и потребовал дополнительной информации. Результаты повторных (так же как и последующих) исследований все равно не смогли убедить министра, что улучшение эксплуатационных качеств авианосца за счет использования атомного реактора оправдывает необходимые затраты.

5. МЕТОДЫ АНАЛИЗА ОРГАНИЗАЦИИ ОБОРОНЫ

Методы исследования операций применяются для выработки военно-политических решений на самом высоком уровне [71, 90]. При этом рассматриваются такие проблемы, как размещение средств обороны (сдерживания) на море (для сути эти проблемы не рассматривались), увеличение или сокращение оборонного бюджета, заключенные соглашения по контролю над вооружениями. Эти задачи не могут рассматриваться только как задачи оптимизации, поскольку они связаны с выбором целей, и при их разрешении приходится сталкиваться как с разумными, так и лишними всякого здравого смысла возражениями. Трудность решения проблем подобного рода обусловлена неопределенностью политического, экономического и технического характера. Поэтому наиболее целесообразно для выработки правильных решений, как правило (но, конечно, не всегда), использовать методы, которые не требуют сложных математических вычислений и могут быть доступны для понимания не только специалистам в области исследования операций. Несогласие с результатами анализа возникает главным образом из-за исходных предпосылок относительно важности выявленных аспектов и из-за принятой точки зрения на роль проводимых расчетов (консерватизм или оптимизм).

В связи с этим желательно применять такие методы исследования, как методы полного анализа систем и методы теории военных игр (причем имеются в виду игры, использующие ЭВМ, но не полностью алгоритмизованные), поскольку они позволяют

определить чувствительность возможных решений во всей области изменений параметров, возникающих из-за неопределенностей политического или поведенческого характера.

5.1. Счетный анализ

Значительное влияние на военную стратегию США оказали результаты анализа, проведенного с целью выбора и определения характера использования стратегических военно-воздушных баз. Краткий отчет [104], имевший первоначально гриф «Совершенно секретно», в настоящее время рассекречен.

В отчете рассматриваются главные факторы, влияющие на выбор стратегических баз, и проводится анализ размещения баз бомбардировщиков на период 1956—1961 гг. При этом учитывалось, как отразится принятое решение на общую стоимость стратегических вооружений, в частности как изменяются расходы в связи с увеличением дальности полета бомбардировщиков, которые не могут достичь цели без дополнительной заправки, или в связи с удлинением маршрута полета над вражеской территорией и возможными из-за этого потерями, или, наконец, в связи с необходимостью возмещения убытков, нанесенных противником во время нападения на базу. При анализе принимаются во внимание также политические соображения относительно заокеанских военных баз и возможных осложнений внутренней обстановки.

При получении окончательных результатов сложные математические методы не использовались, тем не менее они применялись для определения степени изношенности бомбардировщиков, оценки повреждений от бомбежки, выбора маршрутов и заправочных пунктов, анализа целей и путей проникновения к объектам, определения уязвимости наземных установок, выработки контрмер с использованием электроники и ловушек и т. д. Этим объясняется весьма разнообразный характер использованных методов—от простого анализа карт до достаточно тонких процедур оптимизации.

Было подсчитано [58], что благодаря проведению этого анализа США сэкономили более миллиарда долларов только на строительстве баз. Однако более важным результатом проведенного анализа явилось изменение характера оборонной политики США. В проведенных исследованиях впервые был поднят вопрос об уязвимости SAC, в то время самой мощной в мире. Было показано, что эта система оказывалась совершенно неэффективной при атаке противника, и в систему были включены новые элементы, необходимые для защиты SAC от угрозы воздушного нападения и удара баллистических ракет, а также было пересмотрено отношение к роли активной противовоздушной

обороны. Таким образом, анализ, целью которого являлось улучшение организации снабжения и перевозок при размещении военных баз за границей, превратился в исследование американской политики стратегического характера. Более подробное освещение этого вопроса можно найти в работе [83].

5.2. Анализ военно-политических учений

Из методов исследования операций, которые используются при решении проблем выработки стратегии обороны, необходимо выделить имитационное моделирование и методы теории игр. История стратегических игр в рамках решения проблем государственной безопасности и соответствующие методы хорошо описаны в работе [42]. Одна из таких игр — игра SAFE (Strategy and Force Evaluation), не предусматривающая использование вычислительной техники, применяется при решении вопросов снабжения, планирования исследований и разработок стратегической разведки, выбора структуры войск и стратегий проведения операций, а также для проверки ранее составленных планов, возможного ведения стратегической воздушной войны [45]. Данная игра была включена в программу обучения на первом курсе Военно-воздушной академии США.

Помимо военно-политических учений, проведенных на уровне Объединенного комитета начальников штабов США (JCS) и состоящих главным образом в проведении игр без использования ЭВМ и игр с нефиксированной структурой и некоторым математическим обеспечением, возможны и более сложные формы проведения учений, в которых широко используется ЭВМ, работающая в реальном масштабе времени. Одним из примеров игры такого типа является игра XRAY [73]. В игре принимают участие 4 военных подразделения: Голубые, Красные, Желтые и Зеленые. Первые три команды представляют соответствующие органы высшего политического уровня трех стран, принимающие решения. Зеленые выполняют функции штаба, по заданию трех других команд занимаются подготовкой результатов специальных исследований и одновременно олицетворяют союзников Красных, Голубых и Желтых, а также тех, кто не присоединился ни к одной из трех команд, и тех, кто сохраняет нейтралитет.

Игра состоит из 2 основных стадий. На первой стадии команды разрабатывают планы общего стратегического потенциала на ближайшие 12 лет, исходя из бюджетных ассигнований на каждый финансовый год, которые растут пропорционально увеличению валового национального продукта, а также выбирают системы оружия из «богатого меню» существующих и предполагаемых систем. Далее, получив разведывательную информацию и разработав первоначальный план действий, команды модернизи-

руют свой арсенал. На второй стадии команды начинают военные действия друг против друга, сдерживаемые Зелеными, которые выступают в роли персонала штабов и, кроме того, должны отражать обстановку внутри этих стран и на международной арене. Наряду с военными действиями в рамках игры могут иметь место принуждение к сделке, предложения о переговорах и ультиматумы.

Для проведения данной игры крайне необходимо использование ЭВМ, работающей в реальном масштабе времени. При этом задача состоит в том, чтобы иметь в распоряжении заранее составленную программу для планирования уровня состояния вооруженных сил, осуществления мероприятий по противоракетной обороне, размещения авиации, определения направления полетов бомбардировщиков, оценки уровня выпадения осадков от наземных взрывов и т. п. При необходимости эти программы и требуемую информацию вводят в ЭВМ и почти моментально получают результаты.

Наиболее важной из этих программ является программа планирования уровня состояния вооруженных сил [36]. В памяти ЭВМ заложена информация о стоимости запрограммированных и будущих систем оружия. Сюда входят данные относительно издержек на исследование и разработку систем оружия, их производство и эксплуатацию. При этом в процессе планирования намечаются сроки внедрения новых и описания устаревших систем. Кроме того, планируются темпы внедрения и оценивается окончательный уровень вооружений. Программа позволяет вносить поправки при изменениях в процессе производства (например, когда устанавливаются более низкие предельные затраты на производство единицы продукции при увеличении ее выпуска), пересматривать цены, если основной элемент системы, такой, например, как самолет, является общим элементом двух или большего числа систем. Общее количество систем вооружений в рамках данной модели может составлять 30 и более единиц.

6. ПУБЛИКАЦИИ, ПОСВЯЩЕННЫЕ ИСПОЛЬЗОВАНИЮ МЕТОДОВ ИССЛЕДОВАНИЯ ОПЕРАЦИЙ В ВОЕННОЙ ОБЛАСТИ

Большинство работ, посвященных применению методов исследования операций в военной области, засекречено. Поэтому в открытой печати можно найти лишь описание устаревших исследований либо тех, что имеют дело с операциями, проводимыми в мирное время, такими, как организация перевозок и снабжения или управление полетами самолетов. Что касается последних, то описание их можно найти в журналах *Naval Research Logistics Quarterly* и *Operations Research*. В этих же изданиях и значительно реже в иностранных публикациях, посвященных использованию

методов исследования операций, можно найти работы на другие темы.

Другими источниками информации являются публикации предприятий оборонной промышленности и научно-исследовательских центров, таких как Institute for Defense Analysis. Наиболее обширными и легкодоступными являются издания Rand Corp., Санта-Моник, шт. Калифорния, и FOA Reports, Стокгольм, Швеция. Исследовательский институт национальной обороны также издаст Proceedings.

Правительственные агентства и неправительственные организации, заключившие контракты с министерством обороны США, получают информацию из Defense Documentation Center. Секретные документы также представляются на симпозиумы, организуемые дважды в год Military Operations Research Society, а затем публикуются в Proceedings. В конечном итоге секретный материал просачивается в открытую печать, хотя и не сразу, однако именно эта отсрочка дает возможность исследователям изменить проведенный ранее анализ.

В качестве иллюстрации можно привести пример из истории создания препятствий наземным транспортным операциям путем бомбежки.

Общепризнанным является тот факт, что тактическая воздушная бомбардировка железных дорог в Нормандии в значительной мере способствовала успешному вторжению в Европу, подробное описание которого можно найти в работе [22]. В частности, авторы работы рассматривают вопрос, вызвавший многочисленные споры: какой вид атаки является более эффективным — изнуряющие налеты тяжелых бомбардировщиков на железнодорожные узлы и сортировочные станции или прицельная атака истребителей-бомбардировщиков на открытые железнодорожные пути и мосты. Как отмечается в работе [22], этот вопрос остался открытым вплоть до окончания кампании. В конце концов, проделав трудоемкую работу по изучению имеющихся данных и взвесив различные факторы, AAF подготовил доклад, в основу которого легли, главным образом, сведения о железных дорогах Франции. В докладе делалось следующее заключение: не было никакой необходимости в атаке на железнодорожные узлы до начала операций во Франции, и 70 000 бомб могли быть сброшены на другие объекты. Таким образом, предпочтение отдавалось заградительной бомбежке с воздуха, а не палетам на железнодорожные узлы.

Какие бы ни велись споры относительно объектов бомбежки, совершенно очевидно, что атаки на железнодорожный транспорт были эффективными для изоляции поля сражения, что тактическая авиация отличилась осуществлением бомбардировки железных дорог и шоссе и что единое планирование этих атак было необходимо для достижения полной изоляции. Тем не менее исследования, проведенные весной 1951 г., показали, что создание препятствий с воздуха транспортным операциям в Корее не осуществлялось под единым командованием и многие высказывались за

бомбардировку транспортных узлов. Это свидетельствует о том, что результаты анализа палетов на железнодорожные узлы Нормандии не были приняты во внимание.

Год спустя был разработан план единого наступления, который, однако, по ряду причин не привел к разгрому противника, хотя ответственный за операцию офицер докладывал о значительных потерях вражеской артиллерии.

Из публикаций, посвященных исследованию военных операций, можно рекомендовать [107], русский перевод которых содержится в книге Чуева [21]. Пять прекрасных, но гипотетических примеров анализа военных систем описано в работе [85], и библиографический материал по применению методов исследования операций в военной области можно найти в работах [61, 80].

Методы исследования операций широко применяются в военной области, и число соответствующих примеров использования растет. В будущем военные операции станут, по-видимому, еще более дорогостоящими и более опасными с точки зрения последствий, чем сегодня. Поэтому методы исследования операций, какими бы несовершенными они ни были, будут приобретать все большее значение для анализа подобных операций.

ЛИТЕРАТУРА

1. Averch H., Lavin M. M., Simulation of Decisionmaking in Crisis: Three Manual Gaming Experiments, The Rand Corporation, Santa Monica, California, RM-4202 PR (August 1964).
2. Axtell George C., Designing an Integrated Logistics System, *Defense Ind. Bull.*, 5, № 7, 30—33, 44 (July 1969).
3. Ayres C. L., The Uses of Statistics in War, AIC 196 (3/15/40), The Army Industrial College, Washington, D. C. (March 4, 1940).
4. Bell C. F., SAMSON: A Logistics Simulation, in E. S. Quade and W. L. Boucher (Eds.), Systems Analysis and Policy Planning: Applications on Defense, American Elsevier, N. Y., 1968.
5. Bell C. F., Kamins M., Determining Economic Quantities of Maintenance Resources: A Minuteman Application, The Rand Corporation, RM-3308-PR (AD 407200), January 1963.
6. Bell C. F., Smith T. C., The Oxnard Base Maintenance Management Improvement Program, The Rand Corporation, RM-3370-PR (AD 292909), November 1962.
7. Bergey K. H., Letter, *Aviation Week*, 122 (April 24, 1972).
8. Berkovitz L. D., Dresher M., A Game Theory Analysis of Tactical Air War, *Operations Res.*, 7, 499—620 (September — October 1959).
9. Blackett P. M. S., Operations Research, *Adv. Sci.*, 5, № 17 (1948).
10. Blanchard B. S., Cost Effectiveness, Systems Effectiveness, Integrated Logistic Support and Maintainability, *IEEE Trans. Aerospace and Electronic Syst.*, AES-3, 186—194 (March 1967).
11. Brodie B., Strategy in the Missile Age, Princeton University Press, Princeton, N.J., 1959.
12. Brom J. R., Narrative Description of an Analytic Theater Air Ground Warfare System, The Rand Corporation, RM-1428-PR, February 1955.

13. Brothers L. A., Operations Analysis in the United States Air Force, *J. of ORSA*, 2, № 1 (February 1954).
14. Browne S. H., Missile and Spacecraft Support Systems, Lockheed Aircraft Company, LMSD-285107, Sunnyvale, California, December 1959.
15. Bruner J. A., Flourney W. N., A Technique for Evaluating the Structure of U.S. Army Forces in an Area Domination Role, RAC-TP-412, Research Analysis Corporation, McLean, Virginia, December 1970.
16. Burke T. F., Cost-Effectiveness of AICBM for the Defense of ICBM Sites, RM-3515-ARPA, Parts I—V, The Rand Corporation (1963—1964). Official Use Only.
17. Campbell H. S., Procurement and Management of Spares, in Enke, 1967.
18. Canby S. L., Military Manpower Procurement: A Policy Analysis. Heath & Company, Lexington, Massachusetts, Washington. D.C., 1972.
19. Caywood T. E., Thomas C. J., Applications of Game Theory in Fighter Versus Bomber Combat, *J. of ORSA*, 3, 402—411 (1955).
20. Chapman R. L., Kennedy J. L., Newell A., Biel W. C., The Systems Research Laboratory's Air-Defense Experiments, *Management Sci.*, April 1959.
21. Чув Ю. В. Исследование операций в военном деле.— М.: Воениздат, 1970.
22. Craven W. F., Cate J. L. (Ed.), The Army Air Forces in World War II, Vol. 3, Europe—Argument to V. E. Day (Jan. 44 to May 45), University of Chicago Press, 1951.
23. Dalkey N. C., Helmer O., An Experimental Application of the Delphi Method to the Use of Experts, *Management Sci.*, 9, № 3, 458—467 (April 1963).
24. Dalkey N. C., Wegner L. H., The Strategic Operations Model: A Summary Report (U), The Rand Corporation, RM-2221, (DDC № AD 304988), July 28, 1958 (Confidential).
25. Deemer R. L., Insurance Items Stockage Policies, U.S. Army Logistics Center, Fort Lee, Virginia, AD-691732, (1969).
26. Deems C. P., War Gaming and Exercises, *Air Univ. Quart. Rev.*, 9, № 1, 98—126 (Winter 1956—57).
27. Dresher M., Games of Strategy: Theory and Applications, Prentice-Hall, Englewood Cliffs, N. J., 1961.
28. Echler R. A., Burr S. A., Mathematical Modes of Target Coverage and Missile Allocation, MORs, Alexandria, Virginia, 1972.
29. Ellis J. W., Jr., Greene T. E., The Contextual Study: A Structure to the Study of Political and Military Aspects of Limited War, *Operations Res.*, 8, 639—651 (September—October 1960).
30. Enke S., Defense Management, Prentice-Hall, N.J. (1967).
31. Enthoven A. C., Operations Research at the National Policy Level, Address before Operations Evaluation Group, Vicennial Conference, Washington, D.C. (May 16, 1962) Also in Tucker (1966).
32. Enthoven A. C., Smith K. W., How Much is Enough: Shaping the Defense Program, 1961—1969, Harper & Row, N.Y. 1971.
33. Feeney G. J., Petersen J. W., Sherbrooke C. C., An Aggregate Base Stockage Policy for Recoverable Space Parts, RM-3644-PR, The Rand Corporation, AD 408943, May 1963.
34. Fawcett C. D., Jones C. D. Effectiveness Determination of Bombers Penetrating Through an Air-to-Air Defense, *Operations Res.*, 18, 516—525 (May—June 1970).
35. Field R. G., Jr., A Basic Approach to Weapon Support Planning, MINUTEMAN Prog. Office, Space Technology Laboratories, Los Angeles, Calif., 1961.
36. Fisher G. H., Cost Considerations in Systems Analysis, American Elsevier, N.Y., 1971.

37. Flood M. M., Applications of Transportation Theory to Scheduling a Military Tanker Fleet, *J. of ORSA*, 2, № 2, 150—162 (May 1954).
38. Fulkerson D. R., Johnson S. M., A Tactical Air Gaim, *Operations Res.*, 5, № 5, 705—712 (October 1957).
39. Gajalo S., NCP Forecasting Techniques, U.S. Army Logistics Management Center, Fort Lee, Virginia, AD-822934, 1967.
40. Germond H. H., Hastings C., Jr., Scatter Bombing of a Circular Target, Applied Mathematics Panel Report, 10.2R, May 1944.
41. Glazer H., Engel J. H., Steinhardt J., The Selection of Cargo Air Transport, in Proceedings of the Second International Conference on Operations Research, English Universities Press Ltd., London, 1961.
42. Goldhamer H., Speier H., Some Observations on Political Gaming, *World Politics*, 12, 71—83 (1959).
43. Hausrath A. H., Utilization of Negro Manpower in the Army, *J. of ORSA*, 2, № 1 (February 1954).
44. Haywood O. G., Jr., Military Decision and Game Theory, *J. of ORSA*, 2, № 4 (November 1954).
45. Helmer O., Bickner R. E., How to Play SAFE: Book of Rules of the Strategy and Force Evaluation Game, The Rand Corporation, RM-2865-PR (ASTIA № 286900), November 1961.
46. Hitch C. J., McKean R. N., The Economics of Defense in the Nuclear Age, Harvard University Press, Cambridge, 1960.
47. Hoffman R., List W., et al., War Games (trans. by P. Luetzkendorf), Robinson F. B. (Ed.), MS. No. P-094. Department of the Army, Office of Military History, Washington, D. C., 1952.
48. Holtz J. N., An Analysis of Major Scheduling Techniques in the Defense Systems Environment, in D. I. Cleland and W. R. King (Eds.), Systems, Organizations, Analysis, Management, McGraw-Hill, N.Y., 1969.
49. Isaacs R., Differential Games; A Mathematical Theory with Applications to Warfare and Pursuit, Control and Optimization, Wiley, N.Y., 1965.
50. Kamins M., McCall J. J., Jr., Rules for the Replacement of Aircraft and Missile Parts, The Rand Corporation, RM-3810-PR, 1961.
51. Kamrass M., Navarro J. A., An Analysis of Tactical Air Systems in Cost-Effectiveness Analysis, T. A. Goldman (Ed.), F. A. Praeger, N.Y., 1967, pp. 116—130.
52. Kaplan A. J., Economic Retention Limits, U.S. Army Logistics Center, Fort Lee, Virginia AD-6917311, 1969.
53. Kennedy L. J., Chapman R. L., The Background and Implications of the Systems Research Laboratory Studies, The Rand Corporation, September 1955, p. 740.
54. Kirillov V. I., Bombometanie, Moscow Translation, AD-287-703, 63-13182 (1960).
55. Klein B., Meckling W., Application of Operations Research to Development Decisions, *Operations Res.*, 6, 3, 352—363 (May—June 1958) (also The Rand Corporation).
56. Kolmogorov A. N. (Ed.), Collection of Articles on the Theory of Firing I, Translation T-14, The Rand Corporation, October 1948.
57. Langner H., Number of Rounds Required and Destruction Probability in the Form of Tables and Graphs, Institute für Theoretische Ballistik in der Luftfahrtforschungsanstalt Herman Goring, 1944, Translation T-11, The Rand Corporation, April, 1948.
58. Life Magazine, p. 101 (May II, 1959).
59. Lind J. R., FAST-VAL: A Model for Forward Air Strike Evaluation, P-3076, The Rand Corporation, March 1965.
60. Logistics Management Institute, Guidelines for Making Repair Expenditure Decisions, Task 68-6 Report, Washington, D. C., 1968.
61. Malik H. J., Bibliography on Military Applications of Operations

- Research, Computation and Analysis Section, Defence Research Board, Ottawa, Canada, March, 1969.
62. Manne A. S., Allocating MATS Equipment with the Aid of Linear Programming, RM-1612, The Rand Corporation, January 1956.
 63. Martino J. R., Technological Forecasting for Decisionmaking, American Elsevier, N.Y., 1972.
 64. Matlin S., A Review of the Literature on the Missile Allocation Problem-*Operations Res.*, 18, № 2, 334-373 (March-April) 1970.
 65. McHugh F. J., Fundamental of War Gaming, 3rd ed., U.S. Naval War College, Newport, R. I., 1966.
 66. Milgram G., A Cost Effectiveness Study for Strategic Airlift, *Transport. Sci.*, 4, № 1, 79-96 (February 1970).
 67. Morse P. M., Trends in Operations Research, *J. of ORSA*, 1, № 4, 159-165 (August 1953).
 68. Morse P. M., Kimball G. E., Methods of Operations Research, Massachusetts Institute of Technology, The Technology Press and Wiley, N.Y., 1951.
 69. National Defense Research Committee, Methods of Operations Research, Summary Technical Report of Division 6, Vol. 2A, 1946.
 70. Novick David (Ed.), Program Budgeting: Program Analysis and the Federal Budget, Harvard University Press, Cambridge, Massachusetts, 1965, Paperback, Holt, Rinehart and Winston (1969).
 71. Operations Research Society of America, Guidelines for the Practice of Operations Research, *Operations Res.*, 19, № 5; 1149-1237 (September 1971).
 72. Paxson E. W., War Gaming, The Rand Corporation, RM-3489-PR, February 1963.
 73. Paxson E. W., Computers and National Security, in Computers and the Problems of Society, American Federation of Information Processing Societies, N.Y., 1972, pp. 65-92.
 74. Quade E. S. (Ed.), Analysis for Military Decisions, North-Holland Publishing Co., Amsterdam, 1964.
 75. Quade E. S., Boucher W. I. (Eds.), Systems Analysis and Policy Planning: Applications in Defense, American Elsevier Publishing Co., N.Y., 1968.
 76. Schilling G. F., Analytic Model for Border Control, The Rand Corporation, RM-6250-ARPA, December 1970.
 77. Schilling G. F., Turner M., On-Line Computer Programs for the Analysis of Border Control Problems, The Rand Corporation, R-687-ARPA, February 1971.
 78. Schlesinger J. R., Quantitative Analysis and National Security, *World Politics*, 15, № 2 (January 1963).
 79. Schrady D. A., Inventory Models, in Zehna (1971).
 80. Shubik M., Brewer G. D., Models, Simulations, and Games - A Survey, The Rand Corporation, R-1060-ARPA/RC, May 1972.
 81. Shubik M., Savage E., The Literature of Gaming, Simulation, and Model Building: Index and Critical Abstract, The Rand Corporation, R-620-ARPA, June 1972.
 82. Siska C. B., Gramboni L. A., Lind J. R., Analytic Formulation of a Theater Air Ground Warfare System, The Rand Corporation, RM-1338, September 1954.
 83. Smith Bruce L. R., The Rand Corporation, Harvard University Press, Cambridge, Massachusetts, 1966.
 84. Smith T. C., SAMSON: Support-Availability Multi-System Operations Model. The Rand Corporation, RM-4077-PR(AD 601813), June 1964.
 85. Snyder W. P. (Lt. Col., USA), Case Studies in Military Systems Analysis, Industrial College of the Armed Forces, Washington, D. C., 1967.
 86. Specht R. D., The Nature of Models, in Quade and Boucher, 1968.

87. Spring S. G., Miller S. H., FAST-VAL: Relationships Among Casualties Suppression, and the Performance of Company-Size Units, The Rand Corporation, RM-6268-PR, March 1970.
88. Stedry A. C., Griswold J., The Development of Supply Control Procedures for the Defense General Supply Center, Proceedings, U.S. Army Operations Research Symposium, Part 1, Army Research Office, Durham, March 1962.
89. Strauch R., A Preliminary Treatment of Mobile SLBM Defense: Game Theoretic Analysis, The Rand Corporation, RM-4439-ARPA, June 1965.
90. Subcommittee on National Security, U. S. Senate, Planning, Programming, Budgeting, Government Printing Office, Washington, 1970.
91. Summerfield J. R., A Model for Evaluating Fleets of Transport Aircraft, in Proceedings of the Second International Conference on Operations Research, English Universities Press Ltd., London, 1961.
92. Thomas C. J., Military Gaming, in Progress in Operations Research, Vol. 1, R. L. Ackoff (Ed.) Wiley, N.Y., 1961.
93. Thomas C. J., Some Past Applications of Game Theory in the United States Air Force, Paper presented at the NTO Conference on the Theory of Games and Its Military Applications, Toulon, France (June 29—July 3, 1964).
94. Tucker S. A., A Modern Design for Defense Decision: A McNamara-Hitch-Enthoven Anthology, Industrial College of the Armed Forces, Washington, D. C., 1966.
95. U. S. Department of Army, Mathematics of Military Action, Operations and Systems, DA PAM 70-5, Headquarters, Washington, D. C., January 1968.
96. U.S. General Accounting Office, Computer Simulations, War Gaming and Contract Studies, A Report to the Committee on Appropriations, House of Representatives, Washington, D.C., February 23, 1971, p. 8.
97. War Department, Report of the Board on Fortifications and Other Defense (H. E. 49-1, Vol. 28, Pt. I). January 23, 1886, pp. 11, 12.
98. Weiner M. G., War Gaming Methodology, The Rand Corporation, RM-2413, July 1959.
99. Whitehead C. T., The Uses and Limitations of Systems Analysis, The Rand Corporation, P-3683, September 1967.
100. Whitmore W. F., Edison and Operations Research, *J. of ORSA*, 1, № 2, 83—84 (February 1953).
101. Whitmore W. F., Symposium on Military Operations Research, *Operations Res.*, 9, № 2, 258 (March—April 1961).
102. Wilson N. A. B., Manpower Research in a Defense Context, American Elsevier, N. Y., 1969.
103. Wohlstetter A. J., Strategy and the Natural Scientists, in R. Gilpin and C. Wright (Eds.), Scientists and National Policymaking, Columbia University Press, N. Y., 1963.
104. Wohlstetter A. J., Hoffman F. S., Lutz R. J., Rowen H. S. Selection and Use of Strategic Air Bases, The Rand Corporation, R-266, April 1954.
105. Wolf C., Jr., Defense and Development in the Less Developed Countries, *Operations Res.*, 10, № 6, 828—838 (November—December 1962).
106. Young J. P., A Survey of Historical Developments in War Games, ORO-SP-98, Operations Research Office, The Johns Hopkins University, Bethesda, Maryland, March 1969.
107. Zehna Peter W. (Ed.), Selected Methods and Models in Military Operations Research, Office, of Naval Research, Washington, D. C., 1971.

II-6. Электроэнергетические системы

Л. Гарвер¹⁾

ВВЕДЕНИЕ

Электричество используется для передачи от источников к потребителям энергии, накопленной в ископаемом топливе (уголь, нефть, природный газ), кинетической энергии падающей воды или ядерной энергии. Электрическая энергия может быть преобразована в тепловую (в электронагревательных приборах и устройствах), в механическую (в электромоторах), в химическую (при зарядке аккумуляторных батарей). Процесс производства электроэнергии является непрерывным, причем количество производимой энергии должно быть таким, чтобы обеспечивался необходимый уровень ее потребления.

Основы электроэнергетики довольно подробно изложены в работе [73]²⁾. В настоящей главе основное внимание уделяется тем проблемам, решение которых возможно с помощью методов исследования операций. Следует отметить, что развитие этих методов с целью их практического применения в области энергетики осуществлялось в основном по двум направлениям. Одно из них сформировалось на базе использования таких подходов к решению проблем и реализации принятых решений, которые являлись результатом совместных усилий практиков, инженеров, математиков и специалистов в других областях науки и техники. Примерами подобных исследований являются работы по оперативному управлению энергосистемами с учетом потерь в линиях электропередачи (разд. 3.2) и планирование резерва мощности (разд. 4.1). Другое направление основывается на использовании математического аппарата дисциплины исследования операций. Примерами работ этого направления являются краткосрочное прогнозирование величины нагрузки с помощью метода экспоненциального сглаживания временных рядов (разд. 2.1), управление функционированием энергосистем на базе методов динамики и декомпозиции (разд. 3.6), планирование развития сетей электропередач с применением методов линейного программирования (разд. 5).

Последующие разделы данной главы посвящены анализу проблем, связанных с прогнозированием, оперативным управлением и планированием капитальных затрат для различных энергосистем.

Прогнозирование — это, по существу, определение пиковых значений мощностей и общего количества генерируемой электроэнергии, необходимых для удовлетворения будущих потребностей. Эти величины используются как для разработки долгосрочных программ капитальных затрат, так и для принятия решений по текущему управлению режимами энергосистем.

¹⁾ Leonard L. Garver, Ph. D., Senior Application Engineer in Advanced System Planning, Electric Utility Systems Engineering Department of The General Electric Co., Schenectady, New York.

²⁾ В работе [73] рассматриваются такие вопросы, как структура и развитие электроэнергетики, загрязнение окружающей среды, эстетические аспекты, надежность, координация работы различных типов электростанций, финансирование, научные исследования и разработки в области электроэнергетики.

Оперативное управление заключается в основном в распределении нагрузок, выборе энергоблоков и агрегатов, составляющих систему, координации режимов работы тепло- и гидроэлектростанций, календарном планировании ремонтных работ.

Естественно, что в пределах одной главы не представляется возможным рассмотреть все существующие методы исследований операций, которые могут быть использованы при решении проблем электроэнергетики, и все сферы приложения этих методов. Поэтому мы ограничились изложением только основных сведений по этому вопросу, которые, как нам представляется, помогут читателю самостоятельно осуществить анализ подобных проблем.

1. ПРОГНОЗИРОВАНИЕ НАГРУЗКИ

Характер прогноза нагрузки электроэнергетических систем зависит от прогнозируемого показателя системы, временного интервала, на который делается прогноз, и от используемого метода прогнозирования.

В качестве прогнозируемого показателя могут быть использованы следующие характеристики системы:

- пиковая мощность нагрузки (МВт) в течение суток, сезона или года ¹⁾;

- распределение (график) нагрузки в течение суток, месяца или года;

- количество электроэнергии (МВт · ч), потребляемой в течение суток, месяца или года.

Временной интервал может составлять от одного часа до нескольких недель (краткосрочный прогноз) или от одного сезона до нескольких лет (долгосрочный прогноз).

Что касается методов прогнозирования, то они могут быть следующими:

- использование ретроспективных данных для временных интервалов (сутки или последовательность суток), имеющих ряд одинаковых признаков;

- экстраполяция временных рядов;

- разделение прогнозируемого показателя на отдельные составляющие;

- установление зависимости между прогнозируемым показателем и показателями, характеризующими погодные условия, численность потребителей и деловую активность;

- метод скользящего среднего;

- метод экспоненциального сглаживания.

¹⁾ Единицей измерения электроэнергии является ватт-час, а электрической мощности — ватт. Так как 1 Вт — это слишком малая величина, то нагрузка или генерируемая мощность в электроэнергетике обычно выражается в миллионах ватт, т. е. в мегаваттах. Например, пиковая нагрузка электроэнергетической системы компании Baltimore Gas and Electric в 1973 г. составила 3 334 000 000 Вт, или 3334 МВт, а общий объем потребления электроэнергии в этом же году был равен 15 395 121 МВт·ч.

Прогнозирование электропотребления и мощности нагрузки осуществляется электроэнергетическими компаниями и объединениями. Прогнозирование энергопотребления с целью оценить предполагаемые доходы, затраты на топливо и т. д. является функцией финансовых органов, а прогнозирование пиковых нагрузок для управления режимами работы и планирования развития электроэнергетических систем — функцией органов оперативного управления и плановых органов. Нередко различные отделы и службы одной и той же компании выполняют прогнозные расчеты независимо друг от друга. Существующая атмосфера конкуренции заставляет каждого специалиста, занимающегося прогнозированием, постоянно стремиться к более точному прогнозу, что в свою очередь требует использования наиболее совершенных методов.

Прогноз должен обеспечивать достаточно точное предсказание, и хотя приемлемая ошибка может составлять около 3%, желательно, чтобы она не превышала 2% истинного значения прогнозируемой величины. Максимально возможная точность прогноза зависит от погрешности определения мощности нагрузки, измеряемой на выходе генератора; эта погрешность обычно составляет ~ 1% [149].

1.1. Краткосрочное прогнозирование

Для того чтобы определить величину нагрузки, приходящейся на середину следующего дня, опытный оператор энергосистемы прежде всего уточнит, какая нагрузка была в это же время суток сегодня, неделю назад, попытается установить, имеется ли связь между значениями нагрузки для аналогичного и предшествующего дней на прошлой неделе, и одновременно проверит данные для аналогичных дней прошлого периода, когда погодные условия менялись таким же образом, как и в течение настоящей недели. Выявив определенные тенденции и определив относительную значимость имеющихся данных, а также учитывая специфические показатели деловой активности, оператор может дать соответствующую прогнозную оценку мощности нагрузки [74].

Улучшить подобный метод прогнозирования довольно трудно. Тем не менее сопоставление результатов каждого прогноза с фактическими данными позволяет оператору накопить определенный опыт и научиться отделять важные показатели от второстепенных. В результате формируется система показателей, необходимых для надежного прогноза. Поскольку, однако, наряду с новыми данными могут быть использованы и старые, или менее важные данные, то производится «взвешенная» оценка значимости (или определение относительной значимости) имеющихся данных.

Проиллюстрируем трудности, связанные с автоматизацией процесса краткосрочного прогнозирования, на примере разработки машинной программы, предназначенной для прогнозирования суточной нагрузки одной из электроэнергетических систем, расположенных в районах Среднего Запада США. Чтобы добиться большей точности и эффективности прогноза с помощью соответствующей модели, инженер-специалист по таким системам обратился к оператору с просьбой изложить суть принятого способа прогнозирования, с тем чтобы включить некоторые из применяемых процедур в разрабатываемую программу. Оператор, который имел большой опыт в области прогнозирования, но не был знаком с вычислительной техникой, никак не мог понять, почему его просят рассказать о существующем методе, если ЭВМ сама будет определять величину ожидаемой нагрузки. Позднее, когда новая система была внедрена, ее практическое использование продолжалось всего лишь несколько недель, после чего программа запускалась все реже и реже, так как оператор стал вновь использовать традиционные методы прогнозирования. Это было обусловлено тем, что введение и постоянное обновление банка данных программы требовали от оператора больших трудозатрат, а прогнозы, получаемые с ее помощью, были ничуть не точнее, чем прогнозы, составленные с помощью традиционных методов.

Более успешный метод прогнозирования описывается в работе [39], авторы которой провели анализ неавтоматизированных методов прогнозирования, применяемых в электроэнергетических компаниях, и с учетом этого разработали и опробовали модель для предсказания пиковой нагрузки в течение последующих суток. Обобщенное математическое представление этой модели имеет вид:

Пиковая нагрузка = (Базовая составляющая) + (Составляющая, зависящая от дня недели) + (Составляющая, зависящая от температуры воздуха) + (Составляющая, зависящая от облачности).

Величина базовой нагрузки и составляющая, зависящая от дня недели, определялись путем экспоненциального сглаживания соответствующих временных рядов, а две другие составляющие — с помощью регрессионного анализа. В рассматриваемой работе показано, что почти в 65 % случаев ошибка прогноза не превышала 3 %, в то время как точность традиционных методов, используемых компанией, была вдвое меньше. Метод прогнозирования, основанный на представлении прогнозируемой пиковой нагрузки в виде суммы базовой и ряда других составляющих, был разработан многими электроэнергетическими компаниями независимо от авторов работы [73]. Результаты использования этого метода показали целесообразность применения регрессионного анализа для

определения зависимости нагрузки от погодных факторов, а методов экспоненциального сглаживания временных рядов (с учетом данных за последние сутки) — для прогнозирования базовой нагрузки и составляющей, зависящей от дня недели.

В работе [40] применительно к задаче прогнозирования нагрузки электроэнергетической системы American Electric Power (на следующие 24 часа) описан обобщенный метод экспоненциального сглаживания с использованием разложения в конечный ряд Фурье по девяти гармоникам. Однако этот метод не обеспечил повышения точности прогноза и поэтому не нашел практического применения.

Другие подходы к решению проблемы прогнозирования нагрузки на последующие сутки, базирующиеся на использовании внутренних по сложности моделей, которые, однако, вызывают сомнения с точки зрения возможности их практического применения, описаны в работах [88] и [121]. В обеих указанных работах моделируются базовая нагрузка и составляющие, учитывающие влияние дня недели и погодных условий, причем базовая составляющая определяется с помощью скользящего среднего за период в семь дней. Для определения составляющей, учитывающей влияние дня недели, авторы работы [121] использовали ожидаемое или долгосрочное сезонное среднее значение нагрузки, а при моделировании влияния погодных условий — два наибольших собственных значения ковариационной матрицы. Для определения составляющей, отражающей влияние дня недели, авторы работы [88] применили метод экспоненциального сглаживания, а влияние погодных условий определяли с помощью уравнения регрессии между величиной нагрузки и средней температурой воздуха. Рецензенты обеих работ отметили следующие факторы, ограничивающие применимость предложенных методов: громоздкость моделей (действительно, нужно ли оперировать 1344 значениями параметров, если тот же самый результат можно получить, используя лишь 60 значений?), неоправданную линеаризацию (изменение температуры на 5°C оказывает разное влияние при значениях исходной температуры 20 или 30°C), а также общие трудности восприятия моделей их потенциальными пользователями.

В одной из рецензий на работу [88] предельно ясно показано, что состав исходных данных и требования модели не всегда обоснованы. Территория, обслуживаемая компанией Pacific Gas and Electric, включает четыре климатические зоны, и все же управление системой осуществляется по единственному параметру — суммарной выработке электроэнергии. Таким образом, четыре компоненты нагрузки трудноизмеримы, и для сбора данных об этих составляющих в главных узлах (на подстанциях) потребовалось бы установить дорогостоящее телеметрическое оборудова-

ние. Многие фирмы, такие, например, как Baltimore Gas and Electric Co., вполне справляются с управлением своими распределительными сетями [100], так как обслуживаемая ими территория находится в одном климатическом поясе. Опыт эксплуатации более разветвленных систем, пугающихся в сборе большого числа данных, показывает, что сбор таких данных требует существенных затрат.

1.2. Долгосрочное прогнозирование

Долгосрочное прогнозирование позволяет определить пиковую нагрузку для летнего сезона следующего года или 20 будущих летних сезонов, если известны соответствующие данные за прошлые годы, количество электроэнергии, которое потребуется в июле следующего года, в течение всего года и т. д. Модели, используемые для долгосрочных прогнозов, можно проверить лишь один раз в год, в то время как модели краткосрочного (суточного) прогнозирования могут проверяться и корректироваться почти каждый день. Применение моделей долгосрочного прогнозирования естественным образом связано с решением таких задач, как финансовое планирование или планирование развития мощностей системы. Процедуры соответствующих прогнозных расчетов, как правило, базируются на анализе временных рядов экономических показателей.

В упоминавшемся ранее обзоре состояния национальной энергетики США [73] рассматриваются методы долгосрочного прогнозирования, используемые в четырех электроэнергетических системах: Tennessee Valley Authority, Pennsylvania Power & Light Co., Southern California Edison Co. и Commonwealth Edison Co. (Чикаго). Все эти компании, за исключением последней, начинают соответствующие прогнозные расчеты с экстраполяции величины электропотребления, при этом выделяются различные группы потребителей электроэнергии в зависимости от целей, для которых она используется, например для нужд рекламы и объявлений, бытовых, сельскохозяйственных или промышленных нужд либо каких-то других. Затем прогнозные оценки, полученные для каждой группы потребителей, суммируются, и таким образом определяется общая величина ожидаемого электропотребления в будущем сезоне или в течение всего года. Для определения пиковой величины нагрузки сначала по ретроспективным данным определяется соотношение между пиковой нагрузкой и общей выработкой электроэнергии, а затем найденная величина умножается на предсказуемое значение суммарной потребности. Анализ результатов ограничивается лишь указанием на то, что данный метод позволяет получить ожидаемое значение прогнозируемой характеристики.

Компания Commonwealth Edison определяет пиковое значение нагрузки путем прямых расчетов, применяя прогнозирующую модель, в которой выделены базовая составляющая и составляющая, учитывающая влияние погодных условий. Базовая составляющая нагрузки определяется путем исключения из соответствующих данных за прошлые годы составляющей, зависящей от показателей, характеризующих погодные условия, и составляющей, которая зависит от показателей объема промышленного производства, а также от некоторых специальных показателей, отражающих, например, снижение напряжения в сети или периоды сокращения производства. Составляющая, зависящая от погодных условий, рассчитывается с учетом следующих параметров: суточного снижения температуры за последние трое суток, трех соответствующих значений температуры воздуха (по сухому термометру), относительной влажности воздуха за 2 часа до пиковой нагрузки, а также порядкового номера временного интервала (например, года и суток). В результате прогноза определяется ожидаемая величина пика нагрузки, возможные отклонения от которой рассчитываются с помощью методов теории вероятностей.

В работе [120], посвященной анализу практического опыта компании Commonwealth Edison по прогнозированию месячных пиков нагрузки (в период с 1959 по 1966 г.), показано, что распределение ошибок прогноза в большей степени зависит от месяца, на который составляется прогноз, чем от периода прогнозирования. Так, например, прогноз на декабрь, составляемый в феврале того же года, может быть сделан с такой же точностью, как и прогноз, составляемый на сентябрь в августе. В указанной работе подвергается сомнению целесообразность тех частых корректировок прогнозирующей модели, которые при существующей практике прогнозирования в данной компании производились через каждые 3—6 недель.

Методы, используемые для прогнозирования пиковых значений нагрузок компанией Public Service Electric & Gas (шт. Нью-Джерси), также основаны на выявлении тренда базовой составляющей после исключения из общей величины нагрузки составляющей, которая зависит от погодных условий [97]. В соответствующее уравнение линейной регрессии включены следующие переменные, характеризующие погодные условия: температура воздуха по мокрому и сухому термометру, относительная влажность воздуха, а также коэффициент, учитывающий возможность накопления тепла. В модель включена также еще одна составляющая нагрузки, отражающая влияние необъяснимых факторов. Рассматриваемая модель требует существенной доработки, так как ее проверка по ретроспективным данным за 9 лет показала, что отклонение прогнозируемых значений

нагрузки от фактических достигало 5,8% при средней ошибке прогноза 2,4%. Рецензентами работы [97] было предложено ввести в модель несколько дополнительных факторов, например уровень деловой активности. Обсуждение этой работы специалистами электроэнергетических компаний Кливленда, Северо-Запада США, Рочестера, Нью-Йорка, Филадельфии и Лонг-Айленда показывает, что в настоящее время моделирование влияния погодных условий на величину нагрузки находит широкое применение. Это объясняется тем, что при прогнозировании развития энергосистемы необходимо исключить ту составляющую нагрузки, которая определяется погодными факторами.

Исследования составляющей нагрузки, зависящей от погодных факторов, стали интенсивно развиваться, начиная с 1966 г., в связи с тем, что фактическая пиковая нагрузка в нескольких энергетических системах Среднего Запада США оказалась на 8% выше ожидаемой. Такая ошибка прогноза была во многом обусловлена тем, что после двух относительно прохладных летних сезонов лето в 1966 г. выдалось очень жарким. После этого компания Union Electric (Сент-Луис), которая оказалась в числе тех, кто испытал на себе последствия этой ошибки, пришла к выводу о необходимости представлять составляющую, которая зависит от погодных условий, в виде двух компонент, характеризующих соответственно бытовое и промышленное электропотребление [108]. Базовая составляющая, моделируемая с использованием функции регрессии с экспоненциальным весом, возрастает на 6% в год, в то время как составляющая, зависящая от погодных условий, увеличивается за год на 14—16% и в 1969 г. составляла 3/8 пикового электропотребления за весь летний период. В работе [108] показано, что, хотя в количественном отношении новые результаты мало отличаются от прежних, используемый в настоящее время метод прогноза имеет более солидную основу и обеспечивает оценку степени неопределенности прогноза.

В результате графического анализа имеющихся данных о зависимости мощности нагрузки от температуры воздуха специалисты научно-исследовательского центра компании Perdue Energy установили, что результирующий график можно представить в виде U-образной кривой [74, 161]. Если выбирать два характерных значения температуры воздуха, между которыми нагрузка считается чувствительной к изменению температуры, то такая зависимость может быть изображена в виде трех прямолинейных отрезков. Величина нагрузки в интервале температур, не влияющих на нагрузку, а также изменение нагрузки в летние и зимние сезоны в зависимости от температуры воздуха определяются с помощью уравнений линейной регрессии.

Метод, предложенный в работе [161], был опробован несколькими электроэнергетическими компаниями, однако не нашел

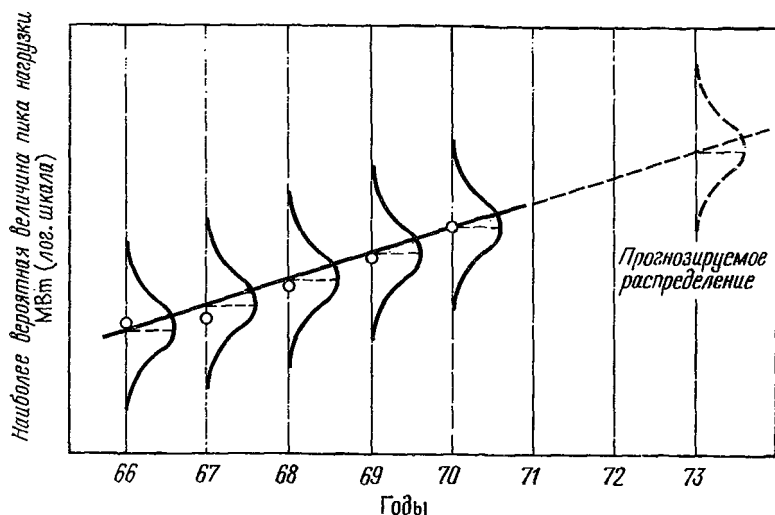


Рис. 1. Прогнозирование пиковых нагрузок на основе ретроспективных данных о погодных условиях.

широкого распространения. Было установлено, что получаемые результаты сильно зависят от выбора диапазона температур U-образной кривой, в котором нагрузка считается постоянной. К тому же по сравнению с существующими методами предлагаемый подход не обеспечивает заметного улучшения точности прогнозов, что делает неоправданным усложнение расчетов, необходимое для его реализации. При обсуждении работы [161] специалистами компании Gulf States Utilities было показано, что при периоде прогнозирования, превышающем 6 лет, пиковая мощность нагрузки, определяемая с помощью метода наименьших квадратов, отклонялась от истинного значения нагрузки не более чем на 1,4% в сторону уменьшения и не более чем на 2,6% в сторону увеличения. Это, по-видимому, свидетельствует о том, что нет никакой необходимости в использовании сложных методов прогнозирования.

Исследования [52], выполненные по инициативе компании Louisiana Power & Light, целью которых являлось прогнозирование дневных пиков нагрузки в течение летнего сезона с помощью методов множественной регрессии, позволили установить следующую зависимость пиковой суточной нагрузки для летнего

сезона от погодных условий:

Нагрузка = -670

- +21,4 (температура воздуха в 3 ч текущего дня, °C)
- +13,1 (температура воздуха в полдень текущего дня, °C)
- +7,4 (температура воздуха средняя за трое суток, °C)
- +1,9 (относительная влажность воздуха в 3 ч текущего дня)
- +2,7 (температура воздуха средняя за трое предыдущих суток, °C)
- +2,9 (скорость ветра в полдень текущего дня, м/с)

Используемые в данном случае переменные не являются независимыми, поэтому статистики могли бы указать на несостоятельность этой модели. Однако ее использование дает хорошие результаты, и она применяется в настоящее время для прогнозирования пиковых суточных нагрузок для летнего периода на предстоящие 5—10 лет.

Подставляя в полученное соотношение данные для каждого года прошлого периода и производя соответствующие расчеты по имеющимся ретроспективным данным о погодных условиях за последние 12 лет, определяют распределение вероятностей величины нагрузки для каждого года. Это распределение может быть представлено графически с использованием полулогарифмической шкалы (рис. 1), причем для каждого из распределений осуществляется экстраполяция средней величины годовой пиковой нагрузки и тренда дисперсии.

В обзорной работе [66], посвященной обсуждению методов прогнозирования нагрузки, сообщалось, что из 57 электроэнергетических компаний 30 включают в используемые ими прогнозирующие модели те или иные погодные факторы.

Для объяснения наблюдаемого роста мощности нагрузки требуется введение дополнительных составляющих. В работе [117] дается описание прогнозирующей модели для определения месячного энергопотребления, применяемой компанией Public Service Electric & Gas. В этой модели используются следующие составляющие:

- составляющая, характеризующая численность потребителей, использующих электроэнергию для бытовых нужд;
- базовая составляющая (кВт · ч), характеризующая количество энергии, необходимое одному потребителю этой группы;
- составляющая, характеризующая расход электроэнергии, потребляемой в целях кондиционирования воздуха в жилых помещениях;
- составляющая, зависящая от переменных погодных факторов.

На основе ретроспективных данных и оценок экспертов для всех четырех составляющих строят нормальное распределение нагрузки во времени. Количественные описания, связанные с экспертными оценками, представляют собой распределения

вероятностей, которые характеризуют субъективные мнения наиболее опытных представителей эксплуатационного персонала, выражаемые в виде утверждений такого рода: «Имеется один шанс из ..., что число потребителей электроэнергии для бытовых нужд возрастет в следующем году до ...». После объединения данных, характеризующих такие распределения, и с учетом корреляции результаты прогноза представляются в столь же трудноинтерпретируемой форме, например: «Имеется один шанс из 2, что энергопотребление в июне следующего года составит X МВт, один шанс из 4, что эта величина составит Y МВт · ч, и т. д.»

Один из авторов работы [117], представляющий компанию Tennessee Valley Authority, описывает применение метода Монте-Карло для объединения указанных распределений, которые в общем случае не обязательно должны подчиняться нормальному закону. В заключение автор рассматриваемой работы отмечает, что, исключая из расчетной схемы метод Монте-Карло, можно существенно уменьшить машинное время, требуемое для проведения расчетов.

Специалисты электроэнергетических компаний, занимающиеся прогнозированием, характеризуя консерватизм, присущий используемым методам долгосрочного прогнозирования, привели данные о том, что прогнозные оценки пиковой нагрузки на 1971—1972 гг., сделанные 57 компаниями за десять лет до этого, оказались в среднем на 6% ниже фактических значений [66]. Прогнозные оценки, сделанные на эти же годы за пять лет до рассматриваемого срока, оказались всего лишь на 1,5% ниже фактических пиковых значений нагрузки, в то время как прогнозы, выполненные непосредственно в предшествующий год, оказались на 2% выше фактических значений.

Характерная особенность долгосрочного прогнозирования, осуществляемого в настоящее время, состоит в том, что, как правило, предсказывается постепенное уменьшение темпов роста к концу предстоящего пятилетнего периода, поскольку считается, что постоянный темп роста электропотребления на уровне 7—8% в год невозможен. Вследствие этого с течением времени и под влиянием того, что на самом деле темпы роста сохраняются или даже возрастают, соответствующие прогнозные оценки неуклонно возрастают. Долгосрочное прогнозирование страдает и от другой формы проявления консерватизма: завышенные прогнозные оценки приводят к менее тяжелым последствиям, чем заниженные, и этот фактор способствует составлению прогнозов, дающих прогнозные оценки пиковых нагрузок, которые превышают фактические значения.

Было предложено много моделей, охватывающих период прогнозирования от одного часа до 30 лет и использующих простую экстраполяцию, методы экспоненциального сглаживания и мно-

жественной регрессии. Практическое применение нашли лишь те модели, для которых характерны следующие общие признаки: простота интерпретации, участие в их разработке самих пользователей и простота получения исходных данных. Независимо от сложности применяемого метода все прогнозные оценки пересматриваются и корректируются с учетом заключений экспертов, основанных на практическом опыте [66].

2. ПЛАНИРОВАНИЕ ПРОИЗВОДСТВЕННЫХ ПРОЦЕССОВ

Задачи текущего планирования связаны с выбором агрегатов, синхронизируемых с системой, и распределением нагрузки между работающими энергоблоками и агрегатами. Затраты на эксплуатацию этих энергоагрегатов независимо от того, установлены они на ТЭС, ТЭС или АЭС, следует минимизировать, соблюдая при этом баланс между нагрузкой и генерируемой мощностью, что необходимо для надежной работы электроэнергетической системы.

В данном разделе будут рассмотрены вопросы, связанные с текущим планированием работы систем, в том числе и эксплуатационная надежность систем. Читатель, желающий узнать, какие методы исследований операций используются при изучении отдельных вопросов, обсуждаемых в данном разделе, может воспользоваться приведенной ниже табл. 1.

2.1. Календарное планирование производства электроэнергии без учета потерь

Подобное планирование предусматривает распределение нагрузки (генерируемой мощности) между работающими электрогенераторами. Если потери электроэнергии внутри системы невелики, что имеет место в сетях с высокой плотностью нагрузки, то задачи календарного планирования могут быть сформулированы в наиболее простом виде. Типичной задачей календарного планирования является, например, задача распределения нагрузки в 60 МВт между двумя генераторами при минимальных затра-

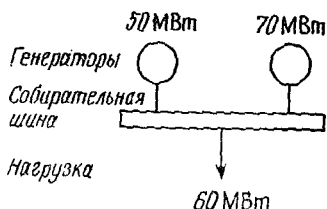


Рис. 2. Схематическое изображение типичной задачи с минимальными затратами

Таблица 1

Методы исследования операций, используемые при решении проблем эксплуатации энергосистем

Проблемы	Методы исследования операций
Календарное планирование производства электроэнергии без учета потерь	ЛП, ЦЛП, ДП
Календарное планирование производства электроэнергии с учетом потерь	ЛП, НЛП
Координация работы ГЭС и ТЭС	ЛП, НЛП, ДП
Планирование использования установленных энергоагрегатов	ЦП, ДП, СП
Надежность системы	ЛП, НЛП, ПС
Календарное планирование производства электроэнергии в больших системах	ДД, ПС
Календарное планирование ремонтных работ	ЦП

Обозначения: ЛП — линейное программирование, НЛП — нелинейное программирование, ЦП — целочисленное программирование, ДП — динамическое программирование, СП — стохастические процессы, ПС — потоки в сетях, ДД — декомпозиция и диаконтинка.

тах на топливо (рис. 2). Иными словами, необходимо минимизировать стоимость расходуемого топлива для покрытия суммарной нагрузки при условии, что мощность, развиваемая каждым из агрегатов, не превышает нижнего и верхнего пределов, устанавливаемых правилами технической эксплуатации. В данном случае задача может быть сформулирована следующим образом:

найти

$$\min \sum_{\text{по всем агрегатам}} \text{Стоимость топлива} \quad (1)$$

при ограничениях

$$\sum_{\text{по всем агрегатам}} P = P_{\text{н}}, \quad (2)$$

$$P_{\min} \leq \text{Мощность} \equiv P \leq P_{\max} \text{ для каждого из работающих агрегатов,} \quad (3)$$

$$P = 0 \text{ для всех неработающих агрегатов.} \quad (4)$$

Стоимость топлива C для каждого агрегата есть функция развиваемой им мощности (рис. 3), причем $C = 0$, если $P = 0$.

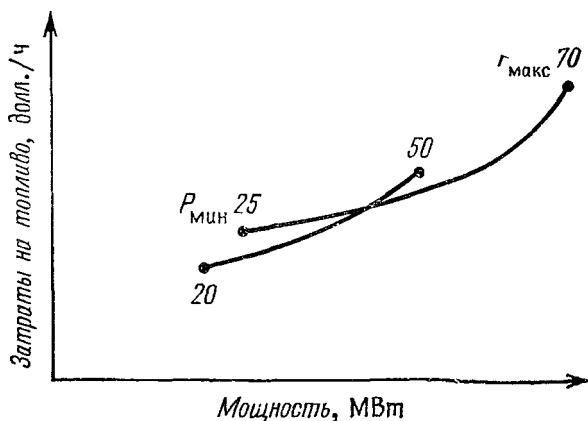


Рис. 3. Зависимость затрат на топливо от мощности, развиваемой турбогенератором.

Еще в 1933 г. было установлено, что в рассматриваемом случае энергоагрегаты должны работать в таком режиме, при котором дополнительные затраты, связанные с дальнейшим увеличением мощности каждого из них, будут одинаковы [162]. В промышленности эти дополнительные затраты ассоциируются с λ -коэффициентами, поскольку исходное доказательство вышеуказанного положения базируется на использовании метода множителей Лагранжа (обозначаемых обычно буквой λ). Рассматриваемая задача решается на ЭВМ с помощью итерационных процедур расчета суммарной мощности по параметру λ , для чего используется зависимость дополнительных затрат от мощности, развиваемой каждым из агрегатов (рис. 4). Как только суммарная

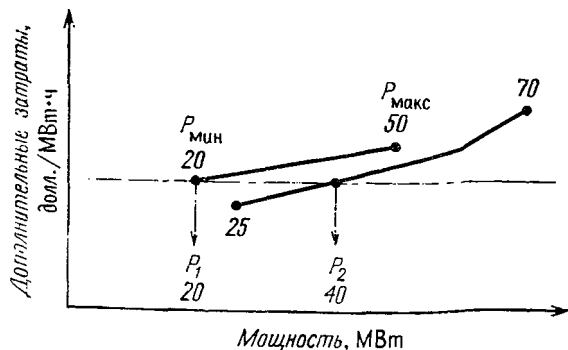


Рис. 4. Зависимость дополнительных затрат от мощности турбогенераторов и системы.

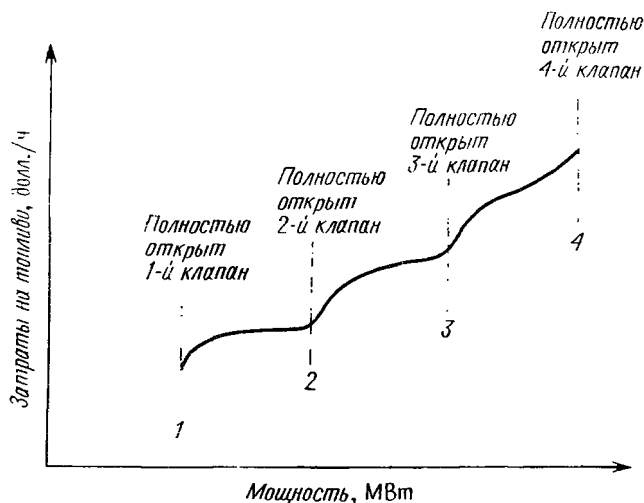


Рис. 5. Зависимость затрат на топливо от мощности, развиваемой паровой турбиной, с учетом характеристик регулирующих клапанов.

мощность агрегатов становится равной общей мощности нагрузки, итерации прекращаются.

Распределение нагрузки, которое базируется на принципе равенства дополнительных затрат, подробно рассматривается в работах [111, 114, 153, 163].

Описанный принцип распределения нагрузки является общепризнанным и используется в современных автоматизированных системах диспетчерского управления энергосистемами. Однако реализация таких систем связана с определенными трудностями: во-первых, необходимо выполнять определенные требования, обеспечивающие надежность работы системы; во-вторых, необходимо учитывать высокие потери в линиях электропередачи с низкой плотностью нагрузки. Кроме того, используемые в расчетах зависимости между входными и выходными параметрами энергоагрегатов не могут быть аппроксимированы с требуемой степенью точности линейными или квадратичными функциями. Рассмотрим более подробно последний из указанных аспектов проблемы.

В США большинство паровых турбин энергоагрегатов снабжено несколькими паровыми клапанами для регулирования мощности. Зависимость затрат на топливо от величины развиваемой мощности при последовательном открытии регулирующих клапанов паровой турбины представлена на рис. 5. Так как при незначительном открытии клапана возникают существенные потери,

связанные с дросселированием, то следует всячески избегать подобных режимов работы турбины.

В 1959 г. группа специалистов в области термодинамики, математики и теории оптимизации приступила к разработке методов расчета точной зависимости между входными параметрами и мощностью турбины и к поиску методов оптимизации распределения нагрузки, использующих точные нелинейные зависимости [146]. Для нахождения строго оптимального режима работы энергоагрегатов были использованы методы динамического программирования, причем результаты реализации оптимального решения сравнивались с результатами расчетов, выполненных с помощью различных приближенных методов. Сравнение показало, что модель, использующая для вычисления зависимость, представленную на рис. 5, позволяет найти более экономичные режимы, чем обычно применявшаяся в то время модель, в которой использовалась зависимость, показанная на рис. 4. Тем не менее эта традиционная модель широко использовалась в электроэнергетике еще в течение длительного времени. Это объясняется тем, что очень трудно точно установить координаты нежелательных участков результирующей кривой (рис. 5), зависящие от фактических характеристик паровых клапанов.

Если, например, в соответствии с предлагаемой моделью оказывается, что мощность 50 МВт может быть получена до открытия следующего регулирующего клапана, но вследствие износа или других причин полное открытие работающего клапана фактически обеспечивает мощность, равную всего лишь 49 МВт, то использование зависимости, которая приведена на рис. 5 и в которой началу следующего этапа соответствует мощность в 50 МВт, приведет к работе турбины с незначительным открытием следующего клапана, т. е. как раз в наименее экономичном режиме. Все дело в том, что не существует простых и надежных способов для точного определения момента открытия следующего клапана. На некоторых электростанциях была установлена специальная измерительная аппаратура, контролирующая давление и температуру пара перед регулирующими клапанами, что позволяет зарегистрировать открытие клапана и посылать соответствующий сигнал или некоторую комбинацию сигналов в центральный диспетчерский пункт. Однако в этом случае требуются дополнительные затраты времени на получение центральным диспетчерским пунктом точных данных о режиме работы каждого агрегата.

2.2. Календарное планирование производства электроэнергии с учетом потерь

Как известно, передача электроэнергии на расстояние сопровождается потерями. Наибольшая часть потерь связана с нагревом проводов линий электропередачи, при этом величина выделяющейся тепловой мощности равна произведению величины сопротивления на квадрат силы тока [163]. Чем больше расстояние, на которое передается электрическая энергия, тем выше сопротивление линии электропередачи; чем больше передаваемая мощность, тем больше сила тока при заданном напряжении. Следовательно, если источники электроэнергии и потребители находятся на большом расстоянии друг от друга (рис. 6), то вырабатывать электроэнергию при больших затратах вблизи пунктов потребления (снижая тем самым потери в линиях электропередачи) оказывается более выгодным, чем использовать режим равных затрат для всех агрегатов, удаленных от пунктов потребления электроэнергии на разные расстояния. В этом случае в ограничение (2) вводится величина потерь, и оно приобретает вид

$$\sum_{\text{по всем агрегатам}} P = P_n - Q, \quad (5)$$

где *потери* представляют собой нелинейную функцию от переменных, определяющих величину *мощности*.

Практический метод учета этого нелинейного ограничения состоял в том, чтобы представить потери как квадратичную функцию от переменных, определяющих мощность, и установить норму штрафа для корректировки дополнительных затрат по каждому агрегату, перед тем как приравнять их друг к другу в соответствии с упомянутым ранее принципом использования системы λ -коэффициентов. Этот способ впервые был предложен в 1943 г.,

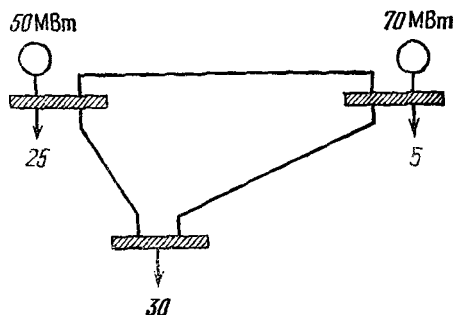


Рис. 6. Схематическое изображение задачи о наиболее экономичном распределении нагрузок с учетом потерь в линиях электропередачи.

но нашел практическое применение только после того, как Габриель Кроп построил серию инвариантных преобразований функции потерь, что позволило рассчитать входящие в формулу потерь постоянные коэффициенты, названные *B*-коэффициентами [111].

Метод, использующий формулу потерь, был разработан комплексной группой: операторы энергосистемы, обычно внедряющие новые методы, работали над ним совместно с математиками и инженерами-электротехниками. Реализация этого метода применительно к энергосистемам, включающим восемь электростанций с паровыми турбинами общей мощностью 800—1900 МВт, позволяет сэкономить почти 150 тыс. долларов в год [111]. Метод, базирующийся на формуле потерь, хотя и получил широкое применение в электроэнергетике [115], все же нуждается в обосновании ряда важнейших моментов и используемых приближений.

Чтобы избежать аппроксимаций, связанных с формулой потерь, требуется решить систему уравнений сетевой задачи, минимизируя затраты на топливо. Уравнение (5) заменяется в этом случае двумя уравнениями для каждого узла сети. Первое уравнение — это уравнение баланса активной мощности для рассматриваемого узла, т. е.

$$P - P_{\text{н}} = \sum_{\substack{\text{по всем линиям} \\ \text{данного узла сети}}} P. \quad (6)$$

Второе уравнение — это уравнение баланса реактивной мощности для рассматриваемого узла сети, т. е.

$$P_r - P_{r\text{н}} = \sum_{\substack{\text{по всем линиям} \\ \text{данного узла сети}}} P. \quad (7)$$

Реактивная мощность расходуется на создание электромагнитных полей, возникающих при работе электрических моторов, трансформаторов, а также вокруг любого проводника. Применяя к этой гораздо более объемной задаче условия Куна — Таккера и дополняя метод Ньютона процедурами наискорейшего спуска, можно найти распределение нагрузки, соответствующее минимальным затратам [135].

После того как в 1968 г. была опубликована работа [135], начался поиск наилучших комбинаций различных методов математического программирования для решения задачи оптимального распределения нагрузки. Были предложены методы линейного программирования [104], квадратичного программирования [129], метод сопряженного градиента [150] и градиентные методы, использующие функцию штрафов [24, 136, 164]. Согласно оценке авторов работы [151], одним из лучших является видоизмененный градиентный метод с коэффициентом штрафа [58].

На основании работ, представленных Англией, Францией, СССР, Швецией и Австралией на III Международной конференции по вычислительным методам, применяемым при исследовании электроэнергетических систем [118], можно сделать вывод о том, что вопросы оптимизации распределения нагрузки с учетом влияния сети электропередачи привлекают внимание специалистов всех стран. Тем не менее лишь представитель французской компании Electricité de France отметил возможность применения методов и при распределении нагрузки в пределах суточного интервала.

Частые пересчеты, необходимые для установления нового режима (осуществляемые с интервалом от 2 мин до 2 с), и ограничения ЭВМ по производительности требуют, чтобы применяемый вычислительный метод имел простую структуру и позволял быстро производить нужные расчеты [64]. Метод, основанный на использовании формулы потерь, обеспечивает простую и быструю реализацию вычислительных процедур. Получаемое с его помощью решение лишь незначительно отличается по величине затрат от решения, получаемого при использовании строгого метода оптимизации распределения нагрузки [95]. Существенный недостаток метода, использующего формулу потерь, состоит в том, что прослеживание изменений в сети возможно только путем составления новой матрицы B -коэффициентов. По-видимому, это обстоятельство, а также необходимость иметь готовое решение относительно распределения нагрузки в сети для непредвиденных случаев будут способствовать более широкому применению методов математического программирования.

2.3. Координация работы гидро- и теплоэлектростанций

До сих пор речь шла о тепловых электростанциях (ТЭС), т. е. станциях, на которых установлены электрогенераторы, приводимые в действие паровыми турбинами. Другой важнейший тип электростанций — гидроэлектростанции (ГЭС), использующие для получения электроэнергии кинетическую энергию падающей воды. Так как гидроресурсы ограничены, минимизация затрат на производство электроэнергии требует использования обоих типов электростанций — ГЭС и ТЭС.

На рис. 7 показана энергосистема, включающая две каскадные ГЭС, при этом выходной поток первой станции является входным потоком для второй ГЭС. Для выбора режимов совместной эксплуатации ГЭС и ТЭС, обеспечивающих минимальные общие затраты на производство электроэнергии, требуется исследование эксплуатационных циклов гидравлических станций продолжительностью от нескольких дней до нескольких недель.

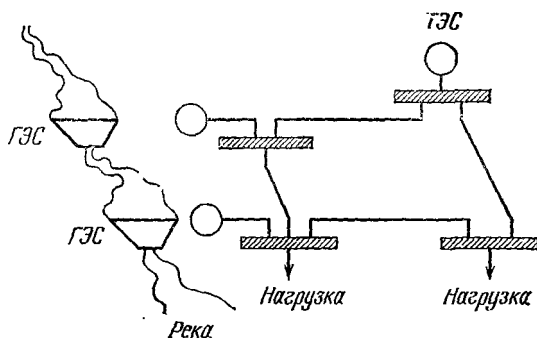


Рис. 7. Схематическое изображение задачи о координации режимов работы ГЭС и ТЭС.

Задача о наиболее экономичном распределении нагрузки в такой системе содержит в качестве переменной время, включает ограничения на количество электроэнергии, вырабатываемой ГЭС, и может быть сформулирована следующим образом:

найти

$$\min \sum_{\text{по времени}} \sum_{\substack{\text{по всем} \\ \text{агрегатам} \\ \text{ТЭС}}} \text{Стоимость топлива} \quad (8)$$

при ограничениях

$$\sum_{\substack{\text{по всем} \\ \text{агрегатам}}} P = P_n \text{ для каждого временного интервала,} \quad (9)$$

$$\sum_{\substack{\text{по} \\ \text{времени}}} \sum_{\substack{\text{по всем} \\ \text{гидроагрегатам} \\ \text{ГЭС}}} E \leq \max E \text{ для каждой совокупности гидроагрегатов,} \quad (10)$$

где энергия определяется путем интегрирования выражения для мощности по времени. Уравнения (8) и (9) заменяют приведенные ранее уравнения (1) и (2), не содержащие временной координаты. Общее количество электроэнергии, генерируемой гидроэнергетической частью системы, ограничивается тем максимальным количеством, которое может быть выработано на ГЭС за рассматриваемый период [ограничение (10)].

В работе [169] описан типичный подход к решению этой задачи, используемый многими электроэнергетическими компаниями. Этот подход основан на использовании коэффициента преобразования кинетической энергии падающей воды в электрическую, что позволяет осуществлять распределение нагрузок по принципу равенства дополнительных затрат и в случае смешанной, т. е. гидро-

тепловой, электроэнергетической системы. В рецензиях на указанную работу обсуждаются многие практические вопросы, связанные с решением рассматриваемой задачи и касающиеся, в частности, пуска паровых агрегатов, факторов, определяющих время подхвата нагрузки, времени передачи упреждающих сигналов, учета потерь; освещается также практический опыт управления режимами гидротепловых систем в США, Канаде, Японии и Швеции.

В Японии при решении этой задачи использовались методы линейного программирования [77], а компания Ontario Hydro исследовала возможности применения градиентного метода для определения оптимальных режимов энергосистемы, состоящей из 16 ГЭС и 4 ТЭС [11].

В работах [21, 26] рассматриваются примеры комплексной задачи, включая задачу оптимизации распределения нагрузки, обсуждавшуюся в предыдущем разделе, и задачу координации работы ГЭС и ТЭС. При решении последней авторы использовали метод Флетчера—Риваса. Предлагаемые в этих работах подходы пока не нашли широкого применения.

Для решения задачи оптимального распределения нагрузки между ТЭС и двумя ГЭС, каскадно расположенными на р. Сасквеханна, был успешно использован метод динамического программирования [7]. Однако в работе [113] высказывается мнение, что в случае более крупных гидроэнергетических систем предпочтительнее использовать вариационные методы.

В работе [27] применительно к теплогидравлической системе, включающей две ГЭС, проведен анализ и сравнение трех различных методов, базирующихся на последовательной аппроксимации решений, получаемых с помощью методов динамического программирования. В работе [143] описано применение этих методов при решении задачи календарного планирования производства электроэнергии в объединенной энергосистеме, включающей гидроаккумулирующую станцию (ГАЭС), две ГЭС и 16 ТЭС. При этом учитывалось наличие трех контрактов, предусматривающих взаимный обмен электроэнергией, и период планирования составлял от двух суток до одного года. Для расчета режимов работы энергосистемы, включающей 38 ГЭС с водохранилищами и 45 ГЭС, работающих на бытовом стоке, авторы работы [99] использовали модель, разработанную компанией Bonneville Power Administration, причем решение было получено методом сопряженных градиентов Флетчера—Риваса. Исследование 49 временных интервалов, на которое было потрачено всего лишь 10% времени, обычно затрачиваемого на эти цели техническим персоналом, позволило найти режимы, обеспечившие прирост выработки электроэнергии почти на 1% по сравнению с тем уровнем, который обеспечивался в результате реализации решений, получаемых методом проб и ошибок.

2.4. Планирование использования установленных агрегатов

Как известно, нагрузка энергосистемы резко увеличивается в дневные часы и падает в ночное время. Поэтому, отключая наименее экономичные агрегаты в период низких нагрузок, можно добиться заметной экономии затрат на производство электроэнергии.

На рис. 8 показано изменение нагрузки в течение суток (от дневного пика одних суток до пика следующих суток). Чтобы покрыть пиковую нагрузку с учетом резерва, необходимого на случай непредвиденных ситуаций, должны работать шесть агрегатов; по мере снижения нагрузки шестой, а затем и пятый агрегаты могут быть отключены. При уменьшении числа работающих агрегатов затраты на производство электроэнергии снижаются. Однако отключение и пуск агрегатов приводит к увеличению составляющих эксплуатационных затрат, обусловленных повышенным расходом топлива в переходных режимах и ускоренным износом оборудования. Эти составляющие затрат возрастают с увеличением времени простоя и достигают максимальных значений при пуске из холодного состояния. Все это приводит к тому, что отключение агрегатов на слишком короткое время оказывается экономически невыгодным. В зависимости от проектно-технологических особенностей электростанций правила технической эксплуатации требуют, чтобы отключение агрегатов производилось не менее чем на 6—8 ч.

Задача планирования использования установленных агрегатов формулировалась как задача целочисленного линейного программирования [80, 127] и как задача целочисленного пирамического программирования [9, 89, 122], а также решалась с по-

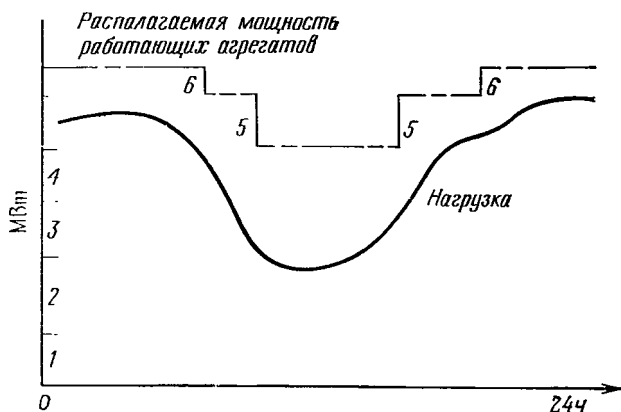


Рис. 8. Суточный график нагрузки и использование установленных агрегатов.

мощью метода ветвей и границ [36]. Однако все попытки использовать методы математического программирования оказывались тщетными из-за необходимости учитывать следующие правила эксплуатации.

1. Любой агрегат может быть выведен из эксплуатации на весь плановый период или его часть.

2. Максимальные или минимальные значения ограничиваемых эксплуатационных характеристик любого агрегата могут изменяться в течение всего планового периода.

3. Контроль надежности режимов системы по величине минимально необходимого вращающегося резерва и допустимых нагрузок линий электропередачи должен осуществляться не реже чем один раз в течение одного часа.

4. Необходимо соблюдать ограничения, устанавливающие минимальное время простоя и продолжительность работы агрегатов в исправном состоянии.

5. В одном и том же пункте одновременно разрешается вводить только ограниченное число агрегатов.

6. Для каждого агрегата необходимо выдерживать допустимую скорость набора нагрузки.

7. Необходимо планировать резерв подхвата нагрузки по всем работающим агрегатам, с тем чтобы мощность, развиваемая системой, непрерывно следовала за изменяющейся нагрузкой, т. е. необходимо, чтобы требуемое повышение мощности достиглось, например, за 1,3 или 5 мин.

8. Необходимо предусмотреть возможности для составления графика работы на сутки и его детализации по четырехчасовым интервалам, а также для составления общего календарного плана на неделю.

9. График работы электрогенераторов ГЭС с собственным водохранилищем длительного регулирования должен составляться с учетом ограничений на рабочий объем водохранилища.

10. Необходимо составлять недельный график эксплуатации ГАЭС.

11. Общая величина затрат на производство электроэнергии должна быть минимальной.

12. Расчеты по составлению календарного графика должны производиться с использованием ограниченного объема оперативного запоминающего устройства ЭВМ.

13. Необходимо, чтобы оптимизация суточного графика нагрузки занимала не более одного часа машинного времени при условии, что ЭВМ используется и для расчетов, связанных с реализацией функций оперативно-диспетчерского управления, имеющих более высокий приоритет.

14. Рассматриваемые системы могут включать до 150 агрегатов.

15. В любом случае программа должна обеспечить получение некоторого допустимого решения (если такие решения существуют) независимо от того, достаточен ли резерв машинного времени для получения оптимального решения.

Для решения рассматриваемой задачи нужны простые и быстрые в реализации расчетные методы, так как размерность задачи слишком велика, а ЭВМ, используемые для ее решения, как правило, имеют процессор небольшой производительности. В практических условиях при решении данной задачи применяются списки приоритетного использования агрегатов, базирующиеся на данных об их средних эксплуатационных затратах при полной нагрузке, и некоторые эвристические процедуры логического характера, с помощью которых в необходимых случаях изменяется очередность ввода агрегатов; для улучшения полученного календарного графика используется ограниченный поиск лучших вариантов [54, 93, 107].

2.5. Надежность энергосистемы

Надежность характеризует функционирующую электроэнергетическую систему с точки зрения влияния возможных возмущений (непредвиденный выход из строя наиболее нагруженного агрегата или линии электропередачи) на работоспособность других элементов системы. Для ликвидации последствий, связанных с непредвиденным выходом из строя электрогенератора, в системе обычно предусматривается наличие достаточного вращающегося (горячего) резерва и соответствующий вариант перераспределения нагрузки, что обеспечивает быстрое увеличение мощности до нужного уровня. Более сложно предусмотреть и ликвидировать последствия внезапного выхода из строя линии электропередачи, так как каждая аварийная ситуация такого рода оказывает специфическое влияние на работу системы, для описания которого требуются расчеты распределения нагрузки в сети.

Уэллс [171] сформулировал задачу о минимизации затрат на производство электроэнергии в виде задачи линейного программирования и использовал модель электрической сети с идеализацией по постоянному току (т. е. рассматривались активные мощности), при условии что изменение дополнительных затрат носит ступенчатый характер. В постановке задачи учитывались ограничения, обусловленные требованиями к надежности системы. В течение ряда лет метод, предложенный Уэллсом, применялся в объединенной электроэнергетической системе шт. Нью-Йорк. С помощью этого метода осуществлялось распределение нагрузки между генерирующими мощностями энергообъединения. При оптимизации режима минимизировались затраты на топливо

и учитывались ограничения, связанные с надежностью сети электропередачи.

В результате дальнейшего развития метода Уэллса было показано, что размерность задачи можно менять в зависимости от количества линий электропередачи, повреждения которых приводят к нарушениям всей системы в целом [155]. В работе приведено детальное решение рассматриваемой задачи на примере электроэнергетической системы, включающей 23 шины.

Другой подход, основанный также на использовании методов линейного программирования, был предложен в работах [106, 165]. Сущность этого подхода заключается в том, что система, испытывающая нарушения, приводится в состояние надежного функционирования путем минимизации отклонения фактического режима работы каждого генератора от того, который в данный момент соответствует наиболее экономичному распределению нагрузки при отсутствии возмущений.

Один из рецензентов работы [51] указал три проблемы, возникающие при моделировании надежности электроэнергетических систем: выбор целевой функции (минимизация перераспределения нагрузки или минимизация затрат), выбор модели энергетической системы (т. е. «рабочей точки» линеаризации) и выбор «сильных» и «слабых» ограничений (т. е. решение вопроса о том, какие из ограничений должны выполняться строго, а какие могут выполняться приближенно). Главная проблема при оценке надежности системы заключается в том, что необходимо точно знать ее рабочее состояние. В настоящее время разрабатываются методы таких оценок по неполной или противоречивой информации о потоках энергии в линиях электропередачи и состоянии оборудования [60, 139].

2.6. Управление режимами крупных энергосистем

Проблемы оперативного управления крупными энергосистемами или системами, образуемыми несколькими самостоятельными электроэнергетическими компаниями, исследовались с помощью тесно связанных между собой методов диакоптики и декомпозиции [15, 35, 91, 145].

В области диакоптики основную роль сыграли работы Крона по разработке алгоритмов оперативного управления, которые используются в энергообъединении шт. Нью-Йорк. Применение этих алгоритмов обеспечивает экономически выгодный обмен электроэнергией внутри объединения, осуществляемый таким образом, что функции оперативного управления соответствующими генерирующими мощностями выполняют непосредственно местные электроэнергетические компании [4].

Разрабатываются и совершенствуются модели объединения энергосистем, в которых для построения требуемых топологических матриц используются методы решения транспортной задачи [131].

2.7. Календарное планирование ремонтных работ

Составление годового календарного плана ремонта энергоагрегатов, включая котлы, турбины и генераторы, заключается в определении сроков (на протяжении года), в течение которых каждый из агрегатов будет выведен из эксплуатации для проведения капитального или текущего ремонта. Для составления оптимального графика ремонта электроагрегатов в основном использовались методы целочисленного линейного программирования, динамического программирования и метод ветвей и границ. Прежде чем перейти к обсуждению методов нахождения оптимального решения, рассмотрим сущность проблемы.

Наиболее просто годовой календарный план ремонта энергоагрегатов системы можно представить в виде графика перевода соответствующих агрегатов в категорию резервных (рис. 9). При построении такого графика каждый агрегат характеризуется мощностью и количеством времени, необходимого для выполнения ремонтных работ. Для пояснения графика приводится табл. 2, согласно которой для ремонта энергоагрегата А мощностью 300 МВт требуется 1 неделя в год. Обычно длительность ремонтных

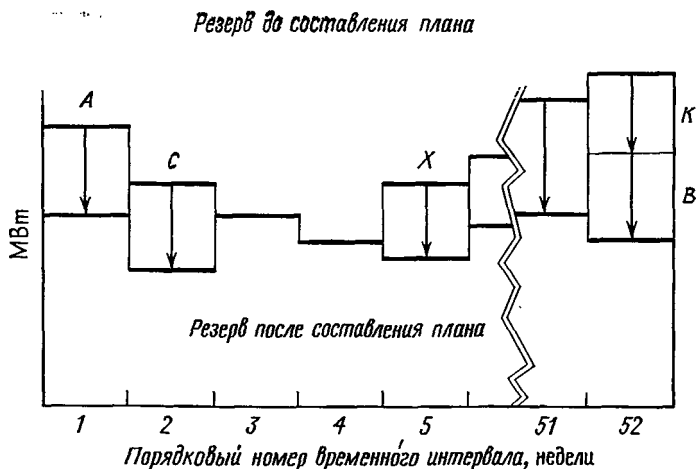


Рис. 9. Распределение резерва мощности в течение недель и график ремонта агрегатов. (На рисунке не показан агрегат Н, который следует за агрегатом X.)

Таблица 2

Энергоагрегаты, требующие ремонта

Индекс агрегата	Мощность, МВт	Продолжительность ремонта, нед.	Порядковый номер временного интервала (недели), соответствующий началу ремонта
A	300	1	Требуется определить
B	300	1	
C	100	6	
H	400	5	
K	250	1	
X	250	1	

работ изменяется в пределах от 1 до 8 недель, причем для большинства агрегатов этот срок превышает две недели. При построении модели год, на который составляется календарный план ремонта агрегатов, характеризуется величиной резерва, имеющегося в системе без учета вывода агрегатов в ремонт. Соответствующие данные приведены в табл. 3, из которой следует, например, что в первую неделю резерв мощности равен 1000 МВт, во вторую неделю 800 МВт и т. д.

Начало ремонта каждого из агрегатов и суммарная мощность агрегатов, выводимых в ремонт, для каждой недели должны

Таблица 3

Сроки ремонтных работ и резервные мощности энергосистемы, которые можно вывести в ремонт

Порядковый номер временного интервала (недели)	Установленная мощность МВт	Пиковая нагрузка, МВт	Резерв до составления календарного плана, МВт	Мощность агрегатов, находящихся в ремонте	Резерв мощности после составления календарного плана
1	4600	3600	1000	Требуется определить	Требуется поддерживать почти постоянным
2	4600	3800	800		
3	4600	3900	700		
4	4600	4000	600		
5	4600	3750	850		
.					
.					
.					
51	4800	3700	1100		
52	4800	3600	1200		

планироваться таким образом, чтобы остающийся резерв мощностей для двух соседних недельных интервалов был приблизительно одинаковым.

Требуемые ремонтные работы графически изображаются в виде ряда прямоугольников, которые располагаются последовательно (по оси ординат откладываются мощность в МВт, по оси абсцисс — продолжительность ремонта в неделях), что, по существу, и определяет резервные мощности для каждой недели рассматриваемого года (рис. 9).

Составление календарного плана ремонта затрудняется совокупностью практических ограничений. Количество агрегатов, которые можно ремонтировать одновременно, обычно ограничивается числом имеющихся ремонтных бригад. Когда некоторый агрегат выходит из строя, то другие агрегаты этой же электростанции обычно нельзя вывести в ремонт либо из-за отсутствия свободных ремонтных площадок, либо в интересах функционирования системы в целом, которые требуют, чтобы один из агрегатов этой станции находился в рабочем состоянии. Календарный план ремонта должен быть составлен так, чтобы стоимость выработки электроэнергии была минимальной. Кроме того, следует стремиться к максимизации надежности системы, а не величины резервных мощностей. Наибольшие осложнения связаны с неизбежным в практических условиях работы непрерывным изменением условий рассматриваемой задачи: величина нагрузки отличается от прогнозируемых значений, а энергоагрегаты в самый неожиданный момент выходят из строя и требуют продолжительного ремонта.

В настоящее время задача календарного планирования ремонтных работ решается, как правило, вручную или с помощью автоматизированной системы управления на базе эвристических методов расчета [33]. Предпринимались попытки решить эту задачу также с помощью методов линейного, целочисленного и динамического программирования [41]. Однако эти методы были признаны слишком громоздкими из-за большого числа переменных, наличия линейных и нелинейных ограничений и больших затрат машинного времени. С учетом того обстоятельства, что при оптимальном решении резервы мощности для всех недельных интервалов остаются приблизительно на одном и том же уровне, были разработаны эвристические методы составления календарного плана, основанные на поэтапной разработке графика ремонтных работ, при которой для каждого этапа соблюдаются ограничения, накладываемые на численность персонала, а также ряд других ограничений и выдерживается постоянство резервных мощностей.

При использовании эвристических процедур поиска наилучшего варианта плана для наиболее крупных электроэнергетических

компаний пришлось иметь дело с многими видами ограничений, связанных, например, с распределением ремонтных бригад и оборудования, необходимостью оставлять на некоторых станциях агрегаты в рабочем состоянии, предельными нагрузками линий электропередачи со сроками, прошедшими со времени проведения предыдущего ремонта агрегатов, и техническими правилами ревизии котельных установок [42]. Исследования вопросов оптимизации графика ремонта агрегатов привели к разработке машинной программы, используемой инженерно-техническими работниками для корректировки графиков в соответствии с новыми, изменившимися условиями и для оценки возможностей изменения имеющихся ограничений за счет сверхурочных работ, использования большего количества ремонтных бригад и некоторых других факторов.

Ряд практически важных аспектов рассматриваемой задачи был учтен в модели, предложенной авторами работы [61], причем сформулированная задача решалась с помощью метода ветвей и границ. Для описания системы авторы выбрали такую целевую функцию (функцию затрат), минимизация которой означала составление графика, согласно которому ремонт каждого агрегата проводился в наиболее удобный период, причем резервные мощности превышали заданный минимальный уровень. Если, например, некоторый агрегат ремонтировался в декабре предыдущего года, то в плане на следующий год его ремонт предусматривался также на это же время. Второй аспект проблемы был связан с изменением календарного плана при вынужденном выходе некоторого агрегата в незапланированный ремонт. В этом случае ставилась задача отыскания такого графика ремонтных работ, при котором необходимые изменения были бы минимальны и сдвиг ранее принятых сроков начала ремонта агрегатов был бы как можно меньше. Например, если выход из строя агрегата K (рис. 4) произойдет на 4-й неделе, то потребуются начать его ремонт на 5-й неделе вместо 52-й. В такой ситуации нельзя просто поменять местами агрегаты X и K , так как это повлечет за собой пересмотр графика ремонта для 47 недель, что связано со слишком большой перестройкой плана. Для минимизации перемещений необходимо, чтобы начало ремонта агрегата X в скорректированном графике было смещено по отношению к первоначальному сроку (т. е. относительно 5-й недели) на минимальный отрезок времени.

В ходе неофициального обмена мнениями выяснилось, что некоторые электросиловые компании при разработке машинных программ, предназначенных для составления календарного плана, основное внимание уделяют возможности введения ограничений и использования простых процедур поиска возможных вариантов решения, удовлетворяющих требованию по-

стоянства резервных мощностей. Большая часть применяемых программ была разработана специалистами-пользователями, которые, естественно, хотели бы использовать более совершенные математические методы. Однако они пришли к заключению, что гибкость программы в отношении вводимых ограничений является более важным фактором, чем собственно математические процедуры, используемые для получения решения.

По-видимому, дальнейшее усовершенствование таких программ связано с поиском способов минимизации затрат на производство при соблюдении всех имеющихся ограничений. Степень риска недоотпуска электроэнергии может быть уменьшена путем эквивалентного изменения нагрузки, что дает возможность не рассчитывать величину резервной мощности на каждом недельном интервале планового периода [83]. Пока еще не найден легко определяемый показатель, эквивалентный стоимости выработки электроэнергии. Наиболее важная характеристика качества календарного плана — это его приспособляемость к изменяющимся условиям, например уровню нагрузки или выходам агрегатов в незапланированный ремонт. Формализация этой характеристики при построении модели остается наиболее трудной проблемой.

2.8. Другие задачи управления

Кроме перечисленных выше следует упомянуть ряд работ в области управления, результаты которых в настоящее время применяются недостаточно широко. Эти работы посвящены задачам определения необходимой избыточной мощности (вращающегося резерва) по отношению к ожидаемой нагрузке, анализ которых проводился с помощью методов теории вероятностей и моделей типа марковских [7, 20, 32]. Планирование выработки электроэнергии на ГАЭС рассматривалось в рамках задач линейного и динамического программирования [3, 43]. На примерах задач небольшой размерности было продемонстрировано применение методов теории оптимального управления, однако трудности, связанные с измерением всех переменных, характеризующих текущее состояние системы, существенно ограничивают применение таких методов [14, 72, 172].

Изменение схемы соединений линий электропередачи в целях более экономичного распределения нагрузки также связано с задачами оперативного управления режимами работы электроэнергетической системы. Решение такой задачи с использованием метода ветвей и границ рассмотрено в работах [47, 154].

2.9. Литература для ознакомления с другими вопросами оперативно-диспетчерского управления

В указанных ниже работах дается обзор широкого круга проблем, касающихся оперативного управления электроэнергетическими системами. В работе [76] излагаются требования к вычислительной системе, предназначенной для автоматизированного центра оперативного управления (без детализации технических аспектов). В работе [115] на основе анализа функционирования дается обзор задач оперативного управления, в котором содержится перечень функций, выполняемых с помощью автоматизированной системы оперативного управления, разработанной для компании Cleveland Electric Illuminating, а проблемы, с которыми пришлось столкнуться при ее внедрении, приводятся в работе [64]. Текущее планирование работы и функционирование объединения энергосистем рассматриваются на некоторых примерах в работе [30]. Материал данной работы базируется на практическом опыте совета по надежности энергосистем центрального и восточного районов США, координационного совета энергосистем запада США и опыте работы энергосистемы Ontario Hydro.

3. ПЛАНИРОВАНИЕ УВЕЛИЧЕНИЯ ПРОИЗВОДСТВА ЭЛЕКТРОЭНЕРГИИ

В области электроэнергетики основная доля капитальных затрат приходится на строительство электростанций, линий электропередач и распределительных сетей. Поэтому именно этим объектам уделяется основное внимание при анализе энергетических систем. Следует заметить, что даже ввод одного турбогенератора (а это может быть единственным мероприятием, требующим капитальных затрат, выделенных компанией на развитие мощностей в течение одного или нескольких лет) требует тщательного планирования, предусматривающего определение сроков ввода турбогенератора, его мощности, вида топлива и места установки.

В разработке процессов планирования увеличения производства электроэнергии важную роль сыграли методы теории вероятностей, которые также эффективно использовались при выборе уровня резервной мощности в случае перехода от изолированных электроэнергетических систем с малой единичной мощностью агрегатов порядка 100 МВт к объединенным энергосистемам, в которых единичная мощность агрегатов превышает 1000 МВт. В последнее время аппарат теории вероятностей нашел применение и в моделировании процесса выработки электроэнергии с учетом случайного характера выхода из строя агрегатов, и, таким образом, отпала необходимость использования дорогостоящего вычислительного метода Монте-Карло.

Что касается планирования развития сетей электропередачи (разд. 5), то в этом случае методы теории вероятностей не получили широкого распространения. Для решения таких задач дос-

таточно эффективными оказались методы линейного программирования и использования теоремы о максимальном потоке и минимальном разрезе.

3.1. Планирование резерва генерирующих мощностей

Если складывается такая ситуация, что потребление энергии превышает существующие возможности электроснабжения, то свет гаснет не только у человека, который последним повернул выключатель, но и у всех потребителей в зоне действия плавкого предохранителя или прерывателя питания данного участка распределительной сети. (Интересно отметить, что подобная ситуация прямо противоположна той, которая имеет место, когда в телефонной трубке слышится сигнал «занято», т. е. когда вызов, перегружающий телефонную сеть, не влияет на предыдущего пользователя.) Поскольку перегрузка или дефицит мощности приводят к отключению потребителей на некотором участке сети, то для электроэнергетической отрасли особенно важно предусмотреть резерв мощностей сверх номинального уровня нагрузки. Принятие решений, касающихся капиталовложений, нельзя откладывать до того момента, когда имеющееся оборудование окажется полностью загруженным, так как дефицит энергии приводит к нарушению пропорциональности развития экономики, а время упреждения, т. е. время между принятием решения о строительстве электростанций и вводом их в эксплуатацию, составляет от двух до десяти лет. Основная часть капиталовложений, которые в 1975 г. для электростанций мощностью 1 млн. кВт достигали 1000 долл./кВт, должна быть освоена еще до того, как материализуется ожидаемый прирост нагрузки. Основное внимание при планировании резерва генерирующих мощностей фокусируется именно на этих трех факторах: величине требуемых капитальных затрат, длительности процесса ввода мощностей и нежелательных последствиях дефицита электроэнергии.

Необходимость ввода новых мощностей обычно возникает тогда, когда имеющиеся мощности не могут обеспечить соответствующего резерва при росте нагрузки до ее прогнозируемых значений. В большинстве случаев для определения требуемого уровня резерва используются методы теории вероятностей, и в частности метод оценки вероятности необеспеченности нагрузки. Этот метод позволяет, например, определить, что для надежного энергоснабжения в соответствии с заданным годовым графиком нагрузки некоторая система общей мощностью 4600 МВт, включающая 32 турбогенератора, должна иметь резерв установленной мощности, равный 600 МВт. Это в свою очередь означает, что если прогнозируемая величина нагрузки достигает 4000 МВт, то в систему должен быть введен дополнительный энергоагрегат.

Метод оценки вероятности необеспеченности нагрузки явился результатом работ, начатых еще в 1946 г. [2], и состоит в определении вероятности того, что установленная мощность окажется недостаточной для покрытия некоторой заданной нагрузки.

Пусть на станции установлены два турбогенератора, каждый мощностью 50 МВт, и один турбогенератор мощностью 100 МВт. Определим вероятность необеспеченности нагрузки, равной, например, 130 МВт. Допустим, что каждый агрегат может находиться в одном из двух состояний — рабочем или аварийном — и что вероятность нерабочего состояния агрегатов мощностью 50 МВт равна 0,02, а агрегата мощностью 100 МВт — 0,03. На основе этих исходных данных можно рассчитать точную и кумулятивную вероятность выхода из строя определенной части установленной мощности системы (табл. 4). Например, выход из строя мощности,

Таблица 4

Значения вероятностей состояния системы¹⁾ из трех агрегатов

Состояние системы	Величина мощности, находящейся в аварийном состоянии, МВт	Вероятность состояния системы	Кумулятивная вероятность выхода из строя данной или большей мощности
1	0	0,931588	1,000000
2	50	0,038024	0,068412
3	100	0,029200	0,030388
4	150	0,001166	0,001188
5	200	0,000012	0,000012

¹⁾ Система состоит из двух агрегатов, каждый мощностью 50 МВт, причем вероятность аварийного состояния каждого из них равна 0,02, и одного агрегата мощностью 100 МВт с вероятностью аварийного состояния, равной 0,03.

равной 100 МВт, может иметь место, если в аварийном состоянии окажутся оба турбогенератора мощностью 50 МВт или турбогенератор мощностью 100 МВт. Вероятность первого состояния системы равна $(0,02) \cdot (0,02) \cdot (0,97)$, а второго — $(0,98) \cdot (0,98) \cdot (0,03)$, что дает вероятность выхода из строя мощности 100 МВт, равную 0,029200. Кумулятивные вероятности, значения которых приведены в табл. 4, можно рассчитать с помощью рекурсивных соотношений, добавляя в систему по одному агрегату [20], обходя все пронумерованные состояния системы и последовательно вычисляя соответствующие вероятности.

Определим теперь предельную величину мощности, при выходе которой из строя рассматриваемая система еще окажется способной покрыть нагрузку, равную 130 МВт.

Система не в состоянии обеспечить заданную нагрузку, если вышедшие из строя мощности больше, чем резерв установленной мощности. Так как в данном случае этот резерв равен $200 - 130 = 70$ МВт, то необходимо определить вероятность выхода из строя мощности, большей чем 70 МВт. В соответствии с результатами расчетов, представленными в табл. 4, эта вероятность равна 0,030388. Если в этом случае из строя выйдет мощность, равная или превышающая 100 МВт, система окажется не в состоянии обеспечить заданную нагрузку, и вероятность необеспеченности нагрузки в течение суток с пиковым значением нагрузки 130 МВт будет равна 0,030388.

С помощью рассмотренного метода проверяется ряд значений нагрузки, характеризующих данный год. В ряде модификаций этого метода нагрузки определяются для каждого часа (8760 значений), в других — для каждого суток (365 значений), а в некоторых — для каждого рабочего дня (260 значений). Независимо от числа нагрузок, характеризующих данный год, вероятности необеспеченности этих нагрузок суммируются, что в результате дает ожидаемое за год число суток с дефицитом мощности. Например, если проанализировать возможности рассмотренной выше системы для покрытия трех значений нагрузки 110, 130 и 90 МВт, то получим результаты, представленные в табл. 5.

Таблица 5

Значения вероятностей необеспеченности нагрузки для системы мощностью 200 МВт

Порядковый номер временного интервала (суток)	Нагрузка, МВт	Резерв, МВт	Вероятность дефицита мощности
1	110	90	0,030388
2	130	70	0,030388
3	90	110	0,001188

Ожидаемое число суток с дефицитом мощности — 0,061964 суток

Таким образом, метод оценки вероятности необеспеченности нагрузки предусматривает, во-первых, перечисление всех возможных состояний системы с выведенными из строя мощностями и расчет кумулятивной вероятности выхода из строя данной или большей мощности и, во-вторых, суммирование вероятностей выхода из строя резерва установленных мощностей для списка заданных значений нагрузки. Результирующее значение, которое в рассматриваемом примере равно 0,061964 суток, называют «вероятностью необеспеченности нагрузки» (неправильный тер-

ми, так как это не вероятность, а ожидаемое значение), степенью риска или ожидаемой потребностью в дополнительной мощности. Однако ни один из этих терминов не является общепринятым.

При планировании резерва установленных мощностей предпринимались попытки использовать метод Монте-Карло, особенно популярный в конце 50-х годов. Поскольку, однако, его применение сопряжено с очень большими затратами машинного времени, то в настоящее время широкое применение находят рекурсивные методы [117].

Посмотря на успешное применение методов теории вероятностей при планировании резерва генерирующих мощностей как в США [73], так и в европейских странах [37], потребовалось почти 20 лет, чтобы планирование резерва с помощью таких методов получило широкое признание. Это объясняется главным образом тем, что те, кто принимает решение, не представляют себе, как следует интерпретировать величину 0,06 суток/год и как оценивать этот результат, т. е. является ли он хорошим или неудовлетворительным. Понимание и трактовка этой новой вероятностной меры необеспеченности нагрузки основывались на сравнении полученного результата для будущего состояния системы с результатом, характеризующим состояние системы в прошлом. Так, например, могло оказаться, что значение 0,06 суток/год — это точно такая же величина степени риска, как и в 1968 г., по меньше, чем степень риска для состояния системы в 1967 и 1970 гг. В этом случае полученный результат свидетельствовал о том, что будущее состояние системы в отношении возможного дефицита мощности аналогично состоянию системы в 1968 г., но лучше, чем в 1967 и 1970 гг.

Правильной трактовке рассматриваемого показателя как новой вероятностной характеристики состояния системы содействовал анализ его чувствительности к параметрам системы. Рассчитав величину степени риска в течение года для различных уровней нагрузки и построив соответствующую зависимость (рис. 10), специалист-плановик мог установить, насколько существенно влияет ввод нового агрегата на способность системы покрыть заданную нагрузку. Так, например, если добавить к системе, имеющей мощность 4600 МВт, первый блок мощностью 600 МВт, то при сохранении прежней степени риска нагрузка может быть увеличена только на 362 МВт. Однако при добавлении пятого блока мощностью 600 МВт степень риска останется неизменной при увеличении нагрузки уже на 508 МВт. Определение рассмотренных нелинейных соотношений между мощностью вводимого агрегата, вероятностью выхода его из строя, мощностью системы и изменением нагрузки как раз и является наиболее важным направлением использования методов теории вероятностей в планировании увеличения производства электроэнергии. Метод опре-

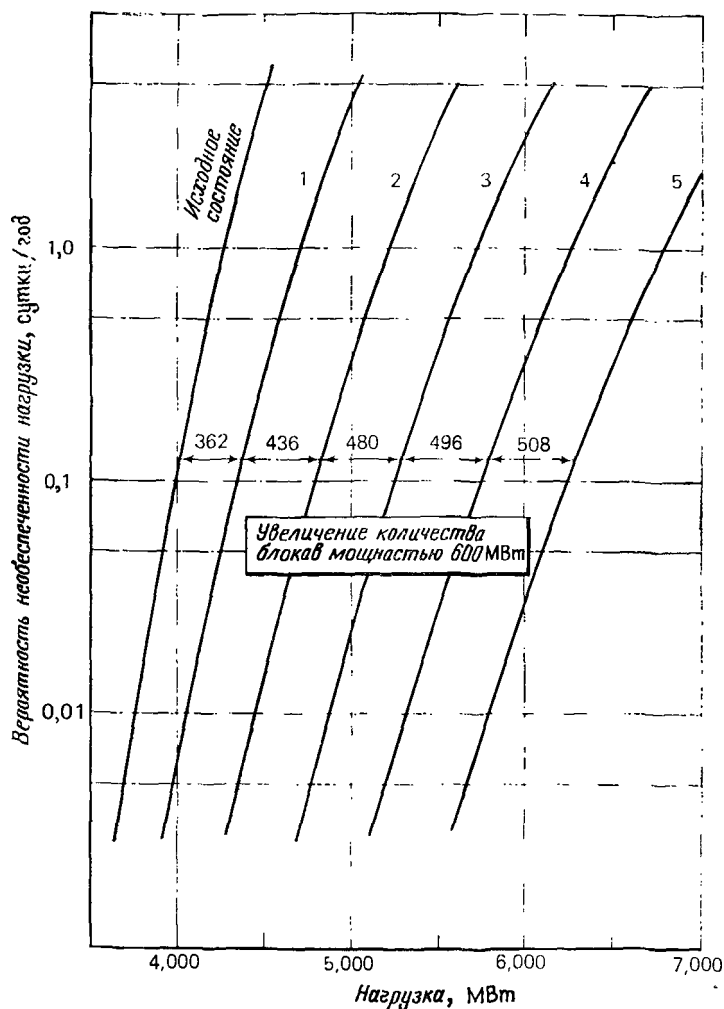


Рис. 10. Зависимость вероятности необеспеченности нагрузки от величины годового пика нагрузки для систем с различным количеством энергоблоков мощностью 600 МВт.

деления этих соотношений, основанный на предположении о том, что зависимость степени риска от нагрузки близка к линейной (рис. 10), предложен в работе [81].

Один из важных результатов применения метода оценки вероятности непокрытия нагрузки состоял в том, что полученные с его

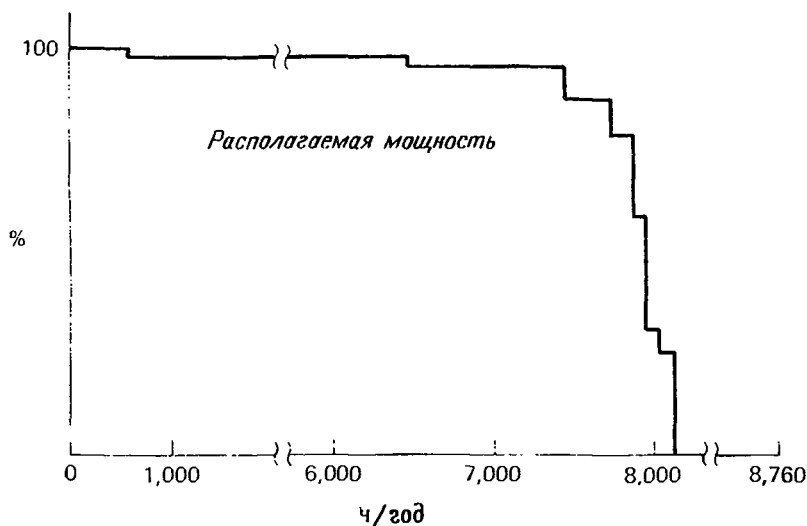


Рис. 11. Кривая длительности нагрузок турбоагрегата.

помощью данные свидетельствовали о целесообразности дальнейшего увеличения единичной мощности агрегатов [109], в особенности при объединении энергосистем [110].

Действительно, в 60-х годах по мере объединения энергосистем единичные мощности агрегатов стали быстро возрастать и достигли величины 1300 МВт, что намного больше той мощности, которая прогнозировалась в 1955 г.

В работе [44] предложен метод анализа изменения величины необходимого резерва при объединении двух энергосистем. Аналогичный метод использовался в ряде исследований, посвященных выбору наилучшего варианта объединения двух энергосистем на основе сопоставления экономии от уменьшения резерва и дополнительных капитальных затрат на развитие линий электропередачи [53, 110, 133]. Результаты всех этих работ свидетельствуют о возможности значительного уменьшения резерва при объединении двух энергосистем приблизительно равной мощности. Попытки моделировать объединение более чем двух энергосистем и ввести ограничения на нагрузку линий электропередачи дали весьма скромные результаты в виде программ, не требующих чрезмерных затрат машинного времени. В работе [157] сообщается об использовании машинной программы, разработанной применительно к объединению по схеме «звезда» трех энергосистем (Новой Англии, шт. Нью-Йорк и Онтарио). С увеличением количества объединяемых энергосистем сложность рассматриваемой задачи резко возрастает. Это обстоятельство отметили, в част-

ности, авторы работы [168], занимавшиеся разработкой методов решения этой задачи применительно к объединенной энергосистеме шт. Огайо и соседних штатов.

Использование методов теории вероятностей при планировании резерва мощностей нередко осложняется отсутствием необходимых исходных данных. Так, все применяемые методы требуют в качестве исходных данных список мощностей турбоагрегатов и вероятностей их вынужденного отключения (называемых в электроэнергетике коэффициентами аварийности), а также данных о продолжительности планового ремонта. Поскольку, однако, располагаемая мощность турбоагрегата может существенно меняться (рис. 11), исследователи вынуждены решать ряд дополнительных задач. Например, следует ли считать в качестве располагаемой мощности агрегата ту максимальную мощность, на которую он был выведен когда-либо в течение года; где граница «располагаемой мощности» агрегата — 100, 80 или 50% его номинальной мощности; как разделить общее время простоя на две составляющие: время «планового» простоя, т. е. длительность выполнения запланированного ремонта, и время «вынужденного простоя». Проблемы сбора данных о мощности, продолжительности вынужденных и запланированных простоев и соответствующей обработки этих данных, получаемых от электроэнергетических компаний, обсуждаются в работе [123].

Модель, в которой используется располагаемая мощность агрегатов, можно расширить, если учесть возможность нахождения агрегатов не только в трех указанных выше состояниях (т. е. в состоянии готовности, в плановом ремонте и в состоянии аварийного отключения). В таких моделях можно рассматривать и состояние частичной (неполной) готовности [45]. Однако эти модели не получили широкого распространения, поскольку их использование требовало увеличения объема исходных данных и времени расчетов.

Модель, базирующаяся на использовании располагаемой мощности агрегата, должна включать в качестве переменной время, так как практический опыт показывает, что в первые годы эксплуатации продолжительность аварийных отключений новых агрегатов в 2—3 раза больше, чем уже освоенных агрегатов.

Помимо возможного количества дней работы с дефицитом мощности планировку электроэнергетической компании необходимо знать, как часто система работает в режиме дефицита мощности и длительность таких режимов. В связи с этим можно привести следующий пример. Когда диспетчеру сообщили, что управляемая им энергосистема была спроектирована исходя из степени риска, равной 0,1 суток/год, то диспетчер, сделав в уме ряд подсчетов, заявил, что ежемесячный 12-минутный интервал с дефицитом мощности — это неприемлемая характеристика системы. Опи-

бочность сделанного вывода была обусловлена попыткой определить частоту и длительность нежелательных режимов только на основе средних значений.

Показатели частоты и продолжительности режимов с дефицитом мощности можно рассчитать с помощью моделей типа марковских или с помощью модели суточной нагрузки, характеризующей двумя состояниями [23]. Введение второй рекурсивной процедуры в машинную программу, составленную с помощью метода оценки вероятности перекрытия нагрузки, позволяет определить степень риска (сутки/год), а также интервал времени между двумя последовательными дефицитами (год) и длительность (ч) режима работы с дефицитом мощности. Степень риска 0,1 суток/год обычно соответствует 7—8-летнему интервалу между наступлением режимов с дефицитом мощности, а их продолжительность составляет от 4 до 6 часов. Эта дополнительная информация существенно не меняет требования к величине необходимого резерва, и поэтому такие расчеты пока еще не находят широкого применения.

Планирование резерва установленных мощностей с помощью методов теории вероятностей развивается в настоящее время в следующих направлениях: учет неопределенностей, касающихся времени установки агрегатов и наличия топлива [23], распределение ответственности за производство электроэнергии между членами энергетического объединения [71], моделирование режимов работы гидроагрегатов, которым в большей степени присущи ограничения по выработке энергии, а не по мощности [86].

Применение этих методов позволяет определить необходимый для системы резерв установленной мощности, а следовательно, и сроки ввода новых генерирующих мощностей. Единичная мощность агрегатов и вид топлива определяются исходя из анализа взаимного влияния капитальных затрат и стоимости производства электроэнергии.

3.2. Сравнительный экономический анализ возможных вариантов решений

Если в результате анализа резерва установленной мощности выявлена необходимость ввода дополнительных генерирующих мощностей, то проводится сравнение возможных вариантов решений относительно выбора единичной мощности агрегатов и вида топлива по величине капитальных затрат и затрат на производство энергии. Сравнительный анализ затрат выполняется, как правило, в рамках технико-экономических расчетов с учетом дисконтированного норматива годовых доходов [102; 103].

В последнее время стали использоваться комплексные модели, более детально описывающие влияние инвестиционных решений на прибыль и движение денежной наличности. Одной из первых

разработку такой модели начала компания Boston Edison с целью оценки возможных вариантов решения относительно ввода новых мощностей и соответствующих финансовых планов [34]. Модель, имитирующая учетно-финансовую деятельность компании, позволяет оценить доходы и наметить систему регулирующих воздействий, необходимых для обеспечения приемлемой величины прибыли. В случае равенства доходов по сравниваемым вариантам программа выводит на печать расчетные характеристики некоторого конкретного плана. В библиографическом сборнике [102] даются ссылки на 26 комплексных моделей, причем четыре из них разработаны непосредственно для нужд компаний Public Service Electric and Gas, шт. Нью-Джерси, Long Island Lighting, Boston Edison и Victoria, Австралия.

Следует отметить, что выгоду от разработки комплексных моделей получают не только электроэнергетические компании. Представляя к публикации одну из статей, трое сотрудников компании сообщили о своем продвижении по службе, которое явилось следствием их участия в разработке такой модели. Работа над созданием модели помогла им найти конструктивные методы изучения деятельности компании, а приобретенные знания открыли новые возможности для продвижения по службе.

Затраты на производство, представляющие собой основную составляющую затрат, рассматриваемых при решении вопросов о единичной мощности агрегата и виде топлива, рассчитываются с использованием детерминированных и вероятностных моделей. Детерминированные модели, в которых для каждого года задан почасовой график нагрузки и не учитывается случайный характер вынужденных отключений агрегатов, дублируют традиционные процедуры оперативного управления энергосистемой, например распределение нагрузки по принципу равенства дополнительных затрат, ввод агрегатов в соответствии со списком приоритетов, планирование режимов работы ГЭС с водохранилищами и ГЭС, работающих на бытовом стоке, оптимизацию режимов ГАЭС [10, 79].

В настоящее время методы линейного программирования используются для прогнозирования затрат на производство электроэнергии с помощью программ для ЭВМ, всесторонне описывающих процесс планирования развития энергосистемы [55, 56]. Формулировка таких задач в виде задач линейного программирования учитывает затраты, связанные с пуском, эксплуатацией в рабочих режимах и ремонтом агрегатов, а также затраты, обусловленные обменом энергией с другими системами. При этом вводятся ограничения на требуемые значения вращающегося резерва, нагрузки, готовности агрегатов, мощности и выработки электроэнергии ГЭС, а также ограничения на нагрузки межсистемных линий электропередачи. С точки зрения затрат машинного времени эти

программы были вполне приемлемыми, однако рассчитанные по ним затраты средств оказывались постоянно заниженными на 7% по сравнению с результатами расчетов по программам оперативно-диспетчерского управления [55]. Тем не менее, когда интерес представляет относительное изменение величины затрат (что характерно для выбора варианта плана), а не абсолютная величина, то применение таких расчетных методов, как показал опыт, вполне оправдано.

Хотя основной результат моделирования заключается в нахождении общей величины затрат, не менее важно при этом определить и режимы работы отдельных агрегатов. В этом отношении детерминированным моделям свойственно систематическое занижение планируемого времени работы менее мощных и реже используемых «пиковых» агрегатов, а также вырабатываемой ими энергии, что постоянно обнаруживается при сравнении расчетных и фактических данных. Это связано с тем, что в таких моделях нельзя учесть случайные аварийные отключения более мощных агрегатов.

В конце 50-х гг. были предприняты попытки учесть аварийные отключения и тем самым получить более точные оценки выработки электроэнергии каждым агрегатом. Для этого предполагалось использовать имитационные модели, предусматривающие применение метода Монте-Карло. Однако выяснилось, что практическая реализация этих методов сопряжена с очень большими затратами, и поэтому применение детерминированных подходов осталось в электроэнергетике общепринятым методом исследования рассматриваемого класса задач [6].

Новый метод учета вынужденных простоев агрегатов при оценке затрат на производство, который в настоящее время уже получает распространение, состоит в «свертывании» кривой продолжительности нагрузок с помощью графика простоев генерирующих мощностей, что позволяет рассчитать ожидаемые значения выработки энергии и продолжительность работы каждого турбогенератора.

Впервые идея этого метода была высказана в работе [12] и впоследствии развита в работах [28, 148]. Создание этого метода привело к появлению самых разнообразных в своем практическом воплощении способов расчета стоимости производства электроэнергии. Электроэнергетические компании начали с использования кривых продолжительности нагрузок для расчетов затрат вручную, а затем, по мере появления все более мощных ЭВМ, стали разрабатывать детальные модели почасового оперативно-диспетчерского управления. В настоящее время кривые продолжительности нагрузок снова находят применение, но уже как составной элемент расчетов, базирующихся на использовании вероятностных методов. Рассматриваемые временные интервалы для

каждой кривой стали более короткими (полусуточными, а не годовыми), а рекурсивное уравнение метода оценки вероятности необеспеченности нагрузки используется для «свертывания» нагрузок, остающихся по мере рассмотрения каждого агрегата.

Использование методов теории вероятностей при планировании резерва мощностей одновременно с методами определения затрат на производство электроэнергии оказалось полезным и в том отношении, что и ожидаемая величина затрат, и показатель надежности, т. е. ожидаемая величина дефицита энергии, рассчитываются в этом случае одновременно. Значения ожидаемого времени использования отдельных агрегатов, рассчитываемые с помощью вероятностных моделей, отличаются от фактических значений обычно не более чем на 5—15% [148].

После того как для последовательности лет рассматриваемого планового периода рассчитаны затраты на производство электроэнергии, их величина совместно с нормативным значением ежегодной прибыли от капиталовложений на развитие генерирующих мощностей используется в комплексной модели или в расчетах «дисконтированных» затрат. По величине этих общих затрат и производится выбор вида топлива, мощности и типа нового турбоагрегата.

3.3. Модели планирования производства электроэнергии

При планировании производства электроэнергии обычно решаются следующие три основные задачи: планирование резерва мощностей, определение затрат на производство электроэнергии и определение капитальных затрат. Для решения этих задач было разработано несколько типов моделей. Первые модели, в которых использовались методы динамического программирования, требовали больших затрат, связанных с выполнением машинных расчетов. Этот недостаток был присущ даже упрощенным моделям, предназначенным для планирования резерва мощностей или затрат на производство электроэнергии [49]. Хотя такие программы и обеспечивали принципиальную возможность выбора наилучшего из нескольких тысяч возможных вариантов решений относительно увеличения мощностей энергосистем, их практическая реализация при исследовании всех возможных комбинаций ввода агрегатов оказывалась слишком дорогостоящей. Во всяком случае, о практическом использовании таких моделей при решении рассматриваемых задач планирования ничего не сообщалось. Для уменьшения размерности соответствующей задачи динамического программирования в работе [28] предложен метод, расчетная схема которого представлена в виде цикла с разомкнутой обратной связью, однако результаты применения этого метода не приводятся.

В энергосистеме Electricité de France методы линейного и нелинейного программирования для целей планирования развития производства начали использоваться с 1954 г. Работа [16] содержит обзор таких методов и результаты исследования, посвященного вопросам развития системы Electricité de France (на период с 1965 по 1985 г.), и в частности выбору оптимального решения относительно капиталовложений в АЭС, ГЭС и тепловые станции с паровыми и газовыми турбинами и размещения этих электростанций (из 6 возможных вариантов). Формулировка рассматриваемых задач в виде задач целинейного программирования, позволяющая учесть затраты, связанные с выходом агрегатов из строя, обсуждается довольно часто и редко используется практически [156]. Следует заметить, что введение этого показателя слабо влияет на результаты оптимизации. В указанной публикации обосновывается также необходимость разработки нескольких плановых решений, каждое из которых базировалось бы на определенной совокупности важнейших допущений и предположений относительно будущей ситуации. После этого нужно проверить, как влияют на результаты решения плановой задачи принятые допущения оптимистического и пессимистического характера.

Как и многие другие специалисты, автор работы [16] приходит к заключению, что уже вскоре после завершения исследований принятые допущения и предположения оказываются не соответствующими действительному положению дел. Так, в системе Electricite de France фактическая величина нагрузки оказалась выше прогнозируемой на 8%. Нужные расчеты по имеющейся сложной модели можно было бы, конечно, провести вновь, но скорее всего необходимые для этого время и трудозатраты оказались бы неоправданными. Отмеченное обстоятельство указывает еще на одну важную задачу операционного анализа, которая состоит в том, чтобы предусмотреть для функционирующей системы разработку моделей планирования, ориентированных на изучение вариантов и допущений, касающихся будущих условий работы системы.

Появление своего рода «книги заказов» на программно-математическое обеспечение задач планирования было обусловлено претенциозным исследованием, посвященным планированию мощности АЭС [85]. Решение сформулированных в ней задач требует разработки комплекса математических моделей, включающего модель развития системы на предстоящие 30 лет, модель среднесрочного планирования на 3—5 лет, модель годового планирования и модель оперативно-диспетчерского управления. Для получения результатов с помощью такой модели многоэтапного процесса принятия решений авторы работы [85] предполагали использовать методы линейного и динамического программирова-

ния. Однако к 1973 г. результаты этих исследований все еще не были опубликованы.

В работе [5] использована упрощенная модель планирования, укладывающаяся в рамки линейных моделей, причем автор исследует проблемы планирования в том аспекте, который не характерен для электроэнергетики США. В этой работе рассматриваются системы, включающие АЭС, ТЭС и ГЭС, при этом учитываются межсистемные связи и размещение электростанций. Для упрощения расчетов затрат на производство используются зависимости распределения нагрузок в течение года, что равносильно предположению о том, что на затраты влияет только величина нагрузки. Однако такое предположение справедливо только при неизменном графике использования установленных агрегатов. Так как на самом деле для рабочих и нерабочих дней и даже для дневных и ночных часов в течение суток эти графики различны, то в более точных моделях для расчета затрат вместо одной зависимости распределения нагрузок в течение года используются четыре или большее число распределений нагрузок для недельного интервала. Требуемый резерв установленной мощности учитывается в модели в виде некоторой дополнительной составляющей, устанавливаемой в процентном отношении к величине нагрузки. С точки зрения реализации оптимизационного метода это допущение практически оправдано, однако для многих специалистов в области планирования развития энергосистем оно неприемлемо.

Более детальная модель линейного программирования для оптимизации планов производства электроэнергии была сформулирована в работе [56]. Для минимизации дисконтированной величины капитальных затрат и затрат на производство электроэнергии при ограничениях на величину нагрузки и резерв мощности была использована стандартная машинная программа линейного программирования. Наиболее важное достоинство предложенной модели заключается в том, что по сравнению с более сложными моделирующими программами ее реализация обеспечивает заметное уменьшение времени расчетов и трудозатрат.

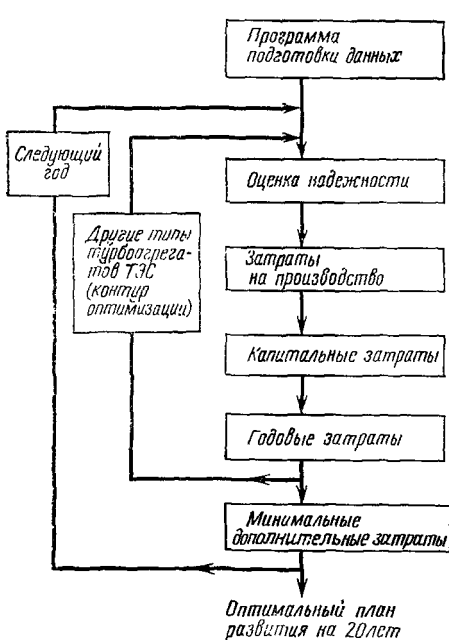
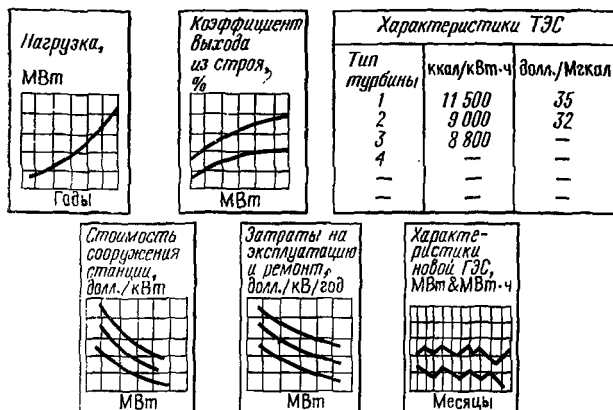
Один из главных факторов, препятствующих применению методов линейного динамического программирования, — это чувствительность решения к усложнению модели. Глобальный оптимум, который находится с помощью упрощенной модели, оказывается неоптимальным решением задачи при совместном использовании такой модели с вероятностной моделью определения необходимого резерва мощностей и затрат на производство. Однако вероятностные модели слишком громоздки и требуют больших затрат при использовании методов динамического программирования. Что касается линейного программирования, то их применение невозможно из-за нелинейности таких моделей.

Альтернативой применению оптимизационных методов, базирующихся на упрощенных моделях, является использование достаточно детализированных моделей и эвристических методов, ориентированных на нахождение решений, близких к оптимальным. В работе [1] дается описание одного из исследований по планированию развития электроэнергетических систем, выполненного с помощью такой модели (рис. 12). Входные переменные модели включают данные о росте нагрузки, характеристики генераторов (например, коэффициенты вынужденных простоев для старых и новых агрегатов, расходы тепла, затраты на топливо), стоимость сооружения электростанций, эксплуатационные расходы и затраты на ремонт, характеристики ГЭС. Заданными при проведении исследований считаются: различные типы генерирующих агрегатов, выбор и сочетание которых подлежат оптимизации; правила, в соответствии с которыми устанавливается величина вращающегося резерва; показатель надежности (сут/год), задаваемый при планировании резерва установленной мощности; относительный объем постоянных затрат при исчислении объема капиталовложений; дисконтированная норма процента на инвестированный капитал, стратегия замены оборудования, предусматривающая его демонтаж после X лет эксплуатации, и инфляционные факторы, учитываемые при расчетах стоимости сооружения электростанций и затрат на топливо.

Во многих практических случаях наиболее экономичный план увеличения генерирующих мощностей системы представляет собой последовательность отдельных годовых планов наращивания мощности, характеризующихся минимальными затратами в течение года при вводе этих дополнительных мощностей в эксплуатацию. В таких ситуациях вместо оптимизационных методов, сложность применения которых обусловлена большими затратами на проведение необходимых расчетов, оказывается возможным использовать простые эвристические процедуры нахождения последовательности наиболее экономичных годовых планов; эти эвристические методы модифицируются таким образом, чтобы можно было учесть предполагаемые в будущем изменения стоимости топлива [124].

Любой «оптимальный» план увеличения производства электроэнергии может оказаться субоптимальным или недопустимым, если проанализировать полученное решение с помощью более точных моделей, например комплексных моделей (учитывающих финансовые ограничения), моделей, описывающих передачу электроэнергии, в которых учитывается ряд дополнительных составляющих затрат, или моделей, характеризующих состояние окружающей среды (указывающих на неприемлемость использования тех или иных природных ресурсов или на чрезмерное загрязнение, обусловленное функционированием энергосистемы). Таким

Входные переменные



Лента с информацией о территории

1. Характеристики нагрузки
2. Характеристики и стоимостные показатели агрегатов существующих ГЭС и ТЭС

Исследуемые факторы

- Оптимизация -----
- Вращающийся резерв -----
- Степень риска -----
- Постоянные затраты -----
- Дисконтированная норма процента на капитал -----
- Выбытие оборудования -----
- Инфляция -----

Рис. 12. Схема планирования увеличения производства электроэнергии.

образом, сложность построения действительно работоспособной модели планирования заключается в том, чтобы учесть в ней большее количество показателей (факторов), важных для принятия решения.

4. ПЛАНИРОВАНИЕ РАЗВИТИЯ СЕТИ ЭЛЕКТРОПЕРЕДАЧИ

Планирование расширения сети электропередачи предусматривает определение новых узлов и ветвей, которые необходимо добавить к существующей сети с целью обеспечения энергией потребителей по мере роста нагрузки и ввода новых электростанций. Для решения этой задачи использовались методы линейного динамического программирования и эвристические методы. Одна из моделей, в которой методы линейного программирования сочетаются с эвристическими процедурами, была опробована семью энергообъединениями США при разработке программы развития сети электропередачи на период от 10 до 25 лет.

В данном разделе рассматриваются общая постановка задачи планирования, различные подходы к ее решению, при этом особое внимание уделяется методам линейного программирования.

4.1. Задача планирования расширения электрической сети

Планирование развития сети электропередачи осуществляется на основе прогнозирования спроса в распределенных по сети узлах электропотребления и планирования ввода новых мощностей. Результаты решения задачи представляются в виде перечня вводимых в данном году новых ветвей сети с указанием конечных пунктов и уровня напряжения. Например, в 1986 г. вводится участок линии напряжением 345 кВ между электростанцией А и подстанцией Н. Таким образом, общий принцип планирования — это изменение конфигурации сети в зависимости от выбранных мест расположения электростанций или сохранения конфигурации (что возможно в том случае, если вместо напряжения 345 кВ для передачи энергии будет использоваться напряжение как 345 кВ, так и 765 кВ). Планирование будущей структуры сети создает также исходную базу для более детальных исследований, предусматривающих анализ потокораспределения мощности и обеспечение устойчивости режимов [159].¹⁾

¹⁾ Обычно рассматриваются три типа моделей потокораспределения: для сети переменного тока, сети с идеализацией по постоянному току и сети с ограниченными пропускными способностями ветвей. Модели первого типа предусматривают решение системы уравнений в комплексной форме для определения потоков активной и реактивной мощности в цепи (ветви), напряжений и фазовых углов напряжений на шинах (в узлах). Модели второго типа предусматривают решение системы уравнений действительных переменных для

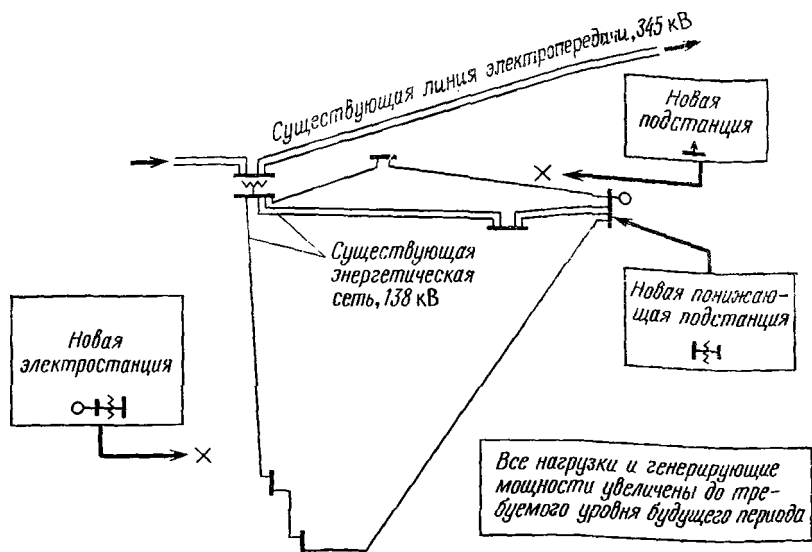


Рис. 13. Схематическое изображение задачи планирования расширения сети электропередачи: существующая сеть, вводимая электростанция и узлы подключения нагрузки.

Типичная задача планирования расширения электрической сети схематически представлена на рис. 13. На схеме показаны фрагмент существующей сети электропередачи напряжением 345/138 кВ и новые элементы сети: питающая подстанция (узел потребления), понижающая подстанция и вводимая электростанция. Все параметры, характеризующие потребности в электроэнергии, приведены к условиям будущего периода, причем их значения могут быть вдвое или даже втрое больше существующих значений. Задача состоит в определении такой комбинации вновь вводимых ветвей напряжением 345 кВ и 138 кВ, которая обеспечивала бы требуемые условия передачи электроэнергии при минимальных затратах. Ввод новых узлов и ветвей сети должен удовлетворять не только частному требованию обеспечения за-

определения потоков активной мощности в цепи и фазовых углов напряжений на шинах. В обоих типах моделей потокораспределения осуществляется минимизация квадратичной функции потерь в сети при обращении в нуль алгебраической суммы напряжений (угловых сдвигов напряжений) в любом замкнутом контуре. Модель потокораспределения в сети с ограниченными пропускными способностями ветвей предусматривает определение приближенных значений потоков мощности; это достигается минимизацией линейной функции затрат, причем требование равенства нулю алгебраической суммы угловых сдвигов напряжений по замкнутым контурам не вводится.

данного уровня нагрузки, но и соответствовать реализации такой стратегии развития сети, при которой каждый из предыдущих этапов развития (расширения) обеспечивает наиболее экономичное расширение сети в будущем.

4.2. Возможные методы решения задачи планирования

В одном из подходов к решению задачи планирования основное внимание уделяется автоматизации инженерных расчетов. Результаты исследований потокораспределения мощности в сети переменного тока и проблемы устойчивости режимов используются планирующими органами при проектировании электрических сетей. Поэтому попытка включить отдельные этапы общей схемы выбора вариантов решения относительно расширения сетей непосредственно в общий процесс планирования можно рассматривать как стремление осуществить частичную автоматизацию этого процесса. В этом направлении был выполнен ряд исследований результатом которых явилась разработка машинной программы, позволяющей осуществлять календарное планирование производства энергии, рассчитывать потокораспределение мощности в сети переменного тока и определять необходимые корректирующие воздействия при возникновении перегрузок [57]. Однако разработки, принятые в этом направлении, оказались не столь успешными, как ожидалось вначале так как реализация данного подхода вызвала существенные затруднения по следующим причинам: во-первых, большой объем необходимых данных затруднял начало исследований с помощью новых программ; во-вторых, точный анализ распределения нагрузок требовал значительных затрат и, в-третьих, нагрузка в каждый последующий год планового периода должна была увеличиваться незначительно, так чтобы обеспечивались сходимость результатов, получаемых при использовании нелинейных уравнений, описывающих распределение нагрузок. Поэтому в случае планирования на период в 10 или 20 лет необходимо было иметь в распоряжении готовый план развития системы на каждый второй или третий год. Сначала это требование казалось приемлемым. Однако было установлено, что даже при анализе небольших изменений электропотребления неизбежно приходится оперировать и с малыми значениями перегрузок, которые в свою очередь могут быть устранены путем незначительных изменений конфигурации сети. Такая последовательность незначительных изменений сети приводит к разработке проекта, неприемлемого для специалиста, занимающегося планированием развития системы. В этом заключается четвертая и наиболее труднопреодолимая причина, препятствующая практической реализации рассматриваемого подхода.

Процесс выбора варианта решения относительно развития электрической сети можно улучшить, используя методы динамического программирования, для исследования возможных путей ее последовательного расширения. В работе [98] описан подобный метод, в котором используется заранее заданное состояние сети в каждом из будущих периодов. Метод динамического программирования в сочетании с более простым способом аппроксимации потокораспределения нагрузок (на основе теоремы о минимальном разрезе и максимальном потоке) используется в Бельгии и США [13]. Вместо определения конфигурации сети в заданные периоды времени решается обратная задача и находится решение для ближайшего из периодов, в котором требуется усиление сети. При этом делается допущение о том, что нагрузка является линейно-возрастающей функцией времени. В процессе расчетов фиксируется несколько вариантов усиления сети. Полученная таким образом информация используется далее в методе усеченного динамического программирования, с помощью которого формируется дерево последовательных решений и осуществляется поиск пути, обеспечивающего развитие системы при минимальных затратах, а также поиск нескольких других путей, близких по затратам к оптимальному.

Авторы работы [105] использовали расчет потокораспределения нагрузки с идеализацией по постоянному току для нахождения ветвей с недостаточной пропускной способностью. Затем с помощью методов линейного программирования было найдено, какие изменения нужно произвести в сети, чтобы устранить перегрузки линий. В работе также обсуждаются возможности использования динамического программирования в качестве основного расчетного метода распределения изменений пропускных способностей на целое число контуров. Авторы отмечают наличие огромного количества возможных вариантов решений рассмотренных задач и те затруднения, которые возникают при попытке выразить в простой и удобной форме ограничения, связанные с надежностью.

Затраты машинного времени на решение задач, связанных с расширением сети, можно снизить за счет совместного использования сортировки, оптимального упорядочения, процедур исключения Гаусса, модифицированного симплекс-метода ЖП и методов программирования, использующих слабую заполненность матриц [132]. Например, для расчетов сети, состоящей из 100 ветвей, затраты машинного времени в этом случае могут сократиться с 20 мин до 20 с. Тем не менее расчетная модель, удовлетворяющая требованиям практического проектирования сетей, пока еще не построена.

Авторы работы [25] использовали ЭВМ в диалоговом режиме, причем специалист-плановик предлагал исходные допустимые

варианты проекта, а ЭВМ осуществляла их проверку и изменения в соответствии с эвристическими логическими правилами. Авторами отмечено, что несколько вариантов проекта обычно характеризуются почти одинаковыми затратами. Это обстоятельство указывает на наличие пологого оптимума. В данной работе приведено решение задачи для сети, содержащей 17 узлов, по авторам считают, что предлагаемый метод можно использовать и при анализе сетей реальных размеров.

Оригинальный метод расчета параметров чувствительности для возможных вариантов развития сети базируется на работе, выполненной в области синтеза электронных схем [142]. Этот метод, получивший название «метод присоединенной сети», используется при планировании, которое осуществляется с помощью автоматизированной системы и предусматривает выполнение расчетов распределения нагрузок с идеализацией по постоянному току и эвристического поиска подсетей, дополняющих существующую сеть в соответствии с заданными требованиями. Метод присоединенной сети представляется особенно многообещающим в силу быстроты выполнения расчетов и приемлемых требований к объему памяти ЭВМ.

В книге Найта [114] обсуждаются критерии надежности, используемые в расчетах сетей, неопределенность, связанная с выбором мест расположения подстанций, а также различные модели линейного и динамического программирования, предложенные специалистами европейских стран.

4.3. Использование методов линейного программирования для решения задачи планирования

Метод, принятый в некоторых энергообъединениях США для начальной стадии концептуального планирования расширения электрических сетей, сочетает в себе модель ЛП для расчета сети с ограниченными пропускными способностями ветвей и эвристические правила выбора вновь вводимых элементов сети [82]. Решение, получаемое с помощью линейной модели, в котором минимизируется величина перегрузки, выражаемая в мегаватт-милиях (где мили могут быть приведены к экономическим параметрам), позволяет определить, имеется ли дефицит пропускных способностей линий, и если имеется, то установить, насколько он велик и какие участки сети требуют увеличения пропускных способностей. Основные преимущества этого метода заключаются в более простых требованиях к составу исходных данных, быстрой реализации расчетов и возможности в любой момент выполнить расчеты для условий будущего периода даже в том случае, если рассматриваемой сети еще не существует. Фактический выбор варианта развития сети для каждого этапа осуществляется после

предварительного определения конфигурации сети на последний год планового периода, т. е. на 10—20 лет вперед. При этом определяются самые существенные потребности развития и, следовательно, обосновывается ввод наиболее мощных линий электропередачи высокого напряжения.

Информация, полученная при анализе развития сети на длительный период, служит основой для планирования последовательного расширения сети на каждом этапе рассматриваемого периода. Такое планирование вызывает недоверие со стороны инженеров-энергетиков, так как они полагают, что задачи, связанные с развитием электрических сетей, не могут быть решены с помощью «неэлектрической» модели. Тем не менее проверка разработанного плана путем анализа распределения активных и реактивных мощностей свидетельствует о том, что получаемые результаты обычно подтверждают работоспособность принятой конфигурации сети. Однако это вовсе не означает, что разработанный вариант развития сети не может быть улучшен.

В первые годы при построении модели основные усилия были направлены на то, чтобы, используя логику оптимизационных методов, найти такой вариант расширения сети, который уже не мог бы быть улучшен специалистом, занимающимся планированием. Для этого прежде всего требовалось четко определить, что понимает специалист-плановик под оптимальным вариантом. С этой целью был проведен следующий контрольный эксперимент. Для энергосистемы одного из штатов Среднего Запада США были выполнены соответствующие проектные работы по развитию электрической сети на предстоящие 20 лет. Так как эти разработки базировались на четырех различных логических концепциях расширения сети, то было подготовлено четыре варианта плана, существенно отличающиеся в отношении трассировки передающих линий высокого напряжения, но почти идентичные с точки зрения стоимостных характеристик. Для оценки предложенных вариантов схем и выбора оптимального плана были приглашены четыре специалиста-плановика. Каждый из них нашел во всех предложенных вариантах те или иные преимущества, однако единого мнения по вопросу о том, какая из схем является наилучшей, достигнуто не было. В связи с этим первоначальная цель разработки модели планирования развития электрической сети была изменена и усилия исследователей были направлены на то, чтобы построить модель, обеспечивающую быстрое составление достаточно экономичного плана, который удовлетворял бы всем необходимым требованиям и мог служить хорошей исходной базой для последующей доработки уже непосредственно самим специалистом-плановиком. Чтобы содействовать разработке дополнительных вариантов плана, возможности всех новых схем оценивались с использованием двойственных пере-

менных, что обусловлено их важной ролью в процессе корректировки дефицита пропускных способностей элементов сети.

После того как стало очевидным, что невозможно пайти оптимальное решение, была начата разработка методов планирования, пригодных для практического использования. Другими словами, разработчик программы должен был быть также и ее основным пользователем в рамках исследований консультативного характера. Первоначальная идея состояла в том, чтобы разработать машинную программу средней сложности, которую можно было бы использовать для анализа развития линий электропередач самого высокого напряжения в рамках общего планирования, а не на завершающей стадии детальной разработки проекта. При построении такой модели следовало стремиться не к исчерпывающей детализации, а к расширению возможностей программы в отношении учета всех аспектов проблемы. Однако при таком подходе надо было оперировать только наиболее фундаментальными характеристиками сети.

Из-за большого числа контрольных требований, которым должны соответствовать параметры сети, масштабы поставленной задачи существенно увеличиваются. Так, например, только в пределах годового интервала может возникнуть необходимость в проверке работоспособности сети в 6—8 различных режимах, таких, как покрытие летнего пика нагрузки, покрытие минимальной нагрузки летнего сезона (при этом более старые агрегаты, расположенные вблизи узлов нагрузки, отключаются и, кроме того, все агрегаты ГАЭС работают в насосном режиме, а не в режиме генератора), а также в режимах с различными комбинациями простоев генерирующих агрегатов, например двух наиболее мощных агрегатов в восточной части системы, затем в центральной части после этого — в южной и т. д. Для других сезонов также могут проверяться некоторые режимы, но обычно это не является необходимым.

Помимо проверки режимов передачи и распределения энергии контрольные требования включают также и требования к надежности электрической сети. В период обычного пика электропотребления выход из строя любой линии электропередачи не должен приводить к перегрузкам других линий. Более того, в настоящее время многие электроэнергетические компании требуют, чтобы к перегрузкам не приводил и одновременный выход из строя любой пары линий электропередач. Таким образом, проверка работоспособности проектируемой электрической сети, необходимая для анализа годового цикла эксплуатации, может потребовать решения свыше 100 сетевых задач. Для исследования развития объединенной энергетической системы на предстоящие 20 лет потребовалась проверка 93 режимов, характеризующих покрытие пиковых и минимальных нагрузок, а также режимов работы сети

при выходе из строя геперирующих мощностей. Затем было необходимо проверить 3600 вариантов отключения линий электропередачи. В целом при проектировании сети электропередачи между 90 узлами, включающей 15 линий напряжением 765 кВ, 39 подземных кабельных линий напряжением 345 кВ и 79 воздушных линий напряжением 345 кВ, потребовалось решить около 4500 сетевых задач.

В результате разработки методов линейного программирования, предназначенных для решения задач о потоках в сетях с ограничениями на пропускные способности дуг стало возможно построение быстрых в реализации и легко адаптируемых моделей, необходимых для обнаружения дефицита пропускных способностей линий передач [38]. Согласно таким моделям, реальная электрическая сеть представлялась в виде совокупности двух типов звеньев, а именно линий, которые характеризуются ограниченной пропускной способностью и низкой стоимостью передачи энергии, и «перегруженных ветвей», которым приписывается неограниченная пропускная способность, но очень высокая стоимость передачи энергии. Подобные модели позволяли получить решение сетевой задачи и таким образом определять дефицит пропускных способностей на участке, наиболее выгодном с точки зрения усиления существующей сети электропередачи. В этом отношении данный класс моделей имел существенное преимущество перед «электрическими моделями», которые позволяли определить лишь перегрузку отдельных ветвей, после чего требовалось отыскать еще и трассу новых линий, предназначенных для разгрузки линий с недостаточной пропускной способностью. При решении рассматриваемой задачи планирования с использованием нового подхода, не требующего дополнительного поиска нового оптимального маршрута передачи электроэнергии и базирующегося на применении упрощенных моделей линейного программирования, затраты, связанные с использованием ЭВМ, снижались почти в 100 раз по сравнению с теми затратами, которые были необходимы для исследования аналогичных проблем с помощью расчетов потокораспределения энергии в сети. Значительное снижение затрат и удобный для практического использования состав входных и выходных переменных позволили применить такой подход и при исследовании гораздо более широкого круга характеристик будущей системы, таких, как размещение вводимых агрегатов, использование новых уровней напряжения в линиях, передача электроэнергии по цепям постоянного тока, использование подземных кабельных линий. Результаты исследований, связанных с расширением сети передачи электроэнергии, нередко становятся предметом изучения, позволяющего выявить неучтенные ограничения или не осознанные ранее предпочтительные характеристики проектного решения. Обычно составляется пе-

сколько предварительных вариантов плана, причем специалист-плановик сам приспособливает исходные данные для улучшения модели проектируемой системы. Таким образом, при анализе любого плана не следует чрезмерно экономить усилия и время; целесообразно провести повторное исследование, изменяя исходные данные для улучшения модели или, если модель вполне приемлема, изменяя эти данные для сравнительной оценки возможных вариантов размещения энергоагрегатов, используемого напряжения и новых трасс линий электропередачи.

Разработка сетевых моделей линейного программирования и соответствующих вычислительных методов, которая началась в 1963 году, способствовала быстрому прогрессу в области общего планирования сетей электропередачи, характеризующегося переходом от использования традиционных методов проб и ошибок к таким методам, которые позволяют быстро оценить варианты развития системы на предстоящие 10—25 лет. Усилия разработчиков были направлены на то, чтобы обеспечить составление плана, который мог бы служить наилучшей исходной базой для дальнейшей доработки в рамках творческого, неформализуемого процесса. В результате таких усилий созданы конструктивные методы, к которым постоянно обращаются электроэнергетические компании и объединения, исследующие возможности будущего развития электрических сетей

ЛИТЕРАТУРА

1. Adamson A. M., Kenney J. F., Moisan R. W., Parametric Sensitivity Method for Establishing Optimum Long Range Generation Mix, *Proc. Am. Power Conf.*, 35, 467—472 (1973).
2. AIEEE Committee Report, Application of Probability Methods to Generating Capacity Problems, *AIEEE Trans.*, 79, Pt. III, 1165—1177 (1960).
3. Akiyama T., Sekine Y., Economic Operation of Power-Generating Systems Comprising Hydroelectric, Thermal and Pump-Up Plants: An Application of Network Flow Theory and the Decomposition Principle of Linear Programming, *Electrical Eng. Japan*, 86, 74—85 (1966).
4. Aldrich J. F., Happ H. H., Leuer J. F., Multi-Area Dispatch, *IEEE Trans.*, PAS-90, 2661—2670 (1971).
5. Anderson D., Models for Determining Least-Cost Investments in Electricity Supply, *Bell J. Economics Management Sci.*, 3, 267—299 (1972).
6. Anderson S. W., Jenkins R. T., Joy D., Merrill E. T., Wolsko T. D., Methods of Planning Generating Unit Operation, *Proc. Am. Power Conf.*, 32, 123—130 (1970).
7. Anstine L. T., Ringlee R. J., Susquehanna River Short Range Hydrothermal Coordination, *IEEE Trans.*, PAS-82, 1885—1891 (1963).
8. Anstine L. T., Burke R. E., Casey J. E., Holgate R., John R. S., Stewart H. G., Application of Probability Methods to the Determination of Spinning Reserve Requirements for the Pennsylvania-New Jersey-Maryland Interconnection, *IEEE Trans.*, PAS-82, 726—735.
9. Ayoub A. K., Patton A. D., Optimal Thermal Generating Unit Commitment, *IEEE Trans.*, PAS-90, 1752—1756 (1971).

10. Bailey E. S., Hawkins E. S., Galloway C. D., Wood A. J., Generation Planning for Interconnected Systems, *IEEE Trans. Special Publication T-152*, 1963, pp. 761—788.
11. Bainbridge E. S., McNamee J. M., Robinson D. J., Nevison R. D., Hydrothermal Dispatch with Pumped Storage, *IEEE Trans.*, PAS-85, 472—485 (1966).
12. Baleriaux H., Jamouille E., Linard F., Simulation de l'exploitation d'un parc de Machines Thermiques de Production d'électricité couplé à des Stations de Pompage, *Review E (ed. SRBE)*, V, 3—24 (1967).
13. Baleriaux H., Douliez P., VanKelecom J., Optimal Investment Policy for a Growing Electrical Network by a Sequential Decision Method, CIGRE International Conference on Large High Tension Electric Systems, 1970 Session Paper 32—08.
14. Bechert T. E., Kwatny H. G., On The Optimal Dynamic Dispatch of Real Power, *IEEE Trans.*, PAS-91, 889—898 (1972).
15. Benthall T. P., Automatic Load Scheduling in a Multiarea Power System, *Proc. IEEE*, 115, 592—596 (1968).
16. Bessiere F., The Investment '85' Model of Electricite de France, *Management Sci.*, 17, B-192—B-211 (1970).
17. Biggerstaff B. E., Jackson T. M., The Markov Process as a Means of Determining Generating Unit State Probabilities for Use in Spinning Reserve Applications, *IEEE Trans.*, PAS-88, 423—430 (1969).
18. Billinton R., Power System Reliability Evaluation, Gordon and Breach, New York, 1970.
19. Billinton R., Jain A. V., Unit Derating Levels in Spinning Reserve Studies, *IEEE Trans.*, PAS-90, 1677—1688 (1971).
20. Billinton R., The Effect of Rapid Start and Hot Reserve Units in Spinning Reserve Studies, *IEEE Trans.*, PAS-91, 511—516 (1972).
21. Billinton R., Sachdeva S. S., Optimal Real and Reactive Power Operation in a Hydrothermal System, *IEEE Trans.*, PAS-91, 1405—1411 (1972).
22. Billinton R., Bibliography on the Application of Probability Methods in Power System Reliability Evaluation, *IEEE Trans.*, PAS-91, 649—660 (1972).
23. Billinton R., Ringlee R. J., Wood A. J., Power-System Reliability Calculations, The MIT Press, Cambridge, Massachusetts, 1973.
24. Billinton R., Sachdeva S. S., Real and Reactive Power Optimization by Suboptimum Techniques, *IEEE Trans.*, PAS-92, 950—956 (1973).
25. Boardman J. T., Hogg B. W., Computer Method for Design of Electricity-Supply Networks, *Proc. IEE*, 19, 851—856 (1972).
26. Bonaert A. P., El-Abiad A. H., Koivo A. J., Optimal Scheduling of Hydro-Thermal Power Systems, *IEEE Trans.*, PAS-91, 263—270 (1972).
27. Bonaert A. P., Effects of Hydrodynamics on Optimum Scheduling of Thermo-Hydro Power Systems, *IEEE Trans.*, PAS-91, 1412—1420 (1972).
28. Booth R. R., Optimal Generation Planning Considering Uncertainty, *IEEE Trans.*, PAS-91, 70—77 (1972).
29. Brewer C., Revington A. M., Linear Programming for Optimizing Generation and Immediate Spare: Development and Application (LPS), IEE Conference on Computers in Power System Operation and Control, Bournemouth College of Technology, England, 1972, pp. 115—134.
30. Brown W. D., Three C's of System Reliability, *Electrical World*, 160, 61—76 (August 26, 1968).
31. Brownlee W. R., Discussion of Improved Loss Formula Computation by Optimally Ordered Elimination Techniques by Meyer and Albertson, *IEEE Trans.*, PAS-90, 67 (1970).
32. Bobenko J. A., Anderson M., Probabilistic Evaluation of the Operation Reserve in a Power System, *IEEE PICA Conf. Proc.*, 1973, pp. 240—250.

33. Buchmeir F. A., Jr., Replacement Energy Costs Set Outages, *Electrical World*, 180, 116—117 (September 15, 1973).
34. Carlin J. F., Lyons R. H., Mitiguy J. R., Murphy E. F., Galloway C. D., Sager M. A., Wood A. J., Corporate Model of an Electric Utility, *IEEE Spectrum*, 6, 75—84 (1969).
35. Carpentier J. L., Differential Injections Method a General Method for Secure and Optimal Load Flows, *IEEE PICA Conf. Proc.*, 1973, pp. 255—262.
36. Carpentier J. L., Total Injections Method, A Method for the Unit Commitment Problem Including Secure and Optimal Load Flow, *IEEE, PICA Conf. Proc.*, 1973, pp. 263—267.
37. Cash P. W., Scott E. C., Security of Supply in the Planning and Operation of European Power Systems, *IEEE Trans.*, PAS-88, 6—21 (1969).
38. Charnes A., Cooper W. W., Management Models and Industrial Applications of Linear Programming, Wiley, New York, 1961.
39. Chen G. K. C., Winters P. R., Forecasting Peak Demand for an Electric Utility with a Hybrid Exponential Model, *Management Sci.*, 12, B-531 — B-538 (1966).
40. Christiaanse W. R., Short-Term Load Forecasting Using General Exponential Smoothing, *IEEE Trans.*, PAS-90, 900—910 (1971).
41. Christiaanse W. R., Palmer A. H., A Technique for the Automated Scheduling of the Maintenance of Generating Facilities, *IEEE Trans.*, PAS-91, 137—144 (1972).
42. Christiaanse W. R., A Program for Calculating Optimal Maintenance Schedules Recognizing Constraints, *IEEE PICA Conf. Proc.*, 1973, pp. 230—239.
43. Cobian M. J., Optimal Pumped Storage Operation with Interconnected Power Systems, *IEEE Trans.*, PAS-90, 1391—1399 (1971).
44. Cook V. M., Galloway C. D., Steinberg M. J., Wood A. J., Determination of Reserve Requirements of Two Interconnected Systems, *AIEE Trans.*, PAS-82, 18—33 (1963).
45. Cook V. M., Ringlee R. J., Wood A. J., Frequency and Duration Methods for Power System Reliability Calculation, Part IV, Models for Multiple Boiler-Turbines and for Partial Outage States, *IEEE Trans.*, PAS-88, 124—1232 (1969).
46. Corpening S. L., Reppen N. D., Ringlee R. J., Experience with Weather Sensitive Load Models for Short and Long-Term Forecasting, *IEEE Trans.*, PAS-92, 1966—1972 (1973).
47. Couch G. H., Morrison I. F., Substation Switching — An Approach to the Determination of Optimal Sequences, 4th PICA Proceedings, Grenoble, September 11—16, 1972.
48. Dahlin E. B., Shen D. W. C., Optimal Solution to the Hydro-Steam Dispatch Problem for Certain Practical Systems, *IEEE Trans.*, PAS-85, 437—458 (1966).
49. Dale K. M., Dynamic Programming Approach to the Selection and Timing of Generation Plant Additions, *Proc. IEE*, 113, 803—811 (1966).
50. Dandeno P. L., Discussion of Direct Calculation of Transmission Loss Formula-I, L. K., Kirchmayer, H. H. Happ, C. W. Stagg, J. F. Hohenstein, *AIEE Trans.*, 79, Pt. III, 968 (1960).
51. Daniels H., Mo-Shing Chen, An Optimization Technique and Security Calculations for Dispatching Computers, *IEEE Trans.*, PAS-91, 883—888 (1972).
52. Davey J., Sacks J. J., Cunningham G. W., Priest K. W., Practical Application of Weather Sensitive Load Forecasting to System Planning, *IEEE Trans.*, PAS-92, 971—977 (1973).
53. Davies M., Probability Studies of Interconnection Capacity, *Proc. IEE*, 117, 1382—1389 (1970).
54. Davidson P. M., Payne T. B., Kohlerman F. L., Masters G. L., Evans

- J. R., Lovewell K. M., Shaefer G. R., Unit Commitment Star-Step Scheduling in the Pennsylvania—New Jersey—Maryland Interconnection, *IEEE Conf. Record, 1967 Power Industry Computer Applications Conference*, 34C69, 1967, pp. 127—131.
55. Day J. T., Forecasting Minimum Production Costs with Linear Programming, *IEEE Trans.*, PAS-90, 814—823 (1974).
56. Day J. T., Federowich A. J., Menge E. E., Optimizing Generation Planning, *Power Eng.*, 117, 32—35 (July, 1973).
57. DeSalvo C. A., Smith H. L., Automatic Transmission Planning with A—C Load Flow and Incremental Transmission Loss Evaluation, *IEEE Trans.*, PAS-84, 156—163 (1965).
58. Dommel H. W., Tinney W. F., Optimal Power Flow Solutions, *IEEE Trans.*, PAS-87, 1866—1876 (1968).
59. Dopazo J. F., Klitin O. A., Stagg G. W., Watson M., An Optimization Technique for Real and Reactive Power Allocation, *IEEE Trans.*, 55, 1877—1885 (1967).
60. Dopazo J. F., Ehrmann S. T., Klitin O. A., Sasson A. M., Justification of the AEP Real Time Load Flow Project, *IEEE Trans.*, PAS-92, 1501—1509 (1973).
61. Dopazo J. F., Merrill H. M., Optimal Generator Maintenance Using Integer Programming, *IEEE Conference Paper C74 162-4*, 1964.
62. Dusonchet Y. P., El-Abiad A., Transmission Planning Using Discrete Dynamic Optimization, *IEEE Trans.*, PAS-92, 1358—1374 (1973).
63. DyLiaccio T. E., The Adaptive Reliability Control System, *IEEE Trans.*, PAS-86, 517—532 (1967).
64. DyLiaccio T. E., Wirtz B. F., Wheeler D. A., Automation of the CEI System for Security, *IEEE Trans.*, PAS-91, 831—844 (1972).
65. EEI, Report of Equipment Availability for the Thirteen-Year Period 1960—1972, EEI Publication No. 73—46, Edison Electric Institute, New York, N. Y., 1973.
66. EEI System Planning Committee, System Planning Committee Surveys—Load Forecasting Methods, *Edison Elec. Inst. Bulletin*, March/April, 1973, p. 89.
67. El-Hawary M. E., Christensen G. S., Functional Optimization of Common-Flow Hydro-Thermal Systems, *IEEE Trans.*, PAS-91, 1833—1840 (1972).
68. Elgerd O. I., *Electric Energy Systems*, McGraw-Hill, New York, 1971.
69. Enns M., Tinney W. F., Eds., *Computers in the Power Industry*, *Proc. IEEE*, 62, 868—1024 (1974).
70. Fanshel S., Lynes E. S., Economic Power Generation Using Linear Programming, *IEEE Trans.*, PAS-83, 347—356 (1964).
71. Firestone L., Masters W. D., Monteith A. H., The CAPCO Group Probability Technique for Timing Capacity Additions and Allocation of Capacity Responsibility, *IEEE Trans.*, PAS-88, 1174—1182 (1969).
72. Fosha C. E., Jr., Elgerd O. I., The Megawatt-Frequency Control Problem, A New Approach Via Optimal Control Theory, *IEEE Trans.*, PAS-89, 563—477 (1970).
73. FPC, 1970 National Power Survey, 4 Vol., U. S. Government Printing Office, Issued between 1970 and 1972.
74. FPC Committee, Descriptions of Forecasting Practices, Appendix A, of the 1970 National Power Survey, Part IV, Federal Power Commission, 1970.
75. Friedlander G. D., Computer-Controlled Power Systems, Part I—Boiler-Turbine Unit Controls, *IEEE Spectrum*, 2, 60—82 (1965).
76. Friedlander G. D., Computer-Controlled Power Systems, Part II—Area Controls and Load Dispatch, *IEEE Spectrum*, 2, 72—92 (1965).
77. Fukao T., Yamazaki T., A Computational Method of Economic Operation

- of Hydro-Thermal Power Systems Including Flow-Interconnected Hydro-Power Plants, *Electrotechnical J. of Japan*, 6, 22—26 (1960).
78. Galloway C. D., Ringlee R. J., An Investigation of Pumped Storage Scheduling, *IEEE Trans.*, PAS-85, 459—465 (1966).
 79. Galloway C. D., Garver L. L., Kirchmayer L. K., Wood K. A. J., Generation-Transmission Planning and Economic Evaluation, 3rd International Power Systems Computation Conference, Rome, Italy, June 23—27, 1969.
 80. Garver L. L., Power Generation Scheduling by Integer Programming-Development, *AIEE Trans. (Power Apparatus and Syst)*, 81, 730—735 (1963).
 81. Garver L. L., Effective Load Carrying Capability of Generating Units, *IEEE Trans.*, PAS-85, 910—919 (1966).
 82. Garver L. L., Transmission Network Estimation Using Linear Programming, *IEEE Trans.*, PAS-89, 1688—1697 (1970).
 83. Garver L. L., Adjusting Maintenance Schedules to Levelize Risk, *IEEE Trans.*, PAS-91, 2057—2063 (1972).
 84. Gent M. R., Lamont J. W., Minimum-Emission Dispatch, *IEEE Trans.*, PAS-90, 2650—2660 (1971).
 85. Gilleland J. E., Hoskins R. E., Kiefer W. M., Koncel E. F., Fitting Nuclear Capacity into Utility Planning, *Proc. Am. Power Conf.*, 32, 116—122 (1970).
 86. Graham E. G., Gunter G. E., Muttart D. C., Reppen N. D., Ringlee R. J., Wood A. J., Generation System Planning Models for Hydro-Thermal Systems, *Am. Power Conf. Proc.*, 33, 1160—1168 (1971).
 87. Gungor R. B., Tsang N. F., Webb B., A Technique for Optimizing Real and Reactive Power Schedules, *IEEE Trans.*, PAS-90, 1781—1790 (1971).
 88. Gupta P. C., Yamada K., Adaptive Short-Term Forecasting of Hourly Loads Using Weather Information, *IEEE Trans.*, PAS-91, 2085—2094 (1972).
 89. Guy J. D., Security Constrained Unit Commitment, *IEEE Trans.*, PAS-90, 1385—1390 (1971).
 90. Hajdu L. P., Peschon L., Tinney W. F., Piercy D. S., Optimum Load-Shedding Policy for Power Systems, *IEEE Trans.*, PAS-87, 784—795 (1968).
 91. Hamam Y. M., Brameller A., Diakoptical Optimization of Electrical Power Systems, IEE Conference on Computers in Power System Operation and Control, Bournemouth College of Tech., England, 1972, pp. 158—167.
 92. Hano I., Tamura Y., Narita S., An Application of the Maximum Principle to the Most Economical Operation of Power Systems, *IEEE Trans.*, PAS-85, 486—494 (1966).
 93. Happ H. H., Large Scale Hydro-Thermal Unit Commitment-Method and Results, *IEEE Trans.*, PAS-90, 1373—1384 (1971).
 94. Happ H. H., Power Pools and Superpools, *IEEE Spectrum*, 10, 54—62 (March 1973).
 95. Happ H. H., Optimal Power Dispatch, *IEEE Trans.*, PAS-93, 820—830 (1974).
 96. Hara K., Kimura M., Honda N., A Method for Planning Economic Unit Commitment and Maintenance of Thermal Power Systems, *IEEE Trans.*, PAS-85, 427—436 (1966).
 97. Heinemann G. T., Nordman D. A., Plant E. C., The Relationship Between Summer Weather and Summer Loads — A Regression Analysis, *IEEE Trans.*, PAS-85, 1144—1154 (1966).
 98. Henault P. H., Eastvedt R. B., Peschon J., Hajdu L. P., Power System Long-Term Planning in the Presence of Uncertainty, *IEEE Trans.*, PAS-89, 156—164 (1970).
 99. Hicks R. H., Gagnon C. R., Jacoby L. S., Kowalik J. S., Large Scale

- Nonlinear Optimization of Energy Capability for the Pacific Northwest Hydroelectric System, *IEEE Trans.*, PAS-93, 1604—1612 (1974).
100. Hubbard J. C., Monitoring Electric-Load Forecasts, *IEEE Trans.*, PAS-88, 797—801 (1969).
 101. IEEE Committee Report, Present Practices in the Economic Operation of Power Systems, *IEEE Trans.*, PAS-90, 1768—1775 (1971).
 102. IEEE Committee, Bibliography on Engineering Economics, *IEEE Trans.*, PAS-91, 1976—1984 (1972).
 103. Jaynes P. H., Profitability and Economic Choice, The Iowa State University Press, Ames, Iowa, 1968.
 104. Jolissaint C. H., Arvanitidis N. V., Luenberger D. G., Decomposition of Real and Reactive Power Flows: A Method Suited for On-Line Applications, *IEEE Trans.*, PAS-91, 661—670 (1972).
 105. Kaltenbach J., Peschon J., Gehrig E. H., A Mathematical Optimization Technique for the Expansion of Electric Power Transmission Systems, *IEEE Trans.*, PAS-89, 113—119 (1970).
 106. Kaltenbach J. C., Hajdu L. P., Optimal Corrective Rescheduling for Power System Security, *IEEE Trans.*, PAS-90, 843—851 (1971).
 107. Kerr R. H., Scheidt J. L., Fontana A. J., Jr., Wiley J. K., Unit Commitment, *IEEE Trans.*, PAS-85, 417—421 (1966).
 108. Kind C. G., Probability Technique Adds New Meaning to Demand Forecasting, *Electric Light & Power*, January 1969, pp. 78—82.
 109. Kirchmayer L. K., Mellor A. G., O'Mara J. F., Stevenson J. R., An Investigation of the Economic Size of Steam-Electric Generating Units, *AIEE Trans.*, 74, Pt. III, 600—614 (1955).
 110. Kirchmayer L. K., Mellor A. G., Simming H. O., Jr., The Effect of Interconnections on Economic Generation Expansion Patterns, *AIEE Trans.*, 76, Pt. III, 203—214 (1957).
 111. Kirchmayer L. K., Economic Operation of Power Systems, Wiley, New York, 1958.
 112. Kirchmayer L. K., Economic Control of Interconnected Systems, Wiley, New York, 1959.
 113. Kirchmayer L. K., Ringlee R. J., Optimal Control of Thermal System Operation, IFAC Congress Basel, Switzerland, September 1963.
 114. Knight U. G., Power Systems Engineering and Mathematics, Pergamon Press, Oxford, England, 1972.
 115. Lamont J. W., Tudor J. R., Survey of Operating Computer Applications, *Proc. Am. Power Conf.*, 32, 1142—1148 (1970).
 116. Lamont J. W., Gent M. R., Environmentally-Oriented Dispatching Techniques, *PICA Conf. Proc.*, 1973, pp. 421—427.
 117. Latham J. H., Nordman D. A., Voorhis J. S., Probability Approach to Electric Utility Load Forecasting, *IEEE Trans.*, PAS-87, 496—504 (1968).
 118. Laughton M. E. (Ed.), 3rd Power Systems Computation Conference Proceedings, Rome, June 23—27, 1969.
 119. Laughton M. A. (Ed.), Fourth Power Systems Computation Conference, University of Grenoble, September 11—16, 1972.
 120. Lev B., A Note on the Analysis of Peak-Demand Forecasts of an Electrical Utility, *Operations Res.*, 18, 174—179 (1970).
 121. Lijesen D., Rosing P. J., Adaptive Forecasting of Hourly Loads Based on Load Measurements and Weather Information, *IEEE Trans.*, PAS-90, 1757—1767 (1971).
 122. Lowery P. G., Generating Unit Commitment by Dynamic Programming, *IEEE Trans.*, PAS-85, 422—426 (1966).
 123. Lynskey J. P., Sloane R. J., Computer Analysis of Outage Data, *IEEE Trans.*, PAS-83, 386—396 (1964).
 124. Marsh W. D., Moisan R. W., Murrell T. C., Perspectives on the Design

- and Application of Generation Planning Programs, Proceedings of the Nuclear Utilities Methods Symposium, Chattanooga, Tennessee, January 16—18, 1974, Sponsored by the American Nuclear Society, ORNL-TM-4443, 1974.
125. McDaniel G. H., Gabrielle A. F., Dispatching Pumped Storage Hydro, *IEEE Trans.*, PAS-85, 465—471 (1966).
 126. Merrill H. M., Schweppe F. C., Bad Data Suppression in Power System Static State Estimation, *IEEE Trans.*, PAS-90, 2718—2726 (1971).
 127. Muckstadt J. A., Wilson R. C., An Application of Mixed-Integer Programming Duality to Scheduling Thermal Generating Systems, *IEEE Trans.*, PAS-87, 1968—1978 (1968).
 128. Nabona N., Freris L. L., Optimisation of Economic Dispatch Through Quadratic and Linear Programming, *Proc., IEEE*, 120, 574—580 (1973).
 129. Nicholson H., Sterling M. J. H., Optimum Dispatch of Active and Reactive Generation by Quadratic Programming, *IEEE Trans.*, PAS-92, 644—645 (1973).
 130. Noakes F., Arismunandar A., Bibliography on Optimum Operation of Power Systems, 1919—1959, *AIEE Trans.*, 81, Pt. III, 864—878 (1963).
 131. Nour N. E., Happ H. H., The Application of Transportation Techniques in Modelling Power Pools, *J. Computers and Operations Res.*, 1, 15—29 (1974).
 132. Ogbuobiri E. C., Sparsity Techniques in Power-System Grid-Expansion Planning, in: Large Sparse Sets of Linear Equations, J. K. Reid (Ed.), Academic Press, New York, 1971, pp. 219—230.
 133. Paris L., Valtota M., Planning the Installation of Large Units in Interconnected Systems, CIGRL 22nd International Conference on Large Electric Systems at High Tension, Vol. II, 1968.
 134. Patton A. D., Short-Term Reliability Calculation, *IEEE Trans.*, PAS-89, 509—513 (April 1970).
 135. Peschon J., Piercy D. S., Tinney W. F., Tveit O. J., Cuenod M., Optimum Control of Reactive Power Flow, *IEEE Trans.*, PAS-87, 40—48 (1968).
 136. Peschon J., Dommel H. W., Powell W., Bree D. W., Jr., Optimum Power Flow for Systems with Area Interchange Controls, *IEEE Trans.*, PAS-91, 898—905 (1972).
 137. PICA, 8th Power Industry Computer Application Conference, Minneapolis, Minnesota, June 4—6, 1973.
 138. Pitcher W. J., Kirchmayer L. K., Mellor A. G., Simmons H. O., Jr., Generator Unit Size Study for the Dayton Power and Light Company, *AIEE Trans.*, 77, Pt. III, 558—563 (1958).
 139. Porretta B., Dhillon R. S., Performance Evaluation to State Estimation from Line Flow Measurement on Ontario Hydro Power System, *IEEE Trans.*, PAS-92, 1702—1712 (1973).
 140. Prewett J. N., Applied Algorithms to Load Dispatching, IEEE Conference on Computers in Power System Operation and Control, Bournemouth College of Technology, England, 1972, pp. 309—317.
 141. PSCC, Power Systems Computation Conference.
 142. Puntel W. R., Reppen N. H., Rignlee R. J., Platts J. E., Ryan W. A., Sullivan P. J., An Automated Method for Long-Range Planning of Transmission Networks, *IEEE Trans.*, PAS-93, 3, 12 (1974).
 143. Rees F. J., Larson R. E., Computer-Aided Dispatching and Operations Planning for an Electric Utility with Multiple Types of Generation, *IEEE Trans.*, PAS-90, 891—899 (March/April 1971).
 144. Reid G. F., Hasdorff L., Economic Dispatch Using Quadratic Programming, *IEEE Trans.*, PAS-92, 2015—2023 (1973).
 145. Revington A. M., Application of Dual Decomposition Methods to Power System Loading, IEEE Conference on Computers in Power System Operation and Control, Bournemouth College of Technology, [England,] May 1972.

146. Ringlee R. J., Williams D. D., Economic System Operation Considering Valve Throttling Losses, II—Distribution of System Loads by the Method of Dynamic Programming, *AIEE Trans.*, PAS-81, 615—622 (1963).
147. Sager M. A., Wood A. J., Corporate Model Programs for System Planning Evaluations, *IEEE Trans.*, PAS-91, 1079—1084 (1972).
148. Sager M. A., Ringlee R. J., Wood A. J., A New Generation Production Cost Program to Recognize Forced Outages, *IEEE Trans.*, PAS-91, 2114—2124 (1972).
149. Sandiford P. J., Bernhotz B., Shelson W., Three Applications of Operations Research in a Large Electric Utility, *Operations Res.*, 4, 663—673 (1956).
150. Sasson A. M., Nonlinear Programming Solutions for Load-Flow, Minimum-Loss, and Economic Dispatching Problems, *IEEE Trans.*, PAS-88, 399—409 (1969).
151. Sasson A. M., Aboytes F., Cardenas R., Gomes F., Vilorio F., A Comparison of Power System Static Optimization Techniques, *IEEE PICA Conf. Proc.*, 1971, pp. 329—337.
152. Sasson A. M., Ehrmann S. T., Lynch P., Van Slyck L. S., Automatic Power System Network Topology Determination, *IEEE Trans.*, PAS-92, 610—618 (1973).
153. Sasson A. M., Merrill H. M., Some Applications of Optimization Techniques to Power Systems Problems, *Proc. IEEE*, 62, 959—972 (1974).
154. Sekin Y., Kawakawi J., Optimal Line Dispatching of Transmission Lines and Transformers, 4th International Power Systems Computation Conference, Grenoble, France, September 11—16, 1972.
155. Shen C. M., Laughton M. A., Power-System Load Scheduling with Security Constraints Using Dual Linear Programming, *IEEE Proc.*, 117, 2117—2127 (1970).
156. Shipley R. B., Patton A. D., Denison J. S., Power Reliability Cost Worth, *IEEE Trans.*, PAS-91, 2204—2212 (1972).
157. Spears H. T., Hicks K. L., Lee S. T. Y., Probability of Loss of Load for Three Areas, *IEEE Trans.*, PAS-89, 521—526 (1970).
158. Stadlin W. O., Economic Allocation of Regulating Margin, *IEEE Trans.*, PAS-70, 1776—1781 (1971).
159. Stagg G. W., El-Abiad A. H., Computer Methods in Power System Analysis, McGraw-Hill, New York, 1968.
160. Stagg G. W., Dopazo J. F., Klitin O. A., VanSlyck L. S., Techniques for the Real-Time Monitoring of Power System Operations, *IEEE Trans.*, PAS-89, 545—555 (1970).
161. Stanton K. N., Gupta P. C., Forecasting Annual of Seasonal Peak Demand in Electric Utility Systems, *IEEE Trans.*, PAS-89, 951—959 (1970).
162. Steinberg M. J., Smith T. H., Economy Loading of Power Plants and Electric Systems, Wiley, 1943.
163. Stevenson W. D., Elements of Power System Analysis, McGraw-Hill, N. Y., 2nd Ed., 1962.
164. Sullivan R. L., Controlling Generator MVAR Loadings Using a Static Optimization Techniques, *IEEE Trans.*, PAS-91, 906—910 (1972).
165. Thanikachalam A., Tudor J. R., Optimal Rescheduling of Power for System Reliability, *IEEE Trans.*, PAS-90, 2186—2192 (1971).
166. Theilsiefje K., Wagner H., Digital Method of Calculation for Optimum Instantaneous Reserve in Interconnected Operation, *Electrotechnische Zeitschrift(A)*, 89, 397—402 (1968).
167. Tinney W. F., Hart C. E., Power Flow Solution by Newton's Method, *IEEE Trans.*, PAS-86, 1449—1460 (1967).
168. Vassell G. S., Tibberts N., Analysis of Generating-Capacity Reserve Requirements for Interconnected Power Systems, *IEEE Trans.*, PAS-91, 638—649 (1972).

169. Watchorn C. W., Inside Hydrothermal Coordination, *IEEE Trans.*, PAS-86, 106—117 (1967).
170. Weedy B. M., Electric Power System, 2nd Ed., Wiley, New York, 1972.
171. Wells D. W., Method for Economic Secure Loading of a Power System, *IEEE Proc.*, 115, 1190—1194 (1968).
172. Yao-nan Yu. Moussa H. A. M., Optimal Stabilization of a Multi-Machine System, *IEEE Trans.*, PAS-91, 1174—1182 (1972).

II-7. Производственные и технологические процессы

Г. Блау¹⁾, Р. Климпел²⁾

ВВЕДЕНИЕ

Данная глава посвящена вопросам применения методов исследования операций для анализа производственных и технологических процессов. Прежде чем приступить к обсуждению формальной постановки задачи, необходимо уяснить ее природу и специфику. Без четкого понимания содержания термина «управление производственным процессом» вряд ли можно заниматься исследованием конкретного процесса.

Согласно толковому словарю, слово «процесс» означает «последовательность действий, направленных на достижение некоторой цели, которая может быть сформулирована как достижение определенного состояния или получение определенных свойств». В соответствии с этим определением «управление производственным процессом» рассматривается как управление последовательностью технологических операций, конечная цель которых состоит в получении желаемого продукта (или нескольких видов продукции) при минимальных денежных, временных или трудовых затратах. Очевидно, что для системного анализа данных производственного процесса все входящие в него технологические операции должны быть полностью конкретизированы.

Результатом любой технологической операции, как правило, является превращение одного объекта (или совокупности таких объектов), обладающего определенными химическими и физическими свойствами, в другой объект с иными свойствами. Так как технологические операции реализуются на основе фундаментальных законов химии, физики, биологии и т. д., то теоретически можно создать исчерпывающую математическую модель исследуемого процесса с учетом законов природы, которым подчиняется каждая из составляющих его операций. После этого можно использовать методы исследования операций для анализа совокупности частных моделей, входящих в состав обобщенной модели исследуемого процесса. Обычно подробные модели технологических операций содержат большое число переменных, носят ярко выраженный нелинейный характер, нередко записываются в виде дифференциальных уравнений и могут включать в себя соотношения, в которых фигурируют величины, имеющие вероятностный характер.

В общем виде любая технологическая операция (n -я операция из множества N различных операций) может быть представлена в виде простой схемы (рис. 1). Обычно каждая операция характеризуется переходной функцией $\mathcal{T}_n$, которая является функцией управляющих переменных и описывает переход системы из одного состояния в другое как результат технологического процесса. При этом начальное и конечное состояния системы описываются с помощью переменных состояния x_n и y_n соответственно. Переход системы из одного состояния в другое может представлять собой, например, химическое превращение одной совокупности веществ в другую, изменение степени полимеризации того или иного вещества и т. д., а управляющими

¹⁾ Gary E. Blau, Ph. D., Senior Research Specialist in Mathematical Applications, Central Research at Dow Chemical Co. in Midland, Michigan.

²⁾ Richard R. Klimpel, Ph. D., Research Manager of Mathematical Applications, Central Research at Dow Chemical Co. in Midland, Michigan.

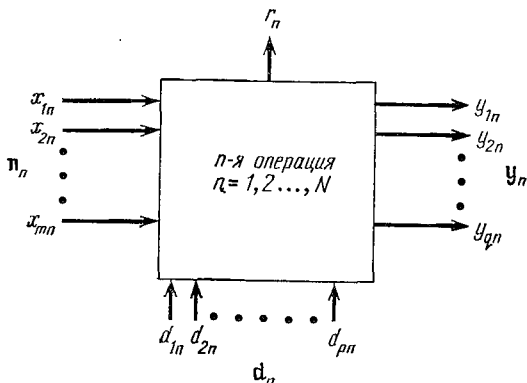


Рис. 1. Схематическое представление технологической операции. $y_n = \mathcal{T}_n(d_n, x_n)$, $r_n = R_n(d_n, x_n)$. (На рисунке вместо n_n должно быть x_n).

переменными могут быть температура, продолжительность реакции, давление и т. д. Помимо переходной функции для описания каждой операции используется функция R_n (норма прибыли, чистый доход или иной экономический показатель).

Различие технологических операций, их конечных целей и стоимости обуславливает разнообразие в постановке задач исследования производственных и технологических процессов. Вопросы классификации такого рода задач будут рассмотрены достаточно подробно несколько позднее.

На практике мы никогда не располагаем моделью, которая описывала бы полностью всю совокупность элементов процессов, и, по-видимому, такую модель вряд ли можно построить при разумных временных, стоимостных и других ограничениях (а возможно, просто из-за отсутствия требуемых знаний). Поэтому на данном этапе основной целью является изучение возможностей построения приближенной математической модели технологического процесса.

Возможность построения подобной модели зависит от многих факторов, в том числе и таких, как современные теоретические и инженерные представления об исследуемых технологических операциях, способы получения данных о реализации того или иного процесса (включая измерения и опрос исполнителей, ответственных за проведение данного процесса), цели исследования (разработка новой или анализ возможностей усовершенствования существующей технологии), ограничения ресурсов, финансов, времени и т. д., наличие и возможность практического использования подходящих математических методов (обеспечивающих решение задачи для моделей различного типа и различной размерности), степень подготовленности соответствующих специалистов и согласованность их действий. Поскольку полезность модели с точки зрения ее практического применения определяется тем, насколько полно учтено взаимодействие всех перечисленных факторов, то именно этому вопросу уделяется основное внимание в данной главе. На примере наиболее типичных и представляющих практический интерес задач иллюстрируется эффективность использования для их решения методов исследования операций, причем вопросы, связанные с реализацией алгоритмов решения, обсуждаются менее подробно. Сначала рассматриваются простейшие детерминированные линейные (или квазилинейные) модели технологических операций, предназначенные для решения задач планирования и/или анализа их практической реализуемости. Далее рассматриваются более сложные модели (как правило, нелинейного типа), предназначенные для решения задач управления

реальными технологическими процессами и/или их проектирования, а также методы построения нелинейных моделей со случайными параметрами.

Вопросам применения методов исследования операций для анализа и оптимизации различных производственных и технологических процессов посвящено большое число работ, среди которых следует отметить [1, 21—24 и 31]. В данной главе основное внимание уделяется методам математического программирования, поскольку применение других методов исследования операций для анализа подобных процессов вызывает существенные затруднения.

Фирмы могут по-разному использовать методы исследования операций, даже если речь идет о почти аналогичных технологических процессах. Помимо вышеуказанных факторов вероятность успешного применения в той или иной сфере методов исследования операций и преодоления организационных трудностей существенным образом зависит от позиции, занимаемой руководством фирмы.

Во многих промышленных фирмах существующее отношение к исследованию операций частично определяется результатами применения этих методов специалистами, привлеченными фирмой к работе в конце 50-х — начале 60-х годов. Большинство из них не имело практических навыков использования методов исследования операций. Именно в силу этого обстоятельство первоначальные попытки ряда фирм получить положительный эффект за счет применения методов исследования операций не увенчались успехом. Однако аналогичные действия следующего поколения руководства промышленных фирм оказались более успешными, и, по-видимому, можно утверждать, что в настоящее время почти все промышленные фирмы используют по крайней мере отдельные методы, обсуждаемые в данной главе (в частности, линейное программирование).

Следует иметь в виду, что не всегда можно непосредственно оценить ту практическую пользу, которая может быть получена (или должна быть получена) фирмой за счет применения методов математического моделирования в производственной и технологической сфере. Однако личный опыт авторов данной главы свидетельствует о том, что применение методов оптимизации приводит к существенному усовершенствованию как производственных, так и технологических процессов, эффект от которого оценить достаточно просто. Более того, даже самые неблагоприятные результаты анализа позволяют лучше понять исследуемый технологический процесс, степень его чувствительности к изменениям тех или иных факторов и отчетливо представить возможные варианты решений рассматриваемой проблемы.

1. ЗАДАЧИ РАСПРЕДЕЛЕНИЯ РЕСУРСОВ ПРИ НЕПРЕРЫВНОМ ПРОИЗВОДСТВЕ

В условиях непрерывного производства необходимость решения задачи рационального распределения ресурсов возникает почти ежедневно. Классическим примером является задача размещения пунктов отгрузки готовой продукции потребителям, причем постановка задачи зависит от конкретных условий. Минимизация затрат на доставку потребителям готовой продукции обычно рассматривается в комплексе с другими задачами, имеющими непосредственное отношение к управлению поставками, например составлением производственно-технологического графика, получением необходимых химических материалов для составления смесей, хранением запасов (рациональным распределением складских

помещений) и т. д. Довольно часто материально-стоимостные связи в задачах распределения хорошо описываются с помощью линейных соотношений, что позволяет для их решения использовать методы линейного программирования. В тех случаях, когда приходится иметь дело с нелинейными соотношениями, возникает необходимость применения методов нелинейного программирования, в частности квадратичное и сепарабельное программирование.

1.1. Анализ распределения ресурсов в угольной промышленности

Типичная задача распределения ресурсов в угольной промышленности рассматривается в работе [29] применительно к условиям добычи угля открытым способом в районе Центральной Пенсильвании (США). Главными потребителями угля в этом районе являются теплоэлектростанции, функционирование которых возможно при гарантированных поставках больших объемов угля с относительно однородностью состава. Проблема в данном случае состоит в том, что объем добычи угля в большинстве карьеров не настолько значителен, чтобы можно было говорить об экономической выгоде крупномасштабных перевозок. Значительно более важным является то, что уголь, получаемый в разных участках региона, различается по своему качеству, что существенно усложняет решение проблемы распределения и затрудняет эффективное управление процессами добычи угля в различных угольных карьерах. Чтобы добиться максимальной практической отдачи и стабильности добычи угля в условиях конкуренции с внешними поставщиками данного вида топлива, можно, например, объединить разрозненные ресурсы и организовать централизованное управление уже действующими технологическими циклами и таким образом избежать технического переоснащения угольных карьеров.

Следовательно, задача сводится к определению таких уровней добычи (например, в течение суток), которые обеспечивали бы максимизацию чистой прибыли фирмы с учетом стоимости угля, определяемой в соответствии с его теплотой сгорания, выраженной в БТЕ ($1 \text{ БТЕ} = 0,252 \text{ ккал}$), затратами на хранение и доставку угля потребителю. Кроме того, поставляемый уголь должен обладать надлежащим качеством.

Обозначим через d_{in} количество угля, отправляемого с i -го карьера (при залежах на большой глубине из i -й шахты) в n -й пункт хранения, и предположим, что в пределах рассматриваемого региона имеется N пунктов централизованного хранения угля, добываемого в различных карьерах (или шахтах), число которых равняется p . Введем следующую систему ограничений: количество

угля, отправляемого с i -го карьера, не должно превышать суточной добычи на данном карьере, т. е.

$$\sum_{n=1}^N d_{in} \leq q_i, \quad i = 1, 2, \dots, p; \quad (1)$$

суммарное количество угля, поступающего со всех карьеров в пункты централизованного хранения, должно равняться суммарному объему ежедневных поставок потребителю (ежедневная потребность в угле предполагается известной), т. е.

$$\sum_{i=1}^p \sum_{n=1}^N d_{in} = y. \quad (2)$$

Если a_i — показатель, характеризующий относительное содержание золы в угле из i -го карьера, и если a — предельно допустимое содержание золы в угольной смеси, то необходимо, чтобы

$$\sum_{i=1}^p \sum_{n=1}^N a_i d_{in} \leq ay. \quad (3)$$

Аналогичные соотношения могут быть получены для относительного содержания в угле таких компонентов, как сера, влага, летучие вещества, шлаки после сгорания и связанный углерод:

$$\begin{aligned} \sum \sum s_i d_{in} &\leq sy && \text{(сера),} \\ \sum \sum w_i d_{in} &\leq wy && \text{(влага),} \\ \sum \sum v_i d_{in} &\geq vy && \text{(летучие вещества),} \\ \sum \sum f_i d_{in} &\geq fy && \text{(шлаки после сгорания),} \\ c_{iy} &\leq \sum \sum c_i d_{in} \geq c_{uy} && \text{(связанный углерод).} \end{aligned} \quad (4)$$

Требование неотрицательности d_{in} выражается неравенством

$$d_{in} \geq 0, \quad i = 1, 2, \dots, p, \quad n = 1, 2, \dots, N. \quad (5)$$

Целевая функция, подлежащая максимизации, записывается в следующем виде:

$$\sum_{i=1}^p \sum_{n=1}^N (\lambda_i - \gamma_{in}) d_{in}, \quad (6)$$

где λ_i — цена (включая затраты на доставку) на уголь, доставляемый из i -го карьера в пункты хранения, а γ_{in} — транспортные расходы, связанные с перевозкой тонны угля от i -го карьера до n -го пункта централизованного хранения.

Следует отметить, что данная модель неоднократно использовалась для определения оптимальных вариантов закрепления за карьерами пунктов централизованного хранения каменного угля, расположенных в пределах региона.

2. ЗАДАЧИ СОСТАВЛЕНИЯ СМЕСЕЙ

Многие технологические процессы предусматривают выполнения операций по приготовлению смесей как с целью создания благоприятных условий для протекания определенных химических реакций, так и с целью получения продуктов, отвечающих установленным фирмой требованиям. Математическим моделям составления смесей чаще всего придают линейный вид, причем связи между переменными модели считаются полностью детерминированными. Использовать такого рода модели следует крайне осторожно, поскольку физические закономерности процессов смешения описываются существенно нелинейными функциями. Однако если эти функции являются выпуклыми, соответствующая модель оказывается легко линеаризуемой. Иногда для адекватного описания исследуемого процесса необходимо прибегнуть к помощи квадратичного или сепарабельного программирования. Типичные подходы к решению задачи составления химической смеси можно проиллюстрировать на следующих примерах.

2.1. Приготовление гранулированных смесей

Рассмотрим задачу приготовления смеси различных химически однородных компонентов, каждый из которых используется в гранулированном виде, при этом необходимо, чтобы полученная смесь обладала некоторыми заданными свойствами [22]. (Обсуждаемая ниже постановка задачи допускает статистический разброс исходных данных.)

Исходный продукт для приготовления V ед. веса смеси представлен p компонентами, каждый из которых имеет k -ю характеристику распределения частиц по размерам и общий вес z_k . Весовые доли частиц компонента k , принадлежащих каждому из m интервалов размеров частиц, равны $a_{1k}, a_{2k}, \dots, a_{mk}$ ($\sum_{i=1}^m a_{ik} = 1, 0$), а для смеси — $b_1, b_2, \dots, b_m$ ($\sum_{i=1}^m b_i = 1, 0$). Результат технологической операции, обозначаемый через y_i (рис. 1), характеризует вес частиц в смеси, имеющих размеры, соответствующие i -му интервалу, т. е. $y_i = V \cdot b_i$. Каждая из управляющих переменных x_j представляет собой количество p -го компонента, которое используется для приготовления смеси. Критерием качества под-

лежащей приготовлению химической смеси будем считать сумму абсолютных значений погрешностей, обусловленных необходимостью сохранения материального баланса в процессе смешения. Теперь задача может быть сформулирована следующим образом:

найти

$$\min \sum_{i=1}^m (u_i + v_i) \quad (7)$$

при ограничениях

$$a_{i1}d_1 + a_{i2}d_2 + \dots + a_{ip}d_p + u_i - v_i = y_i, \quad i = 1, 2, \dots, m. \quad (8)$$

[u_i и v_i — слабые переменные, характеризующие положительное и отрицательное отклонения в ограничениях (8)]. Условие постоянства количества вещества запишем в виде

$$\sum_{j=1}^p d_j = V, \quad (9)$$

где

$$d_j \leq z_j, \quad j = 1, 2, \dots, p, \quad (10)$$

показывает, что ни одно из смешиваемых химических веществ не может быть взято в количестве большем, чем то, которым мы фактически располагаем. Кроме того, следует ввести ограничения, отражающие требование неотрицательности d_j , u_j и v_j , т. е. $p + 2m$ стандартных ограничений вида

$$\begin{aligned} d_j &\geq 0 \quad (j = 1, 2, \dots, p), \\ u_i &\geq 0, \quad v_i \geq 0 \quad (i = 1, 2, \dots, m), \end{aligned} \quad (11)$$

а если необходимо, то и ограничение вида

$$\sum_{i=1}^m u_i - \sum_{i=1}^m v_i = 0. \quad (12)$$

Решение поставленной задачи позволяет определить также значения управляющих переменных d_j ($j = 1, 2, \dots, p$), которые обеспечивают наименьшее отклонение качества химической смеси (приготовление которой предусматривается в технологическом процессе) от идеального, что позволяет найти оптимальные по критерию (7) значения управляющих переменных рассматриваемой технологической операции.

Другая модель, назначение которой аналогично рассмотренной модели, была построена на основе минимаксного критерия.

В этом случае задача формулируется следующим образом:
найти

$$\min \gamma \quad (13)$$

при ограничениях

$$\begin{aligned} a_{i1}d_1 + a_{i2}d_2 + \dots + a_{ip}d_p + \gamma &\geq y_i, \\ a_{i1}d_1 + a_{i2}d_2 + \dots + a_{ip}d_p - \gamma &\leq y_i, \\ i &= 1, 2, \dots, m. \end{aligned} \quad (14)$$

Помимо $2m$ ограничений (14) остаются в силе сформулированные выше ограничения (9) и (10). Кроме того, согласно требованию неотрицательности d_j и γ , имеем

$$\gamma \geq 0, d_j \geq 0 \quad (j = 1, 2, \dots, p). \quad (15)$$

При анализе технологических операций по приготовлению химических смесей нередко приходится одновременно решать задачу составления графика загрузки различных видов оборудования, используемого в рассматриваемом технологическом процессе. Так, например, в работе [25] предлагается использовать целочисленное программирование для оптимизации графиков работы дробильных установок, используемых для подготовки материалов перед приготовлением смесей.

2.2. Задача приготовления смеси для получения сплавов металлов

В связи с обсуждением технологических операций по приготовлению смесей различных химических компонентов следует упомянуть ставшую уже классической модель линейного программирования, предназначенную для оптимизации процесса составления смеси с целью получения сплавов металлов. При этом речь идет об операции, состоящей в загрузке печи. Задача заключается в определении наиболее выгодного с эконоимической точки зрения способа использования сырья при загрузке печи с целью получения алюминиевых сплавов, отвечающих спецификациям по химическому составу и физическим свойствам, когда известно, какими количествами различных видов сырья мы располагаем, каков химический состав каждого из имеющихся сырьевых материалов и сколько стоит единица их массы.

Управляющими переменными в этой задаче являются веса (в килограммах) имеющегося в наличии сырья, которое может быть использовано для приготовления смеси. При этом в качестве целевой функции берется полная стоимость сырья, линейно зависящая

от упомянутых выше управляющих переменных, и вводятся ограничения на количество химических компонентов, составляющих сплав, на имеющееся в наличии сырье и на объем загрузки печи, который не должен быть ниже заданного минимального значения. В качестве аналогичного примера можно привести предложенную в работе [20] модель, относящуюся к выплавке алюминиевых сплавов.

3. ЛИНЕЙНЫЕ МОДЕЛИ ТЕХНОЛОГИЧЕСКИХ ОПЕРАЦИЙ

В предыдущих разделах данной главы была дана краткая характеристика технологических процессов, анализ которых возможен с помощью методов линейного программирования. Рассмотренные модели были разработаны для анализа частных задач, при постановке которых не учитывалась специфика тех технологических операций, при реализации которых возникают эти задачи. Так, например, в приведенных задачах не фигурируют такие понятия, как химическая реакция, технологическое оборудование и т. п. Детализированные модели технологических операций, как правило, носят нелинейный характер. Однако для практических целей нередко требуются и оказываются вполне достаточными приближенные модели. В данном разделе будет рассмотрена конкретная технологическая операция, представленная в виде линейной модели. Технологические операции, аналогичные той, которую мы собираемся здесь рассмотреть, как правило, являются составными частями некоторого сложного производственного процесса. Таким образом, автономное моделирование той или иной технологической операции ориентировано на поиски некоторого субоптимального решения, которое должно учитываться в схеме оптимизации производственного процесса в целом. При этом анализ производственной или технологической операции в целом осуществляется либо строгими количественными методами, либо на чисто экспертной основе (если количественный анализ оказывается по той или иной причине затруднительным).

3.1. Задачи минимизации затрат при производстве чугуна

К числу моделей простой технологической операции можно отнести линейную модель производственного планирования операции выплавки чугуна в доменной печи. Эта модель учитывает требования материального баланса, содержит ограничения, вытекающие из необходимости поддерживать в процессе выплавки чугуна химическое равновесие, и предназначена для минимизации суммарной стоимости используемых материалов [17]. Модель позволяет осуществлять распределение некоторых исходных материалов для получения четырех видов конечного продукта (метал-

ла, шлака, колосникового газа и пыли) и рассчитывать потребности в коксе, флюсах и воздухе. Управляющие переменные обозначим через d_j ($j = 1, 2, \dots, p$), где d_j — количество используемого в технологической операции материала j -го наименования, причем материал каждого вида состоит либо из нескольких химических элементов, либо из нескольких химических соединений.

Система ограничений обсуждаемой модели включает в себя соотношения, связанные с химическим балансом масс различных компонентов, а также с учетом различного рода эмпирически установленных соотношений между факторами, влияющими на исход рассматриваемой технологической операции, и целого ряда других обстоятельств. Из-за сложности модели, предложенной в работе [17], мы не можем представить ее здесь полностью и поэтому рассмотрим лишь отдельные виды ограничений, что кажется вполне достаточным для уяснения сути дела. Тем, кто заинтересуется полной структурой данной модели, рекомендуем обратиться непосредственно к работе [17]. Здесь мы подчеркнем лишь, что в рамках модели процесса выплавки чугуна необходимо учитывать баланс большого числа химических элементов. Так, например, если q_j — относительное содержание j -го химического компонента в расплавленной массе металла, а z_j — доля фосфора, входящего в состав компонента j -го вида, то суммарное количество фосфора в единичной плавке определяется из соотношения

$$y = \sum_{j=1}^p q_j z_j d_j, \quad (16)$$

Ясно, что в реальных условиях могут возникнуть ограничения, обусловленные недостатком некоторых из требуемых материалов или иными обстоятельствами.

Многие другие ограничения, связанные с расходом кокса, обусловлены потребностями в тепловой энергии. Для построения такого типа ограничений необходимо знать, как протекают тепло-энергетические процессы в ходе различных химических реакций, каковы потери тепловой энергии в рассматриваемом процессе и т. д. Так, например, теплосодержание z расплавленной массы металла в доменной печи определяется соотношением

$$z = \sum_{j=1}^s y_j h_j, \quad (17)$$

где h_j — теплосодержание, приходящееся на единицу массы j -го материала, а s — число компонентов ($s < p$), вошедших в состав плавки. Одним из материалов, необходимых для протекания химической реакции, является кислород. Его расход должен тщательно контролироваться. Если t_j — процентное содержание кислорода в сырье j -го наименования, то суммарная масса кисло-

рода (О), используемого для выплавки, с учетом кислорода, поступающего в домну за счет дутья, определяется как

$$y_0 = \sum_{j=1}^n t_j d_j + 0,23 d_{\text{дутья}}. \quad (18)$$

Количество воздуха, вдуваемого в домну в процессе выплавки, также ограничено как мощностью обогревательного и воздушодувного оборудования, так и тем, что состав смеси газов (включаящей воздух) должен удовлетворять вполне определенным требованиям.

Подлежащая минимизации целевая функция представляет собой суммарную стоимость сырья, используемого для выплавки чугуна, т. е.

$$\sum_{k=1}^v \lambda_k d_k, \quad (19)$$

где v — число видов сырьевых материалов, которые нужно закупить ($v < p$), а λ_k — стоимость сырьевого материала k -го вида, отнесенная к единице его измерения (тонна, кг и т. п.). Минимизация целевой функции (19) осуществляется с учетом упомянутых ограничений, а также при условии неотрицательности управляющих переменных, т. е.

$$d_i \geq 0, \quad i = 1, 2, \dots, p. \quad (20)$$

В результате решения описанной выше задачи минимизации получаем оптимальный по стоимости режим использования сырьевых материалов для выплавки чугуна, физико-химические свойства которого (так же как и объем единичной выплавки) нам заданы. Целевую функцию рассмотренной модели можно модифицировать, учитывая суммарные эксплуатационные расходы и доход, получаемый от выплавки чугуна.

4. ЛИНЕЙНЫЕ МОДЕЛИ РЯДА ОДНОВРЕМЕННО ВЫПОЛНЯЕМЫХ ТЕХНОЛОГИЧЕСКИХ ОПЕРАЦИЙ

В данном разделе мы рассмотрим несколько постановочных задач, иллюстрирующих методику анализа технологического процесса в целом. Именно в связи с решением задач комплексного анализа сложных технологических процессов в полной мере начинают раскрываться возможности методов исследования операций. Комплексный анализ одновременно выполняемых технологических операций во многих случаях сопряжен с серьезными трудностями как математического, так и технического характера. Это обусловлено прежде всего большой размерностью такого рода задач (поскольку приходится иметь дело сразу со многими моделя-

ми, в которых фигурирует большое число переменных). Кроме того, как правило, возникают проблемы организационного характера, связанные, например, с необходимостью изменения традиционно сложившихся организационных связей, с учетом отношения руководящих работников фирмы к математическому моделированию технологических процессов и т. п. Если говорить о математической стороне дела, то можно отметить, что в связи с комплексным анализом технологических операций наибольшее распространение получили методы линейного программирования, что объясняется доступностью и простотой использования программ для ЭВМ, реализующих эти методы. Однако при использовании линейных аппроксимаций для описания сложных процессов необходимо соблюдать определенную осторожность. Оценивая степень адекватности комплекса моделей с точки зрения его целевого назначения, необходимо проанализировать каждую технологическую операцию в отдельности. Только после этого можно ставить вопрос о комплексном анализе одновременно выполняемых технологических процедур. Без такого рода предосторожностей нетрудно прийти к физически бессмысленным результатам. Поэтому при анализе реальных технологических процессов главная проблема состоит в том, чтобы система линейных моделей учитывала известные нелинейные зависимости физического и/или экономического происхождения.

Прежде чем перейти к рассмотрению примеров, иллюстрирующих постановку задачи моделирования одновременно выполняемых технологических операций, желательно хотя бы кратко затронуть вопрос об оценке получаемых оптимальных решений. Результаты анализа чувствительности решения путем введения в рассмотрение скрытых цеп и/или путем расположения коэффициентов, входящих в выражение для целевой функции по степени их значимости, являются неотъемлемой частью многих вычислительных алгоритмов линейного программирования. То обстоятельство, что для линейных моделей проверка чувствительности получаемых оптимальных решений к изменению значений параметров моделей не представляет сколько-нибудь серьезных трудностей, служит дополнительным доводом в пользу применения методов линейного программирования при комплексном анализе технологических операций. К сожалению, анализ чувствительности сопряжен со значительными трудностями в тех случаях, когда в модель включены нелинейные соотношения или требуется, чтобы все или, по крайней мере, некоторые переменные принимали лишь целочисленные значения.

Авторы данной главы на собственном опыте убедились в том, что использование скрытых цеп и расположение коэффициентов, входящих в выражение для целевой функции, по степени их значимости играют при моделировании технологических процессов

исключительно важную роль. Многие специалисты этим приемом анализа, к сожалению, часто пренебрегают.

В производственной практике лишь в отдельных случаях удастся внедрить в полном объеме результаты решения задачи оптимизации. Это объясняется целым рядом причин: трудностями, связанными с техническим обеспечением технологического процесса в том виде, в каком он представлен моделью; несовместимостью полученного решения с общей технической политикой предприятия; неточностью задания численных значений коэффициентов модели и т. д. Поэтому весьма важно выяснить, как зависит оптимальное решение от численных значений параметров модели (в частности, от значений констант в правых частях ограничений и от коэффициентов в выражении для целевой функции). При решении большинства задач оптимизации технологических процессов требуется выявить область целесообразной оптимальности. Без этого вероятность успешного использования методов исследования операций на практике существенно снижается. Многие специалисты в области оптимизации технологических процессов применяют имитационный подход (см., например, [22]), так как использование имитационных моделей часто позволяет более адекватно описать исследуемые процессы и исследовать чувствительность полученного оптимального решения. Результаты такого анализа могут быть использованы при решении самых разнообразных практических задач: при установлении цен на сырье и готовую продукцию, при оптимизации производственных мощностей промышленных предприятий и определении характеристик складских помещений, выявлении потребностей в рабочей силе, выработке требований к контролю качества выпускаемой продукции и т. п.

Пример. Задача выбора объема капиталовложений по комплексу переработки минерального сырья [27]. В этой задаче нужно определить объем необходимых капиталовложений и выбрать оптимальные производственные мощности технологических агрегатов химического комбината, перерабатывающего минеральное сырье. При рассмотрении названной задачи мы сразу сталкиваемся с двумя специфическими проблемами, общими почти для всех многостадийных технологических процессов: с целочисленными переменными и сочетанием выпуклых и невыпуклых функций, описывающих эти процессы. Представление процессов в целочисленных переменных возникает в оперативно-календарном планировании работы химических реакторов, которые могут либо функционировать, либо простаивать, в задачах определения капитальных затрат на пуско-наладочные работы и т. п. Нелинейные функциональные соотношения описывают процессы химической кинетики, составления смесей, поддержания соотношения жидких потоков, повышение эффективности капитальных затрат в промышленности и т. д.

Так, например, постоянные капиталовложения в одну установку меньше вложений в две установки, производственная мощность каждой из которых составляет половину от мощности первой установки. Таким образом, имеет место следующее соотношение:

$$\text{Капитальные затраты} = aV^b, \quad (21)$$

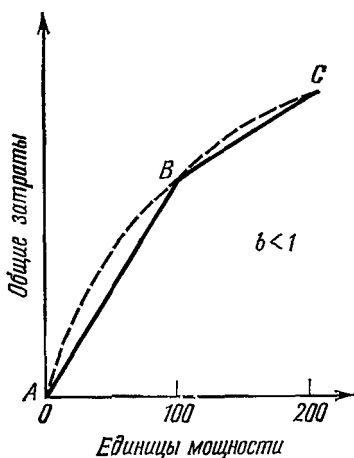


Рис. 2. Аппроксимация выпуклой функции стоимости.

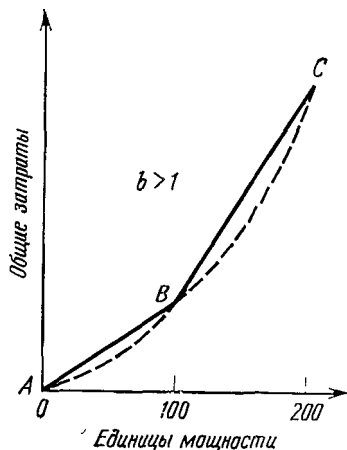


Рис. 3. Аппроксимация вогнутой функции стоимости.

где V — мощность, a — коэффициент затрат, b (для химических установок) лежит в пределах 0,5—0,8. На рис. 2 показан пример линеаризации функции затрат, которая потребовалась бы в этом случае. На рис. 3 приведен пример, когда для отыскания оптимального решения приходится многократно решать задачи линейного программирования (в точках B , C и т. д.) либо использовать методы сепарабельного линейного программирования или нелинейного программирования.

Рассмотрим проблему оценки капиталовложений в новое предприятие (рис. 4) по переработке крупных запасов минерального сырья x_{11} , трудно поддающегося обработке. Поскольку целью анализа является определение рациональных объемов капиталовложений в строительство такого предприя-

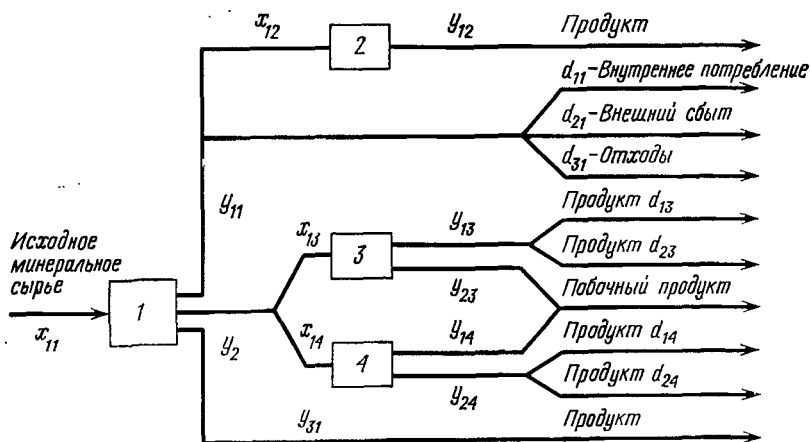


Рис. 4. Схема комплексной переработки минерального сырья.

тия, соотношения, описывающие модель, максимально упрощены. Основной процесс на установке 1 дает три потока продукции — y_{11} , y_{21} и y_{31} , связанные с x_{11} стационарными функциями материального баланса

$$y_{11} = 41,62x_{11}, y_{21} = 5,59x_{11}, y_{31} = 0,555x_{11}. \quad (22)$$

Сбыт продукта y_{31} не ограничен, в то время как на часть продукта y_{11} , где

$$y_{11} = (d_{11} + d_{21} + d_{31} + x_{12}), \quad (23)$$

наложены следующие ограничения по сбыту:

$$d_{11} \leq 1200, d_{12} \leq 3000. \quad (24)$$

Продукт y_{11} может быть также использован (полностью или частично) в установке 2, выпускающей продукт y_{12} .

Продукт y_{21} , состоящий из двух компонентов

$$y_{21} = x_{13} + x_{14}, \quad (25)$$

может быть использован для производства продуктов:

$$y_{13} \text{ — в установке 3 и } y_{24} \text{ — в установке 4.}$$

Уравнения материального баланса для этих продуктов записываются в следующем виде:

$$y_{13} = 0,855x_{13}, y_{24} = 2,62x_{14}. \quad (26)$$

Предполагается, что продукты y_{24} и y_{13} поступят на два рынка: внешний (d_{14} и d_{13}) и внутренний (d_{24} и d_{23}), где на единицу продукта будут получены существенно различные доходы:

$$y_{24} = d_{14} + d_{24} \text{ и } y_{13} = d_{13} + d_{23}. \quad (27)$$

На сбыт этих продуктов наложены следующие ограничения:

$$d_{13} \leq 300, d_{14} \leq 250, d_{24} \leq 300. \quad (28)$$

Спрос на продукт d_{24} отличается значительной изменчивостью. Для переработки продукта, идущего в отход ($y_{23} + y_{24}$), требуется дорогостоящее оборудование, причем затраты на удаление отходов резко возрастают при росте их количества. Образование отходов определяется соотношениями

$$y_{23} = 4010 x_{13}, y_{14} = 1340 x_{14}. \quad (29)$$

Поскольку ожидается, что рост спроса и изменение цен во времени будут умеренными из-за большого запаса исходного сырья, предполагается, что строящиеся установки рассчитаны на полную мощность и сразу же будут функционировать при этой мощности. Принято также предположение, что установки морально устаревают в течение десяти лет. Поэтому оценка всего проекта проведена для периода в десять лет. Потребность в постоянном и оборотном капитале для установки n мощностью V_n описывается выражением

$$\text{Общий объем капитала для установки } n = (a_n) + b_n V_n + c_n V_n^{q_n}. \quad (30)$$

Годовые эксплуатационные затраты для установки n приближенно определяются как

$$(e_n) + f_n V_n + g_n V_n^2 + h_n V_n^{q_n}. \quad (31)$$

Константы a_n и e_n определяют постоянные затраты на пуско-наладочные работы, которые производятся в период пуска установки в эксплуатацию. Доходы представлены в виде линейных функций, за исключением доходов

от продукта d_{24} , как уже упоминалось ранее. Денежная прибыль подсчитывается как алгебраическая сумма годовых доходов и расходов, которая приводится к начальному моменту времени. Полученные результаты суммируются для десятилетнего периода, и в итоге получается алгебраическая приведенная функция чистого дохода.

Таким образом, задача сводится к максимизации функции дохода, учитывающей различные характеристики продукции и оборудования, о которых говорилось выше:

$$\begin{aligned} \max R = & -16\,000x_{11}^{0,9} - 1950x_{11} + 0,17x_{11}^2 + \\ & + 840d_{11} + 672d_{21} + 100y_{31} - \\ & - 24\,000y_{12}^{0,8} + 7000y_{12} - 3\,000\,000y_{12}^* - \\ & - 38,000y_{13}^{0,85} - 12\,000y_{13} + 1,5y_{13}^2 + \\ & + 51\,000d_{13} + 26\,000d_{23} - 1\,000\,000y_{13}^* - \\ & - 12\,000y_{24}^{0,87} - 26\,000y_{24} + 54\,000d_{14} + \\ & + 47\,500d_{24} - 63,3d_{24}^2 - 2\,000\,000y_{24}^* - \\ & - 0,75(y_{23} + y_{14})^{1,05} - 0,33(y_{23} + y_{14}) \end{aligned}$$

при ограничениях на технологию и сбыт, а также при условии неотрицательности всех переменных. Переменные y_{12}^* , y_{13}^* и y_{14}^* характеризуют пусконаладочные работы, принимают значения, равные либо нулю, либо единице ($y_{12}^* = 0$, если $y_{12} = 0$, и $y_{12}^* = 1$, если $y_{12} > 0$ и т. д.). Эта задача была решена с помощью алгоритма нелинейного программирования при линейных ограничениях [27], а позднее с помощью метода целочисленного программирования [26].

Такая постановка задачи позволяет получить ответы на ряд вопросов. Например, целесообразно ли с точки зрения экономики производства вообще осуществлять предложенный проект? Если ответ на этот вопрос положителен, то каков должен быть объем переработки сырья x_{11} ? Какие дополнительные перерабатывающие установки следует соорудить и каковы должны быть их производственные мощности?

Пример. Оперативно-календарное планирование работы нефтеперерабатывающего завода [19]. Нефтеперерабатывающий завод состоит из ряда установок для разделения, преобразования и смешения различных фракций сырой нефти. В рассматриваемом случае фигурируют только две установки: нефтеперегонный аппарат для первичной перегонки и теплообменник для каталитического крекинга нефти. Нефтеперегонный аппарат по существу представляет собой ректификационную колонну, в которой сырая нефть разделяется на три фракции: легкую нефть, средний дистиллят¹⁾ и тяжелый остаток (мазут). При каталитическом крекинге происходит переработка среднестиллятной фракции химическим путем на три конечных продукта: нафта, легкие и тяжелые масляные фракции. Два последних продукта возвращаются в крекинг-установку для повторной перегонки. Термины нафта и дистилляты прямой отгонки относятся к фракциям, которые в дальнейшем не перерабатываются, а используются непосредственно в качестве топлива.

С помощью этих двух процессов получают семь фракций, которые можно смешивать, в результате чего образуются четыре конечных продукта (табл. 1). Для облегчения понимания задачи в рассматриваемом примере используется мнемоническое обозначение переменных.

¹⁾ Дистиллят — печное топливо.

Таблица 1

**Мнемонические обозначения фракций перегонки нефти
нефтеперерабатывающего завода**

CR1F	Сырая нефть 1, перерабатываемая на топливо, MB/D
CR1D	Сырая нефть 1, перерабатываемая на дистиллят, MB/D
CR2F	Сырая нефть 2, перерабатываемая на топливо, MB/D
CR2D	Сырая нефть 2, перерабатываемая на дистиллят, MB/D
CATVD	Дистиллят после прямой отгонки, MB/D
CATLC	Легкая масляная фракция для каталитической переработки, MB/D
CATHC	Тяжелая масляная фракция для каталитической переработки, MB/D
RSFUL	Остаточное топливо (мазут), MB/D
HCFUL	Тяжелые масляные фракции для получения мазута, MB/D
LCFUL	Легкие масляные фракции для получения мазута, MB/D
VDFUL	Дистиллят после прямой отгонки, MB/D
VNAPD	Нафта после прямой отгонки для получения дистиллята (печного топлива), MB/D
VDISD	Дистиллят после прямой отгонки в качестве печного топлива, MB/D
LCCYD	Легкие масляные фракции для получения печного топлива, MB/D
VN1R	Нафта (из сырой нефти 1) после прямой отгонки для получения обычного бензина, MB/D
VN2R	Нафта (из сырой нефти 2) после прямой отгонки для получения обычного бензина, MB/D
CNPR	Нафта (из сырой нефти 1) после каталитического крекинга для получения высококачественного бензина, MB/D
VN1P	Нафта (из сырой нефти 1) для получения высококачественного бензина, MB/D
VN2P	Нафта (из сырой нефти 2) для получения высококачественного бензина, MB/D
CNPP	Нафта после каталитического крекинга для получения высококачественного бензина, MB/D

MB/D—условное обозначение тысячи баррелей в сутки (1 баррель = 158, 758—158, 989 л).

Очевидно, что рентабельность нефтеперегонного завода зависит от стоимости сырья и эксплуатационных затрат, качества процессов и цен на конечные продукты. Как и на любых перерабатывающих предприятиях, на нефтеперегонном заводе технологические процессы могут осуществляться в различных режимах в зависимости от выбора сырья, характеристик установки для крекинга, выбора соотношений фракций при составлении смесей

Таблица 2

Мнемонические обозначения технологических процессов
нефтеперерабатывающего завода

VNAP1	Материальный баланс по нефти, получаемой из сырой нефти 1 в результате прямой отгонки, МВ/Д
VNAP2	Материальный баланс по нефти, получаемой из сырой нефти 2 в результате прямой отгонки, МВ/Д
VDIS	Материальный баланс по дистилляту после прямой отгонки, МВ/Д
RESID	Материальный баланс по остаточному топливу (мазуту), МВ/Д
CATN	Материальный баланс по нефти, получаемой в результате каталитического крекинга, МВ/Д
LCCY	Материальный баланс по легкой масляной фракции, получаемой в результате каталитического крекинга, МВ/Д
HCCY	Материальный баланс по тяжелой масляной фракции, получаемой в результате каталитического крекинга, МВ/Д
CRBAL	Материальный баланс по сырой нефти (расход сырой нефти 1 равен $\frac{1}{3}$ общего расхода нефти), МВ/Д
PSCAP	Производительность ректификационной колонны, МВ/Д
CR2AV	Максимальный суточный запас сырой нефти 2, МВ/Д
TFCAT	Общая производительность установки каталитического крекинга, МВ/Д
FREQ	Потребность в мазуте, МВ/Д
DREQ	Потребность в дистилляте, МВ/Д
DSPEC	Показатели качества дистиллята (печного топлива), М СОН. В/Д
RREQ	Потребность в обычном бензине, МВ/Д
RSPEC	Показатели качества обычного бензина, М ОСТ. В/Д
PREQ	Потребность в высококачественном бензине, МВ/Д
PSPEC	Показатели качества бензина высокого сорта, М ОСТ. В/Д

МВ/Д—тысячи баррелей в сутки; М/Д—тысячи долларов в сутки; М СОН. В—данные о содержании примесей других нефтепродуктов на тысячу баррелей; М ОСТ. В/Д—октановое число на тысячу баррелей.

конечных продуктов и их номенклатуры. В табл. 2 дается перечень технологических процессов, которые схематически показаны на рис. 5. Набор ограничений и характеристики целевой функции (функции дохода) приведены в табл. 3.

Целевая функция представляет собой подлежащую максимизации прибыль и выражается в виде обычной алгебраической суммы затрат и доходов по различным составляющим продуктам процесса. Уравнения материального баланса по потоку продуктов: сумма входных потоков должна быть равна сумме выходных потоков. Так, например, ограничение VNAP1 относится к нефти, полученной из сырой нефти 1 путем прямой отгонки (рис. 5, поток

Целевая экономическая функция нефтеперерабатывающего

		Технологические операции нефтеперегонного аппарата				Технологические операции установки каталитического крекинга		
		CR1F	CR1D	CR2F	CR2D	CATVD	CATLC	CATHC
Целевая функция	PROFT	-3,0	-3,0	-3,1	-3,1	-0,1	-0,15	-0,16
Ограничения материального баланса	VNAP1	-0,1	-0,15					
	VNAP2			-0,2	-0,25			
	VDIS	-0,25	-0,4	-0,2	-0,35			
	RESID	-0,6	-0,4	-0,5	-0,3	-0,1		
	CATN					-0,7	-0,6	-0,3
	LCCY					-0,3	1	-0,7
	HCCY					-0,5	-0,5	1
	CR2AV			1	1			
Ограничения для нефтеперегонного аппарата	CRBAL	1	1	-0,5	-0,5			
	PSCAP	1	1	1	1			
Ограничения для установки каталитического крекинга	TFCAT					1,6	1,2	1,1
Ограничения по смешиванию топлива	FREQ							
Ограничения по смешиванию печного топлива	DREQ							
	DSPEC							
Ограничения по смешиванию стандартного бензина	RREQ							
	RS EC							
Ограничения по смешиванию высококачественного бензина	PREQ							
	PSPEC							

продуктов, поступающих на производство двух сортов бензина, должен быть равен потоку выходящих продуктов (сырой нефти 1 и 2)

$$0,1 \text{ CR1F} + 0,15 \text{ CR1D} = \text{VN1R} + \text{VN1P}.$$

(32)

Таблица 3

завода и наложенные ограничения

Операция по смешиванию топлива				Операции по смешиванию печного топлива			Операции по смешиванию стандартного бензина			Операции по смешиванию высококачественного бензина			RHS
RSFUL	HCFUL	LCFUL	VDFUL	VNARD	VDISD	LCCYD	VNIR	VN2R	CNPR	VN1P	VN2P	CNPP	
+2,5	+2,5	+2,5	+2,5	+4	+4	+4	+4,5	+4,5	+4,5	+5	+5	+5	
0.9			1	1	1		1	1		1	1		= 0 = 0 = 11 = 0 = 0 = -11 = 0 ≤ 75
0.1	1	1				1			1			1	
													≤ 0 ≤ 100
													≤ 46
1	1	1	1										≤ 50
				1 -5	1 -2	1 10							≥ 30 ≤ 0
							1 0	1 1	1 -7				≤ 10 ≤ 0
										1 4	1 5	1 -3	≥ 25 ≤ 0

Четыре первых коэффициента уравнений материального баланса сведены в табл. 4. Следующие три коэффициента уравнений материального баланса по рабочим потокам установки каталитического крекинга приведены в табл. 5. Так, например, баланс по выходам бензина из крекинг-установки

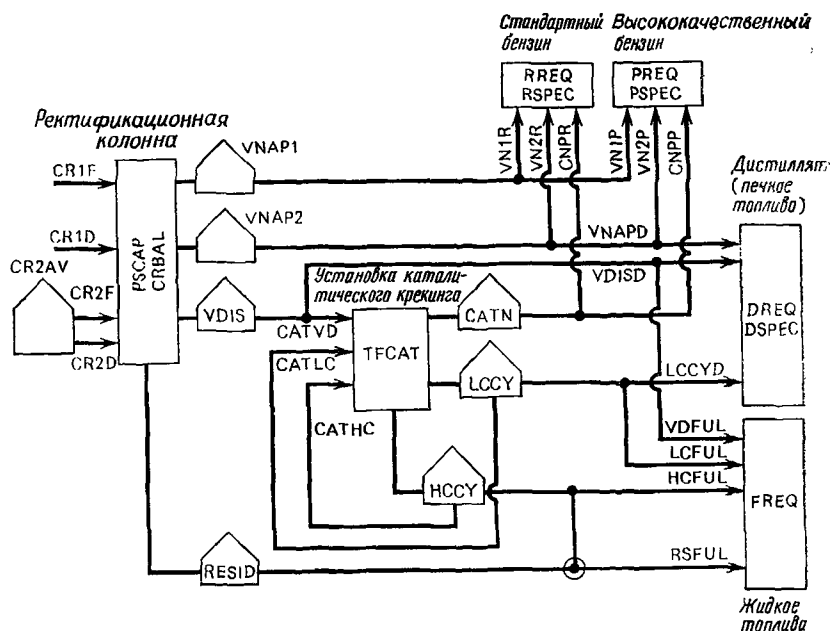


Рис. 5. Схема процесса переработки нефти [наименования, указанные в блоках, являются ограничения (строки), а наименования у стрелок — переменными (столбцы)].

Таблица 4

Выход погонных фракций на предельное количество топлива или дистиллята при работе перегонного аппарата

	Сырая нефть 1		Сырая нефть 2	
	топливо	дистиллят	топливо	дистиллят
Бензин прямой отгонки	0,10	0,15	0,20	0,25
Дистиллят прямой отгонки	0,25	0,40	0,20	0,35
Остаток	0,60	0,40	0,50	0,30

определяется соотношением

$$0,7 \text{ CATVD} + 0,6 \text{ CATLC} + 0,3 \text{ CATC} = \text{CNPR} + \text{CNPP}. \quad (33)$$

Ограничение CR2AV определяет максимальный суточный запас сырой нефти 2, составляющий 75 000 баррелей.

Таблица 5

Рабочие характеристики установки каталитического крекинга

	Дистиллят прямой отгонки, баррель	Легкая масля- ная фракция, баррель	Тяжелая масляная фрак- ция, баррель
Исходное сырье	1,1	1,0	1,0
Резиновое сырье	0,6	0,2	0,1
Нафта каталитического крекинга	0,7	0,6	0,3
Легкая масляная фракция	0,3	—	0,7
Тяжелая масляная фрак- ция	0,5	0,5	—

Ограничения, относящиеся к нефтеперегонному аппарату, заключаются в том, что расход сырой нефти 1 может составлять не более одной трети общей подачи всей сырой нефти

$$CR1F + CR1D \leq 0,5 \quad CR2F + 0,5 \quad CR2D \quad (34)$$

и что его общая производительность составляет 100 000 баррелей в сутки

$$CR1F + CR1D + CR2F + CR2D \leq 100. \quad (35)$$

Ограничением для установки каталитического крекинга также является ее максимальная производительность с учетом рециркулирующих потоков

$$1,6 \text{ CATVD} + 1,2 \text{ CATLC} + 1,1 \text{ CATHC} \leq 46. \quad (36)$$

Ограничение на максимальный выпуск жидкого топлива имеет вид

$$RSFUL + HCFUL + LCFUL + VDFUL \leq 50. \quad (37)$$

Характеристики смесей дистиллятов приведены в табл. 6. Из них

Таблица 6

Характеристики смесей дистиллятов

	Показатель загряз- нения другими нефтепродуктами	Объем
Дистиллят прямой отгонки	54	Весь продукт плюс 11 МВ/Д из запасов
Нафта прямой отгонки из нефти 2	50	Весь продукт
Легкая масляная фракция	65	Весь продукт за выче- том 11 МВ/Д
Технические требования	55	Не менее 30 МВ/Д

вытекают ограничения, указанные в табл. 3. Первое определяет минимальный объем выпуска печного топлива

$$VNAPD + VDISD + LCCYD \geq 30, \quad (38)$$

второе задает предельно допустимое значение коэффициента загрязнения печного топлива другими нефтепродуктами, равное 55:

$$50 VNAPD + 54 VDISD + 65 LCCYD \leq 55 VNAPD + 55 VDISD + 55 LCCYD. \quad (39)$$

Производительность по выпуску бензина и требования к октановым числам задаются точно так же, как и для печного топлива. Необходимые характеристики для этого приведены в табл. 7.

Таблица 7

Октановые числа бензиновых смесей

	Октановое число	Объем
Нафта прямой отгонки из нефти 1	85	Не менее 25 MB/D Не более 10 MB/D
Нафта прямой отгонки из нефти 2	84	
Нафта, получаемая в результате каталитического крекинга	92	
Высококачественный бензин (по техническим условиям)	89	
Обычный бензин (по техническим условиям)	85	

Тщательный анализ технологического процесса, подобный тому, что был проведен при построении модели нефтеперерабатывающего завода, играет важнейшую роль в текущем планировании, так как часто меняются затраты, спрос, наличие сырья и т. п. Сложность этих процессов не позволяет эффективно учитывать влияние изменений такого рода без математического анализа. Изложенный подход вошел в повседневную практику текущего планирования крупных перерабатывающих предприятий.

5. ОПТИМИЗАЦИЯ ТЕХНОЛОГИЧЕСКИХ ПРОЦЕССОВ

Часто *N* технологических операций (рис. 1) связаны таким образом, что выход первой операции (или этапа) является входом второй, выход второй является входом третьей и т. д. Такой процесс принято называть последовательным. В системах подобного рода решение, принятое на каком-либо этапе, влияет только на поток информации на последующих этапах. Для последовательного технологического процесса принцип оптимальности Беллмана можно сформулировать следующим образом: *процесс последовательно оптимизирован, если решения, принимаемые на его последующих этапах, составляют оптимальную стратегию относительно входных данных, поступающих с предшествую-*

щего этапа. Применяя этот принцип, можно расчленить исходную задачу оптимизации, описываемую $N \otimes p$ переменными, на N последовательных подзадач оптимизации, каждая из которых содержит p переменных. Такая декомпозиция особенно важна в перерабатывающих отраслях, когда N часто достаточно велико, например $N > 10$. Выбранный для иллюстрации этого подхода пример относится к химическим реакторам. Этот пример не только освещает важную сферу практического применения метода, но позволяет также наглядно ввести основные понятия, которые будут использованы в дальнейшем.

5.1. Химическая реакция в смешительном аппарате

Химические реакции, в результате которых сырье превращается в более ценные продукты, часто проводятся в крупных смешительных аппаратах, называемых реакторами. В реактор подается сырье, а из реактора выходит поток готовых продуктов (рис. 6). Состав и температура содержащихся в реакторе продуктов поддерживаются однородными за счет достаточно энергичного перемешивания. Предположим, что в реакции участвуют m различных компонентов $A_1, A_2, \dots, A_m$, т. е. они потребляются или образуются в ходе химических реакций, происходящих в реакторе. Пусть $r \equiv (r_1, r_2, \dots, r_m)$ представляет собой m -мерный вектор-столбец, описывающий скорость реакции между различными компонентами (т. е. изменение количества компонентов в единице объема). Из элементарной теории химических реакций следует, что $r = r(x, T)$, где $x \equiv (x_1, x_2, \dots, x_m)$ есть m -мерный вектор-столбец, определяющий количества или концентрации компонентов в реакторе, а T — температура. Строгая аналитическая модель вида $r(x, T)$, описывающая простую химическую реакцию, сложна, и ее трудно построить. Поэтому на практике используется

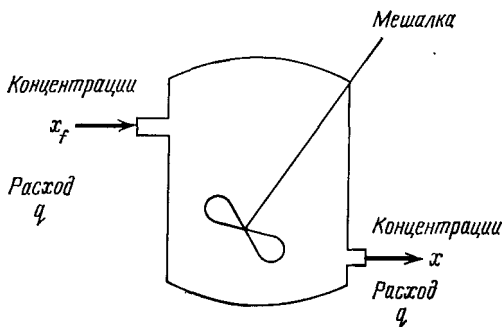


Рис. 6. Химический реактор проточного типа.

эмпирическая модель вида

$$r_i(x, T) = k_i \prod_{j=1}^m x_j^{a_{ij}} - k'_i \prod_{j=1}^m x_j^{b_{ij}}, \quad i = 1, 2, \dots, m, \quad (40)$$

где a_{ij} и b_{ij} — положительные действительные числа, обычно целые, включая нуль, а k_i и k'_i — константы (значения которых зависят от температуры). Первая группа членов в правой части уравнения (40) описывает скорость прямой реакции (например, при образовании компонента A_i), а вторая группа — скорость обратной реакции. Первоначально скорость прямой реакции выше скорости обратной. Затем по мере увеличения (уменьшения) количества A_i эта скорость падает до тех пор, пока не наступает равенство скоростей. При равновесных условиях имеем

$$r(x_e, T) = 0, \quad (41)$$

или

$$k_i \prod_{j=1}^m (x_j)_e^{a_{ij}} = k'_i \prod_{j=1}^m (x_j)_e^{b_{ij}}, \quad i = 1, 2, \dots, m, \quad (42)$$

где индекс e означает максимальную или минимальную концентрацию компонента при равновесии в зависимости от того, образуется или потребляется он в ходе химической реакции.

Если объем V реактора постоянен, то поток продуктов, выходящих из реактора, должен быть равен потоку поступающих в него продуктов. Поскольку содержимое аппарата равномерно перемешано, концентрации x компонента в реакторе те же, что и в конечном продукте. Следовательно, скорость образования (потребления) любого компонента A_i представляет собой разность между скоростью его подачи в реактор и скоростью выхода из реактора. Скорость реакции определяется как скорость увеличения A_i на единицу объема. Для входного потока x_f уравнение материального баланса по каждому компоненту определяет уравнение процессов в реакторе

$$q(x - x_f) = V r(x, T), \quad (43)$$

что равносильно

$$x = x_f + \theta r(x, T), \quad (44)$$

где $\theta \equiv V/q$ — среднее время пребывания продуктов в реакторе, т. е. время, необходимое для полной переработки содержимого реактора.

Предположим теперь, что требуется максимизировать количество компонента A_i , находящегося в реакторе при постоянной температуре T , начиная с входного потока с заданной концентра-

цией x_{fi} . В соответствии с уравнением (44) этого можно добиться, увеличивая θ , r_i или оба параметра одновременно. Но θ возрастает либо при увеличении объема реактора V , либо при сокращении его производительности q , что в экономическом отношении не выгодно. Скорость r_i снижается по мере роста количества образующегося продукта A_i и падает практически до нуля, когда концентрация x_{ie} приближается к максимальной. Отсюда следует, что реакцию нужно проводить в двух или более реакторах, подавая продукт реакции первого реактора во второй и т. д. При этом, поскольку выходная концентрация продукта первого реактора будет ниже, чем у второго и последующих, скорость реакции в первом реакторе будет выше, чем в последующих. Таким образом, осуществляя реакцию в нескольких реакторах, можно получить либо более высокую выходную концентрацию при том же объеме V , либо ту же выходную концентрацию, что и при использовании одной установки, но при меньшем общем объеме. Задача оптимизации заключается в выборе размеров реакторов исходя из некоторого экономического критерия.

5.2. Химическая реакция в последовательно работающих реакторах

Рассмотрим технологический процесс, осуществляемый в N реакторах (рис. 7). Для удобства рассуждений установки пронумерованы от конца к началу, так что входной поток поступает в реактор N , а конечный продукт выходит из реактора 1. Это — последовательный процесс, в котором N технологических операций (рис. 1) осуществляются в N реакторах. Переменными состояниями являются концентрации компонентов, а выходы обобщенной модели представляют собой входы последующей ступени.

Управляющей переменной $d_n = \theta_n$ является время пребывания продукта в n -м реакторе. Уравнение (44) для реактора N выступает в качестве переходной функции, поскольку выход x_n определяется входом x_{n+1} и управляющей переменной θ_n , хотя, вообще говоря,

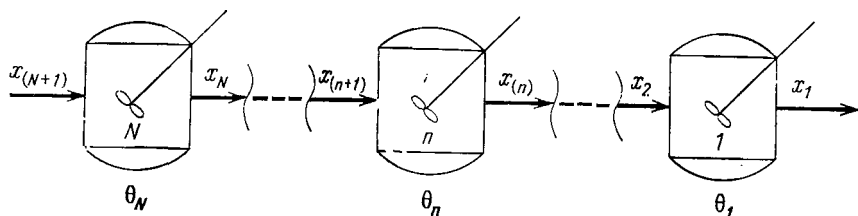


Рис. 7. Последовательное соединение химических реакторов.

для этого может потребоваться применение численного итерационного метода. Следовательно, можно записать

$$x_n = \mathcal{T}_n(x_{n+1}, \theta_n). \quad (45)$$

Целевую функцию этого процесса можно представить множеством способов. Часто используется функция, заданная в виде суммы прибыли, получаемой в результате работы каждой установки. Эти прибыли, как следует ожидать, зависят от x_{n+1} , x_n и θ_n . Однако уравнение (45) позволяет выразить x_n через x_{n+1} и θ_n :

$$P = \sum_{n=1}^N R_n(x_{n+1}, \theta_n), \quad (46)$$

где R_n — прибыль, получаемая на n -й ступени процесса. Одной из важных задач проектирования реактора является максимизация выходной концентрации компонента x_{i1} при входном потоке известного состава x_{N+1} и расходе q . Если α — отношение стоимости установки к стоимости перерабатываемых в ней материалов, то задача сводится к нахождению максимального значения целевой функции

$$x_{i1} - \alpha \left(\sum_{n=1}^N \theta_n \right) \text{ при условии, что } R_n = x_{in} - x_{i, n+1} - \alpha \theta_n. \quad (47)$$

Поскольку сумма содержит постоянный член $-x_{i, N+1}$, максимальное значение необходимо скорректировать, но оптимальная стратегия $\{\theta_n^* : n = 1, 2, \dots, N\}$ будет неизменной.

В терминах математического программирования задача оптимизации может быть записана следующим образом:

найти

$$\max_{\{\theta_n, n=1, 2, \dots, N\}} P = \sum_{n=1}^N R_n(x_{n+1}, \theta_n) \quad (48)$$

при ограничениях $N \oplus m$ в форме равенств

$$x_n = \mathcal{T}(x_{n+1}, \theta_n) = 0, \quad (49)$$

где x и $\mathcal{T}$ — переходная матрица размерности n .

Во всех реальных задачах переменные состояния и управляющие переменные неотрицательны и в большинстве случаев ограничены сверху и снизу. Таким образом, на них также накладываются ограничения

$$\theta_n^l \leq \theta_n \leq \theta_n^u \text{ и } x_n^l \leq x_n \leq x_n^u. \quad (50)$$

5.3. Декомпозиция и последовательная оптимизация методом динамического программирования

Вместо того чтобы непосредственно решать N -мерную задачу оптимизации с ограничениями в виде равенств и неравенств, пользуясь принципом оптимальности, можно разбить ее на N одномерных оптимизационных задач с одним ограничением. Выбор переменной θ_n влияет только на последующие состояния $x_n, x_{n-1}, \dots, x_1$ системы и не оказывает воздействия на предшествующие состояния $x_N, x_{N-1}, \dots, x_{n+1}$. Таким образом, это последовательная система. По принципу оптимальности все n -шаговые последовательные системы, т. е. системы, состоящие из шагов $1, 2, \dots, n$ ($n \leq N$), оптимальны по отношению к своим входным состояниям. Пусть

$$f_1(x_2) = \max_{\theta_1} [R_1(x_2, \theta_1)] \quad (51)$$

есть максимальная прибыль, получаемая от последнего реактора. Оптимальная стратегия для этого реактора зависит от входного состояния x_2 и может быть записана в виде $\theta_1^*(x_2)$. Она не влияет ни на один из $N - 1$ предшествующих этапов и определяется оптимизацией по одной переменной при различных значениях x_2 . Рассмотрим далее два последних реактора. Оптимальное значение целевой функции дается выражением

$$\max_{\theta_2, \theta_1} [R_2(x_3, \theta_2) + R_1(x_2, \theta_1)]. \quad (52)$$

При любом значении θ_2 можно определить R_2 и вычислить значение x_2 по уравнению (45). Однако оптимальной стратегией θ_1 при любом x является $\theta_1^*(x_2)$, так что значения суммы $R_1 + R_2$ определяются выбором θ_2 . Следовательно, уравнение (52) можно записать так:

$$f_2(x_3) = \max_{\theta_2} [R_2(x_3, \theta_2) + f_1(x_2)], \quad (53)$$

что представляет собой очередную задачу оптимизации по одной переменной, результатом решения которой является оптимальная стратегия $\theta_2^*(x_3)$.

Последовательно решая задачи оптимизации, получим, что оптимальным решением для n -го шага является стратегия $\theta_n^*(x_{n+1})$, удовлетворяющая уравнению

$$f_n(x_{n+1}) = \max_{\theta_n} [R_n(x_{n+1}, \theta_n) + f_{n-1}(x_n)], \quad (54)$$

где

$$x_n = \mathcal{T}_n(x_{n+1}, \theta_n), \quad n = 1, 2, \dots, N, \quad (55)$$

при условии что $f_0(x_1) \equiv 0$. Таким образом, оптимальная стратегия для всей последовательности $\{\theta_n^*(x_{n+1}), n = 1, 2, \dots, N\}$

формируется путем последовательного решения этого уравнения при $n = 1, 2, \dots, N$ начиная с $f_0 = 0$. Эту стратегию для простых систем можно представить в табличной или графической форме. Практические способы применения метода динамического программирования изложены в работах [2, 7, 8, 30].

5.4. Последовательные химические реакции с регулированием температуры

В изложенном выше примере предполагалось, что температура в реакторе постоянна и на ее регулирование не требуется никаких энергетических затрат. Если реакция идет с выделением тепла, то оно должно отводиться с определенной скоростью. Положим, что скорость отвода тепла охлаждающей системой равна $\dot{Q}_n$; C_p — теплоемкость реагирующей смеси и теплота, выделяющаяся при реакции i -й компоненты, равна $(-\Delta H_i)$. Тепловой баланс реактора, т. е. тепло, поступающее в него, в сумме с теплом, выделяющимся в ходе реакции, равно теплу, отводимому выходящим потоком, плюс тепло, отводимое охлаждающей системой, описывается уравнением

$$\dot{Q}_n = qC_p(T_{n+1} - T_n) + V_n(-\Delta H')r(x_n, T_n), \quad (56)$$

или

$$Q_n = T_{n+1} - T_n + \theta_n \left(\frac{-\Delta H'}{qC_p} \right) r(x_n, T_n), \quad (57)$$

где $Q_n = \dot{Q}_n/(qC_p)$. Это уравнение характеризует процесс изменения температуры.

При заданных значениях переменной состояния T_{n+1} , x_{n+1} и решениях θ_n и Q_n совместное решение этого уравнения с уравнением материального баланса (45) позволяет определить значения T_n и x_n . (Температура рассматривается как переменная состояния, так как она влияет на требуемый отвод тепла.) Затраты на отвод тепла пропорциональны Q_n . В результате целевая функция уравнения (47) принимает вид

$$R_n = x_{in} - x_{i, n+1} - \alpha \theta_n - \mu Q_n. \quad (58)$$

Отметим, что теперь Q_n стала управляющей переменной, т. е. $d_n = (\theta_n, Q_n)$, так что на каждом шаге приходится решать задачу последовательной оптимизации по двум переменным.

5.5. Применение динамического программирования для оптимизации систем более сложной структуры

Многие технологические процессы построены таким образом, что поток продуктов, выходящий из одной или нескольких установок, разветвляется или возвращается обратно в предшествую-

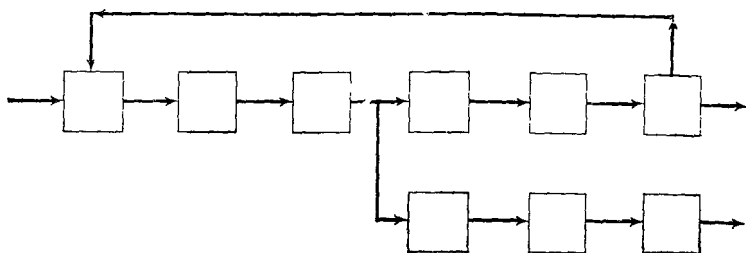


Рис. 8. Сложная последовательно-параллельная система с обратными связями.

ющие по технологической цепочке установки. Типичная структура такого процесса показана на рис. 8. Для оптимизации систем этого класса разработано несколько способов применения метода динамического программирования [4, 6]. В их основе лежит введение дополнительных переменных состояния, соответствующих различным замкнутым контурам процесса. Далее задача большой размерности расчленяется на задачи меньшей размерности с помощью метода динамического программирования [35].

5.6. Выводы

В этом разделе было показано, что как в случае последовательной структуры, так и в системах большой размерности можно использовать метод динамического программирования путем декомпозиции задачи и ее оптимизации в виде задач меньшей размерности. Хотя общая идея динамического программирования, выраженная в принципе оптимальности, достаточно проста, при постановке конкретных задач возникают существенные трудности. Это объясняется тем, что метод динамического программирования представляет собой лишь общую стратегию оптимизации, а не точно определенные алгоритмы, подобные алгоритмам линейного программирования. Поэтому в каждом случае приходится проводить конкретный анализ, а успех или неудача применения метода динамического программирования определяются умением поставить конкретную задачу.

Вопрос о целесообразности применения динамического программирования для декомпозиции последовательных систем исследован рядом авторов [2, 30]. Ценой, которую приходится платить за расчленение системы на подсистемы меньшей размерности, является необходимость решения большого числа задач оптимизации, хотя результаты таких решений не позволяют получить оптимальное решение для всей задачи в целом. Недостатки такого погружения в класс задач меньшей размерности необходимо

сопоставить с вычислительными преимуществами, достигаемыми за счет уменьшения числа переменных. Наилучшим признаком, определяющим границы применимости метода динамического программирования, является число переменных состояния m . С точки зрения вычислительных возможностей задача оптимизации ограничена тремя, максимум четырьмя, такими переменными. Поиск оптимума при большом числе переменных состояния становится неосуществимым даже при использовании современных высокопроизводительных ЭВМ. Однако имеются два подхода, позволяющие устранить это ограничение. Один из них, называемый динамическим программированием эволюции состояний, существенно снижает объем вычислений за счет решения рекуррентного функционального уравнения (54) принципиально иным способом [28]. Другой подход сводится к применению принципа декомпозиции, получившего название метода декомпозиции, основанного на последовательном анализе вариантов. Однако за исключением частных случаев, он не обеспечивает отыскания однозначного оптимума, что присуще динамическому программированию [32]. Этот подход не требует большой памяти ЭВМ, так что число переменных состояния не столь существенно.

6. ОПТИМИЗАЦИЯ ПРОЕКТНЫХ РЕШЕНИЙ МЕТОДОМ ГЕОМЕТРИЧЕСКОГО ПРОГРАММИРОВАНИЯ

В предыдущих разделах было показано, как можно сформулировать проблемы выбора экономических технологических схем и режимов путем решения задач оптимизации, в которых отыскивается экстремум некоторого экономического критерия при определенных технологических или проектных ограничениях. Однако до последнего времени методы оптимизации не применялись при рабочем проектировании ввиду отсутствия математических методов оптимизации нелинейных динамических систем, характерных для данного класса задач.

Рабочее проектирование в отличие от выбора принципиальных технологических схем, рассмотренного в разд. 5, включает определение детальных конструктивных размеров насосов, двигателей, теплообменников и т. п., а не общей мощности установки или времени пребывания продукта в реакторе. Описывающие их соотношения не только нелинейны, но и певыпуклы. Так, например, в разд. 4 показано, что соотношение, по которому определяется объем капитальных затрат на новые предприятия и оборудование, имеет вид: Капитальные затраты $= aV^b$, где V — производственная мощность, a — коэффициент удельных затрат, b — положительный показатель степени, значение которого обычно лежит в интервале 0,5—0,8. Экономический критерий, выраженный в терминах общей модели (рис. 1), применяется в рабочем

(техническом) проектировании, и почти целиком состоит из сумм и разностей соотношений такого вида. Значения коэффициентов определяются по данным экономических наблюдений, приведенным в литературе, а показатели обычно отражают опыт, накопленный разработчиками аналогичного оборудования.

В задачах технического проектирования переходные функции определяются в результате расчетов материальных и энергетических балансов как по отдельным технологическим операциям, так и по процессу в целом. Эти балансы, как правило, включают соотношения, полученные в результате анализа размерностей либо путем подгонки степенных функций к экспериментальным данным. В общем виде такая функция h , выражаемая через P безразмерных переменных, записывается следующим образом:

$$h(x) = k \prod_{i=1}^P \left(\prod_{j=1}^m x_j^{\alpha_{ij}} \right)^{a_i}, \quad (59)$$

где $x \equiv (x_1, x_2, \dots, x_m)$ есть m -мерный вектор конструктивных и/или физических переменных, α_{ij} — целое число, k и a_i — константы (действительные числа), определенные в результате проведения экспериментов на реальной системе.

Поскольку как экономические, так и конструктивные соотношения представляют собой полиномы, то следует предположить, что выбор оптимального проектирования может быть произведен методами геометрического программирования. Однако вследствие того, что постановка задачи достаточно сложна, успех применения метода зависит от способностей проектировщика. Каждую конкретную задачу необходимо рассматривать особо, поскольку нет общего подхода к постановке задач подобного рода. Одним из приемлемых методов геометрического программирования является правильный выбор переменных конструкции на основе аппроксимации проектных соотношений полиномами и преобразования сложных ограничений материальных балансов в полиномиальные ограничения в виде равенств и неравенств.

Изложенные соображения лучше всего проиллюстрировать на примере.

6.1. Проектирование теплообменника с передачей тепла от жидкости к жидкости

Проектирование установки для передачи тепла, регулирующей температуру процесса или обеспечивающей эффективное использование тепла, представляет собой важную конструкторскую задачу. Так, например, задача оптимального проектирования последовательности реакторов включает отыскание оптимальной

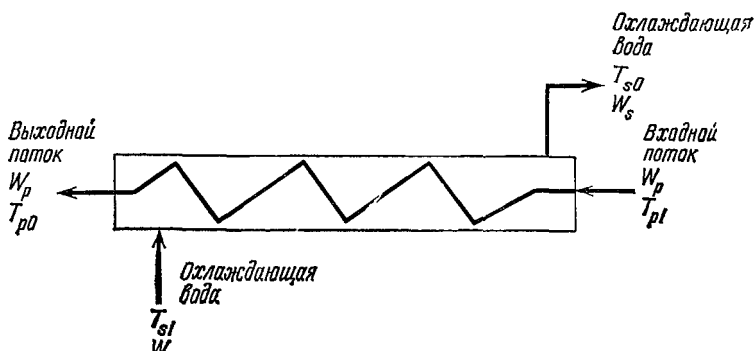


Рис. 9. Схема теплообменника.

стратегии изменения скорости отвода тепла от каждого реактора. Определив эту скорость (Q), переходим к решению проектной задачи выбора габаритов оборудования, обеспечивающего передачу тепла с найденной скоростью при минимуме затрат.

Пример. Рассмотрим теплообменник жидкость — жидкость, в котором рабочая жидкость с расходом W_p охлаждается от температуры T_{pI} до T_{p0} потоком воды W_s , имеющей определенную температуру T_{sI} [9] (рис. 9). Скорость теплопередачи от технологического потока к охлаждающей воде равна

$$Q = W_p C_p (T_{pI} - T_{p0}), \quad (60)$$

где C_p — теплоемкость технологического потока. Задача оптимального проектирования сводится к выбору конструктивных параметров теплообменника, т. е. к определению поверхности теплопередачи, числа и диаметра труб, положения перегородки и т. д., а также к определению расхода охлаждающей воды W_s , необходимого для обеспечения заданной скорости теплопередачи Q , при минимизации годовых затрат на теплообменник.

Целевая функция для теплообменника включает постоянные затраты на теплообменник c_F , затраты на охлаждение воды c_w , а также затраты на прокачку воды через теплообменник c_s и технологической жидкости через теплообменник c_p .

Постоянные затраты. В предположении, что используется теплообменник обычного кожухотрубчатого типа, соотношение, определяющее оценку постоянных затрат, имеет вид

$$c_F = c_1 A^{0,55}, \quad (61)$$

где c_1 — удельные капитальные затраты, включающие годовую норму амортизации, A — поверхность теплообмена.

Затраты на воду. Этот вид затрат, как правило, является функцией расхода воды, т. е.

$$c_w = c_2 H W_s, \quad (62)$$

где c_2 — удельные затраты на единицу объема охлаждающей воды, H — время работы теплообменника в течение года.

Затраты на прокачку. Эти затраты являются следующими линейными функциями потребляемой мощности:

$$c_s = c_3 H \text{ (HP}_s\text{)}, \quad (63)$$

$$c_p = c_3 H \text{ (HP}_p\text{)}, \quad (64)$$

где c_3 — удельные затраты на единицу мощности, а HP_s и HP_p — потери мощности на охлаждающую воду и технологические потоки соответственно.

Если просуммировать эти четыре величины, то задача выбора оптимальной конструкции сводится к определению значений A , W_s , HP_s и HP_p , минимизирующих

$$R = c_F + c_w + c_s + c_p. \quad (65)$$

Для расчета примем следующие параметры теплообменника: 1) используются стандартные дюймовые трубы марки 16BMG (бирмингемского производства) длиной 16 футов; 2) число ходов внутри труб, равное числу ходов в межтрубном пространстве, составляет 1; 3) разбивка по квадрату с шагом $1\frac{1}{4}$ дюйма, расстояние между перегородками равно $\frac{2}{5}$ диаметра кожуха. Потребление мощности HP_s и HP_p можно выразить через параметры конструкции A , W_s и W_p следующим образом. Поскольку охлаждающая вода проходит вне труб теплообменника, можно определить потери мощности на внешней поверхности труб уравнением

$$HP_s = K_s A h_s^{3,5}, \quad (66)$$

а на внутренней поверхности труб — уравнением

$$HP_p = K_p A h_p^{4,75}, \quad (67)$$

где K_s и K_p зависят от теплофизических свойств жидкостей, а h_s и h_p — коэффициенты теплоотдачи для внешних и внутренних поверхностей труб теплообменника соответственно. На основе анализа размерностей (т. е. использования теории подобия) определяем соотношения для коэффициентов теплоотдачи:

$$h_s = 0,023 \left(\frac{k_s}{D_s} \right) (Re_s)^{0,8} (Pr_s)^{1,3} \left(\frac{\mu_s}{\mu_w} \right)^{0,14}, \quad (68)$$

$$h_p = 0,23 \left(\frac{k_p}{D_p} \right) (Re_p)^{0,6} (Pr_p)^{0,33}, \quad (69)$$

где Re и Pr — безразмерные числа Рейнольдса и Прандтля, теплопроводность и вязкость жидкостей определяются коэффициентами k и μ соответственно, а D — диаметр труб. Эти уравнения аналогичны по форме уравнениям (59).

Объединяя члены, описывающие теплофизические свойства жидкостей, и выражая уравнения (68) и (69) через массовые расходы и площади поверхности, получаем

$$h_s = e_1 \phi_s \left(\frac{W_s}{A} \right)^{0,8}, \quad (70)$$

где

$$\phi_s \equiv k_s \left(\frac{1}{\mu_s} \right) (Pr_s)^{1,3} \left(\frac{\mu_s}{\mu_w} \right)^{1,04}, \quad (71)$$

и

$$h_p = e_2 \phi_p \frac{W_p^{0,6}}{A^{0,57}}, \quad (72)$$

где

$$\phi_p \equiv \frac{k_p}{\mu_p} \text{Pr}_p. \quad (73)$$

Коэффициенты e_1 и e_2 определяются выбранной системой единиц. Подставляя соотношения (70) и (71) в уравнения (66) и (67) соответственно, получаем следующее выражение целевой функции (64):

$$R = [c_1] A^{0,55} + [c_2 H] W_s + [c_3 H K_s (e_1 \phi_s)^{3,5}] \frac{W_s^{2,8}}{A^{1,8}} + \\ + [c_4 H K_p (e_2 \phi_p)^{4,75}] \frac{W_p^{2,85}}{A^{1,71}}, \quad (74)$$

где члены, стоящие в квадратных скобках, являются постоянными.

Необходимо учесть два ограничения. Во-первых, поверхность теплоотдачи, необходимая для передачи тепловой нагрузки Q , определяется соотношением

$$Q = UA \Delta T_{LM}, \quad (75)$$

где U — коэффициент теплопередачи, а ΔT_{LM} — среднелогарифмическое значение перепада температур. Величина U связана с коэффициентами h_s и h_p и коэффициентом загрязнения R_D рациональной функцией

$$\frac{1}{u} = \frac{1,15}{h_s} + \frac{1}{h_p} + R_D. \quad (76)$$

Среднелогарифмическое значение перепада температур не выражается в виде логарифма. Однако для систем жидкость — жидкость с аналогичными температурными условиями на концах теплообменника, т. е. $(T_{pO} - T_{sI}) \sim \sim (T_{pI} - T_{sO})$, величину ΔT_{LM} можно приближенно принять равной ΔT_{AM} — среднсрифметическому значению перепада температур. Таким образом,

$$\Delta T_{LM} \approx \frac{(T_{pO} - T_{sI}) + (T_{pI} - T_{sO})}{2}. \quad (77)$$

Подставляя выражения для h_s и h_p из уравнений (70) и (72), получаем (74) в виде

$$Q \left(\frac{1,15}{e_1 \phi_s A^{0,2} W_s^{0,8}} + \frac{1}{e_2 \phi_p A^{0,45} W_p^{0,6}} + \frac{R_D}{A} \right) = \\ = \frac{T_{pO}}{2} - \frac{T_{sI}}{2} + \frac{T_{pI}}{2} - \frac{T_{sO}}{2}. \quad (78)$$

Второе ограничение обусловлено энергетическим балансом: потери тепла технологическим потоком равны количеству тепла, поглощенному охлаждающей водой, т. е.

$$Q = W_p C_{p_p} (T_{pI} - T_{pO}) = W_s C_{p_s} (T_{sO} - T_{sI}). \quad (79)$$

Итак, задача оптимального проектирования сформулирована теперь как нелинейная задача математического программирования с тремя переменными A , W_s , T_{sO} и ограничениями в виде равенств. Целевая функция (74) является

рациональной. Ограничения также можно представить в виде следующих полиномов:

$$\left[\frac{C_{pI} W_p (T_{pO} - T_{pI})}{C_{pS}} \right] \frac{W_p}{T_{sO}} + [T_{sI}] \frac{1}{T_{sO}} = 1, \quad (80)$$

$$\left[\frac{1,15Q}{e_1 \phi_s E_T} \right] \frac{1}{A^{0,2} W_s^{0,8}} + \left[\frac{Q}{e_2 \phi_p E_T W_p^{0,45}} \right] \frac{1}{A^{0,45}} +$$

$$+ \left[\frac{R_D}{E_T} \right] \frac{1}{A} + \left[\frac{1}{2E_T} \right] T_{sO} = 1, \quad (81)$$

где

$$E_T = (T_{pO} - T_{sI} + T_{pI})/2.$$

Разработаны эвристические алгоритмы преобразования этих ограничений в неравенства при сохранении выпуклости [11]. Можно также использовать алгоритм, специально разработанный для решения задач с ограничениями в виде равенств, однако при этом не гарантируется отыскание глобального оптимума, обеспечиваемого условием выпуклости [10].

7. МАТЕМАТИЧЕСКИЕ МОДЕЛИ ТЕХНОЛОГИЧЕСКИХ ОПЕРАЦИЙ

В предыдущих разделах этой главы основное внимание уделялось постановке оптимизационных задач, связанных с различными технологическими процессами. Во всех рассмотренных примерах предполагалось, что переходные функции представлены либо в виде физических моделей, достаточно адекватно описывающих процессы (разд. 5 и 6), либо в виде линейных аппроксимаций таких моделей. Однако при моделировании технологических процессов единственным надежным способом определения адекватности модели является ее проверка с помощью данных эмпирических измерений параметров самих процессов либо дополнительное экспериментальное исследование моделируемых процессов. Если эмпирические данные имеются, то для оценивания основных параметров модели необходимо использовать наименее методы наименьших квадратов. Неопределенность полученных оценок параметров следует учитывать во всех последующих расчетах при использовании модели.

Часто описание технологической операции можно представить в виде различных достаточно адекватных моделей. Это особенно характерно для моделирования процессов, включающих химические реакции, ибо такие процессы по своей природе чрезвычайно сложны, а эмпирические данные о них бедны и/или получение таких данных сопряжено с большими затратами. Из-за существенной нелинейности для определения наиболее подходящей модели нельзя применить обычные статистические испытания, основанные

на линейных моделях. В связи с этим были разработаны нелинейные статистические методы, предназначенные для оценки качества «конкурирующих» моделей по имеющимся данным (см. работы [12—14, 33]), а также методология постановки экспериментов, предназначенных для получения оценок качества (адекватности) моделей [15]. В рамках этой методологии, основанной на максимизации информации, извлекаемой из полученных данных, можно ставить задачи математического программирования.

Однако: даже в случае, когда подходящая модель выбрана, неопределенность оценок параметров может оказаться столь велика, что такую модель не удастся использовать как основу для постановки любой задачи математического программирования. В связи с этим разработаны статистические методы планирования экспериментов, максимизирующих математическое ожидание точности оценок параметров [5, 16]. Такую задачу максимизации можно сформулировать в терминах математического программирования.

Постановки упомянутых выше трех задач математического программирования будут приведены в настоящем разделе. Они рассматриваются в самом общем виде. Аналитическая форма конкретных моделей не указывается. Будет лишь показано, что необходимы эффективные алгоритмы решения задач такого класса, что обусловлено большим количеством параметров и необходимостью оценки значительного числа моделей и многократного решения задач математического программирования по мере поступления новых данных.

7.1. Нелинейная оценка параметров

Предположим, что произведено s измерений $(y_n)_{n,k} = 1, 2, \dots, s$, выходных состояний технологической операции при различных значениях входных состояний $(x_n)_k$, а также $(d_n)_k$. (Для упрощения записей индекс n , обозначающий стадию процесса, опускается, если смысл выражения очевиден.) Эти измеренные значения связаны с истинными значениями w_k соотношением

$$y_k = w_k + \varepsilon_k, \quad k = 1, 2, \dots, s, \quad (82)$$

где $\varepsilon_k \equiv (\varepsilon_{1k}, \varepsilon_{2k}, \dots, \varepsilon_{qk})$ есть q -мерный вектор, отображающий ошибку измерения y_k . Рассмотрим переходную функцию $f(\theta, d, x)$, где $\theta \equiv (\theta_1, \theta_2, \dots, \theta_t)$ есть t -мерный вектор неизвестных параметров. Если функция f адекватно описывает данные, то существует набор параметров θ^* такой, что вектор остатков

$$\varepsilon_k = y_k - f(\theta^*, d_k, x_k), \quad k = 1, 2, \dots, s, \quad (83)$$

определяет оценки ошибок измерения ε_k . Если совместная плотность распределения ошибок измерения известна, то подстановка

остатков ε_k в эту функцию дает известную функцию правдоподобия $L(\theta)$. Набором значений параметров, которые наилучшим в статистическом смысле образом описывают экспериментальные данные, являются значения θ^* такие, что

$$\max L(\theta). \quad (84)$$

Во многих случаях вектор θ неотрицателен и ограничен сверху. Однако достаточно предположить, что θ удовлетворяет ограничениям

$$\theta_l \leq \theta \leq \theta_u \quad (85)$$

где индексы u и l обозначают верхнюю и нижнюю грани соответственно. Следовательно, для определения вектора параметров θ^* необходимо решить нелинейную оптимизационную задачу с однозначными ограничениями.

Вид функции $L(\theta)$ зависит от структуры ошибок измерения. Так, например, если $\varepsilon \sim N(0, I\sigma^2)$, то задача максимизации функции правдоподобия преобразуется в известную задачу взвешенных наименьших квадратов

$$\min L(\theta) = \sum_{k=1}^s \sum_{j=1}^q \frac{\varepsilon_{jk}^2}{\sigma_h^2} \quad (86)$$

при однозначных ограничениях (85). Вид функции правдоподобия при других распределениях описан в работе [18]. Однако, чтобы убедиться в адекватности модели $f(\theta^*, d, x)$, следует выполнить проверку по статистическому критерию согласия. Если такая проверка дает неудовлетворительные результаты, необходимо построить другие модели и осуществить оценку их параметров по критерию максимального правдоподобия. При моделировании химических процессов с помощью подобной модели часто удается, исключив неудачные модели, оставить для последующего анализа одну или несколько потенциально адекватных моделей [12—14]. Как правило, для сопоставления их качества требуется проведение дополнительных экспериментов.

7.2. Планирование экспериментов для оценки качества моделей

Для сравнения моделей могут быть использованы критерии оценки целесообразности проведения дополнительных экспериментов $x_k, d_k, k = s + 1, s + 2, \dots$ [33]. Согласно наиболее часто используемому критерию, эксперименты, максимизирующие ожидаемое убывание энтропии, проводятся последовательно [15]. Используемое далее понятие энтропии заимствовано из теории информации. Предположим, что предложено P моделей $f_1, f_2, \dots, f_P$ и имеется s экспериментальных точек $S \equiv \{(x_k, d_k, y_k),$

$k = 1, 2, \dots, s\}$. Тогда можно применить теорему Байеса для вычисления нормированной апостериорной вероятности адекватности каждой модели $\Pr(f_i | S)$ на основе информации, полученной по S экспериментальным точкам:

$$\Pr(f_i | S) = \frac{\Pr(f_i) L_i(\theta_i^*)}{\sum_{l=1}^P \Pr(f_l) L_l(\theta_l^*)}, \quad (87)$$

где $\Pr(f_i)$ — априорная вероятность и $L_i(\theta_i^*)$ — функция правдоподобия i -й модели. Энтропия системы в этой точке определяется выражением

$$E = - \sum_{i=1}^P \Pr(f_i | S) \ln [\Pr(f_i | S)]. \quad (88)$$

Значение E равно нулю, если адекватной является одна модель, и максимально, когда вероятности адекватности всех P моделей одинаковы. Наибольшее ожидаемое изменение энтропии в $(s + 1)$ -й экспериментальной точке достигается, если x_{s+1}^* , d_{s+1}^* удовлетворяют критерию

$$\max_{x_{s+1}, d_{s+1}} D, \quad (89)$$

где выражения для D приведены в табл. 8. Более подробная инфор-

Таблица 8

Критерии оценки качества моделей

$q = 1$ — выходное состояние, две конкурирующие модели ($P = 2$)

$$D = \max_{x_{s+1}, d_{s+1}} \left(\frac{1}{2} \Pr(f_1 | S) \right) \Pr(f_2 | S) \times \\ \times \left\{ \frac{(\sigma_2^2 - \sigma_1^2)^2}{(\sigma^2 + \sigma_1^2)(\sigma^2 + \sigma_2^2)} + (f_1 - f_2)^2 \left[\frac{1}{\sigma^2 + \sigma_1^2} + \frac{1}{\sigma^2 + \sigma_2^2} \right] \right\}$$

$q = 1$ — выходное состояние, P конкурирующих моделей

$$D = \max_{x_{s+1}, d_{s+1}} \left(\frac{1}{2} \sum_{j=1}^P \sum_{i=1}^P \Pr(f_i | S) \Pr(f_j | S) \right) \times \\ \times \left\{ \frac{(\sigma_j^2 - \sigma_i^2)^2}{(\sigma^2 + \sigma_j^2)(\sigma^2 + \sigma_i^2)} + (f_j - f_i)^2 \left[\frac{1}{\sigma^2 + \sigma_j^2} + \frac{1}{\sigma^2 + \sigma_i^2} \right] \right\},$$

где σ^2 — дисперсия ошибок по совокупности, относящейся к одному измерению,

$$\sigma_i^2 = (d_{s+1}, x_{s+1}) (X_i' X_i)^{-1} (d_{s+1}, x_{s+1})' \sigma^2$$

X_i — якобиан первых производных f_i , оцененных в точке (x_{s+1}, d_{s+1})

мая относительно многомерных моделей содержится в работе [15].

Отметим, что при решении рассмотренной оптимизационной задачи с $m + p$ переменными после определения оценок правдоподобия θ^* по первым s точкам производится поиск по всем допустимым значениям переменных состояния и управляющих переменных. При появлении данных по каждой дополнительной экспериментальной точке оптимальные оценки параметров необходимо скорректировать, пересчитать апостериорные вероятности и вновь целиком решить всю задачу. Хотя приходится выполнять огромный объем вычислений, опыт показал, что для достижения удовлетворительной сравнительной оценки адекватности моделей достаточно четырех-пяти экспериментальных точек. Экономическая целесообразность применения изложенного подхода, по крайней мере в условиях промышленного производства, определяется возможностью сопоставления затрат на вычисления с затратами на проведение экспериментов.

7.3. Планирование экспериментов для оценки параметров

Если адекватная модель найдена, то может возникнуть необходимость проведения дополнительных экспериментов с целью повышения точности оценок параметров. При этом предполагается, что каждый последующий эксперимент обеспечивает более желательную форму апостериорного распределения оценок параметров [5, 16]. Таковым, к примеру, является распределение, максимально сужающее совместную доверительную зону для всех t параметров, фигурирующих в модели $f(0, d, x)$. Если модель линеаризована, то минимизация такой совместной доверительной зоны сводится к отысканию значений d^* , x^* , которые обеспечивают выполнение условия

$$\max \Delta = \left| \sum_{i=1}^q \sum_{j=1}^q \sigma_{ij} X_i X_j \right|, \quad (90)$$

где X_i — якобиан f_i , элементами которого являются некоторые оценки θ^* . В сущности матрица X_i является матрицей моментов остатков, если предположить, что распределение ошибок эксперимента близко к нормальному, а модель с оценками максимального правдоподобия, полученными после s экспериментов, близка к линейной. Отыскание значений d^* , x^* путем максимизации определителя Δ составляет содержание третьей задачи математического программирования. И в данном случае вследствие итеративного характера схемы проведения экспериментов и анализа полученных результатов для ее реализации необходимо использование эффективного вычислительного алгоритма.

7.4. Выводы

В данном разделе сформулированы три задачи математического программирования. Метод решения первой из них, представляющей собой нелинейную задачу оценивания параметров с однозначными ограничениями, разработан в полном объеме. Однако для решения двух других задач [уравнения (89) и (90)] в период написания этой книги были известны методы, позволяющие отыскивать решение с помощью высокопроизводительных ЭВМ лишь при условии, что число переменных состояния и управляющих переменных невелико, т. е. ≤ 3 , и число рассматриваемых моделей также мало, т. е. ≤ 4 . Чтобы добиться широкого практического внедрения методов математического программирования для построения моделей технологических процессов в обрабатывающей промышленности, требуется разработка новых эффективных алгоритмов.

ЛИТЕРАТУРА

1. A Decade of Digital Computing in the Mineral Industry, Post City Press, Baltimore, 1969.
2. Aris R., Discrete Dynamic Programming, Blaisdell Publishing Co., New York, 1964.
3. Aris R., The Optimal Design of Chemical Reactors, Academic Press, New York, 1961.
4. Aris R., Nomhauser G. L., Wilde D. J., *AIChE J.*, 10, No. 6, 913—919 (1964).
5. Atkinson A. C., Hunter W. G., The Design of Experiments for Parameter Estimation, *Technometrics*, 10, No. 2, 271—289 (1968).
6. Beightler C. S., Johnson D. B., Wilde D. J., *J. Math. Anal. and Appl.*, 12, № 1, 96—104 (1965).
7. Bellman R., Dynamic Programming, Princeton University Press, 1957. [Имеется перевод: Беллман Р. Динамическое программирование.— М.: ИИ, 1960.]
8. Bellman R., Dreyfus S. E., Applied Dynamic Programming, Princeton University Press, 1962.
9. Blau G. E., Optimization of Models Derived by Dimensional Analysis Using Generalized Polynomial Programming, *J. Franklin Institute*, 292, No. 6, 519—526 (1971).
10. Blau G. E., Wilde D. J., A Lagrangian Algorithm for Equality Constrained Generalized Polynomial Programming, *Am. Inst. Chem. Eng. J.*, 17, 235—242 (1971).
11. Blau G. E., Generalized Polynomial Programming, *Canadian J. Chem. Eng.*, 47, 317—326 (1969).
12. Blau G. E., Klimpel R. R., Steiner E. C., Equilibrium Constant Determination by Nonlinear Optimization, *Ind. Eng. Chem. Funda.*, 9, 334—341 (1970).
13. Blau G. E., Nonlinear Parameter Estimation and Model Distinguishability of Physicochemical Models at Chemical Equilibrium, *Canadian J. Chem. Eng.*, 50, 399—409 (1972).
14. Blau G. E., Equilibrium Constant Estimation and Model Distinguishability, *Ind. Eng. Chem., Funda.*, 11, 324—332 (1972).

15. Box G. E. P., Hill W. J., Discrimination Among Mechanistic Models, *Technometrics*, 9, No. 1, 57—71 (1967).
16. Box G. E. P., Hunter W. G., Sequential Design of Experiments for Non-linear Models, Proceedings of the IBM Scientific Computing Symposium on Statistics, 1963, p. 113.
17. Fabian T., Balst Furnace Production Planning, *Management Sci.*, 14, No. 2, 1—26 (1967).
18. Hogg R. V., Craig A. T., Introduction to Mathematical Statistics, 2nd Ed., MacMillan, New York, 1967.
19. IBM Data Processing Application: An Introduction to Linear Programming, International Business Machines Corporation, White Plains, New York, 1964.
20. IBM Data Processing Application: Linear Programming-Aluminum Alloy Blending, International Business Machines Corporation, White Plains, New York.
21. IBM KWIC Index in Operations Research, International Business Machines Corporation, White Plains, New York, 1967.
22. Klimpel R. R., Operations Research in the Chemical Industry, Parts I and II, *Chem. Eng.*, 80, No. 9, 10, 103—108, 87—94 (1973).
23. Klimpel R. R., Recent Advances in the Use of Mathematical Optimization in the Mineral Industries, *Minerals Sci. and Eng.*, 1, No. 15—23 (1969).
24. Klimpel R. R., The Use of Mathematical Modeling to Evaluate Grinding and Blending Alternatives in an Industrial Comminution Facility, Proceedings of the International Conference in Particulate Technology, Illinois Inst. of Tech., 1973.
25. Klimpel R. R., Austin L. G., The Optimization of an Industrial Rotary-Cutter Milling Facility, Proceedings of the Third European Symposium on Grinding, Cannes, France, 1971.
26. Klimpel R. R., Burroughs W. J., Some Industrial Experiences in Plant Site Selection, 1974 ORSA Meeting in Boston.
27. Klimpel R. R., Klein M., An Application of Nonlinear Optimization Using the Gradient Projection Method, Proceedings of the Symposium on O. R. in the Mineral Industries, University Park, Pennsylvania, 1966.
28. Larson R. E., State Increment Dynamic Programming, Elsevier, New York, 1968.
29. Manula G., Kim Y., A Linear Programming Simulator for Coal Distribution Problems, Proceedings of the Symposium on O. R. in the Mineral Industries, University Park, Pennsylvania, 1966.
30. Nemhauser G., Introduction to Dynamic Programming, Wiley, New York, 1966.
31. Operations Research/Management Science Literature Digest, Executive Sciences Institute, Whippany, New Jersey, 1961.
32. Погорячий Л. С., Болтянский В. Г., Гамкрелидзе Р. В., Мищенко Е. Ф. Математическая теория оптимальных процессов.— М.: Наука, 1961.
33. Reilly P. M., Statistical Methods of Model Discrimination, *Canadian J. Chem. Eng.*, 48, 168—173 (1970).
34. Shannon C. E., *Bell Syst. Tech.*, 1, 27 (1948).
35. Wilde D. J., Beightler C. S., Foundations of Optimization, Prentice-Hall, Englewood Cliffs, New Jersey, 1967.

II-8. Организация досуга (туризм, спорт, развлечения)

У. Суот¹⁾, Ч. Гиэринг²⁾, Т. Ва³⁾

ВВЕДЕНИЕ

Увеличение свободного времени у населения промышленно развитых стран обусловлено рядом факторов, в том числе такими, как сокращение рабочей недели, увеличение числа бытовых электроприборов и готовых к употреблению продуктов питания, улучшение средств связи и транспорта. Согласно данным работы [64], относительная доля свободного времени от общего количества активного времени человека составляла в 1910 г. всего лишь 27%, в 1950 г. — около 34%, а в 2000 г. достигнет 38% (табл. 1).

Таблица 1

Распределение свободного времени в 1950 и 2000 гг.

Род деятельности	Относительная доля свободного времени, %	
	1950	2000
Работающие		
Рабочие дни	42	34
Выходные дни	39	44
Отпуск	8	16
Пенсионеры	5	5
Прочие	6	1

Увеличение свободного времени сопровождается одновременным увеличением доходов в сфере организации досуга, причем наблюдается смещение в балансе доход — досуг в сторону первого. Авторы работы [29] полагают, что подобное смещение в балансе способствует расширению возможностей для населения наилучшим образом использовать свое свободное время, что в свою очередь стимулирует развитие «отрасли организации досуга».

Под «отраслью организации досуга» понимается экономическая деятельность, направленная на создание возможностей и условий для проведения свободного времени вне дома. Подобного рода деятельность могут осуще-

¹⁾ William W. Swart, Ph. D., Professor of Management Science and Associate Dean of the School of Business Administration and Economics at the California State University at Northridge.

²⁾ Charles E. Gearing, Ph. D., Professor and Acting Dean of Management at the State University of New York at Binghamton, New York.

³⁾ Turgut Var, Ph. D., Professor of Economics and Commerce at the Simon Fraser University in Burnaby, British Columbia, Canada.

ствлять государственные (например, U. S. National Park Service), общественные (культурные и зрелищные муниципальные учреждения) или частные (отели, парки отдыха) организации.

Административная деятельность, связанная с организацией досуга, пока что не стала объектом всестороннего изучения для исследователей систем, хотя полученные результаты свидетельствуют об эффективности применения методов исследования операций и при организации некоторых видов отдыха вне дома. Существующее положение дел в рассматриваемой сфере деятельности напоминает ситуацию, характерную для 1966 г., когда наблюдалось увеличение научно-исследовательских работ, посвященных организации отдыха вне дома, что давало основания для предсказания возрастающей роли профессиональных и научных знаний при решении проблем данной области экологии [28]. Сегодня аналогичный протез может быть сделан относительно применения методов исследования операций при организации досуга [51, 75].

До сих пор исследователи систем основное внимание уделяли организации развлечений, включая путешествия и туризм, и организации спорта. Интересно отметить, что первые попытки привлечь внимание тренеров к методам исследования операций как к эффективному средству улучшения спортивной работы относятся еще к 1954 г. [94], а подобная попытка в области туризма была сделана лишь в 1969 г. [20].

Характерной особенностью первых количественных исследований в области развлечений и спорта является то, что в них почти не используются модели принятия решений. Очевидно, это объясняется тем, что в ряде случаев для моделей принятия решений необходимы данные, которые в свою очередь являются результатом прогноза, и требуется количественные измерения эффективности (оценка прибыли) каждого возможного решения. Поскольку в ранних работах основное внимание уделялось статистическому описанию различных проблем (например, проблемы определения относительной частоты перехвата и потери мяча в баскетболе), разработке моделей прогнозирования и способов оценки прибыли, получаемой от некоторых видов организации досуга, то, следовательно, эти работы можно рассматривать как необходимый этап процесса создания моделей принятия решений.

Все вышесказанное, а также тот факт, что вопросы применения методов исследования операций в области путешествий, туризма и спорта регулярно включаются в повестку дня объединенных совещаний ORSA и TIMS, свидетельствуют о важной роли методов исследования операций в деле организации досуга.

В данной главе рассматривается применение методов исследования операций при организации развлечений, путешествий, туризма и спорта, при этом особое внимание уделяется методам, разработанным для оценки результативности и выбора стратегии, определения состава команд и оптимальной пластов тренировок; обсуждаются наиболее новые и важные подходы к прогнозированию, оценке выгоды и построению моделей принятия решений.

1. ПРИМЕНЕНИЕ МЕТОДОВ ИССЛЕДОВАНИЯ ОПЕРАЦИЙ ПРИ ОРГАНИЗАЦИИ РАЗВЛЕЧЕНИЙ, ПУТЕШЕСТВИЙ И ТУРИЗМА

Прежде чем приступить к рассмотрению примеров применения методов исследования операций при организации развлечений, путешествий и туризма, необходимо уточнить, какую смысловую нагрузку несут эти три понятия.

Понятие «развлечение» определить довольно трудно, так как оно включает все виды занятий, которым посвящается досуг, или

свободное время [16]. Сущностью этого понятия является свободный выбор занятия, т. е., иными словами, развлечение — это свободно выбираемый вид деятельности.

Туризм можно рассматривать как один из видов развлечений, предполагающий передвижение. По определению словаря Вебстера [128], «турист — это тот, кто передвигается, или путешествует, для собственного удовольствия». Туристами, по определению Международного союза официальных туристических организаций (МСОТО), являются визитеры, останавливающиеся в посещаемой или стране не менее чем на 24 ч, причем целью такого посещения могут быть отдых, решение деловых или семейных вопросов, выполнение какого-то поручения или какая-то встреча.

Путешествие, в противоположность туризму, означает сквозной проезд или перемещение от пункта к пункту с использованием любого вида транспорта. В данном разделе рассматриваются, как правило, путешествия с целью развлечения, которые, по некоторым сведениям, составили в 1970 г. 75% всех путешествий в международном масштабе и почти половину всех путешествий в масштабе США. В то же время, по оценке МСОТО, ~75% расходов на туризм приходится на путешествия по стране, причем для Швейцарии эти расходы составляют 44%, а для Соединенных Штатов — 74% [84].

Между рассматриваемыми понятиями несмотря на различия существует тесная взаимосвязь, особенно это относится к туризму и развлечениям. Важность подобных «спиритических» связей подробно обсуждается в работе [43], автор которой в течение более чем 30 лет принимал участие в исследованиях по развитию туризма в различных зонах земного шара.

Первоначально предполагалось [20], что применение методов исследования операций при организации развлечений вне дома ограничивается в основном прогнозированием интенсивности людского потока из центров сосредоточения населения в места отдыха, и развлечений и оценке основного экономического эффекта, получаемого в сфере организации досуга. Однако, по мере того как повышался уровень исследований, проводимых в сфере организации досуга, появилась реальная возможность для разработки и использования модели принятия решений, и в первую очередь при планировании туризма.

1.1. Модели прогнозирования и модели распределения потоков туристов

Обзор моделей для предсказания ожидаемых потребностей в развлечениях вне дома дается в работе [130], в которой представлены модели прогнозирования, модели, использующие временные

ряды, экономические модели, модели притяжения и системного анализа и инерционные модели.

В данном разделе модели прогнозирования рассматриваются в соответствии с классификацией, предложенной в работе [22], т. е. модели, использующие временные ряды, причинно-следственные и качественные модели.

Анализ временных рядов. В работах [8—10] анализируется зависимость интенсивности туризма от фактора сезонности и иллюстрируется применение программы анализа временных рядов, разработанной в U. S. Bureau of the Census, для численной оценки сезонных колебаний и периодичности различных компонент туризма в Израиле. В частности, в работе [9] был проведен анализ временных рядов, содержащих информацию относительно следующих показателей: количество туристов, пользующихся услугами авиофлота, количество приезжающих туристов, объем иностранной валюты, поступающей в Израиль в результате туризма, число постоянных жителей Израиля, покидающих страну с целью туризма, число мест в гостиницах, предназначенных для туристов (только для иностранных туристов, только для местных туристов и общее число мест).

В работе [8] результаты анализа сравниваются с соответствующими данными по некоторым другим странам. Кроме того, показано, каким образом министерство по туризму Израиля использует результаты прогнозов для определения коэффициента максимального ежегодного «использования», величина которого зависит от показателя «сезонность» и представляет собой процент мест отдыха и развлечений, которые могут функционировать в течение всего года, от общего их числа, используемых в разгар сезона. На основе полученной таким образом информации проводятся исследования, целью которых является удлинение туристического сезона путем организации мероприятий в период межсезонья, т. е. введения второго (и по возможности третьего) сезона и т. д. К преимуществам удлиненного туристического сезона можно отнести следующие:

- лучшие условия для отдыха большего числа людей (имеется в виду отсутствие перепасленности, характерной для разгара сезона);

- лучшее использование средств обслуживания туристов и более равномерное распределение нагрузки на экономику и экологию соответствующего района;

- более равномерная занятость обслуживающего персонала гостиниц и других туристских учреждений, что позволяет улучшить качество ресторанный и других видов обслуживания.

Причинно-следственные модели. Баттелевский научно-исследовательский центр в Женеве создал одну из наиболее совершенных причинно-следственных моделей прогнозирования

потока международного туризма [4]. Модель является ступенчатой: вначале делается допущение о существовании простой зависимости между числом туристов, ежегодно прибывающих в j -ю страну из i -й страны, и численностью населения последней. Затем в модель последовательно вводится ряд независимых переменных, таких, как доход на душу населения, общность стран (например, языка), привлекательность (или популярность) данной страны для туристов. Согласно модели, между переменными существует следующая взаимосвязь:

$$N_{ij} = \frac{A_j P_i^{\alpha_j} G_i^{\beta_j} L_{ij}^{\gamma_j} T_n^{\epsilon_j}}{D_{ij}^{\sigma_j}},$$

где N_{ij} — число туристов из i -й страны, прибытие которых в j -ю страну зарегистрировано; P_i — численность населения i -й страны; G_i — доход на душу населения в i -й стране (валовой национальный продукт на душу населения); D_{ij} — расстояние между i -й и j -й странами; L_{ij} — оценка показателя общности, если таковая имеется (например, языка), i -й и j -й стран; A_j — оценка показателя популярности j -й страны у туристов; T_n — фиксированные моменты времени (в данном случае $n = 1973, 1967, 1975$ и 1980); $\alpha_j, \beta_j, \gamma_j, \epsilon_j, \sigma_j$ — коэффициенты эластичности соответствующих переменных для j -й страны.

Переменная T_n служит для оценки прогрессивных изменений за определенный период (до 1980 г.) таких факторов, как размещение населения, уровень образования и количество свободного времени.

Несмотря на то что в модели практически используются только те переменные, относительно которых имеется достаточный объем информации, прогнозы, полученные с помощью этой модели, достаточно хорошо согласуются с прогнозами, составленными организацией МСОТО. По-видимому, улучшение систем сбора информации о туризме позволит включить в модель другие переменные.

В работах [35, 36] перечисляются факторы, влияющие на распределение потоков туристов в районе Тихого океана. В частности, показано, как можно определить:

— «фактор места назначения», характеризующий особенности места назначения, благодаря которым оно становится привлекательным для туристов в течение определенного периода времени;

— «фактор связи», характеризующий различные виды связей, существующих между пунктом отправления и пунктом назначения в течение определенного периода времени. Фактор связи учитывает также средства передвижения и информации, т. е. два вида связи, необходимой для реализации возможности путешествий из одного пункта в другой.

Фактор места назначения и фактор связи вместе с такими фак-

торами. как общее число туристов, выезжающих из данного исходного пункта, и географическая удаленность пункта отправления от пункта назначения, становятся независимыми переменными в модели потоков, предназначенной для прогнозирования числа туристов, которые покидают пункт отправления и должны достичь пункта назначения за определенный период времени. Подобная модель была разработана и применена для изучения определенных районов Тихого океана с точки зрения возможного их использования для туризма.

Для предсказания будущего роста числа туристов в районах Тихого океана была использована модель множественной регрессии [30]. Эта модель является более традиционной в том смысле, что переменные причинно-следственной модели, по предположению, являются членами различных временных рядов экономического характера. (Полученные результаты были использованы в качестве вспомогательных данных при планировании.)

Качественные модели. Большинство традиционных потоковых моделей [125, 126] разрабатываются с учетом субъективных (качественных) данных и демонстрируют эффективность методов поиска согласованного решения и коллективного (группового) принятия решений. Впервые качественный подход к решению проблем организации досуга был предложен в работе [51]. Авторы работы использовали утверждение [14], согласно которому действие, подобное поездке одного лица или семьи к какому-нибудь озеру, можно рассматривать как результат процесса принятия решений. Такой процесс является умственным познавательным процессом, т. е. представляет собой цепь логических рассуждений и оценку результатов возможного образа действия с учетом потребностей, целей и возможностей данного человека или семьи. В соответствии с этим утверждением, процесс принятия решений можно представить в виде формализованной программы действий. Для этого необходимо сформулировать ряд таких утверждений, возражений или распоряжений, придерживаясь которых, можно помочь индивидууму сделать выбор, соответствующий его желаниям.

Эта программа, или схема исключений, может оказаться полезной для органов планирования и управления при прогнозировании действительного числа потенциальных туристов из множества возможных, т. е. при определении степени использования туристических объектов.

1.2. Определение выгод, получаемых от реализации мер по организации досуга

В процессе развития методов планирования туризма был разработан ряд подходов к оценке выгод. Для развивающихся стран, основной задачей которых является увеличение покупательной способности, создание отрасли организации досуга, и в особен-

ности туризма, можно рассценивать как одну из важных тактических мер, обеспечивающих приток иностранной валюты в страну. Объем иностранной валюты, поступающий в страну таким способом, можно рассматривать как непосредственную оценку прямых выгод, получаемых от туризма.

Для всестороннего планирования туризма в Израиле была предложена смешанная модель целочисленного программирования [7]. Целевая функция этой модели отражает вклад иностранной валюты, поступающей в страну вследствие туризма, в общую чистую прибыль. Согласно данным работы [104], дополнительная прибыль от туризма может быть получена в результате экономии средств на развитие местного туризма. Кроме того, автор работы предлагает для оценки вариантов решения относительно развития сектора туризма использовать приведенную чистую стоимость.

Наряду с прямыми выгодами от развития туристического сектора необходимо учитывать также и косвенные выгоды. В связи с этим возникает проблема определения эффекта мультипликации, обусловленного притоком денежных средств в результате туризма [3]. Исследуя состав косвенных и индуцированных потоков денежных средств, можно проанализировать, какие области экономики имеют приток средств в результате этого эффекта [4]. Целесообразность анализа, основанного на понятии мультипликатор, при разработке стратегии в области туризма обсуждается в работе [2].

Методы оценки выгод, не имеющих денежного выражения (например, от капиталовложений в организацию и строительство национальных парков, государственных мест отдыха и развлечений и т. д.), носят более субъективный характер. В настоящее время не существует единого мнения о возможности и целесообразности осуществления оценки таких выгод. Так, в работах [20, 29] стоимость какого-нибудь места отдыха определяется «готовностью заплатить», т. е. суммой, которую потребитель готов заплатить за пользование этим местом (стоимость проезда, стоимость входного билета и т. д.). В то же время авторы работы [85] считают, что не существует надежного и достоверного способа оценки выгод, не имеющих денежного выражения, поскольку трудно «оценить» всю значимость и полезность какого-нибудь места отдыха. Кроме того, они полагают, что правительственные капиталовложения в отрасль организации досуга должны способствовать максимальному социальному обеспечению. В качестве меры такого обеспечения можно использовать взвешенную сумму человеко-дней отдыха, при этом весовые множители определяются на основе политических, экономических и приоритетных показателей, одобренных правительством.

В настоящее время интенсивные исследования в рассматриваемой области проводятся по инициативе организации CORDS [42, 49].

Одним из возможных подходов к определению выгод, связанных с развитием туристического сектора, является метод оценки полезности, которая отражала бы степень привлекательности данного места для туристов. Метод такой оценки, рассмотренный в работе [2], сводится к определению полезности многих показателей, причем для нахождения весовых множителей используется процедура, предложенная в работе [25]. Данный подход может быть представлен следующим образом:

- выбор критерия оценки показателей привлекательности;
- определение относительной значимости показателей, т.е. определение весовых множителей;
- использование оценок экспертов для уточнения правильности выбора критерия оценки;
- определение численного значения функции полезности, или «относительной привлекательности», туристического района.

Данные, полученные в результате применения такого подхода, были использованы министерством по туризму Турции при разработке соответствующих планов и при распределении по категориям 65 географических районов Турции, представляющих интерес с точки зрения туризма. При этом численные значения функции полезности составляли как «инвентарную опись», или оценку состояния дел в области туризма в Турции.

1.3. Модели принятия решений

Модели принятия решений составляют основу дисциплины исследования операций [127]. Тем не менее первая попытка использовать подобные модели для организации досуга была предпринята лишь в 1972 г. [54] и представляла собой комплексное исследование в области планирования туризма в развивающейся стране. В частности, в рамках этого исследования была создана модель, результаты анализа которой явились основой для разработки Турцией соответствующей стратегии капиталовложений в сферу туризма на 3-ю пятилетку.

Задача была сформулирована следующим образом:
найти

$$\max \sum_{i=1}^N \sum_{j=1}^{K_i} d_{ij} X_{ij}$$

при ограничениях

$$\sum_{i=1}^N \sum_{j=1}^{K_i} C_{ij} X_{ij} \leq b,$$

$$X_{iL} - X_{iM} \geq 0 \text{ для } i = \overline{1, N} \text{ и } (L, M) \in P_i,$$

$$X_{ij} = 0 \text{ или } 1 \text{ для } i = \overline{1, N}, j = \overline{1, K_i},$$

где N — число туристических районов, на которые подразделяется страна (для Турции $N = 65$); K_i — число проектов, предложенных для реализации в i -м туристическом районе в течение планируемого периода (3-й пятилетний план для Турции); X_{ij} — переменная, принимающая значение 1, если j -й проект выбран для реализации в i -м туристическом районе, и значение 0 в противном случае; d_{ij} — оценка выгод, получаемых при условии, что j -й проект реализован в i -м туристическом районе (см. ниже); c_{ij} — ориентировочная стоимость (по завершении) j -го проекта, предложенного для реализации в i -м районе; P_i — индексное множество относительной значимости проектов, предложенных для реализации в i -м туристическом районе; b — ассигнования на развитие сектора туризма в течение планируемого периода.

Величина d_{ij} должна отражать относительную значимость возможных проектов и удовлетворять свойству аддитивности для всех проектов отдельного плана распределения, что позволяет определить выгоды от реализации общего плана. На первый взгляд может показаться, что в качестве оценки выгоды (прибыли) можно использовать приведенную чистую прибыль, получаемую от реализованного проекта. Однако при более внимательном рассмотрении такая оценка оказывается неприемлемой. Это связано с тем, что, во-первых, существует большое число различных проектов, рассматриваемых при планировании в региональном или национальном масштабе, а во-вторых, с тем, что помимо проектов гостиниц, дачных поселков и т. д., которые действительно вызывают приток средств, существуют такие, как проекты сооружения дорог, проведения археологических раскопок или строительства музеев, которые не дают выгоды в денежном выражении. Тем не менее реализация подобных проектов связана с определенными затратами, которые должны быть учтены при распределении и так довольно скудных ассигнований на развитие туризма.

Выбранная оценка выгоды представляла собой усредненное значение величины прибыли от различных возможных проектов. При рассмотрении совокупного воздействия развития туризма на страну через получаемую валюту использовалась смешанная оценка. В частности, рассматриваемой оценкой является предельная полезность, которая отражает вклад данного проекта в «усиление» показателя привлекательности данного региона для туристов [53]. Если $A_i^{(M)}$ — вектор, характеризующий привлекательность (по каждому из 17 критериев, как в примере с Турцией) i -го туристского района после завершения проекта M , включая все проекты, непосредственно ему предшествовавшие, то предельный вклад проекта M по каждому критерию выражается вектором «воздействия» $T_i^{(M)}$, который определяется следующим образом:

$$T_i^{(M)} = A_i^{(M)} - A_i^{(L)},$$

где L — множество проектов, непосредственно предшествовавших проекту M . Поясним сказанное на примере. Проект строительства гостиницы в определенном районе может рассматриваться только в том случае, если имеется решение о постройке дороги к месту, где будет располагаться гостиница. Следовательно, проект строительства дороги непосредственно предшествует проекту строительства гостиницы. Аналогичным образом любому проекту может предшествовать более чем один проект. Необходимо подчеркнуть, что для каждого туристического района существует некоторый запланированный проект и связанный с ним проект инфраструктуры и оба эти проекта являются предшественниками любого проекта для данного района. Задача состоит в том, чтобы, прежде чем предпринимать реальные шаги, обеспечить надлежащий уровень планирования развития района и таким образом гарантировать каждому туристическому району надлежащую инфраструктуру, которая удовлетворяла бы по крайней мере минимальным требованиям, предъявляемым к качеству обслуживания туристов.

Если через W обозначить вектор весовых множителей (или относительной значимости показателей привлекательности), то выгоды от реализации j -го проекта в i -м туристическом районе определяются следующим образом:

$$d_{ij} = W T_i^{(j)}, \quad i = \overline{1, N}, \quad j = \overline{1, K_i}.$$

Численное решение задачи осуществлялось с помощью методов линейного программирования [416], а решение общего вида приводится в работе [56].

Рассматриваемая модель была предметом многочисленных исследований, в результате которых были сделаны следующие выводы [55]:

- модель является статической одноперiodной моделью принятия решения;

- модель учитывает перспективу развития «туристической программы» страны в целом, но не учитывает различий географических районов с точки зрения интенсивности потока туристов;

- модель не отражает ограниченных возможностей развития туризма в отдельных географических районах. Одним из результатов исследований в области планирования туризма явилась разработка универсальной модели планирования туризма в Израиле [7, 56]. В данном случае задача формулировалась следующим образом:

найти максимальное значение целевой функции, представляющей собой общий объем чистой прибыли от туризма в иностранной валюте при условии, что критерием экономической эффективности является стоимость эквивалента добавочной чистой прибыли иностранной валюты.

при ограничениях на предложение, определяемое местными факторами планирования (характер сооружений, наличие рабочей силы, расположение мест отдыха и развлечений),

спрос в течение периода планирования (который определяется на основе экзогенного прогноза, проводимого при различных предположениях).

На основе анализа результатов применения данной модели можно осуществлять следующие операции:

- определение вероятного и оптимального выходного эффекта и его качественной структуры с точки зрения национальной экономики;

- уравнивание максимального спроса и предложения и определение соответственно излишка или дефицита;

- определение уровня затрат, необходимых для достижения оптимального выходного эффекта, контролируемого для основных факторов планирования;

- определение регионального распределения средств в зависимости от потенциальных возможностей развития туризма в каждом регионе;

- определение возможного и предпочтительного сезонного распределения различных показателей;

- определение скрытой цены, т. е. предельной стоимости, связанной с «изменением» основных предположений, входных данных и оптимальных результатов.

К важным моделям планирования туризма относится динамическая модель математического программирования, предназначенная для выбора проектов развития туризма в Пуэрто-Рико при ограничении на средства и численность квалифицированной рабочей силы [88].

Модель линейного программирования, представленная в работе [106], может быть использована управляющим отдельных государственных парков при планировании расширения предоставляемых этим парками услуг для удовлетворения возрастающих потребностей посетителей. Анализ этой модели даст информацию относительно числа дополнительных сооружений, которые можно построить при различном объеме капиталовложений, различных эксплуатационных затратах, спросе на земельные участки и общей потребности в местах развлечений.

1.4. Выводы

Наличие большого числа работ, посвященных вопросам прогнозирования в сфере организации досуга, можно объяснить в основном тем, что именно на этих вопросах было сконцентрировано внимание экономистов. Однако повышенный интерес к прогнозу привел к тому, что в рассматриваемой сфере деятельности в основ-

ном применяются экономические методы прогнозирования, классификация которых приводится в работе [26]. Очевидно, только этим можно объяснить тот факт, что до сих пор при решении проблем, связанных с организацией досуга вне дома, в должной мере не используются методы временных рядов [9, 10], и даже в тех случаях, когда требуется прогноз на короткий период или период средней продолжительности.

Новый подход к моделированию потоков туристов был предложен в работе [41]. Согласно этому подходу, туристический район рассматривается как совокупность привлекающих туристов мест (узлов), соединенных дугами, которые представляют собой потоки туристов. Подобное представление дает возможность проводить анализ полученной сети с помощью хорошо разработанных методов исследования операций, которые, однако, до сих пор не применялись в рассматриваемой области. Представляется, что в будущем этот подход найдет более интересное применение при решении проблем, связанных с прогнозированием потоков туристов.

В последнее время большое внимание уделяется проблеме создания эффективной системы сбора информации о местах отдыха и развлечений, которую можно было бы использовать для улучшения прогнозирования, управления и планирования [417].

Что касается методов оценки выгоды (прибыли) от реализации мер по организации отдыха и развлечений, путешествий и туризма, то здесь возникает ряд проблем, с которыми и теоретики и практики в области исследования операций уже сталкивались при исследовании систем городской сферы обслуживания и службы здравоохранения. Главная трудность заключается в том, что субъективные факторы, которые следует учесть при оценке выгоды, должны иметь численное выражение.

Существенным пробелом в области применения методов исследования операций в сфере отдыха и развлечений, путешествий и туризма является отсутствие фундаментальных работ по моделированию процессов принятия решений. Вместе с тем уже результаты первых исследований в этом направлении свидетельствуют о возможности применения формальных моделей. В настоящее время реализация такой возможности не вызывает никаких сомнений, если учесть, что в решении вопросов прогнозирования и оценки выгод (прибыли) достигнут настолько большой прогресс, что соответствующие результаты могли бы быть использованы как исходные данные для введения в математические и имитационные модели.

Современные исследования в области отдыха и развлечений вне дома ведутся по следующим направлениям [412]:

1. *Географические исследования на региональном (национальном) уровне*

Сбор данных на местах:

эстетическая оценка ландшафта, выраженная через показатель физической среды;

привлекательность места — естественная (природная) и искусственная (созданная человеком);

— степень изменения человеком естественного ландшафта с целью придания ему привлекательности.

Сбор и анализ статистических данных:

-- путешествия с целью развлечения;

-- организация туризма;

-- перспективная оценка потребностей в местах отдыха и развлечений для конкретного района;

— распространение новых веяний в проектировании средств обслуживания туристов.

2. Экономические, политические и социологические исследования

Изучение предложения и спроса в сфере отдыха и развлечений.

Анализ распределения ресурсов, анализ типа затраты — выгоды и соответствующих стратегий в области экономики.

Экономические аспекты проблемы отдыха и развлечений.

Социологическое и психологическое воздействие путешествий.

Другие возможные исследования экономического, политического и социологического характера.

3. Другие возможные направления исследований

Биологические исследования (экология).

Анализ технической осуществимости различных проектов, изучение особенностей местности с целью оптимального их использования при реализации проекта.

Пропускная способность мест отдыха.

Природа эстетической оценки данного места.

2. ПРИМЕНЕНИЕ МЕТОДОВ ИССЛЕДОВАНИЯ ОПЕРАЦИЙ В ОБЛАСТИ СПОРТА

Как уже отмечалось, первая работа, посвященная использованию методов исследования операций в области спорта, была опубликована в 1942 г. Автор этой работы писал: «Возможность применения научных методов при изучении спортивных игр до сих пор не изучена должным образом. Между тем предварительные данные по использованию методов исследования операций при изучении футбола и баскетбола показывают, что основные показатели подобных видов спорта могут быть описаны в терминах исследования операций. Вполне возможно, что использование научных методов в тренерской работе позволило бы существенно улучшить спортивные показатели и в частности игру команд». Аналогичная мысль была высказана в 1974 г. в работе [73]: «Спортивные состязания

могут оказаться плодотворной сферой применения методов теории управления».

Анализ видов спорта, по которым проводятся соревнования, с точки зрения исследования операций позволяет сделать следующие выводы: банки данных находятся в сравнительно хорошем состоянии и содержат новейшие сведения (люди проявляют громадный интерес к этим данным); действия повторяются, поэтому можно проводить неоднократные наблюдения процесса приблизительно в одних и тех же условиях; правила игр «точно определены», что исключает возможность «изменить игру» в самом ее разгаре: руководство в этой сфере, очевидно, достаточно восприимчиво к техническим нововведениям, поскольку тренеры активно ищут способы обеспечить так называемый «конкурентный перевес».

Можно предположить, что именно поиск конкурентного перевеса, который ведут тренеры, является причиной того, что многие особенно успешные случаи применения методов исследования операций в области спорта до сих пор остаются секретами, которые ревностно охраняются [27].

Тем не менее мы рассмотрим различные примеры применения методов исследования операций в области спорта, о которых сообщалось в печати.

2.1. Состав команд

Роберт Макол [86] одним из первых предложил использовать алгоритм решения задачи назначения для формирования команды. В частности, тренеру по плаванию при формировании команды для смешанной эстафеты следует отобрать пловцов, среди которых было бы несколько спортсменов, хорошо владеющих более чем одним стилем плавания. Усложненный вариант этой же идеи был использован при формировании основного состава команды из 38 кандидатов [99]. В этой же работе был предложен заслуживающий внимания способ определения коэффициентов целевой функции, который сводится к следующему.

Вначале каждый тренер составляет список показателей (как физических, так и умственных), которыми должен обладать тот или иной спортсмен, чтобы во время игры занять в команде место, которое для него отвел тренер. На основании этих списков делается попытка создать множество, состоящее из n показателей, взаимно исключающих друг друга и вместе с тем образующих исчерпывающий перечень, другими словами, делается попытка определить независимые показатели.

Затем каждый тренер определяет относительную значимость (весовой множитель) каждого показателя, включенного в перечень, с учетом места, которое занимает игрок во время игр, используя при этом метод, предложенный в работе [25]. Обозначим сводную

матрицу весовых множителей через B (b_{ij} — весовой множитель, приписываемый i -му показателю, если игрок занимает j -е место в команде).

После этого главный тренер определяет относительную значимость каждого из M мест в команде с точки зрения выявления потенциальных возможностей команды, вновь используя метод Черчмена — Аккофа (j -му месту приписывается весовой множитель W_j).

Наконец, тренерский состав оценивает каждого из L игроков по всем показателям, учтенным в перечне. Результаты представляются в виде матрицы A (a_{ij} — оценка i -го игрока по j -му показателю).

В результате получают оценку c_{ij} возможности i -го игрока занять j -е место в команде, т. е.

$$c_{ij} = \sum_{k=1}^n a_{ik} b_{kj}.$$

При необходимости можно также оценить вклад d_{ij} в игру команды i -го игрока, занимающего j -е место:

$$d_{ij} = c_{ij} W_j.$$

Таким образом, задача сводится к задаче назначения, решение которой находится с помощью стандартного метода линейного программирования. Другими словами, определяются некоторые переменные X_{ij} (которые принимают значения 0 или 1) так чтобы максимизировать сумму величин $d_{ij} X_{ij}$ по всем i -м местам и всем j -м игрокам при условии, что все места заняты и каждый игрок назначается не более чем на одно место. Этот метод оказался весьма эффективным при решении проблемы расстановки игроков.

2.2. Оценка стратегии в командных видах спорта

Среди всех видов спорта, пожалуй, бейсбол наиболее часто анализировался с помощью формальных методов. Тем не менее можно перечислить всего несколько работ, которые действительно можно расценивать как серьезные исследования, посвященные использованию применения методов исследования операций при решении проблем этого вида спорта [81]. Одна из наиболее известных книг о бейсболе [32], обзор которой был опубликован в журнале *Operations Research*, была расценена как работа, которая интересна для болельщиков, но не должна ни в коем случае попадаться на глаза тем, кто изучает теорию вероятностей [81]. Это замечание справедливо не только по отношению к данной книге, но и ко всем работам о бейсболе в области исследования операций, выполненных до настоящего момента. По существу, все отмеченные

работы применяют методы статистики лишь для анализа данных. Например, в работе [81] вычисляется математическое ожидание трех показателей качества игры команды: зависимости среднего числа подач от мастерства ловящего и подающего игроков; биномиальное распределение попаданий; попадание при полной защите. В одной из более поздних работ [80] исследуются некоторые особенности стратегии максимизации вероятности выигрыша, а в работе [50] описано построение с помощью метода Монте-Карло имитационной модели на основе данных *Sports Illustrated*, раздел «Бейсбол». (Заметим, что сведения о бейсболе являются наиболее доступными по причине его выгодности с коммерческой точки зрения.) Эта имитационная модель была использована для проверки некоторых гипотез, основанных на предположениях, высказанных Куком [32].

Исследования, посвященные бейсболу, побудили авторов работы [18] проанализировать 8373 партий из первых 56 туров, включенных в таблицу розыгрыша 1969 г. Национальной футбольной лиги. В частности, было вычислено ожидаемое время владения мячом при условии, что игра начинается из положения «вне игры» и игрок проходит десять ярдов. В результате проведенного анализа были сделаны два важных вывода относительно стратегии нападения. Во-первых, было показано, что в ситуации, когда было объявлено положение вне игры и забит гол, необходимо стремительно проводить мяч в ворота противника, а не перебрасывать его, как это обычно делается в случае атаки при попытке забить гол. Прямая проводка мяча может применяться и в особых обстоятельствах, например при необходимости забить победный гол, когда до конца игры остается всего несколько секунд, и т. д. Во-вторых, было показано, что, если команда проигрывает не более чем 7 очков, она не должна требовать минутного перерыва в следующих случаях: до конца игры остается более 30 с, и при этом мячом владеет команда; до конца игры остается более 1 мин, и при этом мячом владеет противник.

В работе [11] анализируется случай, имевший место в 1967 г., когда тренер футбольной команды Cornell в ходе игры с командой Harvard принял решение добиться выигрыша в два очка, причем Harvard выигрывал со счетом 14 : 6. Анализ дерева решений продемонстрировал, что при наличии трех вполне разумных предположений тренер принял правильное решение, хотя, к сожалению, действительный результат игры противоречил данным анализа, поскольку попытка выиграть два очка не удалась и Cornell проиграл со счетом 14 : 12. Этот случай был подробно проанализирован в работе [89]. Хотя результаты анализа в основном согласуются с данными работы [11], тем не менее было показано, что решение тренера Cornell сыграть вничью могло быть нелогичным.

Было показано, что оптимальная стратегия выигрыша чемпио-

ната Национальной футбольной лиги может включать такой вариант, как поражение в отдельных играх [111]. В качестве основного довода приводятся игры сезона 1971 г., когда команде Washington Rid было гарантировано участие в решающих встречах. Накануне последней игры с командой Cleveland, независимо от исхода игры, Washington Rid должна была участвовать в чемпионате центральной группы. Однако, если бы они победили Cleveland, им бы пришлось играть против сильной команды в первом туре решающих встреч, в то время как в случае поражения (так, кстати, и произошло) они играли бы против сравнительно слабой команды. Другими словами, войдя в центральную группу, команда Washington Rid должна была бы провести две трудные встречи вместо одной. Данную ситуацию можно проанализировать с использованием цепей Маркова. В рамках такого анализа можно разработать правила принятия решений, которым команда должна следовать, чтобы определить, при каких обстоятельствах стремиться к победе и при каких смириться с поражением.

2.3. Оценка стратегии в индивидуальных видах спорта

В работах [73, 74]¹⁾ приводится модель оптимизации начальных условий, которые выбирает спортсмен во время соревнований в прыжках с шестом или в высоту. Задача состоит в том, чтобы оценить максимально возможный ожидаемый результат спортсмена в зависимости от первоначально установленной высоты. В соревнованиях по прыжкам борьба начинается с того, что кто-то из участников пытается взять сравнительно небольшую высоту, на которой устанавливается планка. Затем в ходе соревнований планка постепенно поднимается. Каждому участнику предоставляются три попытки для преодоления данной высоты, и после этого спортсмену разрешается попробовать свои силы в следующем зачете. Спортсмену засчитывается последняя (наибольшая) взятая высота независимо от результатов предыдущих попыток.

Каждому участнику соревнований разрешается начать прыжки с любой высоты, но не меньшей, чем некоторая начальная (квалификационная). Начиная прыжки с большой высоты, спортсмен меньше устает, и вероятность успеха преодоления последующей высоты повышается, однако в случае неудачи его результат будет равен нулю. Данная модель позволяет определить оптимальную высоту, с которой спортсмену целесообразно начать свое выступление.

¹⁾ Интересно отметить, что автор работ [73, 74], будучи специалистом-практиком в области исследования операций, известен как спортсмен и участник Олимпийских игр.

Основная проблема, с которой приходится сталкиваться штабистам в ходе соревнований, заключается в правильном распределении сил с учетом того, что каждому участнику соревнований предоставляются всего лишь три возможности взять вес. При этом за результат принимается наибольший из взятых спортсменом весов. Каждый последующий вес, назначаемый спортсменом, не может быть меньше предыдущего. В том случае, когда спортсмен начинает соревнования со слишком малого веса, вероятность того, что он победит, практически равна нулю. Напротив, если спортсмен начинает выступление со слишком большого веса, не исключено, что он получит нулевой результат. Процесс принятия решений можно представить в виде дерева решений. На основании такого представления формируется модель нелинейного программирования, целевая функция которой представляет собой ожидаемую полезность и ее максимизация проводится с учетом ограничений, налагаемых ходом соревнований.

Рассмотрим вероятностную модель игры в гольф [100], в основе которой лежат следующие предположения:

- мастерство игрока (или качество игры) характеризуется исключительно распределением вероятностей попадания мяча в каждую из 18 лунок игровой площадки;

- это распределение является стационарным;

- количество очков, выигранных по каждой лунке, является независимым;

- очки по лунке, выигранные одним игроком, являются независимыми по отношению к очкам других игроков.

Эта модель позволяет определить вероятность того, что игрок A победит игрока B в игре, в которой счет ведется по числу ударов, или в матчевой встрече; вероятность того, что игроки A и B закончат встречу в смешанной игре вничью; вероятность того, что игроки A и B закончат встречу в матчевой игре вничью. Все эти величины могут быть определены также для любого игрока A , имеющего фору H_A и играющего против игрока B , имеющего фору H_B . (Наоборот, зная вероятностное распределение очков для одного игрока, можно определить целесообразную величину предназначенной для него форы.)

В работе [107] предлагается метод определения величины форы при игре в гольф с гандикапом, основанный на моделировании более миллиона матчей. При моделировании были использованы данные, хранящиеся в картотеках гольф-клуба. Полученные зависимости отражают процент игр, выигранных сильнейшим игроком, как функцию от разности установленных гандикапов и форы, действительно предоставленной сильнейшим игроком слабейшему. Результаты исследования, проведенного с помощью данного метода, свидетельствуют о том, что современная система правил игры в гольф с гандикапом, принятая United States Golf Association,

даст более сильному игроку несправедливое преимущество в соревнованиях, в которых принимают участие два игрока.

В работе [65] дается описание модели для определения вероятности победы в игре двух противников; модель позволяет оценить воздействия правил игры на вероятность завоевания победы данным игроком, в качестве примеров рассматриваются теннис и бадминтон. В более поздней работе [57] приводится вероятностная модель, позволяющая представить вероятности выигрыша в теннисе как функцию от различной стратегии подачи мяча. В частности, показано, что не всегда оптимальным решением является осуществление первой и второй подачи различными способами. Анализ иллюстрируется данными турнира профессионалов.

2.4. Организация тренировок и формирование команд

В работе [66] рассматриваются проблемы, связанные с осуществлением руководства футбольной командой основной лиги, при этом главное внимание уделяется формированию состава команды. Эта проблема приобретает особую актуальность в связи с тем, что в состав футбольной команды на международных соревнованиях может быть заявлено, самое большее, 20 игроков, а в одной игре могут участвовать не более 14 спортсменов. Поэтому в работе детально рассматриваются вопросы «оптимальной» избыточности. Кроме того, в ней анализируются вопросы, связанные с оптимальной возрастной структурой команды, с ее иерархической структурой, с определением полупериода восстановления, т. е. полного цикла игрока.

Ладани [75] разработал модель, которая, очевидно, может оказаться эффективной при разработке программ тренировок. В работе рассмотрен пример применения этой модели при разработке оптимальной программы еженедельных тренировок для пятиборца. Переменные модели (X_i) представляют собой количество часов в неделю, которые спортсмен должен отводить для i -го типа тренировок. Тренировки делятся на физические упражнения, которые улучшают общую физическую подготовку спортсмена ($i = 1, 5$), и на специальные физические упражнения ($i = 6, 10$).

На первом этапе формирования модели необходимо было точно определить причинно-следственные связи между результатом, показанным в каждом виде соревнования (пятиборье включает 5 видов), и временем, потраченным на каждый вид тренировок. В модели было гипотетически принято, что такая связь может быть выражена следующим соотношением:

$$Y_j = a_{0j} + \sum_{i=1}^{10} a_{ij}x_i,$$

где Y_j — результат в j -м виде соревнований, выраженный в соответствующих единицах (т. е. в секундах для беговых видов, в метрах для прыжков и метаний). Коэффициенты в приведенном соотношении вычислялись путем использования пошаговой линейной регрессии на собранных данных. Интересно, что для всех пяти видов соревнований самый низкий коэффициент корреляции был равен 0,98.

Следующий этап формирования модели состоял в том, чтобы представить достигнутый в сумме пятиборья результат в виде линейной функции от результатов в каждом виде соревнования. Это было сделано только после того, как вклад от прыжков в длину в общий результат для пяти видов был аппроксимирован кусочно-линейной функцией. Полученная линейная функция затем максимизировалась при следующих условиях:

- общее время, которое в течение недели спортсмен может уделить тренировкам, ограничено;
- общее время, посвященное в течение недели силовым тренировкам, не должно превышать точно установленного максимума;
- объем тренировок на скорость должен быть по крайней мере равен объему тренировок на выносливость;
- объем тренировок по общей физической подготовке должен быть больше, чем объем тренировок по отработке техники;
- количество часов в неделю, посвященных тренировкам по отработке техники, должно составлять 50—200 % среднего количества часов, посвященных этим тренировкам к данному моменту времени;
- максимально и минимально допустимое количество часов, посвященных развитию мускульной силы, аэробному и анаэробному развитию тела для увеличения выносливости, должны быть определены так, чтобы предотвратить перенапряжение и его вредные последствия.

Все эти ограничения были выражены линейными функциями, и оптимальные значения переменных X_i модели были определены с помощью методов линейного программирования. Полученные результаты оказались вполне приемлемыми. Однако при их использовании необходимо принимать во внимание ограничения, заложенные в модели.

2.5. Выводы

Нельзя не согласиться с мнением авторов работ [78, 94], что спорт является благодатной областью для применения методов исследования операций. Тем не менее, если судить по числу опубликованных работ, то именно эта область до сих пор не привлекала

к себе должного внимания со стороны исследователей систем (как практиков, так и теоретиков).

Любое спортивное соревнование по своей природе является игрой. Естественно было бы проводить анализ соответствующих ситуаций с помощью хорошо известных и разработанных методов теории игр. Однако до настоящего времени эти методы практически не применялись в области спорта. Это не может не вызывать удивления, особенно если учесть, что в области спорта мы почти всегда располагаем большим количеством данных, а иногда и «живой лабораторией» в виде коллектива спортсменов во главе с благожелательно настроенным и заинтересованным тренером.

ЗАКЛЮЧЕНИЕ

В пределах одной главы не представляется возможным рассмотреть все проблемы, связанные с организацией досуга вне дома, с точки зрения их возможного решения с помощью методов исследования операций. Поэтому мы ограничились анализом основных проблем, с которыми приходится сталкиваться в области туризма и спорта. При этом были упомянуты наиболее интересные прикладные работы.

В конце главы приводится перечень работ, которые в какой-то мере могут оказаться полезными для читателя, если у него возникнет необходимость в дополнительной или более подробной информации по затронутым вопросам.

ЛИТЕРАТУРА

1. Archer B. H., The Primary and Secondary Beneficiaries of Tourist Spending, *Tourist Rev.*, 27, 42—45 (1972).
2. Archer B. H., The Uses and Abuses of Multipliers, A Paper presented the Fourth Annual Travel Research Conference held in Sun Valley, Idaho, August 12—15, 1973.
3. Archer B. H., Owen C. B., Towards a Tourist Regional Multiplier, *Regional Stud.*, 5, 224—294 (1971).
4. Armstrong C. W. G., International Tourism: Coming or Going, Futures, June 1972, pp. 115—125.
5. Artus R., An Economic Analysis of International Travel, Staff Papers—International Monetary Fund, XIX, № 3, 579—614 (November 1972).
6. Auger J., Analyse Factorielle et Evaluation d'Attraction de Sites Recreatifs, Rapport Technique № 2, Gouvernement du Quebec, Ministere du Tourisme, de la Chasse et de la Peche, Service de la Recherche Quebec, Canada, January 1974.
7. Bargar J., Arbel A., A Comprehensive Approach to the Planning of the Tourist Industry, Paper presented at the ORSA/TIMS Joint National Meeting, San Juan, Puerto Rico, October 16—18, 1974.
8. Baron M., Shechter N., Simultaneous Determination of Visits to a System of Visits to a System of Outdoor Recreation Parks with Capacity Limitations, *Regional and Urban Econ.*, 3, № 4, 327—359 (1973).
9. Baron R. R., Seasonality in Tourism—Part 1, *Int. Tourism Quart.*, Special Article № 6 (1972).

10. Baron R. R., Seasonality in Tourism — Part 2, *Int. Tourism Quart.* Special Article No. 6 (1973).
11. Bierman H., Jr., Letter to the Editor, *Management Sci.*, 14, No. 6, B281—B282 (1968).
12. Beaman J., Distance and the Reaction to Distance as a Function of Distance, *J. Leisure Res.*, 6, 220—231 (Summer 1974).
13. Beaman J., Three Methods for Measuring the Attractiveness of a Park — a Comparison, CORD Technical Note, No. 9, Parks Canada, Ottawa, Canada, May 1974.
14. Bowman E. H., Consistency and Optimality in Managerial Decision Making, *Management Sci.* (January 1963).
15. Burd O. R., Brewer D., Estimation of Net Social Benefits from Outdoor Recreation, *Econometrica*, 9, No. 5 813—827.
16. Burton L., Noad P. A., Recreation Research Methods, Occasional Paper No. 3, Centre for Urban and Regional Studies, University of Birmingham, Birmingham, England, 1968.
17. Canadian Government Travel Bureau, Canada-Bibliography Tourism Research Studies, Ottawa, Canada, March 1972.
18. Carter W., Machol R. E., Operations Research in Football, *Operations Res.*, 19, 541—544 (1971).
19. Cesario J., Final Report on Estimating Park Attractiveness, Population Centre Eissiveness and the Effect of Distance (Location) on Outdoor Recreation Travel, Cord Technical Note No. 4 (Mimeo), Department of Indian and Northern Affairs, Ottawa, Canada, 1973.
20. Cesario J., Operation Research in Outdoor Recreation, *J. Leisure Res.*, 1, No. 1, 33—51 (Winter 1969).
21. Cesario J., Goldstone S. E., Knetsch J. L., A Report on Outdoor Recreation Demands and Values to Middle Atlantic Utility Group, Battelle Memorial Institute, Columbus, Ohio, 1969.
22. Chamber J. C., Mullick S. K., Smith D. D., How to Choose the Right Forecasting Technique, *Harvard Bus. Rev.*, 57—86 (July — August 1971).
23. Cheung H. K., A Day-Use Park Visitation Model, *J. Leisure Res.*, 3, 139—142 (Spring 1972).
24. Chubb M., Outdoor Recreation Planning in Michigan by a Systems Analysis Approach, Part III — The Practical Application of Program RECSYS and SYMAP, Technical Report No. 11, Department of Resource Development, Michigan State University, East Lansing, Michigan, 1968.
25. Churchman C. W., Ackoff R. L., An Approximate Measure of Value, *Operations Res.*, 2, No. 2, 172—187 (1954).
26. Cicchetti C. J., Fisher C., Smith V., Economic Models and Planning Outdoor Recreation, *Operations Res.*, 21, No. 5, 1104—1113 (September — October 1973).
27. Clarke R., Trengove A., Are we Creating Super Athletes — Or Monsters? Readers Digest, October 1975, pp. 141—145.
28. Clawson M., Measuring Outcomes in Terms of Economic Implication for Society, in Recreation Research, American Association for Health, Physical Education, and Recreation, 1966.
29. Clawson M., Knetsch J. L., Economics of Outdoor Recreation, The Johns Hopkins Press, Baltimore, Maryland, 1966.
30. Cline R. S., Measuring Travel Volumes and Itineraries and Forecasting Future Travel Growth to Individual Pacific Destination, Paper presented at the ORSA/TIME Joint National Meeting, San Juan, Puerto Rico, October 16—18, 1974.
31. Cohen E., Toward a Sociology of International Tourism, *Social Res.*, 164—181 (Spring 1972).
32. Cook E., Percentage Baseball, M. I. T. Press, Cambridge, Massachusetts, 1967.

33. Cotton W. R., Intervening Opportunities Barriers or Stepping Stones?, *Pacific Soc. Rev.*, 8, 75-81 (1965).
34. Crackwell B., Accessibility to the Countryside as a Factor in Planning for Leisure, *Regional Stud.*, 1, 50 (1967).
35. Crampon L. J., Tau K. T., A Model of Tourism Flow into the Pacific, *Tourist Rev.*, 98-104 (July - Sept. 1973).
36. Crampon L. J., Factors Influencing Travel Flow into and Within the Pacific Basin, Paper presented at the ORSA/TIMS Joint National Meeting, San Juan, Puerto Rico, October 16-18, 1974.
37. Crevo C. C., Characteristics of Summer Week-End Travel, *Highway Res. Record*, 44, 51-60 (1963).
38. Croft D. J., An Application of Two-Phase Discriminant Analysis to Employee Selection, Paper presented at the ORSA/TIMS Joint National Meeting, San Juan, Puerto Rico, October 16-18, 1974.
39. Davis D. H., Potentials for Tourism in Developing Countries, *Finan. and Dev.*, 5, No. 4, 35-39 (December 1968).
40. Deasy G. F., Gress P. R., Impact of Tourist Facility on its Hinterland, *Ann. Assoc. Am. Geographers*, 56, No. 2, 250-306 (1966).
41. Demers J., Tourist Flows, Paper presented at the ORSA/TIMS Joint National Meeting, San Juan, Puerto Rico, October 16-18, 1974.
42. Dubos R., The Human Environment, *Ekistics*, 178, 170-173 (1969).
43. Egan M., Interfaces Between Tourism and Outdoor Recreation. Paper presented at the Western Economic Association Annual Conference, San Diego, California, June 25-28, 1975.
44. El Hodiri M., Quirk J., An Economic Model of a Professional Sports League, *J. Pol. Economics*, No. 79, 1302-1319 (November - December 1971).
45. Ellis J. B., Analysis of Socio-Economic Systems by Physical Systems Techniques, unpublished Ph. D. dissertation, Department of Electrical Engineering, Michigan State University, East Lansing, Michigan, 1965.
46. Ellis J. B., Milstein M. D., Michigan Outdoor Recreation Study, Vol. 1, Methods and Models, Technical Report No. 6, Michigan Department of Conservation, Lansing, Michigan, 1966.
47. Ellis J. B., Systems Analysis of Provincial Park Camping - 1966 Park User Survey, Parks Branch, Ontario Department of Lands and Forest, Toronto, Ontario, Canada, 1968.
48. Ellis J. B., A System Model for Recreational Travel in Ontario - Further Results, Ontario Department of Highways, RR 148, Toronto, Ontario, Canada, July 1969.
49. Forster J., The Sociological Consequences of Tourism, *Int. J. Comparative Sociology*, 5, No. 2, 217-227 (1964).
50. Freeze R. A., An Analysis of Baseball Batting Order by Monte Carlo Simulations, *Operations Res.*, 22, No. 4, 728-735 (1974).
51. Gaumnitz J. E., Swinth R. L., Tollefson J. O., Simulation of Water Recreation User's Decisions, Paper presented at the Time XX International Meeting, Tel Aviv, Israel, June 26, 1973; to appear in *Land Econ.*
52. Gearing C. E., Swart W. W., Var T., A Decision Structure for Touristic Investment Allocation, *Tourist Rev.*, 27, No. 1, 2-13 (1972).
53. Gearing C. E., Determining the Optimal Investment Policy for the Tourism Sector of a Developing Country, *Management Sci.*, 20, No. 4, 487-497 (December 1973). Part I.
54. Gearing C. E., Establishing a Measure of Touristic Attractiveness, *J. Travel Res.*, XII, No. 4, 1-8 (Spring 1974).
55. Gearing C. E., A Multi-Period Planning Model for Tourism Development, Paper presented at the TIMS XXII International Meeting, Kyoto, Japan, July 26, 1975.
56. Gearing C. E., Planning for Tourism Development: Quantitative Approaches, Praeger Publishers, New York, 1976.

57. George S. L., Optimal Strategy in Tennis — A Simple Probabilistic Model, *Appl. Stat.*, **22**, 97—104 (1973).
58. Goeldner C. R., Gearing C. E., Swart W. W., Var T., Bibliographic Sources for Travel and Tourism Research, *Tourist Rev.*, **29**, No. 3, 197, Geneva (1974).
59. Goeldner C. R., General Information Regarding Organizations Engaged in Tourism and Travel Research Activities, in: Planning for Tourist Development: Quantitative Approaches, by C. E. Gearing, W. W. Swart and Turgut Var. Praeger Publishers, N.Y., 1976.
60. Grubb H., Goodwin J., Economic Evaluation of Water Oriented Recreation in the Preliminary Texas Water Plan, Report No. 84, Texas Water Development Board, Austin, Texas, 1968.
61. Harmston F. K., Differences in Secondary Impact of Through and Recreational Travelers on a Small Western Community, A paper presented to the Western Economic Association, Annual Meeting San Diego, Calif., June 1975.
62. Hecock H. H., Public Beach Recreational Opportunities and Patterns of Consumption on Cape Cod, Unpublished Ph. D. Dissertation, Department of Geography, Clark University, 1966.
63. Hill M., Schechter M., Optimal Goal Achievement in the Development of Outdoor Recreation Facilities, in: Urban and Regional Planning, ed. by A. G. Wilson, Pion Ltd., London, 1971, pp. 110—112.
64. Holman M. A., A National Time — Budget for Year 2000, *Sociology and Social Res.*, **42**, 1 (October 1961).
65. Hsi B. P., Burych D. M., Games of Two Players, *Appl. Stat.*, **20**, 86—92 (1974).
66. Huska A. M., An Experimental Model for the Management of a Major League Soccer Team. Paper presented at the ORSA/TIMS Joint National Meeting, Boston, Mass., April 22—24, 1974.
67. Johansson J. K., Sunday A. A., Simon J. L., International Advertising and Tourism, Paper presented at the ORSA/TIMS Joint National Meeting, San Juan, Puerto Rico, October 16—18, 1974.
68. Jud G., Drause W., Evaluating Tourism in Developing Areas: An Exploratory Inquiry. A paper presented to the Western Economic Association, Annual Meeting, San Diego, California, June 1975.
69. Kahn T. D., Employment in the Texas Tourist Industry, A paper presented to the Western Economic Association, Annual Meeting, San Diego, Calif., June 1975.
70. Kain J. K., Meyer J. R., Computer Simulations, Physio-Economic Systems and Interregional Models, *The American Economic Review* (Papers and Proceedings of the 80th Annual Meeting in Washington, D.C. on December 28—30, 1967) LVIII, No. 2. 171—181 (May 1968).
71. Kottas J. F., Spectator Seating Preferences as a Determinant in Design and Pricing Decisions, DMG-DRS Journal, Design and Methods, 1975.
72. Ladany S. P., Optimal Motel Room Rental Policy, Paper presented at the ORSA/TIMS Joint National Meeting San Juan, Puerto Rico, October 16—18, 1974.
73. Ladany S. P., Optimal Starting Height for Pole Vaulting, *Operations Res.*, **23**, No. 5, 968—978 (1975).
74. Ladany S. P., Optimization of Pentathlon Training Plans, *Management Sci.*, **21**, No. 10, 1144—1155 (1975).
75. Ladany S. P., (Ed.), Management Science Application to Leisure-Time Operations, North-Holland Publishing Co., 1975.
76. Lentneck B., et al., Spatial Behavior in Recreational Boating, *J. Leisure Res.*, **1**, 103—124 (Spring 1969).
77. Leopold L. B., Esthetics of Landscape, *Natural History Magazine*, **78**, No. 4, 37—44 (October 1969).

78. Lilien G. L., Optimal Lift Selection: The Weightlifter's Dilemma, Paper presented at the Joint National ORSA/TIMS Meeting, Boston, Massachusetts (April 22—24, 1974).
79. Lindsey G. R., Statistical Data Useful for the Operation of a Baseball Team, *Operations Res.*, 7 (1959).
80. Lindsey G. R., An Investigation of Strategies in Baseball, *Operations Res.*, 16, No. 5, 1088—1089 (1963).
81. Lindsey G. R., Book Review: Percentage Baseball by E. Crock, *Operations Res.*, 16, No. 5, 1088—1089 (1968).
82. Arthur D. Little, Inc., Tourism and Recreation: A State-of-the Art Study, U. S. Department of Commerce, Washington, D. C., 1967.
83. Lucas R., Wilderness Perception and Use, *Natural Resources J.*, 3, No. 3, 393—411 (January 1964).
84. Lundberg D. E., The Tourist Business, Institutions/Volume Feeding Magazine, 1972, p. 11.
85. Mack R. P., Myers S., Outdoor Recreation from Measuring Benefits From Government Investments, R. Dorfman (Ed.), Brooking Institute, Washington, D. C., 1964.
86. Maxhol R. E., An Application of the Assignment Problem, *Operations Res.*, 18, No. 4, 569—760 (1970).
87. Maki D. R., Beck R. A. D., Recreational Programming: A British Columbia Example Working Paper, Department of Economics and Commerce, Simon Fraser University, Burnaby, B. C., No. 75-5-2., 1975, p. 1.
88. McMahon J. J., Tourism Planning and Project Evaluation — The Puerto Rican Case, Paper presented at the ORSA/TIMS Joint National Meeting, San Juan, Puerto Rico, October 16—18, 1974.
89. Means E. H., More on Football Strategy, *Management Sci.*, 15, No. 2, B-15, B-16 (1968).
90. Mercer D., Discretionary Travel Behavior and the Urban Mental Map, *Australian Geographic Studies*, October 1, 1971.
91. Mercer D., Urban Recreational Hinterland, *The Professional Geographer*, 22, 77 (March 1970).
92. Ministère du Tourisme, de la Chasse et de la Pêche, Service de la Recherche, Gouvernement du Québec, Le Flux Touristiques: Project de Recherche et de Développement de Mesure de d'Analyse, Part I, Prepared by Urbatique, Inc. Québec, Canada, 1973.
93. Mitchell L. S., The Facility Index as a Measure of Attendance at Recreational Sites, *The Professional Geographer*, 20, No. 4, 276—278 (1968).
94. Mottley C. M., The Application of Operations Research Methods to Athletic Games, *Operations Res.*, 2, No. 3, 335—338 (1954).
95. Nagashima K., The Human Environment: Evolution of Social and Leisure Space with Reference to Japan, *Eskistics*, 168, 1461—1468 (1970).
96. Nijkamp P., Environmental Attraction Factors and Regional Tourist Effects, Research Memorandum No. 14, Free University, Amsterdam The Netherlands, 1975.
97. OECD (Organization for Economic Cooperation and Development), Government Policy in the Development of Tourism, Paris, 1974.
98. O'Riordan T., Planning to Improve Environmental Capacity: A Case Study in Broadland, *Town Plan. Rev.*, 40, No. 1, 39—58 (1969).
99. Patterson J. H., Wolfe R. A., An Application of the Assignment Algorithm to Football Player/Position Selection, Paper presented at the 41st National ORSA Meeting, New Orleans, Louisiana, April 1972.
100. Pollock S. M., A Model for the Evaluation of Golf Handicapping, Paper presented at the ORSA/TIMS Joint National Meeting, Boston, Massachusetts, April 1974.
101. Powers T. A., Appraising International Tourism Projects, Papers on Project

- Analysis No. 1, Inter-American Development Bank, Washington, D. C., July 1974.
102. Quandt R. E. (Ed.), *The Demand for Travel: Theory and Measurement*, D. C. Heath Company, Lexington, Massachusetts, 1970.
 103. Renoux M., *Rapport Methodologique No. 1 Technique Econometriques de la Demand touristiques et Amorce de leurs Interfractions dans un Systeme Decisionnal*, Gouvernement du Quebec Ministere du Tourisme, de la chasse et de la Peche, Service de la Recherche Quebec, Canada, 1972.
 104. Renoux M., *Rapport Methodologique, No. 3. Vol. 1, Les Modeles de Simulation Appliques au Tourisme et aux Laisirs de Plein Air et leur Integration dans un System Decisionnel*, Gouvernement du Quebec, Ministers du Tourisme, de la Chasse et de la Peche, Service de la Recherche, Quebec, Canada, 1973.
 105. Ross J. H., *A Measure of Site Attraction*, Occasional Paper No. 2, Environment Canada, Ottawa, Canada, 1973.
 106. Saitta W. W., Schedeman I. W., *New Dimension in Park Management through Linear Programming*, *G. Leisure Res.*, 4, No. 4, 333—340 (1972).
 107. Scheid F., *A Least Squares Family of Cubic Curves with an Application to Golf Handicapping*, *SIMA J. Appl. Math.*, 22, No. 1, 77—83 (1972).
 108. Scully G. W., *Pay and Performance in Major League Baseball*, *Am. Econ. Rev.*, LXIV, No. 6, 913—930 (1974).
 109. Shafer E. L., Thompson R. C., *Models that Describe Use of Adirondack Campgrounds*, *Forest Service*, 14, 383—391 (December 1968).
 110. Shelley M., *Competition for Satisfaction Resouces between Tourist and the Inhabitant*, a paper submitted in the XXth TIMS International Meeting in Tel Aviv, June 25, 1973.
 111. Silverman D., Schwartz B., *How to Win by Losing*, *Operations Res.*, 21, No. 2, 639—643 (1973).
 112. Smith S. L., *An Exegesis of Outdoor Recreation Research*, Special Report, College of Agriculture Texas Agricultural Experiment Station, Texas A. M. University: Department of Recreation and Parks, College Station, Texas, 1971.
 113. State of Florida, Department of Natural Resources, *Outdoor Recreation in Florida*, Tallahassee, Florida, August 1971, p. 7.
 114. Stouffer S. A., *Intervening Opportunities and Competing Migrants*, *J. Regional Sci.*, 2, 1—26 (1960).
 115. Swart W. W., Gearing C. E., Var T., *A Dynamic Programming — Integer Programming Algorithm for Allocating Touristic Investments*, *Tourist Res.*, 27, No. 2, 52—61 (1972).
 116. Swart W. W., Cann G., *Investment Planning for the Development of a National Resource — Linear Programming Based Approaches*, *Computers and Operations Res.*, 1, 247—262 (1974).
 117. Taylor G., *The Canadian Travel Industry Management Improvement Program*, Paper presented at the TIMS XX International Meetings Tel-Aviv, Israel, June 26, 1973.
 118. Thompson M., *On Any Given Sunday: Fair Competitor Orderings with Maximum Likelihood Methods*, *J. Amer. Stat. Assoc.*, 70, No. 351, 536—541 (1975).
 119. Tiedman C. E., Minstein D., *Travel Models*, Chapter 4 in Volume 1, *Methods and Models*, Michigan Outdoor Recreation Demand Study, State Resource Planning Program, Michigan Department of Commerce, Lansing, Michigan, 1966.
 120. Ullman E. L., Volk D. J., *An Operational Model for Predicting Reservoir Attendance and Benefits*, *Papers of the Michigan Academy of Science, Arts and Letters* 47, 473—484 (1962).
 121. Van der Smissen B., *Effects of Recreation on Individuals and Society*, a

- chapter in *Recreational Research*, National Education Association, Washington, D. C., 1966.
122. Van der Smitten B., Joyce D., (Ed.), *Bibliography of Theses and Dissertations in Recreation Parks, Camping and Outdoor Education*, National Recreation and Park Association, Arlington, Virginia, 1970.
 123. Van Doren C. S., *An Interaction Travel Model for Projecting Attendance of Campers at Michigan State Parks: A Study in Recreational Geography*, Unpublished Ph. D. Dissertation, Department of Geography, Michigan State University, 1967.
 124. Van Doren C. S., Lentnek B., *Activity Specialization Among Ohio's Recreation Boaters*. *J. Leisure Res.*, 1, 296—315 (Autumn 1969).
 125. Var T., Gearing C. E., Swart W. W. *O/R and MS in Travel and Tourism*. A paper presented at the ORSA/TIMS Joint National Meeting, San Juan, Puerto Rico, October 16—18, 1974.
 126. Var T., *Outdoor Recreation and Travel Research: Present and Future*, Presented at the Travel Research Workshop at the Annual Travel Research Association Conference, San Diego, California, September 10, 1975.
 127. Wagner H., *Principles of Operations Research*. Prentice Hall, Englewood Cliffs, N. J., 1969.
 128. Webster's New Twentieth Century Dictionary, 2d Edition, The World Publishing Company, Cleveland, Ohio, 1966.
 129. Wennergren E. C., Nielsen D. B., *A Probabilistic Approach to Estimating Demand for Outdoor Recreation*, Bulletin 478, Utah State University, Logan, Utah, December 1968.
 130. Wilkinson P. F., *The Use Models in Predicting the Consumption of Outdoor Recreation*, *J. Leisure Res.*, 5, No. 3 (Summer 1973).
 131. Williams A., Zelisky W., *On Some Patterns in International Tourist Flows*, *Econ. Geography*, 46, No. 4, 549—567 (1970).
 132. Wolfe R. I., *Communications*, *J. Leisure Res.*, 2, 85—87 (Winter 1970).
 133. Wolfe R. I., *Inertia Model*, *J. Leisure Res.*, 4, 73—76 (Winter 1974).
 134. Wolfe R. I., *Perspective on Outdoor Recreation — A Bibliographical Survey*, *Geographical Rev.*, LIV, No. 2, 203—235 (April 1964).
 135. Wolfe R. I., *Vacation Homes and the Gravity Model*, *Eristics*, 174, 352—363 (1970); *Communication*, *J. Leisure Res.*, 2, No. 1, 85—87 (1970).

Предметный указатель

- Аварийная замена 348
Агрегированное планирование 71—116
Агрегированные показатели 71
Анализ боевых операций 517—519
— военно-политических учений 529, 530
— военных небоевых операций 514—517
— корреляции 43
— надежности сложных систем 390
— процесса формирования учебных групп 457—460
— распределения возможных видов транспортных средств 475—480
— распределения ресурсов в угольной промышленности 603—605
— системы образования 439—454
— структурных моделей процесса обучения 452—454
— типа «стоимость — эффективность» 523—525
Аналитические модели см. Модели аналитические
Аппаратные средства 174
- Бейсовские прогнозы 19, 20, 28—30
- Вектор минимального пути 319
— пути 318
— сечения 319
— состояния 316, 318
- Вероятность безотказной работы устройства 314
— исправности или пригодности устройства к эксплуатации 314
— неисправности устройства 314
- Весовой множитель 7, 198
Весовые множители одинаковые 17
— — оптимальные 17, 18
— — сглаживающие 18
- Виртуальная память 181
Время безотказной работы 324—332
— задержки 15, 16
— упреждения 6, 9, 145
Вычислительные системы 174, 176—182
- Дежурная бригада 368—370
Декомпозиция модульная 323
Доверительные границы оценок надежности 315—323
- Задача оптимального назначения 48
— Ферма 195
Задачи управления 73
Заказ фиксированный 165
Заказы объединенные 166
— пополняющие 166
— специальные 232
— условные 163
Замена по параболке 350—354
— при удобном случае 354
— профилактическая 334
Запас избыточный 171, 172
— общий располагаемый 171, 172
Запас остаточный 156, 166
— резервный 75, 152
Запасы, классификация 123—127
— систематизация 123—127
Зона реакции 366, 368, 374, 375
- Имитационное моделирование 378, 379, 519—521
Индикатор неисправности 314
Индикаторная функция 315
Интеграция производственного процесса 13
Интенсивность реализации продукции 132, 133
— управления 145

- Информационные системы 174, 187, 188
- Календарное планирование 232
 — — производства электроэнер-
 гии без учета потерь 548—552
 — — — с учетом потерь 553—
 555
 — — ремонтных работ 562—565
- Классификация задач календарного
 планирования 252
 — запасов 123—127
- Комплексное проектирование учеб-
 ных заведений 463, 464
- Комплектование штата 34—45
- Конечный продукт 113
- Контроль за загрязнением атмосфе-
 ры 387—390
 — качества медицинского обслу-
 живания 412
- Координация работ гидро- и электро-
 станций 555—5
- Коэффициент дисконтирования 58
- Кривая жизненного цикла 8
- Кривая распределения 147—149
 — — затрат 147, 148
- Критерий распределения ресурсов
 155, 156
- Медицинское обслуживание 410—
 423
 — — амбулаторное 414, 415
 — — стационарное 414, 415
- Мера расстояния 193
- Минимальное сечение 319
- Минимальный путь 319
- Метод адаптивного сглаживания
 25—27
 — Бокса — Джеккинса 27, 28
 — «вариационной матрицы» 53
 — «все или ничего» 477, 478
 — гиперболической аппроксима-
 ции 202
 — Дельфи 42
 — множественной регрессии 468
 — наименьшей себестоимости еди-
 ницы продукции 142, 143
 — оценки списка 56, 57
 — следящих сигналов 22—25
 — фильтрации 7
 — Фратера 469
 — частичного сбалансирования пе-
 риода планирования 142, 143
- Методы анализа организации оборо-
 ны 527—530
 — прогнозирования 7
- Модели аналитические 518, 519
 — вероятностные 16
 — имитационные 53, 54, 182
 — линейные ряда одновременно
 выполняемых операций 60—62
 — минимизации времени пробега
 372
 — многоканальной системы мас-
 сового обслуживания 371, 372
 — одноканальной системы мас-
 сового обслуживания 372
- Модели планирования производства
 электроэнергии 578—583
 — последовательного назначения
 исполнителей 51—53
 — прогнозирования 11—16
 — регрессионные 446, 447
 — сезонные 13—15
 — стохастические процессов обу-
 чения 454
 — технологических операций
 635—640
 — транспортного типа 48—50
 — целевого программирования 50,
 51
- Модель времени пробега
 — для определения оптимального
 размера партии продукции 84—
 97
 — конфликтующих возможностей
 471—474
 — марковская перемещения пер-
 сонала 40
 — нелинейная аналитическая 98—
 103
 — оптимального вознаграждения
 60—63
 — оптимального управления с об-
 ратной связью 106—108
 — планирования производства
 с переменным фондом рабочей
 силы 77—80
 — — — постоянным фондом
 рабочей силы 74—77
 — предпочтения 472—474
 — принятия решений 649—652
 — притяжения 472—474
 — прогнозирования потоков
 транспорта 644—647
 — равных возможностей 472—474
 — распределения потоков транс-
 порта 644
 — — ресурсов в системе здраво-
 охранения 420—422

Модель времени трудовых ресурсов 460, 461

— с вогнутой функцией затрат 103—106

— — квадратичной функцией затрат 81

— — линейной функцией затрат 73—81

— — произвольной функцией затрат 97—113

Модель типа максимума/минимума 158, 159

— точки заказа и размера партии 157, 158

Надежность связанных систем 320, 324

— систем 322, 323

— теория см. Теория надежности 314—343

— энергосистем 560, 561

Неотложная медицинская помощь 417, 418

Несущественный элемент системы 317

Оперативная бригада 367, 370, 371

Оперативное управление 215

Оптимизация проектных решений методом геометрического программирования 630—635

— технологических процессов 622—630

Организация тренировок 660, 661

Основная теорема декомпозиции 317

Остаток 11

Охрана водных ресурсов 385—387

— окружающей среды 379—390

Оценка стратегии в индивидуальных видах спорта 658—660

— — — командных видах спорта 656—658

Ошибки прогноза 20—25

Параметры модели 74

Переменная интенсивности 5

— состояния 5

Переменные входные 40

— выходные 40

— модели 74

Период «замораживания» 163

Планирование 75

— использования установленных агрегатов 558—560

— потребностей в медицинском персонале 418, 419

— в перевозках людей и грузов 466

— производственных процессов 548—566

— развития сети электропередачи 583—591

— расширения электрической сети 583—591

— резерва генерирующих мощностей 568—576

— тактическое 114, 115

— стратегическое 113, 144

— увеличения производства электроэнергии 567—583

Показатели процесса обучения 443—446

— основные 8

— сводные 71

— уровня обслуживания 145, 146

Полезность 57

— реализуемая 44

— условная 44

Подход «научного управления» 54, 55

Приготовление гранулированных смесей 605—607

— смесей для получения сплавов металлов 607, 708

Принцип декомпозиции 57

— обратной связи 28, 31, 32

— «оптимизация для потребителя» 476—480

— «оптимизация для системы» 476—480

Проблема «запасных частей» 334

Прогнозирование 5

— научно-технического прогресса 8

— нагрузки электроэнергетических систем 539—548

Процесс восстановления 331

Размещение баз экстренных служб 376

— объектов на плоскости с евклидовой метрикой 200

— — — метрикой Минковского 196

Разработка макетов транспортной сети 487

— методов исследования операций 503—505

- Распределение Вейбулла 327
— водных ресурсов 385—387
— времени безотказной работы 325
γ-распределение 327, 328
— оперативных бригад 376
— Парето 125
— трудовых ресурсов 47—54
— усеченное нормальное 325—329
- Санитарные службы 380—393
Сбор и удаление твердых отходов 381—384
Связанные системы 317—320
— случайные величины 321
Сглаживание ошибок 23
Себестоимость единицы продукции 134
Сезонное сечение 13—15
Системы обработки данных 182—187
Службы уличного движения 390—393
Способы поставки заказа продукции 131, 132
Срок реализации продукции 132, 133
Стратегические игры 521, 522
Структура задачи управления за-
срами 126
Структура производственных реше-
ний 110
Структурная функция 316—319
Структурные свойства системы 315
- Теорема Блекуэлла 332
— восстановления узловая 322
— — элементарная 332
Теория надежности 314
Технологическая операция 600
Точка заказа 71
- Управление водоснабжением города
— запасами 121
— трудовыми ресурсами 34—70
— системой медицинского обслу-
живания 419
Упорядочение работ 232
- Факторный анализ 460
Формирование команд 660, 661
Функция восстановления 331
— целевая 38
— реакции организации 37, 38, 56
- Характеристики аппаратных средств 183
— программных средств 183
- Экстремальная точка 8
Экстренные службы 365—373

Оглавление

Часть I

ПРИМЕНЕНИЕ МЕТОДОВ ИССЛЕДОВАНИЯ ОПЕРАЦИЙ К ОРГАНИЗАЦИОННО-ЭКОНОМИЧЕСКИМ СИСТЕМАМ

I-1. Прогнозирование, Р. Браун	5
Введение	5
1. Виды временных рядов	5
2. Исходные данные	9
3. Модели	11
4. Система сглаживающих весовых множителей.	16
5. Ошибки прогноза	20
6. Методы, применяемые для уточнения прогноза.	25
7. Изучение сферы сбыта	30
Литература	32
I-2. Управление трудовыми ресурсами, Р. Мезон, Э. Фламгольц	34
Введение	34
1. Комплектование штата организации	34
2. Повышение квалификации персонала	45
3. Распределение трудовых ресурсов	47
4. Использование трудовых ресурсов	54
5. Оценка эффективности работы персонала	57
6. Оплата работы и поощрение персонала	58
7. Модели сложных систем управления трудовыми ресурсами	63
Заключение	65
Литература	66
I-3. Агрегированное планирование производства, А. Хэкс.	71
Введение	71
1. Математические модели агрегированного планирования производства	73
2. Модели с квадратичной функцией затрат	81
3. Модели для определения оптимального размера партии про- дукции (модели с постоянными затратами)	84
4. Модели с произвольной функцией затрат	97
5. Интеграция производственного процесса	113
Литература	116
I-4. Управление запасами, Р. Браун.	121
Введение	121
1. Структура задачи управления запасами.	121

2. Правила принятия решений относительно выбора размера партии продукции для пополнения ее запасов	130
3. Правила принятия решений относительно уровня резервного запаса	144
4. Распределение ограниченных ресурсов	155
5. Архитектура систем управления запасами	157
6. Реализация систем управления запасами	169
Литература	172
4-5. Вычислительные и информационные системы, А. Мертен.	174
Введение	174
1. Вычислительные системы.	176
2. Системы обработки данных.	182
3. Информационные системы.	187
Заключение	189
Литература	189
4-6. Планировка и размещение объектов, Р. Френсис, Дж. Уайт.	192
Введение	192
1. Задачи размещения с непрерывным пространством решений	195
2. Задачи размещения — распределения с использованием непрерывного пространства решений	202
3. Задачи планировки с использованием непрерывного пространства решений	204
4. Задачи планировки с использованием дискретного пространства решений	209
5. Задачи размещения с использованием дискретного пространства решений	213
6. Задачи размещения — распределения с использованием дискретного пространства решений	215
Заключение	226
Литература	227
4-7. Календарное планирование и упорядочение работ, М. Сальвадор	232
Введение	232
1. Классификация задач календарного планирования	232
2. Некоторые общие критерии оценки графика работ	236
3. Детерминированные задачи календарного планирования	237
4. Стохастические задачи календарного планирования	251
5. Перспективы развития теории и методов календарного планирования и упорядочения работ	258
Литература	259
4-8. Выбор проекта. Планирование работ над проектом и руководство проектом, У. Саудер	264
Введение	264
1. Задачи выбора проекта	266
2. Модели выбора проекта	269
3. Сетевое планирование	281
4. Руководство проектом	299
5. Интегрированные системы планирования и руководства	305
Заключение	308
Литература	309

I-9. Теория надежности, Э. Шойер.	314
Общие положения	314
1. Структурные свойства систем	315
2. Надежность связанных систем	320
3. Время безотказной работы и соответствующие случайные величины	324
4. Некоторые оценки надежности	332
5. Интервальная надежность и эксплуатационная готовность	340
Литература	343
I-10. Техническое обслуживание оборудования, Э. Жардин.	344
Введение	344
1. Замена (обновление) оборудования	344
2. Контроль за состоянием оборудования	355
3. Профилактический текущий ремонт и восстановление оборудования	357
4. Организационная структура службы технического обслуживания	359
5. Перспективы развития технического обслуживания оборудования	361
Литература	362

Часть II

ПРИМЕНЕНИЕ МЕТОДОВ ИССЛЕДОВАНИЯ ОПЕРАЦИЙ К СОЦИАЛЬНЫМ И ПРОИЗВОДСТВЕННЫМ СИСТЕМАМ

II-1. Системы городской сферы обслуживания, И. Савае, Г. Грэмме	364
Введение	364
1. Экстренные службы	365
2. Охрана окружающей среды	379
3. Санитарные службы. Службы уличного движения	390
4. Другие внутригородские (местные) сферы обслуживания.	393
5. Обратная связь в системе управления городской сферой обслуживания	395
6. Проблемы применения методов исследования операций в городской сфере обслуживания	396
7. Новые направления в использовании методов исследования операций для решения проблем городской сферы обслуживания	401
Литература	404
II-2. Система здравоохранения, Р. Фиттер.	410
Введение	410
1. Краткая характеристика и проблемы системы здравоохранения	410
2. Анализ проблем системы здравоохранения с помощью методов исследования операций	414
3. Планирование национальных и региональных систем медицинского обслуживания	418
4. Перспективность использования методов исследования операций для решения проблем системы здравоохранения	420
5. Модель распределения ресурсов в системе здравоохранения	420

6. Пути дальнейшего усовершенствования системы медицин- ского обслуживания	423
Литература	423
II-3. Процессы обучения, А. Муд.	439
Введение	439
1. Анализ системы образования	439
Литература	454
2. Аннотированная библиография	455
II-4. Транспорт, Б. Нийл, Э. Кэтенис, Д. Хайндс.	466
Введение	466
1. Прогнозирование потребностей в перевозках людей и грузов	468
2. Анализ распределения возможных видов транспортных средств	476
3. Определение параметров задачи и описание различных ва- риантов решения	480
4. Оценка эффективности возможных вариантов решения	480
5. Разработка макетов транспортной сети и оценка возможных вариантов решения	487
6. Реализация выбранного варианта решения	488
7. Классический пример практического использования общей схемы исследования транспортной системы	489
8. Типичные проблемы, возникающие при исследовании общей схемы анализа транспортной системы	493
Заключение	496
Литература	497
II-5. Военные системы, Э. Квейд.	499
Введение	499
1. Исторические аспекты использования методов исследования операций в военных системах	499
2. Анализ военных небоевых операций	514
3. Анализ боевых операций	517
4. Различные типы анализа состояния вооружения	523
5. Методы анализа организации обороны	527
6. Публикации, посвященные использованию методов исследова- ния операций в военной области	530
Литература	532
II-6. Электроэнергетические системы, Л. Гаргер	537
Введение	537
1. Прогнозирование нагрузки	538
2. Планирование производственных процессов	548
3. Планирование увеличения производства электроэнергии	567
4. Планирование развития сети электропередачи	583
Литература	591
II-7. Производственные и технологические процессы, Г. Блау, Р. Климпел	600
Введение	600
1. Задачи распределения ресурсов при непрерывном производ- стве	602

2. Задачи составления смесей	605
3. Линейные модели технологических операций	608
4. Линейные модели ряда одновременно выполняемых техно- логических операций	610
5. Оптимизация технологических процессов	622
6. Оптимизация проектных решений методом геометрического программирования	630
7. Математические модели технологических операций	635
Литература	640

II-8. Организация досуга (туризм, спорт, развлечения), У. Сюот, Ч. Гиэринг. Т. Ва.	642
Введение	642
1. Применение методов исследования операций при организа- ции развлечений, путешествий и туризма	643
2. Применение методов исследования операций в области спорта	654
Заключение	662
Литература	662
Предметный указатель	669

УВАЖАЕМЫЙ ЧИТАТЕЛЬ!

Ваши замечания о содержании книги, ее оформлении, качестве перевода и другие просим присылать по адресу: 129820, Москва, И-110, ГСП, 1-й Рижский пер., д. 2, изд-во «Мир».

Под редакцией
ДЖ. МОУДЕРА И С. ЭЛМАГРАВИ

ИССЛЕДОВАНИЕ ОПЕРАЦИЙ

2

Модели и применения

Научный редактор Т. П. Сапожкова
Младший научный редактор Н. Н. Титова
Художник В. А. Медников
Художественный редактор Л. Е. Безрученков
Технический редактор Е. С. Потапенкова
Корректор В. И. Постнова

ИБ № 2221

Сдано в набор 25.12.80.
Подписано к печати 25.03.81.
Формат 60×90¹/₁₆.
Бумага типографская № 1.
Гарнитура обыкновенная. Печать высокая.
Объем 21,25 бум. л. Усл. печ. л. 42,50. Усл. кр.-отт. 45,86.
Уч.-изд. л. 45,86. Изд. № 20/0698.
Тираж 20 000 экз. Зак. 01182. Цена 3 р. 60 к.

ИЗДАТЕЛЬСТВО «МИР»
Москва, 1-й Рижский пер., 2.

Ордена Трудового Красного Знамени
Московская типография № 7 «Искра революции»
Союзполиграфпрома Государственного Комитета СССР
по делам издательств, полиграфии и книжной торговли.
Москва 103001, Трехпрудный пер., 9.

***Издательство «Мир»
выпустило в свет в 1979 году
книгу***

Дэниел К. Применение статистики в промышленном эксперименте. Пер. с англ. — М.: Мир, 18,69 л., 1 р. 70 к.

Книга посвящена новому направлению прикладной статистики, широко используемому в экспериментальных исследованиях. Подробно рассмотрена техника факторного планирования эксперимента. Особое внимание уделено вопросам интерпретации результатов эксперимента и проверки правильности исходных предпосылок, приемам планирования в промышленных условиях, в том числе блочному планированию и планированию в условиях дрейфа. Изложение сопровождается детальным анализом большого числа практических примеров.

Рассчитана на инженеров, занимающихся прикладными исследованиями, и представляет несомненный интерес для аспирантов и студентов вузов различного профиля.