
ТЕОРИЯ И МЕТОДЫ
СИСТЕМНОГО АНАЛИЗА

Б.Г. ЗАСЛАВСКИЙ
Р.А. ПОЛУЭКТОВ

УПРАВЛЕНИЕ
ЭКОЛОГИЧЕСКИМИ
СИСТЕМАМИ



Москва "Наука"
Главная редакция
физико-математической литературы
1988

ББК 28.071
З-36
УДК 519.876.2

Управление экологическими системами./Заславский Б.Г., Полуэктов Р.А. — М.: Наука. Гл.ред.физ.-мат.лит., 1988. — 296 с. — (Теория и методы системного анализа). ISBN 5-02-013890-8

Модели экологических систем обладают рядом специфических особенностей — высокой размерностью, знакоопределенностью фазовых переменных, однонаправленностью управляющих воздействий.

Дается качественный анализ динамики экологических процессов. Рассматриваются вопросы управляемости при ограничении на знак управления и фазовых переменных. Определяется место имитационного моделирования в теории экологических систем и формулируются его основные принципы. Излагаются методы реализации задач управления экологическими системами на ЭВМ.

Для специалистов по математической экологии, имитационному моделированию и теории управления сложными системами.

Табл. 16. Ил. 30. Библиогр. 237 назв.

Рецензент

доктор физико-математических наук *В.В. Калашников*

Борис Григорьевич Заславский
Ратмир Александрович Полуэктов

УПРАВЛЕНИЕ ЭКОЛОГИЧЕСКИМИ СИСТЕМАМИ

Серия "Теория и методы системного анализа", № 23

Редактор *А.С. Левина*

Художественный редактор *Т.Н. Кольченко*

Технический редактор *С.В. Геворкян*

Корректоры *Г.В. Подвольская, Т.В. Обод, Т.А. Печко*

Набор осуществлен в издательстве
на наборно-печатающих автоматах

ИБ № 12902

Сдано в набор 25.01.88. Подписано к печати 05.05.88. Т—12501

Формат 60 × 90/16. Бумага писчая

Гарнитура Пресс-Роман. Печать офсетная

Усл.печ.л. 18,50 Усл.кр.-отг. 18,50 Уч.-издл. 22,79

Тираж 2800 экз. Тип. зак. 45 . Цена 4 р. 90 к.

Ордена Трудового Красного Знамени

издательство "Наука"

Главная редакция физико-математической литературы

117071 Москва В-71, Ленинский проспект, 15

Четвертая типография издательства "Наука"

630077 г. Новосибирск-77, ул. Станиславского, 25

3 $\frac{150200000-131}{053(02)-88}$ 162-88

ISBN 5-02-013890-8

© Издательство "Наука".
Главная редакция
физико-математической
литературы, 1988

ОГЛАВЛЕНИЕ

Предисловие	5
Глава 1	
Введение	7
§ 1.1. Экосистемы. Основные понятия	7
§ 1.2. Динамические модели искусственно культивируемых экосистем и биотехнология	11
1.2.1. Модели систем непрерывного культивирования (11). 1.2.2. Модели биохимической кинетики (16).	
§ 1.3. Динамические модели природных экосистем	17
1.3.1. ELM-имитационная модель природного биома (18). 1.3.2. Имитационная динамическая модель "лес – вредитель" (19).	
§ 1.4. Задачи, возникающие при исследовании динамики экосистем	28
Глава 2	
Общие динамические свойства моделей эксплуатируемых экологических систем	32
§ 2.1. Равновесие и устойчивость	32
2.1.1. Общий вид моделей динамики численности эксплуатируемых экосистем (33). 2.1.2. Существование и основные свойства стационарных решений (35). 2.1.3. Устойчивость стационарных решений (40). 2.1.4. Асимптотическая ограниченность решений (46). 2.1.5. Существование устойчивых стационарных и ограниченных режимов в системах непрерывного культивирования (49). 2.1.6. Многофакторное лимитирование (57). 2.1.7. Анализ возрастной структуры непрерывно культивируемой популяции микроорганизмов (65).	
§ 2.2. Автоколебания	73
2.2.1. Бифуркация рождения периодических решений (73). 2.2.2. Устойчивость периодических решений (80). 2.2.3. Существование колебаний в нелокальном случае (89).	
§ 2.3. Хаос	96
2.3.1. Модели динамики численности популяций с неперекрывающимися поколениями (97). 2.3.2. Общие свойства хаотических режимов (100). 2.3.3. Условия рождения хаотических структур в моделях динамики численности экосистем (106).	

Глава 3	
Задачи управления в экологии	111
§ 3.1. Управление нелIMITИрованными экологическими системами.	111
3.1.1. Управление вблизи равновесного состояния (114). 3.1.2. Управление в положительном конусе (131).	
§ 3.2. Управление лимитированными экологическими системами	138
3.2.1. Управление вблизи равновесного состояния (138). 3.2.2. Стабилизированность равновесного состояния (141).	
§ 3.3. Оптимальное управление экологическими системами.	148
§ 3.4. Системы управления клеточными популяциями. Прямое управление	156
§ 3.5. Системы управления клеточными популяциями. Непрямое управление.	163
§ 3.6. Наблюдение переменных состояния экологических систем	171
3.6.1. Положительная наблюдаемость (171). 3.6.2. Динамическое наблюдение (175).	
Глава 4	
Имитационное моделирование экосистем	179
§ 4.1. Сущность метода	179
§ 4.2. Структура имитационных моделей экосистем	183
4.2.1. Модель агрофитоценоза пшеницы (184).	
§ 4.3. Имитационные системы	196
4.3.1. Иерархическая структура имитационной системы (197).	
4.3.2. Банк моделей (199). 4.3.3. Технология имитационного моделирования (202).	
§ 4.4. Планирование имитационного эксперимента	206
4.4.1. Понятие и цели планирования эксперимента (207). 4.4.2. Планирование дисперсионных экспериментов (212). 4.4.3. Планирование регрессионных экспериментов (229).	
Глава 5	
Математическое и программное обеспечение задач моделирования и управления.	236
§ 5.1. Общая структура программного обеспечения имитационной системы	237
5.1.1. Структура и состав программного комплекса (237).	
5.1.2. Прикладное математическое и программное обеспечение (240).	
5.1.3. Общесистемное внутреннее программное обеспечение (244).	
5.1.4. Автоматизация машинного эксперимента (255).	
§ 5.2. Внешнее программное обеспечение имитационной системы	259
5.2.1. Управление базой данных (260). 5.2.2. Управление экосистемами (263). 5.2.3. Интерфейс пользователя (268).	
Приложение 1. Вспомогательные леммы	272
Приложение 2. Генераторы псевдослучайных чисел и последовательностей	274
Приложение 3. Факторный эксперимент 2^k и его анализ	279
Приложение 4. Критерии сравнения планов регрессионного эксперимента.	284
Список литературы	286
Список основных обозначений	296

ПРЕДИСЛОВИЕ

В последней трети нынешнего столетия в различных областях человеческих знаний появились и продолжают появляться новые и все более сложные проблемы. Возникающие при этом трудности принципиально отличаются от тех, которые многократно приходилось преодолевать ранее. Ясно, например, что проблема автоколебаний нелинейных систем в свое время представлялась достаточно сложной большинству исследователей — и математикам и инженерам. Однако создание качественной теории дифференциальных уравнений, а затем и многих приближенных методов (в частности, метода гармонического баланса) привело к тому, что решение многих, казавшихся ранее неразрешимыми, задач стало почти тривиальным.

Трудности нынешнего периода носят иной характер. Суть дела здесь сводится к тому, что для решения выдвигаемых практикой задач приходится иметь дело с объектами, поведение которых управляется огромным числом взаимосвязанных величин. Эффекты внешнего воздействия на такие объекты часто оказываются непредсказуемыми или, как говорят, эти объекты обладают контринтуитивным поведением. Сложившаяся ситуация хорошо обрисована в книге Дж. Гига "Прикладная общая теория систем" (М.: Мир, 1981, кн. 1), где говорится: "Будучи не в силах справиться со сложными и трудными проблемами, мы пытаемся заменить их более легкими. После такого упрощения решения могут потерять свою реальную основу. Так появляются трудности, с одной стороны, из-за невозможности решить сложные проблемы, а с другой — из-за непригодности решений, полученных при использовании упрощенных моделей".

Это замечание целиком и полностью относится к проблеме изучения экологических систем для решения задач прогноза и управления. Нет никаких сомнений в том, что экосистемы следует относить к этому классу объектов: для них не существует готовых рецептов и методов исследования. Более того, можно утверждать, что опыт развития "классической науки" здесь зачастую оказывается неэффективным, а при решении конкретных задач во многих случаях приходится скорее опираться на житейскую мудрость и здравый смысл, чем на традиционную аксиоматику математических построений. Точнее говоря, при исследовании сложных систем "аксиоматический" и "житейский" подходы должны дополнять друг друга. Будучи оптимистами, авторы считают, что из любого безвыходного положения

всегда найдутся по крайней мере два выхода. Именно поиск этих двух выходов и определил содержание этой книги.

Мы исходили из того, что хорошая математика является украшением любой монографии. Кроме того, в экологических задачах уже используются методы управления, не имеющие аналогов в технике, однако соответствующая новым задачам теория недостаточно развита. Поэтому в первой части книги (гл. 2 и 3) предпринята попытка сформулировать специфические особенности экосистем как объектов управления. Здесь же излагаются новые результаты, касающиеся математической структуры возникающих при этом проблем.

Другая сторона вопроса касается решения задач, связанных с управлением реально существующими природными экосистемами. Единственным выходом здесь является построение машинных (имитационных) моделей. При разработке таких моделей приходится считаться с тем фактом, что современные ЭВМ являются вычислителями дискретного действия, изначально не предназначенными для решения задач моделирования. Поэтому в процессе создания имитационных моделей приходится разрабатывать для них специальное математическое и программное обеспечение. Описанию структуры имитационных систем, предназначенных для решения задач управления, посвящены гл. 4 и 5.

Книга состоит из пяти глав и приложений, из которых Р.А. Полуэктовым написаны первый и третий параграфы гл. 1, а также гл. 4, 5 и приложения 2, 3 и 4. Второй параграф первой главы, гл. 2, 3 и приложение 1 написаны Б.Г. Заславским. Четвертый параграф гл. 1 написан авторами совместно.

Авторы отдают себе отчет в том, что появление настоящей монографии стало возможным благодаря тому, что в лаборатории математического моделирования Агрофизического института в течение ряда лет проводилось целенаправленное исследование по созданию "большой" имитационной модели продуктивности агроэкосистем. Только в процессе создания и последующего использования такой модели можно было понять некоторые тонкости процесса имитации, а постоянно возникавшие трудности стимулировали появление новых решений. В разработке указанной модели наряду с авторами участвовали И.А. Вол, Л.А. Заславская, Г.М. Коваль, Ю.А. Пых, А.И. Смысловский и С.М. Финтушал. Поэтому перечисленных лиц следовало бы считать неофициальными соавторами данной книги. В то же время формулировка всех высказанных ниже положений принадлежит Б.Г. Заславскому и Р.А. Полуэктову, за что они согласны нести полную ответственность.

Авторы выражают глубокую признательность рецензенту книги В.В. Калашникову, внимательно прочитавшему рукопись и сделавшему ряд ценных замечаний, способствовавших ее улучшению. Авторы благодарят также Э.И. Смоляра и Ф.М. Ланду, ознакомившихся с рукописью книги и высказавших ряд замечаний редакционного характера.

§ 1.1. Экосистемы. Основные понятия

Одной из характерных черт интенсификации общественного производства в последние десятилетия является значительное расширение масштабов перестройки человеком существующих на Земле природных комплексов. Разрушение складывавшихся веками естественных связей и создание новых позволяют растущему населению нашей планеты обеспечивать себя необходимым для гармоничного развития количеством ресурсов. Но этот сложный и многогранный процесс взаимодействия Человека и Природы имеет и обратную сторону. Не зная детальной картины взаимодействия многочисленных живых и неживых компонентов природного комплекса и не умея точно оценить то новое состояние, в котором он окажется после нашего вмешательства, мы подчас получаем в конечном счете эффект, существенно отличающийся от ожидаемого. Эта вполне реальная опасность и выдвинула на первый план задачу разработки теории динамического поведения как биосферы в целом, так и биогеоценозов, т.е. тех элементарных природных комплексов, из которых состоит биосфера Земли.

Изучающая эти вопросы наука — экология — оформилась в самостоятельную отрасль человеческих знаний сравнительно недавно, хотя ее отдельные разделы формировались в рамках общих представлений об окружающей нас природе с древних времен. Термин "экология"¹⁾ обозначает совокупность знаний и теоретических представлений о взаимодействии растений и животных как между собой, так и с абиотическим (неживым) окружением. Совершенно очевидно, что роль экологии в нашей жизни и в теоретическом и в прикладном плане будет только возрастать.

Поясним термин "экосистема". Биологические особи (или индивидуумы), будь то многоклеточные растения и животные или микроорганизмы, сразу же после своего появления на свет включаются в сложную цепь взаимодействия с окружающими их другими организмами и средой. Более того, они сами воздействуют на среду своего обитания, меняя ее в том или ином направлении.

Экология изучает все эти взаимодействия в совокупности. Другими словами, она изучает вопросы о том, как сообщества живых организмов, используя органические, минеральные и энергетические ресурсы, создают (продуцируют) новое органическое вещество, как это вещество распреде-

¹⁾ В переводе с греческого *οικος* — дом, жилище, *λογος* — понятие, учение.

ляется между другими организмами в системе и каким образом оно в конце концов вновь распадается на свои минеральные компоненты. При этом экологи стараются понять, за счет каких механизмов создается и поддерживается удивительная стабильность природных комплексов, что обеспечивает сохранение и изменение их пространственно-временной структуры, как складываются в них пищевые цепи и круговорот питательных веществ (биогеохимические циклы), какими факторами определяются их развитие и эволюция. Таким образом, экология изучает надорганизменный уровень функционирования биологических систем и "совокупность или структуру связей между организмами и их средой" (см. [30]).

Экологию традиционно принято считать одним из разделов науки о жизни — биологии [30]. В то же время еще классические представления В.И. Вернадского об экосистеме свидетельствуют о том, что проблемы экологии тесно связаны с рассмотрением всей совокупности взаимодействующих между собой геофизических, геохимических и биологических процессов в природе.

В экосистеме живые организмы и их неживое (абиотическое) окружение связаны между собой и находятся в постоянном взаимодействии. Биологическую часть системы составляет биоценоз или биологическое сообщество, т.е. совокупность живых организмов, обитающих на одной территории (акватории), связанных между собой трофическими и иными связями и участвующих в общих процессах самовоспроизведения. В частности, различают "фитоценоз" (сообщество растений), "зооценоз" (сообщество животных) и "микробиоценоз" (сообщество микроорганизмов). Термин "биоценоз" употребляется также для обозначения искусственных совокупностей живых организмов (например, одноклеточных), разводимых человеком с целью производства биомассы или продуктов жизнедеятельности этих организмов в рамках биотехнологий. Сообщество, состоящее из особей одного биологического вида, называют популяцией.

Понятие "биогеоценоз" — почти синоним термина "экосистема". По В.Н. Сукачеву, "биогеоценоз можно определить как участок земной поверхности, где на известном протяжении биоценоз (фитоценоз, зооценоз и микробиоценоз) и отвечающие ему части атмосферы, литосферы, гидросферы и педосферы остаются однородными, тесно связанными между собой также однородными взаимодействиями и поэтому в совокупности образующими единый внутренне взаимообусловленный комплекс" [42]. Таким образом, под биогеоценозом понимается естественный природный комплекс, через который не проходит ни одна существенная биоценотическая, геоморфологическая, гидрологическая, микроклиматическая и почвенногеохимическая граница [42, 43].

Любой биогеоценоз может рассматриваться как экосистема. В то же время термин "экосистема" является более общим. Он используется также и применительно к искусственным системам, созданным человеком (например, к системам биологической очистки вод, загрязненных органическими отходами, к системам промышленного культивирования микробных сообществ и т.п.).

Рассмотрим структуру экосистем более подробно. На рис. 1.1. приведена упрощенная схема взаимодействия различных элементов в природной экосистеме (биогеоценозе), и на рис. 1.2 — в простейшей системе искус-

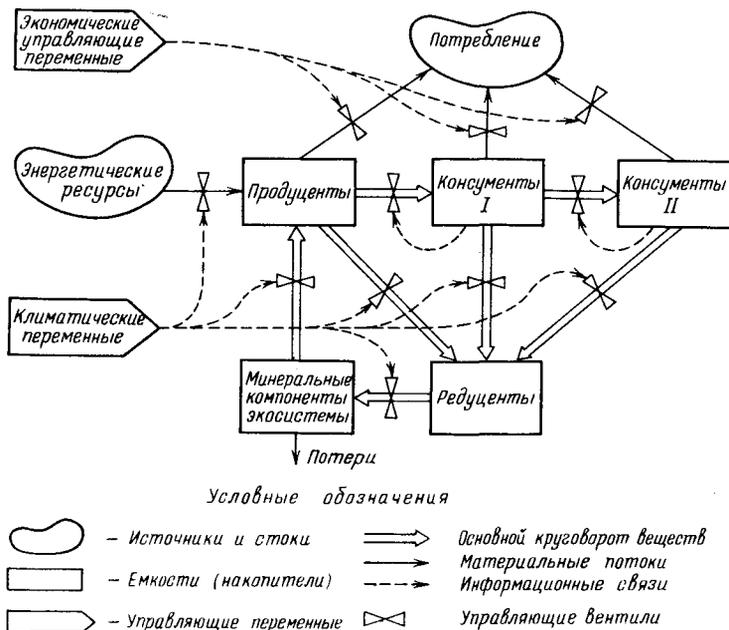


Рис. 1.1. Структурная схема модели травяного биома

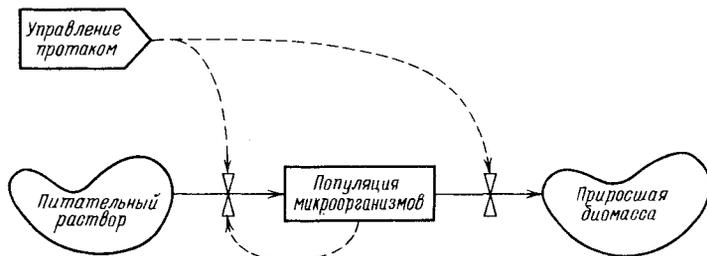


Рис. 1.2. Структурная схема модели системы искусственного культивирования микроорганизмов

ственного культивирования микроорганизмов¹⁾). С термодинамической точки зрения любая экосистема является открытой: для своего функционирования она нуждается в постоянном снабжении энергией и веществом. Более точно, если можно представить себе экосистему с замкнутым и сбалансированным циклом круговорота вещества (например, самая большая из известных нам экосистем – биосфера Земли), то приток, а следовательно, и рассеяние энергии необходимы для существования любой экосистемы. На рис. 1.1 источники энергии и вещества представлены в виде совокупности материальных и энергетических ресурсов экосистемы. Для

¹⁾ На рис. 1.1 и 1.2 использованы обозначения системной динамики, широко распространенные в последнее время в мировой литературе; эти обозначения будут использоваться далее без специальных пояснений.

естественных биогеоценозов эти ресурсы находятся под воздействием климатических управляющих факторов и часто — под воздействием человека.

Биологическая часть экосистемы с точки зрения трофических (пищевых) отношений делится на два компонента: 1) автотрофный компонент, к которому относятся фото- и хемосинтезирующие организмы (продуценты), и 2) гетеротрофный, питающийся за счет органических веществ, созданных на первом уровне. В наземных экосистемах основную часть автотрофного уровня составляют зеленые растения, а в водных экосистемах — одноклеточные (фитопланктон) и многоклеточные водоросли.

На автотрофном уровне происходит фиксация световой энергии и превращение простых неорганических веществ в сложные органические соединения. К гетеротрофной части биоценоза относятся макроконсументы (или просто консументы) — хищники различного уровня, питающиеся либо за счет автотрофов, либо за счет консументов более низкого уровня, и микроконсументы (или редуценты) — гетеротрофные организмы, питающиеся отмершими остатками продуцентов и консументов. Редуценты в процессе своей жизнедеятельности разлагают сложные органические вещества (сахара, белки, целлюлозу, лигнин и др.) и высвобождают неорганические питательные вещества, которые могут быть вновь использованы продуцентами. Тем самым замыкается круговорот веществ в биосфере.

Природные экосистемы в значительной степени находятся под воздействием климатических и погодных факторов. Наряду с человеческой деятельностью именно тип климата, а также гидрологические и ландшафтные особенности определяют видовой состав и структуру биогеоценоза [3]. Основными климатическими факторами являются суммарная солнечная радиация, температурный режим, который, помимо непосредственной связи с приходящей к земной поверхности радиацией, определяется морскими течениями и крупномасштабными перемещениями воздушных масс, а также водный режим, связанный для наземных экосистем с количеством осадков и их распределением в течение года.

Под влиянием климатических и антропогенных воздействий сложились разные растительные формации (фитоценозы), в числе которых различают леса, степи, сельскохозяйственные угодья, пустыни, растительность океанов и пресноводных водоемов. В совокупности с населяющими их животными (консументами), грибами и бактериями (редуцентами) они составляют мозаику разнообразных биологических группировок — биомов, различающихся по структуре и видовому составу, по характеру и скоростям процессов создания и утилизации органического вещества.

Сделаем в заключение настоящего параграфа следующее замечание. Построение математических моделей любых объектов связано с формализацией их описания и с выделением существенных черт рассматриваемой эмпирической ситуации. Эта формализация, естественно, предъявляет определенные требования к строгости используемой терминологии и в первую очередь к основным определениям. В то же время в современной экологии сложилась такая ситуация, что отдельные термины, являющиеся кратким выражением тех или иных концепций, часто совпадают лишь по своему словесному выражению, а на самом же деле за ними скрываются различные и не вполне определенные понятия. Принимая во внимание

это обстоятельство, авторы стремились к однозначному использованию экологической терминологии, сопровождая по возможности появление каждого термина соответствующим разъяснением. Ниже приводятся наиболее важные и часто используемые термины, необходимые для понимания содержания книги в целом.

Популяция – группа особей одного вида, населяющих определенное пространство и обладающих способностью обмена генетической информацией в процессе размножения и передачи ее потомкам.

Биологическое сообщество – любая совокупность популяций, населяющих определенную территорию (акваторию).

Экосистема – функциональная система, включающая в себя сообщество живых организмов и их среду обитания. Таким образом, она сочетает биом (*фито-, зоо- и микроценозы*) и совокупность факторов среды, т.е. *экотон*. При этом каждый из компонентов экосистемы влияет на особенности друг друга и каждый необходим для поддержания слаженного хода жизни.

§ 1.2. Динамические модели искусственно культивируемых экосистем и биотехнология

Биотехнология – это методы получения необходимых человеку продуктов из живых клеток разного происхождения [2]. В современных биотехнологических процессах основными продуцентами являются микроорганизмы. Инженерное обеспечение их жизнедеятельности представляет собой весьма сложную задачу. В промышленных условиях эта задача решается с помощью автоматических систем управления. Внедрение вычислительной техники сделало эффективным использование математических моделей для управления биотехнологическими процессами.

В научной сфере микроорганизмы представляют собой прекрасные объекты, используемые для понимания многих явлений живой природы. На них изучаются процессы, протекающие внутри живой клетки, процессы внутрипопуляционных и межпопуляционных взаимодействий, процессы эволюции и т.п. Таким образом, построение и исследование моделей динамики численности микроорганизмов имеют большое научное и практическое значение.

1.2.1. Модели систем непрерывного культивирования. В настоящее время используются два основных способа организации биохимического процесса – периодическое и непрерывное культивирование. При периодическом способе производства исходный материал и микроорганизмы помещаются в замкнутый объем (ферментер), где совершаются биохимические превращения. После завершения процесса ферментер опорожняют, продукт очищают, и цикл повторяется.

При непрерывном производстве подача сырья и выход готового продукта происходит безостановочно (см. рис. 1.2). Это достигается за счет того, что в резервуар, содержащий микроорганизмы, постоянно поступает раствор свежих питательных веществ, а вытекающий поток уносит приросшую биомассу. Благодаря этому размер размножающейся популяции остается неизменным [23, 31, 44]. Непрерывный процесс может продолжаться неограниченно долго, поэтому участвующие в нем микроорганизмы имеют возможность пройти все фазы своего развития от рождения до оставления потомства.

Для описания динамики процесса биосинтеза при непрерывном культивировании обычно используется модель Моно [23, 31, 32]

$$\dot{x} = \mu(s)x - vx,$$

$$\dot{s} = -g\mu(s)x - vs + vs_p,$$

где x — концентрация биомассы, s_p и s — концентрация питательного субстрата на входе в реактор и внутри него, $\mu(s)$ — удельная скорость размножения, v и g — положительные константы. Эта модель проста, наглядна и имеет ясный биологический смысл. К ее недостаткам следует отнести введение упрощающих предположений о том, что все клетки культивируемой популяции (молодые и старые) в равной степени способны к размножению и одинаково чувствительны к питательному субстрату. Для того чтобы избавиться от этих недостатков, необходимо учесть в модели внутреннюю структуру популяции.

В настоящее время хорошо известно, что в течение своей жизни клетки проходят последовательные фазы развития. Жизненный цикл клеток животной и растительной ткани, а также некоторых микроорганизмов в соответствии со специфической чувствительностью их к различным физико-химическим воздействиям в простейшем случае разбивают на четыре последовательные фазы развития. Это — предсинтетическая фаза G_1 , период синтеза ДНК — фаза S , постсинтетическая фаза G_2 и фаза митоза M , во время которой клетки делятся на две дочерние. Вновь образованные клетки находятся сначала в фазе G_1 , а затем повторяют тот же цикл. Иногда вводят в рассмотрение фазу G_0 , предшествующую G_1 , в которой, как считают в настоящее время, клетки обладают способностью к дифференцированию [6, 15].

При исследовании одноклеточной фотосинтезирующей водоросли хлореллы было установлено [9, 59], что клетки этой водоросли в течение своей жизни проходят семь последовательных стадий. Стадии развития хлореллы образуют так называемый цикл Тамия, состоящий из следующей совокупности фаз: D_n , D_a , L_D , L_1 , L_2 , L_3 , L_4 . Клетки стадий D_n и D_a (dark) — темные клетки — наиболее фотосинтетически активны и могут развиваться только на свету, L_D — переходная стадия от темных к светлым (light) клеткам, $L_1 - L_4$ — стадия светлых клеток, причем в периоды L_3 и L_4

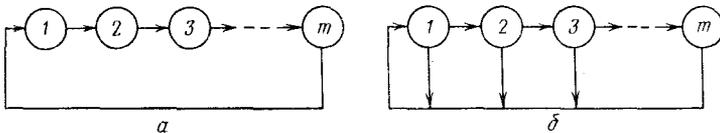


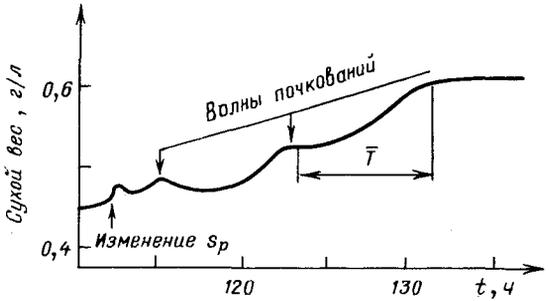
Рис. 1.3. Схема разбиения жизненного цикла клетки на отдельные периоды

клеткам не требуется освещение и они не способны к фотосинтезу. После того как цикл развития завершен, клетка делится на 2, 4, 8 или 16 дочерних, каждая из которых находится на первой фазе D_n (рис. 1.3, а).

Дрожжевые клетки размножаются почкованием [44]. После отделения почки на материнской клетке остается шрам, и поэтому число шрамов на дрожжевой клетке соответствует числу произведенных ею потомков [70, 71]. Возраст клетки (номер фазы развития) определяется числом шрамов на ней (рис. 1.3, б).

Таким образом, мы видим, что в жизненном цикле живой клетки можно выделить конечное число фаз развития. Эти фазы развития различаются как морфологически, так и по специфической чувствительности клеток к факторам окружающей среды. Модель Моно не учитывает этого фактора и поэтому имеет ограниченную область применения. Рассмотрим в качестве примера случай ступенчатого увеличения скорости разбавления или концентрации подаваемого в ферментер субстрата [60]. Наблюдаемый в эксперименте волнообразный переходный процесс можно объяснить частичной

Рис. 1.4. Изменение сухого веса непрерывно культивируемой популяции *S. Cerevisiae* при ступенчатом изменении концентрации глюкозы на входе в реактор



возрастной синхронизацией популяции в результате возникновения волны деления [60]. На рис. 1.4 представлено изменение сухого веса культуры *S. Cerevisiae* при ступенчатом изменении концентрации глюкозы s_p на входе реактора с 0,46 г/литр · час до 2,21 г/литр · час при скорости разбавления $v = 0,1 \text{ час}^{-1}$. Стрелками помечены первая и вторая волны почкований. Если разбить культивируемую популяцию хотя бы на дочерние и материнские клетки, то этот эффект уже можно объяснить.

Влияние лимитирующего субстрата на скорость перехода дочерних клеток в репродуктивную фазу обычно задают формулой Михаэлиса —

Ментен $p_1(s) = \mu_1 \frac{s}{s + k_1}$, а влияние лимитирующего субстрата на скорость

деления материнских клеток, в результате которого образуются дочерние почки, аналогичной формулой $p_2(s) = \mu_2 \frac{s}{s + k_2}$.

Предположим, что чувствительность дочерних клеток к лимитирующему субстрату больше, чем чувствительность материнских, т.е. величина s в реакторе такова, что функция $p_1(s)$ находится почти на линейном участке изменения, в то время как $p_2(s)$ близко к насыщению. Резкое увеличение s ведет в первый момент к тому, что большинство клеток относительно синхронно переходит в репродуктивную фазу, после которой следует волна делений. Действительно, при изменении s скорость роста дочерних клеток

меняется мало: $\dot{x}_1 = \mu_2 \frac{s}{s + k_2} x_2 - v x_2$, а скорость изменения числа мате-

ринских клеток возрастает при увеличении s : $\dot{x}_2 = \mu_1 \frac{s}{s + k_1} - \mu_2 \frac{s}{s + k_2} - v x_2$

(см. ниже формулы (1.2.1) и (1.2.3)). Клетки первой фазы переходят во вторую по истечении времени $T_1 = p_1^{-1}(s) = \frac{s + k_1}{\mu_1 s}$, а клетки второй фазы

дают потомство по истечении времени $T_2 = p_2^{-1}(s) = \frac{s + k_2}{\mu_2 s}$. Отсюда понятно, почему среднее время между волнами деления равняется $T_1 + T_2 = p_1^{-1}(s) + p_2^{-1}(s)$.

Перейдем к построению модели динамики численности непрерывно культивируемой популяции клеток, имеющих конечное число фаз развития. Обозначим через m число фаз развития этих клеток и численный состав популяции зададим вектором $X = (x_1, \dots, x_m)^*$, где x_i — численность клеток i -й фазы развития. Обозначим через s концентрацию лимитирующего субстрата в ферментере, а через s_p — концентрацию этого субстрата на входе в ферментер. Пусть $v > 0$ — скорость разбавления. В условиях идеального перемешивания скорость удаления приросших клеток равна vX . Пусть величина $f_{ij}(s)$ задает скорость, с которой одна клетка j -й фазы

производит клетки i -й фазы. Тогда $f_{ij}(s) = \frac{q_{ij}}{T_i}$, где T_i — средняя для клеток данной популяции продолжительность i -й фазы, а q_{ij} — среднее число потомков j -го типа, оставляемых клеткой при завершении j -й фазы развития. Как правило, число q_{ij} генетически наследуемо и не зависит от условий культивирования (например, тканевые клетки делятся на две дочерние в последней фазе развития, т.е. $q_{1m} = 2$). Концентрация питательных веществ определяет лишь продолжительность фаз развития клеток, т.е. $T_i = T_i(s)$ ($i = 1, \dots, m$). В связи с этим естественно считать, что $f_{ij}(s) = a_{ij} p_j(s)$, где $a_{ij} \geq 0$ при $i \neq j$ и $p_j(s) \geq 0$ ($i, j = 1, \dots, m$). Для упрощения записи введем в рассмотрение квадратные матрицы с неотрицательными недиагональными элементами $F(s) = \{f_{ij}(s)\}_{i,j=1}^m$ и $A = \{a_{ij}\}_{i,j=1}^m$ и диагональную матрицу $P(s) = \text{diag} \{p_i(s)\}_{i=1}^m$, где $p_i(s) = 1/T_i(s)$ — скорость прохождения клеткой i -й фазы развития. Очевидно, что $F(s) = AP(s)$.

Процесс утилизации питательных веществ клетками имеет ферментативную природу, и поэтому следует считать, что одна клетка i -й фазы развития потребляет лимитирующий субстрат со скоростью $g_i p_i(s)$, где g_i — эмпирический коэффициент утилизации ($i = 1, \dots, m$). Таким образом, динамика биосинтетического процесса задается системой [18, 19, 35]

$$\begin{aligned} \dot{X} &= AP(s)X - vX, \\ \dot{s} &= -G^*P(s)X - vs + vs_p, \end{aligned} \quad (1.2.1)$$

где $G^* = (g_1, \dots, g_m)$.

Если клетки в течение своего жизненного цикла проходят m фаз развития и в последней фазе делятся на q дочерних, то матрица A имеет вид [17, 19]

$$A = \begin{pmatrix} -1 & 0 & \dots & 0 & q \\ 1 & -1 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & -1 & 0 \\ 0 & 0 & \dots & 1 & -1 \end{pmatrix}. \quad (1.2.2)$$

Поскольку величина q обозначает среднее число потомков для клеток данной популяции, она задается не обязательно целым числом. Например, в популяции хлореллы, где клетки могут давать 2, 4, 8 и т.д. потомков, величина q зависит от соотношения клеток разной плодовитости.

Для популяции почкующихся клеток матрица A принимает вид [17, 19]

$$A = \begin{pmatrix} 0 & 1 & 1 & \dots & 1 & 1 \\ 1 & -1 & 0 & \dots & 0 & 0 \\ 0 & 1 & -1 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 1 & -1 \end{pmatrix}. \quad (1.2.3)$$

Скорость прохождения клеткой i -й фазы не убывает при увеличении концентрации питательных веществ в среде обитания. Поэтому будем предполагать, что $\partial p_i(s)/\partial s \geq 0$ [20, 27, 28]. Если продолжительность хотя бы одной фазы развития клеток зависит от концентрации лимитирующего субстрата, то имеет место неравенство

$$\sum_{i=1}^m \partial p_i(s)/\partial s > 0.$$

При отсутствии субстрата в реакторе скорость его потребления, очевидно, равняется нулю. Поэтому $\sum_{i=1}^m g_i p_i(s) = 0$ при $s = 0$ и $\sum_{i=1}^m g_i p_i(s) > 0$ при $s > 0$ ($p_i(s) > 0$ при $s > 0$). Кроме того, концентрация субстрата в ферментере не превосходит концентрации субстрата, подаваемого в него. В силу этого будем считать в дальнейшем, что $s_p \geq s \geq 0$ и $x_i \geq 0$ ($i = 1, \dots, m$). Легко убедиться в том, что данное множество является положительно инвариантным.

При многоступенчатом культивировании ряд ферментеров последовательно соединяется в батареи [31]. Многопоточный процесс характеризуется тем, что в каждый ферментер наряду с биомассой и субстратом, поступающими из предыдущей ступени, независимо подается также питательная среда (рис. 1.5).

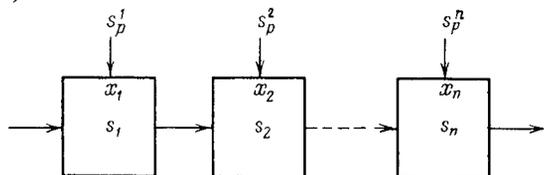


Рис. 1.5. Схема многоступенчатого многопоточного процесса непрерывного культивирования

Поэтому динамика процесса биосинтеза при таком культивировании описывается системой

$$\begin{aligned} \dot{X}^1 &= AP(s^1)X^1 - v_1 X^1, \\ \dot{s}^1 &= -G^*P(s^1)X^1 - v_1 s^1 + v_1 s_p^1, \\ \dot{X}^i &= AP(s^i)X^i - v_i X^i + u_i X^{i-1}, \\ \dot{s}^i &= -G^*P(s^i)X^i - v_i s^i + u_i s^{i-1} + (v_i - u_i) s_p^i \quad (i = 2, \dots, n), \end{aligned} \quad (1.2.4)$$

где индекс i ($1 \leq i \leq n$) обозначает номер ферментера, u_i — скорость потока из $(i-1)$ -го в i -й ферментер, v_i — скорость разбавления в i -м ферментере ($v_i \geq u_i$, $v_i - u_i$ — скорость подачи свежего питательного субстрата).

1.2.2. Модели биохимической кинетики. Функционирование живой клетки, по сути дела, сводится к изменению концентраций различных химических соединений. В результате возникают не только пространственные, но и временные структуры, наиболее важными из которых являются колебательные режимы. Изучение этих режимов проводилось как на экспериментальном, так и на теоретическом уровне. Теоретические исследования сводятся к построению и анализу систем нелинейных дифференциальных уравнений, переменные которых положительны во все моменты времени, так как задают концентрации веществ.

Рассмотрим одну модель внутриклеточного синтеза. Считается общепризнанным, что ДНК является основным генетическим компонентом клетки, передающим в закодированной форме информацию о структуре биохимических реакций. Процесс транскрипции состоит в считывании этой информации и передаче ее на другой носитель — информационную РНК. С информационной РНК осуществляется трансляция этой информации в различные белковые компоненты клетки. На рис. 1.6 представлена упрощенная схема последовательности биохимических реакций, в которой конечный продукт ингибирует начальный участок цепи метаболизма. Модель этого процесса имеет вид [11, 12, 16, 29, 36, 41, 45, 47, 51, 52, 58, 61, 62]

$$\begin{aligned} \dot{x}_1 &= -a_1 x_1 + f(x_m), \\ \dot{x}_i &= k_i x_{i-1} - a_i x_i \quad (i = 2, \dots, m), \end{aligned} \quad (1.2.5)$$

где x_i — концентрация i -го метаболита, a_1, a_i и k_i ($i = 2, \dots, m$) — положительные константы и $f(x_m)$ — монотонно убывающая функция, описывающая процесс подавления транскрипции информационной РНК. Данная модель интенсивно изучалась, для нее получены условия рождения периодических решений [24, 25, 41, 48, 50, 52, 58, 61, 62].

Следует отметить, что к подобным уравнениям приводит модель пролиферации (т.е. роста) и дифференцировки (т.е. спецификации структуры и функции) клеточной ткани [72]. К аналогичному виду приводятся

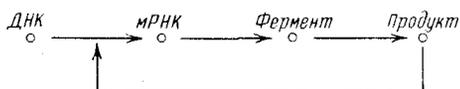


Рис. 1.6. Схема последовательности биохимических реакций, при которой конечный продукт ингибирует начальный участок цепи метаболизма

модели динамики биохимических и структурных превращений, учитывающие механизм распределенного запаздывания при наличии эрланговского ядра. Действительно, если

$$\dot{x} = ax + f\left(\int_{-\infty}^t x(\tau) G(t-\tau) d\tau\right),$$

где $G(t) = \frac{\lambda(\lambda t)^{m-1} e^{-\lambda t}}{m!}$, то введение переменных $x = x_1, x_i$ ($i = 2, \dots, m$)

приводит нас к эквивалентной схеме решения системы обыкновенных дифференциальных уравнений [67]

$$\begin{aligned}\dot{x}_1 &= ax_1 + f(x_m), \\ \dot{x}_i &= \lambda(x_{i-1} - x_i).\end{aligned}$$

Формальное представление нелинейности в системе (1.2.5) в виде $f(x_m) = p(x_m)x_m$ приводит эту систему к виду (2.1.2), который рассматривается в следующей главе, где $s = x_m$ и $\varphi(x) = C^*X$, $C = (0, \dots, 0, 1)$. Перепишем систему (1.2.5) в векторной форме

$$\dot{X} = A(s)X, \quad s = C^*X,$$

где

$$X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix}, \quad A(s) = \begin{bmatrix} -a_1 & 0 & \dots & p(s) \\ & k_2 - a_2 & & \\ & & \ddots & \\ & & & 0 & k_m & -a_m \end{bmatrix}.$$

Отсюда видно, что при каждом фиксированном значении s матрица $A(s)$ имеет неотрицательные недиагональные элементы. Ингибирование или снижение скорости протекания процесса на начальном участке метаболической цепи означает, что выполнено условие

$$\frac{\partial}{\partial s} p(s) = \frac{\partial}{\partial s} \left(\frac{f(s)}{s} \right) < 0.$$

Это неравенство имеет место по крайней мере в окрестности стационарной точки и соответствует требованию, предъявляемому к модели общего вида (см. (2.1.6), (2.2.4)).

Аналогичные выводы относятся к моделям биохимических процессов, протекающих в условиях изъятия и поступления извне:

$$\begin{aligned}\dot{x}_1 &= -a_1x_1 + f(x_m) - vx_1 + b_1, \\ \dot{x}_i &= k_ix_i - a_ix_i - vx_i + b_i \quad (i = 2, 3, \dots, m).\end{aligned}$$

В заключение отметим, что вопросы устойчивости и неустойчивости равновесных состояний, а также вопросы рождения периодических решений в нелинейных моделях биологических систем в настоящее время интенсивно разрабатываются [24, 25, 41, 48, 52]. В этом отношении результаты § 2.2 дополняют или обобщают известные результаты [24, 56, 50]. Результаты исследования устойчивости периодических решений, приведенные ниже в п. 2.2.2, являются совершенно новыми.

§ 1.3. Динамические модели природных экосистем

Описанные в предыдущем параграфе модели появляются в задачах исследования динамики искусственно культивируемых популяций микроорганизмов и их сообществ, что составляет основу ряда биотехнологий. Природные экосистемы составляют другую обширную ветвь, для которой также возникают задачи, связанные с их эффективной эксплуатацией и

(или) сохранением, что приводит к необходимости использования метода моделирования.

Далее будут рассмотрены два примера. Поскольку обе модели подробно описаны в литературе, мы приведем лишь краткое их описание. Нас будет в первую очередь интересовать структура возникающих здесь проблем, которые следует рассмотреть в дальнейшем. "Расшифровка" этих проблем как раз и составляет основное содержание книги.

1.3.1. ELM-имитационная модель природного биома. Естественные фитоценозы земного шара подразделяются на классы, основными из которых являются дождевые тропические леса, светлые и саванновые леса умеренного климата, вечнозеленые хвойные леса, степи и саванны, луга, жаркие и холодные (полярные) пустыни. Рассматриваемые вместе с группировками обитающих в них животных, грибов и бактерий, они образуют биомы – весьма сложные биологические совокупности. Одной из наиболее развитых моделей биома является разработанная в рамках Международной биологической программы модель ELM¹⁾ [68, 69], блок-схема которой представлена на рис. 1.1.

Это общая блок-схема для моделей луговых биомов, типичных для ряда западных штатов США. Целью построения модели явилось исследование внутрисезонной динамики как лугового биома в целом, так и его частей – растительного покрова, животных и почвы с учетом влияния климата и антропогенных воздействий. Для характеристики поведения системы выделены константы и переменные модели. Последние подразделяются на внешние и внутренние переменные. Внешние управляющие переменные оказывают влияние на характеристики внутри системы. К важнейшим внешним управляющим переменным относятся осадки, солнечная радиация, температура и скорость ветра над растительным покровом. Далее рассматривается набор внутренних переменных состояния (уровней). Эти переменные меняются от шага к шагу в ответ на изменение внешних управляющих воздействий или других внутренних переменных. Примерами таких переменных являются содержание влаги в почвенных слоях, зеленая масса, численность животных. Темпы (скорости процессов), т.е. перетоки из одной "емкости" в другую, могут находиться под воздействием либо физических, либо физиологических процессов. Так, например, для переменной состояния "наземная биомасса растений" скорость входного потока определяется интенсивностью фотосинтеза и перемещения воды и минеральных элементов из корней. Скорость оттока из этой "емкости" связана с поеданием травы животными, отмиранием, оттоком метаболитов в корни и вытаптыванием травяного покрова животными. На каждую переменную темпа воздействует несколько переменных уравнений и управляющих воздействий. Скорость фотосинтеза, например, зависит от листового индекса доступной почвенной влаги, солнечной радиации и температуры воздуха. Человек также рассматривается как управляющая переменная или внешняя сила, воздействующая на потоки в систему или из системы.

Модель в целом содержит следующие подмодели:

1) Подмодель абиотических процессов, имитирующую динамику тепло- и влагопереносов в системе.

¹⁾ ELM – Ecosystem Level Model, модель экосистемного уровня.

2) Подмодель продуцентов, описывающую круговорот углерода, фенологию растительного покрова как наземной, так и подземной частей экосистемы.

3) Подмодель редуцентов, описывающую разложение отмерших органических частей растений и животных.

4) Подмодель консументов (млекопитающих и насекомых), в которой имитируются процессы организменного уровня, внутри- и межпопуляционные взаимодействия.

5) Подмодели азота и фосфора, имитирующие динамику перемещения и трансформаций соединений этих элементов в почве и процессы минерального питания растений.

Каждый блок модели взаимодействует с остальными, в целом образуя общую имитационную модель экосистемы. Модель позволяет оценивать влияние на динамику протекающих в экосистеме процессов неконтролируемых погодных условий, а также обеспечивает возможность анализа различных способов воздействия на экосистему.

1.3.2. Имитационная динамическая модель "лес – вредитель". В этом пункте мы переходим к описанию модели, описывающей динамику развития листогрызущих вредителей леса при учете их взаимодействия с развитием и формированием листвы деревьев. Целью моделирования является выработка мероприятий по охране лесов от вредителей.

Лесные листогрызущие насекомые уничтожают листву, что приводит к потере прироста древесины, нарушению водного, температурного и светового режимов, ослаблению древостоя, а при многократных сплошных обеданиях листвы – к усыханию и гибели деревьев [8, 13, 39, 53].

Это делает задачу описания динамики вредителей исключительно актуальной. Мы проиллюстрируем принципы построения и использования модели на одном частном, но важном примере – описании динамики зеленой дубовой листовертки. При этом будем следовать в основном работам [37 – 40].

Зеленая дубовая листовертка является одним из основных вредителей дубрав. При этом в многолетней динамике относительно "спокойные" периоды, когда плотность популяции насекомых невелика, сменяются периодами массового размножения. Механизмы этих колебаний очень сложны, а при их анализе нельзя абсолютизировать действие одного какого-либо фактора. Согласно существующим экологическим теориям [5] все действующие на популяцию факторы делятся на две группы: 1) факторы абиотической среды, которые относятся к группе модифицирующих воздействий; 2) биотические факторы, играющие роль регуляторов численности насекомых и действующие по принципу "отрицательной обратной связи". Последнюю группу образуют воздействия со стороны энтомофагов, хищников, болезней, поддерживающих популяцию в состоянии депрессии. Кроме того, к биотической группе относятся также внутривидовые взаимодействия, приводящие к снижению плодовитости в ответ на возрастание численности (недостаток пищи, миграции, физическое уничтожение особей своего вида, диапаузы, выделение репеллентов и др.). Включение в действие различных факторов в зависимости от плотности популяции схематически изображено на рис. 1.7. Отметим, что влияние совокупности

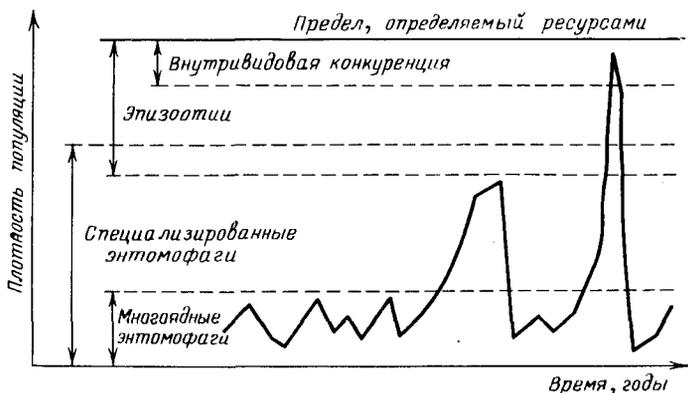


Рис. 1.7. Зависимость лимитирующих факторов от плотности популяции насекомых

регулирующих и модифицирующих воздействий несовершенно и не может обеспечить постоянства плотности популяции.

Перейдем к рассмотрению конкретного вредителя — дубовой листовертки и к описанию имитационной модели. Зеленая дубовая листовертка (*Tortrix viridana* L.) — листогрызущее насекомое. Его отличительной чертой является хорошая приспособленность к неблагоприятным условиям среды и затяжной характер всплеск численности (5 — 10 лет и более). Листовертка имеет годичный цикл развития, в течение которого она проходит следующие фазы: фазу яйца, фазу личинки (гусеницы), фазу куколки и фазу бабочки (имаго). Продолжительность активных фаз (гусеница, куколка, бабочка) составляет полтора-два месяца в году. Питание дубовой листовертки происходит практически только в стадии личинки и продолжается около месяца. За это время личинка проходит пять возрастов развития, переход в которые сопровождается линькой.

Рассмотрим кратко важнейшие факторы, определяющие выживаемость популяции в каждой фазе (рис. 1.8). Ведущим фактором, определяющим продолжительность фаз личинки и куколки, а в течение фазы личинки — сроки перехода гусениц из младшего возраста в более старший, является температурный режим. Сумма эффективных температур воздуха, необходимая для развития, например, в стадии личинки, колеблется незначительно в разных климатических зонах, а заметная разница в среднесуточных температурах воздуха в этих зонах приводит к различным периодам пребывания листовертки в данной фазе. Так, например, если в Воронежской области длительность фазы гусеницы находится в пределах 18 — 26 дней, то в более прохладном Подмоскowie она составляет уже в среднем 23 — 35 дней.

Одним из важных моментов, определяющих гибель гусениц первого возраста, которая иногда решающим образом сказывается на судьбе всей популяции листовертки, является фактор совпадения между временем раскрытия почек дуба и временем отрождения из яиц гусениц [1, 39, 40]. Основные причины несинхронности этих процессов две: 1) различные температурные требования дуба к началу раскрытия почек и листовертки для отрождения личинок; 2) зеленая дубовая листовертка как биологи-

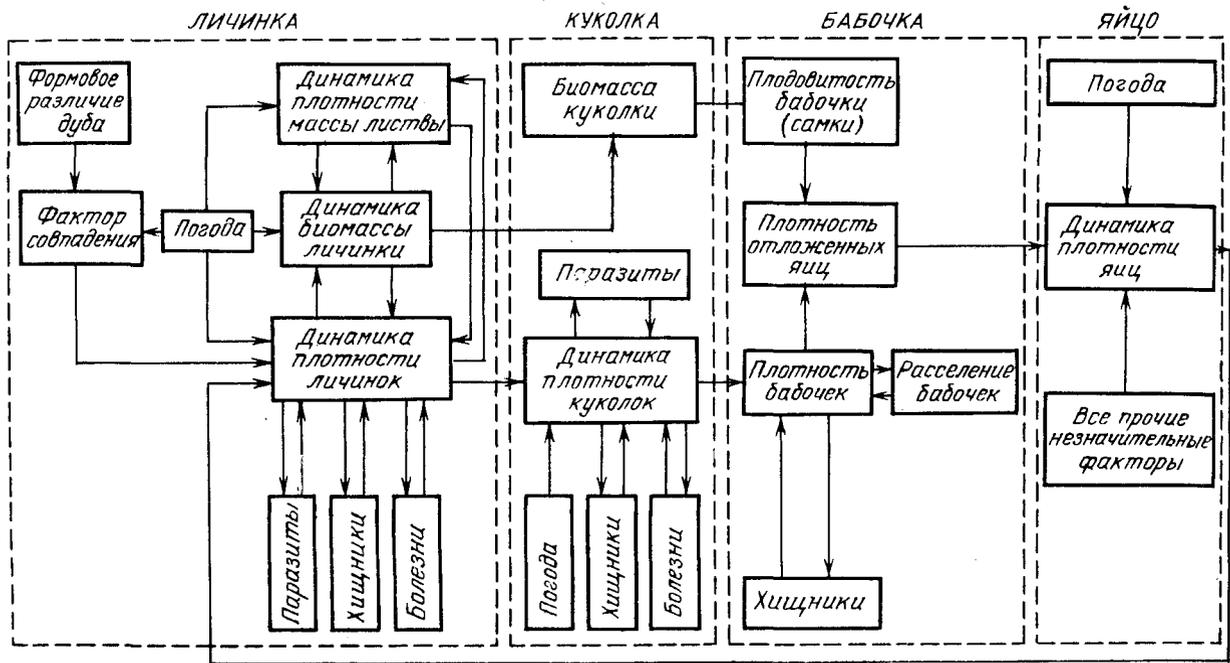


Рис. 1.8. Факторы, определяющие выживаемость дубовой листовертки

ческий вид хорошо приспособлена к ранней форме дуба. Однако в природе часто наблюдается гетерогенность дубового древостоя в отношении его фенологических форм, что и приводит к указанному явлению.

Важность фактора совпадения объясняется повышенной требовательностью маленьких гусениц к качеству пищи и весьма ограниченным временем возможной их жизни без питания. Задержка в раскрытии почек или их вымерзание приводит к миграциям личинок в поисках пищи и, следовательно, к значительным тратам энергии и голодной смерти, что подтверждается экспериментальными материалами многих авторов [1, 14, 21, 63–66]. В более старших возрастах основная гибель личинок вызывается паразитами, болезнями, хищниками, а в случае недостатка корма (большой перенаселенности) — различными биотическими факторами, связанными с конкуренцией за пищу.

Личинки пятого возраста переходят в стадию куколки. При окукливании личинок происходит потеря ими части веса. Величина уменьшения веса зависит от откормленности особей к моменту окукливания. Имеется существенная разница в весе куколок самцов и самок. Снижение плотности популяции в этой фазе развития обусловлено воздействиями со стороны паразитов, болезней и хищников.

Гибель взрослых особей (бабочек) обычно не бывает большой. Бабочка живет всего несколько дней. За это время самки откладывают на ветках деревьев яйца. В определенных пределах плодовитость бабочки-самки пропорциональна массе куколки, из которой эта бабочка вылетела, и, следовательно, определяется предшествующей интенсивностью питания личинки. При очень высокой плотности популяции листовертки в условиях недостатка корма средний вес куколок становится маленьким и плодовитость популяции резко снижается. Это объясняется неплодовитостью части самок и неполовозрелостью самцов, что уменьшает вероятность оплодотворения самки при спаривании.

В стадии яйца зеленая дубовая листовертка находится в среднем 10,5 месяца в году. Отложенные на ветках дуба яйца зимуют. На сохранность яиц листовертки во время их зимовки решающее влияние оказывают опасные понижения температуры воздуха и их продолжительность [1, 13]. Гибель яиц от всех других воздействий незначительна. Весной из яиц отрождаются маленькие личинки и начинается новый цикл развития листовертки.

Таковы в общих чертах основные причины, вызывающие изменение численности дубовой листовертки в течение одной генерации. Как мы упоминали, рост численности популяции насекомых сдерживается паразитами, болезнями, хищниками. Энтомофаги, количество которых возрастает при обилии пищи, т.е. в периоды массовых размножений насекомых-вредителей, становятся часто действенной силой в подавлении вспышки. Однако, в отличие от других листогрызущих насекомых, зеленая дубовая листовертка оказывается очень стойкой к воздействию энтомофагов. Список энтомофагов дубовой листовертки весьма длинен (более 60 паразитов), но не было зарегистрировано ни одного случая полной гибели от них популяции листовертки.

Эпизоотии, вызванные бактериальными, вирусными или грибковыми заболеваниями, часто являются причиной гибели вредителей и прекращения их массового размножения. Очень часты различные бактериальные

и вирусные заболевания у гусениц листовертки, особенно в годы неблагоприятных погодных условий. Но и они обычно не приводят даже к такому снижению плотности вредителя, которое бы повлекло затухание вспышки. Все это говорит об исключительной приспособленности листовертки к непрерывно изменяющимся условиям и факторам среды. Именно поэтому с ней трудно бороться. По этим причинам возникает необходимость искусственных эффективных и рациональных воздействий на популяцию этого насекомого. Имитационная модель должна в той или иной мере отражать все перечисленные выше особенности цикла развития дубовой листовертки.

Перейдем к описанию модели. Отметим, что популяция вредителя неоднородно размещена в ареале своего обитания. При построении модели необходимо учесть эту неоднородность. Неоднородность расселения популяции в горизонтальной плоскости учитывается тем, что вся территория леса разбивается на участки, в пределах которых пространственные колебания плотности незначительны. Поскольку миграции особей из одного участка леса в другой незначительны, в первом приближении их в модели можно рассматривать независимо. Вертикальная неоднородность плотности яйцекладок и личинок должна быть непосредственно учтена в модели, поскольку, например, личинки по мере объедания листвы опускаются с верхних ветвей на более низкие. При этом темп продвижения гусениц определяется переменными самой модели.

Для выделенного участка леса модель функционирует следующим образом. Основные переменные модели (плотность насекомых, биомасса листвы, плотности хищников, паразитов, возбудителей болезней) изменяются от фазы к фазе в зависимости от изменения всех остальных переменных и складывающейся метеообстановки. Эти изменения описываются системой рекуррентных соотношений. Динамика переменных модели внутри отдельных фаз развития насекомого описывается системой нелинейных дифференциальных уравнений или конечными (алгебраическими) соотношениями. Основной фазой, внутрифазная динамика которой существенно влияет на многолетние колебания численности, является фаза личинки (гусеницы). Поэтому именно эта фаза развития насекомого представлена в виде дифференциальных уравнений. Мы первоначально рассмотрим описание отдельных фаз, а затем перейдем к характеристике модели в целом.

а) Система уравнений, описывающая динамику процессов в стадии личинки

Уравнения, описывающие взаимодействие популяции вредителя с его "кормовой базой", следуя [4, 38, 39, 54, 55, 57, 73], можно записать в виде

$$\dot{X} = F(X, S), \quad (1.3.1)$$

где X – вектор состояния моделируемой системы, S – совокупность воздействий со стороны окружающей среды. Как показал анализ [39], в качестве основных переменных модели в стадии личинки можно принять плотность популяции личинок x_1 , биомассу листвы в кронах деревьев, отнесенную к единице площади леса x_2 , и средний вес одной личинки x_3 . Таким образом, $X = (x_1, x_2, x_3)^*$ (* – знак транспонирования). Конкретизируем вид зависимости $F(X, S)$. Основным показателем среды, влияющим на динамику листовертки, так и на скорость формирования листьев,

является температурный фактор. Известно, что каждому насекомому для его развития необходимо определенное количество энергии, определяемое суммой эффективных температур, которая складывается из суммы среднесуточных температур воздуха, превышающих некоторый порог (нижний порог развития насекомого). С таким же параметром (с точностью до порога) связана и скорость формирования листвы. Чем выше среднесуточные температуры, тем быстрее формируется листва и развиваются личинки. В модели этот факт можно учесть, введя экспериментально определяемую переменную ξ , характеризующую продолжительность развития вредителя в фазе личинки. Как очевидно, величина ξ является косвенной характеристикой погодных условий в этот период.

Другой механизм, который следует учесть в модели, связан с влиянием пищевого лимитирования. Естественный путь учета этого механизма связан с допущением, что интенсивность лимитирования в каждый момент времени зависит от количества пищи, приходящегося на единицу биомассы гусениц, т.е. от величины $y = \frac{x_2}{x_1 x_3}$. При этом существуют некоторые "ком-

фортные" условия, когда $y \geq y_*$ и пищи достаточно. При снижении величины y ниже определенного предела развитие популяции ингибируется, что приводит к дополнительной гибели личинок и снижению скорости накопления биомассы выжившими гусеницами. Наконец, для "замыкания" модели следует воспользоваться уравнением баланса, введя коэффициент утилизации g , численно равный биомассе потребленной личинками листвы в отношении к приросту их собственной биомассы.

Учет всех этих вполне существенных допущений приводит к следующей специализации исходной системы уравнений (1.3.1):

$$\begin{aligned} \dot{x}_1 &= k_1(\xi)\mu(\xi, S_1)\nu(y)x_1, \\ \dot{x}_2 &= k_{21}(\xi)x_2 - k_{22}(\xi)x_2^2 - gk_3(\xi)x_1x_3, \\ \dot{x}_3 &= k_3(\xi)\sigma(y)x_3. \end{aligned} \quad (1.3.2)$$

В этих уравнениях $\nu(y) = \nu\left(\frac{x_2}{x_1 x_3}\right)$ и $\sigma(y) = \sigma\left(\frac{x_2}{x_1 x_3}\right)$ — функции лимитирования, зависимости $k_i(\xi)$ учитывает влияние метеоусловий, а функция $\mu(\xi, S_1)$ — учет влияния паразитов, хищников, болезней. Второе соотношение представляет собой модификацию известного уравнения Ферхюльста — Пирла. Переход к явному виду включенных в модель зависимостей (параметризация модели) выполнен в [39, 40], где приведены также результаты параметрической идентификации.

Отметим, что в природе часто наблюдается расхождение между сроками раскрытия почек дуба и сроками отрождения гусениц листовертки. Это отражается как на динамике плотности вредителя, так и на степени дефолиации. Если гусеницы отродились раньше раскрытия почек, то часть из них неизбежно гибнет от голода и усиленных миграций в поисках пищи. Появление гусениц позже срока раскрытия почек также приводит к их гибели, поскольку к этому времени молодые листья уже огрубевают и становятся непригодными для питания гусениц первого возраста. Интенсивность появления гусениц зависит от погодных условий. При теплой и

сухой погоде их выход происходит интенсивно и в короткие сроки. Холодная и влажная погода, напротив, приводит к затяжке сроков отрождения личинок. Гибель личинок листовертки от несовпадения сроков лежки обычно в пределах 0–50% для ранних форм дуба. Для других форм дуба она может достигать 100%. В модели фактор несовпадения учитывается зависимостью скоростей появления личинок и раскрытия почек дуба от метеорологических условий, главным из которых является сумма эффективных температур.

б) Подмодели фаз куколки, бабочки яйца

Куколки, как и личинки, уничтожаются хищниками, паразитами, подвержены болезням. Степень влияния этих воздействий определяется метеорологическими условиями, плотностью популяции листовертки и фазой вспышки. При перенаселенности часть гусениц окукливается, недокормившись. Эти куколки имеют малый вес и погибают. Динамика отпада куколок выражается соотношением

$$\dot{x}_4 = -k_4(\xi)\varphi(S_1)x_4, \quad (1.3.3)$$

где вид зависимостей $k_4(\xi)$ и $\varphi(S_1)$ аналогичен соответствующим зависимостям для стадии личинки. Вес особи при окукливании снижается вследствие затрат энергии. Продолжительность жизни бабочки составляет 4–6 дней. Гибель бабочек обычно незначительна (5–20%). Основная причина гибели — хищники.

в) Общая характеристика модели

Выше рассмотрены особенности описания отдельных фаз развития популяции листовертки. В заключение этого пункта следует охарактеризовать модель в целом. Как видно из изложенного, годовой цикл динамики популяции описывается сложной системой нелинейных дифференциально-разностных уравнений. При этом состояние популяции в конце очередной фазы развития служит для расчета начальных условий для описания динамики последующей фазы. Отметим, что от фазы к фазе изменяется размерность пространства состояния. В фазе личинки вектор состояния является трехмерным, он включает в качестве компонент биомассу листьев, среднюю биомассу особи и плотность популяции. В фазе куколки этот вектор превращается в двумерный, а в фазе имаго и яйца — в одномерный. Состояние яйцекладки после перезимовки используется для расчета начальных параметров развития личинок в последующем году. Это позволяет рассчитывать многолетнюю динамику популяции листовертки. В модели учитывался также влияние метеорологических условий как на выживаемость особей, так и на их плодовитость. Это влияние учитывается дифференцированно по фазам развития листовертки.

г) Использование модели для прогноза и управления плотности популяции

Как уже указывалось, имитационная модель служит для расчета динамики реальных популяций в конкретных условиях их обитания. Поэтому работа по созданию такой модели не может считаться завершенной, пока не определены численные значения всех входящих в модель коэффициентов. В соответствии с этим можно выделить следующие этапы разработки имитационной модели:

1) Определение цели исследования. Разработка общей структуры модели и структуры отдельных блоков.

2) Описание отдельных блоков на основе анализа теоретических представлений о динамике их функционирования и имеющихся экспериментальных данных.

3) Постановка и проведение специальных экспериментальных лабораторных и полевых исследований по идентификации отдельных блоков модели.

4) Постановка и проведение комплексного полевого эксперимента по идентификации модели в целом. Корректировка модели.

5) Сравнение на независимом материале результатов моделирования с натурными наблюдениями (верификация модели).

6) Предложения по практическому использованию модели.

Выше описано применительно к данной модели содержание первого и второго этапов исследования. В этом пункте мы кратко охарактеризуем остальные этапы.

Идентификация модели выполнена В.В. Рубцовым и Н.Н. Рубцовой на основе экспериментальных данных, собранных специально для этой цели в Теллермановском опытном лесничестве Лаборатории лесоведения АН СССР [39]. Для наблюдения за состоянием популяции было выбрано 40 эталонных деревьев, на которых проводились учеты численности листовертки и сопутствующих вредителей на всех стадиях развития — яйцекладок, гусениц всех возрастов, куколок, бабочек. Учитывалась степень зараженности особей всех стадий энтомофагами, уничтожение их хищниками и т.д. Подсчитывались также побеги (почки) и листья, степень их зараженности в различных частях кроны и другие параметры. Анализ этих данных позволил произвести идентификацию всех параметров модели. Кроме того, сравнение экспериментальных данных с результатами моделирования привело к необходимости корректировки модели. В первом варианте модели не учитывалась динамика распределения популяции листовертки и дефолиации по высоте деревьев. Оказалось, что такое упрощение справедливо при низкой плотности популяции. Повышение численности гусениц выше определенного порога приводит к необходимости учета как вертикального начального распределения листовертки по высоте кроны, так и динамики этого распределения. Таким образом, в окончательный вариант модели включена вертикальная стратификация кроны, в которой выделяется 20 слоев.

Идентификация модели позволяет перейти к заключительной стадии ее отработки — сравнению результатов моделирования с экспериментом. С этой целью автором модели В.В. Рубцовым выполнена имитация на ЭВМ процесса развития листовертки с 1969 по 1976 год в двух очагах, расположенных в Теллермановском опытном лесничестве. Экспериментальные данные о состоянии популяции в метеорологической обстановке получены Лабораторией лесоведения АН СССР. Действительная картина протекания вспышки численности в очагах "А" и "Б" и результаты моделирования представлены на рис. 1.9 и 1.10. Сплошные линии соответствуют наблюдениям, штриховые — расчету. Как видно из рисунков, модель обеспечивает удовлетворительную точность расчета динамики численности популяции листовертки.

Перечислим задачи, которые можно решать с использованием описанной имитационной модели. Во-первых, это задачи теоретического плана. "Прогризывая" на ЭВМ многообразные ситуации, соответствующие различным

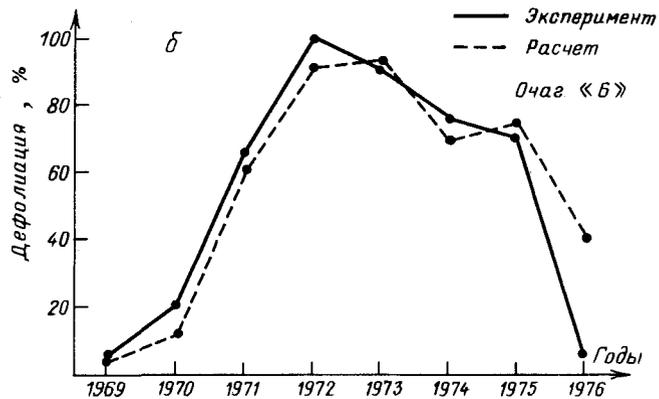
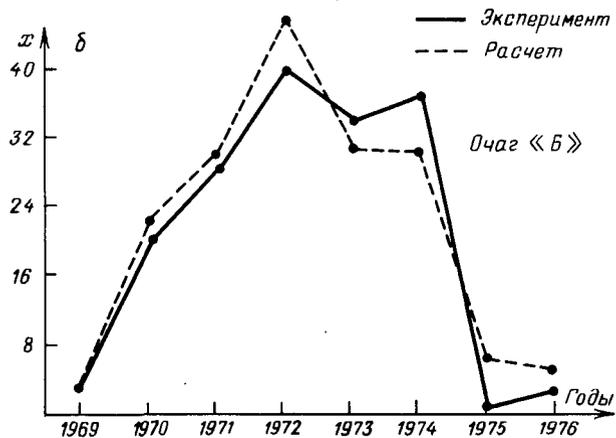
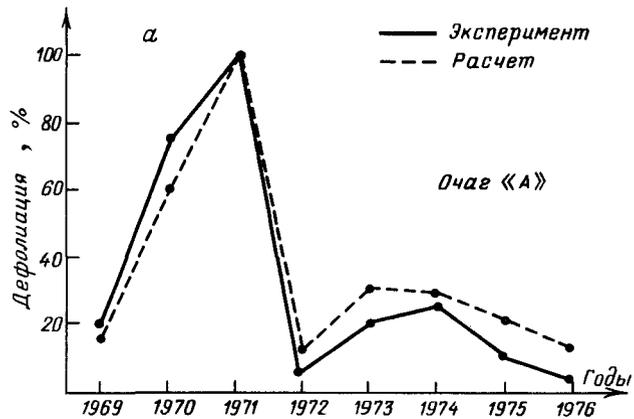
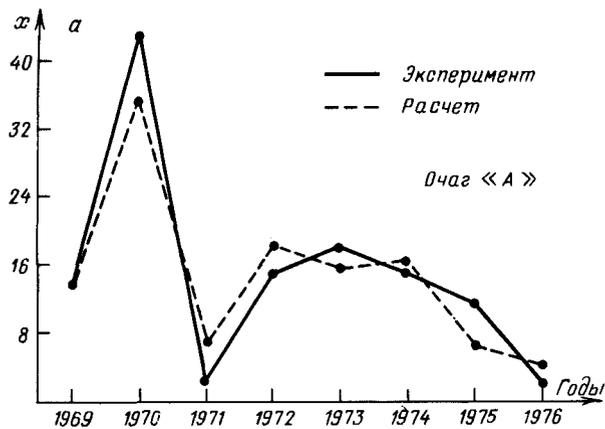


Рис. 1.9. Динамика плотности дубовой листовертки в очагах "А" и "Б"

Рис. 1.10. Степень повреждения листьев популяцией вредителей в очагах "А" и "Б"

метеорологическим условиям, состоянию популяции, разной степени "давления" на популяцию со стороны биотического окружения, можно выявить типичные картины поведения траекторий системы, условия, при которых наступают массовые размножения или депрессия популяции. Сложность модели заставляет относиться к ней как к некоторому природному объекту, обладающему, как и все сложные системы, контринтуитивным поведением. При этом для получения соответствующих выводов следует применять к машинному эксперименту такие же подходы, которые выработаны в настоящее время к постановке любых экспериментов.

§ 1.4. Задачи, возникающие при исследовании динамики экосистем

Изучение динамики экосистем средствами моделирования имеет своей конечной целью выработку методов прогнозирования влияния на экосистемы антропогенных воздействий, решение задач рационального использования природных ресурсов, что в конечном счете сводится к постановке проблемы управления в широком смысле этого слова. В то же время подход к решению этих задач не может быть вполне однозначным. Так, приведенные выше примеры показывают, что в относительно простых ситуациях, возникающих в рамках биотехнологий при культивировании микробных экосистем в искусственных условиях, описание их динамики может быть выполнено весьма детально, что позволяет надеяться на получение ряда аналитических выводов. В то же время при управлении природными комплексами соответствующие модели оказываются гораздо более сложными и получение тех или иных осмысленных рекомендаций оказывается возможным только на основе анализа результатов машинных экспериментов. Тем не менее в обоих случаях структура проблемы управления обладает рядом общих свойств, отличающих ее от задачи управления техническими объектами.

Специфику задач управления, связанных с динамикой эксплуатируемых популяций, можно проследить на простейшей модели рыболовного промысла Е. Шефера [49]. Пусть x — биомасса рыбной популяции, $G(x)$ — скорость биологического роста, которая в простейшем случае задается формулой $G(x) = rx(1 - x/k)$. Если через u обозначить интенсивность отлова, а через q — коэффициент вылова, то уравнение динамики численности эксплуатируемой популяции можно записать в виде

$$\dot{x} = G(x) - qux.$$

Ставится вопрос, при какой стратегии вылова u можно получить урожай, максимизирующий скорость получения прибыли. Для данной модели решение задачи существования и устойчивости стационарной точки, соответствующей оптимальной стратегии, тривиально и центр тяжести исследования переносится на выбор оптимального значения параметра u . Для популяции с возрастной структурой или для многокомпонентного сообщества вопрос существования и устойчивости равновесных состояний перерастает в самостоятельную задачу, требующую специального исследования.

Другой подход к эксплуатации рыбных ресурсов возникает в предположении, что число рыболовных судов растет пропорционально доходу от

рыболовства. В этом случае мы приходим к модели Смита [49]

$$\dot{x} = G(x) - qux,$$

$$\dot{u} = \theta(x, u),$$

где θ – некоторая функция, зависящая от плотности эксплуатируемой популяции и размера рыболовного флота. Приведенная система является простейшей системой непрямого управления, при котором управляющее воздействие ищется как решение некоторого дифференциального уравнения.

Несмотря на большое число работ, посвященных оптимальному управлению динамикой численности популяций и сообществ, проблеме разрешимости такого рода задач уделялось очень мало внимания. Дело в том, что при отыскании вида оптимального управления обычно используется какой-либо необходимый критерий оптимальности, например принцип максимума Л.С. Понтрягина. Получаемое в результате управление наверняка доставляет решение исходной задачи лишь в линейных моделях. В общем случае такое управление лишь "подозрительно на оптимальность" и специальной проблемой является доказательство того, что оно действительно является оптимальным решением. Для ответа на этот вопрос не существует готовых алгоритмов. Однако первым шагом на этом пути является доказательство существования хотя бы одного управления, переводящего объект в требуемое конечное состояние. Таким образом, встает проблема управляемости экологического объекта [7, 22].

Характерная особенность экосистем как объектов управления состоит в том, что, как правило, допустимыми являются лишь однонаправленные входные воздействия. Это отбор или отлов особей из популяции, добавление субстрата и изъятие приростной биомассы при культивировании, это внесение удобрений, поливы, укусы, прореживания в процессе формирования урожая и т.п. Конечно, легко представить себе воздействия на экосистему противоположного знака: добавление особей в эксплуатируемую популяцию, изъятие избыточного субстрата из культиватора, осушающие мероприятия на мелиорированных землях и т.п. Однако, если учесть затраты ресурсов на такие альтернативные воздействия, становится совершенно очевидной практическая ценность рассмотрения именно однонаправленных воздействий на объект. В такой постановке классические критерии управляемости неприменимы даже в случае линейных моделей [7, 22].

Имеется и еще одно обстоятельство, отличающее задачу управления экосистемами от классической. В данном случае фазовые переменные, означающие численность, концентрацию, плотность популяций, сухой вес биомассы растений и животных и т.п., могут принимать лишь неотрицательные значения и допустимые управления должны приводить к сохранению этого свойства. Отсюда естественным образом возникает необходимость привлечения новых методов, ориентированных на процессы управления объектами, эволюционирующими в конусах и имеющими такие ограничения на допустимые управления, при которых нулевые значения управляющих воздействий лежат на границе области их допустимых значений.

Перечисленные вопросы возникают в первую очередь при рассмотрении задач управления искусственно культивируемыми сообществами микроорганизмов. Ситуация существенно усложняется, когда мы обращаемся к управлению природными экосистемами. Действительно, хотя в обоих

случаях экосистемы включают в себя совокупности живых организмов (биоту) и их неживое ("косное") окружение, протекающие в этих частях процессы имеют различные уровни сложности. Так в природных системах в биоценозе складывается сложная иерархия трофических взаимодействий продуцентов, консументов различного уровня и редуцентов [10]. Эти взаимодействия характерны и для агроэкосистем, нацеленных "внешне" на культивирование популяций лишь одного вида – пшеницы, картофеля, сахарной свеклы или какой-либо другой культуры. В посевах культурных растений всегда присутствуют сорняки, насекомые-вредители, популяции болезнетворных организмов, не говоря уже о сложных экологических процессах в почвах. Значительно сложнее также протекает динамика и в косной части экосистемы. Если при исследовании биотехнологии мы имеем дело с одним или несколькими видами питательного вещества – субстрата, то в экосистемах следует рассматривать динамику трансформаций и переноса многих видов энергии и вещества, т.е. решать задачу моделирования процессов энерго- и массообмена. Это приход, отражение и поглощение солнечной радиации зелеными растениями, перенос тепла и влаги в почвах, растениях и приземном воздухе, трансформация и миграция соединений азота, фосфора, калия и других химических элементов в почвах и в растениях, утилизация органических остатков и гумуса, фотосинтез и дыхание растений и т.д. Все это говорит о том, что при исследовании природных экосистем аналитические результаты можно получить лишь при очень большом огрублении их описания, а в качестве основного метода следует использовать метод имитационного моделирования. Центр тяжести здесь переносится на выбор и обоснование численных методов решения задачи и организацию вычислений на ЭВМ.

Использование метода моделирования как мощного способа изучения динамики экосистем приносит ощутимую пользу уже на ранних стадиях, связанных с выбором структуры и уровня детализации модели. Действительно, уже на этих этапах приходится производить инвентаризацию как теоретических знаний и представлений об объекте исследования, так и той структуры экспериментальной информации, которой располагает исследователь или которая может быть доступна в ближайшем будущем. Уже этот вид деятельности позволяет более глубоко проникнуть в существо подлежащей изучению проблемы, целенаправленно спланировать эксперименты по дальнейшему изучению системы, хотя, разумеется, наибольший эффект достигается тогда, когда модель уже создана и доведена до уровня, допускающего ее практическое использование.

Охарактеризуем вкратце те проблемы, которые возникают в процессе разработки имитационных моделей. В первую очередь, как очевидно, они связаны с организацией работы на ЭВМ сложных программных комплексов. Это – организация исходных данных, организация "прогонов" модели и имитационного эксперимента в целом и организация общения пользователя с данными и с моделью. Все эти вопросы подлежат рассмотрению в данной книге (гл. 4 и 5). Важным, разумеется, является вопрос выбора численной схемы. Однако его рассмотрение целиком зависит от вида исходного математического описания (дифференциальные уравнения в обыкновенных или в частных производных, интегральные уравнения и т.д.), что выходит за рамки тематики монографии.

Имитационная модель, как известно, представляет собой реализованную на ЭВМ комплексную модель сложной системы, предназначенную для получения информации о поведении этой системы путем проведения машинных экспериментов. В некотором смысле имитационная модель является упрощенным аналогом реального объекта, исследование которого возможно только экспериментальным путем. Существенное преимущество имитационного эксперимента по сравнению с натурным состоит в скорости, сравнительной дешевизне и доступности самого богатого многообразия возможных вариантов. Кроме того, имитационная модель позволяет изучать ситуации, которые в естественных условиях привели бы к летальному исходу. Возникает вопрос, нужны ли аналитические модели и их качественное исследование в момент, когда создано столь мощное орудие анализа сложных экологических проблем. Опыт работы с имитационными моделями [33, 34] говорит о том, что качественное исследование обобщенных аналитических моделей совершенно необходимо для эффективного использования и осмысливания результатов как имитационного, так и натурального эксперимента. Дело в том, что аналитические модели носят концептуальный характер и позволяют отказаться от слепого перебора бесконечного числа вариантов, показать, какие задачи разрешимы, а какие принципиально не имеют решения в силу некорректности постановки. В то же время предварительное исследование позволяет сформулировать условия, при которых попытки решить проблему с помощью имитационного или натурального эксперимента должны в конечном счете привести к успеху. Примером подобной ситуации является исследование управляемости экологической системы однонаправленными воздействиями, представленное в главе 3.

ОБЩИЕ ДИНАМИЧЕСКИЕ СВОЙСТВА МОДЕЛЕЙ ЭКСПЛУАТИРУЕМЫХ ЭКОЛОГИЧЕСКИХ СИСТЕМ

Проблема управления экологическими системами прежде всего сводится к изучению возможностей выведения таких систем в требуемое конечное состояние с помощью физически реализуемых воздействий. Это — либо добавление в систему и только добавление, либо изъятие из нее и только изъятие. Переменные состояния, как правило, принимают значения лишь одного знака. Допустимые управления не должны нарушать это условие. Иными словами, встает задача управляемости при специфичных ограничениях как на фазовые переменные, так и на допустимые управления.

Решение этой задачи естественным образом распадается на два этапа. Вначале проводится анализ динамики биологических объектов при постоянных значениях управляющих параметров. Это — исследование устойчивых равновесных структур, анализ процесса потери устойчивого равновесия и рождения автоколебаний, исследование случая крайней неустойчивости или так называемых хаотических структур. Второй этап решения состоит в изучении реакций биологических объектов на изменение управляющих параметров. Данная глава содержит результаты первого этапа исследования.

§ 2.1. Равновесие и устойчивость

В этом параграфе приведены модели управляемых биологических систем общего вида, которые в качестве частных случаев содержат в себе ранее рассмотренные примеры. Исследованы вопросы существования, устойчивости равновесных состояний и условия возникновения устойчивых ограниченных структур. Полученные результаты применены к анализу процесса непрерывного культивирования микроорганизмов.

Устойчивость биологического объекта означает его способность реагировать на изменения окружающей среды и сохранять при этом качественно неизменное поведение. Наиболее простым типом такого поведения является постоянное поддержание состояния системы около положения равновесия. В исходных моделях каждому равновесному состоянию соответствует стационарное решение. В силу биологического смысла эти решения должны иметь неотрицательные компоненты и хотя бы одна компонента вектора стационарного состояния должна быть положительной. Последнее требование связано с тем, что нулевое решение соответствует случаю полного вырождения экологической системы и должно быть исключено.

2.1.1. Общий вид моделей динамики численности эксплуатируемых экосистем. Несмотря на большое содержательное различие моделей, приведенных в пп. 1.2.1 и 1.2.2, с точки зрения математического анализа они принадлежат к одному классу и их качественный анализ является специальной задачей, имеющей приложение во многих, довольно далеко отстоящих друг от друга областях математической биологии. В связи с этим в данной книге наряду с моделями численности популяций и сообществ в качестве примеров будут рассматриваться модели транспорта и преобразования веществ, относящиеся к моделям компартментального типа, т.е. образованным совокупностью камер или реакторов с перетоками.

Когда воспроизводство популяций или сообществ протекает в условиях, при которых ограниченностью питательных ресурсов, ограниченностью жизненного пространства и т.д. можно пренебречь, соответствующие модели с математической точки зрения совпадают с моделями расширяющейся экономики и подробно изучены [22].

Если выйти за рамки линейного подхода, то динамика численности экологической системы при нелимитированном воспроизводстве задается уравнением вида [42]

$$\dot{X} = F(X), \quad (2.1.1)$$

где $X \in R_+^m$ — численный состав популяции или сообщества. Функция $F: R_+^m \rightarrow R^m$ обычно предполагается непрерывной и такой, что при любом начальном условии $X(0) \in R_+^m$ уравнение (2.1.1) имеет единственное решение задачи Коши, определенное при $t \geq 0$.

Специфика модели сводится к следующим свойствам правой части уравнения (2.1.1). Функция F квазиположительна [26], т.е. каждая компонента $f_i = (F)_i$ оператора F такова, что

$$f_i(x_1, \dots, x_{i-1}, 0, x_{i+1}, \dots, x_m) \geq 0 \quad (i = 1, \dots, m),$$

квазимоноотонна [33] или положительно гомогенна [26], т.е. $f_i(X) \geq f_i(Y)$ при $X \geq Y$ и $(X)_i = (Y)_i$ и положительно однородна [22, 26], т.е. $F(\alpha X) = \alpha F(X)$ при $\alpha > 0$.

Случай отсутствия лимитирования является вырожденным и представляет собой лишь первый шаг в изучении сложных экологических систем. Влияние лимитирующих факторов состоит в том, что скорости преобразования вещества или энергии могут зависеть от количества этих субстанций как в некоторой части системы, так и во всей системе. Например, естественно считать, что скорость воспроизводства популяций или сообщества определяется обилием источников питания, обеспеченностью жизненным пространством и другими ограничивающими размножение факторами. Обычно среди них удается выделить главный или ведущий фактор, который отличается от остальных тем, что его изменение около некоторой средней величины приводит к значительно большим откликам в поведении экологической системы, чем малые вариации остальных факторов [13, 14, 25]. В зависимости от условий существования самые разные причины могут оказывать лимитирующий эффект, и поэтому главный лимитирующий фактор меняется при изменении состояния системы. Однако во всех случаях степень лимитирования задается скалярной величиной, являющейся функцией этого состояния.

Данное представление впервые было сформулировано Н.Д. Иерусалимским при анализе биохимических превращений [13, 14] и с математической точки зрения означает, что на множестве факторов, определяющих динамику процесса, задана так называемая функция предпочтения [34]. Вид соответствующей функции предпочтения обычно следует из биологических соображений. Тем не менее математические результаты [34] о существовании функции предпочтения на упорядоченном множестве могут быть полезны при моделировании механизмов лимитирования в случае, когда необходимые интуитивные представления оказываются недостаточными.

В общем виде уровень лимитирования определяется размерами экологической системы по формуле $s = \varphi(X)$, где функция предпочтения $\varphi: R_+^m \rightarrow R_+^1$ монотонна и положительно однородна, т.е. $\varphi(X) > \varphi(Y)$ при $X > Y$ и $\varphi(\alpha X) = \alpha \varphi(X)$ при $\alpha > 0$. Например, лимитирующий фактор может задаваться функцией вида $\varphi(X) = \min(x_1, x_2, \dots, x_m)$ или функцией

$$\varphi = \sum_{i=1}^m c_i x_i, \text{ где } c_i \geq 0 \text{ (} i = 1, \dots, m \text{) и } \sum_{i=1}^m c_i > 0. \text{ Скорость воспроизводства}$$

экологической системы в условиях лимитирования определяется функцией $F(X, s)$, которая при каждом фиксированном s , меньшем некоторой величины s_∞ , обладает теми же свойствами, что и правая часть уравнения (2.1.1). Если уровень лимитирования больше критического, то экологическая система теряет способность к воспроизводству, т.е. при фиксированном $s \geq s_\infty$ все решения уравнения $\dot{X} = F(X, s)$, начинающиеся в положительном конусе R_+^m , ограничены и не выходят из этого конуса.

Рассмотренный нами механизм лимитирования позволяет учитывать богатое разнообразие возможных природных ситуаций, поскольку при смене внешних условий может происходить смена лимитирующего фактора. Существенным здесь является лишь предположение о том, что скорости процессов в экологической системе определяются скалярной величиной, а не набором двух или более факторов. В последнем случае мы имеем дело с векторным представлением механизма лимитирования.

Итак, мы приходим к системе уравнений

$$\dot{X} = F(X, s), \quad s = \varphi(X).$$

Динамика эксплуатируемых экологических систем, в которых скорость изъятия i -й ($i = 1, \dots, m$) компоненты пропорциональна ее величине, задается уравнением вида

$$\dot{X} = F(X, \varphi(X)) - vX, \tag{2.1.2}$$

где неотрицательный скаляр v определяет скорость изъятия.

Если экологическая система пополняет свои запасы из внешних источников, то динамика ее численности задается уравнением

$$\dot{X} = F(X, \varphi(X)) - vX + Bu, \tag{2.1.3}$$

где вектор $B \geq 0$ определяет структуру внешнего притока, а u — его интенсивность. Уравнение типа (2.1.3) возникает при изучении динамики численности популяций, в которых имеется иммиграция из соседних регионов или добавление особей для активизации воспроизводства запасов.

Для того чтобы гарантировать существование и единственность решений задачи Коши при всех биологически осмысленных начальных условиях,

положим, как это обычно принято, что функции $F(X, s)$ и $\varphi(X)$ удовлетворяют условию Липшица

$$\begin{aligned} \|F(X_1, s_1) - F(X_2, s_2)\| &\leq l(\|X_1 - X_2\| + |s_1 - s_2|), \\ \|\varphi(X_1) - \varphi(X_2)\| &\leq l\|X_1 - X_2\|. \end{aligned}$$

Более того, будем считать, что константа Липшица l одна для всех $X_i \in R_+^m$ и $s_i \in R_+^1$ ($i = 1, 2$) на каждом ограниченном множестве из произведения $R_+^m \times R_+^1$.

2.1.2. Существование и основные свойства стационарных решений. Динамика нелимитированных экологических систем вида (2.1.1) хорошо изучена. Ее характерная особенность состоит в том, что суммарная численность популяции или сообщества экспоненциально возрастает. В то же время отношение численности любой компоненты экосистемы к общему числу особей стремится к постоянной величине. Если в игру вступают лимитирующие факторы, то происходит утрата этих свойств. В результате усложняется качественная картина поведения системы. Предварим анализ динамики численности систем с лимитированием рассмотрением случая, когда уровень лимитирования фиксирован и не меняется в процессе движения.

Пусть $X(t, X_0)$ — решение уравнения (2.1.1) с начальным условием $X(0, X_0) = X_0$. Определим оператор сдвига вдоль траекторий $\Gamma(t)$ по следующему правилу: $\Gamma(t)X_0 = X(t, X_0)$ [15]. В силу квазимонотонности и положительной однородности правой части уравнения (2.1.1) оператор $\Gamma(t)$ обладает следующими свойствами: оставляет инвариантным положительный конус, монотонен и положительно однороден, т.е.

$$\begin{aligned} \Gamma(t): R_+^m &\rightarrow R_+^m, \\ \Gamma(t)X &\geq_{\neq} \Gamma(t)Y \text{ при } X \geq Y \text{ и } X \neq Y \end{aligned}$$

и

$$\Gamma(t)\alpha X = \alpha\Gamma(t)X.$$

Доказательство положительной инвариантности приведено в книге [15]. Доказательство монотонности является тривиальным следствием леммы сравнения [33]. Доказательство положительной однородности состоит в переходе от дифференциального уравнения к его интегральному аналогу и использовании процедуры последовательных приближений для построения решения.

Содержательный смысл вышеупомянутых свойств состоит в следующем: во-первых, решения уравнений во все моменты времени принимают лишь неотрицательные значения, и поэтому модель не приводит к бессмысленным результатам, во-вторых, на стадии нелимитированного воспроизводства система с большей начальной численностью за конечное время достигает больших размеров, и, в-третьих, одинаковое изменение масштаба измерения для всех компонент вектора состояния нелимитированной экологической системы не изменяет общего характера ее динамики.

Может случиться, что экологическая система содержит меньшую подсистему, функционирование которой не зависит от ее остальных компонент. Исследование таких систем естественным образом распадается на два этапа. Вначале проводится анализ динамики независимой подсистемы.

Затем полученные данные используются в качестве исходной информации для изучения системы в целом. Продолжая выделение меньших подсистем, по индукции мы приходим к анализу неразложимой системы, для которой подобное разбиение невозможно. Сформулируем строгое математическое определение понятия неразложимости.

Определение [22, 24]. Пусть даны векторы X и Y , удовлетворяющие условию $X \geq Y \geq 0$. Введем в рассмотрение множество индексов $N(X, Y) = \{i \mid x_i > y_i\}$. Квазимонотонный оператор $F = (f_1, f_2, \dots, f_m) : R_+^m \rightarrow R_+^m$ называется *неразложимым*, если для любой пары векторов $X \not\geq Y$ имеет место неравенство $f_i(X) \neq f_i(Y)$ при некотором $i \in N(X, Y)$. Здесь и всюду далее неравенство векторов понимается покомпонентно.

Неразложимость правой части уравнения (2.1.1) позволяет усилить свойство монотонности. Покажем, что увеличение хотя бы одной компоненты вектора состояния системы в начальный момент времени приводит к увеличению через конечный промежуток времени всех компонент, составляющих данную экологическую систему.

Предложение 2.1.1. Если функция $F(X)$ неразложима, то оператор сдвига вдоль траекторий уравнения (2.1.1) примитивен [22, 24], т.е. истинно следующее утверждение:

$$\Gamma(t)X > \Gamma(t)Y \text{ при } X \geq Y, X \neq Y \text{ и } t > 0. \quad (2.1.4)$$

Доказательство. Пусть $X(t)$ и $Y(t)$ – решения уравнения (2.1.1) с начальными условиями X_0 и Y_0 соответственно и $X_0 \geq Y_0$. Обозначим $W = (w_1, w_2, \dots, w_m)^* = X - Y$. Тогда $\dot{W} = F(Y + W) - F(Y) = G(t, W)$. Оператор $G(t, W)$ квазиположителен, квазимонотонен и неразложим при любом $t \geq 0$. Следовательно, $W(t) \geq 0$ при $t \geq 0$. Пусть $w_i(0) > 0$ при $i \in I$ и $w_j(0) = 0$ при $j \in J$, где $I \cup J = \{1, 2, \dots, m\}$. По непрерывности $w_i(t) > 0$ при достаточно малом t . Неразложимость означает, что найдется $j \in J$, для которого j -я компонента $G(0, W(0))_j > 0$. Следовательно, $w_j(t) > 0$ при $t > 0$. Рассуждая по индукции, легко показать, что $w_i(t) > 0$ при всех $i \in \{1, 2, \dots, m\}$. Предложение доказано.

Если экологическая система неразложима, то независимо от соотношения численностей особей разных видов в начальный момент времени при достаточно длительном периоде размножения общее число особей экспоненциально увеличивается, а отношение численности любой компоненты экосистемы к общему числу особей устанавливается постоянным. Таким образом, достигается состояние, называемое в математической экономике сбалансированным ростом. Строго говоря, это означает следующее [22, 24].

При выполнении условий монотонности, положительной однородности и примитивности оператор сдвига вдоль траекторий $\Gamma(t)$ уравнения (2.1.1) имеет единственный с точностью до постоянного множителя собственный вектор H , которому отвечает собственное значение $e^{\lambda t}$: $\Gamma(t)H = e^{\lambda t}H$. Более того, для любого неотрицательного вектора $X \neq 0$ найдется такое положительное число c , что

$$\Gamma(t)X/e^{\lambda t} \rightarrow cH \text{ при } t \rightarrow \infty. \quad (2.1.5)$$

Очевидно, что H является собственным вектором оператора F , которому отвечает собственное значение λ . Свойство (2.1.5) позволяет определить следующее понятие.

Определение. Любой неотрицательный вектор, принадлежащий лучу cH ($c > 0$), называется *точкой сбалансированного роста* [22, 24].

Учет механизма лимитирования приводит к тому, что условия однородности и монотонности оператора сдвига вдоль траекторий уравнения (2.1.1) перестают выполняться. В моделях (2.1.2) и (2.1.3) появляется возможность рождения стационарных решений. Эти решения соответствуют равновесному состоянию, при котором экологическая система сохраняет постоянную численность. Естественно возникают вопросы о единственности стационарных решений, о непрерывной зависимости их от параметров и т.д.

Прежде чем перейти к рассмотрению поставленных вопросов, изложим подробно свойства оператора $F(X, s)$, стоящего в правой части уравнений (2.1.2), (2.1.3). В дальнейшем наряду с обозначением $F(X, s)$ будем пользоваться обозначением $F_s(X) = F(X, s)$.

В соответствии с результатом (2.1.5) заключаем, что при всех $s < s_\infty$ определено отображение $s \rightarrow \lambda_s, H_s$, сопоставляющее каждому фиксированному значению уровня лимитирования s собственный вектор и собственное значение нелинейного оператора $F_s(X)$.

Предложение 2.1.2 Отображение $s \rightarrow \lambda_s, H_s$, определение при $s < s_\infty$, непрерывно.

Доказательство. Множество собственных векторов, принадлежащих единичному симплексу $\sum_{i=1}^m x_i = 1$, компактно. По технической лемме 1 (приложение 1) отображение $s \rightarrow H_s$ непрерывно. Если s принадлежит компактному интервалу $s_1 \leq s \leq s_2 < s_\infty$, то множество $\lambda_s = \sum_{i=1}^m \{(F(H_s, s))_i | s \in [s_1, s_2]\}$ также компактно. По технической лемме 1 (приложение 1) получаем непрерывность. Предложение доказано.

Замечание. Как видно из доказательства, условие неразложимости не является необходимым требованием для непрерывности отображения. Достаточно гарантировать единственность собственного вектора и собственного значения оператора $F_s(\cdot)$ при любом фиксированном значении s из области определения.

Остановимся подробнее на свойствах операторов $F(X, s)$ как функции второго аргумента. Естественно предположить, что увеличение уровня лимитирования понижает скорость воспроизводства в точке сбалансированного роста, т.е.

$$F_{s_2}(H_{s_1}) \leq F_{s_1}(H_{s_1}) \text{ при } s_1 < s_2 < s_\infty, \quad (2.1.6)$$

где H_{s_1} — собственный вектор оператора F_{s_1} , отвечающий собственному значению λ_{s_1} при фиксированном $s = s_1$.

Приведем более наглядную трактовку биологического смысла данного предположения. Собственное значение оператора является основной характеристикой способности экологической системы к воспроизводству. Если величина λ_s отрицательна и уровень лимитирования постоянен, то сообщество вырождается. Если эта величина положительна и время наблюдения за сообществом достаточно велико, то его численность экспоненциально возрастает. Оба эти утверждения являются очевидными следствиями формулы (2.1.5). Отношение скорости роста численности экосистемы к самой численности задает относительную скорость воспроизводства этой экосистемы. Покажем, что относительная скорость воспроизводства убывает при увеличении лимитирования.

Предложение 2.1.3. Если выполняется условие (2.1.6), то собственное значение λ_s является монотонно убывающей функцией параметра s :

$$\lambda_{s_2} < \lambda_{s_1} \quad \text{при } s_1 < s_2 < s_\infty. \quad (2.1.7)$$

Доказательство. Техническая лемма 2 (см. приложение 1) позволяет, не умаляя общности, считать операторы F_{s_1} и F_{s_2} положительными и монотонными. Воспользовавшись результатами, приведенными в книгах [22, 24], завершаем доказательство предложения.

В дальнейшем условие (2.1.6) всегда будем считать выполненным. Обратимся теперь к вопросу существования стационарных решений уравнений (2.1.2) и (2.1.3).

Стационарное решение уравнения (2.1.2) при фиксированном v удовлетворяет соотношению

$$F(X, s) = vX, \quad s = \varphi(X). \quad (2.1.8)$$

Отсюда видно, что задача (2.1.8) является обобщением на системы с лимитированием проблемы существования положительного собственного вектора v монотонного и однородного нелинейного оператора [22, 24].

Стационарное решение уравнения (2.1.3) при фиксированных v и u удовлетворяет соотношению

$$vX - F(X, s) = Bu, \quad s = \varphi(X). \quad (2.1.9)$$

Доказательство разрешимости этой системы является обобщением на системы с лимитированием результатов Х. Никайдо о резольвенте нелинейного монотонного и однородного оператора [24]. Напомним, что в обоих случаях нас интересуют неотрицательные ненулевые решения приведенных выше систем. Это обстоятельство вызвано тем, что переменные в моделях имеют биологический смысл лишь в том случае, когда они принимают неотрицательные значения, а нулевое решение соответствует полному вырожждению системы.

Во всех последующих рассуждениях будем предполагать, что при $s < s_\infty$ оператор F_s неразложим, имеет в положительном конусе R_+^m единственный собственный вектор, компоненты которого строго положительны, и этому вектору отвечает собственное значение, которое удовлетворяет условию (2.1.7).

Теорема 2.1.1. Для того чтобы система (2.1.8) имела нетривиальные решения X ($X \neq 0$ и $\varphi(X) < s_\infty$), необходимо выполнение неравенства

$$\lambda_0 > v, \quad (2.1.10)$$

где λ_0 – собственное значение оператора $F_0(\cdot)$.

Доказательство. Пусть X – решение системы (2.1.8). Так как $F_{\varphi(X)}(X) = vX$, то $\lambda_{\varphi(X)} = v$ и $X > 0$ – собственный вектор оператора $F_{\varphi(X)}$. Следовательно, $0 < \varphi(X)$ и $\lambda_0 > \lambda_{\varphi(X)} = v > 0$. Теорема доказана.

Содержательный смысл условия (2.1.10) состоит в том, что при отсутствии лимитирования асимптотическая скорость воспроизводства должна превышать скорость изъятия. Нарушение этого условия приводит к тому, что экологическая система вырождается и единственным стационарным решением оказывается нулевое значение вектора состояния. Условие (2.1.10) является также достаточным для существования стационарного решения уравнения (2.1.2).

Теорема 2.1.2. Если выполнено условие (2.1.10), то при каждом фиксированном $0 < v < \lambda_0$ существует единственное решение X_v системы (2.1.8). Если, кроме того, $\lambda_s < 0$ при некотором $s < s_\infty$, то решение существует при $v = 0$. Функция $v \rightarrow X_v$ непрерывна, принимает строго положительные значения, и отображение $v \rightarrow \lambda_\varphi(X_v)$ — тождественное.

Доказательство. Так как $\lim_{s \rightarrow s_\infty} \lambda(s) \leq 0$, то по условию (2.1.7) найдется $s(v) < s_\infty$ такое, что $\lambda_{s(v)} = v > 0$. Собственный вектор X_v оператора $F_{s(v)}$, отнормированный так, что $\varphi(X_v) = s(v)$, является искомым решением. Если $\lambda_s < 0$ при $s < s_\infty$, то в силу (2.1.10) и предложения 2.1.2 $\lambda_{s_0} = 0$ при некотором $0 < s_0 < s$. Непрерывность отображения следует из предложения 2.1.2. Теорема доказана.

Обратимся теперь к рассмотрению вопроса о существовании стационарных решений в системах с внешними источниками. Напомним, что $v \geq 0$ и $u \geq 0$. Разобьем все множество ненулевых и неотрицательных решений системы (2.1.9) на три взаимно непересекающихся подмножества. На каждом из них будет выполняться одно из следующих неравенств: $\lambda_s > v$, $\lambda_s = 0$ и $\lambda_s < v$.

Теорема 2.1.3. Если $\lambda_s > v$, то система (2.1.9) неразрешима. Если $\lambda_s = v$, то система (2.1.9) разрешима тогда и только тогда, когда $u = 0$. Множество таких решений описано в теореме 2.1.2. Если $\lambda_s < v$, то система (2.1.9) разрешима. Множество таких решений задается графиком непрерывного отображения $s, v \rightarrow u, X$, которое определено на множестве $\{s \in (0, s_\infty), v \in \mathring{R}_+^1 \mid \lambda_s < v\}$ и принимает значения из области $\mathring{R}_+^1 \times \mathring{R}_+^m$.

Если зафиксировать s и интересоваться лишь значениями u , то из исходного отображения мы получим неубывающую функцию $v \rightarrow u$.

Доказательство. Вследствие технической леммы 2 (приложение 1) каждый оператор семейства $\{F_s(X) \mid 0 \leq s \leq s_1, s_1 < s_\infty\}$ при любом конечном s_1 можно считать монотонным. Поскольку s_1 можно выбрать сколь угодно близким к s_∞ , получаемый ниже результат верен для любого $s < s_\infty$. Дальнейшие рассуждения разобьем на ряд этапов.

Лемма 2.1.3.1. Существует определенная на $\Omega = \{s, v, B \mid 0 < s < s_\infty, \lambda_s < v, B \geq 0\}$ непрерывная функция $s, v, B \rightarrow (\alpha, H) \in \mathring{R}_+^1 \times \mathring{R}_+^m$ такая, что первое уравнение системы (2.1.9) при $u = \alpha(s, v, B)$ имеет единственное решение

$$H = H(s, v, B) \quad (\rho(H) = 1) \text{ и } \alpha \rightarrow 0 \text{ при } v \rightarrow \lambda_s.$$

Доказательство. Пусть Y — единственное решение уравнения $vY - F_s(Y) = B$ [22, 24]. Вектор $H = \alpha Y$, где $\alpha = 1/\rho(Y)$, удовлетворяет условию $\rho(H) = 1$ и является решением первого уравнения системы (2.1.9) при $u = \alpha$. Из определения H следует, что $\alpha = (v - \varphi(F_s^-(H))) / \rho(B)$. Так как H принадлежит единичному симплексу и $1/\rho(B)$ ограничено в окрестности каждой ненулевой точки B_0 , то условие технической леммы 1 (приложение 1) выполнено. Непрерывная зависимость H от s, v и B следует из условия $\rho(H) = 1$ и технической леммы 1 (приложение 1).

Лемма 2.1.3.2. При любых $s, v, B \in \Omega$ и любом $u > 0$ решение первого уравнения системы (2.1.9) представимо в виде $(u/\alpha)H$, где α и H — функции из леммы 2.1.3.1. Отображение $v \rightarrow 1/\alpha$ — монотонно невозрастающее. Доказательство этого факта следует из [24, с. 223], так как $\alpha = 1/\rho(Y)$, и непрерывности отображения $B \rightarrow \alpha$, доказанной в предыдущей лемме.

Лемма 2.1.3.3. Система уравнений (2.1.9) при $s, v, B \in \Omega$ однозначно разрешима относительно u и X . Отображение $s, v \rightarrow u, X$ непрерывно, $u \rightarrow 0$ при $v \rightarrow \lambda_s$ и $u(s, v_1) > u(s, v_2)$ при $v_1 > v_2 > \lambda_s$. Если $v < \lambda_s$, то система (2.1.9) неразрешима.

Доказательство. По лемме 2.1.3.2 решение системы (2.1.9) ищется из условия $u\varphi(H)/\alpha = s$. Так как H и s — фиксированные величины, дальнейшие рассуждения очевидны [22, 24]. Теорема доказана.

Таким образом, мы видим, что нетривиальными стационарными решениями уравнений динамики численности эксплуатируемых экологических систем являются векторы со строго положительными компонентами. В системах без внешних источников интенсивность потребления не должна превышать уровень, при котором наступает вырождение, поскольку стационарному режиму соответствует такое значение лимитирующего фактора, при котором асимптотическая скорость воспроизводства равняется скорости потребления. При наличии внешних источников интенсивность потребления сверху неограничена. Стационарному режиму в этом случае соответствует такое значение лимитирующего фактора, при котором интенсивность потребления превышает асимптотическую скорость размножения. Увеличение притока извне допускает увеличение интенсивности потребления при сохранении уровня лимитирования.

Обычно правая часть уравнений (2.1.2) и (2.1.3) зависит от определенного набора параметров, численные значения которых можно определить приближенно с некоторой ограниченной степенью точности. Корректность модели состоит в том, что малые ошибки в определении значений этих параметров не должны вести к значительным изменениям в доставляемых ею решениях. Поэтому несомненный интерес представляет вопрос о характере зависимости стационарных решений систем (2.1.2) и (2.1.3) от их параметров. Ответ на него дает

Теорема 2.1.4. Пусть функция, задающая механизм воспроизводства, зависит от некоторого набора параметров: $F = F(X, s, \beta_1, \dots, \beta_n)$. Если она непрерывна по совокупности всех своих $m + n + 1$ переменных, при $s < s_\infty$ и всех допустимых значениях параметров β_1, \dots, β_n имеет единственный собственный вектор единичной нормы, принадлежащий положительному конусу R_+^m , то стационарное решение уравнения (2.1.2) непрерывно зависит от $v, \beta_1, \dots, \beta_n$, а стационарное решение уравнения (2.1.3) непрерывно зависит от $v, u, B, \beta_1, \dots, \beta_n$.

Доказательство. Доказательство непрерывной зависимости стационарного решения уравнения (2.1.2) от параметров состоит в дословном повторении рассуждений предложения 2.1.2 и теоремы 2.1.2.

Доказательство непрерывной зависимости стационарного решения уравнения (2.1.3) от параметров состоит в дословном повторении рассуждений предложения 2.2.2 и теоремы 2.1.3. Теорема доказана.

Таким образом, малые вариации параметров системы должны приводить к небольшим изменениям величины стационарного решения, и в этом смысле задача поставлена корректно.

2.1.3. Устойчивость стационарных решений. В предыдущем пункте были получены условия существования нетривиальных стационарных решений в системах вида (2.1.2) и (2.1.3). Вопрос практической реализуемости этих решений сводится к определению условий их устойчивости. Проведение оценки области притяжения устойчивых стационарных режимов представляет самостоятельный интерес. Если стационарное решение не является устойчивым, то возникает задача об условиях ограниченности траекторий системы и возможности введения управления, стабилизирующего заданный режим ее функционирования. В данном пункте приведены результаты исследований устойчивости нетривиальных стационарных решений уравнений (2.1.2) и (2.1.3).

Напомним, что фазовым пространством этих уравнений является положительный конус m -мерного евклидова пространства. Как обычно, будем предполагать, что увеличение уровня лимитирования снижает скорость воспроизводства в точке сбалансированного роста, т.е. выполнено условие (2.1.6). Выделим ряд характерных типов обратной связи, которую оказывает механизм лимитирования на динамику воспроизводства популяции или сообщества.

1. Равномерное лимитирование:

$$F(X, s) = p(s)F(X), \quad (2.1.11)$$

$$p(s_2) < p(s_1) \text{ при } s_1 < s_2 < s_\infty \text{ (см. формулу (2.1.6)).}$$

Содержательный смысл этого условия состоит в одинаковом влиянии механизмов лимитирования на все компоненты экологической системы. Такой тип обратной связи иллюстрирует граф 1, представленный на рис. 2.1.

2. Автолимитирование:

$$s = x_i \text{ для некоторого } 1 \leq i \leq m,$$

$$f_j(X, s) = f_j(X) \text{ при } j \neq i, \quad (2.1.12)$$

$$f_i(H_1, s_2) < f_i(H_1, s_1) \text{ при } s_1 < s_2 < s_\infty \text{ (см. формулу (2.1.6)),}$$

где H_1 – собственный вектор оператора $F(X, s_1)$. Содержательный смысл условия (2.1.12) состоит в том, что лимитирующий фактор влияет только на i -ю компоненту экологической системы и зависит только от этой компоненты. Такой тип обратной связи иллюстрирует граф 2, представленный на рис. 2.1.

3. Интегральное лимитирование:

При таком типе лимитирования постоянное значение лимитирующего фактора является интегралом системы (2.1.2). Это означает, что $\dot{s} = 0$ при некотором $s = s_0$. Вычислим скорость изменения уровня лимитирования вдоль траекторий системы (2.1.2). Пусть функция $\varphi(X)$ непрерывно дифференцируема в $R_+^m \setminus 0$. Тогда имеем $\dot{s} = \langle \partial\varphi/\partial X, F(X, s) \rangle - v \langle \partial\varphi/\partial X, X \rangle$. Обозначим $f(X, s) = \langle \partial\varphi/\partial X, F(X, s) \rangle$. Воспользовавшись формулой Эйлера для однородных функций, получаем $\dot{s} = f(X, s) - v\varphi(X)$. Интегральное лимитирование характеризуется тем, что

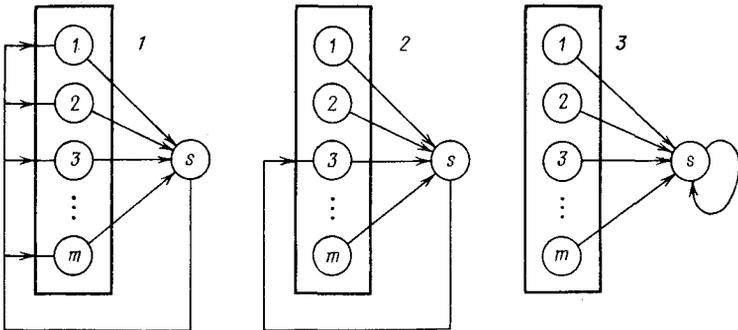


Рис. 2.1. Типы лимитирования

выполнены условия

$$f(X, s_0) = \lambda_{s_0} \varphi(X) \text{ в некоторой точке } 0 < s_0 < s_\infty \quad (2.1.13)$$

и

$$(f(X, s) - f(X, s_0))(s - s_0) < 0 \text{ при } X > 0 \text{ и } s \neq s_0.$$

Содержательный смысл условия (2.1.13) состоит в том, что система стремится "стабилизировать" некоторое стационарное значение лимитирующего фактора. Соответствующий случаю (2.1.13) тип обратной связи иллюстрирует граф З, представленный на рис. 2.1.

Перечисленные механизмы лимитирования гарантируют стремление к равновесному состоянию в модели динамики численности популяций и сообществ при отсутствии внешних источников.

Теорема 2.1.5. Пусть уравнение (2.1.2) имеет ненулевое стационарное решение $X_0, s_0 = \varphi(X_0)$ ($s_0 < s_\infty$). Если выполнено условие (2.1.11), (2.1.12) или (2.1.13), то всякое решение уравнения (2.1.2), начинающееся в $R_+^m \setminus 0$, стремится к стационарной точке X_0 при $t \rightarrow \infty$.

Доказательство. 1. Пусть выполнено условие (2.1.11). Обозначим $\hat{p}(t) = p(\varphi(X(t)))$, где $X(t)$ — решение уравнения (2.1.2) при нетривиальных начальных условиях. Докажем, что $\int_0^t \hat{p}(\tau) d\tau \rightarrow \infty$ при $t \rightarrow \infty$. Допустим противное. Предположим вначале, что $v = 0$. Тогда собственное значение оператора $p(s_0)F(X)$ равняется нулю. В силу условия (2.1.10) это означает, что $p(s_0) = 0$. Воспользовавшись формулой (2.1.11), заключаем, что R_+^m не является положительно инвариантным множеством системы при $s_0 < s < s_\infty$. Таким образом, $v > 0$.

Воспользовавшись однородностью оператора $F(X)$, найдем такое число M , что $|\langle F(X), X \rangle| \leq M \|X\|^2$. Решение системы удовлетворяет равенству $\dot{X}(t) = \hat{p}(t)F(X(t)) - vX(t)$. Отсюда имеем $0,5 \| \dot{X} \|^2 \leq -v \| X \|^2 + M \hat{p}(t) \| X \|^2$. Следовательно, $\|X(t)\| \rightarrow 0$ и $\hat{p}(t) \rightarrow p(0)$ при $t \rightarrow \infty$. Получили противоречие. Итак, $\int_0^t \hat{p}(\tau) d\tau \rightarrow \infty$. Делая замену переменных $Y = e^{vt}X$ и пользуясь положительной одно-

родностью $F(X)$, имеем $\dot{Y} = \hat{p}(t)F(Y)$. Делая замену времени $\tau = \int_0^t \hat{p}(t) dt$, получаем по формуле (2.1.5): $Y(\tau) = ce^{\lambda\tau}X_0 + o(e^{\lambda\tau})$, где $c > 0$ и λ — собственное значение оператора $F(X)$, отвечающее собственному вектору X_0 . Возвращаясь к исходным переменным, имеем $X(t) = x(t)X_0 + o(x(t))$, где $\|o(x(t))\|/x(t) \rightarrow 0$ при $t \rightarrow \infty$. Поскольку на инвариантном луче $\{\alpha X_0 | \alpha > 0\}$ стационарное решение X_0 асимптотически устойчиво, очевидно, $X(t) \rightarrow X_0$ и в общем случае.

2. Пусть выполнено условие (2.1.12). По построению оператор $F(X, \varphi(X))$ — квазимонотонный. Поэтому оператор сдвига вдоль траекторий $\Gamma(t)$ — монотонный. Покажем, что $\Gamma(t)X_* \in R_+^m$ при любом $t > 0$ и $X_* \in R_+^m$. Действительно, оператор $F(X, t) = F(X, \varphi(\Gamma(t)X_*))$ — квазимонотонный, положительно однородный и неразложимый. Следовательно, любое решение уравнения $\dot{X} = F(X, t) - vX$, начинающееся в R_+^m , принадлежит R_+^m при $t > 0$ (рассуждения те же, что и при доказательстве предложения 2.2.1). Функция $X(t) = \Gamma(t)X_*$ удовлетворяет этому уравнению, следовательно, $\Gamma(t)X_* > 0$ при $t > 0$. В дальнейших рассуждениях можно считать, что начальное условие X_* любого решения принадлежит R_+^m , и поэтому $\varphi(X_*) = s_* > 0$. Обозначим через H_0 собственный вектор оператора $F(\cdot, 0)$. По непрерывности и из условия (2.1.10) имеем, что $F(H_0, s) - vH_0 > 0$ при достаточно малых s . Так как $X_* > 0$ и $s_* > 0$, найдется $0 < \epsilon < s_*$ такое, что $\epsilon H_0 / \varphi(H) < X_*$. Пусть $H = \epsilon H_0 / \varphi(H_0)$. Тогда $\dot{X}|_{X=H} > 0$, и в силу монотонно-

сти оператора $\Gamma(t)$ имеем $\Gamma(t)H \leq \Gamma(t)X_*$. Функция $\Gamma(t)H$ монотонно возрастает со временем.

Действительно, $\Gamma(\tau)H > H$ при достаточно малом τ , так как $\dot{X}|_{X=H} > 0$. Рассуждая по индукции и пользуясь монотонностью имеем $\Gamma^n(\tau)H = \Gamma(n\tau)H \geq \Gamma(k\tau)H \geq H$ при любых $n \geq k \geq 1$.

Построим теперь мажоранту решения уравнения (2.1.2). Доопределим правую часть уравнения (2.1.2) по правилу $F(X, s) = F(X, s_\infty)$ при $s > s_\infty$. В этом случае точка X_0 остается единственным нетривиальным стационарным решением уравнения (2.1.2). По условию теоремы найдется $s_0 < s < s_\infty$ такое, что $\lambda_s - v < 0$. Выберем собственный вектор Z оператора $F(\cdot, s)$, отвечающий λ_s , так, чтобы $Z > X_0$ и $\varphi(Z) \geq s$. Тогда по условию (2.1.12) имеем $\dot{X}|_{X=Z} < 0$. Рассуждая так же, как и в предыдущем случае, имеем

$$\Gamma(t)H \leq \Gamma(t)X_* \leq \Gamma(t)Z,$$

где $\Gamma(t)H$ монотонно возрастает, а $\Gamma(t)Z$ монотонно убывает. Следовательно, эти функции стремятся к некоторому пределу при $t \rightarrow \infty$. В силу единственности стационарного решения уравнения (2.1.2) стремятся к этому решению. Утверждение доказано.

3. Пусть выполнено условие (2.1.13). Как показано в предыдущем пункте доказательства теоремы, можно считать, что $\dot{X}(t) > 0$ при $t \geq 0$. Из формулы (2.1.2) и тождества Эйлера для однородных функций имеем

$$\frac{d}{dt} \varphi(X) = f(X, \varphi(X)) - f(X, \varphi(X_0)) + f(X, s_0) - v\varphi(X). \quad (2.1.14)$$

Воспользовавшись формулой (2.1.13), получаем $f(X, s_0) - v\varphi(X) = 0$ и $d/dt (\varphi(X) - \varphi(X_0)) (\varphi(X) - \varphi(X_0)) < 0$ при $\varphi(X) \neq s_0$, и, следовательно, $\varphi(X(t)) \rightarrow \varphi(X_0)$ вдоль любого решения уравнения (2.1.2). Если $\varphi(X(0)) = s_0$ в начальный момент времени, то $X(t)$ удовлетворяет уравнению $\dot{X} = F(X, s_0) - vX$. В силу формулы (2.1.5) имеем $X(t) \rightarrow X_0$ при $t \rightarrow \infty$. Отсюда и из предыдущих рассуждений, очевидно, следует, что $X(t) \rightarrow X_0$ при любом начальном условии $X(0) \in R_+^m \setminus 0$ и $t \rightarrow \infty$. Теорема доказана.

Чтобы установить факт устойчивости по Ляпунову стационарного решения в изучаемых моделях, заменим условие непрерывности правых частей уравнений (2.1.2) и (2.1.3) на более сильное условие непрерывной дифференцируемости. Непрерывная дифференцируемость однородного оператора в окрестности начала координат эквивалентна линейности. Об этом говорит

Предложение 2.1.4. Если оператор $F(X, s)$ непрерывно дифференцируем по X в точке $X = 0$, то имеет место представление $F(X, s) = A(s)X$, где $A(s)$ при каждом фиксированном s является матрицей с неотрицательными недиагональными элементами.

Доказательство. Выпишем тождество $F(\epsilon X, s) = \epsilon D_X F(0, s)X + o(\epsilon \|X\|)$, где $o(\epsilon \|X\|)/(\epsilon \|X\|) \rightarrow 0$ при $\epsilon \|X\| \rightarrow 0$. Поделив правую и левую части на ϵ , воспользовавшись положительной однородностью и переходя к пределу при $\epsilon \rightarrow 0$, имеем $F(X, s) = D_X F(0, s)X$. Предложение доказано.

Чтобы не исключать из рассмотрения нелинейные модели, условие дифференцируемости оператора $F(X, s)$ по X в точке $X = 0$ в дальнейшем не будет использоваться. Предположим, что $F(X, s)$ непрерывно дифференцируем по X и s на множестве $R_+^m \times [0, \infty)$. Класс таких функций достаточно богат, и включение его в рассмотрение значительно расширяет область использования результатов данной главы. Приведем пример монотонной и положительно однородной нелинейной функции, удовлетворяю-

щей условию (2.1.6):

$$x = \sqrt{x^2 + y^2} \frac{s^p}{s^p + k_1},$$

$$y = x^{1-\alpha} y^\alpha \frac{s^r}{s^r + k_2} \quad (p, \alpha, r, k_1, k_2 > 0).$$

Теорема 2.1.6. Если функции $F(X, s)$ и $\varphi(X)$ непрерывно дифференцируемы в окрестности стационарной точки $X_0, s_0 = \varphi(X)$ уравнения (2.1.2) и выполнены условия теоремы 2.1.5, то стационарное решение асимптотически устойчиво по Ляпунову в $R_+^m \setminus 0$.

Доказательство. Вначале докажем следующий технический результат.

Лемма 2.1.6.1. Если оператор $F(X)$ квазимонотонен, положительно однороден и неразложим, то его собственный вектор H , отвечающий собственному значению λ , является собственным вектором неразложимой матрицы с неотрицательными недиагональными элементами $D_X F(H)$ и λ — собственное значение этой матрицы с максимальной вещественной частью.

Доказательство. Выпишем тождество

$$F(H + X) - F(H) = D_X F(H)X + o(\|X\|),$$

где $o(\|X\|)/\|X\| \rightarrow 0$ при $\|X\| \rightarrow 0$. При $X = \epsilon H$ имеем $\epsilon \lambda = \epsilon D_X F(H)H + o(\epsilon \|H\|)$. Разделив на ϵ и перейдя к пределу при $\epsilon \rightarrow 0$, получаем, что H — собственный вектор матрицы $D_X F(H)$. В силу неразложимости этой матрицы H является единственным положительным собственным вектором. Лемма доказана.

Перейдем к доказательству теоремы. В силу теоремы 2.1.5 достаточно ограничиться констатацией устойчивости "в малом".

1. Пусть выполнено условие (2.1.11). Матрица линейного приближения имеет вид

$$D = p(s_0)D_X F(X_0) - \nu I + \frac{\partial p}{\partial s}(s_0)F(X_0) \left(\frac{\partial \varphi}{\partial X} \right)^* (X_0). \quad (2.1.15)$$

Воспользовавшись леммой 2.1.6.1, получаем

$$D = A - aX_0 C^*,$$

где матрица $A = p(s_0)D_X F(X_0) - \nu I$ имеет одно нулевое собственное значение, а остальная часть спектра лежит в левой полуплоскости, $a = -\partial p/\partial s(s_0) > 0$, X_0 — собственный вектор матрицы A и $c = \partial \varphi/\partial X(X_0)$. Вычислим характеристический многочлен матрицы D . Для этого воспользуемся тождеством

$$\det[\lambda I - A + aX_0 C^*] = \det \begin{bmatrix} \lambda I - A & aX_0 \\ -c & 1 \end{bmatrix},$$

откуда $(\lambda + aC^*X_0) \det(\lambda I - A)/\lambda = 0$. Следовательно, матрица D гурвицева.

2. Пусть выполнено условие (2.1.12). Матрица линейного приближения имеет вид

$$D = A - QC^*, \quad (2.1.16)$$

где $A = D_X F(X_0, s_0) - \nu I$, $Q = -\partial F/\partial s(X_0, s_0) \not\equiv 0$ и $c = \partial \varphi/\partial X(X_0) \not\equiv 0$. По предположению матрица QC^* имеет одну отличную от нуля компоненту, равную C^*Q . Следовательно, матрица D имеет неотрицательные недиагональные элементы и совпадает с матрицей A по всем элементам, кроме одного диагонального, который меньше, чем у матрицы A . Поэтому [3] ее собственное значение с максимальной вещественной частью меньше $\lambda(A) = 0$.

3. Пусть выполнено условие (2.1.13). Разложим правую и левую части тождества $\langle \partial \varphi/\partial X, F(X, s_0) \rangle = \lambda_{s_0} \varphi(X)$ по степеням X в окрестности точки X_0 . Приравнявая

члены разложения при одинаковых степенях, получаем

$$\left\langle \frac{\partial \varphi}{\partial X}(X_0), D_X F(X_0)X \right\rangle = \lambda_{s_0} \left\langle \frac{\partial \varphi}{\partial X}(X_0)X \right\rangle$$

при любых $X \in R_+^m$. Отсюда и из леммы 2.1.6.1 следует, что $C = \partial \varphi / \partial X(X_0)$ – собственный вектор матрицы $(D_X F(X_0))^*$, отвечающий собственному значению с максимальной вещественной частью. Вычислив характеристический полином матрицы D по формуле

$$\det[\lambda I - D] = \begin{bmatrix} \lambda I & Q \\ -C & 1 \end{bmatrix}$$

и воспользовавшись формулой окаймленного определителя [18], имеем

$$\det[\lambda I - D] = (1 + C^*(\lambda I - A)^{-1}Q) \det(\lambda I - A) = 0.$$

Поскольку C – собственный вектор матрицы A , характеристическое уравнение имеет вид $(\lambda + C^*Q) \det(\lambda I - A) / \lambda = 0$. Следовательно, матрица D гурвицева. Теорема доказана.

В системе с внешними источниками стационарное решение при фиксированных v и u может быть не единственным. Тем не менее имеет место

Теорема 2.1.7. Предположим, что функции $F(X, s)$ и $\varphi(X)$ непрерывно дифференцируемы в окрестности стационарной точки $X_0, s_0 = \varphi(X_0)$ уравнения (2.1.3). Пусть в этой окрестности выполнено одно из условий: (2.1.11) и вектор внешних источников модели (2.1.3) B является собственным вектором оператора $F(X)$, (2.1.12) или (2.1.13); тогда стационарное решение уравнения (2.1.3) асимптотически устойчиво по Ляпунову.

Доказательство. 1. Пусть выполнено условие (2.1.11) и B – собственный вектор оператора $p(s)F(\cdot) = F(\cdot, s)$ ($0 < s < s_\infty$). Матрица линейного приближения задается формулой $A = QC^*$, где $Q = -aX_0$ и $a = \partial p / \partial s(s_0)$. Матрица A в этой формуле гурвицева, так как ее собственное значение с максимальной вещественной частью равно $\lambda_{s_0} - v < 0$ при $B \neq 0$ (см. теорему 2.1.3).

Вычисляя характеристический полином, имеем

$$\det[\lambda I - D] = (\lambda + v - \lambda_{s_0} + aC^*X_0) \det(\lambda I - A) / (\lambda + v - \lambda_{s_0}).$$

Следовательно, матрица линейного приближения уравнения (1.2.5) гурвицева.

2. Пусть выполнено условие (2.1.12). Доказательство дословно совпадает с соответствующим пунктом теоремы 2.1.6, за исключением того, что A – гурвицева матрица (см. лемму 2.1.6.1 и теорему 1.1.3).

3. Пусть выполнено условие (2.1.13). Доказательство повторяет рассуждения соответствующего пункта теоремы 2.1.6, за исключением того, что матрица A гурвицева и ее собственное значение, соответствующее строке C^* , равно $\lambda_{s_0} - v < 0$. Отсюда следует, что характеристический полином матрицы линейного приближения уравнения (2.1.3) имеет вид

$$(\lambda + v - \lambda_{s_0}) \det(\lambda I - A) / (\lambda + v - \lambda_{s_0})$$

и, следовательно, эта матрица гурвицева. Теорема доказана.

В экспериментальных работах еще не проводилась классификация различных типов лимитирования. Однако в сложной экологической системе далеко не всегда реализуется стационарный режим. Поэтому приведенные типы лимитирования наверняка не исчерпывают всего богатства встречающихся в природе возможностей. Нами разобраны наиболее характерные случаи. Возможность расширения типов лимитирования и проведения классификации случаев устойчивости остается открытой.

2.1.4. Асимптотическая ограниченность решений. В рассматриваемых моделях управляемых экологических систем стационарные режимы далеко не всегда являются устойчивыми и тем более притягивающими множествами для всех начальных состояний системы. Поэтому самостоятельный интерес представляет вопрос о том, в каких случаях численность экологической системы независимо от начальных условий со временем попадает в некоторый ограниченный диапазон, а также вопрос о существовании ограниченного множества в фазовом пространстве системы, в котором траектории остаются во все моменты времени. В отличие от обычной для подобных задач ситуации, особенность наших рассмотрений состоит в том, что динамические системы, задаваемые уравнениями (2.1.2) и (2.1.3), определены в положительном конусе евклидова пространства и при положительном времени не покидают его. Кроме того, необходимо исключить из притягивающей области тривиальную точку нулевой численности, установив тем самым факт невырождения биологической системы. Для уравнения (2.1.2) точка $X = 0$ является стационарным решением. Условия, при которых она является отталкивающей, содержит

Теорема 2.1.9. Если выполнены условия теоремы 2.1.1 о существовании нетривиального стационарного решения уравнения (2.1.2), то асимптотически устойчивое множество ω -предельных точек [23] решений этого уравнения, начинающихся в $R_+^m \setminus \{0\}$, не содержит начала координат.

Доказательство. Будем рассуждать от противного. Пусть $X(t)$ – решение уравнения (2.1.2), для которого точка 0 является ω -предельной. Поскольку

$$(F(X, s))_i \geq (F(0, \dots, x_i, \dots, 0, s))_i \geq x_i (F(0, \dots, 1, \dots, 0, s))_i,$$

то $\dot{X} \geq \text{const} \cdot X(t)$ при $\|X(t)\| \leq \epsilon$. Поэтому $X(t)$ приближается к началу координат не быстрее, чем экспонента. Следовательно, существует сколь угодно большой промежуток времени, в течение которого $\|X(t)\| \leq \epsilon$.

Пусть $s = \varphi(X) \leq \delta$ при $\|X\| \leq \epsilon$. В силу технической леммы 2 (приложение 1) операторы семейства $\{F_s(X) | 0 < s \leq \delta\}$ можно считать положительными. Отсюда по непрерывности имеем $f_i(X, s) \geq (1 - \alpha)f_i(X, 0)$ при $s \leq \delta$, $\rho(X) \leq 1$ и $1 < i \leq m$. Следовательно, $\dot{X}(t) = F(X(t), \varphi(X(t))) - vX(t) \geq (1 - \alpha)F(X(t), 0) - vX(t) = = F(X(t))$.

В силу предложения 2.1.2 и условия $\lambda_0 - v > 0$ собственное значение μ оператора $\Phi(\cdot)$ положительно при достаточно малом α . Так как правая часть дифференциального неравенства квазимоноotonно возрастающая, то по одному из вариантов леммы сравнения [33] имеем $X(t) \geq Y(t)$, где $\dot{Y}(t) = \Phi(Y(t))$ при $t \in [t_1, t_2]$. В то же время $Y(t) = e^{\mu t}(H + o(1)) \rightarrow \infty$ при $t_2 - t_1 \rightarrow \infty$, где H – положительный собственный вектор оператора $\Phi(X)$. Полученный результат противоречит условию $\|X(t)\| \leq \epsilon$ при $t \in [t_1, t_2]$. Таким образом, 0 не является предельной точкой решения $X(t) \geq 0$. Теорема доказана.

Для того чтобы доказать существование притягивающего ограниченного множества в фазовом пространстве системы (2.1.2), приходится предполагать конечность значения лимитирующего фактора, при котором наступает вырождение системы. Кроме того, уровень лимитирования должен формироваться всеми компонентами вектора состояния системы. Если же это условие не выполняется, то ограниченность решений гарантировать невозможно. Исключение составляет случай, когда имеется некоторый "критический" уровень лимитирования, при котором прекращается воспроизводство всех компонент экологической системы. Более того, замедление процессов роста вблизи "критического" значения лимитирующего фактора

должно происходить с одинаковой скоростью во всех компонентах системы. Формализацией последнего свойства является

Определение. Оператор F_s равномерно вырождается при $s \rightarrow s_\infty$, если $F_{s_\infty} \leq 0$, и существует непрерывная неотрицательная функция $f(s)$ такая, что оператор $F_s/f(s)$ определен, непрерывен и неразложим при всех $s \leq s_\infty$, включая s_∞ .

Приведем результат об асимптотической ограниченности или диссипативности решений [29] системы (2.1.2).

Теорема 2.1.10. Если s_∞ – конечное число, множество $\{X \in R_+^m \mid \varphi(X) \leq s_\infty\}$ асимптотически устойчиво и выполнено одно из условий:

а) $\varphi(X) > 0$ при $X \neq 0$,

б) оператор F_s равномерно вырождается и $v > 0$, то система (2.1.2) диссипативна.

Если выполнено условие б) и $v = 0$, то любое решение уравнения (2.1.2) остается ограниченным при всех $t \geq 0$.

Доказательство. 1. Пусть выполнено условие а). Доказательство следует из ограниченности множества $\{X \in R_+^m \mid \varphi(X) \leq s_\infty\}$.

2. Пусть выполнено условие б) и $v \equiv 0$. Предположим, что $\varphi(X) > 0$ при $x_i > 0$ ($i = 1, \dots, k < m$) и $\varphi(X) = 0$ при $x_i = 0$ ($i \leq k$) и $x_j > 0$ ($m \geq j > k + 1$). Тогда компоненты $x_1(t), \dots, x_k(t)$ вектора $X(t)$ принадлежат ограниченному множеству. Пусть $\max\{x_j(t) \mid j > k\}$ при $t \geq 0$ принимает сколь угодно большие значения. Следовательно, $\rho(X(t))$ и $\ln \rho(X(t))$ также принимают сколь угодно большие значения. Пусть $\ln \rho(X(t_1)) = N$, $N \leq \ln \rho(X(t))$ при $t_1 \leq t \leq t_2$ и $\ln \rho(X(t_2)) = N^2$.

Из уравнения (2.1.2) имеем

$$\ln(\rho(X(t_2))/\rho(X(t_1))) = \int_{t_1}^{t_2} \frac{\rho(F(X(t), s(t)))}{\rho(X(t))f(s(t))} f(s(t)) dt \leq a \int_{t_1}^{t_2} f(s(t)) dt,$$

где $s(t) = \varphi(X(t))$ и произвольная положительная константа

$a \geq \sup\{\rho(F(X, s))/f(s) \mid \rho(X) \leq 1, s \leq s_\infty\}$. Отсюда заключаем, что $\int_{t_1}^{t_2} f(s(t)) dt \geq N/a$.

Покажем теперь, что при достаточно больших N условие ограниченности первых k компонент вектора $X(t)$ нарушается. Пусть $C^* = \underbrace{(1, \dots, 1, 0, \dots, 0)}_k$ и

$x_j(t) = \max\{x_i(t) \mid k + 1 \leq i \leq m\}$ (индекс $j(t)$ зависит от t). Так как $\sum_{i=1}^m x_i(t) \geq e^N$, то $x_j(t) \geq e^N/m \geq N$ при $t \in [t_1, t_2]$ и достаточно большом N . Оценим величину $C^*X(t)$. Из формулы (2.1.2) имеем

$$\begin{aligned} C^* \dot{X} &\geq x_j(t) \sum_{i=1}^k f_i(0, \dots, x_i/x_j(t), \dots, 0, 1, \dots, 0, s(t)) \geq \\ &\geq \frac{1}{2} N f(s(t)) \sum_{i=1}^k f_i(0, \dots, 1, \dots, 0, s(t))/f(s(t)) \geq \frac{1}{2} b N f(s(t)), \end{aligned}$$

где множитель $1/2$ появился при замене x_i/x_j на нуль, а величина $b = \inf\{\sum_{i=1}^k f_i(0, \dots, 1, \dots, 0, s)/f(s) \mid 0 \leq s \leq s_\infty\}$ строго положительна ввиду условия равномерной неразложимости. Пронтегрировав последнее неравенство и воспользовавшись усло-

вием $\int_{t_1}^{t_2} f(s(t)) dt \geq N/a$, имеем

$$C^*X(t_2) - C^*X(t_1) \geq 0,5abN^2.$$

Так как $C^*X(t_1)$ ограничено при $\varphi(X) \leq s_\infty$, получили противоречие. Ограниченность решений доказана.

3. Пусть выполнено условие б) и $v \neq 0$. Проводя те же рассуждения, что и в пункте а), имеем

$$a \int_{t_1}^{t_2} f(s(t)) dt \geq v(t_2 - t_1) + \ln \rho(X(t_2))/\rho(X(t_1))$$

для любого решения $X(t)$ уравнения (2.1.2), если $X(t_1) \neq 0$. По теореме 2.1.9 величина $\rho(X(t_2))$ не может принимать сколь угодно малые значения. Поэтому

$\int_{t_1}^{t_2} f(s(t)) dt \rightarrow \infty$ при $t_2 - t_1 \rightarrow \infty$ и $X(t_1) \neq 0$. Как и в предыдущем случае, обозначим $x_{j_i}(t) = \max\{x_i(t) \mid k \leq i \leq m\}$ и выберем N настолько большим, что $\sum_{i=1}^k f_i(0, \dots,$

$\dots, x_i/N, \dots, 0, \dots, 1, \dots, 0, s) \geq 0,5 \sum_{i=1}^k f_i(0, \dots, 1, \dots, 0, s)$ при $\varphi(0, \dots, x_i, \dots,$

$\dots, 0) \leq s_\infty$ ($i = 1, \dots, k$) и $s \leq s_\infty$. Докажем теперь, что всякое решение может находиться вне ограниченного множества $\{X \in R_+^m \mid 0 \leq x_i \leq N \ (i = 1, \dots, m)\}$ лишь конечное время. Допустим противное. Тогда найдутся такие моменты времени t_1 и t_2 , что $t_2 - t_1$ сколь угодно велико и $\max\{x_i(t) \mid k+1 \leq i \leq m, t \in [t_1, t_2]\} \geq N$. Про-

интегрировав неравенство для $d/dt C^*X$ и воспользовавшись тем, что $\int_{t_1}^{t_2} f(s(t)) dt \rightarrow \infty$

при $t_2 - t_1 \rightarrow \infty$, получаем неограниченность первых k компонент вектора $X(t)$. Это противоречие завершает доказательство теоремы.

В системе с внешними источниками точка $X = 0$, очевидно, не является стационарным решением. Все решения покидают ее окрестность за конечное время, и поэтому наша задача ограничивается доказательством диссипативности решений уравнений (2.1.3). Решение этой задачи достигается почти при тех же условиях, что и в предыдущей теореме.

Теорема 2.1.11. Если $v > 0$ и выполнены условия теоремы 2.1.10, то все решения системы (2.1.3) при достаточно больших t попадают в ограниченное множество и не покидают его.

Доказательство. 1. Пусть выполнено условие а) теоремы 2.1.10. Если X лежит вне ограниченного множества $\varphi(X) \leq s_\infty$, то

$$\frac{d}{dt} \rho(X) \leq -v\rho(X) + \rho(B)u < 0$$

при достаточно большом $\rho(X)$.

2. Пусть выполнено условие б) теоремы 2.1.10. Используя обозначения доказательства теоремы 2.1.10, из формулы (2.1.3) получаем

$$\frac{d}{dt} C^*X \leq -vC^*X + C^*Bu < 0$$

при достаточно больших C^*X . Следовательно, C^*X — ограниченная величина. Предположим теперь, что компоненты вектора $X(t)$, которые не входят в линейную форму C^*X , принимают сколь угодно большие значения. Если решение уравнения (2.1.3) находится вне ограниченного множества $\{X \in R_+^m \mid 0 \leq x_i \leq N \ (i = 1, \dots, m)\}$, построенного в пункте 3 теоремы 2.1.10, сколь угодно большое время, то формула

для C^*X по-прежнему верна, а вместо оценки $\int_{t_1}^{t_2} f(s) dt$ имеем

$$a \int_{t_1}^{t_2} f(s(t)) dt \geq v(t_2 - t_1) + \ln \rho(X(t_2))/\rho(X(t_1)) - \rho(B) \int_{t_1}^{t_2} \frac{dt}{\rho(X)}$$

Так как $\rho(X) \geq N$ и N может быть сколь угодно большой величиной, то $\int_{t_1}^{t_2} f(s(t)) dt \rightarrow \infty$ при $t_2 - t_1 \rightarrow \infty$. Повторяя рассуждения пункта 3 теоремы 2.1.10, завершаем доказательство теоремы.

Таким образом, мы нашли условия, при выполнении которых численность эксплуатируемой популяции независимо от начального состояния попадает в некоторый ограниченный диапазон, в котором она остается во все последующие моменты времени. Следствием этого свойства является существование ограниченного инвариантного множества в фазовом пространстве изучаемых систем [29]. Все решения, начинающиеся в $R^m \setminus \emptyset$, стремятся к нему при $t \rightarrow \infty$, а движения, начинающиеся вблизи этого множества, остаются в его малой окрестности во все положительные моменты времени. Это означает, что имеет место асимптотическая устойчивость инвариантного множества [29]. Если это множество содержится в шаре малого диаметра, то экологическую систему можно считать практически устойчивой.

2.1.5. Существование устойчивых стационарных и ограниченных режимов в системах непрерывного культивирования. Применим полученные результаты к исследованию моделей динамики численности микробных популяций (1.2.1) и (1.2.4). Легко видеть, что матрица A , задаваемая формулами (1.2.2) и (1.2.3), является неразложимой. Ее собственное значение с максимальной вещественной частью $\lambda(A)$ определяет асимптотическую скорость роста численности при нелимитированных питательных ресурсах и отсутствии оттока особей из популяции. Следовательно, неравенство $\lambda(A) > 0$ означает условие невырождения популяции. Для матрицы (1.2.2) это условие, очевидно [3], выполнено, если среднее число потомков, оставляемых каждым микроорганизмом, больше единицы, т.е. $q > 1$. Для матрицы (1.2.3) это условие выполнено всегда. В дальнейших рассуждениях конкретный вид матрицы фигурировать не будет, а приводимые ниже результаты верны для произвольной неразложимой матрицы с неотрицательными недиагональными элементами и $\lambda(A) > 0$. Мы будем предполагать также, что $\det A \neq 0$. Для матрицы $AP(0)$, напротив, будем считать выполненным условие $\lambda(AP(0)) \leq 0$. Это предположение мотивируется тем, что при отсутствии питательного субстрата популяция микроорганизмов теряет способность к размножению.

Система (1.2.1) обладает асимптотически устойчивой инвариантной гиперплоскостью $-G^*A^{-1}X + s = s_p$. Этот интеграл задает закон сохранения вещества и позволяет исключить из рассмотрения динамику субстрата. Поскольку знакопостоянство матрицы A^{-1} не гарантировано, проведем замену переменных $X = AY$. В результате на инвариантной гиперплоскости приходим к уравнению вида (2.1.2):

$$\dot{Y} = P(s_p - G^*Y)AY - vY. \quad (2.1.17)$$

Фазовое пространство этого уравнения образовано векторами $Y \geq 0$ и $AY \geq 0$ и, следовательно, содержится в фазовом пространстве исходной системы. Поэтому полный анализ качественного поведения динамической

системы (1.2.1) не может быть сведен к исследованию уравнения (2.1.17). Локальный анализ уравнения (2.1.17) в окрестности стационарной точки эквивалентен анализу исходной системы, поскольку стационарное решение системы (1.2.1) является, очевидно, внутренней точкой фазового пространства уравнения (2.1.17).

Сформулируем теоремы о существовании стационарного режима при хемостатном способе непрерывного культивирования, когда скорость разбавления постоянна.

Теорема 2.1.12. Для того чтобы система (1.2.1) имела нетривиальное решение, необходимо и достаточно, чтобы выполнялось неравенство

$$\lambda(AP(s_p)) > v. \quad (2.1.18)$$

Это решение единственно.

Доказательство. Стационарное решение уравнения (1.2.1) существует, если оно удовлетворяет системе

$$\begin{aligned} 0 &= AP(s_0)X_0 - vX_0, \\ s_p &= GA^{-1}X_0 + s_0. \end{aligned} \quad (2.1.19)$$

Поэтому $Y_0 = A^{-1}X_0 > 0$ и принадлежит области определения уравнения (2.1.17). По теоремам 2.1.1 и 2.1.2 условие (2.1.18) необходимо и достаточно для существования единственного решения уравнения (2.1.17). Теорема доказана.

Условие асимптотической эквивалентности системы (1.2.1) и уравнения (2.1.17) принимает вид:

Предложение 2.1.5. Если выполнено условие (2.1.18) и система (1.2.1) диссипативна, то любое нетривиальное решение системы (1.2.1), начинающееся на множестве $X \in R_+^m \setminus 0$, при достаточно больших t принадлежит области определения уравнения (2.1.17) и даже попадает во внутренность конуса

$$\{Y = A^{-1}X \geq 0 \mid X \in R_+^m\}.$$

Доказательство.

Лемма. Если выполнено условие (2.1.18), то $\|X(t)\| \geq \alpha > 0$ при достаточно больших t .

Доказательство леммы. Из второго уравнения системы (1.2.1) видно, что $s(t) \leq s_p + \epsilon$ (ϵ — сколь угодно малая константа) для достаточно больших t . Более того, если в начальный момент времени $s(0) \leq s_p$, то $s(t) \leq s_p$ для достаточно больших t . Обратимся теперь к переменной $X(t)$. Пусть L ($L > 0$) — собственный вектор матрицы A^* , отвечающий $\lambda(A)$ и такой, что $\lambda(A)L > G$. Умножив первое уравнение системы (1.2.1) на L^* и воспользовавшись неравенством $\lambda(A)L > G$, имеем

$$L^* \dot{X} = \lambda(A)L^*PX - vL^*X,$$

$$\dot{s} \geq -\lambda(A)L^*PX - v(s - s_p).$$

Отсюда получаем $d/dt(L^*X + s) + v(L^*X + s) \geq vs_p$. Интегрируя неравенство, имеем

$$L^*X + s \geq s_p + e^{-vt}(L^*X(0) + s(0) - s_p).$$

Докажем теперь, что существует малое число $\alpha > 0$ такое, что $\|X(t)\| \geq \alpha$ для достаточно больших t . Допустим противное, тогда для любого $\epsilon > 0$ найдется такое t_0 , что $L^*X(t_0) < \epsilon$. Поэтому на основании предыдущего неравенства можно счи-

тать, что $s > s_p - 2\epsilon$ для достаточно больших t_0 . Выберем ϵ из условия, что $\lambda(AP(s_p - 2\epsilon)) - v > 0$. В силу предположения (2.1.18) такое ϵ существует. Пусть $R > L > 0$ — собственный вектор матрицы $AP(s_p - 2\epsilon)$, отвечающий $\lambda(AP(s_p - 2\epsilon))$. Из первого уравнения системы (1.2.1) имеем $R^* \dot{X} \geq \geq RX(\lambda(AP(s_p - 2\epsilon)) - v) > 0$ при $L^*X < \epsilon$ и тем более при $R^*X < \epsilon$. Следовательно, $R^*X \geq \epsilon$, и так как $R > 0$, то $\|X(t)\| \geq \alpha$. Лемма доказана.

Теперь можно завершить доказательство предложения. Если система (1.2.1) диссипативна, то для достаточно больших t имеет место ограниченность $\|X(t)\| \leq M$. В этом случае из второго уравнения системы (1.2.1) следует, что $s(t) \geq \epsilon > 0$. Тогда $P(s(t)) \geq \alpha I$ ($\alpha > 0$). Умножим первое уравнение системы (1.2.1) на A^{-1} . Воспользовавшись обозначением $Y = A^{-1}X$, имеем

$$\dot{Y} = P(s)X - vY.$$

Отсюда получаем

$$\exp(vt)Y = \int_0^t \exp(v\tau)P(s)X d\tau + Y(0) \geq$$

$$\geq \alpha \int_0^t \exp(v\tau)X d\tau + Y(0).$$

Для того чтобы установить неравенство $Y(t) > 0$, достаточно показать, что

$$\int_0^t x_j(\tau) d\tau \rightarrow \infty \text{ при } t \rightarrow \infty \quad (j = 1, \dots, m). \text{ Допустим противное. Тогда } \int_0^t x_j(\tau) d\tau \leq M$$

при $1 \leq i \leq k$ и $\int_0^t x_j(\tau) d\tau \rightarrow \infty$ при $k+1 \leq j \leq m$ и $t \rightarrow \infty$. В силу леммы имеем

$k < m$. Воспользовавшись неразложимостью матрицы A , получаем для некоторого $1 \leq i \leq k$:

$$\dot{x}_i = \sum_{j=1}^m a_{ij}p_j(s)x_j - vx_i \geq a_{ij}p_j(s)x_j + (a_{ii}p(s) - v)x_i,$$

где $a_{ij} > 0$ ($k \leq j \leq m$). Интегрируя неравенство и воспользовавшись тем, что

$$p_j(s_p) \geq p_j(s) \geq \alpha, \quad \int_0^t x_j(\tau) d\tau \rightarrow \infty \text{ и } \int_0^t x_i d\tau < M \text{ при } t \rightarrow \infty, \text{ получаем противоречие.}$$

Предложение доказано.

Перейдем к рассмотрению модели динамики численности одноклеточных при многоступенчатом многопоточном культивировании.

Система (1.2.4) в фазовом пространстве $R_+^m \times R_+^1 \times \dots \times R_+^m \times R_+^1$ имеет асимптотически устойчивую интегральную гиперповерхность размерности mn , задаваемую соотношениями

$$G^*A^{-1}X^1 + s^1 = s_p^1, \tag{2.1.20}$$

$$G^*A^{-1}X^i + s^i = \frac{u_i}{v_i} s_p^{i-1} + \left(1 - \frac{u_i}{v_i}\right) s_p^i \quad (i = 2, \dots, n).$$

Сделаем замену переменной $X^i = AY^i$ ($i = 1, \dots, n$). Тогда на инвариантной гиперповерхности система (1.2.4) имеет вид

$$\begin{aligned} \dot{Y}^1 &= P(s_p^1 - G^* Y^1)AY^1 - vY^1, \\ Y^i &= P\left(\frac{u_i}{v_i} s_p^{i-1} + \left(1 - \frac{u_i}{v_i}\right)s_p^i - G^* Y^i\right)AY^i - \\ &- v_i Y^i + u_i Y^{i-1} \quad (i = 2, \dots, n). \end{aligned} \quad (2.1.21)$$

Фазовое пространство системы (2.1.21) образовано векторами $Y^i \geq 0$ и $AY^i \geq 0$ ($i = 1, \dots, n$) и является подконусом исходного фазового пространства. Стационарное решение системы (1.2.4), если оно существует, удовлетворяет неравенствам

$$\begin{aligned} Y^1 &= A^{-1}X^1 = X^1/v_1 > 0, \\ Y^i &= A^{-1}X = (P(s^i)X^i + u_i Y^{i-1})/v_i > 0 \quad (i = 2, \dots, n), \end{aligned} \quad (2.1.22)$$

т.е. является внутренней точкой этого подконуса. Следовательно, локальный анализ системы (2.1.21) в окрестности стационарной точки эквивалентен анализу исходной системы. Условие асимптотической эквивалентности систем (1.2.4) и (2.1.21) имеет вид:

Предложение 2.1.6. Если система (1.2.4) асимптотически ограничена (диссипативна) и ее первая подсистема имеет нетривиальное стационарное решение ($X^1 \neq 0, s^1 \neq s_p$), то любое решение системы (1.2.4) принадлежит области определения системы (2.1.21) при достаточно больших t .

Доказательство.

Лемма. Если для первой подсистемы системы (1.2.4) выполнено условие (2.1.18), то $\|X^i(t)\| \geq \alpha > 0$ при $i = 1, \dots, n$ и достаточно больших t .

Проведем доказательство леммы для $i = 2$. Из второго и четвертого уравнений

системы (1.2.4) следует, что $s^2(t) \leq \frac{u_2}{v_2} s_p^1 + \left(1 - \frac{u_2}{v_2}\right)s_p^2 + \epsilon$ при достаточно больших t .

Воспользовавшись обозначениями предложения 2.1.5 из третьего уравнения системы (1.2.4), имеем

$$L^* \dot{X}^2 = \lambda(A)L^*PX^2 - v^2 L^*X^2 + u_2 L^*X^1.$$

Так как $L^*X^1 > \alpha_1 > 0$ на основании леммы из предложения 2.1.5 и s — ограниченная величина, имеем $L^*X^2 \geq \alpha_2 > 0$. Лемма доказана.

Завершим доказательство предложения для случая $i = 2$. Так же, как и в предложении 2.1.5, получаем $s^2(t) \geq \epsilon > 0$ при достаточно больших t . Из третьего уравнения системы (1.2.4) имеем

$$\dot{Y}^2 \geq \alpha X^2 - v_2 Y^2 + u_2 Y^1,$$

где $Y^i = A^{-1}X^i$ ($i = 1, 2$) и $Y^1 > 0$ в силу предложения 2.1.5. Рассуждая так же, как и в предложении 2.1.5, завершаем доказательство утверждения.

Сформулируем теорему о существовании стационарного режима при многопоточном культивировании. На вопрос о существовании стационарного режима в первом ферментере отвечает теорема 2.1.12. Для того чтобы определить стационарные значения переменных во всех последующих ступенях батареи, достаточно указать алгоритм решения задачи для второго ферментера и продолжить рассуждения по индукции. Пусть условие (2.1.18) выполнено при $i = 1$ для системы (1.2.4) и X^1 и s^1 — стационарные значения переменных первой подсистемы. Тогда при $i = 2$ имеем

$$\begin{aligned} \dot{X}^2 &= AP(s^2)X^2 - v_2 X^2 + u_2 X^1, \\ s^2 &= -G^*P(s^2)X^2 - v_2 s^2 + u_2 s^1 + (v_2 - u_2)s_p^2. \end{aligned} \quad (2.1.23)$$

Теорема 2.1.13. Существует такая функция $\omega(s, u): R_+^1 \times R_+^1 \rightarrow R_+^1$, что для любого $s^2 > 0$, любого $u_2 > 0$ и любого $v_2 > \omega(s^2, u_2)$ найдется такая единственная концентрация субстрата на входе во второй ферментер s_p^2 , что в этом ферментере реализуется стационарный режим с заданными s^2, u^2, v^2 .

Доказательство. Фиксируем s^2, u^2 и $v^2 \geq \max\{\lambda_{s^2}, u^2\}$. Тогда уравнение $0 = AP(s^2)X^2 - v_2 X^2 + u_2 X^1$

имеет единственное решение на основании технической леммы 2 (приложение 2) и [22, 24]. Подставим это решение во второе уравнение системы (2.1.23). Из этого уравнения и условия стационарности имеем

$$s_p^2 = (G^*P(s^2)X^2 + v^2 s^2 - u_2 s^1)/(v_2 - u_2).$$

Функция $\omega(u_2, s^2)$ должна удовлетворять условиям $\omega(u_2, s^2) \geq \lambda_{s^2}$, $\omega(u_2, s^2) \geq u_2$ и $s_p^2 \geq 0$ при $v_2 \geq \omega(u_2, s^2)$. Такая функция существует, так как при $v_2 \geq u_2 s^1/s^2$ последнее неравенство выполняется. Теорема доказана.

Обратимся теперь к изучению вопроса устойчивости стационарных режимов в культивационных системах. Предположим, что существует единственное нетривиальное стационарное решение системы (1.2.1). Тогда имеет место

Теорема 2.1.14. Пусть выполнено неравенство (2.1.18) и одно из условий:

- а) $P(s) = p(s)I$, где $p(s)$ — скалярная монотонно убывающая функция;
- б) $P(s_0)G$ — положительный собственный вектор матрицы $(AP(s_0))^*$, который отвечает собственному значению с максимальной вещественной частью;
- в)

$$P(s) = \begin{bmatrix} 1 & & & & 0 \\ & \ddots & & & \\ & & p_i(s) & & \\ & & & \ddots & \\ 0 & & & & 1 \end{bmatrix}, \quad G = \begin{bmatrix} 0 \\ \vdots \\ g_i \\ \vdots \\ 0 \end{bmatrix}.$$

Тогда стационарное решение системы (1.2.1) асимптотически устойчиво при $X \in R_+^m \setminus 0$ и $s \in R_+^1$.

Доказательство. Пусть выполнено условие а). Доказательство состоит в повторении рассуждений теорем 2.1.5 и 2.1.6, использовании леммы из предложения 2.1.5 и асимптотической устойчивости инвариантной гиперплоскости $G^*A^{-1}X + s = s_p$.

Пусть выполнено условие б). Обозначим $G_1^* = G^*P(s_0)$ и $P_1(s) = P(s)P^{-1}(s_0)$. Перейдем к новым обозначениям в системе (1.2.1). Тогда $P_1(s_0) = I$. В дальнейших рассуждениях опустим индекс "1" и будем считать, что $P(s_0) = I$ и $G^*A = \lambda(A)G^*$. Умножим первое уравнение в системе (1.2.1) на G^* и ограничим наше рассмотрение на асимптотически устойчивом инвариантном множестве $G^*A^{-1}X + s = s_p$. По предположению $G^*A^{-1} = G^*/\lambda(A)$. Отсюда $G^*P(s_p - G^*X/\lambda(A))X - vG^*X = G^*X$. Функция $G^*P(s_p - \sigma)X$ удовлетворяет условиям (2.1.13) по σ , и при функции предпочтения $\sigma = G^*X/\lambda(A)$ имеет место теорема 2.1.6. В силу асимптотической устойчивости гиперплоскости $G^*X/\lambda(A) + s = s_p$ результат верен для всех $X \geq 0$ и $s \geq 0$.

Пусть выполнено условие в). Покажем, что $Y = A^{-1}X(t) \geq 0$ для любого нетривиального решения при достаточно больших t . Для определенности предположим, что $p_1 = p_1(s)$. Умножим первое уравнение (1.2.1) на A^{-1} . Имеем

$$y_1 = p_1 x_1 - v y_1,$$

.....

$$y_m = x_m - v y_m.$$

Убедимся вначале, что $\int_{t_0}^t p_1 x_1 d\tau \rightarrow \infty$ при $t \rightarrow \infty$. Допустим противное. Интегрируя второе уравнение системы (1.2.1), имеем

$$s(t) = s(t_0) - g_1 \int_{t_0}^t p_1 x_1 d\tau - v \int_{t_0}^t s d\tau + v s_p(t - t_0).$$

В то же время $s(t) \leq s_p + ce^{-vt}$ ($c > 0$). Обозначим $T(t) = [t, 2t]$ и $\Delta(t) = \{\tau \in T(t) \mid s(\tau) \leq s_p - \delta\}$, где δ — малое число. Тогда мера $\mu(\Delta(t)) \rightarrow 0$ при $t \rightarrow \infty$. Если это не так, то

$$\begin{aligned} s_p t + \text{const} &\leq \int_t^{2t} s(t) dt \leq \mu(\Delta(t))(s_p - \delta) + \\ &+ \mu(T(t))(s_p + ce^{-vt}) = s_p t - \mu(\Delta(t))\delta + ce^{-vt}, \end{aligned}$$

и получаем противоречие.

Пусть L — собственный вектор матрицы $(AP(s_p - \delta))^*$, отвечающий собственному значению с максимальной вещественной частью. Обозначим $\alpha = \lambda(AP(s_p - \delta)) - v > 0$. Из первого уравнения системы (1.2.1) получаем, что

$$L^* \dot{X} \geq \alpha L^* X \quad \text{при } \tau \in T(t),$$

$$L^* \dot{X} \leq -v L^* X \quad \text{при } \tau \in \Delta(t).$$

Следовательно, $\ln(L^*X(2t)/L^*X(t)) \rightarrow \infty$ при $t \rightarrow \infty$. В силу леммы из предложения 2.1.5 $L^*X(t) \geq \epsilon > 0$ при достаточно больших t . Поэтому $L^*X(2t) \rightarrow \infty$ при $t \rightarrow \infty$. Используя неразложимость, малость $\mu(\Delta(t))$ и рассуждая так же, как и при доказательстве предложения 2.1.5, легко показать, что $x_i(2t) \rightarrow \infty$ при $i = 1, \dots, m$ и $t \rightarrow \infty$.

Следовательно, $\int_0^{2t} p_1 x_1 d\tau \geq \int_0^t p_1 x_1 d\tau \rightarrow \infty$ при $t \rightarrow \infty$. Из полученного противоречия

закключаем, что $\int_0^t p_1 x_1 d\tau \rightarrow \infty$ при $t \rightarrow \infty$. Опять повторяя рассуждения предложе-

ния 2.1.5, докажем, что $y_i \geq 0$ при $i = 1, \dots, m$ и достаточно больших t . Устойчивость следует из формулы (2.1.17) и теоремы 2.1.5. Теорема доказана.

Таким образом, при равномерном лимитировании, автолимитировании и интегральном лимитировании стационарный режим культивирования в хемостате является устойчивым.

В многопоточной системе стационарное решение при фиксированных v_i и u_i ($i = 1, \dots, n$), вообще говоря, может быть не единственным. Однако верна

Теорема 2.1.15. Пусть система (1.2.4) имеет нетривиальное стационарное решение X_0^i, s_0^i ($i = 1, \dots, n$) ($X^1 \neq 0$). Если в некоторой окрестности стационарной точки выполнено условие а) или в) теоремы 2.1.14, то это решение асимптотически устойчиво.

Доказательство следует из теоремы 2.1.7 и формул (2.1.20) и (2.1.21).

Таким образом, мы показали, что условия равномерного лимитирования и автолимитирования достаточны для устойчивости стационарных режимов как при одноступенчатом, так и при многоступенчатом культивировании.

В случае, когда устойчивость стационарного режима при непрерывном культивировании гарантировать не удастся, приходится удовлетворяться доказательством существования устойчивого ограниченного множества, в которое со временем попадают переменные состояния системы. При этом желательно исключить точку вырождения популяции. По аналогии с теоремой 2.1.9 имеет место

Теорема 2.1.16. Если выполнено условие (2.1.18) теоремы 2.1.12 о существовании нетривиального стационарного решения в системе (1.2.1) и в начальный момент времени $X(0) \neq 0$, то суммарная численность популяции $\sum_{i=1}^m x_i(t)$ остается больше некоторого положительного числа при достаточно больших t . Это число не зависит от начальной численности популяции и начальной концентрации субстрата.

Доказательство содержится в лемме из предложения 2.1.5.

Результат о диссипативности системы (1.2.1) дает следующая

Теорема 2.1.17. Пусть выполнено одно из условий:

а) $G > 0$;

б) $G^*A^{-1} > 0$;

в) существует функция $p(s)$, такая, что $p(s) > 0$ при $s > 0$ и $p(0) = 0$, и такая, что $0 \neq \lim_{s \rightarrow 0} p_i(s)/p(s) < \infty$ ($i = 1, \dots, m$).

Для существования ограниченного и асимптотически устойчивого в R_+^{m+1} множества системы (1.2.1) достаточно выполнения условия а) и б). Для диссипативности на положительно инвариантном множестве $A^{-1}X \geq 0$ достаточно выполнения условия в).

В случае многопоточного культивирования достаточно рассмотреть двухступенчатую систему и доказать для второй ступени диссипативность, считая, что для первой ступени диссипативность доказана.

Доказательство. Пусть выполнено условие а). Обозначим через L ($\lambda(A)L < G$) собственный вектор матрицы A , отвечающий $\lambda(A)$. Из системы (1.2.1) имеем

$$L^* \dot{X} = \lambda(A)L^*P(s)X - vX,$$

$$s \leq \lambda(A)L^*P(s)X - vs + vs_p.$$

Отсюда получаем $L^*X + s \leq s_p + e^{-vt}(L^*X(0) + s(0) - s_p)$. Так как $X \geq 0$ и $s \geq 0$, утверждение доказано.

Пусть выполнено условие б). Поскольку $G^*A^{-1}X(t) + s(t) = s_p + e^{-vt}(G^*A^{-1}X(0) + s(0) - s_p)$, утверждение доказано.

Пусть выполнено условие в). Результат следует из формулы (2.1.17) и теоремы 2.1.10.

Теорема 2.1.18. Если выполнены условия а) и б) теоремы 2.1.17, то система (1.2.4) диссипативна в R_+^{m+1} . Если выполнено условие в), то система диссипативна на положительно инвариантном множестве $A^{-1}X$.

Доказательство. Ход рассуждений тот же, что и при доказательстве теоремы 2.1.17.

В качестве примера применим теоремы об асимптотической ограниченности к моделям динамики численности делящихся и почкующихся клеток. Обратная матрица в случае (1.2.2) и при выполнении условия невырожденная $q > 1$ строго положительна:

$$A^{-1} = \frac{1}{1-q} \begin{vmatrix} 1 & q & \dots & q \\ 1 & 1 & \dots & \dots \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \vdots & q \\ 1 & 1 & \dots & 1 \end{vmatrix} \quad (2.1.24)$$

Поэтому пункт б) в теореме 2.1.17 всегда выполняется.

В случае (1.2.3) обратная матрица при $m \geq 3$ не является знакоопределенной:

$$A^{-1} = -\frac{1}{m-1} \begin{vmatrix} 1 & 2 & \dots & m \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 2 & \dots & m \end{vmatrix} + \frac{2}{m-1} \begin{vmatrix} 1 & \dots & 1 \\ \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots \\ 1 & \dots & 1 \end{vmatrix} + \begin{vmatrix} 0 & 1 & \dots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \vdots & 1 \\ 0 & \dots & \dots & 0 \end{vmatrix}. \quad (2.1.25)$$

Поэтому условие б) в теореме 2.1.17 имеет вид

$$g_1 > 0,$$

$$g_1 + g_2 > \left(\sum_{i=3}^m g_i \right) / (m - 2),$$

$$g_1 + g_2 + g_3 > 2 \left(\sum_{i=4}^m g_i \right) / (m - 3), \tag{2.1.26}$$

.....

$$g_1 + g_2 + \dots + g_{m-1} > (m - 2)g_m.$$

2.1.6. Многофакторное лимитирование. В предыдущих пунктах были рассмотрены модели, соответствующие случаю однофакторного лимитирования. Эти модели иллюстрируют все основные свойства динамики численности экологических систем, которые характерны как при однофакторном, так и при многофакторном лимитировании. Для того чтобы в этом убедиться, приведем модель динамики численности экосистемы, подверженной одновременному влиянию многих лимитирующих факторов.

Пусть динамика численности эксплуатируемой экосистемы зависит от n лимитирующих факторов, оказывающих совокупное влияние на нее так, что нельзя выделить главный или ведущий фактор. Обозначим через s_i величину i -го лимитирующего фактора. По аналогии с предыдущим считаем, что эта величина определяется размерами экологической системы по формуле $s_i = \varphi_i(X)$, где функции $\varphi_i(X)$ ($i = 1, \dots, n$) монотонны и положительно однородны, т.е. $\varphi_i(X) < \varphi_i(Y)$ при $X < Y$ ($i = 1, \dots, n$) и $\varphi_i(\alpha X) = \alpha \varphi_i(X)$ при $\alpha > 0$ ($i = 1, \dots, n$). Введем в рассмотрение вектор $S = (s_1, \dots, s_n)^*$. Скорость воспроизводства экологической системы определяется функцией $F(X, S)$. Эта функция при каждом фиксированном $S \geq 0$ как функция первого аргумента по-прежнему задает динамическую систему с инвариантным положительным конусом, квазимоноотонна, положительно однородна и при S , принадлежащем порядковому интервалу $0 \leq S \leq S_\infty$, где S_∞ – вектор, имеющий либо положительные, либо бесконечные компоненты, неразложима (см. п. 2.1.2). Если величина хотя бы одного лимитирующего фактора больше критического, то экологическая система теряет способность к воспроизводству, т.е. при каждом фиксированном $S \notin S_\infty$ все решения дифференциального уравнения $\dot{X} = F(X, S)$, начинающиеся в положительном конусе, ограничены и не выходят из этого конуса.

Таким образом, динамика численности экосистемы определяется следующим векторным уравнением:

$$\dot{X} = F(X, \Phi(X)) - \nu X, \tag{2.1.27}$$

где $\Phi(X) = (\varphi_i(X))_{i=1}^n$.

По-прежнему будем считать, что увеличение уровня лимитирования понижает скорость воспроизводства в точке сбалансированного роста (см. формулу (2.1.6)):

$$F_{S_2}(H_{S_1}) \leq F_{S_1}(H_{S_1}) \quad \text{при } S_1 < S_2 < S_\infty,$$

где H_{S_1} – собственный вектор оператора $F(\cdot, S_1)$, отвечающий собственному значению λ_{S_1} . Легко видеть, что предложение 2.1.3, в котором скалярные значения лимитирующего фактора следует заменить на векторные, по-прежнему имеет место.

Сформулируем теперь теоремы о стационарных решениях.

Теорема 2.1.19. Для того чтобы система (2.1.27) имела нетривиальное стационарное решение X ($X \neq 0$ и $\Phi(X) < S_\infty$), необходимо выполнение неравенства

$$\lambda_0 > v. \tag{2.1.28}$$

Доказательство состоит в дословном повторении рассуждений теоремы 2.1.1 с заменой скалярной величины лимитирующего фактора на векторную.

Содержательная трактовка условия (2.1.28) в точности совпадает с трактовкой условия (2.1.10). Требуется, чтобы при отсутствии лимитирования асимптотическая скорость воспроизводства превосходила скорость изъятия. Нарушение этого условия приводит к тому, что экологическая система вырождается и единственным стационарным решением оказывается нулевой вектор состояния.

При многофакторном лимитировании не выполняется свойство единственности стационарного решения системы (2.1.27). Для доказательства существования этого решения требуется определить дополнительное свойство функции, связывающей уровень лимитирования с численностью экологической системы. Оно состоит в том, что набор лимитирующих факторов представлен в модели достаточно полно. Другими словами, это означает, что каждая компонента экологической системы влияет на величину одного или нескольких лимитирующих факторов. В связи с этим в дальнейшем будем предполагать, что при ограниченных значениях лимитирующих факторов численность популяции или сообщества ограничена, т.е. $\{X \in R_+^m \mid \|\Phi(X)\| \leq 1\}$ – ограниченное множество.

Теорема 2.1.20. Если выполнено неравенство (2.1.28), то при каждом фиксированном $0 < v < \lambda_0$ существует стационарное решение уравнения (2.1.27) со строго положительными компонентами. Если, кроме того, $\lambda_S < 0$ при некотором $S < S_\infty$, то нетривиальное решение существует при $v = 0$.

Доказательство. Поскольку все собственные значения оператора не больше нуля при $S \notin S_\infty$, по непрерывности можно считать, что $\lambda_{S_*} - v < 0$ при некотором конечном значении вектора $S_* < S_\infty$ и $\lambda_S - v < 0$ при $S \notin S_*$. Пусть M – такой шар в R^m , что

$$\Phi^{-1}([0, S_*]) \subset M.$$

В силу ограниченности множества $\{S \mid \Phi^{-1}(S) \leq S_*\}$ он существует. В соответствии с технической леммой 2 оператор $F^+(X, S) = F(X, S) + (N - v)I$ монотонен при каждом фиксированном значении $S \leq S_*$, если N — достаточно большое число. Решение уравнения $F(X, \Phi(X)) - vX = 0$ соответствует отысканию неподвижной точки оператора $F^+(X, \Phi(X))/N$, оставляющего инвариантным положительный конус R_+^m . Перейдем к изучению этого оператора. Поскольку требуется доказать существование отличной от 0 неподвижной точки полученного отображения, воспользуемся усиленным вариантом теоремы о неподвижной точке [41, 49], опирающимся на понятие изгоняющей неподвижной точки. Напомним определение этого понятия. Неподвижная точка Y_0 непрерывного отображения Γ некоторого банахова пространства в себя называется изгоняющей [41, 49, 50], если существует окрестность $U \ni Y_0$ такая, что для любой точки $Y \neq Y_0$ и $Y \in U$ найдется целое $m(Y)$, при котором $\Gamma^{m(Y)}Y \notin U$. Нами будет использоваться следующая

Теорема [49, 50]. Пусть M — замкнутое выпуклое множество, $\Gamma: M \rightarrow M$ — непрерывное отображение и Y_0 — изгоняющая неподвижная точка. Пусть существует такое число r , что для любой точки $X \in M$ и $\|X\| = r$ равенство $\Gamma X = tX$ может иметь место лишь при $t < 1$. Тогда Γ имеет неподвижную точку $X_0 \neq Y_0$, $X_0 \in M$ и $\|X_0\| < r$.

Воспользуемся этой теоремой. В силу условия (2.1.28) собственное значение оператора $F^+(\cdot, 0)$ больше единицы. Поэтому, рассуждая от противного и строя в некоторой малой окрестности начала координат миноранту для этого оператора, не зависящую от S , легко показать, что O — изгоняющая неподвижная точка. Очевидно, что $\Phi(X) \notin S_*$ на границе шара M и, следовательно, все собственные значения оператора $F^+(X, \Phi(X))$ при X , принадлежащих границе M , меньше единицы. По цитированной теореме существует неподвижная точка $X_v \neq 0$ оператора $F^+(X, \Phi(X))$, лежащая внутри шара M .

Величина $\Phi(X_v) < S_\infty$, так как в противном случае все собственные значения оператора $F(\cdot, \Phi(X_v))$ неположительны и условие $F(X_v, \Phi(X_v)) = vX$ не выполняется при $X = X_v$. При $S < S_\infty$ оператор $F^+(\cdot, S)$ неразложим и его единственное собственное значение X_v имеет строго положительные компоненты. Теорема доказана.

Перейдем к рассмотрению экологической системы, пополняющей свои запасы из внешних источников. Включая в модель (2.1.27) внешние источники, имеем

$$\dot{X} = F(X, \Phi(X)) - vX + Bu, \quad (2.1.29)$$

где вектор B определяет структуру внешнего притока, а u — его интенсивность. Стационарное решение этого уравнения может быть не единственным при фиксированных значениях параметров v и u . Существование этого решения гарантирует

Теорема 2.1.21. Если функции $\varphi_i(X)$ ($i = 1, \dots, n$), образующие оператор $\Phi(X) = (\varphi_i(X))_{i=1}^n$, выпуклы (например, линейны), то при любых $S < S_\infty$ и $v > \lambda_S$ найдется такое значение управляющего параметра u_0 , что уравнение (2.1.29) имеет стационарное решение X с положительными компонентами при всех $u \leq u_0$ и уровень лимитирования в стационарной точке не превосходит S , т.е. $\Phi(X) \leq S_\infty$.

Доказательство. Оно состоит в отыскании неподвижной точки непрерывного отображения конуса R_+^m в себя. Пусть выполнено условие $\lambda_{S_0} - v = -\epsilon$. В силу предположения 2.1.3 $\lambda_S - v \geq -\epsilon$ при $S = \Phi(X)$ и $\Phi(X) \leq S_0$. В силу технической леммы 2 оператор $F^+(\cdot, S) = (F(\cdot, S) + (N - v)I)$ монотонен и положителен при каждом фиксированном S из компактного множества $\Gamma = \{X \in R_+^m \mid \Phi(X) \leq S_0\}$, если N достаточно велико. Стационарное решение уравнения (2.1.29) является неподвиж-

ной точкой оператора $F^* + Bu_*$, где $u_* = u/N$, если оно принадлежит множеству Γ . Воспользуемся теоремой о неподвижной точке, приведенной в доказательстве теоремы 2.1.20. Убедимся в том, что при достаточно малом u_* , $\lambda \geq 1$ и $X \in \{X \in R_+^m \mid \Phi(X) = S_0\}$ уравнение $F^*(X, \Phi(X)) + Bu_* = \lambda X$ не имеет решения. Предположим противное. Пусть $X_0 \in R_+^m$ и $\Phi(X_0) = S_0$. Собственное значение оператора $F^*(\cdot, \Phi(X_0))$ меньше единицы. Величину u_* можно задать сколь угодно малой. Поскольку X_0 является собственным вектором оператора $F^*(X, \Phi(X_0)) + \frac{u_*}{\| \Phi(X_0) \|} B \| \Phi(X) \|$, полученного малым возмущением из непрерывного монотонного неразложимого оператора $F^*(X, \Phi(X_0))$, его собственное значение также меньше единицы. Это противоречит предположению $\lambda \geq 1$. Ввиду выпуклости функций φ_i ($i = 1, \dots, m$) множество Γ выпукло. На основании теоремы о неподвижной точке, использованной при отыскании стационарного решения модели (2.1.27) в теореме 2.1.19, множество Γ может фигурировать в качестве шара радиуса r . Начало координат не является неподвижной точкой отображения $F^* + Bu_*$. Таким образом, неподвижная точка $X_0 \in \Gamma$ существует и $\Phi(X_0) > 0$. Из неразложимости оператора следует, что $X_0 > 0$.

Перейдем к рассмотрению условий устойчивости стационарных решений моделей (2.1.27) и (2.1.29).

При многофакторном лимитировании типа обратной связи, приводящей к устойчивости стационарных режимов в эксплуатируемых экологических системах, принимают следующий вид:

1. Равномерное лимитирование:

$$\begin{aligned} F(X, S) &= p(S)F(X), \\ p(S_1) &< p(S_2) \text{ при } S_1 > S_2. \end{aligned} \quad (2.1.30)$$

2. Автолимитирование:

$$\begin{aligned} F(X, S) &= (f_i(X, S))_{i=1}^m, \\ f_i(X, S) &= f_i(X, s_i) \text{ при } 1 \leq i \leq n \leq m, \\ s_i &= x_i \quad \text{при } 1 \leq i \leq n, \\ f_i(H_0, S_0) &< f_i(H_0, S) \text{ при } S > S_0, \end{aligned} \quad (2.1.31)$$

где H_0 — собственный вектор оператора $F(X, S_0)$.

3. Интегральное лимитирование:

Функция $\Phi(X) = (\varphi_i(X))_{i=1}^m$ непрерывно дифференцируема на $R_+^m \setminus 0$, и вспомогательные функции

$$h_i(X, S) = \langle \partial \varphi_i / \partial X, F(X, S) \rangle \quad (i = 1, \dots, n)$$

зависят от величины лимитирующего фактора так, что

$$h_i(X, S_0) = \lambda_{S_0} \varphi_i(X) \text{ в некоторой точке } 0 < S_0 < S_\infty \quad (2.1.32)$$

и

$$\begin{aligned} (h_i(X, S) - h_i(X, S_0))(s_i - s_{i_0}) &< 0 \text{ при} \\ S \neq S_0, X > 0 \quad (i = 1, \dots, n), \end{aligned}$$

где λ_{S_0} — собственное значение оператора $F(\cdot, S_0)$.

Содержательный смысл перечисленных выше условий приведен в п. 2.1.3. Тип обратной связи (2.1.30) состоит в одинаковом влиянии механизмов лимитирования на все компоненты экологической системы. Условие

(2.1.31) означает, что лимитирующий фактор влияет лишь на i -ю компоненту экологической системы и в свою очередь зависит только от этой компоненты. Условие (2.1.32) состоит в том, что система стремится "стабилизировать" некоторое стационарное значение лимитирующего фактора.

Приведенные механизмы лимитирования гарантируют сходимость к стационарной точке в модели (2.1.27).

Теорема 2.1.22. Пусть уравнение (2.1.27) имеет стационарное решение $X_0, S_0 = \Phi(X_0)$. Если выполнено условие (2.1.30), (2.1.31) или (2.1.32), то всякое решение этого уравнения, начинающееся в $R_+^m \setminus 0$, стремится к стационарной точке при $t \rightarrow \infty$.

Доказательство состоит в тривиальном перенесении рассуждений теоремы 2.1.5 на многомерный случай.

Теорема 2.1.23. Если функции $F(X, S)$ и $\Phi(X)$ непрерывно дифференцируемы в окрестности стационарной точки $X_0, S_0 = \Phi(X_0)$ уравнения (2.1.27) и выполнены условия теоремы 2.1.22, то стационарное решение асимптотически устойчиво по Ляпунову в R_+^m .

Доказательство состоит в тривиальном перенесении рассуждений теоремы 2.1.6 на многомерный случай.

Для модели с внешними источниками перечисленные типы лимитирования гарантируют устойчивость стационарных решений "в малом".

Теорема 2.1.24. Предположим, что функции $F(X, S)$ и $\Phi(X)$ непрерывно дифференцируемы в окрестности стационарного решения $X_0, S_0 = \Phi(X_0)$ уравнения (2.1.29). Пусть в этой окрестности выполнено одно из условий: (2.1.30) и B – собственный вектор оператора $F(X)$, (2.1.31) или (2.1.32); тогда стационарное решение асимптотически устойчиво по Ляпунову.

Доказательство состоит в тривиальном перенесении рассуждений теоремы 2.1.7 на многомерный случай.

В качестве примера рассмотрим динамику численности популяции микроорганизмов культивируемой на многосубстратной среде.

При непрерывном культивировании микроорганизмов лимитирующее вещество может содержаться в нескольких подаваемых в реактор субстратах. В этом случае их потребление обычно происходит последовательно [27, 28]. Причина появления предпочтения при утилизации субстратов состоит в том, что один из них ингибирует потребление других [27, 28]. Если ингибирования не происходит, то одновременно потребляется несколько источников лимитирующего вещества [27, 28, 37].

Пусть для жизнедеятельности клеток необходимо присутствие в культуральной среде n различных веществ, концентрацию каждого из которых будем обозначать через s_i ($i = 1, \dots, n$). Введем в рассмотрение вектор $S = (s_1, \dots, s_n)^*$. Продолжительности фаз развития клеток, как и в предыдущем случае, являются функциями S . Поэтому по аналогии с п. 2.1.4 процесс роста популяции описывает системой

$$\begin{aligned} \dot{X} &= AP(S)X - vX, \\ \dot{S} &= -G^*P(S)X - vS + vS_p, \end{aligned} \quad (2.1.33)$$

где вектор X задает численный состав популяции, вектор S задает концентрации питательных субстратов в культуральной среде, столбец $G_{.j} = (g_{1j} \dots g_{mj})^*$ прямоугольной матрицы $G = \{g_{ij}\}_{m \times n}$ задает коэффи-

циенты потребления клетками популяции j -го субстрата, v — скорость разбавления, $P(S)$ — матрица продолжительностей фаз развития, A — матрица, задающая структуру размножения популяции (например, см. формулы (1.2.2) и (1.2.3)). Как и в предыдущем параграфе, предполагаем, что A имеет в положительном конусе R_+^m единственный собственный вектор с положительными компонентами, которому отвечает вещественное собственное значение $\lambda(A) > 0$. Последнее неравенство по-прежнему означает, что выращивается невырождающаяся популяция. Если в культуральной среде присутствуют субстраты, по отношению к которым клетки нейтральны, то включение их в модель нецелесообразно, поскольку не вносит дополнительной информации. На этом основании считаем, что $G \geq 0$ и

$\sum_{i=1}^m g_{ij} > 0$ ($j = 1, \dots, n$). Одновременно с этим предполагаем, что набор

субстратов, рассматриваемых в модели, является достаточно полным. Другими словами, это означает, что клетки каждой фазы развития потреб-

ляют хотя бы один субстрат, т.е. $\sum_{j=1}^n g_{ij} > 0$ ($i = 1, \dots, m$). Напомним,

что аналогичное предположение делалось по отношению к функции $\Phi(X)$ в модели (2.1.27). Из соображений "минимальности" модели будем считать, что $S_p > 0$, т.е. исключены из рассмотрения субстраты, которые могли бы присутствовать в питательной среде, но в данной конкретной ситуации не подаются в реактор.

Естественно предполагать, что скорость прохождения клетками i -й фазы развития не убывает при увеличении концентрации питательных веществ в культуральной среде и хотя бы одна фаза развития чувствительна к концентрации питательных веществ. Поэтому будем считать, что

$$\frac{\partial p_i(s)}{\partial s_k} \geq 0 \quad (i = 1, \dots, m) \quad (k = 1, \dots, n), \quad \sum_{i=1}^m \sum_{k=1}^n \frac{\partial p_i}{\partial s_k}(s) > 0. \quad (2.1.34)$$

Так как биосинтез невозможен, если отсутствует какая-либо компонента питательной среды, то будем считать, что

$$p_i(s_1, \dots, s_{k-1}, 0, s_{k+1}, \dots, s_n) = 0 \quad (i = 1, \dots, m). \quad (2.1.35)$$

Остановимся подробнее на конкретном виде функций $p_i(S)$. При последовательном потреблении субстратов важную роль может играть принцип "узкого места" [13, 14], состоящий в том, что скорость самой медленной ферментативной реакции лимитирует ход всех остальных биохимических превращений. При изменении соотношения питательных веществ в культуральной среде происходит смена "узкого места" метаболизма [14, 27], в результате чего меняется лимитирующий фактор. Механизм смены лимитирующего фактора может быть использован для объяснения отклонений от теоретической кривой функции зависимости удельной скорости роста микроорганизмов от концентрации лимитирующего субстрата [4, 27, 35, 37].

Приведенные рассуждения позволяют задать удельную скорость прохождения клетками каждой фазы развития по следующему правилу. Пусть $\mu_{ij}(s_j)$ — скорость прохождения i -й фазы развития в случае, когда лимити-

рующим фактором является j -й субстрат, а все прочие компоненты питательной среды присутствуют в культуральной жидкости в избытке (например, $\mu_{ij}(s_j) = \mu_{ij}^0 \frac{s_j}{s_j + k_j}$). Тогда при произвольном соотношении питательных субстратов в среде культивирования [7]

$$p_i(S) = \min \{ \mu_{i1}(s_1), \dots, \mu_{in}(s_n) \}.$$

При параллельном влиянии нескольких компонент питательной среды на скорость биосинтеза в качестве первого приближения может быть использована функция вида [2, 28, 32]

$$p_i(S) = \mu_{i1}(s_1) \mu_{i2}(s_2) \dots \mu_{in}(s_n),$$

где $\mu_{ij}(s_j)$ задает независимое влияние j -го субстрата на ход всего процесса биосинтеза.

В дальнейших рассуждениях конкретный вид функций $p_i(S)$ ($i = 1, \dots, m$) не используется и фигурируют лишь предположения (2.1.34) и (2.1.35), которые в обоих приведенных выше примерах, очевидно, выполняются для почти всех S .

Сформулируем основные свойства хеомстатного режима культивирования в многосубстратном случае.

Теорема 2.1.25. Система (2.1.33) диссипативна. Она имеет нетривиальное стационарное решение тогда и только тогда, когда выполняется неравенство

$$\lambda(AP(S_p)) > v. \quad (2.1.36)$$

Если матрица $P(S)$ представима в виде $P(S) = P(\varphi(S))$, где $\varphi(S)$ — положительно однородная и монотонная функция предпочтения, как в п. 2.1.1, то стационарное решение единственно.

Если условие (2.1.36) выполнено и в начальный момент времени $X(0) \neq 0$, то суммарная численность популяции $\sum_{i=1}^m x_i(t)$ при достаточно больших t устанавливается большей некоторого положительного числа. Это число не зависит от начального состояния системы.

Доказательство. Так как $\sum_{i=1}^m g_{ij} > 0$, то существует собственный вектор L матрицы A^* , отвечающий $\lambda(A)$, для которого $0 < L < \sum_{i=1}^n G_i/\lambda(A)$. Из формулы (2.1.33) имеем

$$L^* \dot{X} = \lambda(A) L^* P(S) X - v L^* X,$$

$$\sum_{i=1}^n \dot{s}_i \leq -\lambda(A) L^* P(S) X + v \sum_{i=1}^n (s_{ip} - s_i).$$

Отсюда получаем

$$L^* X + \sum_{i=1}^n s_i \leq \sum_{i=1}^n s_{ip} + e^{-vt} (L^* X(0) + \sum_{i=1}^n s_i(0) - \sum_{i=1}^n s_{ip}).$$

Пусть константа $\beta > 0$ такова, что $G_i < \beta L$ ($i = 1, \dots, n$). Тогда $\dot{s}_i \geq -\beta L^* P(S) X - v S + v S_p$, и из системы (2.1.33) получаем

$$\beta L^* X + s_i \geq s_{ip} + e^{-vt} (\beta L^* X(0) + s_i(0) - s_{ip}).$$

Таким образом, система (2.1.33) имеет асимптотически устойчивое инвариантное выпуклое множество

$$L^*X + \sum_{i=1}^n s_i \leq s_{ip}, \quad \beta L^*X + s_i \geq s_{ip} \quad (i = 1, \dots, n).$$

Исключим из него точку $X = 0$. Выберем $\epsilon > 0$ настолько малым, что

$$\lambda(AP(S_p)) - \nu + \sum_{i=1}^m (P_i(S_p)/P_i(S) - 1) \geq \delta > 0$$

при $s_i \geq s_{ip} - \epsilon$ ($i = 1, \dots, n$).

Пусть R ($R > L$) – собственный вектор матрицы $(AP(S_p))^*$, отвечающий собственному значению с максимальной вещественной частью и $\alpha \leq \epsilon/\beta$. Тогда $s_i \geq s_{ip} - \epsilon$ ($i = 1, \dots, n$) при $L^*X \leq R^*X \leq \alpha$. Из системы (2.1.33) имеем $R^*X \geq \delta R^*X$ при $R^*X \leq \alpha$ и $X \neq 0$. Асимптотически устойчивое положительно инвариантное множество $L^*X + \sum_{i=1}^n s_i \leq s_{ip}$, $R^*X \geq \alpha$ решениями системы (2.1.33) отображается в себя и, следовательно, содержит стационарную точку. Если $P(S) = P(\varphi(S))$, то единственность стационарного решения следует из теоремы 2.1.2. Теорема доказана.

Условия равномерного лимитирования, автолимитирования и интегрального лимитирования гарантируют устойчивость стационарного решения.

Теорема 2.1.26. Пусть система (2.1.33) имеет единственное стационарное решение X_0, S_0 и выполнено одно из условий:

- $P(S) = p(S)I$, где $p(S)$ – скалярная монотонно убывающая функция;
- каждый столбец матрицы $P(S_0)G = P(S_0)\{G_{.1}, \dots, G_{.n}\}$ является собственным вектором матрицы $(AP(S_0))^*$, который отвечает собственному значению $\lambda(AP(S_0))$;
- размерность вектора S не превосходит размерности вектора X , т.е. $n \leq m$.

Матрица $P(S)$ такова, что

$$\begin{aligned} p_i(S) &= p_i(s_i) && \text{при } 1 \leq i \leq n, \\ p_j(S) &= 1 && \text{при } n+1 \leq j \leq m. \end{aligned}$$

Матрица G диагональна при $n = m$ и в общем случае, когда $n < m$, $g_{ji} = 0$ при $i \neq j$, $g_{ii} \neq 0$ и $1 \leq i \leq n$, $1 \leq j \leq m$. Тогда стационарное решение асимптотически устойчиво в $R_+^m \setminus 0$.

Доказательство. Вначале докажем сходимости к стационарной точке. Из системы (2.1.33) имеем

$$G^*A^{-1}X + S = S_p + e^{-\nu t}(G^*A^{-1}X(0) + S(0) - S_p).$$

Пусть выполнено условие а). Доказательство сходимости решения к стационарной точке проводится так же, как и в теореме 2.1.5, если использовать результаты теоремы 2.1.25.

Пусть выполнено условие б). Поскольку все строки матрицы G^* равны с точностью до постоянного множителя, система (2.1.23) на инвариантной гиперповерхности $G^*A^{-1}X + S = S_p$ приводится к виду (2.1.17) и для нее верна теорема 2.1.5.

Докажем теперь устойчивость по линейному приближению. Переходя к переменным $Y = A^{-1}X$ и воспользовавшись соотношением $S = S_0$, получаем, что матрица линейного приближения в окрестности стационарной точки имеет вид

$$D = AP(S_0) - \nu I + D_S P(S_0) X_0 G^*$$

Легко показать, что

$$\det[\lambda I - D] = \left| \frac{\lambda I - AP(S_0) + vI}{G^*} \middle| \frac{D_S P X_0}{I_n} \right|,$$

где I_n — единичная матрица размерности $n \times n$. Дальнейшее доказательство проводится так же, как и в теореме 2.1.6, с использованием формулы окаймленного определителя вида [18]

$$\det[\lambda I - D] = \det[\lambda I - A_1] \det[I_n + G^*(\lambda I_m - A)^{-1} Q],$$

где $A_1 = AP(S_0) - vI$ и $Q = D_S P(S_0) X_0$. Теорема доказана.

Для изучения динамики процесса биосинтеза при многопоточном культивировании рассмотрим пример двухступенчатой системы. В соответствии с результатами п. 2.1.4 для второго ферментера имеем

$$\begin{aligned} \dot{X}^2 &= AP(S^2) X^2 - v_2 X^2 + u_2 X^1, \\ \dot{S}^2 &= -G^* P(S^2) X^2 - v_2 S^2 + u_2 S^1 + (v_2 - u_2) S_p^2, \end{aligned} \quad (2.1.37)$$

где верхний индекс при векторе X или S обозначает номер ферментера, v_2 — скорость разбавления во втором ферментере, u_2 — скорость потока из первого во второй ферментер. Динамика величин X^1 и S^1 задается системой (2.1.33).

Теорема 2.1.27. Система (2.1.37), решаемая совместно с системой (2.1.33), диссипативна и при выполнении условия (2.1.36) имеет нетривиальное ($X^1 \neq 0$) стационарное решение. Если выполнено условие а) или б) теоремы 2.1.26, то это решение асимптотически устойчиво по Ляпунову.

Доказательство. Воспользуемся обозначениями из теоремы 2.1.25. На основании формулы (2.1.37) имеем

$$\begin{aligned} L^* \dot{X}^2 &= \lambda(A) L^* P(S^2) X^2 - v_2 L^* X^2 + u_2 L^* X^1, \\ \sum_{i=1}^n \dot{s}_i^2 &\leq -\lambda(A) L^* P(S^2) X^2 - v_2 \sum_{i=1}^n s_i^2 + u_2 \sum_{i=1}^n s_i^1 + (v_2 - u_2) \sum_{i=1}^n s_{ip}^2, \end{aligned}$$

откуда получаем

$$L^* \dot{X}^2 + \sum_{i=1}^n \dot{s}_i^2 \leq -v_2 (L^* X^2 + \sum_{i=1}^n s_i^2) + u_2 (L^* X^1 + \sum_{i=1}^n s_i^1) + (v_2 - u_2) \sum_{i=1}^n s_{ip}^2.$$

По теореме 2.1.25 величина $L^* X^1 + \sum_{i=1}^n s_i^1$ при достаточно больших t ограничена не зависящей от начальных условий константой. Отсюда, очевидно, следует асимптотическая ограниченность $L^* X^2 + \sum_{i=1}^n s_i^2$. Диссипативность доказана. Стационарное решение существует по теореме о неподвижной точке в силу выпуклости и инвариантности

множества $L^* X^2 + \sum_{i=1}^n s_i^2 \leq M$.

Докажем устойчивость стационарного решения системы (2.1.37). В системе (2.1.37) сделаем замену переменной $Y = A^{-1} X$. Матрица линейного приближения в этом случае принимает вид, приведенный в доказательстве теоремы 2.1.26. Дальнейшее доказательство проводится так же, как и в теореме 2.1.7, с использованием формулы окаймленного определителя [18]. Теорема доказана.

2.1.7. Анализ возрастной структуры непрерывно культивируемой популяции микроорганизмов. Использование метода синхронизации процесса размножения клеточной популяции позволило доказать, что чувствительность клеток к условиям внешней среды существенно зависит от возраста

[30]. Синхронизирующие воздействия имеют физико-химическую природу и тормозят процессы, протекающие в определенной фазе развития клеток. При этом на процессы, происходящие в остальных фазах развития, они оказывают значительно меньшее влияние. В результате большинство клеток данной популяции задерживается в требуемой возрастной группе. Разумеется, полная синхронизация никогда не достигается и в популяции всегда остаются клетки, имеющие другие фазы развития. Таким образом, формируется распределение клеток по возрастным группам, имеющее явно выраженный максимум и сравнительно небольшую дисперсию. При последующих генерациях дисперсия быстро возрастает.

Данный метод позволяет определять продолжительность фаз развития клеток культивируемой популяции. Его основными недостатками являются сравнительно небольшая точность и методическая необоснованность перенесения результатов, полученных при стрессовых синхронизирующих воздействиях на процессы, протекающие при стабильных внешних условиях. В этом пункте излагается методика, которая позволяет идентифицировать параметры модели динамики численности непрерывно культивируемой популяции и, в частности, определить продолжительность различных фаз развития клеток на основании экспериментальных данных, полученных в условиях непрерывного культивирования.

Несмотря на то, что некоторые биологические явления можно объяснить лишь с помощью модели, учитывающей возрастную структуру популяции, обработка и планирование эксперимента с ее помощью имеют свои недостатки. Основной из них состоит в том, что модель (1.2.1) содержит значительно большее число констант, чем модель Моно. Поэтому естественно возникает проблема использования модельных соображений, позволяющих заменить прямые эксперименты по оценке параметров модели сравнительно простыми косвенными.

Приведем метод определения коэффициентов потребления субстрата клетками разных фаз развития с помощью модели (1.2.1). В обозначениях п. 2.1.5 рассматривается задача определения m компонент g_1, \dots, g_m строки G .

В п. 2.1.5 показано, что система (1.2.1) имеет интеграл $G^* A^{-1} X = s_p - s$. Предположим, что, варьируя скорость разбавления или концентрацию поступающего в реактор субстрата, можно организовать $n \geq m$ таких режимов культивирования, при которых распределение клеток по возрастам не повторяется по крайней мере в m экспериментах, т.е. векторы численного состава популяции в этих экспериментах удовлетворяют условиям

$$X_{i_k} / \|X_{i_k}\| \neq X_{i_p} / \|X_{i_p}\| \quad \text{при} \quad i_k \neq i_p,$$

где i_k, i_p принимают значения из некоторого подмножества $\{i_1, \dots, i_m\} \subset \{1, \dots, n\}$. Пусть s_i и s_{p_i} ($i = 1, \dots, m$) — соответствующие концентрации субстрата в реакторе и на его входе, а $y_i = C^* X_i / (s_{p_i} - s_i)$ — коэффициенты утилизации, где $C^* X_i$ — суммарная биомасса культуры, а $C = (c_1, \dots, c_m)^*$ — вектор, компоненты которого c_i ($i = 1, \dots, m$) задают массу клеток в различных фазах развития. В итоге имеем следующую систе-

му n линейных уравнений относительно $m \leq n$ неизвестных:

$$G^* A^{-1} X_i / C^* X_i = Y_i^{-1} \quad (i = 1, \dots, n).$$

При $n > m$ подобная система, вообще говоря, неразрешима. Метод наименьших квадратов дает "наилучшее" в некотором смысле решение подобной системы (подробнее см. п. 4.4.3). Воспользуемся им. Вводя обозначения $\Xi = [X_1, X_2, \dots, X_n]$ и $Z = (y_1^{-1} C^* X_1, \dots, y_n^{-1} C^* X_n)$, перепишем эту систему в виде $G^* A^{-1} \Xi = Z^*$, или

$$\Xi^* (A^*)^{-1} G = Z, \quad (2.1.38)$$

откуда $G = A^* (\Xi^*)^+ Z$, где $(\Xi^*)^+$ – псевдообратная матрица к матрице Ξ^* . Решение существует и единственно, так как ранг матриц A и Ξ равен m .

Проиллюстрируем изложенный метод конкретным примером. В работе [48] культура *S. Cerevisiae* размножалась в хемостате, при различных скоростях разбавления измерялась концентрация биомассы $C^* X$, концентрация субстрата s и вычислялся коэффициент утилизации $Y = C^* X / (s_p - s)$. Кроме того, варьировался состав культуральной среды и измерялось распределение клеточной популяции по размерам. На рис. 2.2, а представлено распределение клеток по размерам при трех скоростях разбавления в среде без глутаминовой кислоты, а на рис. 2.2, б – распределение по размерам при добавлении в среду 0,075% раствора глутаминовой кислоты. Результаты, представленные в табл. 2.1, соответствующей опыту без глутаминовой кислоты, и в табл. 2.2, соответствующей опыту с глутаминовой кисло-

Таблица 2.1

	Диаметр, мкм	$v = 0,02$	$v = 0,055$	$v = 0,095$
x_1 (%)	3,3–4	26	17	15
x_2 (%)	4–5	36	39	39
x_3 (%)	5–8,6	38	44	46
Коэффициент утилизации		0,414	0,491	0,498
Суммарная биомасса г/л		3,93	4,7	4,75

Таблица 2.2

	Диаметр, мкм	$v = 0,011$	$v = 0,029$	$v = 0,102$
x_1 (%)	3,3–4	30	23	15
x_2 (%)	4–5	36	37	38
x_3 (%)	5–8,6	34	40	47
Коэффициент утилизации		0,43	0,461	0,5
Суммарная биомасса, г/л		4,08	4,39	4,77

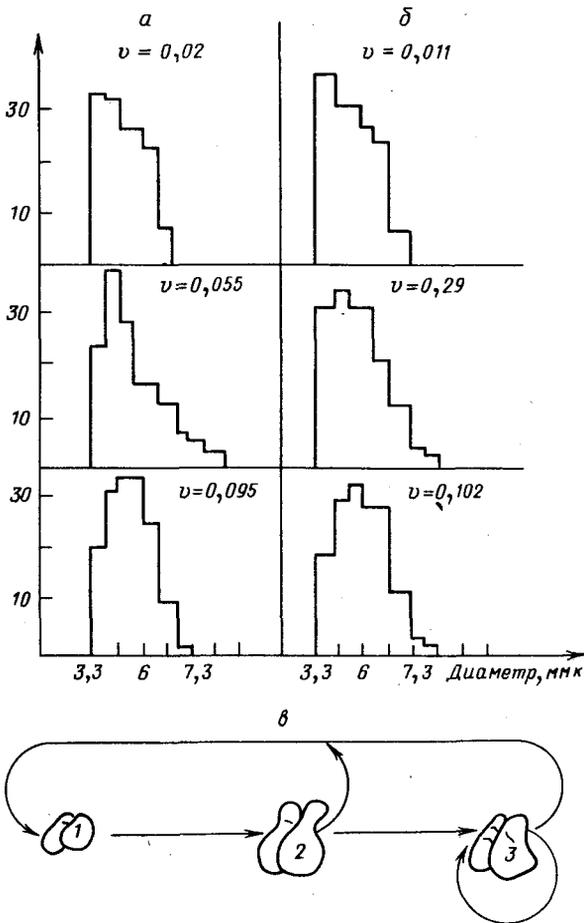


Рис. 2.2. а, б) Распределение по размерам клеток непрерывно культивируемой популяции *S. Cerevisiae*; в) схема размножения клеток *S. Cerevisiae*

той, говорят о том, что коэффициент утилизации увеличивается с ростом скорости разбавления.

Попытаемся связать этот процесс с перестройкой возрастной структуры популяции. В соответствии с данными, приведенными в работе [48], естественно ввести следующую возрастную структуру популяции: молодые почки, клетки, способные к почкованию, и зрелые клетки, давшие хотя бы одного потомка. Считаем, что молодые клетки не дают потомков до тех пор, пока не достигнут размеров взрослых особей. Почка в начальный момент имеет наименьший размер, а способные к генерации клетки мало меняют свои размеры [52, 53]. Такая схема размножения представлена на рис. 2.2, в, и ей соответствует матрица в уравнении (1.2.1) вида

$$A = \begin{bmatrix} -1 & 1 & 1 \\ 1 & -1 & 0 \\ 0 & 1 & 0 \end{bmatrix}. \quad (2.1.39)$$

Легко видеть, что

$$A^{-1} = \begin{bmatrix} 0 & 1 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}.$$

Пусть x_1 , x_2 и x_3 — число клеток каждой возрастной группы. Распределение численного состава популяции по возрастам, полученное на основании рис. 2.2, представлено в табл. 2.1 (среда без глутаминовой кислоты) и в табл. 2.2 (среда с глутаминовой кислотой). Перейдем к относительным единицам измерения численности клеток в ферментере, так что общее число клеток в единице объема реактора будет равняться единице и x_i ($i = 1, 2, 3$) — доля клеток каждой возрастной группы. В итоге система (2.1.38) в первом случае принимает вид

$$0,01(74g_1 + 38g_2 + 62g_3) = 9,49,$$

$$0,01(83g_1 + 44g_2 + 56g_3) = 9,57,$$

$$0,01(85g_1 + 46g_2 + 54g_3) = 9,54,$$

а во втором — вид

$$0,01(70g_1 + 34g_2 + 66g_3) = 9,49,$$

$$0,01(77g_1 + 40g_2 + 60g_3) = 9,54,$$

$$0,01(85g_1 + 47g_2 + 53g_3) = 9,54.$$

Обе системы с точностью до значения в третьем знаке имеют решение $g_1 = 6,9$, $g_2 = 0$ и $g_3 = 6,9$. Отсюда заключаем, что коэффициенты потребления субстрата клетками разных возрастных групп являются устойчивыми характеристиками данной популяции.

Чтобы связать изменение коэффициента утилизации с перестройкой возрастной структуры популяции, необходимо определить удельные скорости потребления субстрата клетками разных возрастных групп, которые в соответствии с системой (1.2.1) задаются выражениями $-\frac{d}{dt} \frac{\partial s}{\partial x_i} = g_i p_i(s)$ ($i = 1, \dots, m$). Таким образом, мы приходим к задаче определения скорости смены фаз развития клеток p_i ($i = 1, \dots, m$) или, что эквивалентно, определения средней продолжительности этих фаз: $\bar{T}_i = 1/p_i$ ($i = 1, \dots, m$).

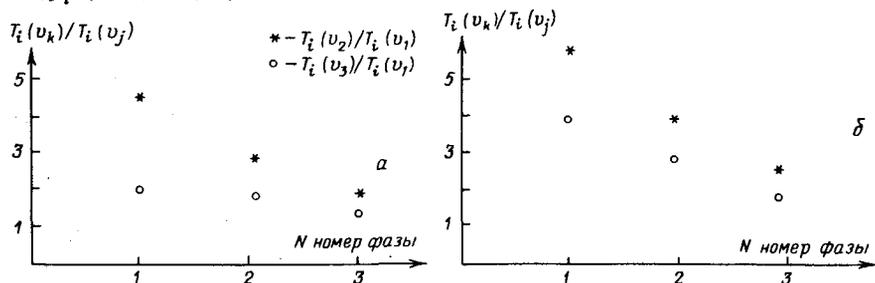


Рис. 2.3. Отношение продолжительностей фаз развития клеток при различных скоростях разбавления

Изложим метод решения этой задачи в общем виде, а затем применим результаты к анализу работы [28].

Пусть динамика численности культивируемой популяции задается системой вида

$$\begin{aligned} \dot{x}_1 &= b_1 p_1 x_1 - d_1 x_1 + b_2 p_2 x_2 + \dots + b_m p_m x_m - v x_1, \\ \dot{x}_i &= c_{i-1} p_{i-1} x_{i-1} - p_i x_i - d_i x_i - v x_i \quad (i = 2, \dots, m), \\ \dot{s} &= - \sum_{i=1}^m g_i p_i x_i - v s + v s_p, \end{aligned} \quad (2.1.40)$$

где b_i ($i = 1, \dots, m$) – число потомков каждой особи, p_i ($i = 1, \dots, m$) – скорость перехода из фазы в фазу, c_i ($i = 1, \dots, m - 1$) – доля клеток, перешедших в следующую фазу развития, d_i ($i = 2, \dots, m$) – смертность. Эта система является частным случаем системы (1.2.1) с матрицей размножения

$$A = \begin{pmatrix} b_1 & \dots & b_{m-1} & b_m \\ c_1 - 1 & & \cdot & \cdot \\ 0 & c_2 - 1 & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ 0 & & c_{m-1} - 1 & \cdot \end{pmatrix} \quad \text{при } d_1 = d_2 = \dots = d_m,$$

которая охватывает случай (1.2.2) и (1.2.3). Обозначим через x_i ($i = 1, \dots, m$) долю клеток i -й фазы развития или i -й возрастной группы от общего числа клеток в единице объема ферментера, так что $\sum_{i=1}^m x_i = 1$.

Определив эти величины экспериментально и приравняв нулю правую часть системы (2.1.40), получаем соотношения для вычисления скоростей перехода из одной возрастной группы в другую [6]:

$$\begin{aligned} p_i &= \frac{x_{i+1}}{c_i x_i} (p_{i+1} + v_{i+1}) \quad (i = 1, \dots, m - 1), \\ p_m &= (v_1 \bar{x}_1 - \sum_{i=1}^{m-1} b_{m-i} \sum_{j=1}^i v_{m+1-j} x_{m+1-j}) / \prod_{k=j}^i c_{m-k} / (x_m (\sum_{i=0}^{m-1} b_{m-i} / \prod_{k=1}^i c_{m-k})) \end{aligned} \quad (2.1.41)$$

$$\text{и } \bar{T}_i = 1/p_i \quad (i = 1, \dots, m),$$

где по определению $\prod_{k=1}^0 = 1$ и $v_i = v + d_i$ ($i = 1, \dots, m$).

В случае, когда структура размножения популяции задается матрицей (2.1.39), алгоритм отыскания скоростей перехода из одной возрастной группы в другую получается из формул (2.1.41) следующим образом. Положим $m = 3$, $b_1 = -1$, $b_2 = 1$, $b_3 = 1$, $c_1 = 1$, $c_2 = c$, $v_1 = v_2 = v$ и $v_3 = cv$. Тогда искомые величины являются пределом решения (2.1.41) при $c \rightarrow \infty$. Отсюда имеем

$$p_3 = v(x_1 + x_2)/x_3, \quad p_2 = vx_3/x_2, \quad p_1 = (p_2 + v)x_2/x_1.$$

Применим эти формулы к результатам эксперимента [48]. Данные, полученные при культивировании без глутаминовой кислоты, представлены в табл. 2.3, а при добавлении глутаминовой кислоты – в табл. 2.4. На рис. 2.3 и в табл. 2.5 показано относительное изменение продолжи-

Таблица 2.3

	1			2			3		
	$\bar{T}_1, \text{час}$	$P_1, \text{час}^{-1}$	$-\frac{\partial^2 s}{\partial r \partial x_1}$	$\bar{T}_2, \text{час}$	$P_2, \text{час}^{-1}$	$-\frac{\partial^2 x}{\partial r \partial x_2}$	$\bar{T}_3, \text{час}$	$P_3, \text{час}^{-1}$	$-\frac{\partial^2 s}{\partial r \partial x_3}$
v									
$v_1 = 0,02$	17	0,057	0,039	47	0,021	0	27	0,033	0,22
$v_2 = 0,055$	3,7	0,27	1,9	16	0,062	0	14,3	0,07	0,48
$v_3 = 0,095$	1,8	0,53	3,66	9,1	0,11	0	9,1	0,11	0,76

Таблица 2.4

	1			2			3		
	$\bar{T}_1, \text{час}$	$P_1, \text{час}^{-1}$	$-\frac{\partial^2 s}{\partial r \partial x_1}$	$\bar{T}_2, \text{час}$	$P_2, \text{час}^{-1}$	$-\frac{\partial^2 s}{\partial r \partial x_2}$	$\bar{T}_3, \text{час}$	$P_3, \text{час}^{-1}$	$-\frac{\partial^2 s}{\partial r \partial x_3}$
v									
$v_1 = 0,011$	38	0,026	0,18	97	0,0103	0	47	0,0213	0,14
$v_2 = 0,029$	10	0,096	0,65	32	0,631	0	23,2	0,04	0,29
$v_3 = 0,102$	1,73	0,578	3,92	7,94	0,126	0	8,7	0,115	0,79

Таблица 2.5

а)	v_3/v_1	v_2/v_1	v_3/v_2
		4,75	2,5
$T_1(v_1)/T_1(v_3)$	9,4	4,6	2
$T_2(v_1)/T_2(v_3)$	5	2,9	1,8
$T_3(v_1)/T_3(v_3)$	3	1,8	1,6

б)	v_3/v_1	v_2/v_1	v_3/v_2
		9,2	2,6
$T_1(v_1)/T_1(v_3)$	22	3,8	5,8
$T_2(v_1)/T_2(v_3)$	12	3	4
$T_3(v_1)/T_3(v_3)$	5	2	2,7

тельности пребывания в различных возрастных группах при смене скорости разбавления на среде без глутаминовой кислоты а) и на среде с глутаминовой кислотой б). Отсюда видно, что наибольшей чувствительностью к концентрации субстрата в ферментере обладают молодые клетки и что эта чувствительность падает с возрастом. Воспользовавшись данными, приведенными в табл. 2.1, 2.2., 2.3, 2.4, легко убедиться в том, что во всех шести режимах культивирования выполняется равенство

$$y = \frac{C^* X}{s_p - s} = vG^* X / (g_1 p_1 x_1 + g_2 p_2 x_2 + g_3 p_3 x_3) = - \frac{vC^* X}{\sum_{i=1}^3 \frac{\partial^2 s}{\partial t \partial x} x_i}$$

Следовательно, доказано, что изменение коэффициента утилизации связано с перестройкой возрастной структуры популяции.

Предложенный метод определения продолжительностей фаз развития клеток был использован [28] для анализа динамических характеристик двух штаммов сахаромисцетов. Разбиение по возрастным группам проводилось в соответствии с числом почкований, которое прошла клетка. Принадлежность клеток той или иной возрастной группе определялась визуально по числу родовых шрамов на стенке клетки. Для данного случая алгоритм вычисления скоростей смены фаз развития получается из формул (2.1.41) и имеет вид [6]

$$p_m = v(x_1 - \sum_{i=3}^m (i-2)x_i / ((m-1)x_m)),$$

$$p_i = \frac{x_{i+1}}{x_i} (v + p_{i+1}) \quad (i = 1, \dots, m-1),$$

или

$$p_1 = \frac{v}{(m-1)x_1} \left[\sum_{i=2}^{m-1} (m-i)x_i + 1 \right],$$

$$p_i = \frac{x_{i-1}}{x_i} p_{i-1} - v \quad (i = 2, \dots, m).$$

Для делящихся клеток, структура размножения которых задается матрицей (1.2.2), алгоритм вычисления скоростей смены фаз принимает вид [6]

$$p_m = \frac{v}{(q-1)x_m},$$

$$p_i = \frac{x_{i+1}}{x_i} (v + p_{i+1}) \quad (i = 1, \dots, m-1).$$

Данный алгоритм был применен для анализа продолжительностей фаз развития популяции.

Проведенное выше исследование динамических свойств культивируемой популяции с помощью модели (1.2.1) убедительно показывает, что учет возрастной структуры популяции необходим для правильного понимания целого ряда экспериментальных фактов. При этом привлечение модельных соображений позволяет отказаться от прямых измерений и тем самым облегчает задачу исследователя. Прямые эксперименты по анализу чувствительности различных возрастных групп культивируемой популяции дрожжей *S. Utilis* показали [52, 53], что клетки разных фаз развития по-разному реагируют на измерение условий лимитирования. В результате автор этих работ также пришел к выводу о необходимости учета возрастной структуры популяции для правильного понимания экспериментального материала.

§ 2.2. Автоколебания

В этом параграфе изучаются условия возникновения колебаний в моделях эксплуатируемых экологических систем. Исследуется бифуркация рождения периодических решений и рассматривается вопрос устойчивости этих решений. Доказывается колебательный характер движений в случае неустойчивости стационарного режима и рассматриваются условия рождения нелокальных колебаний. Полученные результаты применяются к модели биохимической кинетики.

Приводимые в этом параграфе результаты касаются m -мерных систем, сведение которых к классическим двумерным осцилляторам невозможно хотя бы потому, что в моделях динамики численности эксплуатируемых популяций периодические режимы рождаются при размерностях, не меньших трех. Специфика изучаемых систем позволяет модифицировать метод Малкина — Лурье применительно к задачам популяционной динамики и благодаря этому получить для целого класса моделей аналитический критерий устойчивости периодических решений. Основным методом доказательства существования колебательных режимов в изучаемых системах является метод теории бифуркаций при малых изменениях параметров систем. В то же время имеется класс задач, для которых предположение о малости бифуркационного параметра не выполняется. В результате возникают проблемы нелокальной теории колебаний. Привлекая топологические методы, в ряде случаев удается установить факт существования периодического решения даже при нелокальной постановке задачи.

2.2.1. Бифуркация рождения периодических решений. Использование методов теории бифуркаций накладывает достаточно сильные требования на гладкость векторного поля динамической системы. Поэтому будем

предполагать, что правые части уравнений (2.1.2) и (2.1.3) столько раз непрерывно дифференцируемы в $R_+^m \setminus 0$, сколько этого будут требовать применяемые в дальнейшей теоремы.

Переформулируем свойства изучаемых моделей в терминах производных (см. п. 2.1.1). Условие квазимонотонности функции $F_s(X)$ ($F_s(X) \equiv F(X, s)$) принимает вид

$$\partial f_i(X, s) / \partial x_j \geq 0 \text{ при } i \neq j \text{ (} i, j = 1, \dots, m \text{)}. \quad (2.2.1)$$

По-прежнему будем предполагать неразложимость оператора F_s при всех $s < s_\infty$ (см. п. 2.1.2) или, что эквивалентно, неразложимость матрицы $D_x F_s(X) = \{ \partial f_i(X, s) / \partial x_j \}_{i,j=1}^m$. Напомним, что матрица называется неразложимой, если путем перестановок одноименных строк и столбцов ее нельзя привести к блочно-треугольному виду (см. ниже (3.1.6)) [3].

В этом случае собственное значение $\lambda(D_x F_s(X))$ матрицы $D_x F_s(X)$ с максимальной вещественной частью [3] совпадает с собственным значением λ_s оператора F_s (см. лемму 2.1.6.1):

$$\lambda(D_x F_s(X)) = \lambda_s. \quad (2.2.2)$$

Условие замедления скорости воспроизводства в точке сбалансированного роста (2.1.6) принимает вид

$$D_s F(H_s, s) \not\leq 0, \quad (2.2.3)$$

где H_s – положительный собственный вектор оператора F_s , отвечающий собственному значению λ_s . Прекращение воспроизводства при $s \geq s_\infty$ означает, что матрица $D_x F_s(X)$ при $s \geq s_\infty$ не имеет собственных значений в правой полуплоскости.

В дальнейшем будем предполагать, что уравнения (2.1.2) и (2.1.3) имеют стационарное решение, и займемся локальным анализом этих уравнений в окрестности стационарной точки. По аналогии с условием (2.2.3) в модели динамики численности с внешними источниками будем предполагать, что лимитирующий фактор оказывает замедляющее воздействие на скорость воспроизводства в стационарной точке, т.е.

$$D_s F(X_0, s_0) \not\leq 0, \quad (2.2.4)$$

где $s_0 = \varphi(X_0)$ и X_0 – стационарное решение уравнения (2.1.3). Покажем, что для оператора, задающего динамику численности микроорганизмов при непрерывном культивировании, условия (2.2.3) и (2.2.4) всегда выполняются.

Предложение 2.2.1. Если $F(X, s) = AP(s)$, где A – неразложимая матрица с неотрицательными недиагональными элементами, $P(s) = \text{diag} \{ p_i(s) \}_{i=1}^m$ и $\partial p / \partial s \not\leq 0$, то замена переменных $Y = AX$ приводит уравнения (2.1.2) и (2.1.3) к виду, при котором условия (2.2.3) и (2.2.4) выполняются. В уравнении (2.1.17) и системе (2.1.21) при всех $i = 1, \dots, n$ они также выполняются.

Доказательство. Делаем замену $Y = AX$ в уравнении (2.1.2). Из условия стационарности имеем

$$\frac{\partial p}{\partial s} AY_0 = v \left(\frac{\partial p}{\partial s} \right) P^{-1}(s_0) Y_0 \not\leq 0,$$

где Y_0 — стационарное решение. Условие (2.2.3) доказано.

Делаем замену $Y = AX$ в уравнении (2.1.3) и, не умаляя общности, считаем в стационарной точке $P(s_0) = I$. Из условия стационарности получаем

$$AY_0 = (vI - A)^{-1}B.$$

По теореме 2.1.3 $v > \lambda(A)$, и поэтому [3, 24] $(vI - A)^{-1} > 0$. Так как $B \not\equiv 0$ и $\partial P/\partial s \not\equiv 0$, условие (2.2.4) выполняется. В формулах (2.1.17) и (2.1.21) условия (2.2.3) и (2.2.4) выполняются по доказанному выше. Предложение доказано.

Монотонность функции предпочтения означает, что

$$\partial\varphi/\partial x_i \geq 0 \quad (i = 1, \dots, m) \quad \text{и} \quad \sum_{i=1}^m \partial\varphi/\partial x_i > 0. \quad (2.2.5)$$

Следствием однородности функции предпочтения является формула Эйлера $\langle \partial\varphi/\partial X, X \rangle = \varphi(X)$ при $X \geq 0$. (2.2.6)

Будем предполагать, что функция $F(X, s)$ имеет в окрестности стационарной точки $X_0, s_0 = \varphi(X_0)$ уравнений (2.1.2) и (2.1.3) разложение по s в последовательность Тейлора вида

$$F(X, s) = F(X, s_0) + \mu F_1(X) (s - s_0) + F_2(X, s) (s - s_0),$$

где $\mu F_1(X) = D_s F(X, s_0)$, $\mu > 0$ — постоянный параметр и $F_2(X, s_0) = 0$.

Тогда матрица линейного приближения уравнений (2.1.2) и (2.1.3), вычисленная в стационарной точке, имеет вид

$$D = D_x F(X_0, s_0) - vI - \mu QC^*, \quad (2.2.7)$$

где $Q = -F_1(X_0) \not\equiv 0$ и $C^* = (\partial\varphi/\partial X)^* \not\equiv 0$. В силу теорем 2.1.2 и 2.1.3 и условия (2.2.2) собственное значение с максимальной вещественной частью матрицы $D_x F(X_0, s_0) - vI$ в случае (2.1.2) равно нулю, а в случае (2.1.3) отрицательно.

Отсюда следуют дополнительные критерии устойчивости "в малом" стационарных решений уравнений динамики численности эксплуатируемых популяций [8, 11]. В некоторых важных для практики случаях они позволяют доказать, что матрица линейного приближения является гурвицевой, т.е. обладает лишь собственными значениями с отрицательной вещественной частью.

Предложение 2.2.2. Если фазовое пространство систем (2.1.2) и (2.1.3) двумерно, т.е. $m = 2$, то матрица линейного приближения (2.2.7) гурвицева при любых $\mu > 0$.

Доказательство предложения 2.2.2, очевидно, получается из следствия технической леммы 3.

Предложение 2.2.3. Матрица линейного приближения (2.2.7) гурвицева при всех $\mu > 0$, меньших некоторого положительного числа ϵ .

Доказательство предложения 2.2.3. Известно [18], что ближайшее к нулю собственное значение матрицы (2.2.7) уравнения (2.1.2) с точностью до положительного множителя равняется $-\mu C^* B(0) Q + o(\mu)$, где $B(0)$ — присоединенная матрица для матрицы $A = D_x F(X_0, s_0) - vI$ ($\lambda(A) = 0$) при $\lambda = 0$ и $o(\mu)/\mu \rightarrow 0$ при $\mu \rightarrow 0$. Так как $B(0) > 0$, то $\lambda_1 < 0$. В уравнении (2.1.3) матрица $D_x F(X_0, s_0) - vI$ по теореме 2.1.3 и формуле (2.2.2) гурвицева. Ее малое возмущение (2.2.7) также сохраняет это свойство. Предложение доказано.

Содержательный смысл полученного результата состоит в том, что при достаточно малом лимитировании в эксплуатируемой популяции возни-

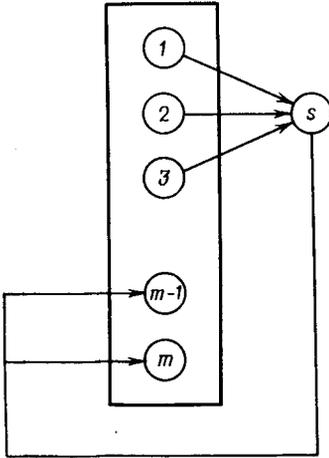


Рис. 2.4. Схема несогласованного лимитирования

кает асимптотически устойчивая стационарная структура, которая при отсутствии лимитирования может не существовать.

В теоремах 2.1.7 и 2.1.8 предыдущего параграфа было показано, что автолимитирование (2.1.12) любой интенсивности приводит к гурвицевости матрицы линейного приближения (2.2.7). Противоположность этому явлению составляет несогласованное лимитирование k -го порядка (рис. 2.4), точной формулировкой которого является следующее

Определение. Лимитирование является

несогласованным k -го порядка, если лимитирующий фактор влияет только на первые i компонент системы, а величина лимитирующего фактора формируется компонентами системы, имеющими номера с $(i + k)$ -го по m -й т.е.

$$\partial\varphi/\partial X = (0, 0, 0, \dots, \partial\varphi/\partial x_{i+k}, \dots, \partial\varphi/\partial x_m)^*, \quad (2.2.8)$$

$$D_s F(X_0, s_0) = (\partial f_1/\partial s, \dots, \partial f_i/\partial s, 0, 0, \dots, 0)^*.$$

Сформулируем критерий неустойчивости для матрицы, задающей воспроизводство популяции с возрастной или фазовой структурой (см. формулы (1.2.2) и (1.2.3)):

$$D_x F(X_0, s_0) = \{a_{ij}\}_{i,j=1}^m,$$

где

$$a_{ij} \geq 0 \text{ при } i \neq j, \quad a_{ij} = 0 \text{ при } i \geq j + 2. \quad (2.2.9)$$

Предложение 2.2.4. Пусть оператор $F(X, s)$ системы (2.1.2) или (2.1.3) таков, что матрица $D_x F(X_0, s_0)$ имеет структуру (2.2.9). Предположим также, что в системе реализовано несогласованное лимитирование порядка два или более. Тогда при достаточно больших значениях параметра μ стационарное решение системы (2.1.2) или (2.1.3) неустойчиво. Более того, спектр матрицы линейного приближения (2.2.7) пересекаются как с правой, так и с левой полуплоскостями.

Доказательство. По формуле окаймленного определителя [18] характеристический полином матрицы (2.2.7) можно представить в виде

$$\Delta(\lambda) = \lambda^m + a_1 \lambda^{m-1} + \dots + a_{m-1} \lambda + a_m + \mu C^* Q \lambda^{m-1} + \mu C^* B_1 Q \lambda^{m-2} + \dots$$

$$\dots + \mu C^* B_{m-1} Q,$$

где a_i ($i = 1, \dots, m$) — коэффициенты характеристического полинома матрицы $D_x F(X_0, s_0) - \nu I$ и B_i ($i = 1, \dots, m-1$) — коэффициенты матричного полинома присоединенной матрицы. Имеем $C^* Q = 0$ и $C^* B_1 Q = 0$, так как $B_1 = D_x F(X_0, s_0) - \nu I - aI$ [3]. Таким образом,

$$\Delta(\lambda) = \lambda^m + \mu b_{k-1} \lambda^{m-k} + \mu (b_k \lambda^{m-k-1} + \dots) + a_m,$$

где $b_{k-1} = C^* B_k Q > 0$ и $k \geq 3$. Пусть $\nu > 0$ таково, что $\nu^m = \mu$. Сделаем замену пере-

менной $\lambda = \nu^{m/k} x$. Характеристическое уравнение принимает вид

$$\nu^{m^2/k} x^m + \nu^{m^2/k} b_{k-1} x^{m-k} + (\nu^{m^2/k} b_k) \nu^{-m/k} x^{m-k-1} + \dots + a_m = 0.$$

Разделив на $\nu^{m^2/k}$, имеем $x^m + b_{k-1} x^{m-k} + \epsilon p(x, \epsilon) = 0$, где $\epsilon = \nu^{-m/k} \rightarrow 0$ при $\mu \rightarrow \infty$.

Так как $b_{k-1} > 0$ и $k \geq 3$, уравнение $x^{n-k}(x^k + b_{k-1}) = 0$ имеет корни как в правой, так и в левой полуплоскостях. То же верно при достаточно большом μ для $\Delta(\lambda)$. Предложение доказано.

Отметим структурную устойчивость полученного результата.

Следствие. Если $Q = (q_1, q_2, \dots, q_i, \epsilon q_{i+1}, \dots, \epsilon q_m)$, $C = (\epsilon c_1, \dots, \epsilon c_{i+k-1}, c_{i+k}, \dots, c_m)$ и $k \geq 2$, то при достаточно малом ϵ выводы предложения 2.2.4 сохраняют силу. Доказательство очевидно.

Каков бы ни был характер лимитирования, если стационарная точка уравнения (2.1.2) или (2.1.3) неустойчива, то в окрестности этой точки движения имеют колебательный характер. Точнее, выполнено

Предложение 2.2.5. Характеристический полином матрицы линейного приближения (2.2.7) имеет неотрицательные коэффициенты и положительный свободный член. Поэтому след матрицы неположителен и собственные значения, которые не лежат в левой полуплоскости, имеют ненулевую мнимую часть [11].

Доказательство очевидно, если воспользоваться следствием технической леммы 3.

В предложении 2.2.3 показано, что матрица линейного приближения является гурвицевой при малых положительных значениях параметра μ . Если она не является гурвицевой при всех значениях этого параметра, как, например, при выполнении условий предложения 2.2.4, то найдется такое число μ , при котором она теряет это свойство. Нижняя граница таких чисел положительна (см. предложение 2.2.3), и, не умаляя общности, будем считать ее равной единице.

Предположим, что правая граница спектра матрицы (2.2.7) при μ , близких к единице, имеет вид $\alpha(\mu) + j\beta(\mu)$, где $\alpha(1) = 0$, $\beta(1) = \omega$, j — мнимая единица и $\partial\alpha/\partial\mu(1) = \nu \neq 0$. Тогда по теореме Хопфа существует непрерывная функция $\mu(\epsilon)$ ($\mu(0) = 1$) такая, что при достаточно малых ϵ имеет место бифуркация рождения периодического решения $X(t, \mu(\epsilon))$ периода $2\pi/\omega + o(\epsilon)$, где $o(\epsilon)/\epsilon \rightarrow 0$ при $\epsilon \rightarrow 0$.

Рассмотрим пример рождения периодического решения при несогласованном лимитировании $(m-1)$ -го порядка в непрерывно культивируемой популяции делящихся клеток, имеющих m фаз развития. Динамика численности такой популяции задается системой (1.2.1), где матрица A имеет вид (1.2.2),

$$P(s) = \frac{1}{m\sqrt{q}-1} \begin{bmatrix} p_1(s) & & 0 \\ & 1 & \\ & & \ddots \\ 0 & & & 1 \end{bmatrix}, \quad G = (0, 0, \dots, 0, g_n)^* \quad (2.2.10)$$

и $\nu = 1$. Эта система в окрестности стационарной точки приводится к виду (2.1.17).

При $a = r^m ((\cos \pi/m)^{-m} + 1)/(r - 1)$ собственные значения $\lambda_{1,2} = \pm j \frac{r}{r-1} \operatorname{tg} \pi/m$ являются мнимыми, а остальные лежат в левой полуплоскости. Собственный вектор, отвечающий λ_1 , ищется из системы уравнений

$$x_k - r x_{k+1} = -j r \operatorname{tg} \pi/m \quad (k = 1, \dots, m-1)$$

и имеет компоненты $x_1 = 1$, $x_2 = (1/r)(1 + j \operatorname{tg} \pi/m)$

$$x_k = (1/r)^{k-1} (1 + j \operatorname{tg} \pi/m)^{k-1} \quad (k = 3, \dots, m).$$

Приводя комплексные числа к канонической форме, получаем собственный вектор

$$S_0 = (x_k)_{k=1}^m \left[\frac{\cos^{k-1} \pi/m}{r^{k-1}} \left(\cos \left(\frac{\pi}{m} (k-1) \right) - j \sin \left(\frac{\pi}{m} (k-1) \right) \right) \right]_{k=1}^m$$

Отсюда

$$\operatorname{Re} S_0 = \left[\frac{\cos^{k-1} \pi/m}{r^{k-1}} \cos \left(\frac{\pi}{m} (k-1) \right) \right]_{k=1}^m,$$

$$-\operatorname{Im} S_0 = \left[\frac{\cos^{k-1} \pi/m}{r^{k-1}} \sin \left(\frac{\pi}{m} (k-1) \right) \right]_{k=1}^m.$$

Любое вещественное периодическое решение линеаризованной системы имеет вид

$$X_\omega(t) = c(S_0 e^{j\omega t} + \bar{S}_0 e^{-j\omega t}) = 2c(\operatorname{Re} S_0 \cos \omega t - \operatorname{Im} S_0 \sin \omega t) =$$

$$= 2c((\cos \pi/m)^{k-1} \cos(\omega t - \pi(k-1)/m))_{k=1}^m,$$

где c — некоторая константа и $\omega = r/(r-1) \operatorname{tg} \pi/m$. Следовательно [21], периодическое решение системы (2.1.17) при малых значениях бифуркационного параметра задается выражением $X_0 + \epsilon X_\omega(t) + o(\epsilon)$. Формула доказана.

Сформулируем биологический смысл полученных результатов. Из формул (1.2.2) и (2.2.10) следует, что общая продолжительность жизненного цикла равна $m(r-1)$. Отсюда заключаем, что в популяции делящихся клеток, имеющих достаточно большое число фаз развития m , отношение периода колебаний к общей продолжительности жизненного цикла равно двум. На рис. 2.5 представлено отношение периода колебаний T к продолжительности жизненного цикла клетки \bar{T} при разном числе фаз развития и разном количестве потомков.

Сделаем некоторые выводы. Если сообщество или популяция является источником биомассы, то чаще всего оптимальным оказывается стационарный режим эксплуатации. Совершенно другая картина наблюдается в случае, когда организмы выращиваются для переработки субстрата в некоторый продукт. Может случиться, что необходимой ферментативной активностью обладают клетки лишь в определенной фазе развития. Например, популяция дрожжей *Schizosaccharomyces pombe* проявляла периодичность в гликолитической активности с максимумом во второй трети жизненного цикла [30]. С помощью синхронной популяции дрожжей *Sach. Cerevisiae* показана периодичность в синтезе трех ферментов [30] и т.д. Поскольку при малой численности клеток в фазе наибольшей активности возможны большие потери субстрата, непрерывный процесс культивирования оказывается неприемлемым. Используется периодический процесс культивирования, на который накладываются различные синхронизирующие воздействия. Такой метод трудоемок, малопроизводителен и слабо поддается автоматизации [30].

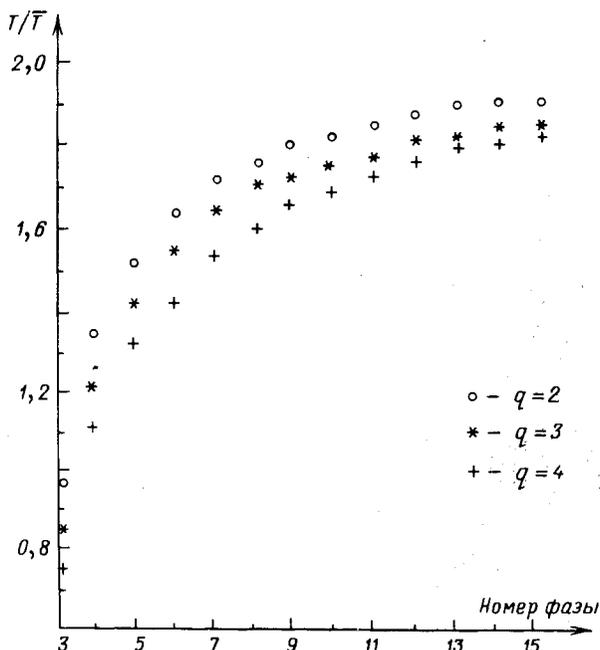


Рис. 2.5. Отношение периода колебаний численности клеточной популяции к продолжительности жизненного цикла клетки

Предположим, что с целью переработки некоторого вещества в конечный продукт реализована двухступенчатая система непрерывного культивирования. В первом реакторе клетки размножаются, а во втором они не размножаются и используются для переработки некоторого субстрата в конечный продукт. Пусть в первом реакторе реализуется приведенный выше периодический режим и способностью к переработке субстрата, подаваемого во второй реактор, обладают лишь клетки k -й фазы развития. Из приведенного примера видно, что численности $(k - p)$ - и k -фаз развития колеблются со сдвигом по фазе на $\pi p/m$. Поэтому с целью экономии субстрата имеет смысл подавать его во второй реактор импульсно со скважностью T в моменты времени, когда концентрация клеток k -й фазы развития близка к максимальной. Задача управления процессом в первом ферментере сводится к выбору режимов культивирования, при которых амплитуда колебаний максимальна и эти колебания устойчивы.

2.2.2. Устойчивость периодических решений. Теорема Хопфа гарантирует существование семейства периодических решений уравнений (2.1.2) и (2.1.3) вида $X(t, \mu(\epsilon))$, где $\mu(\epsilon)$ — непрерывная функция малого параметра ϵ . Если функция $\mu(\epsilon)$ принимает постоянное значение, то результат не допускает практической интерпретации и является вырожденным. Это связано с тем, что эксперимент любой степени точности позволяет указать лишь некоторый диапазон значений параметров модели. Следовательно, наряду с фактом существования периодического решения следует установить структурную устойчивость этого решения и прежде всего непрерывность значений функции $\mu(\epsilon)$. Однако даже в этом случае практический

вывод состоит лишь в том, что динамика реального объекта в окрестности стационарной точки имеет колебательный характер. Для того чтобы говорить о существовании периодических колебаний реального объекта, необходимо установить асимптотическую устойчивость периодического решения.

В случае, когда при вариации бифуркационного параметра мнимую ось пересекает лишь пара собственных значений матрицы линейного приближения, а остальная часть ее спектра лежит в левой полуплоскости, разработан алгоритм решения этой задачи. Он состоит в определении устойчивости стационарного решения системы при таких значениях бифуркационного параметра, когда матрица линейного приближения имеет пару мнимых корней. С этой целью в собственном подпространстве мнимой части спектра матрицы линейного приближения вводят полярные координаты r и θ с центром в стационарной точке. Затем в этом подпространстве вблизи стационарной точки фиксируют некоторую точку r_0, θ_0 и строят решение с этим начальным условием. Пусть $r(2\pi + \theta_0, r_0), 2\pi + \theta$ – фазовые координаты проекции полученного решения на подпространство, вычисленные при таком значении аргумента t , когда переменная θ возросла на 2π . Коэффициенты разложения функции последования $r(2\pi + \theta_0, r_0) - r$ по степеням r_0 отличны от нуля лишь при нечетных степенях: $r(2\pi + \theta_0, r_0) = l_1 r_0^3 + l_2 r_0^5 + \dots + l_p r_0^{2p+1}$ и носят название ляпуновских величин [19].

Если первая отличная от нуля величина $l_p < 0$, то периодическое решение асимптотически устойчиво; если $l_p > 0$, то неустойчиво и в обоих случаях результат существования периодического решения структурно устойчив, так как $\mu(\epsilon)$ принимает значения из некоторого интервала [21]. Частота колебаний с точностью до малых величин не зависит от ляпуновской величины и приближенно равняется частоте колебаний линеаризованной системы ω .

Алгоритм вычисления ляпуновской величины состоит в проведении последовательной нелинейной замены переменных, разложении правых частей уравнения в ряды Тейлора и вычислении коэффициентов при одинаковых степенях [20, 21]. Ввиду большой трудоемкости и чрезвычайной громоздкости вычислений к настоящему времени для уравнений общего вида вычислить первую ляпуновскую величину удалось лишь для систем не выше четвертого порядка [21]. Для систем произвольной размерности это удалось сделать лишь в некоторых специальных случаях [21] либо получить алгебраическое представление l_1 , зависящее от величин, алгоритм вычисления которых также чрезвычайно трудоемок [21]. Величина l_1 вычисляется через производные правой части уравнения до третьих степеней включительно и условие $l_1 \neq 0$ является "типичным" в S^3 -топологии, так как сколь угодно малые возмущения уравнений в этой топологии приводят к условию $l_1 \neq 0$. Поэтому для практических целей знания знака l_1 достаточно. Формулы, приведенные в работе [21], позволяют выписать характеристики периодического решения в терминах ляпуновской величины. Воспользуемся ими в наших исследованиях.

Согласно предыдущему пункту правая часть спектра матрицы линейного приближения уравнений (2.1.2) и (2.1.3) имеет вид $\alpha(\mu) + j\beta(\mu)$, где $\alpha(1) = 0, \beta(1) = \omega$ и $\alpha'(1) = \nu > 0$. Значения бифуркационного параметра μ

можно представить в виде функции вспомогательного малого параметра ϵ по формуле $\mu = 1 - (l_1/\nu T)\epsilon^2$, где $T = 2\pi/\omega + o(\epsilon)$ – период колебаний. Отсюда видно, что в случае устойчивости ($l_1 < 0$) бифуркация рождения цикла происходит при значениях $\mu > 1$. Допустим, что бифуркационный параметр принимает такое значение, при котором периодическое решение существует. В соответствии с первым методом Ляпунова линеаризуем исходную систему в окрестности этого периодического решения. В результате получим линейную систему с периодическими коэффициентами. Путем линейной замены переменных с периодическими коэффициентами она приводится к линейной системе с постоянными коэффициентами [3]. Собственные значения фундаментальной матрицы решений этой системы, вычисленные при $t = T$, где T – период колебаний, называются мультипликаторами линейной системы с периодическими коэффициентами. Таким образом, доказательство устойчивости периодического решения сводится к доказательству того факта, что все мультипликаторы, за исключением одного, который равен единице, меньше единицы [20]. Если $\lambda_1, \lambda_2, \dots, \lambda_{m-2}$ – собственные значения матрицы линейного приближения уравнения (2.1.2) или (2.1.3), вычисленные при $\mu = 1$, то мультипликаторами периодического решения являются величины $\exp(\lambda_i T + o(1))$ ($i = 1, \dots, m-2$), $1 + l_1 \epsilon + o(\epsilon)$ и единица. Отсюда видно, что при $\text{Re } \lambda_i < 0$ ($i = 1, \dots, m-2$) и $l_1 < 0$ периодическое решение, полученное в результате бифуркации, асимптотически устойчиво. Амплитуда колебаний r равна $r = \epsilon + o(\epsilon)$. Фазовый портрет системы в окрестности стационарной точки и амплитуда колебаний как функции параметра μ представлены на рис. 2.6, где a – случай $l_1 < 0$ и b – случай $l_1 > 0$.

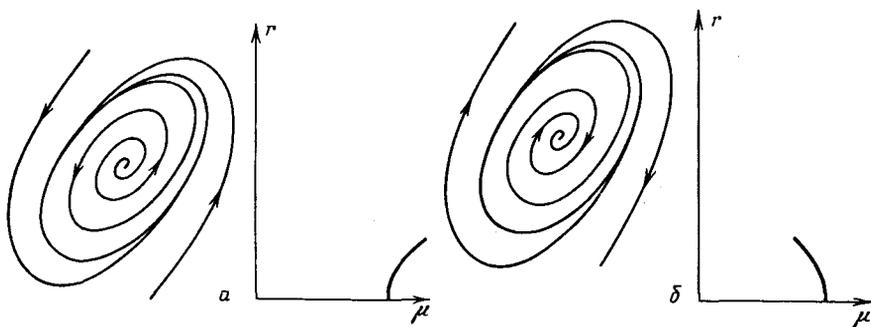


Рис. 2.6. Фазовый портрет системы в окрестности стационарной точки и функция зависимости амплитуды колебаний от величины бифуркационного параметра при (а) отрицательной и (б) положительной ляпуновской величине

Приведем вид первой ляпуновской величины для уравнений (2.1.2) и (2.1.3) в предположении, что оператор, задающий динамику воспроизводства популяции, и функция предпочтения имеют вид, использованный в уравнениях динамики культивируемых популяций из п. 2.1.5. Таким образом, исследуется уравнение вида

$$\dot{X} = P(s_\infty - C^* X) A X - \nu X + B, \quad (2.2.11)$$

где s_∞ – некоторая положительная константа, C – ненулевой неотрицатель-

ный вектор, B – постоянный вектор, который может иметь нулевые компоненты. К виду (2.2.11) приводится уравнение динамики культивируемой популяции при однопоточном (уравнение (2.1.17)) и многопоточном (любое уравнение системы (2.1.21)) культивировании. Сделаем еще одно упрощающее предположение. Будем считать, что лимитирующий фактор влияет лишь на одну компоненту биологической системы, т.е. $p_j(s) = 1$ при $j \neq i$ и $p_i = p_i(s)$ – монотонно возрастающая функция. Стационарное решение по-прежнему будем обозначать через X_0 и через $s_0 = s_\infty = C^* X_0$. Не умаляя общности, будем предполагать, что $p_i(s_0) = 1$. С помощью замены переменной $Y = X - X_0$ и представления $P(s) = p_i(s)EE^*$, где $E^* = (0, \dots, 1, 0, \dots, 0)$, уравнение (2.2.11) приводится к нелинейной системе

прямого регулирования вида

$$\dot{Y} = DY + \psi_0(\sigma_0)(1 + \psi_1(\sigma_1))E,$$

где $\sigma_0 = C^* Y$, $\sigma_1 = R^* Y$, $\psi_0(\sigma_0)$ и $\psi_1(\sigma_1)$ – скалярные функции скалярного аргумента $\psi_0(0) = \psi_1(0) = 0$, $Y, C, R, E \in R^m$ и D – квадратная матрица. Явный вид ляпуновской величины для системы прямого регулирования при $\psi_1 \equiv 0$ получен А.И. Лурье [19]. Нами разработана модификация метода, пригодная для исследования приведенного выше уравнения [11].

Введем вспомогательные обозначения. Обозначим i -ю компоненту вектора AX_0 через ρ_0 ($\rho_0 = (AX_0)_i$) и через

$$\rho_1 = \frac{\partial}{\partial s} p_i(s),$$

$$\rho_2 = \frac{1}{2} \frac{\partial^2}{\partial s^2} p_i(s_0) \quad \text{и} \quad \rho_3 = \frac{1}{6} \frac{\partial^3}{\partial s^3} p_i(s_0).$$

Матрицу линейного приближения, вычисленную по формуле (2.2.7) при $\mu = 1$, будем обозначать через D . Введем линейные формы, определенные на векторах $Y \in R^m$ следующими формулами:

$$\sigma_0 = C^* Y, \quad \sigma_1 = (\rho_2 - \rho_1^2)\rho_0\sigma_0 + \rho_1 E^*(D + uI)Y,$$

$$\sigma_2 = (\rho_3 - \rho_2\rho_1)\rho_0\sigma_0 + \rho_2 E^*(D + uI)Y.$$

Исходной системе (2.2.11) сопоставим систему управления вида

$$\dot{Y} = DY + Eu(t), \tag{2.2.12}$$

где $E = (0, \dots, 1, \dots, 0)$ и $u(t)$ – скалярная функция скалярного аргумента.

По-прежнему будем предполагать, что матрица D имеет пару мнимых собственных значений $\pm j\omega$ и остальная часть ее спектра лежит в левой полуплоскости. Обозначим через w_k передаточную функцию системы (2.2.12) при наблюдении по сигналу $\sigma_k(Y)$:

$$w_k(s) = \sigma_k((sI - D)^{-1}E) \quad (k = 0, 1, 2).$$

Вычислим константы

$$a_k = 2 \lim_{s \rightarrow j\omega} \operatorname{Re}(s - j\omega) w_k(s),$$

$$b_k = -2 \lim_{s \rightarrow j\omega} \operatorname{Im}(s - j\omega) w_k(s)$$

при $s \rightarrow j\omega$ и $k = 0, 1, 2$. Обозначим

$$\begin{aligned}\alpha_k &= (a_0 a_1 + b_0 b_1) (w_k(0) - b_k/\omega)/2, \\ \beta_k &= (a_0 a_1 - b_0 b_1) (\operatorname{Re} w_k(2j\omega) + b_k/3\omega)/2 + \\ &+ (a_1 b_0 + a_0 b_1) (\operatorname{Im} w_k(2j\omega) + 2a_k/3\omega), \\ \gamma_k &= -(a_1 a_0 - b_1 b_0) (\operatorname{Im} w_k(2j\omega) + 2a_k/3\omega)/2 + \\ &+ (a_1 b_0 + a_0 b_1) (\operatorname{Re} w_k(2j\omega) + b_k/3\omega),\end{aligned}$$

где $k = 0, 1, 2$.

Теорема 2.2.1. Пусть пара (D, E) управляема ($\det(E, DE, \dots, D^{m-1}E) \neq 0$) и матрица D несобственным линейным преобразованием приводима к диагональному виду. Тогда первая ляпуновская величина задается формулой

$$\begin{aligned}\frac{1}{2\pi} l_1 &= \frac{1}{8\omega^2} [a_1 b_1 (a_0^2 + b_0^2) + a_0 b_0 (a_1^2 + b_1^2)] + \\ &+ \frac{1}{2\omega} \left[a_1 \alpha_0 + a_1 \frac{\beta_0}{2} + b_1 \frac{\gamma_0}{2} + a_0 \alpha_1 + a_0 \frac{\beta_1}{2} + b_0 \frac{\gamma_1}{2} \right] + \\ &+ \frac{1}{8\omega} [3a_0^2 a_2 + b_0^2 a_2 + 2a_0 b_0 b_2].\end{aligned}\quad (2.2.13)$$

Если $l_1 < 0$, то периодическое решение уравнения (2.2.11), лежащее в малой окрестности стационарной точки, устойчиво; если $l_1 > 0$, то неустойчиво.

Доказательство. Проведем разложение правой части уравнения (2.2.11) в окрестности стационарной точки до полиномов третьей степени включительно. Вначале с помощью линейной замены переменных упростим линейную часть этого разложения.

Перенесем начало координат в нетривиальную стационарную точку с помощью сдвига $X - X_0 = Y$. Имеем

$$\begin{aligned}\dot{Y} &= DY + \left[\sigma_0 \frac{\partial P}{\partial s} AY + \frac{\sigma_0^2}{2} \frac{\partial^2 P}{\partial s^2} AX_0 \right] + \\ &+ \left[\frac{\sigma_0^2}{2} \frac{\partial^2 P}{\partial s^2} AY + \frac{\sigma_0^2}{6} \frac{\partial^3 P}{\partial s^3} AX_0 \right] + \dots,\end{aligned}$$

где $\sigma_0 = C^* Y$, $D = A - \nu I + \frac{\partial P}{\partial s} AX_0 C^*$ и, следовательно, $A = D + \nu I - \frac{\partial P}{\partial s} AX_0 C^*$. Группируя члены при одинаковых степенях, окончательно получаем

$$\begin{aligned}\dot{Y} &= DY + [\rho_2 \rho_1] \rho_0 \sigma_0^2 E + \rho_1 \sigma_0 J (D + \nu I) Y + \\ &+ [\rho_3 - \rho_2 \rho_1] \rho_0 \sigma_0^3 E + \rho_2 \sigma_0^2 J (D + \nu I) Y + \dots,\end{aligned}$$

где $J = EE^*$.

Не умаляя общности, предположим, что порядок системы равен $n + 2$ и $j\omega, -j\omega, \mu_t$ ($t = 1, \dots, n$) — корни характеристического полинома матрицы D . По условию $\omega \neq 0$ и $\operatorname{Re} \mu_t < 0$. Пусть S_0, S_1, \dots, S_n ($S_0^-, S_1^-, \dots, S_n^-$) — собственные векторы матрицы $D(D^*)$, отвечающие собственным значениям $j\omega, -j\omega, \mu_t$ ($t = 1, \dots, n$) соответственно. Образует матрицы

$$S = [\operatorname{Re} S_0, -\operatorname{Im} S_0, S_1, \dots, S_n] \text{ и } S^{-1} = [\operatorname{Re} S_0^-, \operatorname{Im} S_0^-, S_1^-, \dots, S_n^-]^*.$$

Воспользовавшись предположением об управляемости пары (D, E) , выберем векторы S_0^-, \dots, S_n^- так, чтобы $(S_0^-)^* E = (S_n^-)^* E = 1$. Тогда $S^{-1} E = (1, 0, 1, \dots, 1)$. Сделаем замену

переменных $Y = SZ$, где $Z = (x, y, z_1, \dots, z_n)^*$. Получаем

$$\begin{aligned}x &= -\omega y + (\rho_2 - \rho_2^2)\rho_0\sigma_0^2 + \rho_1\sigma_0\xi + (\rho_3 - \rho_2\rho_1)\rho_0\sigma_0^3 + \\&+ (\rho_3 - \rho_2\rho_1)\rho_0\sigma_0^3 + \rho_2\sigma_0^2\xi + o(\|Z\|^3), \\ \dot{y} &= \omega x + o(\|Z\|^3), \\ \dot{z}_t &= \mu_t z_t + (\rho_2 - \rho_1^2)\rho_0\sigma_0^2 + \rho_1\sigma_0\xi + (\rho_3 - \rho_2\rho_1)\rho_0\sigma_0^3 + \rho_2\sigma_0^2\xi + \\&+ o(\|Z\|^3),\end{aligned}$$

где $o(\|Z\|^3)/\|Z\|^3 \rightarrow 0$ при $\|Z\| \rightarrow 0$, $t = 1, \dots, n$, $\xi = E^*(D + \nu J)SZ$, $\sigma_0 = C^*SZ$, $\rho_0 E = AX_0$. При выводе формул мы воспользовались тем, что $J = EE$. Напомним обозначения

$$\sigma_1 = (\rho_2 - \rho_1^2)\rho_0\sigma_0 + \rho_1\xi, \quad \sigma_2 = (\rho_3 - \rho_2\rho_1)\rho_0\sigma_0 + \rho_2\xi.$$

Тогда

$$\begin{aligned}\dot{x} &= -\omega y + \sigma_0\sigma_1 + \sigma_0^2\sigma_2 + \dots, \\ \dot{y} &= \omega x + \dots, \\ \dot{z}_t &= \mu_t z_t + \sigma_0\sigma_1 + \sigma_0^2\sigma_2 + \dots,\end{aligned}\tag{2.2.14}$$

Таким образом, в первых двух уравнениях полученной системы нет линейных членов от переменных z_t ($t = 1, \dots, n$).

Второй этап доказательства состоит в том, чтобы избавиться в правой части этих уравнений от полиномов второй степени, содержащих переменные z_t ($t = 1, \dots, n$). Делаем замену переменных $z_t = v_t(x, y) + \xi_t$, где $v_t = M_t x^2 + 2P_t xy + N_t y^2$ и коэффициенты квадратного многочлена выбраны в соответствии с условиями [19, с. 196, 202]

$$\frac{\partial v_t}{\partial x} \dot{x} + \frac{\partial v_t}{\partial y} \dot{y} = \dot{z}_t(v_t) + o(x^2 + y^2),$$

где $o(x^2 + y^2)$ – ряд от x и y степени выше двух. Таким образом, имеем

$$\begin{aligned}-2(M_t x + P_t y)\omega y + 2(P_t x + N_t y)\omega x &= \\ = \mu_t(M_t x^2 + 2P_t xy + N_t y^2) + \sigma_0\sigma_1.\end{aligned}$$

Введем обозначения $\sigma_i = a_i x + b_i y + \sum_{t=1}^n c_t^i z_t$ ($i = 0, 1, 2$). Отсюда получаем

$$\begin{aligned}-\mu_t M_t + 2\omega P_t &= a_0 a_1, \\ -\mu_t N_t - 2\omega P_t &= b_0 b_1, \\ -\omega M_t + \omega N_t - \mu_t P_t &= (a_1 b_0 + a_0 b_1)/2,\end{aligned}$$

откуда находим

$$\begin{aligned}M_t &= -\frac{a_0 a_1 + b_0 b_1}{2\mu_t} - \frac{1}{\mu_t^2 + 4\omega^2} \left[\frac{1}{2}\mu_t(a_0 a_1 - b_0 b_1) + \omega(a_1 b_0 + a_0 b_1) \right], \\ N_t &= -\frac{a_0 a_1 + b_0 b_1}{2\mu_t} - \frac{1}{\mu_t^2 + 4\omega^2} \left[\frac{1}{2}\mu_t(a_0 a_1 - b_0 b_1) + \omega(a_1 b_0 + a_0 b_1) \right], \\ P_t &= \frac{\omega}{\mu_t^2 + 4\omega^2} [(a_0 a_1 - b_0 b_1) - \mu_t(a_1 b_0 + a_0 b_1)/2].\end{aligned}\tag{2.2.15}$$

По построению разложения правой части уравнения $\dot{\xi}_t = \mu_t \xi_t + \dots$ ($\xi_t = z_t - v_t(x, y)$) ($t = 1, \dots, n$) по x и y при условии $\xi_t = 0$ ($t = 1, \dots, n$) начинается членами не менее третьего порядка. В связи с этим при исследовании устойчивости в малом ограничимся первыми двумя уравнениями системы (2.1.14), в которых следует положить $z_t = v_t(x, y)$ ($t = 1, \dots, n$) [20]. Сделаем в системе (2.2.14) подстановку $z_t = v_t(x, y)$.

Для первых двух уравнений получаем

$$\begin{aligned}\dot{x} &= -\omega y + (a_0 x + b_0 y + R_0 x^2 + 2Q_0 xy + T_0 y^2)(a_1 x + \\ &+ b_1 y + R_1 x^2 + 2Q_1 xy + T_1 y^2) + (a_0 x + b_0 y)^2 (a_2 x + b_2 y) + \dots, \\ \dot{y} &= \omega x + \dots,\end{aligned}$$

где использованы обозначения

$$\begin{aligned}\sigma_k &= a_k \dot{x}_k + b_k \dot{y}_k + \sum_{t=1}^n c_k^t z_k, \\ R_k &= \sum_{t=1}^n c_k^t M_k, \\ T_k &= \sum_{t=1}^n c_k^t N_t \text{ и } Q_k = \sum_{t=1}^n c_k^t P_t \quad (k=0, 1, 2).\end{aligned}\tag{2.2.17}$$

Перейдем к полярным координатам $x = r \cos \theta$, $y = r \sin \theta$. Имеем

$$R_k x^2 + 2Q_k xy + T_k y^2 = r^2 (\alpha_k + \beta_k \cos 2\theta + \gamma_k \sin 2\theta),$$

где

$$\alpha_k = \frac{R_k + T_k}{2}, \quad \beta_k = \frac{R_k - T_k}{2} \text{ и } \gamma_k = Q_k \quad (k=0, 1).$$

Система (2.2.16) преобразуется к виду

$$\begin{aligned}\frac{dr}{dt} &= (\sigma_0 \sigma_1 + \sigma_0^2 \sigma_2) \cos \theta + \dots, \\ \frac{d\theta}{dt} &= \omega - \frac{1}{r} [\sigma_0 \sigma_1 + \sigma_0^2 \sigma_2] \sin \theta + \dots\end{aligned}$$

или

$$\frac{dt}{d\theta} = \frac{1}{\omega} + \frac{1}{r\omega^2} (\sigma_0 \sigma_1 \sin \theta + \sigma_0^2 \sigma_2 \sin \theta) + \dots,$$

где в линейных формах σ_0 , σ_1 и σ_2 на месте переменных z_t ($t = 1, \dots, n$) стоят соответствующие квадратичные формы v_t . Вычислим величины $dr/d\theta$ и соберем члены, которые содержат r^3 . Имеем

$$\frac{1}{r\omega^2} \sigma_0^2 \sigma_1^2 \cos \theta \sin \theta + \frac{1}{\omega} [\sigma_0 \sigma_1 + \sigma_0^2 \sigma_2] \cos \theta.$$

Пусть g – коэффициент при r^3 . Раскрывая предыдущее выражение, имеем

$$\begin{aligned}g &= \frac{1}{\omega^2} (a_0 \cos \theta + b_0 \sin \theta)^2 (a_1 \cos \theta + b_1 \sin \theta)^2 \cos \theta \sin \theta + \\ &+ \frac{\cos \theta}{\omega} [(a_1 \cos \theta + b_1 \sin \theta) (\alpha_0 + \beta_0 \cos 2\theta + \gamma_0 \sin 2\theta) + \\ &+ (a_0 \cos \theta + b_0 \sin \theta) (\alpha_1 + \beta_1 \cos 2\theta + \gamma_1 \sin 2\theta) + \\ &+ (a_0 \cos \theta + b_0 \sin \theta)^2 (a_2 \cos \theta + b_2 \sin \theta)].\end{aligned}$$

Отсюда $\frac{1}{2\pi} \int_0^{2\pi} g d\theta$ задается формулой (2.2.13).

В выражении для ляпуновской величины фигурируют параметры

$$a_k, b_k \quad (k=0, 1, 2), \quad \alpha_k = \frac{R_k + T_k}{2}, \quad \beta_k = \frac{R_k - T_k}{2} \text{ и } \gamma_k = Q_k,$$

где R_k , T_k и Q_k вычисляются по формулам (2.2.16). Выразим эти параметры в явном

виде через коэффициенты исходной системы. Напомним, что

$$\sigma_0 = C^* S Z, \quad \xi = E^* (D + \nu I) S Z, \quad \sigma_1 = (\rho_1 - \rho_1^2) \rho_0 \sigma_0 + \rho_1 \xi, \\ \sigma_2 = (\rho_3 - \rho_2 \rho_1) \rho_0 \sigma_0 + \rho_2 \xi$$

и, кроме того, $\sigma_k = a_k x + b_k y + \sum_{t=1}^n c_k^t z_t$ ($k = 0, 1, 2$).

Воспользовавшись разложением матрицы по компонентам, представим резольвенту в виде

$$(\lambda I - D)^{-1} = S_0 (S_0^-)^* / (\lambda - j\omega) + \bar{S}_0 (\bar{S}_0^-)^* / (\lambda + j\omega) + \sum_{t=1}^n S_t (S_t^-)^* / (\lambda - \mu_t).$$

Напомним, что $(S_t^-)^* E = 1$ ($t = 1, \dots, n$) по построению. Отсюда приходим к следующему способу вычисления коэффициентов линейных форм. Вначале введем некоторые дополнительные обозначения:

$$\Delta(\lambda) = \det(\lambda I - D), \quad \Delta'(\lambda) - \text{производная от } \Delta(\lambda),$$

$$M_0(\lambda) = \Delta(\lambda) C^* (\lambda I - D)^{-1} E,$$

$$M_1(\lambda) = (\rho_2 - \rho_1^2) \rho_0 M_0(\lambda) + \rho_1 M_3(\lambda),$$

$$M_2(\lambda) = (\rho_3 - \rho_2 \rho_1) \rho_0 M_0(\lambda) + \rho_2 M_3(\lambda),$$

$$M_3(\lambda) = \Delta(\lambda) E^* (D + \nu I) (\lambda I - D)^{-1} E.$$

По свойству резольвенты имеем

$$\frac{1}{2} (a_k - j b_k) = \frac{M_k(j\omega)}{\Delta'(j\omega)} \quad (k = 0, 1, 2)$$

и, следовательно,

$$a_k = 2 \operatorname{Re} \frac{M_k(j\omega)}{\Delta'(j\omega)}, \quad b_k = -2 \operatorname{Im} \frac{M_k(j\omega)}{\Delta'(j\omega)} \quad (k = 0, 1, 2),$$

$$c_k^t = \frac{M_k(\mu_t)}{\Delta'(\mu_t)}.$$

Перейдем теперь к вычислению величин R_k , T_k и Q_k ($k = 0, 1$). По формулам (2.2.15) и (2.2.17) получаем

$$R_k = (a_0 a_1 + b_0 b_1) s_k^1 / 2 - (a_0 a_1 - b_0 b_1) s_k^2 / 2 - \omega (a_1 b_0 + a_0 b_1) s_k^3,$$

$$T_k = (a_0 a_1 + b_0 b_1) s_k^1 / 2 + (a_0 a_1 - b_0 b_1) s_k^2 / 2 + \omega (a_1 b_0 + a_0 b_1) s_k^3, \quad (2.2.18)$$

$$Q_k = \omega (a_0 a_1 - b_0 b_1) s_k^3 - \omega (a_1 b_0 + a_0 b_1) / 2,$$

где

$$s_k^1 = - \sum_{t=1}^n \frac{M_k(\mu_t)}{\mu_t \Delta'(\mu_t)},$$

$$s_k^2 = \sum_{t=1}^n \frac{\mu_t M_k(\mu_t)}{\Delta'(\mu_t) (\mu_t^2 + 4\omega^2)},$$

$$s_k^3 = \sum_{t=1}^n M_k(\mu_t) / (\Delta'(\mu_t) (\mu_t^2 + 4\omega^2)) \quad (k = 0, 1, 2).$$

Вычислим эти суммы. Воспользуемся тождествами [19]

$$\sum_{t=1}^n \frac{M_k(\mu_t)}{\mu_t \Delta'(\mu_t)} = \frac{n+2}{\sum_{t=1}^n \mu_t \Delta'(\mu_t)} - \frac{1}{j\omega} \left[\frac{M_k(j\omega)}{\Delta'(j\omega)} - \frac{M(-j\omega)}{\Delta'(-j\omega)} \right],$$

$$\sum_{\rho=1}^{n+2} \frac{1}{\lambda_\rho \Delta'(\lambda_\rho)} = - \frac{1}{\Delta(0)},$$

$$\sum_{\rho=1}^{n+2} \frac{1}{\lambda_\rho^3 \Delta'(\lambda_\rho)} = \frac{\Delta'(0)}{[\Delta(0)]^2},$$

где $\lambda_1 = j\omega$, $\lambda_2 = -j\omega$, $\lambda_{\rho+2} = \mu_t$ ($t = 1, \dots, n$). Отсюда находим

$$\sum_{t=1}^n \frac{M_k(\mu_t)}{\mu_t \Delta'(\mu_t)} = -\frac{M_k(0)}{\Delta(0)} + \frac{1}{\omega} b_k.$$

Далее,

$$\begin{aligned} \sum_{t=1}^n \frac{M_k(\mu_t)}{\Delta'(\mu_t)(\mu_t^2 + 4\omega^2)} &= \sum_{t=1}^n \frac{M_k(\mu_t)}{\Delta'(\mu_t)(\mu_t^2 + 4\omega^2)} - \frac{1}{3\omega^2} a_k = \\ &= \frac{1}{4\omega j} \sum_{t=1}^n \frac{M_k(\mu_t)}{\Delta'(\mu_t)} \left(\frac{1}{\mu_t - 2\omega j} - \frac{1}{\mu_t + 2\omega j} \right) - \frac{1}{3\omega^2} a_k = \\ &= \frac{-1}{4\omega j} \left[\frac{M_k(2\omega j)}{\Delta(2\omega j)} - \frac{M_k(-2\omega j)}{\Delta(-2\omega j)} \right] - \frac{1}{3\omega^2} a_k = -\frac{1}{2\omega} \operatorname{Im} \frac{M_k(2j\omega)}{\Delta(2j\omega)} - \frac{1}{3\omega^2} a_k. \end{aligned}$$

Мы воспользовались здесь формулой разложения рациональной функции на простейшие дроби:

$$\frac{M_k(\mu)}{\Delta(\mu)} = \sum_{t=1}^{n+2} \frac{M(\lambda_t)}{\Delta'(\lambda_t)} \frac{1}{\mu - \lambda_t},$$

где λ_t ($t = 1, \dots, n+2$) – простые корни многочлена $\Delta(\mu)$.

Аналогичным образом найдем

$$\sum_{t=1}^n \frac{\mu_t M_k(\mu_t)}{\Delta'(\mu_t)(\mu_t^2 + 4\omega^2)} = -\operatorname{Re} \frac{M_k(2j\omega)}{\Delta(2j\omega)} - \frac{b_k}{3\omega}.$$

В итоге получаем

$$\begin{aligned} s_k^1 &= M_k(0)/\Delta(0) - b_k/\omega, \\ s_k^2 &= -\operatorname{Re}(M_k(2j\omega)/\Delta(2j\omega)) - b_k/3, \\ s_k^3 &= \operatorname{Im}(M_k(2j\omega)/\Delta(2j\omega))/2\omega - a_k/(3\omega^2). \end{aligned} \quad (2.2.19)$$

Для упрощения формул введем вспомогательные функции

$$w_k(s) = M_k(s)/\Delta(s) \quad (k = 0, 1, 2).$$

Очевидно, что

$$a_k = 2 \lim_{s \rightarrow j\omega} \operatorname{Re}(s - j\omega) w_k(s)$$

и

$$b_k = -2 \lim_{s \rightarrow j\omega} \operatorname{Im}(s - j\omega) w_k(s) \quad (k = 0, 1).$$

Отсюда и из формул (2.2.18) и (2.2.19) получаем окончательные выражения для вычисления параметров ляпуновской величины. Теорема доказана.

Приведем формулы, удобные для вычисления передаточных функций $w_k(s)$ ($k = 0, 1, 2$). Обозначим $\Delta_A(s) = \det \{sI - A + vI\}$, $\Delta(s) = \det \{sI - D\}$. Тогда

$$\begin{aligned} w_0(s) &= (\Delta(s) - \Delta_A(s))/(\Delta(s)\rho_1\rho_0), \\ w_1(s) &= (\rho_2 - \rho_1^2)\rho_0 w_0(s) + \rho_1 w_3(s), \\ w_2(s) &= (\rho_3 - \rho_1\rho_2)\rho_0 w_0(s) + \rho_2 w_3(s). \end{aligned} \quad (2.2.20)$$

Здесь $w_3(s) = (s+1)\Delta_{ii}/\Delta(s) - 1$, Δ_{ii} – алгебраическое дополнение i -го элемента на главной диагонали матрицы.

Доказательство. При доказательствах теоремы 2.1.6 для произвольных матриц D и N , произвольных векторов X и C и скаляра a , удовлетворяющих равен-

ству $M = N - aXC^*$, установлено детерминантное соотношение

$$\det(\lambda I - M) = (1 + aC^*(\lambda I - N)^{-1}X) \det(\lambda I - N).$$

Напомним, что $D = A - \nu I + \rho_0 \rho_1 EC^*$, или $D - \rho_0 \rho_1 EC^* = A - \nu I$. Полагая $N = D$ и $M = A - \nu I$, получаем первую формулу системы (2.2.20). Рассуждая аналогично в случае третьей формулы, имеем

$$w_3(s) = E^*(D + I)(sI - D)^{-1}E = \det\{sI - D + EE^*(D + I)\} / \Delta(s) - 1.$$

Остальное очевидно.

Подведем краткий итог полученным результатам. В первом параграфе данной главы мы показали, что в случае неустойчивости стационарного решения уравнения (2.1.2) или (2.1.3) поведение траекторий в окрестности стационарной точки носит колебательный характер. Существование периодических решений в этой окрестности установлено с помощью бифуркационной теоремы Хопфа. В данном параграфе рассмотрен важный частный случай, когда в цепи взаимодействий экологической системы можно выделить "узкое место", на которое влияет лимитирующий фактор. Для таких систем при условии линейной зависимости величины лимитирующего фактора от численности экосистемы получен аналитический критерий устойчивости периодического решения.

2.2.3. Существование колебаний в нелокальном случае. В первом пункте этого параграфа было показано, что при неустойчивости стационарного решения в моделях динамики численности эксплуатируемых экосистем локальное движение в окрестности стационарной точки носит колебательный характер. Введение малого параметра позволило установить факт рождения периодических решений, которые лежат достаточно близко от этой точки. Естественно возникает задача исследования условий рождения колебательных решений в нелокальном случае, когда введение малого бифуркационного параметра неоправданно. Ввиду большой сложности этой задачи ограничим наше рассмотрение моделью (2.1.17), определенной в инвариантном множестве $R_+^m \setminus 0$, считая, что предположения п. 2.1.5 о правой части уравнения (2.1.17) выполнены.

На протяжении этого пункта будем считать, что диссипативность решений уравнения (2.1.17) установлена (см. теоремы 2.1.10 и 2.1.17). Пусть Y_0 — нетривиальное стационарное решение и $s_0 = s_p - G^*Y_0$.

Предложение 2.2.6. Всякое решение уравнения (2.1.17), начинающееся в $R_+^m \setminus 0$, может находиться на конечном расстоянии от гиперплоскости $G^*Y = s_p - s_0$ лишь конечное время.

Доказательство. Допустим противное. Пусть $G^*Y \leq s_p - s_0 - \epsilon$. Тогда $s = s_p - G^*Y \geq s_0 + \epsilon$. Напомним, что уравнение (2.1.17) задано на положительно инвариантном множестве $Y \geq 0$ и $AY \geq 0$. Пусть L — собственный вектор матрицы $(P(s_0 + \epsilon)A)^*$, отвечающий собственному значению с максимальной вещественной частью. Тогда из уравнения (2.1.17) и из условия $AY \geq 0$ имеем

$$L^*Y = L^*(P(s)A - \nu I)Y \geq L^*(P(s_0 + \epsilon)AY - \nu I) = (\lambda(P(s_0 + \epsilon)A) - \nu I)L^*Y > 0.$$

Следовательно, $L^*Y(t)$ растет экспоненциально. Этот факт противоречит диссипативности уравнения (2.1.17). Предложение доказано.

Стационарная точка Y_0 принадлежит гиперплоскости $G^*Y = s_p - s_0$ и не является репеллером, т.е. существует многообразие ненулевой размерности W_s [45], являющееся инвариантным множеством уравнения (2.1.17) и такое, что всякое принадлежащее ему решение стремится к стационар-

ной точке. Действительно, так как след матрицы линейного приближения в силу предложения 2.2.5 неположителен и нет нулевых собственных значений, при неустойчивости должны существовать собственные значения с положительной и отрицательной вещественной частью.

Если стационарная точка неустойчива, то размерность этого многообразия меньше размерности фазового пространства уравнения и движения, не принадлежащие W^s , обладают колебательными свойствами. Точнее, имеет место [8]

Предложение 2.2.7. Если стационарная точка Y_0 неустойчива и решение $Y(t)$ из $R_+^m \setminus 0$ не стремится к Y_0 при $t \rightarrow \infty$, то число перемен знака скалярной функции $G^*Y(t) - G^*Y_0$ стремится к бесконечности при $t \rightarrow \infty$.

Доказательство. В силу предложения 2.2.6 имеет место одно из условий: $G^*Y(t) \rightarrow G^*Y_0$ при $t \rightarrow \infty$ или функция $G^*Y(t) - G^*Y_0$ меняет знак на любом интервале времени $[T, \infty)$, где T — сколь угодно большое число и $Y(0) \in R_+^m \setminus 0$. Если имеет место первое условие и $Y(t) \neq Y_0$, то множество ω -предельных точек решения $Y(t)$ принадлежит гиперплоскости $G^*Y = s_p - s_0$ и не совпадает с Y_0 . Пусть $Z(t) \neq Y_0$ — такое решение уравнения (2.1.17), которое принадлежит этому множеству ω -предельных точек. Тогда

$$G^*\dot{Z}(t) = G^*(P(s_0)A - \nu I)Z(t) = 0$$

при $t > 0$. В силу единственности положительного собственного вектора имеем $G^*(P(s_0)A - \nu I) = 0$. Тогда по теореме 2.1.14 стационарная точка устойчива. Таким образом, имеет место второе условие. Предложение доказано.

Практическое значение данного результата сводится к тому, что в популяции с возрастной структурой, размножающейся при стационарных внешних условиях и отсутствии каких-либо периодических возмущений, суммарная численность может колебаться около некоторого постоянного уровня. При этом амплитуда колебаний не убывает со временем и может быть достаточно большой. Причиной таких колебаний является существенное различие в реакциях возрастных групп на недостаток источников питания и ограниченности жизненного пространства.

Сформулируем результат о существовании периодического решения в строгом смысле этого слова для модели динамики численности с возрастной или фазовой структурой. Идея доказательства существования периодического решения состоит в построении выпуклого инвариантного компактного множества D , которое фазовым потоком уравнения (2.1.17) отображается в себя [29]. При построении этого множества решающую роль будет играть результат предложения 2.2.7. Неподвижная точка отображения множества D в себя, отличная от начала координат и стационарного решения Y_0 , будет задавать начальное условие для искомого периодического движения. Существование периодического движения будет следствием соответствующего варианта теоремы о неподвижной точке [41, 49, 50].

Введем некоторые обозначения. Через $U(D)$ будет обозначать некоторую окрестность множества D , через H — гиперплоскость $G^*Y = G^*Y_0$. Пользуясь тем, что положительно инвариантным множеством уравнения (2.1.17) является конус $\{Y \in R_+^m \mid AY \geq 0\}$, который содержит внутри себя стационарное решение Y_0 , все условия в формулируемой ниже теореме будем считать выполненными лишь на этом конусе.

Теорема 2.2.2. Пусть стационарная точка Y_0 — гиперболическая [1] неустойчивая. Предположим также, что существуют положительно инвариант-

ное компактное выпуклое множество D и замкнутое множество Q с кусочно-гладкой границей такие, что

1. Точка Y_0 лежит внутри D , т.е. $Y_0 \in \overset{\circ}{D}$.

2. Устойчивое многообразие W^s лежит внутри Q , т.е.

$$W^s \cap U(D) \subset \overset{\circ}{Q}.$$

3. Часть гиперплоскости H , которая лежит внутри $U(D)$ и не пересекается с Q , имеет две компоненты связности H_1 и H_2 , одна из которых является выпуклым множеством, т.е.

$$(U(D) \setminus Q) \cap H = H_1 \cup H_2 \quad (H_1 \cap H_2 = \emptyset).$$

4. Векторное поле пересекает границу множества $Q \cap U(D)$ изнутри наружу.

5. Векторное поле пересекает H_1 и H_2 трансверсально.

Тогда уравнение (2.1.17) имеет периодическое решение ненулевого периода.

Доказательство. В силу замкнутости $Q \in Y_0$ множества H_1 и H_2 являются замкнутыми выпуклыми и пересекаются лишь в точке Y_0 . В силу неустойчивости стационарной точки и пункта б) теоремы 2.1.14 гиперплоскость H не является инвариантным множеством уравнения (2.1.17). Всякая точка, начинающаяся на H_1 должна по предложению 2.2.7 перейти в полупространство $G^*Y > G^*Y_0$ ($G^*Y < G^*Y_0$) и, побыв там конечное время, вернуться в полупространство $G^*Y < G^*Y_0$ ($G^*Y > G^*Y_0$). В силу трансверсальности векторного поля на H_1 и H_2 и связности этих множеств векторное поле пересекает каждое из них в одном направлении. Поскольку $D \setminus Q$ отображается фазовым потоком уравнения (2.1.17) в открытое множество $U(D) \setminus Q$, всякое решение, начавшееся на $H_1 \setminus Y_0$, через конечное время должно пересечься с H_2 в силу односторонней направленности (трансверсальности) векторного поля на этих множествах. Аналогично, решение, начавшееся на H_2 , через конечное время пересечет H_1 . Таким образом, имеется отображение $H_1 \setminus Y_0$ в H_1 . В силу трансверсальности это отображение непрерывно, и так как Y_0 — неподвижная точка, оно доопределяется до непрерывного отображения \bar{H}_1 в себя. Так как по теореме 2.1.9 всякое решение, начинающееся в достаточно малой окрестности нуля, покидает эту окрестность, то $0 \in H_1$ является изгоняющей точкой. Решение из малой окрестности точки Y_0 может приблизиться или оставаться вблизи точки Y_0 , если оно принадлежит многообразию W^s . Поскольку это многообразие не пересекается с H_1 и точка Y_0 неустойчива, решения из малой окрестности Y_0 , принадлежащие H_1 , покинут эту окрестность за конечное время, т.е. Y_0 — тоже изгоняющая точка. Следовательно, существует неизгоняющая неподвижная точка отображения \bar{H}_1 в себя. Эта точка является начальным условием искомого периодического решения. Теорема доказана.

Пример. Рассмотрим динамику эксплуатируемой популяции, когда число возрастных групп равно трем. Будем предполагать, что уровень лимитирования не оказывает прямого влияния на первую возрастную группу и определяет динамику второй и третьей возрастных групп, т.е. $P(s) = \text{diag} \{1, p_2(s), p_3(s)\}$. Будем считать, что первая и третья возрастные группы непосредственно не влияют на уровень лимитирования, т.е. $G = (0, g, 0)$. Интенсивность эксплуатации будем считать постоянной $v = d$ и такой, что выполняется условие существования нетривиального стационарного решения Y_0 (см. теоремы 2.1.1 и 2.1.2). Таким образом, мы приходим к системе

$$\begin{aligned} \dot{y}_1 &= a_{11}y_1 + a_{12}y_2 + a_{13}y_3 - dy_1, \\ \dot{y}_2 &= (a_{21}y_1 + a_{22}y_2)p_2(s_p - gy_2) - dy_2, \\ \dot{y}_3 &= (a_{32}y_2 + a_{33}y_3)p_3(s_p - gy_2) - dy_3. \end{aligned} \quad (2.2.22)$$

Не умаляя общности, будем считать, что $p_2(s_0) = p_3(s_0) = 1$. Так же, как и в п. 2.1.5, будем предполагать, что

$$p'_i(s) \geq 0 \quad (i = 2, 3), \quad p'_2(s) + p'_3(s) > 0,$$

$$G^* P(0) = 0, \text{ т.е. } P_2(0) = 0.$$

Кроме того, будем пользоваться обозначением $y_i^0 = (Y_0)_i$.

К виду (2.2.22) приводится модель динамики численности непрерывно культивируемой популяции микроорганизмов, имеющих три фазы развития и обладающих следующими особенностями: первая и третья фазы не потребляют лимитирующий субстрат, и этот субстрат является индуктором биосинтетического процесса клеток третьей фазы. Действительно, в соответствии с п. 2.2.4 имеем

$$\begin{aligned} \dot{x}_1 &= a_{11}x_1 + a_{12}p_2(s)x_2 + a_{13}p_3(s)x_3 - dx_1, \\ \dot{x}_2 &= a_{21}x_1 + a_{22}p_2(s)x_2 - dx_2, \\ \dot{x}_3 &= a_{32}p_2(s)x_2 + a_{33}p_3(s)x_3 - dx_3, \\ \dot{s} &= -gp_2(s)x_2 + d(s_p - s). \end{aligned} \quad (2.2.23)$$

Делая замену переменных $X = AY$, где

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & 0 \\ 0 & a_{32} & a_{33} \end{bmatrix}, \quad (2.2.24)$$

и органичиваясь рассмотрением на асимптотически устойчивой инвариантной гиперплоскости $G^*A^{-1}X + s = s_p$, приводим систему (2.2.23) к виду (2.2.22). Отметим, что матрицы законов размножения (1.2.2) и (1.2.3) являются частным случаем матрицы (2.2.24).

Будем искать периодическое решение системы (2.2.22) при условии, что стационарная точка неустойчива. Имеет место

Предложение 2.2.8. Существуют такие числа $\epsilon > 0$ и $M > 0$, что стационарное решение неустойчиво, если

$$p'_2(s_0) \leq \epsilon, \quad (2.2.25)$$

$$p'_3(s_0) \geq M. \quad (2.2.26)$$

Доказательство. Предположим вначале, что $p'_2(s_0) = 0$, и найдем условие неустойчивости стационарного решения. В соответствии с формулой (2.2.7) матрица линейного приближения имеет вид

$$D = \begin{bmatrix} a_{11} - d & a_{12} & a_{13} \\ a_{21} & a_{22} - d & 0 \\ 0 & a_{32} - c & a_{33} - d \end{bmatrix},$$

где $c = gp'_3(s_0)x_3^0$. Выпишем характеристический полином, принимая во внимание, что $\lambda(A) = d$. Имеем

$$\begin{aligned} \lambda^3 + \lambda^2(3d - \sum_{i=1}^3 a_{ii}) + \lambda(\sum_{i \neq j} (d - a_{ii})(d - a_{jj}) - \\ - a_{12}a_{21}) + a_{13}a_{21}c = 0. \end{aligned}$$

Условие неустойчивости имеет вид

$$(3d - \sum_{i=1}^3 a_{ii})(\sum_{i \neq j} (d - a_{ii})(d - a_{jj}) - a_{12}a_{21}) - a_{13}a_{21} < 0$$

и выполняется при достаточно большом c . Малое возмущение матрицы D при $0 \neq p'_2(s_0) \leq \delta$ и достаточно малом ϵ не нарушает условие неустойчивости. Предложение доказано.

Содержательный смысл приведенных выше условий сводится к тому, что стационарный режим практически нереализуем, когда обеспеченность особей второй фазы развития лимитирующим субстратом близка к насыщению, в то время как скорость биосинтетических процессов особой третьей фазы сильно меняется при малых изменениях уровня лимитирования.

Существование инвариантного выпуклого компакта D из теоремы 2.2.2 при законе размножения в виде (1.2.2) известно (см. теорему 2.1.17 и формулу (2.1.24)). В случае, когда матрица (2.2.24) имеет вид (1.2.3), условие равномерной ограниченности решений также выполнено. Действительно, в соответствии с формулой (2.1.25) имеем $G^*A^{-1} = g(1, 0, 1)$. Отсюда заключаем, что $0 \leq x_1 + x_3 \leq 2s_p$ на инвариантном множестве $G^*A^{-1}X + s = s_p$. Отсюда и из второго уравнения системы (2.2.23) имеем положительную инвариантность множества $0 \leq x_2 \leq dg/2s_p$. Существование D установлено.

Множество Q будем задавать условиями

$$Q_1: y_1 \geq y_1^0, y_2 \leq y_2^0, y_3 \leq y_3^0$$

и

$$Q_2: y_1 \leq y_1^0, y_2 \geq y_2^0, y_3 \geq y_3^0.$$

Оно пересекается векторным полем изнутри наружу, если $Y \neq Y_0$ и

$$\begin{aligned} p_3(s) &> f(s) && \text{при } s > s_0, \\ p_3(s) &< f(s) && \text{при } s < s_0, \end{aligned} \quad (2.2.27)$$

где $f(s) = dy_3^0 / ((a_{32}/g)(s_p - s) + a_{33}y_3^0)$.

Доказательство. Условия отталкивания границ состоят в следующем:

$$\dot{y}_1|_{y_1^0} < 0, \dot{y}_2|_{y_2^0} > 0, \dot{y}_3|_{y_3^0} > 0 \quad \text{при } Y \in Q_1 \setminus Y_0,$$

$$\dot{y}_1|_{y_1^0} > 0, \dot{y}_2|_{y_2^0} < 0, \dot{y}_3|_{y_3^0} < 0 \quad \text{при } Y \in Q_2 \setminus Y_0.$$

Первые два неравенства в Q_1 и Q_2 выполняются, так как по теореме Котелянского [3] $a_{ii} - d \leq 0$ ($i = 1, 2, 3$). Третье условие выполняется, если

$$f_3(gy_3)(a_{32}y_2 + a_{33}y_3^0) - dy_3^0 > 0$$

при $y_2 < y_2^0$, $y_3 < y_3^0$ и противоположного знака при $y_2 > y_2^0$ и $y_3 > y_3^0$. Поскольку $s_p = s + gy_2$, это условие эквивалентно формуле (2.2.27).

Инвариантное многообразие W^3 принадлежит внутренности Q , если

$$p_2'(s_0) < \delta, \quad (2.2.28)$$

$$p_3'(s_0) \geq N, \quad (2.2.29)$$

где δ достаточно мало и N достаточно велико (рис. 2.7).

Доказательство. Пусть $p_2'(s_0) = 0$, выполнено условие неустойчивости (2.2.26) и μ_1, μ_2 и μ_3 – корни характеристического уравнения. Пусть μ_1 ($\mu_1 < 0$) – вещественный корень. Собственный вектор $Z = (z_1, z_2, z_3)$, отвечающий μ_1 , находится из системы

$$a_{21}z_1 + (a_{22} - d - \mu_1)z_2 = 0,$$

$$(a_{32} - c)z_2 + (a_{33} - d - \mu_1)z_3 = 0$$

и с точностью до постоянного множителя равен

$$Z = \left(1, \frac{a_{21}}{-a_{22} + d + \mu_1}, \frac{a_{21}(a_{32} - c)}{(-a_{22} + d + \mu_1)(-a_{33} + d + \mu_1)} \right).$$

Убедимся в том, что при достаточно большом c прямая $Y_0 + \sigma Z$, где $\sigma \in (-\infty, +\infty)$, лежит в множестве Q .

Действительно, свободный член характеристического полинома стремится к бесконечности при $c \rightarrow +\infty$, и поэтому произведение корней бесконечно возрастает. Сумма корней (коэффициент при λ^2) от c не зависит. Следовательно, $\mu_1 \rightarrow -\infty$, а $\mu_2 + \mu_3 \rightarrow +\infty$ при $c \rightarrow \infty$. Отсюда при достаточно большом c компоненты вектора Z имеют следующее распределение знаков: $Z = (+, -, -)$, и поэтому $Y_0 + \sigma Z \in Q_1 \cup Q_2$ при $\sigma \in (-\infty, +\infty)$. При достаточно малом $P_2'(s_0)$ это условие сохраняется. Формулы доказаны.

В итоге имеет место следующий результат.

Предложение 2.2.9. Существует положительно инвариантное множество системы (2.2.23), на котором функция $f(s(y_2)) = f(s_p - gy_2)$ из формулы (2.2.27) ограничена и принимает лишь положительные значения.

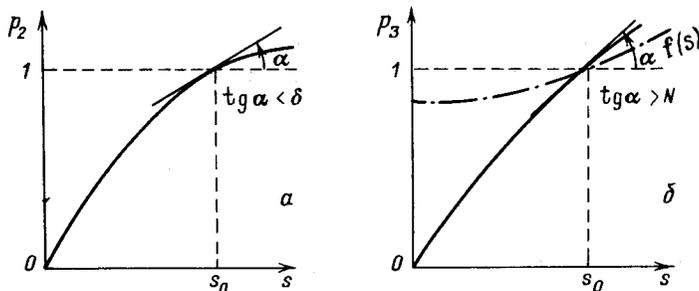


Рис. 2.7. Функции влияния концентрации лимитирующего субстрата на скорости прохождения фаз развития (a – соответствует формуле (2.2.28); b – формуле (2.2.29))

Если скорость изменения функции $p_3(s)$ на этом множестве настолько велика, что выполнены условия (2.2.26), (2.2.27) и (2.2.29), а скорость изменения функции $p_2(s)$ в окрестности стационарной точки настолько мала, что выполнены условия (2.2.25) и (2.2.28), то система (2.2.24) имеет периодическое решение ненулевого периода (см. рис. 2.8, a и b).

Доказательство. Ограничим область определения системы (2.2.22) инвариантным конусом $Y \geq 0$, $AY \geq 0$ и $G^*Y + s = s_p$, на котором системы (2.2.22) и

(2.2.23) эквивалентны. Повторяя рассуждения леммы из предложения 2.1.5 и пользуясь условием $G^*A^{-1}X + s = s_p$, оценим снизу величину $\|X(t)\|$. Пусть $L = (l_1, l_2, l_3)^*$ — собственный вектор матрицы A^* , отвечающий собственному значению $\lambda(A)$, выбран так, что $L \geq G^*A^{-1}$. Заменяв $G^*A^{-1}X$ на L^*X и умножив первые три уравнения системы (2.2.23) на L^* , перейдем к неравенству

$$L^*\dot{X} \geq L^*(\lambda(A)P(s_p - L^*X) - dI)X.$$

Так как $P(s) \geq I$ при $s \geq s_0$, то при $s_p - L^*X \geq s_0$ имеем $L^*X \geq 0$. Отсюда заключаем, что

$\sum_{i=1}^3 l_i x_i \geq s_p - s_0$. Воспользовавшись этим условием и неравенством

$s_0 \leq s \leq s_p$ перейдем от второго и третьего уравнений системы (2.2.23) к неравенствам

$$\dot{x}_2 \geq a_{21} \left[\frac{s_p - s_0}{l_1} - \frac{l_2}{l_1} x_2 - \frac{l_3}{l_1} x_3 \right] + (a_{22} p_2(s_p^*) - d)x_2,$$

$$\dot{x}_3 \geq a_{32} x_2 + [a_{33} p_3(s_p^*) - d]x_3,$$

где $s_p^* = s_p$ при $a_{ii} \leq 0$ ($i = 2, 3$) и $s_p^* = s_0$ при $a_{ii} > 0$ ($i = 2, 3$). Прибавим к первому неравенству второе, умноженное на константу c_1 , которая выбрана так, чтобы избавиться от x_2 . Имеем

$$\dot{x}_2 + c_1 \dot{x}_3 = -[\dots]x_3 + a_{21} \frac{s_p - s_0}{l_1},$$

и так как $x_2 \geq 0$, то

$$\dot{x}_2 + c_1 \dot{x}_3 \geq -[\dots] \frac{x_2 + c_1 x_3}{c_1} + a_{21} \frac{s_p - s_0}{l_1}.$$

Проинтегрировав неравенство, получаем $x_2 + c_1 x_3 \geq \beta > 0$. Подставив неравенство $x_2 \geq \beta - c_1 x_3$ во второе неравенство и проинтегрировав его, легко получаем, что $x_3 = a_{32} y_{22} + a_{33} y_2 > \alpha > 0$. Воспользовавшись тем, что $s_p = s + g y_2$, окончательно имеем ограниченность функции $f(s)$ из формулы (2.2.27). Доказательство предложения завершается использованием предложений 2.2.28, 2.2.9 и теоремы 2.2.2.

Таким образом, когда влияние лимитирующего фактора на разные возрастные группы популяции существенно неодинаково, в популяции могут возникнуть автоколебания большой амплитуды. На рис. 2.8 представлены

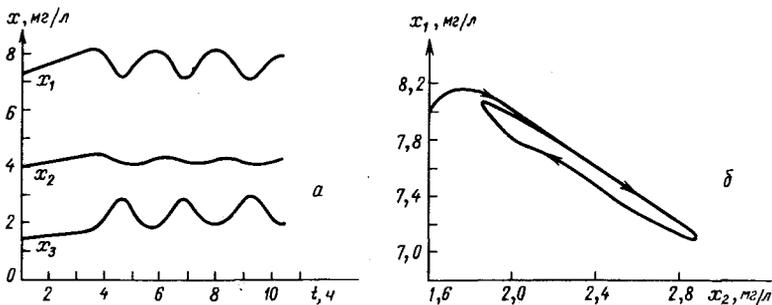


Рис. 2.8. Колебания численности непрерывно культивируемой популяции почкующихся клеток

результаты численных экспериментов с моделью (2.2.22), когда матрица (2.2.24) имеет вид (1.2.3) и задает закон размножения почкующихся клеток. При расчете было принято: скорость разбавления $d = 0,84 \text{ час}^{-1}$, концентрация подаваемого субстрата $s_p = 10 \text{ г/литр}$, коэффициенты потребления задаются вектором $G = (0, 3, 0) \text{ г/(литр} \times \text{час} \times \text{мг)}$, матрица продолжительностей фаз развития имеет вид

$$p(s) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \frac{1,9s}{s + 0,03} \end{bmatrix}.$$

§ 2.3. Хаос

В первом и втором параграфах рассматривались простейшие типы динамики численности экологических систем — равновесие и периодические колебания. Их устойчивость означала регулярное поведение этих систем. В то же время хорошо известны примеры экологических процессов с нерегулярным поведением. Существуют виды насекомых, у которых в течение длительного срока могут наблюдаться слабые изменения численности, сменяющиеся внезапными подъемами. При этом максимальная численность превышает минимальную в десятки тысяч раз. Вспышки размножения происходят с неравными интервалами времени. Такое поведение называется *хаотическим* и характеризуется крайней степенью неустойчивости движения. Состояние системы с хаотическим поведением не поддается прогнозу. Как бы ни задавались начальные данные, судьба биологической системы через конечный промежуток времени практически непредсказуема.

Хаотическая динамика особенно характерна для некоторых моделей экологических систем с дискретным временем. Эти модели не являются тривиальным разностным аналогом соответствующих дифференциальных уравнений, а возникают при изучении заметно синхронизированных процессов. Классическим примером такого процесса является воспроизводство популяции с неперекрывающимися поколениями. В естественных условиях четкая смена фаз развития особенно свойственна насекомым, в искусственных условиях известным примером процесса с дискретным временем является периодическое культивирование.

В моделях динамики численности популяций с неперекрывающимися поколениями изначально не предполагается вероятностный характер механизмов взаимодействия. И все же, несмотря на сугубо детерминистический подход, во многих случаях решения демонстрируют слабую зависимость от начальных состояний и отсутствие локальной устойчивости траекторий. Визуально они напоминают реализацию случайного процесса [5, 12, 38].

Основные свойства хаотических структур обычно сводятся к следующим признакам [44, 46, 47, 51]:

1. Существование бесконечного числа периодических решений.
2. Существование бесконечного несчетного множества решений, которые демонстрируют случайное поведение.

3. Чрезвычайная чувствительность к начальным данным.

К настоящему времени развита специальная математическая теория, позволяющая установить наличие этих признаков. Она называется символической динамикой [1]. Суть этой теории сводится к следующему. Строится отображение конечной последовательности символов в фазовое пространство исходной системы. В настоящем параграфе этими символами являются нуль и единица. Сдвигом по времени нуль переводится либо в нуль, либо в единицу. Таким образом, формируется траектория динамической системы, фазовое пространство которой состоит из конечного числа точек. Отображение символов в фазовое пространство исходной системы сопоставляет "символической" траектории траекторию исходной системы. Другими словами, динамика детерминированной модели моделируется последовательными испытаниями Бернулли. Основная трудность в данном случае состоит в том, чтобы построить описанное выше отображение.

Материал этого параграфа организован следующим образом. Вначале рассмотрим класс моделей динамики численности популяций с непрерывными поколениями, который охватывает случаи конечного, счетного и континуального распределения особей по фазам развития или возрастным группам. После этого сформулировано математическое определение структуры, которое является формализацией интуитивного представления о "хаотическом" поведении и в то же время достаточно широко для того, чтобы быть применимым к рассматриваемым моделям. Затем изучены свойства этой структуры и дано конструктивное условие рождения такой структуры в моделях динамики численности с непрерывными поколениями.

2.3.1. Модели динамики численности популяций с непрерывными поколениями. Рассмотрим модель динамики численности популяции с непрерывным возрастным распределением [31]:

$$\frac{\partial x(a, t)}{\partial t} + \frac{\partial x(a, t)}{\partial a} = -d(a, N)x(a, t),$$

$$x(0, t) = \int_0^{\infty} b(a, s)x(a, t)da,$$

где $x(a, t)\Delta a + o(\Delta a)$ — число особей, возраст которых лежит в интервале $[a, a + \Delta a]$, $s = \int_0^t x(a, t)da$ — общая численность популяции, $d(a, s) -$

коэффициент смертности, $b(a, s)\Delta a + o(\Delta a)$ — число потомков одной особи, возраст которой лежит в интервале $[a, a + \Delta a]$. Будем считать, что способность производить потомство падает до нуля, когда численность популяции приближается к критическому уровню $s = s_{\infty}$.

При численном изучении данной модели на непрерывный возраст накладывается дискретная сетка, интегралы заменяются суммами, а значения функций на промежутке между двумя узлами сетки считаются постоянными и равными значению в некоторой средней точке. Таким образом, в популяции вводят разбиение на конечное число возрастных групп, численность каждой из которых обозначается через x_i ($i = 1, \dots, m$). Пусть

матрица правой части является разложимой. Тем не менее она имеет единственный собственный вектор с положительными компонентами $(1, b_1/[a(1-s)^{q(s)}], \dots, b_1 \dots b_m/[a^m(1-s)^{q(s)m}])$, которому отвечает положительное собственное значение $a(1-s)^{q(s)}$. При $s = s_\infty$ матрица правой части нильпотентна.

Подытожим основные свойства рассмотренных выше моделей. Фазовым пространством обеих моделей является положительный конус евклидова пространства. Оператор сдвига вдоль траекторий задан матрицей с неотрицательными компонентами. В тех случаях, когда лимитирующий фактор не достигает критического уровня, эта матрица обладает единственным собственным вектором, который принадлежит положительному конусу. Этому собственному вектору отвечает положительное собственное значение. При значениях лимитирующего фактора, превосходящих критический уровень, оператор сдвига вдоль траекторий нильпотентен.

Перейдем к построению общей модели динамики численности экосистем с непрерывными поколениями. Большой практический интерес представляет вопрос рождения хаотических движений в моделях экосистем со счетным или континуальным распределением по возрасту или какому-нибудь другому признаку. В связи с этим предположим, что численный состав экологической системы в каждый момент времени задается вектором банахова пространства E . В соответствии с биологическим смыслом траектории модели будут принадлежать положительному конусу K этого пространства. Конус K является порождающим, выпуклым и выступающим (собственным) [16, 17, 36]. Это означает, что $E = K - K$, $K = aK + bK$ ($a, b \in R_+^1$) и $K \cap \{-K\} = 0$. Он задает на E отношение порядка [16, 17, 36]. Это означает, что неравенство $X \geq Y$ имеет место тогда и только тогда, когда $X - Y \in K$. Отношение \geq обладает обычными свойствами знака неравенства [15, 16]. Будем предполагать, что конус K является вполне минимизированным [16, 17]. Это означает, что отношение порядка обладает свойством решетки (структуры) и условной полноты, т.е. для любой пары элементов $X, Y \in E$ определена точная верхняя граница $\text{Sup}(X, Y)$ и каждое (порядково) ограниченное сверху множество также имеет точную верхнюю границу. Таким образом, E является упорядоченным K -пространством (пространством Л.В. Канторовича) [16, 17]. Поскольку E наделено свойствами банахова пространства, конус K является замкнутым множеством. К числу банаховых K -пространств относятся пространства интегрируемых в p -й степени функций, пространства существенно ограниченных функций и т.д. Конус K в этом случае образован неотрицательными функциями. Читатель, который интересуется лишь конечномерным случаем, может считать в дальнейшем, что $E = R^m$ и $K = R_+^m$.

Рассмотрим систему вида [9, 10]

$$X_{t+1} = F(s_t, X_t), \quad s_t = \langle C, X_t \rangle, \quad (2.3.3)$$

где $X \in E$, C — неотрицательный функционал [36], а $F(s, X)$ — нелинейный непрерывный оператор скалярного аргумента s и вектора X . Предположим, что при всех s из некоторого конечного интервала $[0, s_\infty)$

оператор $F(s, \cdot)$ как функция второго аргумента положителен, монотонен в K и положительно однороден [16, 22, 24], т.е.

$$\begin{aligned} F(s, \cdot): K &\rightarrow K, \\ F(s, X) &\geq F(s, Y) \quad \text{при } X \geq Y, \\ F(s, \alpha X) &= \alpha F(s, X) \quad \text{при } \alpha \in R_+^1. \end{aligned} \quad (2.3.4)$$

Иногда будет использоваться условие сублинейности [17]

$$F(s, X + Y) \leq F(s, X) + F(s, Y). \quad (2.3.5)$$

Кроме того, будем считать, что при каждом фиксированном $s \in [0, s_\infty]$ оператор $F(s, \cdot)$ имеет единственный положительный собственный вектор H_s , принадлежащий внутренности K , которому отвечает собственное значение $\lambda(s)$, и для любого вектора из внутренности конуса $X \in \overset{\circ}{K}$ имеет место [24, 36]

$$F^n(s, X) / \lambda^n(s) \rightarrow a H_s \quad (2.3.6)$$

в слабой топологии при $n \rightarrow \infty$, где $a > 0$.

Рассматривая $F(s, X)$ как функцию первого аргумента, будем считать его монотонно убывающим по s на интервале $[0, s_\infty)$, т.е.

$$F(s_1, X) \leq F(s, X) \quad \text{при } s \leq s_1. \quad (2.3.7)$$

Суперпозицию операторов $F(\langle C, X \rangle, X)$ будем обозначать через $F(X)$. Предположим, что гиперплоскость $\langle C, X \rangle = s_\infty$ переводится n -й итерацией оператора $F(X)$ в нуль, точнее,

$$F^i(F(s_\infty, X)) \in K \setminus 0 \quad \text{при } i \leq n - 1$$

(2.3.8)

и

$$F^{n-1}(F(s_\infty, X)) = 0.$$

В случае, когда E — конечномерное пространство и $K = R_+^m$, свойства правой части уравнения (2.3.3) имеют вид

$$\begin{aligned} F(s, \cdot): R_+^m &\rightarrow R_+^m, \\ F(s, X) &\geq F(s, Y) \quad \text{при } X \geq Y, \\ F(s, \alpha Y) &= \alpha F(s, Y), \end{aligned}$$

где $X, Y \in R_+^m$ и $\alpha \in R_+^1$. Формулировка условия сублинейности сохраняется, а условие (2.3.7) принимает вид

$$\|F^n(s, X) / \lambda^n(s) - a H_s\| \rightarrow 0 \quad \text{при } n \rightarrow \infty.$$

2.3.2. Общие свойства хаотических режимов. В работе [46] численно исследовалась двумерная модель вида (2.3.1). Предполагалось, что $b_0(s) = \beta_1(1-s)$, $b_1(s) = \beta_2(1-s)$, $c_1 = 1$ и $s = 1 - \beta_1 x_1 - \beta_2 x_2$. Параметры

системы принимали значения из области $D = \{ \beta_1, \beta_2 \geq 0 \mid \beta_1 + \beta_2 \leq 4 \}$, и поэтому ее решения принадлежали квадрату $\left\{ (x_1, x_2)^* \in R_+^2 \mid x_1 \leq \frac{1}{4}, x_2 \leq \frac{1}{4} \right\}$.

Было замечено, что при β_1 , близком к четырем, и β_2 , достаточно малом, движения системы "случайным образом" ложились на некоторую гладкую кривую, принадлежащую положительному ортанту R_+^2 и проходящую через начало координат. Этот феномен визуально воспринимался как "одномерный хаос" [46]. В тех случаях, когда β_2 было близко к четырем и β_1 достаточно мало, траектории движений системы заполняли некоторую область на плоскости и "хаос" визуально воспринимался как двумерный [46]. Легко видеть, что в двумерной модели вида (2.3.1) при больших β_1 и малых β_2 одно собственное значение матрицы $A(0)$ лежит вне единичного круга, а другое — внутри него. При больших значениях β_2 оба собственных числа матрицы $A(0)$ не принадлежат единичному кругу. Поэтому в первом случае неустойчивое подпространство образовано одним собственным вектором и, следовательно, одномерно, а во втором случае оно образовано двумя собственными векторами и, следовательно, совпадает со всей плоскостью. В данном пункте будут изучены движения, начинающиеся в малой окрестности неустойчивой стационарной точки и визуально воспринимаемые как хаотические. Будет показано, что "квазислучайные" или "хаотические" движения группируются в окрестности неустойчивого инвариантного многообразия исходной системы. Такое многообразие является аналогом неустойчивого подпространства линеаризованной системы и касается этого подпространства в стационарной точке.

Введем вспомогательные обозначения. Пусть E — банахово пространство. Определим расстояние между множествами этого пространства правилом $d(A, B) = \inf \{ \|X - Y\| \mid X \in A, Y \in B \}$. Диаметр произвольного множества пространства E определим правилом $\text{diam } A = \sup \{ \|X - Y\| \mid X, Y \in A \}$. Любое непрерывное отображение $G: E \rightarrow E$ задает динамическую систему с дискретным временем $X_{t+1} = G(X_t)$. Ниже пойдет речь о том, как моделировать движения этой динамической системы движениями более простой системы.

Представим себе, что состояния исходной системы измеряются грубым прибором. Показания этого прибора могут принимать лишь два значения — нуль и единица. С приращением времени на один шаг состояние исходной системы меняется и, следовательно, показание прибора либо сохраняет исходное значение, либо принимает другое. Запись показаний прибора представляет собой бесконечную последовательность нулей и единиц $(\xi_0, \xi_1, \xi_2, \dots)$, где $\xi_i \in \{0, 1\}$ ($i \geq 0$), которую мы будем называть символической последовательностью. Множество таких последовательностей образует счетное декартово произведение пар $\{0, 1\}$, которое мы будем обозначать через $\Sigma = \prod_{i \geq 0} \{0, 1\}$. Рассмотрим динамику показаний

прибора. Если измерения начаты в момент времени $t = 0$, то показания прибора представлены символической последовательностью $(\xi_0, \xi_1, \xi_2, \dots)$. Допустим теперь, что измерения начаты в момент времени $t = 1$. Тогда показания прибора представлены последовательностью (ξ_1, ξ_2, \dots) , которая

получена из предыдущей отбрасыванием первого члена и сдвигом на один индекс влево. Таким образом, сдвигу по времени на один шаг соответствует сдвиг символов на один индекс влево: $\tau: (\xi_0, \xi_1, \xi_2, \dots) \rightarrow (\xi_1, \xi_2, \dots)$. Динамическая система, которая порождается отображением $\tau: \Sigma \rightarrow \Sigma$, является сравнительно простым объектом исследования. Она имеет бесконечное число периодических движений. Например, элемент $(0, 1, 0, 1, 0, 1, \dots)$ переходит в себя за два временных шага. Система имеет бесконечное число движений, которые могут совпадать в некоторый момент времени, а затем расходиться. Например, движения с начальными условиями $(0, 1, 0, 1, 0, 1, \dots)$ и $(0, 0, 1, 0, 1, 0, 1, \dots)$ совпадают лишь при $t = 0$.

Легко видеть, что динамическая система $\eta(t+1) = \tau(\eta(t))$, где $\eta(t) \in \Sigma$, демонстрирует все признаки хаотической динамики, которые упомянуты во введении к данному параграфу.

Такая система моделирует динамику исходного уравнения $X_{t+1} = G(X_t)$ в том случае, когда каждой символической траектории можно сопоставить только одну траекторию исходной системы. Это условие легко гарантировать, если показать, что существует взаимно однозначное отображение (инъекция) h пространства Σ в пространство E , которое вместе с оператором сдвига на один индекс влево τ образует коммутативную диаграмму по отношению к отображению G , т.е. $h\tau = Gh$ [1]. Другими словами, вместо того чтобы вычислять значение функции $G(X)$, где $X = h(\eta)$, достаточно вычислить значение функции $\tau(\eta)$ и потом найти образ этого значения в пространстве E .

Мы рассмотрим более слабое свойство ϵ -коммутативности [10]. Оно состоит в следующем. Существует бесконечная последовательность инъекций $h_k: \Sigma \rightarrow E$ ($k \geq N$) таких, что

$$\|h_k \tau^p(\xi) - G^p h_k(\xi)\| < \epsilon_k, \quad (2.3.9)$$

где $\epsilon_k \rightarrow 0$ при $k \rightarrow \infty$ и $p \geq 1$. Тогда аналогом периодических точек будут инвариантные множества сколь угодно малого диаметра.

1. Существует бесконечная последовательность множеств $\{B_i\}_{i \geq N}$ из E такая, что для любого $k \geq N$ имеет место включение $G^k(B_k) \subset B_k$ и $\text{diam } B_k \leq \epsilon_k$, где ϵ_k определены в формуле (2.3.9).

Условие непредсказуемости поведения несчетного множества траекторий имеет вид:

2. Существует число d_0 , не зависящее от k и такое, что при всех $k \geq N$ имеет место

$$\lim_{n \rightarrow \infty} \text{Sup } d(G^n(X), G^n(Y)) > d_0,$$

где $X, Y \in h_k(\Sigma)$, $X \neq Y$.

3. Для любого $Z \in h_k(\Sigma)$ такого, что $h_k^{-1}(Z)$ не является периодической последовательностью, и любого множества B_k имеет место неравенство

$$\lim_{n \rightarrow \infty} \text{Sup } d(G^n(Z), G^n(B_k)) > d_0,$$

где d_0 определено в п. 2.

Условие того, что разбегающиеся траектории могут сходиться в некоторые моменты времени сколь угодно близко, принимает вид:

$$4. \lim_{n \rightarrow \infty} \text{Inf} d(G^n(X), G^n(Y)) \leq \epsilon_k,$$

если $h_k^{-1}(X)$ и $h_k^{-1}(Y)$ совпадают хотя бы в одной точке.

Определение. Если траектории динамической системы, задаваемой оператором G , удовлетворяют условию (2.3.9) и условиям 1–4, будем говорить, что данная система обладает свойством Ch.

Прежде чем перейти к рассмотрению условий, при которых динамическая система $X_{t+1} = G(X_t)$ обладает свойствами Ch, введем некоторые дополнительные обозначения и сформулируем один вспомогательный результат.

В дальнейшем через $C^{1,1}(A, E)$ будем обозначать класс функций $F: A \rightarrow E$, отображающих подмножество A банахова пространства E в пространство E , непрерывно дифференцируемых по Фреше и таких, что их дифференциал DF удовлетворяет условию Липшица, т.е. $F(X+H) = DF(X)H + o(\|H\|)$, где $H \in E$, $DF(X): E \rightarrow E$ — линейный непрерывный оператор и $\|DF(X) - DF(Y)\| \leq l\|X - Y\|$ при любых $X, Y \in E$. Очевидно, что дважды непрерывно дифференцируемые функции принадлежат классу $C^{1,1}(A, E)$. На протяжении всего пункта будем предполагать, что $G \in C^{1,1}(E, E)$ имеет неподвижную точку 0 и спектр дифференциала $DG(0)$ в неподвижной точке распадается на непересекающиеся множества S_{p_1} и S_{p_2} такие, что

$$\sup_{\rho \in S_{p_1}} |\rho| < \inf_{\rho \in S_{p_2}} |\rho| = \lambda \text{ и } \lambda > 1.$$

Пусть E_i — инвариантное подпространство оператора $DG(0)$, соответствующее S_{p_i} ($i = 1, 2$). Через P_{r_i} ($i = 1, 2$) будем обозначать проекции E на подпространства E_i ($i = 1, 2$). Предположим, что E_2 имеет конечную размерность. Напомним теорему об инвариантном многообразии, являющуюся перенесением хорошо известных результатов [43, 45] на бесконечномерный случай.

Теорема 2.3.1. При сделанных выше предположениях существует локальное многообразие W^U , касающееся E_2 в точке 0, которое является инвариантным относительно отображения G . Это многообразие представимо в виде графика функции $\theta \in C^{1,1}(U^2, E_1)$, где $U^2 \subset E_2$ — некоторая окрестность нуля.

Доказательство. Пусть $P_{r_i}G(X) = L_i X_i + R_i(X)$, где $X_i \in E_i$, $L_i = (DG(0)|_{E_i})$ и $R_i = O(\|X\|^2)$ ($i = 1, 2$). Заменяем оператор G новым оператором $G^0 \in C^{1,1}(E, E)$, который определяется соотношением $P_{r_i}G^0(X) = L_i X_i + R_i^0(X)$, где $R_i(X) = R_i^0(X)$ при $P_{r_2}X \in U^2$, $R_i^0 = 0$ при $P_{r_2}X \notin U^2$ и $U^2 \subset E_2$ — достаточно малая окрестность. Инвариантное многообразие, как это обычно делается, строится для оператора G^0 . Определим отображение графиков функций $f: A \subset E_2 \rightarrow E_1$ правилом

$$\Gamma: (Y, f(Y)) \mapsto (P_{r_2}G^0(Y, f(Y)), P_{r_1}G^0(Y, f(Y))).$$

Следуя [43], введем полное пространство функций

$$B = \{f \in C(E_2, E_1) \mid \sup \|f(Y) - f(Z)\| / \|Y - Z\| \leq \beta \text{ и } f(0) = 0\},$$

наделенное метрикой $d(f, g) = \sup \|f(Y) - g(Y)\| / \|Y\|$. Повторяя рассуждения лемм 10 и 11 работы [43], которые нигде не используют конечномерность простран-

ства E_1 , можно показать, что $\Gamma B \subset B$ при достаточно малом β и $d(\Gamma f, \Gamma g) \leq \gamma d(f, g)$, где $\gamma < 1$. Неподвижная точка θ^0 этого сжатия задает искомое инвариантное многообразие.

Принадлежность функции θ^0 классу $C^{1,1}$ доказывается следующими стандартными рассуждениями. Пусть $f \in B$ и $\|Df(Y) - Df(Z)\| \leq \|Y - Z\|$. Рассуждая так же, как и в теореме о центральной многообразии при оценке константы Липшица у производной, легко показать, что константа Липшица функции $D(\Gamma f)$ не превосходит единицы.

Введем в рассмотрение пространство функций

$$B_1 = \{ (f|U^2) \mid f \in B \text{ и } \|Df(Y) - Df(Z)\| \leq \|Y - Z\| \},$$

где окрестность нуля $U^2 \subset E_2$ достаточно мала. Определим отображение графиков функций Γ_0 правилом $\Gamma_0(f) = (\Gamma(f)|U^2)$.

Так как Sp_2 оператора $(DG(0)|E_2)$ лежит вне единичного круга, можно считать, что область определения $\Gamma(f)$ для $f \in B_1$ содержит окрестность U^2 . Поэтому $\Gamma B_1 \subset B_1$. По теореме Асколи пространство B_1 компактно в C^1 -топологии. Следовательно, всякая сходящаяся в C^0 -топологии последовательность сходится в C^1 -топологии. Отсюда следует, что $\theta = (\theta^0|U^2) \in C^{1,1}(U^2, E_1)$. Теорема доказана.

Вернемся к исследованию динамических свойств траекторий, порождаемых суперпозицией отображения G . Критерием рождения Ш-структуры будет выполнение следующего свойства.

Определение. Пусть 0 — неподвижная точка оператора $G: E \rightarrow E$ и G^n — суперпозиция отображений G , взятых n раз. Будем говорить, что динамическая система, порождаемая отображением G , обладает сильно возвратной структурой в точке 0 , если для любой окрестности U этой точки найдется пара (X, n) , где $X \in U$ и $X \neq 0$, а n — целое положительное число, такая, что $G^n(X) = 0$. Множество всех точек, попадающих за конечное время в 0 , будем обозначать через SR_0 .

Общий результат содержит

Теорема 2.3.2. Предположим, что отображение G таково, что

1. Сильно возвратная структура в точке 0 существует и имеет непустое пересечение с W^U , т.е. $SR_0 \cap W^U \neq \emptyset$.

Для некоторой точки $X \in W^U \cap SR_0$ существует окрестность U_X такая, что $G^n(U_X) \cap E_1 = \{0\}$, где n — номер первого попадания X в 0 .

3. Сужение отображения G^n на W^U трансверсально E_1 в точке X .

Тогда динамическая система $X_{t+1} = G(X_t)$ обладает свойством Ш.

Доказательство. Через $\|\cdot\|_1$ будем обозначать норму, определяющую C^1 -топологию в пространстве функций, отображающих некоторую окрестность нуля из E_2 в E_1 . Кроме того, будем использовать знак трансверсального пересечения: \pitchfork . Если область задания функции f , отображающей некоторую окрестность нуля из E_2 в E_1 , не оговорена специально, будем писать $f: E_2 \rightarrow E_1$, понимая эту запись локально.

Окрестность в пространстве функций $f: (E_2, 0) \rightarrow (E_1, 0)$ определим условием

$$S_\epsilon^\alpha(f) = \{ \psi \in C^{1,1}(E_2, E_1) \mid \|\psi - f\|_1 \leq \epsilon, \psi(0) = 0 \text{ и } \|D\psi(Y) - D\psi(Z)\| \leq \alpha \|Y - Z\| \}.$$

Воспользовавшись теоремой 2.3.1, зададим многообразие W^U графиком отображения $\theta \in C^{1,1}(E_2, E_1)$. Пусть $X \in SR_0 \cap W^U$. Поскольку $W^U \cap G^{-k}(X) \rightarrow 0$ при $k \rightarrow \infty$, можно считать, что X принадлежит сколь угодно малой окрестности нуля. Фиксируем точку X и число n такие, что $G^n(X) = 0$. Определим отображение графиков функций, заданных в некоторой окрестности точки $\text{Pr}_2 X$, правилом

$$\Gamma_X^n: (Y, f(Y)) \mapsto (\text{Pr}_2 G^n(Y, f(Y)), \text{Pr}_1 G^n(Y, f(Y))).$$

Так как по условию теоремы $(G^n | W^U) \uparrow_X E_1 = 0$, то существует такая окрестность $U_X^2 \subset E_2$ точки X , что $\Gamma_X^n(\theta | U_X^2)$ представимо в виде функции $\varphi: E_2 \rightarrow E_1$, причем $\varphi(0) = 0$.

Лемма 2.3.2.1. По любому $\epsilon > 0$ найдутся такие $\delta > 0$ и $\alpha > 0$, что $\Gamma_X^n((f | U_X^2)) \in S_\epsilon^\alpha(\varphi)$ при всех $f \in S_\delta^1(\theta)$.

Доказательство леммы. Пусть $f \in S_\delta^1(\theta)$. Если δ достаточно мало, то $\Gamma_X^n(f)$, близко к φ в C^1 -топологии и $\Gamma_X^n((f | U_X^2))$ пересекает E_1 трансверсально. По условию теоремы $\Gamma_X^n(f)(U_X^2) \cap E_1 = \{0\}$ при достаточно малом δ . Следовательно, $\Gamma_X^n(f | U_X^2)$, так же как и $\Gamma_X^n(\theta | U_X^2)$, представимо в виде $\psi: E_2 \rightarrow E_1$.

Существование константы Липшица α , одной для всех $\Gamma_X^n((f | U_X^2))$ при $f \in S_\delta^1(\theta)$, вытекает из следующих двух факторов: DG^n имеет ограниченную константу Липшица в точке X . Семейство отображений $\text{Pr}_2 G^n f: U_X^2 \rightarrow E_2$ в силу трансверсальности образовано диффеоморфизмами, близкими в C^1 -топологии. Лемма доказана.

Определим отображение Γ_0 графиков функций, заданных в окрестности точки 0 правилом

$$\Gamma_0: (Y, f(Y)) \mapsto (\text{Pr}_2 G(Y, f(Y)), \text{Pr}_1 G(Y, f(Y))).$$

Лемма 2.3.2.2. Если область определения функций из $S_\epsilon^\alpha(\varphi)$ ограничить на достаточно малую окрестность нуля, то по любому $\delta > 0$ найдется номер k такой, что

$$\Gamma_0^k S_\epsilon^\alpha(\varphi) \subset S_\delta^1(\theta).$$

Доказательство леммы. Следуя [43], введем в рассмотрение пространство B функций $f: E_2 \rightarrow E_1$, удовлетворяющих условию $\text{Sup} \|f(Y) - f(Z)\| / \|Y - Z\| \leq \beta$ и наделенных метрикой

$$d(f, g) = \text{Sup} \|f(Y) - g(Y)\| / \|Y\| \text{ при } f, g \in B.$$

При доказательстве теоремы 2.3.1 показано, что $\Gamma_0 B \subset B$, $\Gamma_0 \theta = \theta \in B$ и $d(\Gamma_0 f, \theta) \leq \gamma d(f, \theta)$, где $\gamma < 1$ при достаточно малом β .

Докажем, что существуют такое целое число p и такая малая область задания функций из $S_\epsilon^\alpha(\varphi)$, что выполняется включение $\Gamma_0^p S_\epsilon^\alpha(\varphi) \subset B$. Действительно, для любого $p > 0$ найдется такая окрестность нуля U_p , что имеет место представление $\text{Pr}_1 G^p(X) = L_i^p X_i + R_i(X)$, где $L_i = DG(0) | E_i$ и $R_i(X)$ сколь угодно малы по сравнению с линейной частью при $X \in U_p$. Поскольку $\varphi \in C^{1,1}$ и $\varphi(0) = 0$ для $\varphi \in S_\epsilon^\alpha(\varphi)$, при достаточно большом p функтор Γ_0^p переводит график во внутренность пространства B . Техника оценок та же, что и в теореме об инвариантном многообразии [21, 43].

Следовательно, $S_\epsilon^\alpha(\varphi)$ при достаточно большом p отображается функтором Γ_0^p в B . Кроме того, можно считать, что

$$\|D(\Gamma_0^k \varphi)(Y) - D(\Gamma_0^k \psi)(Z)\| \leq \|Y - Z\| \text{ при } k \geq p$$

и

$$\psi \in S_\epsilon^\alpha(\varphi).$$

Подмножество пространства B , удовлетворяющее предыдущему условию, будем обозначать через B_1 .

Итак, показано, что $\Gamma_0^k S_\epsilon^\alpha(\varphi) \subset B_1$ и $d(\Gamma^i B_1, W^U) \rightarrow 0$ при $i \rightarrow \infty$. Тот факт, что $\Gamma_0^i S_\epsilon^\alpha(\varphi) \rightarrow W^U$ в C^1 -топологии, следует из компактности множества B_1 в этой топологии и доказывается так же, как и в теореме 2.3.1. Лемма доказана.

Завершим доказательство теоремы. Из леммы 2.3.2.1 и 2.3.2.2 следует, что $S_\delta^1(\theta) \xrightarrow{\Gamma_X^n} S_\epsilon^\alpha(\varphi)$ и $S_\epsilon^\alpha(\varphi) \xrightarrow{\Gamma_0^k} S_\delta^1(\theta)$ при всех достаточно больших k . Для краткости суперпозицию отображений

$$f \in S_\delta^1(\theta) \xrightarrow{\Gamma_X^n} \psi \in S_\epsilon^\alpha(\varphi) \xrightarrow{\Gamma_0^k} \psi \in S_\delta^1(\theta)$$

будем обозначать через Γ^{n+k} . Построим последовательность множества $Q_0 =$

$= W^U \cap U_X^2 \times E_1$ и $Q_p = \text{gr}(\Gamma^{(n+k)p}\theta | U_X^2)$, где gr обозначает график функции. По построению

$$Q_0 = W^U \cap U_X^2 \times E_1 \text{ и } Q_{p+1} = G^{n+k}Q_p \cap U_X^2 \times E_1.$$

Пусть $F = G^{n+k}$. Сопоставим последовательности $\{Q_p\}_{p \geq 0}$ последовательность множеств $\{A_i^p\}_{p \geq 0}^{i \leq p}$ по правилу

$$A_0^p = Q_p, A_1^p = F^{-1}A_0^p \cap Q_{p-1}, \dots, A_p^p = F^{-1}A_{p-1}^p \cap Q_0.$$

Множество $A_1^p \neq \emptyset$ при любом p . Если U_X^2 – замкнутое множество, то Q_p – компакт при любом p , так как W^U – конечномерное многообразие. Отсюда следует, что A_p^p – компакт. Пусть in_p – вложение $A_{p+1}^p \rightarrow A_p^p$. Тогда проективный предел $S = \varprojlim (A_p^p, \text{in}_p) \neq \emptyset$. Множество $\bigcup_{i \geq 0} F^i S = B_{n+k}$ инвариантно относительно G^{n+k}

и содержится в некоторой окрестности U_X точки X . Следовательно, $\text{diam } B_{n+k} \leq \text{diam } U_X$.

Существование бесконечной последовательности инвариантных множеств мы доказали. Покажем теперь, что $B_{n+k} \rightarrow 0$ при $k \rightarrow \infty$. Будем считать, что область определения функций из $S_\delta^1(\theta)$ и $S_\epsilon^\alpha(\varphi)$ лежит в шаре радиуса $2\|X\|$ из E_2 , где $X \in SR_0$. Число $\|X\|$ считаем настолько малым, что $\|\text{Pr}_2 G(Y)\| \geq \mu \|\text{Pr}_2 Y\|$ при $\|Y\| \leq \|X\|$ и $1 < \mu$ достаточно близко к $\min_{\rho \in \text{Sp}_2} |\rho|$. Если область определения функций из $S_\epsilon^\alpha(\varphi)$

лежит в шаре D_k радиуса $2\|X\|/\mu^k$, то окрестность $X \in U_X^2$ принадлежит области определения функций $\Gamma_0^k \psi$ для всех $\psi \in S_\epsilon^\alpha(\varphi)$. Поэтому при построении $n+k$ инвариантных множеств B_{n+k} можно считать, что функции из $S_\epsilon^\alpha(\varphi)$ определены на шаре D_k . Для любой функции $f \in S_\delta^1(\theta)$ отображение $\text{Pr}_2 G^n f: U_X^2 \rightarrow E_2$ является локальным диффеоморфизмом и множество таких отображений равномерно непрерывно. Следовательно, обозначив $U_X^2(k) = \bigcap_{f \in S_\delta^1(\theta)} (\text{Pr}_2 G^n f)^{-1} D_k$, имеем, что $U_X^2(k)$ – окрестность точки X и $\text{diam } U_X^2(k) \rightarrow \infty$ при $k \rightarrow \infty$.

Если в лемме 2.3.2.1 отображение Γ_X^n определить по правилу

$$f \mapsto (f | U_X^2(k)) \mapsto \Gamma_X^n(f | U_X^2(k)),$$

то лемма 2.3.2.1 и все последующие рассуждения сохраняют силу. Для полученных таким образом множеств имеет место включение $\text{Pr}_2 B_{n+k} \subset U_X^2(k)$. Следовательно, $\text{diam } \text{Pr}_1 B_{n+k} \leq \text{diam } \theta(U_X^2(k)) + \delta$. В лемме 2.3.2.2 мы показали, что δ можно сделать сколь угодно малым, выбирая достаточно большие значения k . Поэтому, начиная с некоторого N , последовательность множеств $\{B_{n+k}\}_{k \geq N}$ существует и $\text{diam } B_{n+k} \rightarrow 0$ при $k \rightarrow \infty$.

Убедимся в существовании последовательности искомого инъекций $h_k: \Sigma \rightarrow E$. Пусть $k \geq N \geq n$, где n взято из условия 2 теоремы 2.3.2, и $V_0^2 \subset E_2$ – такая окрестность нуля, что $U_X^2 \cap V_0^2 \neq \emptyset$. Инъекции h_k определены следующим образом. Пусть Ξ – отображение, сопоставляющее точкам $U_X^2(k)$ число 0, а точкам из V_0^2 – число 1. Пусть $\xi = (\xi^1, \xi^2, \dots) \in \Sigma$. Проводя те же рассуждения, что и при построении множеств B_{n+k} , найдем непустые множества R_k такие, что $\Xi(G^{ik}(R_k)) = \xi^i$. Тогда $h_k: \xi \rightarrow Y$, где Y – произвольная точка из R_k , является искомым отображением. Теорема доказана.

2.3.3. Условия рождения хаотических структур в моделях динамики численности экосистем.

Цель данного пункта состоит в том, чтобы сформулировать конструктивные условия существования сильно возвратной структуры в моделях ди-

намики численности популяций с неперекрывающимися поколениями. При отсутствии лимитирования $s_t \equiv 0$ ($C \equiv 0$) решения уравнения (2.3.3) стремятся к одному нормальному магистральному множеству, порождаемому собственным вектором H_0 оператора $F(0, X)$. Естественно рассматривать невырожденную биологическую систему, т.е. предполагать, что собственное значение $\lambda(0) > 1$. В некоторой окрестности начала координат, когда лимитирующий фактор достаточно мал, динамическая система (2.3.3) при $C \neq 0$ приближенно ведет себя так же, как и в тривиальном случае $s_t \equiv 0$, т.е. решения стремятся к неустойчивому инвариантному многообразию W^U (см. теорему 2.3.1), являющемуся в данном случае магистральным множеством системы (2.3.3). С ростом лимитирующего фактора скорость воспроизводства падает и при критических значениях $s = s_\infty$ наступает вырождение биологической системы. Поэтому существование сильно возвратной структуры в нуле обеспечивается пересечением некоторого продолжения локального многообразия W^U с гиперплоскостью $\langle C, X \rangle = s_\infty$. В этом состоит основная идея формулируемых ниже теорем.

В дальнейшем будем считать, что выполнены условия (2.3.4) и (2.3.5) и наименьшее значение скалярного аргумента $s = s'$, при котором $\lambda(s') = 1$, больше нуля.

Теорема 2.3.3. Пусть выполнено условие (2.3.5) и $\varphi: (0, 1] \rightarrow \overset{\circ}{K}$, $\varphi(0) = 0$ — непрерывная кривая. Если при некотором $\epsilon > 0$

$$\max \{ \lambda(s) \mid 0 < s < s' - \epsilon \} \geq s_\infty, \quad (2.3.10)$$

то найдется неотрицательный линейный функционал, при котором система (2.3.3) имеет сильно возвратную структуру, причем

$$SR_0 \cap \varphi(0, \tau] \neq \emptyset, \text{ когда } \tau \in (0, 1].$$

Доказательству теоремы предварим лемму.

Лемма 2.3.3.1. Для любой непрерывной кривой φ , отображающей $(0, 1]$ во внутренность K и такой, что $\varphi(0) = 0$, любого $C \geq 0$ и любого ϵ найдется номер n такой, что

$$[0, s' - \epsilon] \subset \langle C, F^n(\varphi[0, 1]) \rangle.$$

Доказательство. Допустим, что найдутся $X \in \overset{\circ}{K}$ и $\epsilon > 0$ такие, что при любом n имеет место $\langle C, F^n(X) \rangle \leq s' - \epsilon$. Пусть $s_m = s' - \epsilon$ и $\lambda_m = \lambda_m(s_m)$. По условию (2.3.6) существует отвечающий λ_m собственный вектор H_m оператора $F_m = F(s_m, \cdot)$ такой, что $F_m^n X = \lambda_m^n (H_m + o(1)) \rightarrow 0$ слабо при $n \rightarrow \infty$. В силу монотонности $F(s, \cdot)$ и предположения $s \in [0, s' - \epsilon]$ имеем $F^n(X) \geq F_m^n X$. Получили противоречие. Следовательно, для любой отличной от начала координат точки кривой $\varphi(\tau)$ найдется n , при котором $\langle C, F^n(\varphi(\tau)) \rangle \geq s' - \epsilon$. По условию $\langle C, F^n(\varphi(0)) \rangle = 0$. Из связности образа $\langle C, F^n(\varphi(t)) \rangle$ следует доказательство леммы.

Перейдем к доказательству теоремы. Пусть $s_* \in [0, s')$ таково, что $\lambda(s_*) s_* \geq a$. Положим $F_* = F(s_*, \cdot)$ и $\lambda_* = \lambda(s_*)$. Пользуясь сублинейностью F_* , найдем такой опорный линейный оператор $A \geq 0$ [17], что $F_*(Y) \geq AY$ при любом $Y \in E$ и $AH_* = F_*H_* = \lambda_*H_*$, где H_* — собственный вектор F_* , отвечающий собственному значению λ_* . Пусть C — собственный вектор сопряженного оператора A^* , отвечающий λ_* [36, с. 335]. По лемме 2.3.3.1 найдутся n и $\tau \in (0, 1]$ такие, что $Y = F^n(\varphi(\tau))$ и $\langle C, Y \rangle = s_*$. Тогда $s_{n+1} = \langle C, F_*(Y) \rangle \geq \langle C, AY \rangle \geq a$. Поскольку 0 — неподвижная точка оператора F , существует $\alpha < \tau$, при котором $s_{n+1}(\varphi(\alpha)) = a$. Используя условие (2.3.8), получаем доказательство теоремы.

Зададим ветвь неустойчивого инвариантного многообразия оператора $F(0, X)$, которое касается собственного вектора H_0 ($F(0, H_0) = \lambda(0)H_0$) в

точке 0 (см. теорему 2.3.1), функцией $\theta: [0, 1] \rightarrow K$. Воспользовавшись предыдущей теоремой, получаем первое условие рождения сильно возвратной структуры в окрестности начала координат.

Теорема 2.3.4. Пусть функционал C задан. Тогда любое решение уравнения (2.3.3), начинающееся в $K \setminus 0$, находится на конечном расстоянии от полупространства $\langle C, X \rangle \geq s'$ лишь конечное время. Если

$$\max \{ \langle C, F^n \theta(\tau) \rangle \mid \tau \in [0, 1] \} \geq s_\infty \quad (2.3.11)$$

при некотором конечном n , то система (2.3.3) имеет сильно возвратную структуру в нуле и $SR_0 \cap \theta(0, \tau]$ при любом $\tau \in (0, 1]$.

Доказательство. Первая часть утверждения доказана в лемме 2.3.3.1. Вторая часть доказывается дословным повторением последних двух предложений из доказательства теоремы 2.3.3.

Теорема 2.3.5. Пусть выполнено условие (2.3.5) и функционал C задан. Если существуют числа $s < s'$, $\alpha > 0$ и $\beta \geq \alpha$ такие, что

$$\alpha C_0 \leq C \leq \beta C_0, \quad (\alpha/\beta) \lambda(s) s \geq s_\infty, \quad (2.3.12)$$

где C_0 — собственный вектор оператора $A^*: E^* \rightarrow E^*$, сопряженного некоторому опорному к $F(s, \cdot)$ линейному оператору A , то система (2.3.3) имеет в нуле сильно возвратную структуру. Эта структура имеет непустое пересечение с любой непрерывной кривой, выходящей из начала координат и лежащей в $K \cup \{0\}$.

Доказательство. Ход рассуждений тот же, что и в теореме 2.3.3.

Замечание. Если уравнение (2.3.3) определено в одномерном пространстве, то условие (2.3.10) является необходимым и достаточным для существования на положительной полуоси сильно возвратной структуры в нуле.

Доказательство необходимости. Точка s' является неподвижной. Если $x_{t-1} > s'$ и $s_{t-1} < s_\infty$, то $\lambda(x_{t-1}) < 1$ и $\lambda(x_{t-1})x_{t-1} < s_\infty$. Поэтому верна импликация

$$x_{t+1} = 0 \Rightarrow x_t = s_\infty \Rightarrow x_{t-1} < s'.$$

Приведем теперь результат об условиях рождения хаотических режимов в системе (2.3.3). Предположим, что оператор F непрерывно дифференцируем по Фреше и его дифференциал удовлетворяет условию Липшица. Обозначим через $W_+^U = W^U \cap K$ ветвь инвариантного многообразия, касающуюся в нуле H_0 .

Теорема 2.3.6. Если выполнены условия:

1. Существует число p такое, что кривая $F^p(W_+^U)$ пересекает гиперплоскость $\langle C, X \rangle = s_\infty$ трансверсально в некоторой точке $Y > 0$.

2. $\partial/\partial s F^{n-1}(F(s, Y))|_{s=s_\infty} > 0 (< 0)$,

где n — наименьшее число шагов, за которое гиперплоскость $\langle C, X \rangle = s_\infty$ переводится в начало координат.

3. Существует окрестность s точки Y такая, что множество $s_+ = \{X \in s \mid \langle C, X \rangle \leq s_\infty (\geq s_\infty)\}$ отображается суперпозициями операторов $F, F \circ F, \dots, F^n$ в конус K , т.е. $F^i(s_+) \subset K$ при $i \leq n$, то динамическая система (2.3.3) обладает свойством Ch и образцы инъекций, фигурирующих в определении Ch, принадлежат K .

Доказательство. Достаточно показать, что условия теоремы 2.3.2 выполнены, и убедиться в принадлежности траекторий положительному конусу. Проверим выполнение условий теоремы 2.3.2. Из монотонности и положительности оператора $F(0, \cdot)$ следует, что $DF(0)$ – положительный линейный оператор. Из сублинейности $F(0, \cdot)$ следует, что $\lambda(0)$ и H_0 оператора $F(0, \cdot)$ являются собственным значением и собственным вектором линейного оператора $F(0, \cdot)$. Зададим ветвь многообразия W_+^U функцией $\varphi: [0, 1] \rightarrow K$ так, чтобы $\varphi(0) = 0$, $\varphi' \neq 0$ и $Y = \varphi(0, 5)$. Из условия 2 теоремы следует, что $D_t F^n(\varphi(0, 5)) > 0$ (< 0). Действительно, $D_t F^n(\varphi(t)) = \langle \partial/\partial \sigma F^{n-1}(F(\sigma, X)) \partial/\partial t < C, \varphi(t) \rangle$ при $t = 0,5$, поскольку $F^n(X) = 0$ при $\sigma = a$ и любом X . Аналогичный результат верен для произвольной кривой, близкой к φ в C^1 -топологии в окрестности точки $\varphi(0,5)$. Так как по предположению (2.3.6) $K \cap E_1 = \{0\}$, из условия 2 теоремы имеем, что $F^{n+k}|_{W_+^U} \uparrow E_1$, т.е. условие 3 теоремы 2.3.2 выполнено. Кроме того, $F^n(Z) \in K$ при всех $Z \in s_+$ и $\langle C, Z \rangle \neq a$. Пусть $X \in (F^k)^{-1}(Y) \cap W_+^U$ – точка из окрестности нуля. Из условия 1 теоремы 2.3.6 следует, что локальное $m-1$ многообразие $U_X \cap (F^k)^{-1}\{X \in K | \langle C, X \rangle = s_\infty\}$, где U_X – окрестность точки X , трансверсально W_+^U . Пусть $U_X^+ = (F^k)^{-1}s_+$. При индуктивном построении множеств в теореме 2.3.2 достаточно ограничиться рассмотрением части U_X^+ окрестности U_X точки X , так как X принадлежит положительной ветви многообразия W_+^U . Поскольку $F^{k+n}(U_X^+) \subset K \cup \{0\}$, то $F^{k+n}(U_X^+) \cap E_1 = \{0\}$ и условие 2 теоремы 2.3.2 выполнено. Теорема доказана.

Пример 1. Динамика популяции с возрастной структурой и конечным числом возрастных групп обычно задается уравнением вида (см. формулу (2.3.1))

$$X_{t+1} = aP(s_t)AX_t, \quad s_t = \langle C, X \rangle,$$

где $X \in R_+^m$, A – неразложимая неотрицательная матрица, $P = \text{diag}\{p_i(s)\}_{i=1}^m$ – диагональная матрица отображений $p_i: [0, s_\infty) \rightarrow (0, \infty)$, $p_i(s_\infty) = 0$ ($i = 1, \dots, m$) и a – положительная константа. Из теоремы 2.3.3 следует, что найдутся такое $a > 0$ и такой набор положительных констант $(c_1, c_2, \dots, c_m) = C$, что система имеет в нуле сильно возвратную структуру. Условие трансверсальности пересечения многообразия W_+^U с гиперплоскостью $\langle C, X \rangle = s_\infty$ из теоремы 2.3.6 всегда можно достичь малыми вариациями вектора C . Поэтому в силу теоремы 2.3.6 найдутся такое число a и такой вектор $C > 0$, что система обладает свойством Ch .

Пример 2. Рассматривается динамика сообщества хозяин – паразит, заданная моделью (2.3.2). Пусть $q(s)$ – дважды непрерывно дифференцируемая функция, удовлетворяющая условиям $q(s) = 0$ при $s \leq 1 - 1/a$ и $q(s) = 1$ при $s \geq 1$. Проверим выполнение условий теоремы 2.3.6. Неустойчивое инвариантное многообразие задается отрезком прямой $t(1, b_1/a, \dots, b_1, \dots, b_m/a^m)$ при $t \in [0, (1 + 1/a)(c + \sum_{i=1}^m b_1 \dots b_i/a^i)]$. Для существо-

вания сильно возвратной структуры, пересекающейся с неустойчивым многообразием, достаточно потребовать (см. теорему 2.3.4), чтобы $as \geq 1$ при $s \leq 1 - 1/a$, т.е. $a \geq 2$. Легко видеть, что в этом случае образ неустой-

чивого многообразия за один шаг итерации трансверсально пересекает гиперплоскость $cx + \sum_{i=1}^m y^i = 1$ в точке $Y = (1, b_1/a, \dots, b_1 \dots b_m/a^m)/(c + b_0)$,

где $b_0 = \sum_{i=1}^m b_1 \dots b_i/a^i$. Пусть множество S_+ из условия теоремы 2.3.6

образовано частью окрестности точки Y , лежащей в полупространстве $s \leq 1$. Тогда условия 2 и 3 теоремы 2.3.6 выполнены, если $s(F^i(Y)) < 1$ при $1 \leq i \leq m$. Легко видеть, что

$$F(Y) = (0, b_1, b_1 b_2/a \dots b_1 \dots b_m/a^{m-1})/(c + b_0),$$

$$F^{m-1}(Y) = b_1 \dots b_m/(c + b_0).$$

Отсюда получаем систему неравенств

$$b_1 + b_1 b_2/a + \dots + b_1 \dots b_m/a^{m-1} \leq c + b_0,$$

$$b_1 b_2 + \dots + b_1 \dots b_m/a^{m-2} \leq c + b_0,$$

.

$$b_1 \dots b_m \leq c + b_0,$$

которая выполняется, если $c > \max \left\{ \sum_{i=j}^m b_1 \dots b_i/a^{i-j} - b_0 \mid 1 \leq j \leq m \right\}$.

Последнее условие вместе с неравенством $a \geq 2$ гарантирует возникновение свойства Ch.

Сформулируем экологический смысл данного результата. Пусть каждая особь в популяции жертвы при благоприятных условиях способна оставлять после себя не менее двух потомков. Скорость воспроизводства популяции жертвы убывает достаточно быстро, если размер этой популяции возрастает. При численности сообщества, большей, чем критическая, популяция жертвы не размножается. Тогда в рассматриваемом сообществе возможны практически непредсказуемые взрывы численности, сменяющиеся почти полным исчезновением вида. Например, пусть в непроточном культиваторе выращивается культура бактерии *E. Coli*. При этом выполняются перечисленные выше условия: число потомков каждой бактерии равно двум и при превышении некоторого уровня численности скорость роста замедляется вплоть до нуля (переход в стационарную фазу). Пусть роль хищника играет бактериофаг, которым заражена бактерия (у хищника одна возрастная группа). Скорость размножения бактериофага пропорциональна числу бактериальных клеток. В такой ситуации у сообщества возможны хаотические изменения численности.

Реальная действительность накладывает ограничения на управления, с помощью которых человек пытается воздействовать на живую природу. В большинстве случаев эти ограничения относятся к амплитуде допустимых воздействий и, что значительно более принципиально, к их направлению. Другими словами, в задачах математической экологии фиксированным оказывается знак допустимых управлений. При этом нулевое управление, т.е. отсутствие всяких воздействий, становится граничной точкой. Примером тому служат отлов или отстрел особей без пополнения извне, удаление приростшей биомассы при непрерывном культивировании простейших, укосы, прореживания, внесение удобрений, поливы без осушения и другие агротехнические мероприятия в сельском хозяйстве.

Естественно возникает вопрос, обладают ли модели динамики численности экосистемы такими свойствами, которые позволили бы эффективно использовать однонаправленные воздействия. В данной главе приводятся результаты, которые дают положительный ответ на поставленный вопрос. Отдельно исследуются задачи локальной и глобальной управляемости. При решении задач стабилизации показано, что естественным является использование управлений переменной структуры и введение скользящих режимов. Приведены примеры такого управления, нашедшие широкое распространение в микробиологической практике.

Задача определения состояния многокомпонентной экологической системы по наблюдениям за частью компонент является двойственной задаче управления. Она разрешима, если в распоряжении исследователя наряду с информацией, которую обеспечивает эксперимент, имеется модель изучаемого объекта. Задача получения биологически осмысленных результатов по искаженной помехами информации является двойственной задаче управления однонаправленными воздействиями. Показано, как решать эту задачу при наличии аналитической или имитационной модели.

§ 3.1. Управление нелимитированными экологическими системами

При малых размерах экологической системы и обилии источников питания влиянием лимитирующих факторов на динамику воспроизводства можно пренебречь. Это означает, что отсутствует отрицательная обратная связь между численностью сообщества и скоростью воспроизводства. Такое состояние называется фазой экспоненциального или логарифмического

роста. Управляющие воздействия в этой фазе состоят в добавлении особей в экосистему (и только добавлении) или изъятии (и только изъятии). Иногда представлено совместное действие этих механизмов. Для моделей (2.1.2) и (2.1.3) отсутствие лимитирования означает, что скорость воспроизводства не зависит от величины лимитирующего фактора, т.е. $F(X, s) = F(X)$ при всех $s \geq 0$.

Рассмотрим пример. Пусть скалярная величина x задает численность популяции вредителей и u — интенсивность уничтожения особей. Простейшая модель процесса имеет вид

$$\dot{x} = ax - u,$$

где $u \geq 0$. Совершенно очевидно, что за конечное время можно добиться того, что численность популяции станет равной нулю. Учет внутренней структуры экосистемы приводит к многомерной модели, и в этом случае задача принципиально усложняется. Рассмотрим простейший пример:

$$\dot{x} = ax + by - b_1 u,$$

$$\dot{y} = cx + dy - b_2 u,$$

где x и y — численность особей первой и второй популяции, $c \geq 0$ и $b \geq 0$ — скорость, с которой особи одной популяции переходят в другую, b_1 и b_2 — влияние управляющего фактора на каждую популяцию (например, $b_1 = b_2 = 1$ или $b_1 = 0$ и $b_2 = 1$), $u \geq 0$ — интенсивность воздействия.

Пусть выполнено условие управляемости по Калману [14]:

$$\det \begin{bmatrix} b_1 & ab_1 + bb_2 \\ b_2 & cb_1 + db_2 \end{bmatrix} \neq 0.$$

Тем не менее задача достижимости начала координат для всех точек $x(0) \geq 0$ и $y(0) \geq 0$ не имеет решения.

Действительно, поскольку система имеет одно вещественное собственное значение, оба ее собственных значения вещественны. Если $c \neq 0$ и $b \neq 0$, то матрица в правой части неразложима. Следовательно, существует вещественный собственный вектор $(v_1, v_2)^*$ транспонированной матрицы, у которого $v_1 < 0$ и $v_2 > 0$. Поскольку выражение $(b_1 v_1 + b_2 v_2)u$ знакопостоянно при $u \geq 0$, недостижимость нуля следует из теоремы 3.1.3 (см. ниже).

Если $b = 0$ и $b_1 = 0$, то $\dot{x} = ax$ и задача попадания в начало координат за конечное время неразрешима. Отсюда $b_1 > 0$. В этом случае при $x(0) = 0$, $y(0) > 0$ и любом $u(t) \geq 0$ движение $x(t)$ становится отрицательным и, следовательно, задача достижимости нуля неразрешима. Утверждение доказано.

Неуправляемость систем второго порядка не означает, что при увеличении размерности задача достижимости нуля тем более неразрешима. В качестве примера рассмотрим модель динамики численности популяции микроорганизмов, имеющих m фаз развития и делящихся в последней фазе, производя q потомков ($q > 1$). Пусть управление направлено на уничтожение популяции путем дозированного введения биологически активного препарата. Таким образом, приходим к модели $\dot{X} = AX - Bu - C$, где A имеет вид (1.2.2), $B \in R_+^m$ — чувствительность клеток разных фаз развития

к препарату, $u(t)$ – интенсивность поступления препарата, $C \in R_+^m$ – интенсивность гибели клеток в результате действия других факторов. Перенося начало координат в точку $A^{-1}C \in R_+^m$ (см. формулу (2.1.24)) с помощью замены переменных $X = Y + A^{-1}C$, приходим к системе $\dot{Y} = AY - Bu$. Отметим, что отрицательные значения вектора $Y \geq -A^{-1}C$ имеют биологический смысл.

Предположим, что пара (A, B) управляема по Калману [14]. Тогда при четном числе фаз развития m задача попадания в стационарную точку за конечное время неразрешима, а при нечетном m разрешима.

Для доказательства воспользуемся теоремой 3.1.3 (см. ниже). Собственные значения матрицы A^* задаются выражением

$$\lambda_k = q^{1/m} \left(\cos \frac{2\pi k}{m} + j \sin \frac{2\pi k}{m} \right) - 1 \quad (k = 0, 1, \dots, m-1).$$

При четных m матрица A^* имеет вещественный собственный вектор $Y \in R_+^m$ и выражение Y^*Bu при $u \geq 0$ не меняет знак. При нечетных m матрица A^* не имеет вещественных собственных векторов.

Обратимся к изучению модели нелимитированной экологической системы общего вида. Пусть численный состав биологического сообщества задается вектором $X \in R_+^m$. Управляющие воздействия на систему предполагаем однонаправленными, т.е. $U \in R_+^p$ ($U \in R_+^p$). Предположение об однонаправленности управлений не является ограничивающим. Действительно, пусть некоторое управление u_i может принимать значения разных знаков, т.е. $-\infty < u_i < +\infty$. Вводя вспомогательные переменные $v_i \leq 0$ и $w_i \leq 0$ и пользуясь представлением $u_i = v_i - w_i$, приходим к однонаправленному случаю. Таким образом, всякое двустороннее управление эквивалентно двум однонаправленным.

По аналогии с моделями (2.1.2) и (2.1.3) рассматриваем кооперативный характер взаимодействий. Математически это означает квазимонотонность оператора правой части модели (см. п. 2.1.1). На протяжении данного параграфа этот оператор считаем непрерывно дифференцируемым. Таким образом, имеем

$$\dot{X} = F(X, U), \tag{3.1.1}$$

где $F(X, U) = (f_i(X, U))_{i=1}^m$ и $\partial f_i / \partial x_j \geq 0$ ($i, j = 1, \dots, m$ и $i \neq j$). Положительная однородность функции $F(X, U)$ (см. п. 2.1.1) в данном параграфе не используется. Пусть X_0 – стационарное решение системы, если оно существует; тогда $F(X_0, 0) = 0$.

Предположим, что в начальный момент времени численный состав экологической системы задается нулевым вектором или вектором, не превосходящим X_0 . Если эту систему необходимо перевести в равновесное состояние за конечное время, то возникает задача достижимости стационарной точки X_0 из точек конуса $X \leq X_0$. В этом случае, когда требуется уменьшить численность сообщества до равновесного, возникает проблема достижимости точки X_0 из точек конуса $X \geq X_0$. Математически эти задачи эквивалентны, и их решению посвящена первая часть параграфа. Вторая часть параграфа излагает решение той же проблемы при дополнительных ограничениях на фазовые переменные. Эти ограничения заключаются в том, что движение системы во все моменты времени должно принадлежать

конусу начальных состояний. Такая ситуация возникает, например, в задаче уничтожения популяции за конечное время. При движении к нулевой численности переменные состояния системы в силу биологического смысла должны принимать неотрицательные значения.

3.1.1. Управление вблизи равновесного состояния. В результате линеаризации системы (3.1.1) в окрестности равновесного состояния получаем уравнение вида

$$\dot{X} = AX + BU, \quad (3.1.2)$$

где $A - m \times m$ -матрица с неотрицательными недиагональными элементами, $B - m \times p$ -матрица и U - управление со значениями в R^p (R^p_+). Обычно матрица B имеет компоненты одного знака ($B \geq 0$). Так как конус R^m_+ (R^m_+) положительно инвариантен в случае $U = 0$, то при $U \in R^p_+$ попадание в нуль возможно для точек R^m_+ и при $U \in R^p_+$ для точек R^m_+ . В дальнейшем для определения будем считать, что $U \in R^p_+$ и движение начинается в R^m_+ . Ранг матрицы $[B, AB, \dots, A^{m-1}B]$ будем обозначать через $\text{rang}(A, B)$.

Определение. Множество точек, из которых решение можно перевести в начало координат за конечное время с помощью допустимых управлений, будем называть множеством достижимости.

При ограничении на знак управления и отсутствии каких-либо ограничений на знак матриц A и B множество точек, для которых начало координат достижимо, образует выпуклый конус [14]. Условие телесности этого конуса [12, 18], т.е. наличия в нем внутренних точек, сводится к управляемости пары (A, B) по Калману [14].

Теорема 3.1.1. Для того чтобы множество достижимости системы (3.1.2) содержало внутренние точки, необходимо и достаточно, чтобы $\text{rang}(A, B) = m$ [29].

Доказательство. Множество достижимости системы (3.1.2) совпадает с множеством концов решений, определенных при $t \leq 0$ и начинающихся в начале координат:

$$K = \left\{ \int_0^t e^{-A(s-t)} BU(s) ds \mid U(s) \geq 0 \text{ при } s \geq 0 \right\}.$$

Варьируя U , получаем $\bar{K} = \{ e^{-At} BU \mid U \geq 0 \text{ и } t \geq 0 \}$. Ввиду выпуклости телесность K и \bar{K} имеет место одновременно [25, с. 53]. Для телесности \bar{K} необходимо и достаточно, чтобы сопряженный конус $K^* = \{ C \in R^m \mid \langle C, K \rangle \geq 0 \}$ был собственным [25] или выступающим, т.е. не содержал линейных подпространств. Доказательство этого факта см. [12, с. 393]. Покажем, что если K^* - несобственный, то пара (A, B) неуправляема. Пусть существует вектор $C \in R^m$ такой, что $-C^* e^{-At} B \geq 0$ и $C^* e^{-At} B \geq 0$ при $t \geq 0$. Отсюда $B^* e^{-A^* t} C = 0$. Последовательным дифференцированием получаем, что $\text{rang}(A, B) < m$. Пусть, наоборот, $\text{rang}(A, B) = m$. Тогда собственный вектор матрицы $[B, AB, \dots, A^{m-1}B]^*$, отвечающий нулевому собственному значению, порождает линейное подпространство в конусе K^* . Теорема доказана.

Таким образом, управляемость по Калману гарантирует, что множество достижимости не является пренебрежимым (множеством нулевой меры), и в дальнейшем всегда будет предполагаться выполненной.

Рассмотрим вначале случай, когда матрица A неразложима. Пусть Y - собственный вектор матрицы A^* , отвечающий собственному значению с

максимальной вещественной частью $\lambda(A)$. Обозначим через V произвольный собственный вектор матрицы A^* , который отвечает отличному от $\lambda(A)$ вещественному собственному значению (если оно существует). Вследствие неразложимости по крайней мере две компоненты вектора V имеют разные знаки. Выпишем условия, при которых исследуется структура множества достижимости:

$$\langle Y, BU \rangle \text{ меняет знак при } U \in R^p_-, \quad (3.1.3)$$

$$\langle Y, BU \rangle \leq 0 \text{ при } U \in R^p_-, \quad (3.1.4)$$

$$\langle V, BU \rangle \text{ меняет знак при } U \in R^p_-. \quad (3.1.5)$$

Теорема 3.1.2. Предположим, что $\text{rank}(A, B) = m$. Для того чтобы начало координат было достижимо для любой точки из R^m , необходимо и достаточно, чтобы выполнялись условия (3.1.3) и (3.1.5) [26].

В экологических задачах матрица B обычно имеет неотрицательные элементы и условие (3.1.3) не выполняется. Рассмотрим более общую ситуацию, когда матрица B может иметь элементы разных знаков, но условие (3.1.3) нарушено. В этом случае применима

Теорема 3.1.3. Пусть $\text{rank}(A, B) = m$, матрица A неразложима и вполне условие (3.1.4). Для того чтобы множество достижимости системы (3.1.2) содержало R^m_+ , необходимо и достаточно, чтобы выполнялось условие (3.1.5).

Доказательство. Достаточность. Ввиду автономности системы множество точек K , из которых начало координат достижимо для процесса (3.1.2), совпадает с множеством концов решений, определенных при $t \leq 0$ и начинающихся в нуле. Переходя к отрицательному времени, получаем систему $\dot{X} = -AX + BU$, где $U \in R^p_+$. Отсюда $K = \{ X = \int_0^t e^{-A(t-s)} BU(s) ds \mid U(s) - \text{измеримые функции } U: R^1_+ \rightarrow R^p_+ \} -$ выпуклый конус. Варьируя U , получаем, что $\bar{K} = \{ e^{-At} BU \mid U \geq 0 \text{ и } t \geq 0 \}$.

Покажем, что сопряженный конус $K^* \not\subset \{ C \in R^m \mid \langle C, K \rangle \geq 0 \} \subset R^m_+$. Будем рассуждать от противного. Пусть $K^* \neq R^m_+$ и $C = -Y \in \bar{K}^*$. Тогда $\langle C, e^{-At} BU \rangle = e^{-\lambda(A)t} \langle C, BU \rangle \geq 0$. Из предположения (3.1.4) и условия управляемости пары (A, B) заключаем, что $\langle C, BU \rangle < 0$ при некотором $U \geq 0$. Следовательно, $-Y \notin \bar{K}^*$. Таким образом, $C \in K^* \setminus R^m_+$ имеет ненулевую проекцию на подпространства, отвечающие другим собственным значениям матрицы A . Воспользуемся разложением экспоненты по компонентам матрицы A :

$$e^{-At} = \sum_{k=1}^s e^{-\lambda_k t} \sum_{i=1}^{n_k} \frac{t^{i-1}}{(i-1)!} Z_{ki} \quad (s \leq m).$$

Пусть $\lambda_1 = \lambda(A) \geq \text{Re } \lambda_2 > \dots \geq \text{Re } \lambda_k$ ($\lambda_i \neq \lambda_j$ при $i \neq j$) — спектр матрицы A и j — максимальный номер подпространства, для которого

$$\left\langle C, \sum_{i=1}^{n_j} \frac{t^{i-1}}{(i-1)!} Z_{ji} \right\rangle \neq 0. \text{ Как показано в работе [26, лемма 2.8],}$$

$$C^* \sum_{i=1}^{n_k} \frac{t^{i-1}}{(i-1)!} Z_{ki} = t^{m-k} (V_k^* + o(1)),$$

где V_k – собственный вектор матрицы A^* , отвечающий λ_k и $m_k \leq n_k$. Таким образом, возможны следующие ситуации:

$$\langle C, e^{-At}BU \rangle = e^{-\operatorname{Re} \lambda_j t} m_j (a_j(t, U) + o(1)),$$

где $\operatorname{Im} \lambda_j \neq 0$, $a_j(t, U) = e^{\operatorname{Im} \lambda_j t} V_j^* BU + e^{\operatorname{Im} \bar{\lambda}_j t} \bar{V}_j^* BU$ – почти периодическая функция [26, лемма 2.8] и либо $\operatorname{Re} \lambda_j > \operatorname{Re} \lambda_{j-1}$, либо $m_j > m_{j-1}$;

$$\langle C, e^{-At}BU \rangle = e^{-\lambda_j t} t^{m_j} (V_j^* BU + o(1)),$$

где $\operatorname{Im} \lambda_j = 0$ и либо $\lambda_j > \operatorname{Re} \lambda_{j-1}$, либо $m_j > m_{j-1}$;

$$\langle C, e^{-At}BU \rangle = e^{-\lambda_j t} t^{m_j} (V_j BU + a_{j-1}(t, U) + o(1)),$$

где $\lambda_j = \operatorname{Re} \lambda_{j-1}$, $\operatorname{Im} \lambda_j = 0$, $\operatorname{Im} \lambda_{j-1} \neq 0$ и $m_j = m_{j-1}$.

В силу управляемости пары (A, B) заключаем, что $V^* BU \neq 0$ для любого собственного вектора V матрицы A^* . Отсюда в приведенных выше формулах $a_j(t, U) \neq 0$ при $\operatorname{Im} \lambda_j \neq 0$ и $V_j^* BU \neq 0$ при $\operatorname{Im} \lambda_j = 0$. В первом случае скалярное произведение меняет знак, поскольку $a_j(t, u)$ – сумма синусов [31, лемма 4.1; 26, лемма 2.8], во втором случае скалярное произведение меняет знак по условию теоремы, в третьем случае по тем же соображениям при подходящих значениях t и u выражение имеет значения различных знаков. Следовательно, $K^* \subset R_+^m$. Отсюда $\bar{K} = K^{**} \supset R_+^m$. Ввиду выпуклости [25, с. 53] $\bar{K} \supset \mathring{R}_+^m$. Из неразложимости матрицы A следует, что для любого ненулевого решения системы (3.1.2) $X(t) > 0$ при $t > 0$ и $U \equiv 0$ (предложение 2.1.1). Поэтому можно считать, что множество решений с начальными условиями из \mathring{R}_+^m совпадает с множеством решений, начинающихся в R_+^m . Достаточность доказана.

Необходимость. Пусть существует вещественный собственный вектор V_j матрицы A^* такой, что $\langle V_j, BU \rangle \geq 0$ при всех $U \in R_+^p$. Тогда $V^* \bar{K} = \{e^{-\lambda_j t} V_j^* BU(t) \mid U(t) \geq 0\} \geq 0$, т.е. $V_j \in K^*$. Ввиду неразложимости матрица A имеет единственный неотрицательный собственный вектор и, следовательно, $V_j \notin R_+^m$. Отсюда $K^* \not\subset R_+^m$. Таким образом, $\bar{K} = K^{**} \not\subset R_+^m$. Теорема доказана.

Следствие. Если $\operatorname{rank}(A, B) = m$, матрица A неразложима и все ее собственные значения, отличные от $\lambda(A)$, имеют ненулевую мнимую часть, то множество достижимости начала координат содержит R_+^m .

Полученный результат имеет следующий биологический смысл. Рассмотрим всевозможные скалярные функции вектора состояний системы. Среди них имеются такие, которые при нулевом управлении сохраняют знак во все моменты времени. Пусть набор управлений настолько богат, что они способны менять знак любой такой функции, за исключением той, которая задает средневзвешенную численность экологической системы. Пусть с помощью доступных управлений средневзвешенная численность может за конечное время быть обращена в нуль. В этом случае любое решение, имеющее в начальный момент времени неотрицательные компоненты, может за конечное время быть переведено в состояние равновесия. Следовательно, для такой экологической системы задача максимального быстрого действия или любая другая эквивалентная ей задача имеет решение.

Рассмотрим проблему управления экологической системой в случае, когда A — разложимая матрица. Биологический смысл разложимости состоит в том, что исходная экологическая система представляет собой объединение нескольких подсистем, находящихся на разных уровнях иерархии. Численность подсистемы, которая находится на высшем уровне иерархии, меняется в зависимости от численности образующих ее особей. Численность подсистемы, которая находится на промежуточном уровне иерархии, меняется в зависимости от численности подсистем высших уровней иерархии и своей собственной численности. Математически свойство разложимости матрицы линейного приближения означает, что с помощью соответствующей перенумерации переменных системы (3.1.2) эта матрица приводится к блочно-треугольному виду [7]. Рассуждая аналогично, будем считать, что матрица управлений B также имеет блочно-треугольный вид. Таким образом, имеем

$$A = \begin{bmatrix} A_1 & 0 & \dots & 0 \\ A_{21} & A_2 & & \\ \vdots & \vdots & & \\ A_{s1} & A_{s2} & \dots & A_s \end{bmatrix}, \quad B = \begin{bmatrix} B_1 & 0 & \dots & 0 \\ B_{21} & B_2 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ B_{s1} & B_{s2} & \dots & B_s \end{bmatrix}, \quad (3.1.6)$$

где $A_i = A_{ii}$ — неразложимая $m_i \times m_i$ -матрица, $B_i = B_{ii}$ — $m_i \times p_i$ -матрица ($i = 1, \dots, s$).

Если биологическая система разложима, то при обнулении компонент, соответствующих некоторому количеству первых блоков, т.е. при $X_i = 0 \in R^{m_i}$ и $U_i = 0 \in R^{p_i}$ ($i = 1, \dots, q < s$), эти компоненты остаются нулевыми во все оставшиеся моменты времени. Таким образом, разложимая система имеет по крайней мере s (считая R^m) инвариантных подпространств.

Для исследования разложимых систем введем некоторые дополнительные понятия.

Определение. Путем $\omega(i)$ с началом $i \in [1, \dots, s]$ называется набор возрастающих номеров $i = i_0 < i_1 < \dots < i_q$ максимальной длины таких, что $A_{i_{k+1} i_k} \neq 0$ ($k = 0, \dots, q - 1$) [32]. Объединение элементов путей, начинающихся в i , будем называть i -й компонентой связности и обозначать через Ω_i .

Приведем несколько примеров. Простейшим представителем разложимых матриц является матрица цепной системы [33]:

$$A = \begin{bmatrix} -a_1 & 0 & \dots & 0 \\ a_1 & -a_2 & \dots & 0 \\ 0 & a_2 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & -a_m \end{bmatrix}.$$

В этом случае компонента связности Ω_i образована единственным путем $\omega(i) = (i, i + 1, \dots, m)$. Матрица совокупности изолированных систем с

одним источником имеет вид [33]

$$A = \begin{bmatrix} a_{11} & 0 & \dots & 0 \\ a_1 & -b_1 & \dots & 0 \\ \vdots & & & \\ a_{m-1} & & & -b_{m-1} \end{bmatrix}.$$

В этом случае

$$\Omega_1 = \{(1, 2), (1, 3), \dots, (1, m)\} \text{ и } \Omega_i = \{i\} \quad (2 \leq i \leq m).$$

Интересный пример разложимости представляет модель динамики численности популяции микроорганизмов при многоступенчатом культивировании. Обращаясь к системе (1.2.4), получаем, что диагональные матрицы в формуле (3.1.6) имеют вид $A_0 - v_i I$ ($i = 1, \dots, s$), где A_0 задается выражением (1.2.2) или (1.2.3), а ненулевые поддиагональные матрицы имеют вид $A_{i+1, i} = u_i I$.

Пусть $X_i \in R^{m_i}$ и $U_i \in R_-^{p_i}$ ($i = 1, \dots, s$). Введем в рассмотрение векторы $P_k(X) = (X_i)_{i \in \Omega_k}$, $P_k(U) = (U_i)_{i \in \Omega_k}$ и матрицы $P_k(A) = \{A_{ij}\}_{i, j \in \Omega_k}$, $P_k(B) = \{B_{ij}\}_{i, j \in \Omega_k}$. Таким образом, оператор $P_k(\cdot)$ задает ограничение исходной системы на инвариантное подпространство, получаемое при $X_j = 0$ и $U_j = 0$, где $j \in \Omega_k$. Обозначим через $P_k(V)$ вещественный собственный вектор матрицы $P_k^*(A)$, имеющий по крайней мере две компоненты разных знаков.

Определение. Разложимую матрицу будем называть полуразложимой, если каждая ее поддиагональная матрица имеет хотя бы один ненулевой элемент, т.е. $A_{i+1, i} \neq 0$ ($i = 1, \dots, s-1$).

В случае полуразложимости $\Omega_1 \supset \Omega_2 \supset \dots \supset \Omega_s$ и оператор $P_k(\cdot)$ обозначает ограничение исходной системы на инвариантное подпространство, полученное при $X_i = 0$ и $U_i = 0$ ($1 \leq i \leq k-1$). Для таких систем имеет место

Теорема 3.1.4. Предположим, что матрица A полуразложима. Для того чтобы множество достижимости системы (3.1.2) содержало положительный конус, необходимо и достаточно, чтобы для любого собственного вектора $P_k(V)$ скалярное произведение $\langle P_k(V), P_k(B) P_k(U) \rangle$ принимало как положительные, так и отрицательные значения при $P_k(U) \leq 0$ и всех $k = 1, \dots, s$.

Доказательство. Пусть ρ – собственное значение матрицы A^* и $Z = (Z_i)_{i=1}^s$, где $Z_i \in R^{m_i}$ ($i = 1, \dots, s$) – отвечающий ему собственный вектор. Легко видеть, что Z удовлетворяет системе

$$(\rho I - A_1^*) Z_1 = \sum_{k=2}^s A_{k1}^* Z_k,$$

$$(\rho I - A_g^*) Z_g = \sum_{k=g+1}^s A_{kg}^* Z_k,$$

.....

$$(\rho I - A_s^*) Z_s = 0.$$

Обозначим через $T_i = (T_{ki})_{k=1}^s$ и $\theta_i = (\theta_{ki})_{k=1}^s$ собственные векторы матрицы A^* , отвечающие собственным значениям $\lambda(A_i)$ и μ_i ($\text{Re } \mu_i < \lambda(A_i)$) матрицы A_i^* ($i = 1, \dots, s$) ($T_{ki} = \theta_{ki} = 0$ при $k > i$).

Лемма 3.1.4.1. Векторы $\theta_{ii} \in R_+^{m_i}$ ($i = 1, \dots, s$) с точностью до постоянно-го множителя. Если $T_{ki} \in R_+^{m_k}$ при $j < k \leq i$, то $T_{ji} \in R_+^{m_j}$ тогда и только тогда, когда $\lambda(A_j) > \lambda(A_i)$. Если $\lambda(A_j) = \lambda(A_i)$ при $i > j$, то собственного вектора T_i ($T_{ii} \neq 0$) не существует. Если $T_{ki} \in R_+^{m_k}$, то $T_{ki} \in \overset{\circ}{R}_+^{m_k}$ ($k \leq i$).

Доказательство. Поскольку $\theta_{ki} = 0$ при $k > i$ и θ_{ii} — собственный вектор неразложимой матрицы A_i , то $\theta_{ii} \in R_+^{m_i}$. Перейдем к рассмотрению T_i , у которого $T_{ki} = 0$ при $k > i$ и $T_{ii} > 0$ ввиду неразложимости матрицы A_i^* ($i = 1, \dots, s$). Пусть j_0 ($j \leq j_0 < i$) — ближайший к i номер, при котором $\lambda(A_{j_0}) \geq \lambda(A_i)$.

Рассуждаем по индукции:

$$T_{i-1,i} = (\lambda(A_i)I - A_{i-1}^*)^{-1} A_{i-1}^* T_{ii} > 0$$

ввиду неразложимости A_{i-1} и неравенства $\lambda(A_i) > \lambda(A_{i-1})$. Далее,

$$T_{j_0+1,i} = (\lambda(A_i)I - A_{j_0+1}^*)^{-1} A_{j_0+1}^* T_{j_0+2,i} + \dots > 0.$$

Тогда

$$(\lambda(A_i)I - A_{j_0}^*) T_{j_0,i} = A_{j_0+1}^* T_{j_0+1,i} + \dots$$

Пусть $X > 0$ — собственный вектор матрицы A_{j_0} , отвечающий $\lambda(A_{j_0})$. Умножим последнее равенство на X^* слева. Имеем $(\lambda(A_i) - \lambda(A_{j_0})) X^* T_{j_0,i} \geq 0$. Отсюда, если $\lambda(A_{j_0}) > \lambda(A_i)$, то $T_{j_0,i} \notin R_+^{m_{j_0}}$. Если $\lambda(A_{j_0}) = \lambda(A_i)$, то получаем противоречие с предположением $T_{ii} \neq 0$. Лемма доказана.

Доказательство достаточности проводится методом от противного. Пусть при некотором $C \in R_+^m$, всех $t \geq 0$ и $U \in R_+^p$ выражение $\Delta = \langle C, e^{-At} BU \rangle \geq 0$ (см. доказательство теоремы 3.1.3). Тогда

$$\Delta = \sum_{i=1}^s \alpha_i e^{-\lambda(A_i)t} \sigma_i(t) (T_i BU + o(1)) +$$

$$+ \beta_i e^{-\rho_i t} \delta_i(t) (a_i(t, U) + o(1)),$$

где $\sigma_i(t) = t^{r_i}$, $\delta_i(t) = t^{l_i}$ и

$$a_i(t, U) = \begin{cases} a_{1i} = \theta_i^* BU, & \text{если } \text{Im } \mu_i = 0, \\ a_{2i} = (e^{\text{Im } \mu_i t} \theta_i + e^{-\text{Im } \mu_i t} \bar{\theta}_i), & \text{если } \text{Im } \mu_i \neq 0, \\ a_{3i} = a_{1i} + a_{2i}, & \text{если } \text{Re } \mu_i = \nu_i, \end{cases}$$

где ν_i — вещественное собственное значение матрицы A_i . Константы α_i и β_i ($i = 1, \dots, s$) принимают значения $+1$, -1 и 0 . Последнее имеет место в случае, когда соответствующие собственные значения кратны или когда проекция вектора C на соответствующее инвариантное подпространство равна нулю. По построению все показатели экспонент в разложении Δ попарно различны.

Поскольку $\theta_{ki} = 0$ при $k > i$, функция $a_i(t, U)$ при $U_j = 0$ и $j < i$ зависит лишь от компоненты θ_{ii} и принимает значения разных знаков при $U_i \in R_+^{p_i}$ и некоторых, достаточно больших t (см. доказательство теоремы 3.1.3).

В соответствии с введенными выше понятиями функция $P_k(Y(t)) = P_k(e^{-At})^* P_k(C)$ обозначает решение уравнения $P_k(Y) = -P_k(A)^* P_k(Y)$ с начальным условием $P_k(Y(0)) = P_k(C)$. Разложим эту функцию по экспонентам, степени которых являются собственными значениями матрицы $P_k(A)$. Обозначим через $L_k(t)$ член ее разложения, содержащий $e^{-\lambda(A_i)t} = e^{-\lambda(A_1)t}$ ($i \geq k$). Он, очевидно, может равняться нулю.

Лемма 3.1.4.2. Пусть $\lambda(A_1)$ является кратным корнем матрицы A . Тогда имеет место одно из двух:

а) либо $L_1(t) = e^{-\lambda(A_1)t} \alpha_1 T_1$,

б) либо $L_k(t) = t^k e^{-\lambda(A_1)t} (\alpha_i P_k(T_i) + \epsilon(t))$ при некотором $k \geq 1$, где $r_k \geq 1$, $\epsilon(t) \rightarrow 0$ при $t \rightarrow \infty$ и собственный вектор $P_k(T_i) = (T_{ij})_{j=k}^s$ матрицы $P_k(A)^*$, отвечающий $\lambda(A_i) = \lambda(A_1)$, обладает свойством $\alpha_i T_{ji} \in R_+^{m_j}$ при некотором $j \leq i$.

Доказательство. Пусть i — максимальный номер, при котором $L_i(t) \neq 0$. Тогда $\lambda(A_1) = \lambda(A_i)$. Если $i = 1$, то выполнено условие а) леммы. Если $\lambda(A_1) = \lambda(A_k)$ для некоторого $1 < k < i$, то перейдем к рассмотрению ограничения $P_k(\cdot)$ на подпространство меньшей размерности. В связи с этим для простоты считаем, что $\lambda(A_1) = \lambda(A_i)$ и $\lambda(A_j) \neq \lambda(A_1)$ для всех $1 \leq j \leq s$, $j \neq i$.

Тогда $L_2(t) = e^{-\lambda(A_1)t} \alpha_2 P_2(T_i)$, где $P_2(T_i)$ — собственный вектор матрицы $P_2(A)$, отвечающий $\lambda(A_1)$. Если $\alpha_2 T_{ii} \in R_+^{m_i}$, т.е. $\alpha_2 = -1$, то лемма доказана при $k = i$. Пусть $\alpha_2 T_{ii} \in R_+^{m_i}$ ($\alpha_2 = 1$). Тогда по лемме 3.1.4.1 либо все векторные компоненты вектора $P_2(T_i) = (T_{ji})_{j=2}^s$ с первой до i -й строго положительны, либо при некотором $T_{ji} \in R_+^{m_j}$. Во втором случае лемма доказана при $k = j$.

Пусть $P_2(T_i) \in R_+^{m_2 + \dots + m_s}$. Найдем член разложения $L_1(t)$. Обозначим через $Z(t) \in R^{m_1}$ первую векторную компоненту этого вектора. Она является решением уравнения

$$\dot{Z} = -A_1^* Z(t) e^{-\lambda(A_1)t} \sum_{j=2}^s A_{j1}^* T_{ji}.$$

Отсюда $Z(t) = t e^{-\lambda(A_1)t} (-T_{11} + \epsilon(t))$, где $T_{11} = a Y_1 \in R_+^{m_1}$ — собственный вектор матрицы A_1^* и $\epsilon(t) \rightarrow 0$ при $t \rightarrow \infty$. Действительно, $Z(t) = e^{-A_1^* t} Z(0) - e^{-A_1^* t} \int_0^t e^{(A_1^* - \lambda(A_1))\tau} F d\tau$, где $F = \sum_{j=2}^s A_{j1}^* T_{ji} \geq 0$. Пусть X_1 ($\langle X_1, Y_1 \rangle = 1$) — собственный вектор матрицы A_1 , отвечающий $\lambda(A_1)$. Тогда $e^{A_1^* \tau} = e^{\lambda(A_1)\tau} (Y_1 X_1^* + \delta(\tau))$, где $\|\delta(\tau)\| \rightarrow 0$ при $\tau \rightarrow \infty$. Подставляя эту оценку в формулу Коши, имеем

$$Z(t) = e^{-\lambda(A_1)t} t (-a Y_1 + \epsilon(t)),$$

где $a = \langle X_1, F \rangle > 0$ и $\epsilon(t) \rightarrow 0$ при $t \rightarrow 0$. Таким образом, $L_1(t) = te^{-\lambda(A_1)t} \left[\begin{pmatrix} -aY_1 \\ 0 \end{pmatrix} + \epsilon(t) \right]$, где $\epsilon(t) \rightarrow 0$ при $t \rightarrow \infty$. Лемма доказана.

Будем говорить, что k -я векторная компонента вектора

$$D_i = (D_{ki})_{k=1}^s = \alpha_i T_i + \beta_i \theta_i = \alpha_i (T_{ki})_{k=1}^s + \beta_i (\theta_{ki})_{k=1}^s = 1$$

обладает свойством перемены знака, если либо $\beta_i > 0$, либо $\beta_i = 0$, $\alpha_i \neq 0$ и $T_{ki} \in R_+^{m \times k}$ при некотором $k \leq i$. Составим матрицу $D = [D_{ki}]_{k,i=1}^s$, где элементами D_{ij} являются m_i -мерные векторы ($i, j = 1, \dots, s$). Рассмотрим последовательность главных подматриц $D_r = \{D_{ij}\}_{i,j=r}^s$. Пусть j_0 — максимальный номер, при котором в подматрице содержится по крайней мере один элемент, обладающий свойством перемены знака. Поскольку $C \in R_+^m$, проекция этого вектора хотя бы на один из векторов со свойством перемены знака не равна нулю.

Действительно, если $\beta_i \neq 0$ при некотором $i = 1, \dots, s$, то этот факт очевиден. Пусть $\beta_i = 0$ ($i = 1, \dots, s$). Если проекция вектора C на каждое инвариантное подпространство, соответствующее собственному значению, совпадает с собственным вектором ($\sigma_i(t) = 1$ ($i = 1, \dots, s$)), то отсутствие перемены знака противоречит условию $C \in R_+^m$, так как сумма неотрицательных векторов дает неотрицательный вектор. Если инвариантное подпространство, соответствующее собственному значению, многомерно и проекция на него не совпадает с собственным вектором ($\sigma_i(t) = t^{r_i}$ и $r_i \geq 1$), то, воспользовавшись леммой 3.1.4.2, можно ограничить систему на инвариантное подпространство меньшей размерности (положив $u_i = 0$ при $i < k$) так, что один из собственных векторов этого ограничения обладает свойством перемены знака.

Пусть j_1 — номер первого ненулевого столбца матрицы D_{j_0} , удовлетворяющий неравенству $j_1 > j_0$. Номерам $i \geq j_1$, у которых $|\alpha_i| + |\beta_i| > 0$, сопоставим последовательность $\lambda(A_{j_1}), \dots, \lambda(A_{j_r})$ ($j_0 \leq j_1 \leq \dots \leq j_r \leq s$). Тогда имеет место

Лемма 3.1.4.3. Подматрица D_{j_0} обладает свойствами

$$\alpha_{j_1} > 0, \dots, \alpha_{j_r} > 0, \quad \beta_{j_1} = \beta_{j_2} = \dots = \beta_{j_r} = 0$$

и

$$\lambda(A_{j_1}) < \lambda(A_{j_2}) < \dots < \lambda(A_{j_r}).$$

Кроме того, возможно одно из двух:

а) вектор T_{j_1} не обладает свойством перемены знака, тогда j_0 -я компонента j_0 -го столбца матрицы D_{j_0} обладает свойством перемены знака и $\lambda(A_{j_0}) < \lambda(A_{j_1})$;

б) вектор T_{j_1} обладает свойством перемены знака, тогда $T_{j_0 j_1} \in R_+^{m j_0}$ и $\lambda(A_{j_0}) > \lambda(A_{j_1})$.

Доказательство. Условие $\beta_{j_l} \neq 0$ или $\alpha_{j_l} < 0$ при некотором $1 \leq l \leq r$ противоречит определению номера j_0 . Если $\lambda(A_{j_l-1}) \geq \lambda(A_{j_l})$

($i = 2, \dots, r$) и $\alpha_{ji} > 0$, то по лемме 3.1.4.1 имеем $T_{j_{i-1}j_i} \in R_+^{mj_{i-1}}$, что противоречит определению номера j_0 .

Рассмотрим случай а). В силу леммы 3.1.4.1 имеем $\lambda(A_{j_0}) < \lambda(A_{j_1})$. Пусть j_0 -я компонента j_0 -го столбца не обладает свойством перемены знака. Так как $T_{lj_0} = 0$ при $l > j_0$, этим свойством обладает какой-то из векторов T_k при $k > j_1$. Пусть $T_{ik} \in R_+^{mi}$. Тогда $i > j_0$, так как $\lambda(A_{j_0}) < \lambda(A_{j_1}) < \lambda(A_{j_k})$. Это противоречит определению числа j_0 .

Рассмотрим вариант б). Пусть $\lambda(A_{j_0}) \leq \lambda(A_{j_1}) < \lambda(A_{j_i})$ ($i = 2, \dots, r$). Тогда $T_{j_0j_i} \in R_+^{mj_0}$ и условие $T_{kq} \in R_+^{mk}$ выполняется при $q \geq j_1$ и $k > j_0$. Это противоречит определению номера j_0 . Следовательно, $\lambda(A_{j_0}) > \lambda(A_{j_1})$ и по лемме 3.1.4.1 $T_{j_0j_1} \in R_+^{mj_0}$.

Перейдем к построению управления, при котором $\Delta(U, t) < 0$. В случае а) положим $U_i = 0$ при $i \neq j_0$. Если $\beta_{j_0} = 0$, то $\Delta = e^{-\lambda(A_{j_1})t} [\alpha_{j_0} T_{j_0j_0}^* B_{j_0} U_{j_0} + o(1)] < 0$ при некоторых $U_{j_0} \in R_+^{pj_0}$ и достаточно больших t в силу условий теоремы 3.1.3. Теорема доказана.

Если $\beta_{j_0} > 0$, то $\Delta = e^{-\rho_{j_0}t} (a_{j_0}(t, 0, \dots, U_{j_0}, \dots, 0) + o(1)) < 0$ при некоторых, достаточно больших t и $U_{j_0} \in R_+^{pj_0}$ (см. доказательство теоремы 3.1.3). Теорема доказана.

В случае б) рассмотрим наиболее сложный вариант, когда j_0 -й столбец не является нулевым. Положим $U_1 = 0, \dots, U_{j_0-1} = 0$. Имеются следующие возможности:

б1) $\beta_{j_0} \neq 0$ и $\rho_{j_0} < \lambda(A_{j_1}) < \lambda(A_{j_0})$. Тогда управление строится так же, как и в случае а). Теорема доказана.

б2) $\beta_{j_1} = 0$ или $\beta_{j_1} \neq 0$ и $\lambda(A_{j_1}) < \rho_{j_0} < \lambda(A_{j_0})$.

Тогда $\Delta = e^{-\lambda(A_{j_1})t} [\alpha_{j_1} \sum_{l=j_0}^{j_1} \sum_{k=j_0}^l T_{kj_1}^* B_{j_1,l} U_l + o(1)] < 0$ при некоторых, достаточно больших t и некоторых $U_l \in R_+^{pl}$ ($j_0 \leq l \leq s$). Теорема доказана.

б3) $\rho_{j_0} = \lambda(A_{j_1})$. Здесь в свою очередь возможны варианты:

б3.1) Матрица A_{j_0} не имеет вещественных собственных значений, равных ρ_{j_0} . Тогда $\Delta = e^{-\lambda(A_{j_1})t} [\beta_{j_0} a_{j_0}(t, 0, \dots, U_{j_0}, \dots, 0) + \alpha_{j_1} T_{j_1}^* B U + o(1)] < 0$ при некоторых, достаточно больших t и некоторых $U_l \in R_+^{pl}$ ($j_0 \leq l \leq s$). Доказательство этого факта дословно совпадает с рассмотрением случая, когда $a_{j_0}(t, U) = a_{3j_0}(t, U)$ (см. доказательство теоремы 3.1.3). Теорема доказана.

б3.2) Матрица A_{j_0} имеет вещественное собственное значение, равное ρ_{j_0} . Если $\lambda(A_{j_1})$ и ρ_{j_0} отвечают различные собственные векторы, то их линейная комбинация также является собственным вектором матрицы A^* и по лемме 3.1.4.1 не принадлежит R_+^m . Следовательно,

$$\Delta = e^{-\lambda(A_{j_1})t} [\beta_{j_0} \theta_{j_0j_0}^* B_{j_0} U_{j_0} + \alpha_{j_1} T_{j_1}^* B U + o(1)] < 0$$

при некоторых $U_l \in R_+^{pl}$ ($j_0 \leq l \leq s$) и достаточно больших t .

Таким образом, доказано, что замыкание множества достижимости содержится в R_+^m . Воспользовавшись теми же соображениями, что и в теореме 3.1.3, получаем включение внутренностей этих множеств. Пусть k — номер первой ненулевой компоненты вектора начальных условий $X(0)$. Тогда $P_k(X(t)) > 0$ при $t > 0$ и $U \equiv 0$. Достаточность доказана.

Необходимость. Пусть $\langle P_k(V), P_k(B)P_k(U) \rangle \geq 0$ для $P_k(V) \in \sum_{i=k}^s R_+^{m_i}$. Рассмотрим ограничение исходной системы на инвариантное подпространство, полученное при $X_i = 0$ и $U_i = 0$ ($i = 1, \dots, k-1$). Дословно повторяя рассуждения, использованные при доказательстве необходимости в теореме 3.1.3, завершаем доказательство теоремы.

Невыполнение условия полуразложимости означает, что исходная система на некотором инвариантном подпространстве распадается на две независимые подсистемы. Действительно, пусть $A_{i+1,i} = 0$. Положим $X_k = 0$, $U_k = 0$ ($k = 1, \dots, i-1$), получаем систему вида

$$\dot{X}_i = A_i X_i + B_i U_i,$$

$$P_{i+1}(\dot{X}) = P_{i+1}(A)P_{i+1}(X) + P_{i+1}(B)P_{i+1}(U).$$

Вторая подсистема при $U_i = 0$ не зависит от первой. Для управляемости таких систем необходимы дополнительные ограничения на вид матрицы B . Для того чтобы в этом убедиться, рассмотрим следующий пример:

$$\dot{X}_1 = A_1 X_1 + B_1 U_1,$$

$$\dot{X}_2 = A_2 X_2 + B_{21} U_1 + B_2 U_2,$$

где $B_1 \geq 0$, $B_2 \geq 0$ и $B_{21} \not\geq 0$. При $X_1(0) \not\geq 0$ и $X_2(0) = 0$ задача достижимости нуля неразрешима. Действительно, при любом $U_1(t) < 0$ движение $X_2(t)$ переходит в $R_-^{m_2}$ и, так как $B_2 U_2 \leq 0$, в начало координат за конечное время попасть не может.

Для независимых подсистем, объединенных лишь совместными управляющими воздействиями, имеет место

Теорема 3.1.5. Пусть $A_{ij} = 0$ и $B_{ij} \leq 0$ при $i \neq j$, $i = 2, \dots, s$ и $j = 1, \dots, i-1$. Тогда множество достижимости начала координат содержит R_+^m .

Доказательство. Вначале построим управление $U_1(t)$ такое, что $X_1(T) = 0 \in R_+^{m_1}$. На промежутке времени $0 \leq t \leq T$ зададим управление в виде $U_1 = U_1(t)$ и $U_i = 0$ ($i = 2, \dots, s$). Так как $\dot{X}_2 = A_2 X_2 + B_{21} U_1 \geq \geq A_2 X_2$ при $t \in [0, T]$, то $X_2(t) \geq 0$. Далее рассуждаем по индукции. Теорема доказана.

Обратимся к изучению разложимых экологических моделей общего вида. Введем в рассмотрение класс дополнений компонент связности $C\Omega_k = \{k, k+1, \dots, s\} \setminus \Omega_k$ ($k = 1, \dots, s$). В соответствии с предыдущим результатом будем считать, что

$$a) B_{ij} \leq 0, \text{ если } j \in \Omega_k \text{ и } i \in C\Omega_k \text{ при некотором } k = 1, \dots, s \text{ и } i \in \bigcup_{p=1}^{k-1} \Omega_p$$

для любого такого k , и $B_{ij} = 0$, если $j \in \Omega_k$, $i \in C\Omega_k$ и $i \in \bigcup_{p=1}^{k-1} \Omega_p$ при

некотором $k = 1, \dots, s$. Заметим, что эти условия выполняются, если ограничиться более простым предположением

б) $B_{ij} = 0$, если $j \in \Omega_k$ и $i \in C\Omega_k$ при некотором $k = 1, \dots, s$. Для выполнения условия б) достаточно потребовать, чтобы $B_{ij} = 0$, если $A_{ij} = 0$ ($i = 2, \dots, s; j = 1, \dots, i - 1$).

Так или иначе, но будем считать, что одно из этих предположений имеет место. Отсюда следует, что управления, принадлежащие некоторой компоненте связности, влияют лишь на фазовые переменные своей компоненты и фазовые переменные компонент с большим порядковым номером. Влияние на последующие компоненты таково, что их фазовые переменные сохраняют принадлежность положительному конусу. Это дает возможность решать задачу попадания в начало координат поэтапно, в порядке возрастания номеров компонент.

Теорема 3.1.6. Пусть A — разложимая матрица общего вида и выполнены условия а) или б). Для того чтобы множество достижимости системы (3.1.2) содержало положительный конус R_+^m , необходимо и достаточно, чтобы выражение

$$\langle P_k(V), P_k(B)P_k(U) \rangle$$

принимало как положительные, так и отрицательные значения при $P_k(U) \leq 0$ и всех $k = 1, \dots, s$.

Доказательство. Рассуждаем по индукции. Допустим противное. Пусть $\Delta = \langle C, e^{-A}BU \rangle \geq 0$ для $C \in R_+^m$ и всех $U \in R_+^p$. Если $C = (0, \dots, 0, C_s)$ и $C_s \not\geq 0$, то полагаем $U_1 = 0, \dots, U_{s-1} = 0$. Имеем $X_1 = 0, \dots, X_{s-1} = 0$, и Δ зависит лишь от X_s . Воспользовавшись теоремой 3.1.3, получаем доказательство теоремы.

Допустим теперь, что теорема верна для $\{A_{ij}\}_{i,j=2}^s, \{B_{ij}\}_{i,j=2}^s$. Докажем ее для A и B . Если $\Delta \not\geq 0$ при $U_1 = 0$ и некоторых $U_i \in R_+^{p_i}$ ($i = 2, \dots, s$), то теорема доказана. Поэтому считаем, что $\Delta \geq 0$ при $U_1 = 0$. Будем предполагать, что $P_1(A) = A$. В противном случае строим управление для компонент $P_1(X)$ и далее пользуемся индуктивным предположением.

Воспользуемся обозначениями, введенными при доказательстве теоремы 3.1.4. Напомним, что T_i и θ_i — собственные векторы матрицы A^* , отвечающие собственным значениям $\wedge(A_i)$ и μ_i ($\text{Re } \mu_i < \lambda(A_i)$) матрицы A_i . По построению $T_{ki} = \theta_{ki} = 0 \in R^{m_k}$ при $k > i$ и $T_{ii} > 0 \in R^{m_i}$. Пусть номер $\omega_i \in \omega(1)$. Обозначим через $\bar{\theta}_{\omega_i} = (\theta_{k\omega_i})_{k \in \omega(1)}$ и $\bar{T}_{\omega_i} = (T_{k\omega_i})_{k \in \omega(1)}$ векторы, полученные из θ_{ω_i} и T_{ω_i} вычеркиванием компонент, не вошедших в путь $\omega(1)$. Для них верна следующая

Лемма 3.1.6.1. Вектор $\theta_{\omega_i \omega_i} \not\geq 0$. Если $T_{k\omega_i} \geq 0$ при $1 \leq j \leq k \leq \omega_i$ и $j \in \omega(1)$, то $T_{j\omega_i} \not\geq 0$ тогда и только тогда, когда $\lambda(A_j) > \lambda(A_{\omega_i})$. Если $\lambda(A_i) = \lambda(A_{\omega_i})$ при некотором $j < \omega_i$, то собственного вектора T_{ω_i} ($T_{\omega_i \omega_i} \neq 0$) не существует. Если $T_{k\omega_i} \geq 0$, то $T_{k\omega_i} > 0$ при $k \in \omega(1)$ и $k \leq \omega_i$.

Доказательство проводится так же, как и доказательство леммы 3.1.4.1. Используется тот факт, что

$$(\lambda(A_{\omega_i})I - A_j^*)T_{j\omega_i} = \dots + A_{kj}^*T_{k\omega_i} + \dots$$

и $A_{kj}^* \neq 0$ при $k, j \in \omega(1)$ и $k > j$. Воспользуемся обозначениями, введенными в теореме 3.1.4.

Лемма 3.1.6.2. Пусть для набора матриц $\{A_i\}_{i \in \omega(1)}$ выполнено условие $\lambda(A_1) = \lambda(A_i)$ при некотором $i \neq 1$ и $i \in \omega(1)$. Тогда имеет место одно из двух:

а) либо $L_1(t) = e^{-\lambda(A_1)t} \alpha_1 T_1$,

б) либо $L_k(t) = t^{r_k} e^{-\lambda(A_1)t} (\alpha_i P_k(T_i) + \epsilon(t))$ при некотором $k \geq 1$, где $\alpha_1, \alpha_i = \pm 1, r_k \geq 1, \epsilon(t) \rightarrow 0$ при $t \rightarrow \infty$ и собственный вектор $\alpha_i P_k(T_i)$ матрицы $P_k(A)^*$, отвечающий $\lambda(A_i) = \lambda(A_1)$, имеет компоненту со свойствами перемены знака.

Доказательство. Пусть $i_\omega \in \omega(1)$ — максимальный номер, для которого $\lambda(A_{i_\omega}) = \lambda(A_1)$. Обозначим через j_ω максимальный на множестве $\omega(1) \setminus i_\omega$ номер, для которого $\lambda(A_{j_\omega}) = \lambda(A_1)$. Пусть $j_0 = \max\{j_\omega \mid \omega(1) \in \Omega_1\}$. Ограничимся рассмотрением $P_{j_0}(Y(t))$. Далее, для простоты считаем, что $j_0 = 1$. Таким образом, на любом пути $\omega(1)$ лежит не более двух равных $\lambda(A_1)$ собственных значений матрицы A . Если в некотором векторе $L_k(t)$ ($k = 2, \dots, s$) имеются компоненты со свойствами перемены знака, то лемма доказана. Предположим, что $L_k(t) \geq 0$ с точностью до $o(t)$ при $k \geq 2$. Образует максимальный набор множеств $\{\Omega_i\}_{i \in I}$, удовлетворяющих условиям

$$\Omega_i \subset \Omega_1, \quad \Omega_i \neq \Omega_1, \quad \Omega_i \cap \Omega_j = \emptyset \quad (i \neq j \text{ и } i, j \in I)$$

и $\lambda(A_{k(i)}) = \lambda(A_1)$ при некотором $k(i) \in \Omega_i$. Случай а) имеет место тогда и только тогда, когда $I = \emptyset$. Пусть $I \neq \emptyset$. Тогда $L_i(t) = e^{-\lambda(A_1)t} P_i(T_{k(i)})$, где $P_i(T_{k(i)}) = (T_{lk(i)})_{l \in \Omega_i}$ — собственный вектор матрицы $P_i^*(A)$, отвечающий $\lambda(A_{k(i)}) = \lambda(A_1)$ ($i \in I$). По построению $A_{i1}^* \neq 0$, и в Ω_i существует путь, соединяющий i и $k(i)$. Следовательно, по лемме 3.1.6.1 $T_{lk(i)} > 0$ и, значит, $A_{i1}^* T_{lk(i)} \neq 0$. Обозначим через $Z(t)$ первую компоненту вектора $L_1(t)$. Она является решением уравнения

$$\dot{Z} = -A_1^* Z - e^{-\lambda(A_1)t} \sum_{i \in I} \sum_{i < l \leq k(i)} A_{il}^* T_{lk(i)}.$$

Повторяя рассуждения леммы 3.1.4.2, имеем, что $Z(t) = te^{-\lambda(A_1)t} \times X(-T_{11} + \epsilon(1)) < 0$, где $\epsilon(t) \rightarrow 0$ при $t \rightarrow \infty$. Лемма доказана.

Разложим Δ по экспонентам так же, как и в теореме 3.1.4:

$$\Delta = \sum_{i=1}^s \alpha_i e^{-\lambda(A_i)t} \sigma_i(t) (T_i^* BU + o(1)) + \beta_i e^{-\rho_i t} \delta_i(t) (a_i(t, U) + o(1))$$

и составим матрицу $D = [\alpha_i T_i + \beta_i \theta_i]_{i=1}^s$. Повторяя рассуждения предыдущей теоремы, получаем, что хотя бы один столбец этой матрицы обладает свойством перемены знака. Пусть максимальный номер строки, в которой существует хотя бы одна векторная компонента, обладающая свойствами перемены знака, равен единице. Как будет видно из дальнейших рассуждений, это предположение эквивалентно индуктивному предположению, что $\Delta \geq 0$ при $U_1 = 0$.

Фиксируем путь $\omega(1) \in \Omega_1$ и возьмем все векторы $\alpha_{\omega_i} T_{\omega_i} + \beta_{\omega_i} \theta_{\omega_i}$ с номерами $\omega_i \in \omega(1)$. Вычеркнем у них компоненты с номерами, не во-

шедшими в $\omega(1)$. Из полученного набора укороченных векторов образуем матрицу

$$D_\omega = [\alpha_1 \bar{T}_1 + \beta_1 \bar{\theta}_1, \dots, \alpha_q \omega_q \bar{T}_{\omega_q} + \beta_q \omega_q \bar{\theta}_{\omega_q}],$$

где $1, \dots, \omega_q \in \omega(1)$ и $\omega_q \leq s$. Для некоторых путей возможны матрицы D_ω со столбцами без перемены знака, и существуют пути $\nu(1) \subset N$, для которых матрицы D_ν имеют столбцы, обладающие свойствами перемены знака. По индуктивному предположению перемена знака возможна лишь в первой строке. Пусть

$$\omega_0 = 1 \text{ и } \omega_i = \min \{ \omega_j \in \omega(1) \mid \omega_j > \omega_{i-1}, |\alpha_{\omega_j}| + |\beta_{\omega_j}| > 0 \}.$$

Аналогично определяются номера ν_i . Предположим для определенности, что этих номеров p штук.

Лемма 3.1.6.3. Для любого пути $\omega(1) \in \Omega_1$ имеют место соотношения $\alpha_{\omega_1} > 0, \dots, \alpha_{\omega_p} > 0, \beta_{\omega_1} = 0, \dots, \beta_{\omega_p} = 0$ и $\lambda(A_{\omega_1}) < \dots < \lambda(A_{\omega_p})$. Если $\omega(1) \in N$, то $\alpha_1 > 0, \beta_1 = 0$ и $\lambda(A_1) < \lambda(A_{\omega_1})$. Если $\nu(1) \in N$, то возможны варианты:

а) если \bar{T}_{ν_1} не обладает свойством перемены знака, то первая компонента первого столбца матрицы D_ν обладает свойствами перемены знака и $\lambda(A_1) < \lambda(A_{\nu_1})$;

б) если \bar{T}_{ν_1} обладает свойством перемены знака, то $T_{1\nu_1} \in R_+^{m_1}$ и $\lambda(A_1) > \lambda(A_{\nu_1})$.

Доказательство этого факта дословно повторяет рассуждения леммы 3.1.4.3.

Перейдем к построению управления, при котором $\Delta(t, u) < 0$. Если при всех $\nu(1) \in N$ во втором столбце нет перемены знака, положим $U_i = 0$ при $i \neq 1$. На основании леммы 3.1.6.3 заключаем, что первое слагаемое в разложении Δ по экспонентам является максимальным при достаточно больших t . Управление $U_1 \in R_+^{p_1}$, при котором $\Delta < 0$, строится так же, как и в теореме 3.1.3.

Пусть $\bar{T}_{\nu_1} \neq 0$ для некоторых $\nu(1) \subset N_1 \subset N$. Если $\lambda(A_{\nu_1})$ для всех $\nu(1)$ попарно различны, выбираем ν_1^0 так, что $\lambda(A_{\nu_1^0}) = \min \{ \lambda(A_{\nu_1}) \mid \nu_1 \in \nu(1) \in N_1 \}$. В силу леммы 3.1.6.3 этому числу при больших t соответствует максимальное слагаемое в Δ . От укороченного вектора $\bar{T}_{\nu_1^0}$ вернемся к вектору $T_{\nu_1^0}$ матрицы A . Если $\beta_1 = 0$ или $\rho_1 > \lambda(A_{\nu_1^0})$, то максимальное слагаемое в Δ при достаточно больших t равно $\alpha_{\nu_1^0} e^{-\lambda(A_{\nu_1^0})t} T_{\nu_1^0}^* BU$. Если $\beta_1 \neq 0$ или $\rho_1 < \lambda(A_{\nu_1^0})$, то максимальным слагаемым в Δ при больших t является $\beta_1 e^{-\rho_1 t} \theta_1 BU$. Если $\beta_1 \neq 0$ и $\rho_1 = \lambda(A_{\nu_1^0})$, то максимальным слагаемым в Δ при больших t является $\beta_1 e^{-\rho_1 t} \theta_1 BU + \alpha_{\nu_1^0} e^{-\lambda(A_{\nu_1^0})t} T_{\nu_1^0}^* BU$. Во всех случаях, строя управление так же, как и в теореме 3.1.4, имеем $\Delta < 0$.

Рассмотрим случай равенства собственных значений: $\lambda(A_{\nu_1^0}) = \lambda(A_{\mu_1^0})$. Пусть для определенности $\nu_1^0 < \mu_1^0$. Применяв лемму 3.1.6.2 к матрице $P_{\nu_1^0}(A)$ и воспользовавшись индуктивным предположением, заключаем, что возможен лишь случай а). Следовательно, в разложении $e^{-A^*t} C$ по экспонентам собственному значению $\lambda(A_{\nu_1^0})$ отвечает собственный вектор $\alpha_{\nu_1^0} T_{\nu_1^0}$. В этом же разложении по лемме 3.1.6.1 собственному значению

$\lambda(A_{\mu_1^0})$ отвечает собственный вектор $\alpha_{\mu_1^0} T_{\mu_1^0}$. Сумма $Z = \alpha_{\nu_1^0} T_{\nu_1^0} + \alpha_{\mu_1^0} T_{\mu_1^0}$ также является собственным вектором матрицы A . Первая компонента Z_1 этого слагаемого удовлетворяет уравнению

$$(\lambda(A_{\nu_1^0})I - A_1^*)Z_1 = + \dots + A_{\nu_1^0}^* Z_{\nu_1^0} + \dots + A_{\mu_1^0}^* Z_{\mu_1^0} + \dots$$

и $\lambda(A_{\nu_1^0}) < \lambda(A_1)$ по лемме 3.1.6.3. Воспользовавшись схемой рассуждений леммы 3.1.6.1, заключаем, что $Z_1 \in R_+^{m_1}$. Таким образом, приходим к предыдущей ситуации.

Таким образом, замыкание множества достижимости содержится в R_+^m . Следовательно, имеется включение внутренностей этих множеств [25, с. 53]. Пусть $(X(0))_k \neq 0$. Тогда в силу связности $P_k(X(t)) > 0$ при $t > 0$ и $U = 0$. Воспользовавшись теми же соображениями, что и в теореме 3.1.4, завершаем доказательство достаточности.

Результаты по управлению экологическими системами при неограниченных внешних воздействиях имеют принципиальное теоретическое значение. На практике управления всегда конечны. В задачах оптимального управления при линейности правой части модели по управлению они принимают граничные значения. В связи с этим важное практическое значение представляют вопросы о структуре множества достижимости при конечных управляющих воздействиях и об управляемости с помощью релейных воздействий. Приводимые ниже результаты говорят о том, что этот вопрос в принципе уже решен.

Следствие. Пусть выполнены условия теоремы 3.1.6 и имеются ограничения на величину управления $\|U\| < 1$. Тогда начало координат достижимо для всех точек, достаточно близких к нему и принадлежащих R_+^m . Для того чтобы множество достижимости начала координат содержало положительный конус R_+^m , необходимо и достаточно, чтобы $\lambda(A_i) \leq 0$ ($i = 1, \dots, s$).

Доказательство. Докажем первую часть теоремы. Пусть непрерывное управление $U^i(t): R_+^1 \rightarrow R_-^m$ за время T_i переводит решение из точки $E_i = (0, \dots, 1, \dots, 0) \in R_+^m$ в начало координат. Обозначим

$T = \max \{T_i | i = 1, \dots, m\}$ и $M = \{\text{Sup} \|U^i(t)\| | 0 \leq t \leq T_i, i = 1, \dots, m\}$. Пусть начальное состояние $X(0)$ принадлежит окрестности $0 \leq x_i(0) \leq$

$\leq 1/M$ ($i = 1, \dots, m$). Разложим $X(0)$ по базису $X(0) = \sum_{i=1}^m \alpha_i E_i$, где $\alpha_i \leq$

$\leq 1/M$. Доопределим функцию $U^i(t)$ так, что при $t \in [0, T_i]$ она совпадает с исходной, а при $t \in [T_i, T]$ принимает нулевое значение. Положим $U(t) =$

$= \sum_{i=1}^m \alpha_i U^i(t)$. Ввиду линейности системы имеем $X(T) = 0$.

Перейдем к изучению глобальной достижимости.

Необходимость. Пусть $\lambda(A_i) > 0$ при некотором $i = 1, \dots, s$ и Y_i — отвечающий ему положительный собственный вектор A_i^* . Положим $X_k = 0$ ($k = 0, 1, \dots, i-1$). При достаточно больших $X_i(0)$ имеем $\langle Y_i, X(t) \rangle = \lambda(A_i) \langle Y_i, X(t) \rangle + Y_i B U > 0$. Поэтому $\langle Y_i, X(t) \rangle \rightarrow \infty$ при $t \rightarrow \infty$.

Достаточность. Рассмотрим ограничение исходной системы на компоненту связности Ω_k ($k = 1, \dots, s$). В дальнейшем для упрощения

записи будем считать, что $P_k(A) = A$. Так же, как и при доказательстве теоремы 3.1.3 – 3.1.6, перейдем к системе с обратным временем. Множество концов траекторий, начинающихся в начале координат, выпукло и задается выражением

$$K = \left\{ \int_0^t e^{-A\tau} BU(t-\tau) d\tau \mid U \in R_+^p, \|U\| \leq 1 \right\}.$$

Предположим, что $\bar{K} \not\subset R_+^m$. По теореме Хана – Банаха найдутся такой функционал $C \in R^m$ и такое число α , что $\sigma(t) = \int_0^t C^* e^{-A\tau} BU(t-\tau) d\tau < \alpha$ при всех $t > 0$ и всех допустимых управлениях. Покажем, что это утверждение ложно. Для этого достаточно показать, что при любом $C \in R_+^m$ существует управление, при котором $\sigma(t) \rightarrow -\infty$, если $t \rightarrow \infty$. Действительно, если $C \in R_+^m$ и имеет компоненты разных знаков, то $-C \in R_+^m$. Если $C \in R_+^m$, то $-C \in R_+^m$ и $C \in R_-^m$, так как $0 \in K$.

Перейдем к доказательству этого факта. Пусть $C \in R_+^m$. При доказательстве теорем 3.1.3, 3.1.4 и 3.1.6 было установлено, что имеет место $C^* e^{-A\tau} BU = e^{-\mu\tau} \tau^k (\varphi(t, U) + \epsilon(t))$, где $-\mu > \lambda(A_i) \geq 0$ ($i = 1, \dots, s$), и возможны варианты:

1) $a(t, U) = \langle T, BU \rangle$ и принимает как положительные, так и отрицательные значения при $U \in R_+^p$. В этом случае считаем, что $a(t, U_0) \leq -\delta < 0$ при $t > 0$ и некотором $U = U_0 \in R_+^p$ и $\|U_0\| \leq 1$.

2) $a(t, U)$ – почти периодическая функция, которая через неравные конечные промежутки времени J_k ($k = 1, 2, \dots$) на интервалах длины α принимает значения, меньшие $-\delta(U) \leq 0$ [26, 31]. Полагая $U_0(t) = 0$ при $t \in J_k$ и $U(t) = U_0 \neq 0$, $U_0 \in R_+^p$ и $\|U_0\| \leq 1$ при $t \in J_k + [0, \alpha]$, можно считать, что $\int_0^t a(\tau, U_0(\tau)) d\tau \rightarrow -\infty$ при $t \rightarrow \infty$ и некотором $U = U_0(t)$. Зада-

дим управление по следующему правилу: $U(t-\tau) = 0$ при $\tau \geq t_0$, где t и t_0 выбираются из условия $|\epsilon(\tau)| < \delta/2$ при $\tau \geq t - t_0$; $U(t-\tau) = U_0(t-\tau)$ при $\tau \leq t_0$. Поскольку $e^{\mu\tau} \tau^k \geq 1$, имеем, что

$$\int_0^t C^* e^{-A\tau} BU(t-\tau) d\tau = \int_{t-t_0}^t (a(t, U(t-\tau)) d\tau + \epsilon(\tau)) \rightarrow \infty$$

при $t \rightarrow \infty$. Теорема доказана.

Полученные результаты допускают перенесение на нелинейные модели экологических систем. Рассмотрим нелинейный процесс управления (3.1.1).

Предположим, что начало координат перенесено в равновесную точку, т.е. $F(0, 0) = 0$. Сопоставим нелинейному процессу (3.1.1) линеаризованную систему (3.1.2), где $A = D_X F(0, 0)$ и $B = D_U F(0, 0)$.

Пусть матрицы A и B имеют вид (3.1.6). Используем $X_i \in R_+^{m_i}$, $U_i \in R_+^{p_i}$ и $F(X, U) = (F_i(X, U))_{i=1}^s$, где $F_i \in R^{m_i}$ ($i = 1, \dots, s$). Будем называть нелинейный оператор разложимым, если матрица линейного приближения разложима и из условия $X_i = 0$ и $U_i = 0$ при всех $i \in \Omega_l$ следует, что $F_k(X, U) = 0$ при $k \in \Omega_l$ ($l = 1, \dots, s$).

Обозначим через Q куб со сторонами единичной длины, все вершины которого, кроме одной, принадлежат граням конуса R_+^p , и через ∂Q – мно-

жество его вершин. Пусть $K(Q)$ ($K(\partial Q)$) — замкнутый выпуклый конус, порожденный множеством

$$\{F(0, U) | U \in Q\} \quad (\{F(0, U) | U \in \partial Q\}).$$

Заметим, что $K(Q) = K(\partial Q)$ при линейной зависимости функции $F(X, U)$ от U .

Теорема 3.1.7. Пусть линейное приближение процесса (3.1.1) удовлетворяет условиям теоремы 3.1.3 или 3.1.6. Тогда при $U(t) \in Q$ начало координат достижимо для всех внутренних точек R_+^m , лежащих достаточно близко от него. Если одновременно с разложимостью матрицы линейного приближения имеет место разложимость нелинейного оператора, то аналогичный результат верен для точек замкнутого конуса R_+^m , принадлежащих малой окрестности нуля. В том случае, когда $K(Q) = K(\partial Q)$, предыдущие результаты имеют место при релейных управлениях

$$U_i(t) \in \{0, 1\} \quad (i = 1, \dots, p).$$

Доказательство. Пусть $E_i = (0, \dots, 1, \dots, 0)$ ($i = 1, \dots, m$) — базисные векторы конуса R_+^m . Поскольку R_+^m является множеством достижимости системы (3.1.2), для каждого вектора E_i найдется время t_i ($i = 1, \dots, m$), за которое точка E_i решением системы (3.1.2) переводится в начало координат. Пусть $T = \max \{t_i | 1 \leq i \leq m\}$. Тогда множество достижимости системы (3.1.2) за время T совпадает с R_+^m . Рассмотрим

теперь замкнутый конус $K_T = \{ \sum_{i=1}^m a_i e^{AT} E_i | a_i \geq 0 \ (i = 1, \dots, m) \} \subset$

R_+^m . В этот конус за время T попадает всякое решение системы (3.1.2), начинающееся в R_+^m при $U(t) = 0$. Если A — неразложимая матрица, то $K_T \subset R_+^m$. Пусть A — разложимая матрица и $X(0) \in R_+^m$ — ненулевое начальное условие. Допустим, что $X_j(0) \neq 0$ при $j \in J \subset \{1, \dots, s\}$ и $X_i(0) = 0$ при $i \notin J$. Тогда $X_k(T)$ принадлежат замкнутому конусу $K_T^k \subset R_+^{mk}$ при $k \in J_1 = \bigcup_{j \in J} \Omega_j$ и $X_i(t) = 0$ при $i \notin J_1$ и $t \geq 0$. Пусть $K_T = \prod_{k \in J_1} K_T^k$ и $R_+^n = \prod_{k \in J_1} R_+^{mk}$. Положим $U_i(t) = 0$ при $i \notin J_1$. Тогда во все моменты

времени движение принадлежит R_+^n и $K_T \subset \overset{\circ}{R}_+^n$. Для нелинейного процесса (3.1.1) в достаточно малой окрестности нуля все решения при $U = 0$ за время T также попадают в некоторый замкнутый конус N , принадлежащий внутренности $R_+^m (R_+^n)$.

Не умаляя общности, будем считать, что $X(0) \in N \subset \overset{\circ}{R}_+^n$. Исследуем достижимость начала координат из точек конуса K_T за время T , что эквивалентно исследованию достижимости начала координат за время $2T$ из точек конуса $R_+^m (R_+^n)$.

По определению [14] решения системы $\dot{X} = AX + W(t)$ с начальным условием $X(0) = 0$, где $W(t)$ — кусочно-постоянные функции со значениями в $K(Q)$ (в $K(\partial Q)$) при $t \in [-T, 0]$, образуют касательный конус возмущений [14] системы (3.1.1) за этот промежуток времени. В силу замкнутости $K(Q)$ имеем $\{BU | U \in R_-^p\} \subset K(Q)$. Поэтому множество достижимости системы (3.1.2) содержится в касательном конусе возмущений системы (3.1.1). По теореме 3.1.3 или 3.1.6 R_+^n принадлежит касательному конусу возмущений за время $[-T, 0]$. По лемме об отображении кону-

сов [14, с. 277] для любого вектора $H \in N$ найдется конус $K_H \subset \mathbb{R}_+^n$, содержащий внутри себя H и такой, что его часть, отстоящая на расстояние, не большее, чем h_H , принадлежит множеству достижимости для нелинейного процесса.

Пусть D — единичная сфера. Пересечение $K_H \cap D$ образует окрестность точки $H/\|H\|$ на этой сфере, и множества $\{K_H \cap D | H \in N\}$ покрывают $N \cap D$. Пользуясь компактностью $N \cap D$, найдем векторы H_1, \dots, H_r такие, что $N \cap D \subset \bigcup_{1 \leq i \leq r} K_{H_i} \cap D$. Тогда множество $S = \{X \in N | \|X\| \leq$

$\leq \min h_{H_i} (i = 1, \dots, r)\}$ образует окрестность нуля в конусе N . Это множество принадлежит множеству достижимости начала координат при $t \in \in [-T, 0]$ для процесса (3.1.1). Следовательно, для всякой начальной точки из S найдется допустимое управление, которое переводит решение уравнения (3.1.1) в начало координат. Всякое решение уравнения (3.1.1) при $U = 0, X(0) \geq 0$ и $t \geq 0$ попадает в S . Теорема доказана.

Для того чтобы исследовать проблему достижимости нелинейных систем "в большом", приведем следующий результат о глобальной устойчивости нелимитированных экологических систем.

Теорема 3.1.8. Пусть $F(0, 0) = 0$ и существуют такие неотрицательные векторы $X_0 \in \mathbb{R}_+^m$ и $U_0 \in Q$, при которых $F(X_0, U_0) \leq 0$. Если матрица $D_X F(X, U_0)$ гурвицева при всех $X \in [0, X_0]$, то начало координат асимптотически устойчиво на этом порядковом интервале при $U = U_0$.

Доказательство. Оператор сдвига вдоль траекторий монотонен [21, с. 245], поэтому для любой точки $X_* \leq X_0$ сохраняется неравенство $X(t, X_*) \leq X(t, X_0)$. Поскольку $\dot{F} = D_X F \times F$ и матрица $D_X F(X, U_0)$ имеет неотрицательные недиагональные элементы, $F(X(t, X_0), U) \leq 0$ при $t \geq 0$. Следовательно, $X(t, X_0)$ монотонно убывает к точке $\text{Inf} \{X(t, X_0) | t \geq 0\}$, которая является стационарной. Вследствие гурвицевости якобиана его главные миноры нечетного порядка отрицательны, а четного — положительны [7]. Следовательно, уравнение $F(X, U_*) = 0$ на порядковом интервале $[0, X_*]$ имеет единственное решение [17, с. 473], которое совпадает с началом координат. Локальная устойчивость следует из гурвицевости матрицы $D_X F(0, U_*)$. Теорема доказана.

Очевидным следствием приведенных выше результатов является

Теорема 3.1.9. Пусть выполнены условия теорем 3.1.7 и 3.1.8. Тогда начало координат достижимо для всех точек порядкового интервала $[0, X_*]$ при $U(t) \in Q$ и, если $K(\partial Q) = K(Q)$, при $U(t) \in \partial Q$.

Нарушение условий теоремы 3.1.8 может привести к неустойчивости экологической системы. Проиллюстрируем этот факт результатом о неограниченности решений.

Теорема 3.1.10. Пусть $F(0, 0) \geq 0$ и в каждой точке $X \in [0, X_0]$ якобиева матрица $D_X F(X, 0)$ имеет на главной диагонали $s(X)$ неразложимых подматриц $A_i(X)$ ($i = 1, \dots, s(X)$) (см. формулу (3.1.6)) и $\lambda(A_i(X)) > 0$ ($i = 1, \dots, s(X)$). Тогда любое ненулевое решение, начинающееся на порядковом интервале $0 \leq X \leq X_0$, за конечное время покидает его при $U = 0$.

Доказательство. Из монотонности оператора сдвига вдоль траекторий имеем, что решение, начинающееся в точке $X(0)$, минорируется решением, начинающимся в точке $X_* \leq X(0)$.

Лемма 3.1.10.1. Для любого решения $X(t, X(0))$, начинающегося в точке $X(0) \in R_+^m \setminus 0$, существует неубывающая во времени миноранта $X(t)$, которая не является стационарной точкой.

Доказательство. Пусть вектор X_* принадлежит достаточно малой окрестности начала координат и удовлетворяет условию $0 \not\leq X_* \leq X(0)$. Обозначим через X_k (Z_k) собственный вектор матрицы $A_k(0)$ ($A_k^*(0)$) ($k = 1, \dots, s(0)$) с положительными компонентами, отвечающий $\lambda(A_k(0))$, и через $\bar{X}_k = (0, \dots, X_k, \dots, 0)$ ($\bar{Z}_k = (0, \dots, Z_k, \dots, 0)$) — вложение вектора из подпространства R^{mk} в пространство R^m . Так как $X_* \neq 0$, то существует номер k , при котором $\langle \bar{Z}_k, X_* \rangle > 0$. Ввиду неразложимости матрицы $A_k(0)$ ограничение нелинейного оператора $F(X, 0)$ на R^{mk} неразложимо. Следовательно (см. предложение 2.1.1), проекция вектора $X(t, X_*)$ на это подпространство имеет строго положительные компоненты при $t > 0$. Выберем t_0 и ϵ из условия $X(t_0, X_*) \geq \epsilon \bar{X}_k$. Очевидно, что $F(\epsilon \bar{X}_k, 0) \not\geq 0$. Обозначим через $X(t)$ решение, начинающееся в ϵX_k . Ввиду квазимонотонности $F(X, 0)$ лемма доказана.

Итак, $X(t) \rightarrow \text{Sup} \{X(t) \mid t \geq 0\} = C$. Обозначим через Y_i собственный вектор матрицы $A_i^*(C)$, отвечающий $\lambda(A_i(C))$, и через \bar{Y}_i — его вложение в пространство R^m . Пусть $C \leq X_0$. Имеем оценку

$$\begin{aligned} \frac{d}{dt} \langle \bar{Y}_i, X(t) - C \rangle &\leq \\ &\leq \lambda(A_i(C)) \langle \bar{Y}_i, X(t) - C \rangle + o(\|X(t) - C\|) < 0 \end{aligned}$$

при некотором $i = 1, \dots, s(C)$, что противоречит монотонной сходимости к точке $X(t)$. Теорема доказана.

3.1.2. Управление в положительном конусе. До сих пор изучалась проблема попадания в начало координат из точек положительного конуса. В нелинейном случае такая постановка была эквивалентна задаче достижения положения равновесия из начальных векторов, все компоненты которых либо превосходят, либо меньше компонент вектора конечного состояния. Никакие ограничения на состояние системы в процессе движения не накладывались. Рассмотрим более сложную ситуацию, когда движение системы во все моменты времени не должно быть меньше или больше своего терминального значения. Подобные ограничения возникают, например, в задаче уничтожения вредной популяции с помощью внешних воздействий. При движении к нулю фазовые переменные модели в силу биологического смысла могут принимать лишь неотрицательные значения. Такой подход приводит к понятию управляемости в конусе.

Определение. Будем говорить, что для точек множества S имеет место управляемость в конусе K , если для любой точки этого множества существует кусочно-непрерывное во времени управление, которое переводит процесс в вершину конуса за конечное время, и движение в течение этого времени не покидает K . Если $S = K$, то будем говорить, что процесс глобально управляем.

Данное ограничение приводит к существенному сужению класса допустимых управлений. В общем случае оказывается, что количество независимых регуляторов не должно быть меньше, чем число степеней свободы системы. В качестве иллюстрации рассмотрим пример.

Пусть динамика управляемой экологической системы задается моделью

$$\dot{X} = F(X) + \sum_{i=1}^m a_i E_i U_i,$$

где $F(x) = (f_i(x))_{i=1}^m$ — квазимоноotonно возрастающая функция, $E_i = (0, \dots, 1, \dots, 0)^*$, $U_i \geq 0$ ($U_i \leq 0$) и $a_i \geq 0$ ($i = 1, \dots, m$). Обозначим через X_0 точку равновесия системы, т.е. $F(X_0) = 0$. Тогда имеет место следующий факт.

Для того чтобы рассматриваемая система была управляема в конусе $X \leq X_0$ ($X \geq X_0$), необходимо и достаточно, чтобы $a_i > 0$ ($i = 1, \dots, m$). Докажем необходимость. Не умаляя общности, будем считать, что $F(0) = 0$, $U_i \leq 0$ ($i = 1, \dots, m$) и $X(0) \in R_+^m$. Пусть $a_i = 0$. Тогда $\dot{x}_i(t) = f_i(X(t)) \geq f_i(0, \dots, x_i(t), \dots, 0)$ при $X(t) \geq 0$. Так как $f_i(0) = 0$, то в силу единственности решения задачи Коши имеем, что $x_i(t) > 0$ во все моменты времени. Необходимость доказана. Достаточность следует из приводимой ниже теоремы 3.1.11.

В дальнейших рассуждениях, не умаляя общности, ограничимся рассмотрением управляемости в конусе R_+^m при неположительных управлениях $U \in R_-^m$. Изучим вначале задачу с линейным управлением:

$$\dot{X} = F(X, t) + B(t)U, \quad (3.1.7)$$

где $X \in R_+^m$, $F(X, t) = (f_i(X, t))_{i=1}^m$ — квазимоноotonно возрастающий по X оператор, $F(0, t) \geq 0$, $B(t)$ — $m \times m$ -матричная функция, $U \in R_-^m$ и $t \geq 0$.

Теорема 3.1.11. Если матрица $B^{-1}(t)$ имеет неотрицательные элементы при всех $t \geq 0$, то процесс (3.1.7) глобально управляем.

Доказательство. Пусть $Z \in R^m$. Определим вектор $P(Z) = (p_i)_{i=1}^m$ по правилу $p_i = z_i$ при $z_i > 0$ и $p_i = 0$ при $z_i \leq 0$. Предположим, что в начальный момент времени $x_i(0) = 0$ при $i \in I$ и $x_j(0) > 0$ при $j \in J$, где $I \cup J = \{1, \dots, m\}$. Построим вектор $E(X) = (e_i(X))_{i=1}^m$ по правилу $e_i = 0$ при $i \in I$ и $e_j = 1$ при $j \in J$. Поскольку $f_i(X, t) \geq 0$ при $x_i = 0$, неположительное управление зададим в виде

$$U(X, t) = -B^{-1}(t)(P(F(X, t)) + E(t)) \in R_-^m.$$

Тогда $\dot{x}_j(t) \leq -1$ при $j \in J$ и $\dot{x}_i(t) = 0$ при $i \in I$ и $t \geq 0$. Следовательно, найдется конечный момент времени, при котором хотя бы одна компонента вектора $X(t)$ с номером из множества J станет равной нулю. Продолжая рассуждения по индукции, завершаем доказательство теоремы.

Следствие. Если матрица $-B(t)$ квазиположительна и является гурвицевой при всех $t \geq 0$, то процесс (3.1.7) глобально управляем.

Доказательство очевидно [7].

В этом случае, когда управление ограничено по величине, получаем результат, аналогичный теореме 3.1.7.

Теорема 3.1.12. Пусть правые части системы (3.1.7) не зависят от времени $F(0) = 0$ и выполнено условие теоремы 3.1.11. Если $\|U\| \leq 1$, то управляемость в конусе имеет место для всех точек малой окрестности начала координат.

Пусть функция $F(X)$ линейна. Для того чтобы процесс был глобально управляем, необходимо и достаточно, чтобы собственное значение опера-

тора $F(X)$ с максимальной вещественной частью было неположительно.

Доказательство. Легко видеть, что управление, построенное при доказательстве теоремы 3.1.11, ограничено на каждом ограниченном множестве. Отсюда следует локальная управляемость. Пусть оператор $F(X)$ имеет вид (3.1.6) и $\lambda(A_i) \leq 0$ ($i = 1, \dots, s$). Зададим управление в виде $U(X) = -\epsilon B^{-1}X$, где ϵ берется настолько малым, что в начальный момент времени $\|\epsilon B^{-1}X\| < 1$. Тогда $X(t) \rightarrow 0$ при $t \rightarrow \infty$. Достаточность доказана.

Необходимость докажем, рассуждая от противного. Пусть $\lambda(A_i) > 0$ при некотором $1 \leq i \leq s$. Предположим, что в начальный момент времени состояние процесса $X(0) = (X_i(0))_{i=1}^s$ таково: $X_k(0) = 0$ при $k < i$ и $X_i(0) \neq 0$. Обозначим через Y_i собственный вектор матрицы A_i^* . Тогда имеем, что

$$\frac{d}{dt} \langle Y_i, \dot{X}_i \rangle = \lambda(A_i) \langle Y_i, X_i(0) \rangle + Y_i^* B U(t) > 0$$

при достаточно большой $\|X_i(0)\|$. Отсюда $\langle Y_i, X_i(t) \rangle \rightarrow \infty$ при $t \rightarrow \infty$. Теорема доказана.

Обобщим полученные результаты на нелинейные системы специального вида

$$\dot{X} = F(X, t) - B(U), \quad (3.1.8)$$

где $X \in R_+^m$, $F(X, t)$ — квазимонотонная по X функция, $F(0, t) \geq 0$, $U \in R_+^m$ и $B(U)$ — квазимонотонная функция, обладающая, кроме того, свойством положительной однородности по U . Предположим, что $B(U)$ недифференцируема в точке $U = 0$, однако удовлетворяет условию Липшица. Как показано в предложении 2.1.4, дифференцируемость в точке $U = 0$ приводит к линейности функции. Нелинейный оператор $B(U)$ имеет собственный вектор с неотрицательными компонентами, которому отвечает максимальное собственное значение λ [17].

Теорема 3.1.13. Если $\lambda < 0$, то процесс (3.1.8) глобально управляем.

Доказательство. Пусть в начальный момент времени $x_i(0) = 0$ при $i \in I$ и $x_j(0) > 0$ при $j \in J$, где $I \cup J = \{1, \dots, m\}$. Воспользовавшись обозначениями, введенными при доказательстве теоремы 3.1.11, зададим управление в виде

$$U(X, t) = B^{-1}[P(F(X, t) + E(X))] \in R_+^m.$$

Функция $B^{-1}(Z)$ определена единственным образом при всех $Z \in \overset{\circ}{R}_+^m$ [17, с. 223] и отображает всякое ограниченное множество в ограниченное [17, с. 223]. Воспользовавшись технической леммой 1 (приложение 1), заключаем, что управление $U(X, t)$ определено и непрерывно. Далее рассуждаем так же, как и в теореме 3.1.11. Теорема доказана.

Прежде чем перейти к изучению нелинейных процессов общего вида, введем некоторые дополнительные понятия.

Рассмотрим задачу синтеза управления, построенного по принципу обратной связи, когда управляющие воздействия являются функциями фазовых координат системы. Преимущество замкнутого контура по сравнению с разомкнутой цепью управления состоит в автоматической коррекции непредвиденных внешних возмущений и неточностей самой модели.

Принцип обратной связи лежит в основе многих регуляторных механизмов биологических систем.

Задачу синтеза управления будем решать, покрывая фазовое пространство конечным или счетным числом множеств $\{Q_i\}_{i \geq 0}$ и определяя на каждом из них управления в виде непрерывных функций координат или в виде таких разрывных функций, при которых на Q_i ($i \geq 0$) возникает скользящий режим [23, 24]. Управления будут заданы так, что множество Q_i отображается решениями дифференциальных уравнений в множество $\bigcup_{j > i} Q_j$ при $t \geq 0$. В классе определенных таким образом управлений зада-

дим иерархию, считая управление, определенное на множестве с большим номером, более приоритетным. В точках пересечения областей Q_j реализуется более приоритетное управление. Такой вид регулирования мы будем называть иерархической системой с переменной структурой [23]. Решение дифференциального уравнения с разрывной правой частью мы будем понимать в смысле А.Ф. Филиппова [24]. Очевидно, имеет место правосторонняя единственность решения задачи Коши. Примером систем управления переменной структуры, в которых реализуются скользящие режимы, являются системы непрерывного культивирования микроорганизмов, основанные на турбидостатном принципе (см. ниже § 3.4).

Перейдем к решению задачи глобальной управляемости в конусе нелинейной квазимонотонной системы. Предположим, что правая часть системы (3.1.1) непрерывно дифференцируема и ее производные удовлетворяют условию Липшица при всех $X \in R_+^m$ и $U \in R_-^m$. Наряду с квази-монотонным возрастанием по первому аргументу предположим квази-монотонное убывание по второму, т.е. $\partial f_i / \partial x_j \geq 0$ и $\partial f_i / \partial u_j \leq 0$ при всех $i \neq j$ ($i, j = 1, \dots, m$), $x \in R_+^m$, $u \in R_-^m$.

Теорема 3.1.14. Пусть $F(0, 0) \geq 0$ и $F(X_0, U_0) \leq 0$ при некоторых $X_0 \in R_+^m$ и $U_0 \in R_-^m$. Если матрица $-D_U F(X, U)$ гурвицева при всех $X \in [0, X_0]$ и $U \in [U_0, 0]$, то процесс (3.1.1) управляем в конусе R_+^m из некоторой окрестности точки X_0 . Управление принадлежит R_-^m и принимает значения из некоторой окрестности множества $[U_0, 0]$.

Предположим, что функция $F(X, U)$ дифференцируема на бесконечности [9] по U и матрица $-D_U F(X, U)$ гурвицева при всех $X \in R_+^m$ и $U \in R_-^m$. Тогда процесс глобально управляем в конусе R_+^m .

В обоих случаях управление можно задать функцией координат, реализующей иерархическую систему переменной структуры. Возникающие скользящие движения определены однозначно.

Доказательство. Пусть $E = (1, \dots, 1)^*$, $\alpha = \max(1, \|D_X F(X_0, U_0)\|)$, $\beta = \max(1, \|D_U F(X_0, U_0)\|^{-1})$ и $l = \max(1, l_0)$, где l_0 — константа Липшица дифференциала функции F в d -окрестности точки X_0, U_0 и $\epsilon = \min\{(5l\alpha\beta^2 m^2)^{-1}, d\}$.

Лемма 3.1.14.1. Пусть $F(X_0, U_0) \leq 0$. Существует такое постоянное управление $U \in R_-^m$, при котором $F(X, U) < 0$ для всех X из множества $\|X - X_0\| \leq \epsilon$.

Доказательство. Линеаризуем функцию $F(X, U)$ в окрестности точки X_0, U_0 . Имеем

$$F(X, U) = F(X_0, U_0) + D_X F \delta X + D_U F \delta U + \varphi(X, U),$$

где δX и δU — малые вариации переменных X и U , $\|\varphi(X, U)\| \leq \frac{l_0}{2} (\|X\|^2 + \|U\|^2)$ [11]. Положим

$$\delta U = -\frac{2\gamma}{m\beta} [D_U F]^{-1} E.$$

Так как $\lambda(-D_U F) < 0$, то $\delta U \leq 0$ и, следовательно, $U_0 + \delta U \in R_-^m$. Кроме того, $\|\delta U\| \leq 2\gamma$. Пусть $\|\delta X\| \leq \frac{\gamma}{\alpha\beta m}$. Положим $\gamma = (5l_0\beta m)^{-1}$. Так как

$$\|\varphi(X, U)\| \leq \frac{5l_0\gamma^2}{2}, \text{ имеем}$$

$$F(X, U) = F(X_0, U_0) + D_X F \delta X - \frac{2\gamma}{\beta m} E + \varphi(X, U) \leq -\frac{\gamma}{\beta m} E + \varphi(X, U) < 0.$$

Лемма доказана.

Лемма 3.1.14.2. Пусть $F(X_0, U_0) \leq 0$. Существует кусочно-постоянное управление, при котором все компоненты решения $X(t)$, начинающегося в точке X_0 , строго монотонно убывают и за конечное время решение $X(t)$ достигает границы R_+^m .

Доказательство. Воспользовавшись леммой 3.1.14.1, будем считать, что $F(X_0, U_0) < 0$. Так как $\dot{F} = D_X F \cdot F$ и $F < 0$ при $t = 0$, то решение $X(t)$, полученное при $U = U_0$, монотонно убывает. Пусть $X_1 = \text{Inf} \{X(t) | t \geq 0\} \in R_+^m$. Очевидно, что $F(X_1, U_0) = 0$ и $X_1 < X_0$. По лемме 3.1.14.1 найдется такое постоянное управление U_1 , при котором $F(X_1, U_1) < 0$ в малой окрестности X_1 . Таким образом, строится последовательность $X_0 > X_1 > \dots > X_n > \dots$. Обозначим через U_n управление в ϵ_n -окрестности точки X_n , построенное в лемме 3.1.14.1. Зададим управление правилом $U(X) = U_n$, где $n = \min \{k | \|X_k - X\| \leq \epsilon_k\}$. По построению $F(X, U(X)) < 0$ почти во все моменты времени. Будем рассуждать от противного. Пусть $\{X_i\}_{i=1}^\infty \subset R_+^m$. Ввиду монотонности последовательности стационарных точек имеем, что $X_i \rightarrow X_0 \in R_+^m$ при $i \rightarrow \infty$. Так как функция F непрерывно дифференцируема, то $\epsilon_n \geq \epsilon > 0$ при достаточно больших n . Следовательно, $\|X_{n+k} - X_n\| \leq \epsilon_n$ при достаточно больших n и k . Отсюда $U_{n+k} = U_n$ и $F(X_{n+k}, U_{n+k}) < 0$, что противоречит стационарности точки X_{n+k} . Поэтому при конечном числе переключений $k < \infty$ управления последовательность $\{X_i\}_{i=1}^k$ покинет R_+^m . Поскольку $X(t) \rightarrow X_k$, лемма доказана.

Введем вспомогательные обозначения. Пусть $Y \in R_+^m$. Сформулируем векторы $\overset{I}{Y} = (y_i)_{i \in I}$ и $\overset{J}{Y} = (y_j)_{j \in J}$, где $I \cup J = \{1, \dots, m\}$.

Лемма 3.1.14.3. Предположим, что $F(X_0, U_0) < 0$, $(X_0)_i = 0$ при $i \in I$ и $(X_0)_j > 0$ при $j \in J$. Тогда существует гладкое управление $U = U(X)$ со значениями в $[U_0, 0]$, при котором $f_i(X, U(X)) = 0$ при $i \in I$ и $f_j(X, U(X)) \leq f_j(X, U_0) < 0$ при $j \in J$. Это управление реализуется в скользящем режиме системой переменной структуры вида

$$U(X) = (U_k(X))_{k=1}^m = \begin{cases} U_k(X) = (U_0)_k & \text{при } x_k > 0, \\ U_k(X) = 0 & \text{при } x_k = 0 \end{cases} \quad (k = 1, \dots, m).$$

Если $F(0, 0) > 0$, скользящий режим устойчив.

Доказательство. Рассмотрим дифференциальное уравнение $\frac{d}{d\tau} U = -F(X_0, U, U)$ с начальным условием $U = U_0$. Ввиду квазимонотонности функции $-F$ по U и условия $F < 0$ при $\tau = 0$ имеем, что $U(\tau)$ монотонно возрастает к вектору U_* . Покажем, что $U_* \leq 0$. Допустим противное. Пусть $U_i(\tau) = 0$ и $U_k(\tau) \leq 0$ ($i, k \in I$) в некоторый момент времени τ . Так как $(X_0)_i = 0$ и $U_i(\tau) = 0$, из квазимонотонности по X имеем $-f_i(X_0, U(\tau)) \leq -f_i(0, 0) \leq 0$.

Отсюда видно, что $U_* \leq 0$ и $U_* < 0$, если $F(0, 0) > 0$. Введем вектор U_* с компонентами $(U_*)_j = (U_0)_j$ при $j \in J$ и $(U_*)_i = (U_*)_i$ при $i \in I$. Тогда $\bar{F}(X_0, U_*) \leq \bar{F}(X_0, U_0) < 0$, так как $U_* \geq U_0$ и $U_* = U_0$. Кроме того, $\bar{F}(X_0, U_*) = 0$. Построим непрерывно дифференцируемую функцию $U = G(X)$, при которой все компоненты вектора $\bar{F}(X, G(X))$ строго отрицательны и $\bar{F}(X, G(X)) = 0$. Пусть $F(X, U_0) < 0$ при $X \in S_{X_0}$, где S_{X_0} — окрестность точки X_0 , построенная в лемме 3.1.14.1. Поскольку матрица $-D_U \bar{F}$ гурвицева, этим же свойством обладает любая ее главная подматрица [7]. Следовательно, матрица $D_U \bar{F}$ обратима. Так как $\bar{F}(X_0, U_*) = 0$, то по теореме о неявной функции в некоторой окрестности $W_{X_0} \subset S_{X_0}$ точки X_0 определена гладкая функция $U = G(X)$ ($G(X) = U_0$), при которой $\bar{F}(X, G(X)) = 0$. Ввиду того, что матрица $-D_U \bar{F}$ гурвицева, матрица $D_U \bar{F}$ имеет положительные главные миноры [7]. Следовательно [17], не существует векторов $U \neq G(X)$, при которых $\bar{F}(X, U) = 0$. Отсюда каждое значение функции $G(X)$ является предельной точкой монотонно возрастающего решения уравнения $\frac{d}{d\tau} U = -F(X, U)$. Как показано ранее, $U \leq 0$ в каждой точке, где $F(X, U) \leq 0$. Следовательно, $G(X) \leq 0$. Поскольку $\bar{F}(X, U) < 0$ при $U = U_0$ и $\bar{F}(X, U) \geq 0$ при $U = 0$, данная функция реализуется в скользящем режиме. Этот режим устойчив при $F(0, 0) > 0$, так как $\bar{F}(X, 0) > 0$.

Покажем теперь, что $G(X) \geq U_0$. Фиксируем некоторую точку $X \in W_{X_0}$ и введем в рассмотрение множество $A = \bar{F}(X, R_-^m)$. Ввиду того, что матрица $D_U \bar{F}$ имеет положительные главные миноры, обратное к $\bar{F}(X, U)$ по U отображение $H: A \rightarrow [U_0, 0]$ существует и является монотонно возрастающим [17, с. 475]. Отсюда и из условий $\bar{F}(X, G(X)) = 0$ и $\bar{F}(X, U_0) < 0$

закключаем, что $\dot{U} = \dot{G}(X) > \dot{U}_0$. Так как $\dot{G}(X) = \dot{U}_0$, имеем $0 \geq G(X) \geq U_0$. Следовательно, функция $G(X)$ определяет допустимые значения управления. Ввиду условия $\partial f_i / \partial u_i \leq 0$ ($i = 1, \dots, m$) имеем $\dot{F}(X, G(X)) \leq \dot{F}(X, U_0) < 0$. Приведенная процедура позволяет продолжить функцию $G(X)$ на всю область S_{X_0} . В силу глобальной единственности эта функция будет гладкой. Лемма доказана.

Доказательство теоремы проводится по индукции. Вначале на основании леммы 3.1.14.2 строится управление, при котором решение достигает границы R_+^m . Далее, на основании леммы 3.1.14.3 строится управление, при котором компоненты решения, попавшие на границу R_+^m , принадлежат ей во все оставшиеся моменты времени, а остальные компоненты монотонно убывают. Как только одна из убывающих компонент становится равной нулю, вновь применяется лемма 3.1.14.3 и т.д.

Если функция $F(X, U)$ дифференцируема на бесконечности по U , то $F(X, U) = D_U(X, \infty)U + \varphi(X, U)$, где $\varphi(X, U) / \|U\| \rightarrow 0$ при $\|U\| \rightarrow \infty$. Ввиду гурвицевости матрицы $-D_U(X, \infty)$ найдется $U \in R_-^m$, при котором $F(X, U) < 0$ в каждой точке $X \in R_+^m$. Теорема доказана.

Замечание. Для нелинейных по управлению систем в настоящее время не существует общих методов, позволяющих однозначным образом строить уравнения скольжения. Такие методы разработаны для некоторых частных случаев. Существует проблема расширения условий, при которых возможно однозначное определение скользящих движений [23]. Теорема 3.1.14 гарантирует существование и единственность скользящих движений для нелинейных по управлению систем. Доказательство этого факта опирается на квазимонотонность правой части. В частном случае, когда $f_i(X, U) = f_i(X, U_i)$ ($i = 1, \dots, m$), управляемость процесса (3.1.1) исследовалась ранее [28]. При таком предположении доказательство единственности скользящих движений проводилось известными методами.

Итогом данного параграфа является следующий результат. Задачи оптимального управления нелимитированными экологическими системами при произвольных, хотя и близких к стационарному начальных состояниях и фиксированном знаке управления не имеют решения. При такой постановке необходимо ограничиться рассмотрением начальных условий, при которых все фазовые переменные либо превосходят, либо меньше своих конечных или терминальных значений. В этом случае задачи становятся разрешимыми. Идея доказательства сводится к следующему.

Если исходные фазовые переменные заменить новыми, согласованными с инвариантами системы, то одна из новых переменных будет либо превосходить, либо окажется меньшей терминального значения, а остальные могут отклоняться от него в обоих направлениях. Условие существования хотя бы одного управления, переводящего систему в заданное конечное состояние, сводится к однонаправленному (соответствующего знака) воздействию на одну переменную и воздействиям произвольного знака на другие переменные системы. Знакопеременность управляющих воздействий на новые фазовые переменные может быть либо результатом внутренних колебательных свойств самой экологической системы, приводящих к

изменению знака отклонения от состояния равновесия, либо результатом богатства исходного набора управлений.

При дополнительных ограничениях на знак отклонения от конечного состояния в процессе движения задача оптимального управления имеет решение, только если число независимых управлений равняется количеству степеней свободы.

§ 3.2. Управление лимитированными экологическими системами

Лимитированные экологические системы способны демонстрировать значительно более сложное поведение, чем свободно воспроизводящиеся популяции и сообщества. В то же время наличие лимитирующих факторов делает такие системы отзывчивыми на управляющие воздействия. Для лимитированных систем некоторые задачи управления доступны более полному решению. Сюда прежде всего следует отнести стабилизируемость объекта при произвольных отклонениях от равновесного состояния. Изложению этого вопроса, а также проблеме управляемости моделей (2.1.2) и (2.1.3) посвящен данный параграф.

3.2.1. Управление вблизи равновесного состояния. Рассмотрим обобщение моделей динамики численности лимитированных популяций, в котором управление входит в правую часть в неявном виде. Динамика чистого потребления задается системой

$$\dot{X} = F(X, s, W) - vX, \quad s = \varphi(X), \quad (3.2.1)$$

где $v \geq 0$ – постоянная скорость изъятия особей, $W \in R_+^p$ – управляющие параметры.

Динамика численности популяции при потреблении и внешних источниках задается системой

$$\dot{X} = F(X, s, W) - vX + Bu, \quad s = \varphi(X), \quad (3.2.2)$$

где $v \geq 0$ – постоянная скорость потребления, $u \geq 0$ – постоянная скорость притока и B – структура этого притока. Предполагается, что функции $F(X, s, W): R_+^m \times R_+^1 \times R_+^p \rightarrow R^m$ и $\varphi(X): R_+^m \rightarrow R_+^1$ непрерывно дифференцируемы при $X \neq 0$ и при $W = 0$ удовлетворяют тем же условиям, что и правые части уравнений (2.1.2) и (2.1.3). Кроме того, будем считать, что постоянные величины v и u выбраны так, что уравнение (3.2.1) или (3.2.2) имеет нетривиальное стационарное решение $X_0, s_0 = \varphi(X_0)$ (см. теоремы 2.1.2 и 2.1.3). Линеаризуем уравнение (3.2.1) или (3.2.2) в окрестности стационарной точки $X = X_0$ и $W = 0$. Имеем

$$\dot{Y} = DY + RW, \quad (3.2.3)$$

где квадратная матрица $D = D_X F(X_0, s_0, 0) + D_s F(X_0, s_0, 0) \partial \varphi / \partial X(X_0)$, $R = D_W F(X_0, s_0, 0)$ – прямоугольная матрица размерности $m \times p$.

В силу предложения 2.2.5 собственные значения матрицы, которые не принадлежат левой полуплоскости, имеют ненулевую мнимую часть. Поэтому неустойчивая компонента решения линеаризованного уравнения обладает свойством "осцилляционности", которое необходимо для положительной управляемости линейной системы [31]. Линейное уравнение (3.2.3) определено в m -мерном евклидовом пространстве E , которое может быть представлено в виде прямой суммы $E = E^1 \oplus E^2$ подпрост-

ранств размерности m_1 и m_2 , соответствующих неустойчивой (включая мнимые значения) и устойчивой частям спектра матрицы D . С помощью невырожденной линейной замены переменных уравнение (3.2.3) приводится к виду

$$\dot{Y}_1 = D_1 Y_1 + R_1 W, \quad (3.2.4)$$

$$\dot{Y}_2 = D_2 Y_2 + R_2 W, \quad (3.2.5)$$

где $D_i(\cdot) + R_i W$ — ограничение линейного оператора $D(\cdot) + RW$ на инвариантное подпространство E^i ($i = 1, 2$).

Как уже отмечалось, матрица D гурвицева, а все собственные значения матрицы D_1 имеют ненулевую мнимую часть. Будем считать в дальнейшем, что ранг матрицы $[R_1, D_1 R_1, \dots, D_1^{m_1-1} R_1]$ равняется m_1 . Тогда множеством достижимости [26] уравнения (3.2.4) при $U \in R_+^p$ является все пространство E^1 , т.е. любая точка E^1 может быть переведена в начало координат за конечное время с помощью кусочно-непрерывного управления $W(t) = (w_1(t), \dots, w_p(t)) \geq 0$. Вид стабилизирующего управления для линейной задачи очевиден: произвольную точку из E следует с помощью неотрицательного управления перевести в подпространство E^2 , а затем занулить управление. В силу асимптотической устойчивости нулевого решения уравнения (3.2.5) при $W = 0$ траектория процесса управления будет стремиться к нулю. Аналогично строится управление для нелинейной системы, где вместо устойчивого подпространства E^2 используется устойчивое многообразие, касательное к E^2 .

Прежде чем сформулировать окончательный результат, введем вспомогательные обозначения. Обозначим через Q единичный куб в R_+^p с вершиной, лежащей в начале координат, и вершинами ∂Q , образованными векторами с нулевыми и единичными компонентами. Пусть $K(R) = \{RW \mid W \in R_+^p\}$ и $K(\partial Q)$ — замкнутый выпуклый конус, порожденный множеством $\{F(X_0, s_0, W) \mid W \in \partial Q\}$. Пусть $X(t, Z)$ — решение системы (3.2.1) или (3.2.2) при начальных условиях $X(0, Z) = Z$.

Теорема 3.2.1. Для любой точки z некоторой окрестности стационарного решения X_0 существует кусочно-непрерывное управление $W(t) \in Q$ при $t \geq 0$, при котором $X(t, Z) \rightarrow X_0$, когда $t \rightarrow \infty$. Если $(K(R) \cap E_1) \subset \subset (K(\partial Q) \cap E_1)$, то можно использовать релейное управление $W(t): R_+^p \rightarrow \partial Q$.

Доказательство. Обозначим через P_{E^1} проекцию пространства E на E^1 и через $K(Q)$ — замкнутый выпуклый конус, порожденный множеством $\{F(X_0, s_0, W) \mid W \in Q\}$. При $W = 0$ система (3.2.1) или (3.2.2) имеет инвариантное устойчивое многообразие M [27]. Это многообразие проходит через стационарную точку X_0 и касается пространства E^2 . Обозначим через $Ac(T, Q)$ ($Ac(T, \partial Q)$) множество концов решений процесса (3.2.1) или (3.2.2), начинающихся при $t = 0$ на многообразии M и определенных на промежутке времени $[-T, 0]$ при управлении $W: [-T, 0] \rightarrow K(Q)$ ($W: [-T, 0] \rightarrow K(\partial Q)$), и через $Ac_L(T, Q)$ ($Ac_L(T, \partial Q)$) — множество концов решений системы

$$\dot{Y}_1 = D_1 Y + W_1(t), \quad (3.2.6)$$

$$\dot{Y}_2 = D_2 Y + W_2(t), \quad (3.2.7)$$

где D_1 и D_2 те же, что и в системах (3.2.4) и (3.2.5), начинающихся из точки $Y_1(0) = 0$ и $Y_2(0) = 0$ и определенных при $t \in [-T, 0]$ и неотрицательных кусочно-

непрерывных функциях

$$(W_1(t), W_2(t)) \in K(Q) \quad ((W_1(t), W_2(t)) \in K(\partial Q)).$$

Лемма 3.2.1.1. Существует $T > 0$, при котором $E^1 \subset \text{Pr}_1 \text{Ac}_L(T, Q)$. Если $K(R) \cap E^1 \subset K(\partial Q) \cap E^1$, то $E^1 \subset \text{Pr}_1 \text{Ac}_L(T, \partial Q)$.

Доказательство. Поскольку уравнение (3.2.4) положительно управляемо [26], при некотором $T > 0$ среди всех кусочно-непрерывных функций $W_1(t): [-T, 0] \rightarrow K(R) \cap E^1$ найдутся такие, что решения уравнения (3.2.6), начинающиеся в начале координат, порождают E^1 [26, лемма 1].

Очевидно, что $K(R) \subset K(Q)$ и, следовательно, первая часть леммы доказана. Если $\text{Pr}_1 K(R) \subset \text{Pr}_1 K(\partial Q)$, то множества $\text{Pr}_1 \text{Ac}_L(T, Q)$ и $\text{Pr}_1 \text{Ac}_L(T, \partial Q)$ совпадают. Лемма доказана.

Лемма 3.2.1.2. Если $E^1 \subset \text{Pr}_1 \text{Ac}_L(T, Q)$, то множество $\text{Ac}(T, Q)$ содержит окрестность начала координат; если $E^1 \subset \text{Pr}_1 \text{Ac}_L(T, \partial Q)$, то множество $\text{Ac}(T, \partial Q)$ содержит окрестность начала координат.

Доказательство. Обозначим через K касательный конус возмущений [14], образованный векторами элементарных возмущений [14] при кусочно-постоянных неотрицательных вариациях управления в окрестности $W = 0$ и вариациях начальных данных на многообразии M . Тогда [4, 14] $K = K^W + K^M$, где K^W – конус возмущений, полученных в результате вариаций управления W , а K^M – конус возмущений, полученных в результате вариаций начальных данных. Так как M – инвариантное многообразие при $W = 0$ и E^2 – касательное к нему пространство, то $K^M = E^2$. По определению K^W совпадает с множеством $\text{Ac}_L(T, Q)$ ($\text{Ac}_L(T, \partial Q)$) и, следовательно, $E^1 \cap \text{Pr}_1 K^W$.

Конус K является выпуклой оболочкой K^W и K^M и поэтому равен E . Лемма об отображении конуса [4] в форме [14, с. 277] применима к настоящему случаю, поскольку в доказательстве нигде не используется природа вариаций. Воспользовавшись этой леммой, можно заключить, что, каков бы ни был луч, исходящий из точки X_0 , всегда найдется достаточно узкий телесный усеченный конус, содержащий внутри себя этот луч и целиком лежащий внутри $\text{Ac}(T, Q)$ ($\text{Ac}(T, \partial Q)$). Продолжив этот усеченный конус до пересечения с единичной сферой, мы получим некоторую окрестность на ней, лежащую внутри продолженного таким образом конуса. Варьируя направление луча, образуем покрытие единичной сферы и пользуясь ее компактностью, выберем конечное подпокрытие. Усеченные конусы, порождающие это подпокрытие, полностью покрывают некоторую окрестность X_0 в R^m и содержатся в $\text{Ac}(T, Q)$ ($\text{Ac}(T, \partial Q)$). Лемма доказана.

Воспользовавшись леммой 3.2.1.2, построим управление, которое переводит точку из малой окрестности начала координат на устойчивое многообразие M . Всякое решение, принадлежащее этому многообразию, стремится к началу координат при $t \rightarrow \infty$ и $W = 0$. Теорема доказана.

Следует отметить, что полученный результат без всяких изменений переносится на системы с многофакторным лимитированием. Рассмотрим модель

$$\dot{X} = F(X, S, W) - vX + Bu, \quad (3.2.8)$$

где $X \in R_+^m$ – численный состав экосистемы, $S \in R_+^n$ – набор лимитирующих факторов, $v \in R_+^1$ – постоянная скорость потребления, $u \in R_+^1$ – постоянная скорость притока, $W \in R_+^p$ – набор управляющих параметров. Предполагается, что функции $F: R_+^m \times R_+^n \times R_+^p \rightarrow R_+^m$ и $\Phi: R_+^m \rightarrow R_+^n$ непрерывно дифференцируемы при $X \neq 0$ и при $W = 0$ удовлетворяют условиям, сформулированным в п. 2.1.6. Кроме того, будем считать, что урав-

нение (3.2.8) имеет нетривиальное стационарное решение $X_0, S_0 = \Phi(X_0)$ (см теоремы 2.1.20 и 2.1.21). Пусть $\bar{K}(\partial Q)$ – замкнутый выпуклый конус, порожденный множеством

$$\{ F(X_0, S_0, W) \mid W \in \partial Q \}.$$

Теорема 3.2.2. Для любой точки Z некоторой окрестности стационарного решения X_0 существует кусочно-непрерывное управление $W(t): R_+^1 \rightarrow Q$, при котором $X(t, Z) \rightarrow X_0$, когда $t \rightarrow \infty$. Если $(K(R) \cap E_1) \subset (\bar{K}(\partial Q) \cap E_1)$, то можно использовать релейное управление $W(t): R_+^1 \rightarrow \partial Q$.

Доказательство. Убедимся в том, что все собственные значения матрицы линейного приближения, которые не лежат в левой полуплоскости, имеют ненулевую мнимую часть. Рассуждая так же, как и в п. 2.2.1, легко получить, что собственное значение матрицы $D_X F(X_0, S_0)$ с максимальной вещественной частью неположительно. Ненулевая матрица $D_S F(X_0, S_0)$ имеет вид

$$[Y_1, \dots, Y_n] [X_1, \dots, X_n]^*, \text{ где } Y_i \in R_+^m \text{ и } X_i \in R_+^m \text{ (} i = 1, \dots, n \text{)}.$$

Воспользовавшись технической леммой 3 (приложение 1) и формулой окаймленного определителя [13], получаем, что все коэффициенты характеристического полинома неотрицательны и свободный член положителен. Следовательно, вещественные собственные значения лежат в левой полуплоскости. Далее, дословно повторяя рассуждения теоремы 3.2.1, завершаем доказательство теоремы.

Прокомментируем приведенный результат. В силу предложения 2.2.5 часть спектра матрицы линейного приближения лежит в левой полуплоскости, и поэтому размерность пространства, для которого приходится строить стабилизирующее управление, меньше размерности исходной задачи. В этом состоит преимущество модели динамики численности популяции как объекта управления. Недостатком ее является невозможность решать задачу оптимального быстрогодействия, поскольку время достижения стационарной точки бесконечно. Осмыслена постановка задачи на оптимизацию функционала, заданного на бесконечном интервале времени, например,

$$\int_0^{\infty} \|X(t) - X_0\| dt \rightarrow \min, \text{ или функционала, зависящего лишь от од-$$

$$\text{ного управления, например, } \int_0^t \|W(t)\| dt \rightarrow \min.$$

3.2.2. Стабилизируемость равновесного состояния. В этом пункте рассмотрен вопрос синтеза в моделях (2.1.2) и (2.1.3) такой обратной связи, при которой фиксированное стационарное решение асимптотически устойчиво. Помимо нелинейности изучаемых процессов, их специфика состоит в том, что управляющие воздействия могут принимать значения лишь одного знака. Поэтому классическая теория линейных стабилизаторов неприменима. Тем не менее задача синтеза стабилизирующего управления в рассматриваемых системах полностью разрешима, если расширить класс допустимых управлений. Проблема управляемости процессов (2.1.2) и (2.1.3) также допускает удовлетворительное решение.

Обозначим через ST множество стационарных решений уравнения (2.1.2), существующих по теореме 2.1.2 при $v \geq 0$, и ST_B – множество

стационарных решений уравнения (2.1.3), существующих по теореме 2.1.3 при $v \geq 0$ и $u \geq 0$. В качестве управлений в данном параграфе будут выступать скорость изъятия особей $v \geq 0$ и скорость притока извне $u \geq 0$.

Определение. Будем говорить, что подмножество множества стационарных решений стабилизируемо в классе кусочно-непрерывных (измеримых) функций, если для любой точки Y из этого множества существует неотрицательное управление, являющееся кусочно-непрерывной (измеримой) функцией координат, при котором точка Y асимптотически устойчива в $R_+^m \setminus 0$.

Стационарную точку, в которой асимптотическая скорость воспроизводства равняется нулю, т.е. $\lambda_{\sigma_0} = 0$, будем в дальнейшем обозначать через X_0 . Вначале рассмотрим процесс управления (2.1.2).

Теорема 3.2.3. Множество $ST \setminus X_0$ глобально стабилизируемо процессом управления (2.1.2) в классе кусочно-непрерывных функций. Управление может осуществляться иерархической системой переменной структуры, порождающей скользящие режимы.

Доказательство. Разобьем доказательство на ряд лемм.

Лемма 3.2.3.1. Для любого ограниченного множества на R_+^m существует такое постоянное значение управления W , при котором начало координат асимптотически устойчиво.

Доказательство. Так как функция $F(X, s)$ непрерывна в точке s_∞ , то в силу однородности по первому аргументу существует такая константа N , при которой $f_i(X, s) \leq N\rho(X)$ ($i = 1, \dots, m$). Отсюда имеем $F(X, s) \leq N\rho(X)I$ и, следовательно, $\dot{X} = F(X, s) - vX \leq (N\rho(X) - v)X$. При $v > N\rho(X)$ начало координат асимптотически устойчиво на ограниченном множестве для системы $\dot{Y} = (N\rho(X) - v)Y$. Применение леммы сравнения [21] завершает доказательство.

Лемма 3.2.3.2. Пусть $F(X)$ – квазимонотонный положительно однородный оператор, H – его собственный вектор $| \rho(H) = 1 |$ и λ – собственное значение. Пусть $v: R_+^m \rightarrow R_+^1$ – функция, при которой уравнение $\dot{X} = F(X) - v(X)X$ имеет единственное решение задачи Коши на некотором инвариантном множестве из R_+^m . Тогда решение этого уравнения с ненулевым начальным условием из инвариантного множества допускает представление $X(t) = \rho(X(t)) \overset{t}{(H + \epsilon(t))}$, где $\epsilon(t) \rightarrow 0$ при $t \rightarrow \infty$.

Доказательство. Обозначим $\tau(t) = \int_0^t v(X(t)) dt$ и умножим исходное уравнение на $e^{\tau(t)}$. Воспользовавшись однородностью, получаем $\dot{Y} = F(Y)$, где $Y = e^{\tau} X$. В силу свойства (2.1.5) имеем $X(t) = e^{\lambda t - \tau(t)} (H + \delta(t))$, где $\delta(t) \rightarrow 0$ при $t \rightarrow \infty$. Дальнейшие рассуждения очевидны.

Лемма 3.2.3.3. Пусть выполнены условия леммы 3.2.3.2 и функции $v(X)$ и $F(X)$ обладают одной константой Липшица для всех точек каждого ограниченного множества. Пусть скалярное уравнение $\dot{\xi} = \lambda\xi - v(\xi H) \xi$ имеет положительное стационарное решение ξ_* , которое асимптотически устойчиво на множестве $(0, \infty)$. Тогда $X(t) \rightarrow \xi_* H$ при любом $X(0) \geq 0$ из инвариантного множества и $t \rightarrow \infty$, где $X(t)$ – решение уравнения из леммы 3.2.3.2.

Доказательство. В силу леммы 3.2.3.2 функция $\rho(t) = \rho(X(t))$ удовлетворяет дифференциальному неравенству

$$(\lambda - v(\rho(t))H - |\epsilon_1(t)|) \rho(t) \leq \frac{d}{dt} \rho(t) \leq (\lambda - v(\rho(t))H + |\epsilon_1(t)|) \rho(t),$$

где $\epsilon_1(t) \rightarrow 0$ при $t \rightarrow \infty$. Воспользовавшись леммой сравнения и убыванием $\epsilon_1(t)$, заключаем, что $\rho(X(t)) \rightarrow \xi_*$ при $t \rightarrow \infty$. Вновь воспользовавшись леммой 3.2.3.2, завершаем доказательство.

Лемма 3.2.3.4. Пусть X_v – решение системы (2.1.8) при $0 < v = \lambda$. Обозначим $\omega(X) = \langle \partial\varphi/\partial X, F(X, \varphi(X)) \rangle$ и зададим в уравнении (2.1.2) управление в виде

$$v(X, \alpha) = \begin{cases} 0 & \text{при } \varphi(X) < \alpha \text{ или } \omega(X) \leq 0, \\ \omega(X)/\alpha & \text{при } \varphi(X) = \alpha \text{ и } \omega(X) > 0, \\ M & \text{при } \varphi(X) > \alpha. \end{cases}$$

Тогда при достаточно большом M стационарное решение X_v асимптотически устойчиво в малом. Область притяжения точки $X_v = X_{\lambda_\alpha}$ содержит шар, радиус которого непрерывно зависит от параметра α .

Доказательство. Управление $v(X, \alpha)$ реализуется в скользящем режиме [23] при управлении переменной структуры

$$v = \begin{cases} 0 & \text{при } \varphi(X) < \alpha, \\ M & \text{при } \varphi(X) > \alpha. \end{cases}$$

Действительно, равенство $v(X, \alpha) = \omega(X)/\alpha$ получается по методу эквивалентного управления [23] из условия $\partial\varphi/\partial t|_{\varphi(X)=\alpha} = 0$ и свойства однородности $\langle \partial\varphi/\partial X, X \rangle = \varphi(X)$. Условие устойчивости скользящего режима в окрестности точки X_v имеет вид $\omega(X) > 0$, так как при $\varphi(X) < \alpha$ и $\omega(X) > 0$ функция $\varphi(X)$ монотонно возрастает, а при $\varphi(X) > \alpha$ монотонно убывает, если M достаточно велико. Поэтому в дальнейшем неравенство $\omega(X) > 0$ при $\varphi(X) = \alpha$ будем называть условием скольжения. В окрестности точки X_v ($v = \lambda_\alpha > 0$) условие скольжения выполняется, так как $\omega(X_v) = \lambda_\alpha \varphi(X_v) > 0$ и рассматриваемые функции непрерывны. Покажем, что матрица линейного приближения уравнения скольжения

$$\dot{X} = F(X, \alpha) - (\omega(X)/\alpha) X,$$

ограниченного на инвариантное множество $\varphi(X) = \alpha$, несмотря на разрывность функции $\omega(X)$, непрерывно зависит от параметра α и является гурвицевой.

Пусть $X = X_v + Y$. Тогда имеем

$$\dot{Y} = [D_X F(X, \alpha) - \lambda_v I] Y + [\omega(X) - \omega(X_v)] X_v + o(\|Y\|),$$

$$\langle \partial\varphi/\partial X(X_v), Y \rangle + o(\|Y\|) = 0.$$

Обозначим через Q линейное преобразование, которое приводит матрицу $D_X F(X, \alpha)$ к жордановой нормальной форме. В силу однородности функции $F(X, \alpha)$ по первому аргументу X_v является собственным вектором матрицы $D_X F(X, \alpha)$, которому отвечает собственное значение $\lambda_\alpha = v$ нелинейного оператора $F(\cdot, \alpha)$. Не умаляя общности, будем считать, что первый столбец матрицы Q совпадает с вектором X_v . Сделаем в линеаризованной системе замену переменных $Y = Q \begin{pmatrix} \eta \\ Z \end{pmatrix}$, где η – скаляр и Z – столбец размерности $m - 1$. Имеем следующую матрицу линейного приближения:

$$Q^{-1} (D_X F(X, \alpha) - \lambda_\alpha I) Q = \left[\begin{array}{c|cccc} 0 & & & & \\ \cdot & & & & \\ \cdot & & & & \\ \cdot & & & & \\ 0 & & & A & \end{array} \right].$$

Первая строка полученной матрицы и подматрица $A_{(m-1) \times (m-1)}$ отличны от нуля. Так как $\langle \partial\varphi/\partial X, X_v \rangle \neq 0$, то в новых координатах инвариантное многообразие можно задать функцией $\eta = \psi(Z)$ ($\psi(0) = 0$). Отсюда получаем, что движение по многообразию определяется системой $\dot{\eta} = \psi(Z)$, $\dot{Z} = AZ + o(\|Z\|)$, где матрица A имеет размерность $(m - 1) \times (m - 1)$ и ее спектр получается из спектра матрицы $D_X F(X, \alpha)$ сдвигом на λ_α влево и удалением нулевого собственного значения.

Как показано в лемме 2.1.6.1 из теоремы 2.1.6, число λ_α является правой границей спектра матрицы $D_X F(X, \alpha)|_{X=X_\alpha}$. Поэтому матрица A гурвицева, что и требовалось доказать.

Завершим доказательство леммы. Пусть $X(0)$ достаточно близко к X_v . Тогда за конечное время $X(t)$ попадет на поверхность скольжения. Поскольку X_v — устойчивая точка уравнения скольжения, решение не выйдет из окрестности X_v , на которой выполнено условие скольжения. Следовательно, начиная с некоторого момента времени, решение удовлетворяет уравнению скольжения.

Пользуясь устойчивостью по линейному приближению, определим функцию Ляпунова в виде квадратичной формы. Поверхность уровня этой квадратичной формы задает положительно инвариантную для уравнения скольжения окрестность точки X_v . Ввиду непрерывной зависимости этой квадратичной формы от параметра α уравнение скольжения имеет непрерывно зависящую от α область притяжения. Непрерывная зависимость от α области притяжения стационарной точки X_v ($v = \lambda_\alpha$) системы (2.1.2) при $\varphi(X) \neq \alpha$ следует из непрерывной зависимости функции $F(X, s)$ от параметра s . Лемма доказана.

Лемма 3.2.3.5. Существует малое $\delta > 0$, при котором множество $\{X_v | \varphi(X_v) \leq \delta\}$ глобально стабилизируемо.

Доказательство. В силу леммы 3.3.1.1 будем считать, что в начальный момент времени $\varphi(X(0)) \leq \delta$. Зададим управление в виде $v(X, \alpha)$ из леммы 3.2.3.4 и покажем, что при достаточно малом δ решение непременно перейдет в скользкий режим. По технической лемме 2 (приложение 1) при всех $s \leq \delta$ имеет место представление $F(X, s) = F_s^+(X) - NI$, где $F_s^+ = (f_{st}^+)^m_{i=1}$ — монотонный оператор. Пусть H — положительный собственный вектор оператора $F(X, 0)$. В силу непрерывности по s и однородности по X имеем $F(X, s) \leq F(X, 0) + \epsilon \rho(X)H$, где $\epsilon \rightarrow 0$ при $s \rightarrow 0$. С другой стороны, $(1 - \epsilon)f_{0t}^+(X) \leq f_{st}^+(X)$ при $\rho(X) = 1$, $i = 1, \dots, m$ и $\epsilon \rightarrow 0$ при $s \rightarrow 0$. Отсюда заключаем, что

$$\Phi^1(X) = (1 - \epsilon)F_0^+(X) - NI \leq F(X, s) \leq F_0(X) + \epsilon \rho(X)H - NI = \Phi^2(X)$$

при $s \leq \delta$.

Построим теперь миноранту решения $X(t)$ уравнения (2.1.2) при управлении $v(X, \alpha)$. Пусть $\omega^1(X) = \langle \partial \varphi / \partial X, \Phi^2(X) \rangle$. Из определения $\Phi^2(X)$ имеем $\omega^1(X) > \omega(X)$. Пусть $v^1(X) = 0$ при $\omega^1(X) \leq 0$ и $v^1(X) = \omega^1(X)/\alpha$ при $\omega^1(X) > 0$ ($\alpha \leq \delta$). Тогда $v^1(X) \geq v(X)$. Отсюда заключаем, что $\Phi^1(X(t)) - v^1(X(t))X(t) \leq D_t X(t)$, где D_t — любая производная Дини [21]. Уравнение $\dot{Y} = \Phi^1(Y) - v^1(Y)Y$ удовлетворяет условию леммы 3.2.3.3. Воспользовавшись формулой $\langle \partial \varphi / \partial X, \dot{X} \rangle = \varphi(X)$, имеем при $Y = \xi H$:

$$\dot{\xi} = ((1 - \epsilon)\lambda_0 - \epsilon N)\xi - (\lambda_0 + \epsilon) \frac{\varphi(H)}{\alpha} \xi^2$$

и, следовательно, $\xi(t) \rightarrow \alpha \frac{(1 - \epsilon)\lambda_0 - \epsilon N}{(\lambda_0 + \epsilon)\varphi(H)} H$ при $\xi(0) > 0$ и $t \rightarrow 0$. По лемме 3.2.3.3 получаем $Y(t) \rightarrow (1 - \beta(\epsilon))(\alpha/\varphi(H))H$ при $Y(0) \neq 0$ и $t \rightarrow \infty$, где $\beta(\epsilon) \rightarrow 0$ при $\epsilon \rightarrow 0$. Напомним, что $\epsilon \rightarrow 0$ при $\delta \rightarrow 0$. По лемме сравнения [21] для достаточно больших t можно считать, что

$$(1 - 2\beta(\epsilon))(\alpha/\varphi(H))H \leq X(t) \text{ при } X(0) \neq 0.$$

Мажоранта решения уравнения (2.1.2) строится аналогично. Пусть

$$v^2(X) = \begin{cases} 0 & \text{при } \varphi(X) \leq \alpha, \\ (\varphi - \alpha)/\alpha^2 & \text{при } \varphi(X) > \alpha. \end{cases}$$

В силу определения $\Phi^2(X)$ и инвариантности множества $\varphi(X) \leq \alpha$ имеем $D_t X(t) \leq \Phi^2(X) - v^2(X)X$. Рассуждая так же, как и в предыдущем случае, получаем, что решение уравнения $\dot{Y} = \Phi^2(Y) - v^2(Y)Y$ стремится к $((\alpha + \alpha^2(\lambda_0 + \epsilon))/\varphi(H))H$ при $t \rightarrow \infty$. Отсюда $X(t) \leq (1 + 2\beta_1(\alpha))(\alpha/\varphi(H))H$ при достаточно больших t , где $\beta_1(\alpha) \rightarrow 0$ при $\alpha \rightarrow 0$ ($\alpha \leq \delta$).

Итак, мы получили, что $X(t) \in (\alpha/\varphi(H))[1 - \beta_-(\delta), 1 + \beta_+(\delta)]H$ при $t \geq T$, где T — достаточно большое число и $\beta_\pm(\delta) \rightarrow 0$ при $\delta \rightarrow 0$. Отсюда $\langle \partial \varphi / \partial X, F(X(t), \varphi(X(t))) \rangle >$

> 0 при $t \geq T$. Если $\varphi(X(t)) < \alpha$ при $t \geq T$, то $\dot{X} = F(X(t))$, $\varphi(X(t)) > 0$ и функция $\varphi(X(t))$ монотонно возрастает до тех пор, пока не выполнится условие $\varphi(X(t)) = \alpha$. Использование лемм 3.2.3.3 и 3.2.3.4 завершает доказательство леммы.

Лемма 3.2.3.6. Пусть $X_\alpha, X_\beta \in ST$ и $\alpha > \beta$. Существует управление, при котором все решения из некоторой окрестности точки X_α стремятся к X_β при $t \rightarrow \infty$.

Доказательство. По формуле (2.1.7) и теореме 2.1.2 существуют единственные $s_\alpha < s_\beta$ такие, что $\lambda_{s_\alpha} = \alpha$ и $\lambda_{s_\beta} = \beta$. Пусть $s_1 = s_\beta$ и d — максимальный диаметр шара, принадлежащего области притяжения точки X_β . Если $\|X_\alpha - X_\beta\| < d_1$, положим $s_2 = s_\alpha$. В противном случае найдется точка $s_\alpha < s_2 < s_1$, в которой $\|X_{s_1} - X_{s_2}\| = d_2/2$. При $i \geq 2$ величины d_i обозначают максимальный диаметр шара, содержащегося в области притяжения точки $X_{\lambda_{s_i}}$. Продолжая построение по индукции, получаем последовательность $s_\beta = s_1 > s_2 > \dots > s_i > \dots$, для которой $\|X_{\lambda_{s_i}} - X_{\lambda_{s_{i+1}}}\| = d_i/2$. Рассуждая от противного, убедимся в том, что $s_k = s_\alpha$ при некотором конечном k . Пусть последовательность $\{s_i\}_{i \geq 1}$ бесконечна. Поскольку

интервал $[s_\alpha, s_\beta]$ компактен, имеем $s_{i-1} - s_i \rightarrow 0$ при $i \rightarrow \infty$. В этом случае $d_i \rightarrow 0$ и $s_i \rightarrow s_* > s_\alpha$ при $i \rightarrow \infty$. Пусть d — диаметр шара, принадлежащего области притяжения точки $X_{\lambda_{s_*}}$. По лемме 3.2.3.4 имеем $d_i > d/2$ при достаточно большом номере i . Получили противоречие. Поэтому построенная нами последовательность конечна. Искомое управление принимает вид

$$v(X) = v(X, s_i), \text{ где } i = \min \{1 \leq j \leq k \mid \|X - X_{\lambda_{s_j}}\| < d_j\},$$

где $v(X, s_i)$ задается формулой из леммы 3.2.3.4, т.е. используется управление иерархической переменной структуры [23], которое функционирует следующим образом. Если движение начинается в области притяжения точки $X_{\lambda_{s_i}}$, то реализуется управление $v(X, s_i)$, стабилизирующее эту точку. По построению стационарная точка $X_{\lambda_{s_i}}$ вместе с некоторой окрестностью принадлежит области притяжения точки $X_{\lambda_{s_{i-1}}}$.

Как только движение оказывается в области притяжения стационарной точки $X_{\lambda_{s_{i-1}}}$, управление переключается на более приоритетное $v(X, s_{i-1})$ и т.д. Таким образом, все решения из множества $\|X - X_\alpha\| \leq d_k$ стремятся к X_β при $t \rightarrow \infty$. Лемма доказана.

Завершая доказательство теоремы, определим управление в фазовом пространстве системы по следующему правилу. В малой окрестности множества $ST \setminus X_0$ используется управление, построенное по лемме 3.2.3.6. В остальных точках фазового пространства используется управление, приведенное в лемме 3.2.3.1. Использование леммы 3.2.3.5 завершает доказательство.

Стационарное решение X_0 реализуется на границе $v = 0$ области допустимых значений управления. Поэтому задача локальной стабилизации этого решения неразрешима классическими методами.

Линеаризуем уравнение (2.1.2) в окрестности стационарного решения X_0 . Имеем

$$\dot{Y} = DY - vX_0, \tag{3.2.9}$$

где $D = D_X F(X, \varphi(X))(X_0)$ (см. формулу (2.2.7)) и $v \geq 0$. Уравнение определено в m -мерном евклидовом пространстве E , которое может быть представлено в виде прямой суммы $E = E^1 \oplus E^2$ подпространств размерности m_1 и m_2 , соответствующих неустойчивой, включая мнимые значения, и устойчивой частям спектра матрицы D . Обозначим через

$D_i(\cdot) - X_{0i}v$ ограничение линейного оператора $D(\cdot) - X_0v$ на инвариантное подпространство E^i ($i = 1, 2$).

Теорема 3.2.4. Если $\text{rank}(D, X_{01}) = m_1$, то точка X_0 глобально стабилизируема в классе измеримых неотрицательных функций $v = v(X)$ с помощью иерархической системы переменной структуры. Число переключений управлений конечно на любом замкнутом множестве, не содержащем точки X_0 .

Доказательство. Воспользуемся теоремой 3.2.1 для доказательства существования стабилизирующего управления в системе (2.1.2) в окрестности точки X_0 , $v = 0$. В обозначениях теоремы 3.2.1 имеем $K(R) = K(X_0)$, $K(Q) = K([0, 1]X_0)$. Пусть $Q_a = [0, a]$ и $\partial Q_a = \{0, a\}$, где a — произвольное положительное число. Тогда $K(Q) = K(Q_a) = K(\partial Q_a)$. Из доказательства леммы 3.2.1.2 видно, что существует $T > 0$, при котором множество концов решений, начинающихся при $t = 0$, на устойчивом многообразии за время $t \in [-T, 0]$ покрывает окрестность нуля, каково бы ни было малое число a . Следовательно, для любой точки из окрестности начала координат $A_a = A_c(T, \partial Q_a)$ можно построить кусочно-постоянное управление $v: [0, T] \rightarrow \{0, a\}$, при котором решение, начинающееся в этой окрестности, попадает на устойчивое инвариантное многообразие за конечное время $t \leq T$, не зависящее от этой точки. Положив на инвариантном многообразии $v = 0$, найдем такое конечное время τ , что решение $X(t)$ уравнения (2.1.2) с начальным условием $X(0)$ принадлежит сколь угодно малой $\delta/2$ -окрестности точки X_0 при $t > \tau \geq T_0$.

Управление $v(t)$ всегда можно выбрать так, что движение $X(t)$ не имеет самопересечений. В этом случае отображение $t \rightarrow X(t)$ взаимно однозначно на траектории системы, и можно построить на этой траектории отображение $X \rightarrow t \rightarrow u(t)$. Используя решение $X(t)$ как эталонное, зададим в его окрестности управление v как функцию координат так, что все решения из окрестности траектории $\{X(t) | t \geq 0\}$ попадают в δ -окрестность точки X_0 за конечное время. Построение проведем, двигаясь вдоль траектории $X(t)$ от конца $X(\tau)$ к началу $X(0)$. Пусть τ_1 ($\tau_1 < \tau$) — момент первого от конца переключения. Положим для определенности $v(\tau_1 - 0) = 1$ и $v(\tau_1 + 0) = 0$. Пусть W — такая окрестность точки $X(\tau)$, что решения уравнения (2.1.2) с управлением $v = 0$, начинающиеся в момент времени τ_1 , в этой окрестности не отходят от траектории $\{X(t) | \tau_1 \leq t \leq \tau\}$ более чем на $\delta/2$. Проведем через точку $X(\tau_1)$ поверхность, которая позволит однозначно задать управление в окрестности W и окажется локальным сечением динамической системы (2.1.2) [2, 16].

Обозначим $\dot{X}_- = -\dot{X}(\tau_1 - 0)$ и $\dot{X}_+ = \dot{X}(\tau_1 + 0)$. Эти векторы отличны от нуля в силу выбора управления и не совпадают. Проведем через точку $X(\tau_1)$ гиперплоскость R так, что точки $X(\tau_1) + \dot{X}_-$ и $X(\tau_1) + \dot{X}_+$ не принадлежат R и лежат по разные стороны от нее. Положим $v = 0$ для точек окрестности W , которые лежат в том же полупространстве, что и \dot{X}_+ и $v = 1$ в другом полупространстве. Поскольку векторы \dot{X}_+ и \dot{X}_- трансверсальны R , множество решений (2.1.2) при $v = 0$ ($v = 1$), начинающихся на $R \cap W$ в нулевой момент времени, порождает при $0 \leq t \leq \alpha$ ($-\alpha \leq t \leq 0$), где α — малое положительное число, множество W_+ (W_-) с непустой внутренностью (теорема [16] о выпрямлении). Поэтому множество $W_- \cup W_+$ является окрестностью точки $X(\tau_1)$. Определим трубку траекторий системы (2.1.2) при $v = 0$ соотношением $P_1 = \{ \bigcup_{\tau_1 \leq t \leq \tau} X(t, Z) | Z \in W_+ \}$ [16], где $X(t, Z)$ —

решение уравнения (2.1.2) при $v = 0$ и $X(\tau_1, Z) = Z$. На множестве P_1 положим управление $v = 0$. Пусть при движении из точки $X(0)$ в точку $X(\tau)$ управление $v(t)$ переключается в моменты времени τ_i ($i = 1, \dots, n$) и пусть $\tau_{n+1} = -\alpha$. Рассуждая по индукции, построим трубки P_i , каждая из которых содержит внутри себя траектории $\{X(t) | \tau_i < t < \tau_{i+1}\}$, и на каждой из них зададим управление $v(X)$, равное $v(t)$ при $\tau_i < t < \tau_{i+1}$ ($i = 1, \dots, n$), где $v(t) = v(0)$ при $-\alpha \leq t \leq 0$. Множество

$P(X(0)) = \bigcup_{i=1}^n P_i$ имеет непустую внутренность, которой принадлежит точка $X(0)$.

Пользуясь компактностью A_1 , найдем конечный набор трубок $P(X_j)$ ($j = 1, \dots, k$), покрывающий A_1 . Пусть на $P(X_j)$ управление равно $v_{X_j}(X)$. Зададим на A_1 управление $v^i(X) = v_{X_r}(X)$, где $r = \max \{1 \leq j \leq k \mid x \in P(X_j)\}$. Тогда всякое решение из A_1 попадает за конечное время в δ -окрестность точки X_0 . Продолжим эти рассуждения по индукции. Как было сказано в начале доказательства, каждому числу 2^{1-i} ($i \geq 1$) можно сопоставить такое число δ_i , что решения из окрестности $A_{2^{1-i}}$ диаметра δ_i за время T попадают на устойчивое инвариантное многообразие, не выходя из этой окрестности, и при $t > T$ асимптотически приближаются к X_0 . По аналогии с предыдущим каждой такой δ_i -окрестности сопоставим управление $v^i(X)$ ($i \geq 1$), переводящее решения из δ_i -окрестности в δ_{i+1} -ю. На множестве A_1 управление зададим следующей иерархией переключений: $v(X) = v^r(X)$, где $r = \max \{i \geq 1 \mid X \in A_{2^{1-i}}\}$. Очевидно, что $\delta_i \rightarrow 0$ при $i \rightarrow \infty$. Отсюда следует асимптотическая устойчивость точки X_0 . Использование теоремы 3.2.3 завершает доказательство.

Перейдем к изучению процесса управления (2.1.3). Пусть ST_B^0 — множество стационарных решений уравнения (2.1.3), реализующихся при $v = 0$.

Теорема 3.2.5. Множество $ST_B \setminus ST_B^0$ глобально стабилизируемо процессом управления (2.1.3) в классе кусочно-непрерывных функций. Управление может осуществляться иерархической системой переменной структуры, порождающей скользящие режимы.

Доказательство. Доказательство состоит в сведении к случаю, рассмотренному в теореме 3.2.3. Пусть стационарный режим $Z \in ST_B \setminus ST_B^0$ реализуется при $0 < s_1 < s_\infty$, $u_1 > 0$ и $v_1 > \max \{0, \lambda_{s_1}\}$ (см. теорему 2.1.3). Обозначим через $\nu(s)$ произвольную непрерывную функцию, удовлетворяющую условиям $\nu(s) > \max \{0, \lambda_s\}$ при $0 < s < s_1$ и $\nu(s_1) \neq v_1$.

Пусть $\beta(s) = u(s, \nu(s))$, где $u(s, v)$ — функция из условия теоремы 2.1.3. Обозначим через ST_B^1 множество стационарных решений, реализующихся при $0 < s \leq s^1$, $v = \nu(s)$ и $u = u(s, \nu(s))$. Пусть $Y \in ST_B^1$ и $\varphi(Y) = \alpha$. Положим в окрестности точки Y управление $u(X) = \beta(\alpha)\varphi(X)$. Тогда Y является стационарным решением уравнения $\dot{X} = F(X, \varphi(X)) - vX - \beta(\alpha)B\varphi(X)$.

где $v = \nu(\alpha)$. Это стационарное решение локально стабилизируемо управлением из леммы 3.2.3.4. Глобальная стабилизируемость решения Z осуществляется построением последовательности локально устойчивых точек $Y_n = 0, \dots, Y_1 = Z$, где $0 = \varphi(Y_n) = s_n < \dots < \varphi(Y_i) = s_i < \dots < \varphi(Y_1) \neq s_1$ с пересекающимися областями притяжения и заданием такой же иерархии управлений, как и в теореме 3.2.3. Теорема доказана.

Обратимся к вопросу стабилизируемости множества ST_B^1 . Пусть Z_0 — стационарное решение уравнения (2.1.3), реализуемое при $0 < s < s_\infty$ и $v = 0$. Пусть $u_0 = u(s, 0)$, где $u(s, v)$ — функция из условия теоремы 2.1.3. Линеаризовав уравнение (2.1.3) в окрестности точки Z_0 , $v = 0$ и $u = u_0$, имеем $\dot{Y} = DY + wB$, где $D = D_X F(X, \varphi(X))(Z_0)$ и $W \geq -u_0$.

Теорема 3.2.6. Все точки множества ST_B^0 , в которых выполнено условие $\text{rank}(D, B) = m$, глобально стабилизируемы в классе кусочно-непрерывных функций с помощью иерархической системы переменной структуры.

Доказательство. Так как $w > -u_0$ может принимать значения обоих знаков, то точка Z_0 стабилизируема линейным управлением в некоторой

окрестности A [14]. Пусть v_1 — настолько малое положительное число, что стационарное решение уравнения (2.1.3) при s, v_1 и $u(s, v_1)$ принадлежит ST_B^0 . Воспользовавшись теоремой 3.2.5, завершаем доказательство теоремы.

Завершим наше рассмотрение результатом об управляемости изучаемых процессов. Вначале сформулируем

Определение. Точка Z из множества стационарных решений ST_B ($Z \in ST$) называется глобально достижимой, если для любой точки $X \in R_+^m \setminus 0$ существуют кусочно-непрерывные функции $v(t)$ и $u(t)$ ($u \equiv 0$), принимающие неотрицательные значения и такие, что процесс управления переводит систему из точки X в точку Z за конечное время.

Пусть стационарное решение Z реализуется при постоянных управлениях $v_1 \geq 0$ и $u_1 \geq 0$ и $D = D_X F(X, \varphi(X))(Z)$. Введем в рассмотрение матрицу $[Z, B]$, первый столбец которой образован вектором Z , а остальные столбцы — столбцами матрицы B . Очевидным следствием полученных результатов является

Теорема 3.2.7. Если $\text{rank}(D, [Z, B]) = m$ и не существует вещественного собственного вектора C матрицы A^* такого, что $C^*(Zv + Bu) \leq 0$ при всех $-v_1 \leq v$ и $-u_1 \leq u$, то точка $Z \in ST_B$ ($Z \in ST$) глобально достижима процессом (2.1.2), (2.1.3).

Доказательство. В силу теорем 3.2.3–3.2.6 для любой точки фазового пространства найдется кусочно-непрерывное по времени управление, которое переводит эту точку в окрестность любой наперед заданной стационарной точки. При выполнении условий теоремы 3.2.7 система (2.1.2) или (2.1.3) локально управляема в окрестности стационарной точки [26]. Теорема доказана.

Подведем итог полученным результатам. Мы рассмотрели динамику численности популяции или сообщества, которые контролируются с помощью механизма потребления в одном случае и с помощью механизмов потребления и добавления в другом. Показано, что любой стационарный режим стабилизируем однонаправленным управлением с обратной связью, имеющим переменную структуру. Для почти всех стационарных режимов существует окрестность, в которой сходимось к стационарной точке осуществляется в скользящем режиме. Показано, что глобальная управляемость рассмотренных процессов с помощью разомкнутого управления сводится к условиям общности положения в окрестности стационарных точек.

Использование скользящих режимов с целью стабилизации процесса воспроизводства популяции в искусственных условиях нашло в настоящее время широкое практическое применение. О моделях таких систем пойдет речь в следующих параграфах.

§ 3.3. Оптимальное управление экологическими системами

Проблема принятия оптимальных решений при эксплуатации экологических ресурсов приводит к следующей математической постановке. В виде системы дифференциальных уравнений строится модель динамики биологического объекта, определяются области допустимых значений управления и фазовых переменных и формулируется критерий качества процесса. Таким критерием может быть объем собираемого урожая, доходы от

производства и т.п. Задача оптимального управления состоит в следующем. Требуется найти кусочно-непрерывное управление, принимающее допустимые значения и такое, что биологическая система переходит из начальной точки в требуемую конечную, а критерий качества принимает при этом минимальное значение.

Процедура решения задачи оптимального управления естественным образом распадается на два этапа. Вначале следует доказать, что конечная точка достижима при некотором допустимом управлении. Этот результат будет означать, что множество управляющих функций, на котором должна проводиться оптимизация, непустое. Следующий этап состоит в построении управления, минимизирующего заданный критерий качества. На этом этапе можно использовать принцип максимума или метод динамического программирования.

Цель данного параграфа состоит в том, чтобы связать проблему управляемости экологических систем с проблемой существования оптимального управления.

Допустим, что динамика экологической системы задается моделью

$$\dot{X} = F(X, U), \quad (3.3.1)$$

где $X \in R_+^m$, $U \in Q = \{U = (u_i)_{i=1}^p \in R_-^p \mid 0 \geq u_i \geq -1 \ (i = 1, \dots, p)\}$ и $F(X, U) = \{f_i(X, U)\}_{i=1}^m$ — непрерывно дифференцируемый по обеим переменным оператор. Пусть в начальный момент времени состояние системы задается вектором X_0 или может принимать произвольные значения из замкнутого ограниченного множества Ξ . Рассматривается задача выведения системы за время, не превосходящее T , в терминальную точку X_1 или в произвольную точку замкнутого ограниченного множества Ξ_1 . Если сформулированы дополнительные ограничения на фазовые переменные системы, например, их неотрицательность, то в процессе движения эти ограничения должны выполняться. Всякая кусочно-непрерывная или измеримая функция $U(t)$, принимающая значения в Q - и удовлетворяющая сформулированным выше условиям, называется допустимым управлением.

Показатель качества процесса задает количественную оценку эффективности каждого допустимого управления и имеет следующий вид:

$$J(U) = \int_0^t f_0(X(t), U(t)) dt,$$

где t — время движения в требуемую точку, $f_0(X, U)$ — непрерывно дифференцируемая функция затрат.

Задача оптимального управления состоит в том, чтобы отыскать такое допустимое управление $U_0(t)$, при котором $J(U_0(t)) \leq J(U(t))$ при всех других допустимых управлениях $U(t)$. В том случае, когда ставится задача максимизации дохода $g(X, U)$, функционал задается в виде

$$- \int_0^t g(X(t), U(t)) dt.$$

Например, задача максимального быстродействия определяется функцио-

налом $\int_0^t 1 dt$, т.е. $f_0 \equiv 1$, задача минимизации затрат на управление имеет вид $\int_0^t C^* U(t) dt$, где $C \in R_+^p$ – вектор цен и т.д.

В предыдущем параграфе рассматривался вопрос о том, когда множество допустимых управлений $\Delta(T)$ непусто. Приведем теперь результат о существовании оптимального управления.

Теорема 3.3.1 [14]. Предположим, что множество допустимых управлений непусто и все решения системы (3.3.1), полученные при $U(t) \in \Delta(T)$, равномерно ограничены, т.е. $\|X(t)\| \leq b$ во все моменты $t \in [0, t_1]$ ($0 \leq t_1 \leq T$) и b не зависит от $U(t)$. Если при любом $X \in R_+^m$ множество $\{F(X, U) | U \in Q\}$ выпукло, то задача оптимального управления имеет решение.

Совершенно очевидно, что при линейности правой части модели (3.3.1) по U последнее условие теоремы выполняется. Докажем равномерную ограниченность движений в моделях нелимитированных экологических систем.

В линейных моделях, а также в системах, линеаризованных в окрестности положения равновесия, условие равномерной ограниченности, очевидно, выполняется. Рассмотрим нелинейную модель с линейным управлением (3.1.7). В данном случае правая часть считается независимой от времени. Тогда $-C + F(X) \leq \dot{X} \leq F(X) + C$, где $C \in R_+^m$ – такой вектор, что $-C \leq BU \leq C$ при $U \in Q$. Отсюда на основании леммы сравнения [21] заключаем, что $Y^-(t) \leq X(t) \leq Y^+(t)$, где $X(t)$ – решение системы (3.1.7), а $Y^\pm(t)$ – решения систем $\dot{Y}^\pm = F(Y^\pm) \pm C$ с тем же начальным условием. Таким образом, равномерная ограниченность за конечное время доказана.

Рассмотрим модель (3.1.1) с ограничениями, задаваемыми условиями теоремы 3.1.14. Пусть $E = (1, \dots, 1)^*$, $P_i = (\underbrace{1, \dots, 0}_{i}, \dots, 1)^*$ и $E_i = (\underbrace{0, \dots, 1}_{i}, \dots, 0)$. Тогда $f_i(X, 0) \leq f_i(X, U) \leq f_i(X, -E)$ при $\partial f_i / \partial u_i \leq 0$, $f_i(X, -E_i) \leq f_i(X, U) \leq f_i(X, -P_i)$ при $\partial f_i / \partial u_i \geq 0$ и $U \in Q$ ($i = 1, \dots, m$). Далее, на основании леммы сравнения [21] получаем равномерную ограниченность решений.

Докажем равномерную ограниченность решений в моделях (2.1.2) и (2.1.3). В силу положительной однородности имеем

$$f_i(X, s) / \rho(X) \leq \text{Sup} \{ f_i(X, s) / \rho(X) \leq 1, x_i \geq 0, i = 1, \dots, m \} \leq c(s).$$

Допустим, что $c(s) \leq c_0$ при всех $s \geq 0$. Тогда

$$\dot{x}_i \leq c_0 \rho(X) + b_i \quad (i = 1, \dots, m).$$

В силу леммы сравнения [21] равномерная ограниченность сверху доказана. Так как решения принадлежат R_+^m , равномерная ограниченность нормы решения установлена.

Проиллюстрируем теорему 3.3.1 примерами.

Пример 1. В § 3.1 для линейной модели нелимитированной экологической системы второго порядка была доказана неразрешимость задачи

попадания в равновесное состояние за конечное время при одном управлении фиксированного знака. Для нелинейных моделей второго порядка этот вопрос оставался открытым. Воспользовавшись теоремой 3.3.1 и принципом максимума, докажем сходный результат в нелинейном случае. Рассмотрим систему

$$\begin{aligned}\dot{x}_1 &= f_1(x_1, x_2) - u, \\ \dot{x}_2 &= f_2(x_1, x_2),\end{aligned}$$

где $\partial f_i / \partial x_j \geq 0$ при $i \neq j$ ($i, j = 1, 2$) и управление U принимает значения из диапазона $u \in [0, u_0]$. Предположим, что начало координат помещено в стационарную точку, т.е. $f_1(0, 0) = f_2(0, 0) = 0$. Тогда имеет место следующее утверждение [28].

Какова бы ни была амплитуда изменения управления u_0 , задача попадания в равновесную точку из любой точки X_0 со строго положительными компонентами не имеет решения.

Действительно, допустим противное. Тогда по теореме 3.3.1 задача максимального быстрогодействия разрешима. В силу принципа максимума на оптимальной траектории управление удовлетворяет условию $\psi_1(t)u(t) \rightarrow \max$, где $\psi_1(t)$ – первая компонента вектора $\psi = (\psi_1, \psi_2)$, являющегося решением системы линейных уравнений $\dot{\psi} = [D_X F]^* \psi$ ($F = (f_1, f_2)^*$).

Матрица $D_X F$ в силу квазимонотонности имеет неотрицательные недиагональные элементы. Следовательно, для любого $t \in [0, T]$ выполняются неравенства $\psi(t) \leq 0$ ($\psi(t) \geq 0$) при $t \leq s$, если $\psi(t) \leq 0$ ($\psi(t) \geq 0$). Поскольку 0 – стационарная точка, попасть в нее можно лишь при ненулевом управлении $u(T-0) > 0$. Следовательно, $\psi(T) > 0$. Отсюда возможны следующие три случая:

1. $\psi_1(T) > 0$ и $\psi_2(T) > 0$. Тогда $\psi_1(t), \psi_2(t) > 0$ и $u(t) > 0$ при $t \in [0, T]$.

2. $\psi_1(t) > 0$ и $\psi_2(t) < 0$ при $t_1 < t < T$ и $\psi_1(t_1 - 0) > 0, \psi_2(t_1 - 0) > 0$ при $t < t_1$. Следовательно, $u(t) > 0$ во все моменты времени.

3. $\psi_1(t) > 0$ и $\psi_2(t) < 0$ при $t_1 < t < T$ и $\psi_1(t_1 - 0) < 0$.

Следовательно, $\psi(t) \leq 0$ при $t < t_1$. Отсюда $u = 0$ при $t < t_1$ и $u(t) > 0$ при $t > t_1$. Одновременно $\psi_1(t)$ и $\psi_2(t)$ в нуль не обращаются в силу единственности решений и однородности системы. В случаях 1 и 2 $\dot{x}_1 < 0$ при $0 < t < T$ и, следовательно, $x_1(t) > 0$ при $0 < t < T$. В случае 3 $u = 0$ при $t < t_1$. Из $X(0) > 0$ получаем, что $X(t_1) > 0$ в силу монотонности оператора сдвига вдоль траекторий, которая следует из квазимонотонности правой части [21]. Поскольку $x_1(t_1) > 0$ и $\dot{x}_1 < 0$ при $t > t_1$, имеем $x_1(t) > 0$ при $t < T$.

Из второго уравнения системы следует, что

$$\dot{x}_2 = f_2(x_1, x_2, t) \geq f_2(0, x_2, t).$$

Отсюда по лемме сравнения [21] $x_2(t) \geq y_2(t)$, где $y_2(t)$ ($y_2(0) = x_2(0)$) – решение уравнения $\dot{y}_2 = f_2(0, y_2, t)$. В силу единственности $y_2(t) > 0$ при $t \in [0, T]$. Таким образом, $X(t) \not\geq 0$ при $t \in [0, T]$. Полученное противоречие доказывает недостижимость точки равновесия.

Пример 2. В п. 2.1.5 рассматривалась модель динамики численности непрерывно культивируемой популяции микроорганизмов. Для простоты предположим, что клетки в популяции не различаются по фазам развития

и их численность задается скалярной величиной $x(t)$. Тогда система (1.2.1) принимает вид

$$\begin{aligned} \dot{x} &= p(s)x - vx, \\ \dot{s} &= gp(s)x - vs + vs_p. \end{aligned} \quad (3.3.2)$$

В дальнейшем будем считать, что $g = 1$. Это условие легко выполнить с помощью замены переменной $gx = y$. в качестве управляющего параметра будет использоваться концентрация подаваемого в реактор субстрата $s_p \in [0, a]$. Пусть при $s_p^0 < a$ выполняется условие $p(s_p^0) > v$. Как показано в п. 2.1.5, в этом случае стационарное решение существует и единственно. Оно задается системой

$$\begin{aligned} s_0 &= p^{-1}(v), \\ x_0 &= s_p^0 - s_0. \end{aligned}$$

Областью притяжения этого решения является множество $\Omega \{ x > 0, s \geq 0 \}$. Построим оптимальное управление $s_p(x, s)$, переводящее решения системы (3.3.2) в точку (x_0, s_0) за наименьшее время и удовлетворяющее ограничениям $0 \leq s_p \leq a$. В связи с тем, что s_p^0 — внутренняя точка области допустимых значений, управление можно представить в виде $s_p = s_p^0 + u$, где u изменяется в пределах $-s_p^0 \leq u \leq (a - s_p^0) = u_0$ и, следовательно, меняет знак. Условие достижимости стационарной точки при малых отклонениях от нее в этом случае гарантируется калмановским критерием [14]

$$\det \begin{bmatrix} 0 & vp'(s_0) \\ v & -v(v + gp'(s_0)x_0) \end{bmatrix} \neq 0,$$

который, очевидно, выполняется. По теореме 3.3.1 оптимальное управление существует. Для его отыскания воспользуемся принципом максимума [4, 14] и покажем, что оно единственно. Поскольку принцип максимума дает необходимое условие оптимальности, отсюда будет следовать, что построенное ниже управление решает поставленную задачу.

Прежде всего найдем те значения, которые может принимать оптимальное управление. Выпишем для этого гамильтониан системы

$$H = \psi_1 [p(s) - v] + \psi_2 [-p(s) - v] + \psi_2 vs_p^0 + \psi_2 u.$$

На основании принципа максимума заключаем, что при $\psi_2 \neq 0$ оптимальное управление принимает лишь два значения:

$$u = \begin{cases} -s_p^0, & \text{если } \psi_2 < 0, \\ u_0, & \text{если } \psi_2 > 0. \end{cases} \quad (3.3.3)$$

Найдем теперь число изменений величины управления u , называемое числом переключений управления. Из формулы (3.3.3) видно, что число переключений равно числу перемен знака функции $\psi_2(t)$. Исследуем поведение $\psi_2(t)$ при движении точки $(x(t), s(t))$ вдоль оптимальной траектории. С этой целью выпишем уравнения, которым удовлетворяют вспомогательные переменные ψ_1 и ψ_2 :

$$\begin{aligned} \dot{\psi}_1 &= -(p(s) - v)\psi_1 + p(s)\psi_2, \\ \dot{\psi}_2 &= -\frac{\partial p}{\partial s}x\psi_1 + \left(\frac{\partial p}{\partial s}x + v \right)\psi_2. \end{aligned}$$

Отсюда следует, что

$$\frac{d}{dt}(\psi_2 - \psi_1) = \left(v + \frac{\partial p}{\partial s} x - p \right) (\psi_2 - \psi_1).$$

Функции $\frac{\partial p}{\partial s}(s(t))$, $x(t)$ и $p(s(t))$, очевидно, являются ограниченными и интегрируемыми вдоль траекторий системы (3.3.2) при $u = -s_p^0$ и $u = u^0$. Поэтому разность $\psi_2 - \psi_1 = \varphi(t)$ представима в виде

$$\varphi(t) = \psi_2 - \psi_1 = (\psi_2(0) - \psi_1(0)) \exp \left(vt + \int_0^t \left(\frac{\partial p}{\partial s} x - p \right) dt \right).$$

Отсюда получаем, что функция φ не меняет знак в течение всего движения. Выразим переменную ψ_1 через φ и ψ_2 :

$$\dot{\psi}_2 = v\psi_2 + \frac{\partial p}{\partial s} x\varphi,$$

откуда имеем

$$\psi_2 = \left[\psi_2(0) + \int_0^t \frac{\partial p}{\partial s} x \varphi e^{-v\tau} d\tau \right].$$

Так как $\frac{\partial p}{\partial s}(s(t)) \geq 0$ и $x(t) > 0$, то ψ_2 меняет знак не более одного раза

Кроме того, при оптимальном решении $(x(t), s(t))$ функция ψ_2 не может тождественно равняться нулю на некотором отрезке времени. Отсюда следует, что оптимальное управление может принимать лишь два значения, заданные формулой (3.3.3).

Чтобы найти окончательный вид оптимального управления, рассмотрим поведение оптимальных траекторий на фазовой плоскости (x, s) . Обозначим через L_1 семейство кривых, соответствующих управлению $u = -s_p$. Эти кривые являются траекториями решений системы вида

$$\begin{aligned} \dot{x} &= (p(s) - v)x, \\ \dot{s} &= -p(s)x - vx. \end{aligned} \quad (3.3.4)$$

На фазовой плоскости (x, s) семейство L_1 образовано решениями уравнения

$$\frac{ds}{dx} = -\frac{p(s)x + vs}{(p(s) - v)x}.$$

Проведем на плоскости (x, s) (рис. 3.1) прямую $s = s_0$, где s_0 удовлетворяет условию $p(s_0) = v$. Поэтому семейство кривых L_1 ортогонально прямой $s = s_0$. Кроме того, из монотонности функции $p(s)$ и положительности величин $p(s)$, x и s следует, что $ds/dx > 0$ при $s < s_0$ и $ds/dx < 0$ при $s > s_0$.

Сложим уравнения системы (3.3.4). Имеем $\dot{x} + \dot{s} = -v(x + s)$. Отсюда видно, что $x + s \rightarrow 0$ при $t \rightarrow \infty$. Кроме того, $\dot{s} < 0$, поэтому $s \rightarrow 0$ при $t \rightarrow \infty$ и, следовательно, $x \rightarrow 0$ при $t \rightarrow \infty$. Итак, решения системы (3.3.4), начинающиеся в области $\Omega \{x > 0, s \geq 0\}$, стремятся к точке $(x = 0, s = 0)$

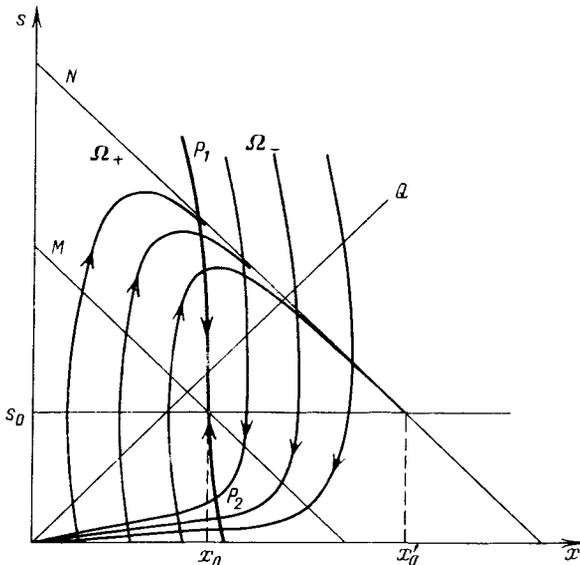


Рис. 3.1. Линии переключения управления и оптимальные траектории в задаче наискорейшего выведения культивируемой популяции в равновесное состояние

при $t \rightarrow \infty$. Поэтому семейство кривых проходит через начало координат, имеет вид, показанный на рис. 3.1, и точка $(x(t), s(t))$ движется вдоль траектории семейства сверху вниз.

В дальнейшем исследовании важную роль будет играть часть траектории семейства L_1 , проходящей через точку (x_0, s_0) , которая лежит выше прямой $s = s_0$. Обозначим эту кривую через P_1 . Рассмотрим теперь семейство оптимальных траекторий, соответствующих управлению $u = u_0$. Обозначим это семейство через L_2 . Оно задается системой

$$\begin{aligned} \dot{x} &= (p(s) - v)x, \\ \dot{s} &= -p(s)x - vs + v(s_p^0 + u_0). \end{aligned} \quad (3.3.5)$$

Легко видеть на основании равенства $p(s_0) = v$, что траектории L_2 ортогональны прямой $s = s_0$, из монотонности функции $p(s)$ получаем также, что $\dot{x} < 0$ при $s < s_0$, и поэтому точка $(x(t), s(t))$ движется влево, а при $s > s_0$ производная $\dot{x} > 0$ и точка $(x(t), s(t))$ движется вправо (см. рис. 3.1).

Рассмотрим кривую P_2 семейства L_2 , при движении по которой точка $(x(t), s(t))$ попадет в (x_0, s_0) . При $x = x_0$ и $s = s_0$ величина s возрастает, так как $\dot{s}|_{x=x_0, s=s_0} = u_0 > 0$, и поэтому кривая P_2 входит в точку (x_0, s_0) снизу.

Покажем, что линия P_2 задается монотонно убывающей функцией и пересекает ось $s = 0$ (см. рис. 3.1). Построим на прямой $s = s_0$ точку x'_0 , найденную из условия $x'_0 = s_p^0 + u_0 - s_0$. В силу теоремы 2.1.14 точка (x'_0, s_0) является асимптотически устойчивым стационарным решением системы (3.3.5).

Проведем из начала координат вектор Q под углом $\pi/4$ к оси x и построим прямую N , перпендикулярную вектору и проходящую через точку (x'_0, s) (см, рис. 3.1). Складывая уравнения системы (3.3.5), получаем

$$\dot{x} + \dot{s} = -v(x+s) + v(s_p^0 + u_0).$$

Отсюда видно, что прямая N является траекторией решения системы (3.3.5), причем любое движение, начавшееся слева (справа) от прямой N , всегда остается слева (справа) от нее. Если в начальный момент времени точка $(x(0), s(0))$ была слева от N , то решение $(x(t), s(t))$ движется по направлению вектора Q , так как $\dot{x} + \dot{s} > 0$. Из системы (3.3.5) имеем

$$\frac{dx + ds}{dx} = \frac{-v(x+s) + v(s_p^0 + u_0)}{(p(s) - v)x}.$$

Числитель правой части уравнения положителен в части плоскости, лежащей слева от N , а знаменатель неположителен при $s \leq s_0$. Поэтому для участков траекторий семейства L_2 , лежащих в части положительного квадранта, отгороженной прямыми $s = s_0$ и N , выполняется условие

$$\frac{ds + dx}{dx} \text{ или } \frac{ds}{dx} \leq -1.$$

Кривая P_2 , лежащая левее N , заходит в точку (x_0, s_0) снизу. В силу неравенства $\frac{ds}{dx} \leq -1$ она остается снизу от прямой $s = s_0$, и поэтому

для нее в каждой точке выполняется неравенство $\frac{ds}{dx} \leq -1$. Проинтегрировав его, получаем

$$s \leq -(x - x_0) + s_0.$$

Отсюда следует, что кривая P_2 пересекает ось $s = 0$ слева от прямой M , перпендикулярной вектору Q и проходящей через точку (x_0, s_0) (см. рис. 3.1).

Подводя итог сказанному, мы получаем, что монотонно убывающая кривая P_1P_2 разбивает область $\Omega \{x > 0, s \geq 0\}$ на множество Ω_+ , лежащее слева от P_1P_2 , и множество Ω_- , лежащее справа от P_1P_2 . Иными словами, $\Omega = \Omega_+ \cup \Omega_- \cup (P_1P_2)$.

Покажем, что оптимальное управление задается следующим правилом:

$$u = \begin{cases} u_0 & \text{при } (x, s) \in \Omega_+ \cup P_2, \\ s_p^0 & \text{при } (x, s) \in \Omega_- \cup P_1, \end{cases} \quad (3.3.6)$$

т.е. P_1P_2 образует линию переключения.

Доказательство. Прежде всего покажем, что решение $(x(t), s(t))$ системы (3.3.5), начинающееся в области Ω_- , всегда в ней и остается (инвариантность Ω_- при $t \in [0, \infty)$). Допустим противное. В силу единственности траектория решения (3.3.5) не может иметь ни одной общей точки с линией P_2 . Поэтому попасть в множество $\Omega \setminus \Omega_1$ движение $(x(t), s(t))$ может лишь через кривую P_1 в области $s > s_0$ положительного квадранта.

Пусть $A(x, s)$ — точка, принадлежащая P_1 , через которую проходит решение $(x(t), s(t))$, начинающееся в Ω_- . Построим в точке A нормаль N ,

направленную из области Ω_+ в область Ω_- (см. рис. 3.1). Так как траектория P_1 задается решением уравнения (3.3.4), то вектор L , касательный к кривой P_1 в точке $A(x, s)$, имеет следующие координаты:

$$L = \left(1, \frac{ds}{dx} \right) = \left(1, -\frac{p(s)x + vs}{(p(s) - v)x} \right).$$

Отметим, что $p(s) \neq v$ при $s > s_0$ и $x \neq 0$. Отсюда

$$N = \left(-\frac{ds}{dx}, 1 \right) = \left(\frac{p(s)x + vs}{(p(s) - v)x}, 1 \right).$$

Пусть $R(t) = (x(t), s(t))$ – решение системы (3.3.5). Мы предположили, что существует такой момент времени t' , при котором

$$R(t') = A(x, y).$$

Определим знак проекции производной $\frac{dR}{dt}$ на нормаль N , имеем

$$\left\langle \frac{dR}{dt}(t'), N \right\rangle = (s_p + u_0)v > 0.$$

Следовательно, $R(t') \in \Omega_-$ и $R(t') \neq A(x, y)$. Полученное противоречие доказывает, что множество Ω_- инвариантно относительно траекторий системы (3.3.5) при $t \in [0, \infty)$.

Аналогично доказывается, что решения системы (3.3.4), начавшиеся в области Ω_+ , остаются в ней при любом $t \in [0, \infty)$ (инвариантность Ω_+ при $t \in [0, \infty)$). Покажем теперь, что P_1P_2 образует линию переключения и оптимальное управление задается правилом (3.3.6).

Действительно, движения, начавшиеся в области Ω_+ и имеющие управление $u = u_0$, пересекают линию P_1 , так как они стремятся к точке (x'_0, s_0) , лежащей в области Ω_- . Движения, начавшиеся в области Ω_- и имеющие управление $u = -s_p^0$, пересекают линию P_2 , так как они стремятся к началу координат, лежащему в области Ω_+ . Из ранее доказанного следует, что кривые семейства L_2 могут иметь с линией P_1 не более одной общей точки, а кривые семейства L_1 также могут пересечь P_2 лишь один раз. В точку (x_0, s_0) решение $(x(t), s(t))$ должно попасть, двигаясь вдоль оптимальной траектории и имея не более одного переключения. Поэтому единственный способ попасть в нее – это выйти на линию P_1P_2 и двигаться вдоль этой линии к стационарной точке. Отсюда следует, что оптимальное управление должно иметь вид (3.3.6).

§ 3.4. Системы управления клеточными популяциями.

Прямое управление

Рассмотрим модель системы автоматического регулирования процессом культивирования, в которой скорость подачи питательных веществ и удаления приросшей биомассы функционально зависит от концентрации клеток в реакторе. В этом случае динамика численности клеточной популяции задается моделью (1.2.1), правая часть которой удовлетворяет всем условиям, сформулированным в п. 2.1.5, за исключением того, что скорость разбавления уже не является постоянной величиной, а зависит от состоя-

ния системы: $v = v(X)$ и допустимы режимы, при которых скорость воспроизводства не лимитируется субстратом.

В системах культивирования, основанных на турбидостатном принципе, поток через реактор включается в моменты, когда измеряемый показатель плотности популяции [20, 22] превосходит уставку. При этом контролируется либо оптическая плотность биомассы (плотностат), либо кислотность среды (pH -стат) [20, 22], которая характеризует концентрацию питательного субстрата в реакторе. Положим $\sigma = C_0^*X + C_{m+1}^0(s_p - s) - \sigma_0$, где

$$C_0^* = (c_1^0, \dots, c_m^0) \geq 0, \quad c_{m+1}^0 \geq 0 \text{ и } \sigma_0 > 0.$$

Ввиду изложенного управление задается выражением

$$v = \begin{cases} 0 & \text{при } \sigma < 0, \\ M & \text{при } \sigma > 0. \end{cases} \quad (3.4.1)$$

Константу M можно считать сколь угодно большой, так как скорость срабатывания технических устройств более чем на порядок превосходит скорость размножения клеток.

Если в системе реализуется скользящий режим, то гиперплоскость скольжения задается выражением

$$C^*X + c_{m+1}(s_p - s) - 1 = 0, \quad (3.4.2)$$

где $C = C_0/\sigma$ и $c_{m+1} = c_{m+1}^0/\sigma_0$, а эквивалентное управление [23] находится из условия $\dot{\sigma} = 0$ и имеет вид

$$v = C^*APX + c_{m+1}G^*PX. \quad (3.4.3)$$

Найдем условия, при которых решения, начинающиеся в плоскости скольжения, не выходят из нее за бесконечное время. Для того чтобы в каждой точке поверхности разрыва выполнялось условие скольжения, необходимо и достаточно, чтобы $\dot{\sigma} > 0$ при $v = 0$ в каждой точке множества (3.4.2). Поэтому в неотрицательном конусе $X \in R_+^m \setminus 0$ условие скольжения таково:

$$(C^*A + c_{m+1}G^*)P > 0.$$

Диагональ функции $P(S)$ положительна при $s > 0$ и может обращаться в нуль лишь при $s = 0$. Отсюда скользящий режим существует для любого нетривиального решения, если

$$C^*A + c_{m+1}G^* > 0 \quad (3.4.4)$$

и $s > 0$.

Выясним, когда $s = 0$ является гиперплоскостью отталкивания на множестве (3.4.2). Если $C = 0$ и $c_{m+1} > 0$, то из второго уравнения системы (1.2.1) и формул (3.4.2) и (3.4.3) имеем

$$G > 0 \text{ и } c_{m+1} > 1/s_p. \quad (3.4.5)$$

Если $C \neq 0$, то условие отталкивания гиперплоскости $s = 0$ таково:

$$G^*(c_{m+1} - 1/s_p) + C^*A > 0. \quad (3.4.6)$$

Легко убедиться в том, что неравенства (3.4.4) и (3.4.5) являются следствием (3.4.6).

Итак, если выполнено условие (3.4.6), то движение, начавшееся в множестве $\Omega = \{X \geq 0, X \neq 0, 0 \leq s \leq s_p, \sigma = 0\}$, целиком лежит в плоскости скольжения и ω -предельные точки любого нетривиального решения системы (1.2.1) при управлении (3.4.2) также принадлежат плоскости скольже-

ния. Добавляя естественное предположение о том, что либо $G > 0$, либо $C > 0$, докажем более сильное утверждение.

Предложение 3.4.1. При выполнении условий (3.4.6) система (1.2.1) с управлением (3.4.3) имеет в множестве Ω положительно инвариантный компакт Π , не содержащий $X = 0$ и $s = 0$. Множество Π является асимптотически устойчивым.

Доказательство. Рассмотрим систему (1.2.1) с управлением (3.4.3). Покажем, что из (3.4.6) следует $v > 0$. Убедимся в том, что X не обращается в нуль и

$$\int_0^t v dt = \int_0^t (C^*A + C_{m+1}G)PX dt \rightarrow \infty \text{ при } t \rightarrow \infty.$$

Действительно, в противоположном случае $X \rightarrow 0$, так как на основании (3.4.6) можно считать, что $s > \delta > 0$. Умножим первое уравнение системы (1.2.1) при управлении (3.4.3) слева на $L^*(L^*A = \lambda(A)L^*)$. Имеем $L^*\dot{X} = L^*PX + o(\|X\|)$. Отсюда, если $L^*X(0) > 0$ (т.е. $X(0) \neq 0$, так как $L > 0$), то существует $\epsilon > 0$ такое, что $L^*X \geq \epsilon$, начиная с некоторого t . Следовательно, $\int_0^t v dt \rightarrow \infty$.

Докажем теперь ограниченность решений. Если $C > 0$, то она следует из условия $C^*X \leq 1$. Пусть $G > 0$. Выберем число $\alpha > 0$ так, что $\alpha L < G$. Умножим первое уравнение системы слева на αL^* и сложим результат со вторым. Воспользовавшись неравенством $\alpha L^* < G^*$, получим

$$d/dt(\alpha L^*X + s - s_p) \leq v(\alpha L^*X + s - s_p).$$

Проинтегрировав его, воспользовавшись условием $\int_0^t v dt \rightarrow \infty$, имеем $\Pi = \{X, s \in \Omega \mid L^*X \geq \epsilon, L^*X + s \leq s_p\}$. Предложение доказано.

Область притяжения скользящего режима образуют все точки множества $\{X \geq 0, 0 \leq s \leq s_p, \sigma \geq 0\}$, а также точки множества $\{X \geq 0, 0 \leq s \leq s_p, \sigma < 0\}$, из которых движения системы (1.2.1) при $v = 0$ пересекают гиперплоскость $\sigma = 0$. В случае, когда $C = 0$ и $c_{m+1} \neq 0$, решение, очевидно, попадает в скользящий режим из любой точки множества $\{X \geq 0, s \geq 0\}$.

Будем в дальнейшем предполагать, что условие скольжения (3.4.5) выполнено.

Предложение 3.4.2. Система (1.2.1) при управлении (3.4.2) имеет в скользящем режиме нетривиальное стационарное решение $X_0 > 0$ и $s_p > s_0 > 0$.

Доказательство, очевидно, следует из выпуклости Π .

Рассмотрим вопрос устойчивости стационарного решения на всем множестве Ω .

Теорема 3.4.1. Если $C = 0$ и $c_{m+1} \neq 0$ или $c_{m+1} = 0$ и $C = \alpha P(s_0)G$ — собственный вектор $(AP(s_0))^*$, отвечающий $\lambda(AP(s_0))$ ($\alpha > 1/(\lambda(AP(s_0))s_p)$), то стационарное решение асимптотически устойчиво на множестве Ω . Если $C \neq 0$ — произвольный вектор, то существуют такие числа ϵ и h , что стационарное решение асимптотически устойчиво на множестве Ω , если одновременно выполняются неравенства

$$|p_i(s)/p_i(s_0) - p_j(s)/p_j(s_0)| \leq \epsilon |s - s_0|,$$

$$|p_i(s)/p_i(s_0) - p_j(s)/p_j(s_0)| \leq h$$

при $i, j \in \{1, \dots, m\}$, $0 \leq s \leq s_p$.

При этом найдутся такие числа r и β , что $\|X(t) - X_0\| + |s(t) - s_0| \leq \leq re^{-\beta t}$ на множестве Π .

Доказательство. а) Пусть $C = 0, c_{m+1} \neq 0$. Тогда на гиперплоскости скольжения $s = \text{const}$ и устойчивость следует из леммы 3.3.1.3.

б) Пусть $C = \alpha P(s_0) G$ – собственный вектор матрицы $(AP(s_0))^*$, отвечающий $\lambda(AP(s_0))$. Из системы (1.2.1), воспользовавшись полученным при доказательстве предложения 3.4.1 результатом $\int_0^t v dt \rightarrow \infty$, имеем $s \rightarrow s_p - G^* A^{-1} X$ при $t \rightarrow \infty$. Отсюда $s \rightarrow s_p - G^* P(s_0) X / \lambda(AP(s_0))$ при $t \rightarrow \infty$. Поскольку $G^* P(s_0) X = \text{const}$ на гиперплоскости скольжения, s стремится к положительной константе. Использование леммы 3.3.1.3 и леммы сравнения [21] завершает доказательство.

в) Не умаляя общности, будем считать, что $\lambda(A) = 1, P(s_0) = p(s_0) I$, где $p(s_0)$ – положительный скаляр. В этом случае условия теоремы принимают вид

$$\|P(s) - p(s) I\| \leq \epsilon \|s - s_0\|, \quad (3.4.7)$$

$$\|P(s) - p(s) I\| \leq h \quad (3.4.8)$$

при $0 \leq s \leq s_p$.

Скалярная функция $p(s)$ такова: $p(0) = 0, p > 0$ при $s > 0$ и $p'_s > 0$. Сделаем в системе (1.2.1) с управлением (3.4.3) замену переменных $x_{m+1} = s_p - s$ и $\underline{X} = \begin{pmatrix} X \\ x_{m+1} \end{pmatrix}$

и обозначим $\underline{C}^* = (C^*, c_{m+1})$. Тогда получаем

$$\dot{\underline{X}} = [pB - (\underline{C}^* B \Phi \underline{X}) I] \underline{X} + B \Phi_1 \underline{X},$$

где

$$B = \begin{bmatrix} A & 0 \\ G^* & 0 \end{bmatrix}, \quad \Phi = \begin{bmatrix} P & 0 \\ 0 & \dots & 0 \end{bmatrix}, \quad \Phi_1 = \begin{bmatrix} (P - pI) & 0 \\ \vdots & \vdots \\ 0 & \dots & 0 \end{bmatrix}.$$

Сделаем в этой системе замену переменных $\underline{X} = Q(\underline{Y} + E)$, где Q – линейное преобразование, приводящее матрицу B к жордановой форме, $\underline{Y} = \begin{pmatrix} \eta \\ Y \end{pmatrix}$, η – скаляр, Y – столбец размерности m и $E = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$. Не умаляя общности, можно считать, что первый столбец Q совпадает с вектором $\begin{pmatrix} X_0 \\ x_{m+1}^0 \end{pmatrix}$, где $x_{m+1}^0 = s_p - s_0$ и X_0, s_0 – стационарное решение системы (2.1.14), поскольку

$$B \begin{pmatrix} X_0 \\ x_{m+1}^0 \end{pmatrix} = \begin{pmatrix} X_0 \\ x_{m+1}^0 \end{pmatrix}.$$

Имеем

$$\dot{\underline{Y}} = (pB - (\underline{C}^* B \Phi \underline{X}) I)(\underline{Y} + E) + \underline{\Psi}(\underline{Y} + E),$$

где $B = \begin{bmatrix} 1 & 0 & \dots \\ 0 & & \\ \vdots & & D \end{bmatrix}_{(m+1) \times (m+1)}$, $\underline{\Psi} = Q^{-1} B \Phi_1 Q = \begin{bmatrix} \psi_1 \\ \Psi \end{bmatrix}$, ψ_1 – первая строка,

Ψ – матрица размерности $m \times (m+1)$.

В новых переменных нетривиальное стационарное решение совпадает с точкой $\eta = 0, Y = 0$. Поскольку $\underline{C}^* \begin{pmatrix} X_0 \\ x_{m+1}^0 \end{pmatrix} \neq 0$, уравнение инвариантной гиперплоскости, по которой происходит движение, разрешимо относительно η . В итоге получаем

уравнение движения, ограниченного на гиперплоскость скольжения, в виде

$$\begin{aligned} \eta &= R(Y), \\ \dot{Y} &= p[D - I]Y + \Psi Y + \Psi E + (p - \underline{C}^* B \Phi \underline{X})Y, \end{aligned} \quad (3.4.9)$$

где R – некоторый линейный функционал. Поскольку собственное значение матрицы B с максимальной вещественной частью равно 1, матрица $D - I$ гурвицева. Пусть v и w – пара положительно определенных квадратичных форм с матрицами V и W , удовлетворяющих соотношению $(D - I)V + V(D^* - I) = -W$. Дифференцируя функцию v в силу системы (3.4.9), имеем

$$\dot{v} = -pw + 2(p - \underline{C}^* B \Phi \underline{X})v + \left\langle \frac{\partial v}{\partial Y}, \Psi Y + \Psi E \right\rangle.$$

Воспользовавшись предложением 3.4.1, ограничимся рассмотрением этого соотношения на инвариантном множестве Π , где $s > \alpha > 0$ и, следовательно, $p > p_\alpha > 0$.

Величины $\underline{C}^* B \Phi Q Y v$ и $\left\langle \frac{\partial v}{\partial Y}, \Psi Y \right\rangle$ имеют порядок малости $O(\|Y\|^3)$ и могут быть сделаны малыми по сравнению с $p_{\alpha w}$ соответствующим выбором константы h в условии (3.4.8). Величины $(p - \underline{C}^* B \Phi Q E)v$ и $\left\langle \frac{\partial v}{\partial Y}, \Psi E \right\rangle$ имеют порядок $O(\|Y\|^2)$ и могут быть сделаны малыми по сравнению с $p_{\alpha w}$ соответствующим выбором константы ϵ в формуле (3.4.9). Таким образом, $v < 0$ и теорема доказана.

Обсудим полученные результаты. Если измеряемая функция σ удовлетворяет условию (3.4.5), то в каждой точке разрыва управления, кроме тривиальных $X=0$ и $s=0$, выполняется условие скольжения. При этом скользящий режим обеспечивает постоянство плотности биомассы и устойчив к произвольным изменениям функции $P(s)$ во времени с единственным ограничением $p_i > 0$ ($i = 1, \dots, m$) при $s > 0$. Величины $1/p_i$ ($i = 1, \dots, m$) имеют смысл продолжительностей фаз развития клеток данной популяции и являются наиболее чувствительными к внешним возмущениям характеристиками биотехнической системы. Таким образом, использование системы управления переменной структуры позволяет обеспечить высокую надежность работы культивационной установки.

При выполнении условий теоремы доказано существование мажорирующей экспоненты. Следовательно, результат об устойчивости стационарного режима, полученный методом эквивалентного управления, переносится на реальную систему [23].

Недостатком рассматриваемого метода является низкая устойчивость скользящего режима при малых концентрациях субстрата, так как $\dot{\sigma} \rightarrow 0$ при $s \rightarrow 0$. Поэтому метод неприменим при культивировании с малыми концентрациями субстрата.

Пример. Сформулируем условия, при которых скользящий режим культивирования делящихся и почкующихся клеток практически реализуем. Если клетки в популяции в течение своего жизненного цикла проходят m фаз развития и в последней фазе делятся, оставляя q дочерних, то в соответствии с формулами (1.2.2) и (3.4.6) измеряемый показатель $\sigma =$

$$\sigma = \sum_{i=1}^m c_i x_i \text{ должен удовлетворять условиям}$$

$$c_1 > c_2 > \dots > c_{m-1} \text{ и } qc_1 > c_m.$$

Если культивируется популяция почкующихся клеток, то измеряемый по-

казатель $\sigma = \sum_{i=1}^m c_i X_i$ должен удовлетворять условиям

$$c_2 > c_3 > \dots > c_{m-1} > c_m.$$

При измерении оптической плотности (плотностат) эти условия, очевидно, выполняются, так как более зрелые, имеющие больший размер клетки представляют большее препятствие при прохождении света через суспензию.

Рассмотрим модель системы культивирования, в которой разбавление ведется с постоянной скоростью (хемостат) [15]. Динамика биосинтетического процесса в такой системе задается уравнением (1.2.1), где полагается $v = d = \text{const}$ и правая часть удовлетворяет всем условиям, сформулированным в п. 2.1.5. Известно, что при достаточно больших значениях s функции $\partial p_i / \partial s$ ($i = 1, \dots, m$) выходят на плато насыщения: $\partial p_i / \partial s < \epsilon$ (ϵ — малое положительное число). В результате правая граница спектра матрицы линейного приближения, оставаясь в левой полуплоскости, стремится к нулю при $\epsilon \rightarrow 0$ (см. формулу (2.2.7) и предложение 2.2.3). Уменьшение запаса устойчивости системы на практике приводит к срыву процесса культивирования и вымыванию микроорганизмов из реактора [1, 19]. Поэтому при высоких концентрациях субстрата и при скоростях разбавления, близких к критическим, возникает задача: найти управление, повышающее степень устойчивости процесса культивирования. Покажем, что введение соответствующего скользящего режима позволяет решить эту задачу. Итак, рассматривается система

$$\begin{aligned} \dot{X} &= APX - dX, \\ \dot{s} &= -G^*PX - ds + u. \end{aligned} \quad (3.4.10)$$

Пусть

$$u = \begin{cases} 0 & \text{при } \sigma = C^*X + s - s_* > 0, \\ ds_p & \text{при } \sigma = C^*X + s - s_* \leq 0. \end{cases} \quad (3.4.11)$$

Если в плоскости разрыва возникает скользящий режим, то метод эквивалентного управления приводит к уравнению

$$\dot{X} = AP(s_* - C^*X)X - dX. \quad (3.4.12)$$

Необходимое и достаточное условие существования стационарного решения в скользящем режиме состоит в следующем:

$$0 < d < \lambda(AP(s_*)), \quad s_* < s_p. \quad (3.4.13)$$

Первое неравенство (3.4.13) необходимо и достаточно для существования стационарного решения уравнения (3.4.12) (см. теорему 2.1.12). Покажем, что второе условие гарантирует существование скользящего режима в окрестности стационарной точки. Пусть X_0 — стационарное решение (3.4.12) и $s_0 = s_* - C^*X_0$. В точке (X_0, s_0) условия скольжения выполнены: $\dot{\sigma} = -G^*PX_0 - ds_0 < 0$ при $\sigma > 0$ и $\dot{\sigma} = (s_p - s_*) > 0$ при $\sigma < 0$.

Сформулируем локальные условия устойчивости стационарного решения. Пусть $C = L$, где L — собственный вектор матрицы $(AP(s_0))^*$, отвечающий собственному значению d с максимальной вещественной частью $\lambda(AP(s_0)) = d$, т.е. $L^*AP(s_0) = dL^*$.

Предложение 3.4.3. Если выполнено условие $C=L$, то стационарное решение уравнения (3.4.12) асимптотически устойчиво. Матрица линейного приближения имеет спектр

$$\lambda_1 = -dL^* (\partial \ln P / \partial s)(s_0) X_0 < 0, \quad \lambda_i = \rho_i(AP(s_0)) - d,$$

где $\rho_i(AP(s_0))$ ($i = 2, \dots, m$) – собственные значения матрицы $AP(s_0)$, не равные $\lambda(AP(s_0)) = d$.

Доказательство. Матрица линейного приближения такова:

$$AP(s_0) - dI + A \frac{\partial P}{\partial s} X_0 L^*.$$

Обозначим $D = AP(s_0) - dI$ и $Q = A \frac{\partial P(s_0)}{\partial s} X_0$. Характеристический полином матрицы задается уравнением

$$\det(D - QL^* - \lambda I) = 0,$$

которое эквивалентно следующему (см. доказательство теоремы 2.1.6):

$$\det \left[\begin{array}{c|c} D - \lambda I & Q \\ \hline L^* & 1 \end{array} \right] = 0.$$

Пользуясь формулой окаймленного определителя [13], получаем

$$\det(D - \lambda I)(L^*Q/\lambda - 1) = 0.$$

По построению собственное значение матрицы D с максимальной вещественной частью равно нулю. Отсюда имеем

$$\lambda_1 = -L^*Q = -dP(s_0)^{-1} \frac{\partial P}{\partial s}(s_0) X_0,$$

$$\lambda_i = \rho_i(AP(s_0)) - d \quad (i = 2, \dots, m).$$

Предложение доказано.

Поскольку вектор L^* можно взять достаточно большим по норме, максимальный достижимый запас устойчивости равен

$$d - \max \{ \operatorname{Re} \rho_i(AP(s_0)) \mid i = 2, \dots, m \}.$$

Таким образом, при специальном выборе плоскости скольжения в культивационной системе реализуется локально устойчивый стационарный режим с конечным запасом устойчивости.

Пусть теперь C – произвольный положительный вектор. Рассмотрим вопрос об устойчивости в целом скользящего режима в положительном конусе $\{X \geq 0, X \neq 0, s \geq 0\}$. Для устойчивости требуется, чтобы

$$\dot{\sigma} = C^*APX - G^*PX - d(C^*X + s) + ds_p > 0 \quad \text{при}$$

$$C^*X + s \leq s_*, \tag{3.4.14}$$

$$\dot{\sigma} = C^*APX - G^*PX - d(C^*X + s) < 0 \quad \text{при}$$

$$C^*X + s > s_*. \tag{3.4.15}$$

Заданием достаточно большой разности $s_p - s$ можно добиться выполнения условия (3.4.14). Условие (3.4.15) выполняется тогда и только тогда, когда

$$(C^*A - G^*)P(s) - dC^* < 0 \tag{3.4.16}$$

при $C^* X + s > s_*$ и $s \leq s_p$. Если строка $C^* A - G^*$ имеет отрицательную компоненту, то для этой компоненты неравенство (3.4.16) заведомо выполняется. Поэтому для выполнения (3.4.16) достаточно потребовать, чтобы

$$(C^* A - G^*)P(s_p) - dC^* < 0. \quad (3.4.17)$$

Рассмотрим подробнее полученное условие. Умножив неравенство (3.4.17) справа на собственный вектор H матрицы $AP(s_p)$, отвечающий собственному значению $\lambda(AP(s_p))$, и воспользовавшись условием стационарности $\lambda(AP(s_0)) = d$, имеем необходимое ограничение на C^* :

$$C^* H < G^* P(s_p) H / (\lambda(AP(s_p)) - \lambda(AP(s_0))).$$

Отсюда видно, что верхняя граница для C^* тем больше, чем меньше изменяется асимптотическая скорость воспроизводства в диапазоне $s_0 \leq s \leq s_p$.

Если $C = L$, то неравенство (3.4.17) принимает вид

$$0 < G,$$

$$l_i < g_i p_i(s_p) p_i(s_0) / (d(p_i(s_p) - p_i(s_0))),$$

где $L^* = (l_1, \dots, l_m)$, $G^* = (g_1, \dots, g_m)$ и $i \in \{1, \dots, m \mid p_i(s_p) \neq p_i(s_0)\}$. При выполнении этого условия, соответствующем выбору величин s_p и s_0 и в силу теоремы 2.1.14, где в пункте в) вектор G следует заменить на C , стационарное решение асимптотически устойчиво на всем множестве $\{X \geq 0, X \neq 0, s \geq 0\}$. Другие условия устойчивости стационарного решения уравнения (3.4.12) при устойчивом скользящем режиме дает теорема 2.1.14, где вектор G следует заменить на C .

Итак, найдены условия, гарантирующие существование скользящего режима, нечувствительного к таким изменениям функции $P(s)$, которые сохраняют ее основные свойства. Управление (3.4.11) может быть реализовано в случаях, когда управление (3.4.1) не удастся реализовать, так как условия (3.4.4), (3.4.6) и (3.4.17) существенно дополняют друг друга.

§ 3.5. Системы управления клеточными популяциями.

Чепрямое управление

В устройствах непрямого управления, используемых для стабилизации стационарного режима в культивационной системе, в отличие от устройств с непосредственным измерением численности популяции, плотность биомассы влияет на производную контролируемого параметра. Скорость разбавления задается в зависимости от этого параметра. Устройство непрямого управления будем рассматривать на примере оксигатора, когда измеряется концентрация растворенного кислорода в среде. На рис. 3.2 схематически представлены основные элементы такой системы. В ферментере 1 находится раствор культуры. Измерительный элемент 4 определяет концентрацию кислорода в среде. Дозирующее устройство 5, связанное с контрольным клапаном, измеряет скорость подачи субстрата в ферментер.

Обозначим через z концентрацию кислорода в среде, через z_p — концентрацию насыщения кислорода, через $q(X, s, z)$ — скорость потребления кислорода клетками, а через b — скорость растворения кислорода. Динамика

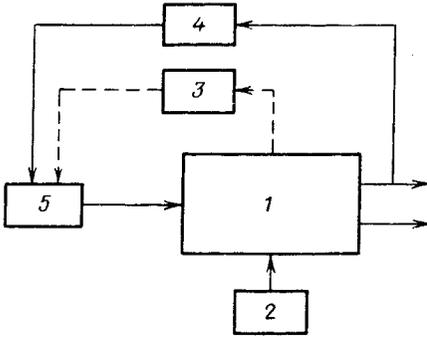


Рис. 3.2. Система непрямого управления для стабилизации численности непрерывно культивируемой клеточной популяции: 1 – реактор; 2 – система подачи кислорода; 3 – датчик концентрации субстрата; 4 – датчик концентрации растворенного кислорода; 5 – дозатор

биосинтетического процесса, протекающего в замкнутом объеме, задается системой

$$\begin{aligned} \dot{x} &= p(s, z)AX - (d + w)X, \\ \dot{s} &= -p(s, z)G^*X - (d + w)(s - s_p), \\ \dot{z} &= -q(x, s, z) - b(z - z_p). \end{aligned} \quad (3.5.1)$$

Здесь d – установленная скорость разбавления, $d + w$ – фактическая скорость разбавления (w – управление), A – неразложимая матрица с неотрицательными диагональными элементами, описывающая структуру размножения клеток культивируемой популяции, $p(s, z)$ – удельная скорость роста биомассы. Если скалярная функция $p = p(s)$, т.е. не зависит от z , первые два уравнения системы (3.5.1) являются частным случаем системы (1.2.1).

Напомним естественные с содержательной точки зрения свойства функций $p(s, z)$ и $q(x, s, z)$. При увеличении концентрации питательного субстрата и концентрации растворенного кислорода в среде скорость роста не убывает. Поэтому будем считать, что $p > 0$ при $sz > 0$, $p'_s \geq 0$ и $p'_z \geq 0$. В отличие от п. 2.1.5, допустимыми считаются такие режимы культивирования, при которых удельная скорость роста не зависит от концентрации лимитирующего субстрата ($p'_s = 0$). При нулевой концентрации субстрата он, очевидно, не потребляется, поэтому $p(0, z) = 0$. Кислород не потребляется при отсутствии кислорода, отсутствии окисляемого субстрата и отсутствии клеток: $q(X, s, 0) = q(X, 0, z) = q(0, s, z) = 0$ и $q > 0$ при $szX > 0$. Кроме того, его потребление не убывает с ростом X, s и z :

$$q'_{x_i} \geq 0 \quad (i = 1, \dots, m), \quad q'_s \geq 0 \quad \text{и} \quad q'_z \geq 0.$$

Будем рассматривать динамику невырождающейся популяции и поэтому будем считать, что $\lambda(A) > 0$. При записи (3.5.1) это предположение эквивалентно условию $\lambda(A) = 1$, которое будем считать выполненным.

Так как величины X, s, z по смыслу могут принимать лишь неотрицательные значения, систему (3.5.1) в дальнейшем будем изучать в положительном конусе $\{X \geq 0, s \geq 0, z \geq 0\}$. Выясним, когда рассматриваемая система автоматического регулирования имеет стационарный режим работы.

Теорема 3.5.1. Для того чтобы система (3.5.1) имела нетривиальное стационарное решение $X_0 \neq 0, s_0, z_0$, достаточно выполнения неравенства $p(s_p, z_p) > d$. Если $p'_s(s, z) \neq 0$ при некоторых $s_0 < s \leq s_p$ и $z_0 \leq z \leq z_p$, то условие необходимо.

Доказательство. Достаточность. Делаем замену переменной $s_p - s = x_{m+1}$. Если $s \leq s_p$, то $x_{m+1} \geq 0$. Легко видеть, что множество $s \leq s_p$ является положительно инвариантным для системы (3.5.1); при достаточно большом t это условие выполняется для любого нетривиального решения системы. Имеем

$$\begin{aligned} \dot{Y} &= p(Y, z)BY - (d + u)Y, \\ \dot{z} &= -q(Y, z) - b(z - z_p), \end{aligned} \quad (3.5.2)$$

где $Y = (X^*, x_{m+1})^*$, $p(Y, z) = p(X, s_p - x_{m+1}, z)$,

$$q(Y, z) = q(X, s_p - x_{m+1}, z), \quad B = \begin{bmatrix} A & 0 \\ G^* & 0 \end{bmatrix}.$$

Систему (3.5.2) будем изучать в конусе $\{Y \geq 0, z \geq 0\}$, являющимся положительно инвариантным множеством данной системы. "Максимальное" собственное значение матрицы B , как и прежде, $\lambda(B) = 1$; ему отвечает собственный вектор $Y_* = (X_*^*, g)^*$, где $X_*^* > 0$ — собственный вектор матрицы A , отвечающий $\lambda(A)$, $g = G^*X_*$. Ищем стационарное решение построенной системы в виде $Y_0 = x_0 Y_*$, где x_0 — скаляр, удовлетворяющий системе

$$\begin{aligned} F(x, z) &= p(s_p - gx, z) - d = 0, \\ \Phi(x, z) &= q(xX_*, s_p - gx, z) - bz - bz_p = 0. \end{aligned} \quad (3.5.3)$$

На основании сделанных предположений заключаем, что $\Phi'_z > 0$, $\Phi(x, 0) < 0$ и $\Phi(x, z_p) \geq 0$. Поэтому существует однозначная функция $z = z(x)$, определенная при $(s_p/g) \geq x \geq 0$. Из (3.5.2) вытекает, что $z' = z_p$ при $x = 0$ и, следовательно, $F(x, z(x))|_{x=0} > 0$. С другой стороны, $F(x, z(x))|_{x=s_p/g} = -d$. Таким образом, стационарная точка x_0 существует. Стационарное решение имеет вид

$$X_0 = x_0 X_*, \quad s_0 = s_p - gx_0, \quad z_0 = z(x_0),$$

где $p(s_0, z_0) = d$ и $q(x_0 X_*, s_p - gx_0, z_0) + bz_0 - bz_p = 0$.

Необходимость. Доказательство очевидно, так как в противном случае "максимальное" собственное значение матрицы $p(s, z)B$ меньше d при всех $0 \leq s \leq s_0$ и $0 \leq z \leq z_0$. Теорема доказана.

Асимптотическую устойчивость стационарного решения на всем положительном конусе обеспечивает управление переменной структуры [23], которое имеет вид произвольной функции, удовлетворяющей условию

$$\begin{aligned} w(z, s) &> p(s_0, z) - d \quad \text{при } (z - z_0)k(s) > 0, \\ -d &\leq w(z, s) < p(s_0, z) - d \quad \text{при } (z - z_0)k(s) < 0, \end{aligned} \quad (3.5.4)$$

где $k(s) = (s - s_0)[q(X_0(s_p - s)/(s_p - s_0), s, z_0) - q(X_0, s_0, z_0)]$.

Отметим, что при $p'_z > 0$ и $k(s) < 0$ нулевое управление удовлетворяет сформулированному условию. Стационарное решение в этом случае является устойчивым в силу свойств системы. В дальнейших рассуждениях будем рассматривать несколько более общий закон управления

$$\begin{aligned} w(z, s) &= p(s_0, z) - d + \frac{1 + \text{sign } k(s)}{2} \varphi(z) + \\ &+ \frac{1 - \text{sign } k(s)}{2} \psi(z), \end{aligned} \quad (3.5.5)$$

где $\varphi(z)$ и $\psi(z)$ — произвольные кусочно-непрерывные функции, удовлетворяющие условиям

$$\begin{aligned} \varphi(z)(z - z_0) &> 0 \quad \text{и} \quad p(s_0, z) + \varphi(z) \geq 0, \\ \psi(z)(z - z_0) &< 0 \quad \text{и} \quad p(s_0, z) - \psi(z) \geq 0. \end{aligned}$$

Если $k(s) = 0$, то управление не определено. В то же время предположение $k(s) \neq 0$ при $s \neq s_0$ недопустимо с содержательной точки зрения, поскольку даже в простейшем случае $q(X, s, z_0) = \sum_{i=1}^m \mu_i s x_i$, где μ_i ($i = 1, \dots, m$) — положительные константы, это условие не выполнено.

Рассмотрим подробнее смысл равенства $k(s) = 0$. Уравнение (3.5.1), определенное в R_+^{m+2} , имеет инвариантное двумерное множество $X = xX_0$, $s = s_p - gx$, $z \geq 0$, где $g = G^*X_0$ и $0 \leq x \leq s_p/g$. На этом инвариантном множестве $k(s) = k(s_p - gx)$, и если $k(s) = 0$ на некотором интервале $s_1 \leq s \leq s_2$, то $q(xX_0, s, z) = q(X_0, s, z)$ при $s_1 \leq s \leq s_2$ и, следовательно, последнее уравнение системы (3.5.1) оказывается независимым от первых двух в некотором диапазоне изменения переменных. Это означает, что система стала ненаблюдаемой. Поэтому предположим, что

$$k(s) = 0 \quad (3.5.6)$$

лишь в изолированных точках. Кроме того, потребуем, чтобы точки вырождения функции $k(s)$ (за исключением стационарной s_0) не являлись инвариантными множествами системы (3.5.1). Как будет видно из дальнейшего, для этого достаточно предположить, что

$$\partial p(s, z)/\partial s \neq 0 \quad \text{при } k(s) = 0. \quad (3.5.7)$$

Правая часть системы (3.5.1) при управлении (3.5.5) является разрывной, и, как будет показано в дальнейшем, не во всех точках разрыва выполнено условие правосторонней единственности решений [24]. В дальнейших рассуждениях мы будем рассматривать все решения и доказывать соответствующие утверждения для всего множества возможных решений.

Теорема 3.5.2. При выполнении условий глобальной наблюдаемости (3.5.6) и (3.5.7) система (3.5.1) при управлении (3.5.4) устойчива в целом на множестве $\{X \neq 0, 0 \leq s \leq s_p, 0 \leq z \leq z_p\}$.

Доказательство. Из вида (3.5.1) следует, что существует инвариантное двумерное множество вида

$$\{xX_0, s = s_p - gx, 0 \leq x \leq s_p/g, 0 \leq z \leq z_p\}, \text{ где } g = G^*X_0.$$

Вначале докажем устойчивость стационарного решения системы, ограниченной на это инвариантное множество. Из (3.5.1) имеем

$$\begin{aligned} \dot{x} &= (p(x, z) - d)x - ux, \\ \dot{z} &= -q(x, z) - b(z - z_p), \end{aligned} \quad (3.5.8)$$

где $p(x, z) = p(xX_0, s_p - gx, z)$, $q(x, z) = q(xX_0, s_p - gx, z)$.

Для дальнейшего нам удобно привести систему (3.5.8) к другим переменным: $x - x_0 = \xi$, $z - z_0 = \eta$. Воспользовавшись тем, что $(p(x_0, z_0) - d)x \equiv 0$, имеем

$$\begin{aligned} \dot{\xi} &= [-s(\xi, \eta) + l(\eta) - u](x_0 + \xi), \\ \dot{\eta} &= -[r(\xi, \eta) + b\eta] - h(\xi), \end{aligned} \quad (3.5.9)$$

где $-s = p(x, z) - p(x_0, z)$, $l = p(x_0, z) - p(x_0, z_0)$,

$$r = q(x, z) - q(x, z_0) \text{ и } h = q(x, z_0) - q(x_0, z_0).$$

Легко убедиться в том, что $s\xi \geq 0$, $l\eta \geq 0$ и $r\eta \geq 0$ в силу свойств функций p и q . Оказывается, что решения системы (3.5.1) асимптотически приближаются к решениям системы (3.5.8).

Лемма 3.5.2.1. При любом нетривиальном начальном условии имеет место следующее асимптотическое свойство системы (3.5.2):

$$Y(t) = x(t) Y_* + \epsilon(t), \quad (3.5.10)$$

где $x(t)$ — скалярная функция времени, а $\|\epsilon(t)\|/x(t) \rightarrow 0$ при $t \rightarrow \infty$.

Доказательство. Так как "максимальное" $\lambda(B) = 1$, достаточно показать, что $\int_0^t P(x, z) dt \rightarrow \infty$ вдоль любого нетривиального решения системы (3.5.1). Допустим противное: $\int_0^t P dt \neq \infty$. Выпишем решение системы (3.5.2) в виде

$$Y(t) = [\alpha \exp\{\int_0^t P dt\} Y_* + o(\exp\{\int_0^t P dt\})] \exp\{-td - \int_0^t u dt\}, \quad (3.5.11)$$

где $\alpha = Q^* X(0)$, Q — биортогональный X_0 собственный вектор матрицы A^* (собственный вектор матрицы B^* , отвечающий $\lambda = 1$, очевидно, равен $(Q^*, 0)^*$). Из допущения следует, что либо $Y(t) \rightarrow 0$, либо $\int_0^t u dt + td < \infty$ при $t \rightarrow \infty$. Если $Y(t) \rightarrow 0$ ($x(t) \rightarrow 0$, $s(t) \rightarrow s_p$), то по свойству функции $p(X, s, z)$ имеем $p \geq c > 0$, так как из последнего уравнения системы (3.5.1) легко видеть, что $z \geq \alpha > 0$, начиная с некоторого t . Получили противоречие с исходным допущением. Значит, $Y(t) \neq 0$, а $\int_0^t u dt + td < \infty$.

Если $\int_0^t p dt < \infty$, то $p \rightarrow 0$, и поэтому в силу свойств функции $p(X, s, z)$ ($z > \alpha$) имеем либо $s \rightarrow 0$, либо $z \rightarrow 0$ и $p(X, s, 0) = 0$. Пусть $s \rightarrow 0$. Тогда из последнего уравнения системы (3.5.1) получаем, что $z \rightarrow z_p$. Из (3.5.4) следует, что $k(s=0) = q(X_0, s_0, z_0) s_0 > 0$, и поэтому получаем $u|_{s=0, z=z_p} > 0$. Противоречие доказывает, что $\int_0^t p dt \rightarrow \infty$. Обозначим $x(t) = \alpha \exp(\int_0^t p dt - td - \int_0^t u dt)$,

$$\epsilon(t) = o(\exp\int_0^t p dt - td - \int_0^t u dt).$$

Очевидно, $\|\epsilon(t)\|/x(t) \rightarrow 0$. Лемма доказана.

Если теперь удастся установить асимптотическую устойчивость решений системы (3.5.8), то устойчивость решений исходной системы доказывается следующим образом. Для любого решения системы (3.5.2) в силу (3.5.10) находится решение (3.5.8), которое в некоторый момент времени достаточно близко к нему. Затем, пользуясь непрерывной зависимостью решений от начальных условий, результат переносим на общий случай. При этом оказывается возможным проводить анализ системы (3.5.2) на ограниченном замкнутом асимптотически инвариантном множестве, не содержащем тривиальной особой точки $X = 0, s = s_p, z = z_p$. Существование этого множества дает следующая

Лемма 3.5.2.2. Начиная с некоторого t , каждое нетривиальное решение системы (3.5.2), начинающееся в положительном конусе, попадает в некоторую окрестность стационарной точки: при этом для каждого решения существует такое t^* , что $x(t) \geq \beta > 0$ в формуле (3.5.10) и $z(t) \geq \alpha > 0$ при $t > t^*$.

Доказательство. В доказательстве используется функция Ляпунова, на которую в основном опирается результат теоремы. Так как $e^A \geq 0$, то, очевидно, в формуле (3.5.10) $x(t) \geq 0$. Из системы (3.5.1), воспользовавшись формулой (3.5.10), легко получить, начиная с некоторого t , что выполняется неравенство $x(t) \leq s_p/g$. Из последнего уравнения системы (3.5.1) видно, что $z \leq z_p$ при больших t . Мы уже отмечали, что $z \geq \alpha$ (см. доказательство леммы 3.5.2.1). Нам осталось доказать, что $x(t) \geq \beta$. Докажем этот факт для системы (3.5.8). Управление для системы (3.5.8) принимает вид

$$u = l(\eta) + \frac{1 - \text{sign } k(\xi)}{2} \varphi(\eta) + \frac{1 + \text{sign } k(\xi)}{2} \psi(\eta), \quad (3.5.12)$$

где $\varphi(\eta) = \varphi(z_0 + \eta)$ и $\varphi\eta > 0$, $\psi(\eta) = \psi(z_0 + \eta)$ и $\psi\eta < 0$ при $\eta \neq 0$, а $k(\xi) = \xi h(\xi) = -k(s)$.

Введем функцию Ляпунова

$$v = \int_0^{\xi} \frac{h(\xi) \operatorname{sign} k(\xi)}{\xi + x_0} d\xi + \operatorname{sign} k(\xi) \int_0^{\eta} [l - u] d\eta, \quad (3.5.13)$$

определенную при $\xi > -x_0$ и $\eta > -z_0$, так как $x > 0$ и $z > 0$. В силу системы (3.5.9) имеем

$$\dot{v} = -s(\xi, \eta) h(\xi) \operatorname{sign}(\xi, h) - [r + b\eta] \left[\frac{1 - \operatorname{sign} k(\xi)}{2} \varphi(\eta) - \frac{1 + \operatorname{sign} k(\xi)}{2} \psi(\eta) \right] \leq 0. \quad (3.5.14)$$

В окрестности $\xi = -x_0$ получаем $k(\xi = -x_0) = -q(x_0, z_0) \neq 0$, и переключений нет. Из формул (3.5.14) и (3.5.12) и свойств функции $p(x, z)$ следует, что $\dot{v} < 0$ при $\xi = -x_0$. С другой стороны, из (3.5.13) имеем, что $v \rightarrow \infty$ при $\xi \rightarrow -x_0$. Функция $v(\xi, \eta)$ непрерывна в окрестности точки $\xi = -x_0$. Поэтому для системы (3.5.9) верно $\xi > -x_0 + \beta$ и, следовательно, для $x \geq \beta > 0$.

Так как в окрестности $\xi = -x_0$ ($x = 0$) нет переключений управления, то выполняется интегральная непрерывность решений. Из вида системы (3.5.2) следует, что ее решения не обращаются в нуль за конечное время (исключая тривиальное решение). Поэтому, предположив, что $Y(t)$ может стать сколь угодно малым при больших t , выбрав достаточно близкое к нему решение системы (3.5.8) (это можно сделать в силу (3.5.10)) и воспользовавшись доказанным свойством системы (3.5.8), мы приходим к противоречию. Отсюда $x(t) \geq \beta$ в формуле $Y(t) = x(t)Y_* + \epsilon(t)$. Лемма доказана.

Займемся теперь исследованием системы (3.5.9).

Лемма 3.5.2.3. Если выполнены условия глобальной наблюдаемости (3.5.6), (3.5.7), то существует ограниченное положительное инвариантное множество, содержащее стационарную точку, в котором решения системы (3.5.9) определены единственным образом и непрерывно зависят от начальных условий. Функция $k(\xi)$ на этом множестве не меняет знак, хотя может обращаться в нуль. За конечное время все решения системы (3.5.9) попадают в данное множество.

Доказательство. а) Рассмотрим случай, когда $\xi = 0$ не является точкой переключения. Пусть для определенности имеет место следующая картина: $\xi h(\xi) > 0$ при $\xi < \xi^*$ и $\xi h(\xi) < 0$ при $\xi > \xi^*$ и $\xi^* > 0$. Тогда $\dot{\xi} = (-s - \psi(\eta))x$ в области $\xi < \xi^*$ и $\dot{\xi} = (-s - \varphi(\eta))x$ в области $\xi > \xi^*$. Движения, начавшиеся слева от $\xi = \xi^*$, очевидно, не выходят за линию переключения, так как при $\xi < \xi^*$ и $\eta < 0$ имеем $\dot{\xi} < 0$, а при $\xi > \xi^*$ и $\eta > 0$ имеем $\dot{\xi} > 0$. На линии переключения могут возникать скользящие режимы и неоднозначность. Если доопределить решения на линии переключения по А.Ф. Филипову [24], то в инвариантном множестве будет выполняться интегральная непрерывность движений и правосторонняя единственность.

Покажем теперь, что всякое решение системы (3.5.9) за конечное время попадает в область, лежащую слева от линии переключения. Для этого достаточно показать, что всякое решение, начинающееся справа от $\xi = \xi^*$, за конечное время θ пересекает линию переключения, т.е. $\xi(\theta - 0) = \xi^*$ и $\dot{\xi}(\theta + \tau) > 0$ при $\tau > 0$.

Допустим противное. Определим новую систему уравнений, заменив в системе (3.5.9) функцию $h(\xi)$ функцией

$$h^* = \begin{cases} h(\xi) & \text{при } \xi > \xi^*, \\ -\frac{1}{M} h(\xi) & \text{при } \xi < \xi^*. \end{cases}$$

При достаточно большой константе M величина η во введенной системе не выходит из области определения $\eta \in [-z_0, z_0 + z_p]$ (см. формулу (3.5.9)). В силу соотношений (3.5.13), (3.5.14) получаем, что положительно определенная функция v на решениях введенной системы непрерывна и имеет неположительную производную. Поэтому $\xi > -x_0$ ($v \rightarrow \infty$ при $\xi \rightarrow -x_0$) и $\xi < -x_0 + s_p/g$, так как при $\xi > \xi^*$ система совпадает

с прежней. К тому же везде, кроме точки $\xi = \xi^*$, имеем $\xi h^* < 0$. Поэтому управление u и функция v для вновь построенной системы непрерывны.

Итак, решения построенной системы устойчивы по Лагранжу и непрерывно зависят от начальных условий. Кроме того, положительно определенная функция v (3.5.13) в силу системы имеет неположительную производную. Множество $\dot{v} = 0$ при выполнении (3.5.7) не содержит целых траекторий. Отсюда стационарная точка $\xi = 0$, $\eta = 0$ для построенной системы является асимптотически устойчивой в целом по Ляпунову [2]. Следовательно, любое решение пересекает ось $\xi = \xi^*$. Решение системы (3.5.9), не попавшее в область $\xi > \xi^*$, удовлетворяет построенной системе. Доопределим функцию v на оси $\xi = \xi^*$ по непрерывности. Отсюда существует θ , при котором решение пересечет ось $\xi = \xi^*$ и $\xi(\theta + \tau) > 0$ при $\tau > 0$.

б) Если $\xi = 0$ является линией переключения, то полуось $\eta > 0$, $\xi = 0$ является множеством скольжения, полуось $\eta < 0$, $\xi = 0$ — множеством отталкивания и здесь решения определены неоднозначно. Решения системы (3.5.9) не могут пересекать линию переключения. В каждой из инвариантных областей имеет место интегральная непрерывность решений и правосторонняя единственность.

Замечание. Пусть число точек перемены знака функции $k(\xi)$ ($h(\xi) = 0$) больше единицы. Обозначим их через $\xi_1 < \xi_2 < \dots < \xi_k$. Проводя рассуждения, аналогичные предыдущим, для самой отдаленной от нуля точки ξ_k , а затем для точки ξ_{k-1} и т.д., можно доказать, что решение попадает в инвариантное множество, содержащее нуль. Лемма доказана.

Теперь мы можем доказать устойчивость системы (3.5.9).

Лемма 3.5.2.4. Если выполнены условия леммы 3.5.2.3, то стационарное решение системы (3.5.9) при управлении (3.5.12) асимптотически устойчиво в целом в положительном квадранте.

Доказательство. Все решения системы (3.5.9) устойчивы по Лагранжу в положительном квадранте в силу леммы 3.5.2.2. С некоторого момента времени они попадают в инвариантное множество, где выполняется свойство правосторонней единственности и интегральной непрерывности решения и почти во всех точках множества $k(\xi) = 0$. Доопределим функцию v на линии переключения по непрерывности. На инвариантном множестве $\dot{v} \leq 0$, так как в скользящем режиме $\xi = \xi^*$ и поэтому $\xi = 0$ и $\dot{\eta}/\eta < 0$, а в остальных точках верна формула (3.5.14). Принимая $\xi = \xi^*$ в силу (3.5.13), (3.5.14) не содержит ω -пределных точек решений. Отсюда следует устойчивость по Ляпунову. Лемма доказана.

Изучаемая система имеет разрывное управление, и поэтому непрерывная зависимость решений от начальных условий не выполняется. Однако имеет место следующее свойство.

Лемма 3.5.2.5. Любое нетривиальное решение системы (3.5.1) таково, что $s(t)$, начиная с некоторого t , попадает в область, где $k(s)$ (см. (3.5.4)) не меняет знак или меняет знак лишь в точках $s = s_0$ и не выходит оттуда. Если решение $Y(t)$ системы (3.5.2) таково, что $s = s_p - x_{m+1}$ не выходит из области постоянства знака $k(s)$, то по любым ϵ и T найдется сколь угодно большой момент времени t^* и $\delta > 0$ такое, что $\|Y(t) - x(t) Y_*\| < \epsilon$ при $t \in [t^*, t^* + T]$, если $\|Y(t^*) - x_* Y_*\| < \delta$, $|z(t^*) - z_*| < \delta$, где $x(t)$ — решение системы (3.5.8) с начальными условиями $x(t^*) = x_*$, $z(t^*) = z_*$.

Доказательство. Докажем первую часть леммы.

а) Пусть $s = s_0$ не является линией переключения. Допустим противное. Заменим систему (3.5.1) так же, как это делалось в лемме 3.5.2.3, системой, у которой нет переключения управления. Далее, пользуясь интегральной непрерывностью и формулой (3.5.9), легко показать, что решения построенной системы попадают в сколь угодно малую окрестность стационарной точки. Следовательно, решения системы (3.5.1) пересекают ось перемены знака $k(s)$.

При достаточно больших t они не покидают множество постоянства знака $k(s)$. Введем оси координат ξ' , η' и определим на решениях системы (3.5.2) функции $\xi' = x_{m+1}/g - x_0$, $\eta' = z - z_0$. Линия переключения имеет ту же координату, что и у системы (3.5.9). Покажем, что поведение функции (ξ', η') и решений системы (3.5.9) в окрестности линии переключения эквивалентно. Действительно, при $\eta < 0$ у системы (3.5.9) в окрестности прямой $\xi = \xi^*$ ($\xi < \xi^*$) выполняется строго неравенство $\dot{\xi} < 0$. Если в выражении $\dot{\xi}' = \frac{1}{g} \dot{x}_{m+1}(y, z)$ положить $Y = x Y_*$, то получит-

ся первое уравнение системы. Аналогичное свойство выполняется для η' . Мы имеем $\xi < 0$ строго. Полагая $Y = xY_* + \epsilon$, мы получаем $\xi' < 0$ при малых ϵ . Следовательно, при $\eta' < 0$ величина ξ' не выходит за линию $\xi' = \xi^*$. При $\eta > 0$ справа от линии переключения имеет место строгое неравенство $\xi < 0$. Рассуждая аналогично, получаем, что переменная ξ не может перейти линию переключения слева направо, а может лишь скользить по ней.

б) Случай, когда $s = s_0$ является линией переключения, разбирается особо ниже.

В заключение отметим, что асимптотическая интегральная непрерывность выполняется в положительно инвариантном множестве. Лемма доказана.

Доказанная асимптотическая непрерывность решений системы (3.5.1) позволяет завершить доказательство.

Завершение доказательства теоремы. Достаточно, очевидно, показать, что теорема выполняется в асимптотически инвариантном множестве.

а) $s = s_0$ не является линией переключения. Окружим стационарную точку системы ϵ -окрестностью и найдем $T = \text{Sup } T(\xi, \eta)$ по множеству $\Gamma\{-x_0 + \beta \leq \xi \leq \xi^*, -z_0 + \alpha \leq z < z_p - z_0\}$, где $T(\xi, \eta)$ — время, необходимое для попадания движения системы (3.5.9) $\{\xi(t), \eta(t) \mid \xi(0), \eta(0) \in \Gamma\}$ в ϵ -окрестность нуля, а α и β — константы, найденные в лемме 3.5.2.2, ξ^* — ось перемены знака $h(\xi)$ (см. доказательство леммы 3.5.2.3). Присоединим к инвариантному множеству Γ ось $\xi = \xi^*$ и определим при этом решение так, чтобы сохранилась инвариантность $\Gamma = \Gamma \cup \{\xi = \xi^*\}$.

Очевидно, $T = \text{Sup } \{T(\xi, \eta) \mid \xi, \eta \in \Gamma\}$. В силу асимптотической устойчивости систем (3.5.9) и замкнутости Γ имеем $T < \infty$. Далее, по T и δ найдем такое t^* , что $|\xi'(t) - \xi(t)| < \delta$ и $|\eta'(t) - \eta(t)| < \delta$ при $t \in [t^*, t^* + T]$, если $\xi'(t^*) = \xi(0)$ и $\eta'(t^*) = \eta(0)$ (здесь $(\xi'(t), \eta'(t))$, как и в лемме 3.5.2.5, находятся в силу системы (3.5.2), а ξ, η — решения системы (3.5.9)). Рассмотрим решение системы (3.5.9), начинающееся из точки $(\xi'(t), \eta'(t))$. За время T оно попадает в ϵ -окрестность нуля. В силу асимптотической интегральной непрерывности движение $(\xi'(t), \eta'(t))$ попадает в ранее выбранную сколь угодно малую окрестность стационарной точки. Рассуждая аналогично при $t > t^* + T$ и принимая во внимание лемму 3.5.2.5, получаем, что система (3.5.2) условно устойчива по координатам x_{m+1} и z . Отсюда с учетом (3.5.10) на множестве $\{Y \geq 0, Y \neq 0, z \geq 0\}$ получаем асимптотическую устойчивость системы (3.5.2).

б) Пусть $k(s)$ меняет знак при $s = s_0$. Предположим, что движение (ξ', η') может находиться вне малой окрестности нуля сколь угодно большое время. Вне этой окрестности неравенство $\xi < 0$ ($\xi > 0$) при $\xi = 0$ является строгим. Следовательно, для движения (ξ', η') вне малой окрестности нуля верна лемма 3.5.2.5, а значит и предыдущие рассуждения теоремы. Поэтому (ξ', η') попадает в сколь угодно малую окрестность нуля и выходит из нее лишь на конечное время. Отсюда в силу асимптотической устойчивости (3.5.9) и асимптотической интегральной непрерывности (3.5.2) вне окрестности нуля следует асимптотическая устойчивость системы (3.5.2). Теорема доказана.

Из доказательства теоремы вытекает следующее очевидное

Замечание. Если технически проще измерять не концентрацию субстрата s , а некоторый показатель концентрации биомассы $\sigma = C^*X$ ($C \geq 0$), то в формуле (3.5.5) функцию $k(s)$ можно заменить на \neq

$k(s_p - g\sigma(X)/\sigma(X_0))$; при этом истинность теоремы сохранится.

Таким образом, построена математическая модель биотехнической системы непрерывного культивирования одноклеточных организмов с непрямым управлением. Модель учитывает динамические характеристики биологического звена системы автоматического регулирования, а также технические характеристики этой системы. Приведено условие существования стационарного режима в культивационной системе. Найдено управление переменной структуры, обеспечивающее асимптотическую устойчивость стационарного режима во всей области задания начальных данных. Из приведенных результатов видно, что существующие системы

культивирования с непрямым измерением (см. рис. 3.2 – сплошная линия), в которых измеряется лишь концентрация кислорода, не могут обеспечивать устойчивость системы при больших возмущениях начальных данных. Такие системы обычно работают при малых скоростях разбавления и, следовательно, при малых концентрациях субстрата (см. формулу (3.5.1)). Благодаря этому добиваются выполнения условия $k(s) > 0$ в окрестности стационарной точки. Из полученных результатов видно, что существующие системы могут работать при высоких скоростях разбавления и высоких концентрациях субстрата, когда $k(s) < 0$. Для этого достаточно поменять знак управления на противоположный. Работу системы в наиболее интересном диапазоне, когда стационарный режим близок к точке перемены знака $k(s)$, обеспечивает лишь система управления переменной структуры.

§ 3.6. Наблюдение переменных состояния экологических систем

Для рациональной эксплуатации экологических систем необходима информация о их состоянии в каждый текущий момент времени. В реальных условиях измерению поддается лишь часть переменных состояния или некоторая функция этих переменных. Так, например, при наблюдении за природной популяцией может быть измерена либо ее суммарная численность, либо число взрослых особей, либо то и другое. Допустим, что популяция имеет ярко выраженную возрастную структуру и число возрастных градаций больше числа измеряемых величин. Возникает вопрос, можно ли, располагая доступной для наблюдения информацией, определить численность каждой возрастной группы в каждый текущий момент времени. На первый взгляд это сделать нельзя. Однако задача оказывается разрешимой, если, помимо результатов измерения, в распоряжении исследователя имеется аналитическая или имитационная модель динамики исследуемого объекта. Перейдем к изложению методов ее решения.

3.6.1. Положительная наблюдаемость. Пусть задана модель экологического объекта. Обозначим через $X(t)$ траекторию системы и через $Y(t) = H(X(t))$ – наблюдаемый сигнал, где $H: R^m \rightarrow R^r$ – некоторая непрерывная функция. Под наблюдаемостью понимается возможность построения такого оператора G , который по функции $Y(t)$, заданной на интервале времени $0 \leq t \leq T$, мог бы восстановить начальное состояние $X(0)$, или конечное состояние $X(T)$, или, так как известен вид модели, траекторию $X(t)$ при $0 \leq t \leq T$.

Если бы доступным для наблюдений был идеальный, неискаженный сигнал, то вид оператора G был бы безразличен, лишь бы он решал поставленную задачу. Однако измерение всегда сопровождается наложением некоторой помехи. В этом случае оператор наблюдения дает не точный ответ, а некоторое приближение реального состояния. Отсюда возникают следующие проблемы:

- 1) Построить оператор наблюдения, который на идеальном сигнале дает точное решение, а на искаженном сигнале – решение с минимальной ошибкой.
- 2) Решение, полученное по искаженному сигналу, должно иметь биологический смысл.

Остановимся на втором условии подробнее. Для этого рассмотрим примеры. Пусть популяция имеет возрастную структуру, а наблюдению поддается лишь общая численность. Задача состоит в том, чтобы восстановить значения численностей всех возрастных групп для некоторого момента времени. Другой пример связан с богатым классом компартментальных моделей биологических систем. Пусть при введении меченого препарата в живой организм измерению поддается динамика этого препарата в отдельном органе или группе органов, но не во всем организме. Требуется определить распределение препарата по органам всего организма. В обоих случаях состояние биологического объекта задается набором неотрицательных чисел и только такие числа имеют биологический смысл. Полученные в эксперименте значения измеряемых параметров также имеют неотрицательные значения. Таким образом, решение задачи наблюдения по неотрицательному сигналу должно быть неотрицательным.

Перейдем к математической формулировке поставленных проблем. Допустим, что динамика экологического объекта задается моделью с непрерывно дифференцируемой правой частью:

$$\dot{X} = F(X), \quad (3.6.1)$$

где $F: R^m \rightarrow R^m$ — квазимоноotonно возрастающая функция

$$\partial(F)_i / \partial x_j \geq 0 \quad \text{при } i \neq j \quad (i, j = 1, \dots, m) \quad \text{и} \quad F(0) = 0.$$

Таким образом, положительный конус R_+^m является инвариантным множеством системы.

Пусть измеряемый сигнал задается непрерывно дифференцируемой функцией

$$Y = H(X), \quad (3.6.2)$$

где $H: R^m \rightarrow R^r$ и $H(0) = 0$. Обозначим через $X(t)$ решение системы (3.6.1). В реальной ситуации величину $Y(t) = H(X(t))$ точно определить невозможно, так как на полезный сигнал накладывается помеха. В эксперименте измеряется искаженный сигнал $\hat{Y}(t) = Y(t) + \Delta(t)$, где $\Delta(t)$ — некоторая помеха. Опишем свойства множества допустимых помех. Обозначим через \mathcal{L} векторное пространство функций скалярного аргумента, принимающих значения R^r . В зависимости от способа нормирования пространства $\mathcal{L}(\|\cdot\|_{\mathcal{L}})$ оно становится пространством \mathcal{L}^∞ существенно или почти всюду ограниченных, \mathcal{L}^1 интегрируемых, \mathcal{L}^p ($p > 0$) интегрируемых в p -й степени функций. Обозначим через \mathcal{L}_+ конус пространства \mathcal{L} , образованный вектор-функциями с почти всюду неотрицательными компонентами. Будем считать, что помеха $\Delta(t)$ принадлежит пространству \mathcal{L} . В тех случаях, когда H отображает R_+^m и R_+^r , измеряемый сигнал должен принадлежать \mathcal{L}_+ . В результате наложения помехи это условие может нарушиться, т.е. $\hat{Y}(t) = Y(t) + \Delta(t) \in \mathcal{L}$.

Сформулируем задачу построения оператора $G: \mathcal{L} \rightarrow R^m$, который по неискаженному сигналу $Y(t)$ точно восстанавливает состояние объекта, дает наименьшую погрешность при ограниченной помехе и на сигналах, принадлежащих множеству неотрицательных функций, дает неотрицательное решение, т.е.

а) $G(H(X(\tau))) = X(t)$, где $X(\tau)$ — решение системы (3.6.1) при $\tau \in [-T + t, t]$, принимающее значение $X(t)$ при $t = \tau$;

б) $\|G(Y(\tau) + \Delta(\tau)) - G(Y(\tau))\| \leq l \|\Delta(\tau)\|_{\mathcal{L}}$
и l принимает наименьшее возможное значение при всех $Y(\tau), \Delta(\tau) \in \mathcal{L}$
и $\tau \in [-T+t, t]$;

в) $G: \mathcal{L}_+ \rightarrow R_+^m$.

Определение. Если выполнено условие а) и функция является липшицевой, то решена задача наблюдаемости. Если выполнены условия а) и б), то решена задача оптимальной наблюдаемости [8, 30]. Если выполнены условия а) и в), то решена задача положительной наблюдаемости.

Перейдем к решению задачи оптимальной положительной наблюдаемости. В дальнейших рассуждениях условие положительности отображения H не используется. Вначале рассмотрим линейный случай. Пусть модель биологического объекта задана системой

$$\dot{X} = AX, \quad (3.6.3)$$

где $A - m \times m$ -матрица и вектор $X(t) \in R^m$ при всех $t > 0$. Измеряемый сигнал задается выражением

$$Y = CX, \quad (3.6.4)$$

где $C - r \times m$ -матрица. Помеха принадлежит пространству \mathcal{L} . Задача наблюдения состоит в построении интегрального оператора $G(Y(\cdot)) =$

$$= \int_{t-T}^t W(\tau)Y(\tau)d\tau, \text{ ядро которого } W(\tau) \text{ является } m \times r\text{-матрицей, каждая}$$

строка которой принадлежит сопряженному пространству \mathcal{L}^* . Выпишем таблицу соответствий [8, 12]:

$$(\mathcal{L}^\infty)^* = \mathcal{L}^1, \quad (\mathcal{L}^1)^* = \mathcal{L}^\infty \quad \text{и} \quad (\mathcal{L}^p)^* = \mathcal{L}^q, \text{ где } 1/p + 1/q = 1.$$

Наблюдение оптимально, если норма оператора $\|G\|$ минимальна [8, 12]. Наблюдение положительно, если $m \times r$ -матрица $W(\tau)$ при почти всех τ имеет неотрицательные компоненты.

Приведем пример. Допустим для определенности, что измеряемый сигнал принадлежит \mathcal{L}^2 . Тогда оптимальным положительным наблюдением

является оператор $\int_{t-T}^t W(\tau)Y(\tau)d\tau$, который удовлетворяет следующим

условиям:

а) $\int_{t-T}^t W(\tau)CX(\tau)d\tau = X(t)$, где $X(t)$ — решение уравнения (3.6.3);

б) $\int_{t-T}^t \|W(\tau)\|^2 d\tau \rightarrow \min$;

в) $W(\tau) \in \mathcal{L}_+^2$.

Решение линейной задачи дает следующая

Теорема 3.6.1. Для того чтобы система (3.6.3) была положительно наблюдаема по сигналу (3.6.4), необходимо и достаточно, чтобы множество достижимости сопряженной системы

$$\dot{X} = A^*X + C^*U(t) \quad (3.6.5)$$

содержало положительный конус R_+^m при $U(t) \in R_-^r$ и $t \geq 0$.

Доказательство. Выпишем решение уравнения (3.6.3), которое проходит через точку $X(t)$: $X(\tau) = e^{A(\tau-t)}X(t)$. Возвращаясь к исходной задаче, получаем $\int_{t-T}^t W(\tau)Ce^{A(\tau-t)}X(t)d\tau = X(t)$. Это векторное соотношение сводится к системе скалярных:

$$\int_{t-T}^t U_i^*(\tau)Ce^{A(\tau-t)}X(t)d\tau = E_i^*X(t) \quad (i = 1, \dots, m),$$

где $E_i^* = \underbrace{(0, \dots, 1, \dots, 0)}_i$ и $U_i^*(\tau)$ — i -я строка $m \times p$ -матрицы $W(\tau)$. Транспонируя

выражение и воспользовавшись произвольностью $X(t)$, получаем, что наблюдаемость имеет место тогда и только тогда, когда $\int_{t-T}^t e^{A^*(\tau-t)}C^*U_i(\tau)d\tau = E_i \quad (i = 1, \dots, m)$,

или, что эквивалентно, $\int_0^T e^{-A^*(T-\tau)}C^*U_i(t-T+\tau)d\tau = E_i$. Данные интегральные соотношения эквивалентны следующим задачам управления:

$$\frac{d}{d\tau}X = -A^*X + C^*U_i(\tau), \quad X(0) = 0, \quad X(T) = E_i$$

и

$$U_i(\tau) \in R_+^r \quad \text{при } \tau > 0 \quad (i = 1, \dots, m).$$

Замена времени $\tau \rightarrow T - t$ окончательно приводит нас к проблеме достижимости нуля из R_+^m при однонаправленных управлениях:

$$\dot{X} = A^*X + C^*U_i(t), \quad X(0) = E_i, \quad X(T) = 0, \quad U_i \in R_-^r.$$

Теорема доказана.

Следствие 3.6.1.1. Пусть для определенности $\mathcal{L} = \mathcal{L}^p$. Оператор $G = \int_{t-T}^t W(\tau)(\cdot)d\tau$, где $W(\tau) = (U_i^*(t-\tau))_{i=1}^m$, является оптимальным положительным наблюдением тогда и только тогда, когда $U_i(\tau) \in R_-^r$ доставляет решение задаче оптимального управления системой (3.6.5):

$$X(0) = E_i, \quad X(T) = 0, \quad \int_0^T \|U(\tau)\|^{1/q}d\tau \rightarrow \min$$

$$(1/p + 1/q = 1, \quad i = 1, \dots, m).$$

Следствие 3.6.1.2. Предположим, что матрица A имеет неотрицательные недиагональные элементы и в случае ее разложимости матрицы A^* и C^* имеют блочно-треугольный вид (3.1.6):

$$A^* = \begin{bmatrix} A_1^* & 0 & \dots & 0 \\ A_{21}^* & A_2^* & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ A_{s1}^* & A_{s2}^* & \dots & A_s^* \end{bmatrix}, \quad C^* = \begin{bmatrix} C_1^* & 0 & \dots & 0 \\ C_{21}^* & C_2^* & \dots & 0 \\ \dots & \dots & \dots & \dots \\ C_{s1}^* & C_{s2}^* & \dots & C_s^* \end{bmatrix},$$

где $A_i^* - m_i \times m_i$ -неразложимая матрица, $C_i^* - m_i \times p_i$ -матрица и $\text{rank}(A_i^*, C_i^*) = m_i$. Обозначим через $X_i \in R_+^{m_i}$ собственный вектор матрицы

A_i ($i = 1, \dots, s$), отвечающий собственному значению $\langle X_i, C^* U_i \rangle \geq 0$ нарушается при некотором $U_i \in R^{P_i}$ и всех $i = 1, \dots, s$. Для того чтобы задача положительной наблюдаемости имела решение, необходимо и достаточно, чтобы пара (A^*, C^*) удовлетворяла условиям теоремы 3.1.6.

Перейдем к изучению нелинейной задачи. На множестве липшицевых функций $G: \mathcal{L} \rightarrow R^m$, $G(0) = 0$ определим норму правилом [30]

$$\|G\| = \sup_{Y_1 \neq Y_2, Y_1, Y_2 \in \mathcal{L}} \frac{\|G(Y_1) - G(Y_2)\|}{\|Y_1 - Y_2\|}.$$

Требование оптимальности наблюдения сводится к отысканию такого наблюдения G , которое обладает наименьшей нормой. Обозначим $A = D_x F(0)$ и $C = D_x H(0)$.

Теорема 3.6.2. Если линейная система (3.6.3) положительно наблюдаема по сигналу (3.6.4), то нелинейная система (3.6.1) оптимально положительно наблюдаема по сигналу (3.6.2) в некоторой окрестности начала координат пространства \mathcal{L} .

Доказательство. Пусть оператор $\int_{t-T}^t W(\tau)(\cdot) d\tau$ решает задачу положительной наблюдаемости для линеаризованной системы. Тогда оператор $G(\cdot) = e^{AT} \int_{t-2T}^{t-T} W(\tau)(\cdot) d\tau$ также является положительным наблюдением для линеаризованной системы. Введем в рассмотрение оператор сдвига вдоль траекторий Γ_τ системы (3.6.1) за время $\tau \in [-2T + t, t]$. Построим последовательность отображений

$\Phi: R^m \xrightarrow{\Gamma_\tau} C^1 \xrightarrow{H} \mathcal{L} \xrightarrow{G} R^m$. Докажем, что это локальный диффеоморфизм. Очевидно, что $\Phi(0) = 0$. Далее, $D_x \Phi(0) = G(D_x H(0) D_x \Gamma_\tau(0)) = I_m$, так как $D_x \Gamma_\tau(0)$ – решение линейной системы (3.6.3). Отсюда по теореме об обратной функции в малой окрестности начала координат определено отображение $\Phi^{-1}: R^m \rightarrow R^m$. Докажем, что $N = \Phi^{-1}(G)$ – положительное липшицево наблюдение. По построению $N(H(\Gamma_\tau(X(t)))) = X(t)$. Убедимся в том, что $N: \mathcal{L} \rightarrow R^m$ – липшицева функция. Действительно, функция Φ^{-1} в силу непрерывной дифференцируемости липшицева с константой l_Φ . Пусть l_g – норма линейного оператора $G: \mathcal{L} \rightarrow R^m$. Тогда

$$\|\Phi^{-1}(G(Y_1)) - \Phi^{-1}(G(Y_2))\| \leq l_\Phi l_g \|Y_1 - Y_2\|.$$

Покажем теперь, что $N(\mathcal{L}_+) \subset R_+^m$. Обозначим через $\mathring{\mathcal{L}}_+$ подмножество строго положительных почти всюду функций из \mathcal{L} . Очевидно, что $G(\mathring{\mathcal{L}}_+) \subset \bar{K}^m \subset \overset{\circ}{R}_+^m$, где \bar{K}^m – некоторый замкнутый конус. Поскольку $N(0) = 0$ и Φ^{-1} близко к тождественному отображению, имеем $\Phi^{-1}(\bar{K}^m) \subset R_+^m$, т.е. $N(\mathring{\mathcal{L}}_+) \subset R_+^m$. По непрерывности получаем $N(\mathcal{L}_+) \subset R_+^m$.

Докажем существование оптимального наблюдения. Конус положительных функций $G: \mathcal{L}_+ \rightarrow R_+^m$ замкнут в слабой топологии пространства \mathcal{L} . Дословное повторение рассуждений теоремы 4 работы [30] завершает доказательство теоремы.

3.6.2. Динамическое наблюдение. Выше изложен метод определения состояния системы по наблюдениям за небольшой долей параметров, характеризующих это состояние. Данный метод сводит задачу наблюдения к решению ряда задач оптимального управления. Размерность вектора состояния задает число таких задач. Численное построение оптимального управления для систем высокой размерности является достаточно трудно-

емкой работой. Поэтому изложенный метод целесообразно применять при анализе моделей малой размерности.

Рассмотрим другой подход к проблеме. Он основан на использовании модели биологического объекта с целью проведения специально сконструированных численных экспериментов. Такая ситуация возникает в имитационном моделировании. Имитационная модель многосвязного экологического объекта имеет огромное число функциональных и параметрических зависимостей. Поэтому, несмотря на то, что такая модель не содержит неизвестных параметров, использование ее для получения каких-либо аналитических результатов не представляется возможным. Другими словами, имитационная модель является "черным ящиком", на вход которого можно подавать различные управляющие воздействия и на выходе наблюдать ответ системы. В отличие от условий реального эксперимента, вектор состояния модели является полностью наблюдаемой величиной. Таким образом, естественно возникает следующая задача. Как, пользуясь частичной информацией о состоянии экологического объекта и имитационной моделью этого объекта, определить его состояние в каждый момент времени? Перейдем к решению этой задачи. Допустим, что динамика экологического объекта задается системой (3.6.1) и наблюдаемый сигнал имеет вид (3.6.2). Модель объекта, которая может быть реализована на машинном носителе, задается выражением

$$\begin{aligned} \dot{Z} &= F_0(Z, U), \\ S &= H(Z), \end{aligned} \quad (3.6.6)$$

где $F_0(0, 0) = 0$ и $U \in R^p$ – вектор управлений. Задача динамической наблюдаемости состоит в следующем. Не используя информацию о конкретном виде правой части системы и предполагая лишь ее принадлежность некоторому классу, считая известным в каждый момент времени значение сигнала Y (3.6.2), сконструировать такое управление $U(t)$, при котором движение модели (3.6.6) асимптотически приближается к движению реального объекта, т.е. $t \rightarrow \infty$.

Решим вначале линейную задачу. Пусть динамика экологического объекта задается системой (3.6.3), наблюдаемый сигнал – выражением (3.6.4) и модель имеет вид

$$\dot{Z} = AZ + BU, \quad (3.6.7)$$

$$S = CZ, \quad (3.6.8)$$

где $m \times p$ -матрица B и $p \times m$ -матрица C образованы неотрицательными элементами и матрица A имеет неотрицательные недиагональные элементы. Если матрица A неразложима, то никакие дополнительные ограничения на вид матриц B и C не накладываются. В том случае, когда матрица разложима, т.е. имеет вид (3.1.6), будем считать, что матрица B также имеет вид (3.1.6), а матрица

$$C = \begin{bmatrix} C_1 & 0 & \dots & 0 \\ C_{21} & C_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ C_{s1} & C_{s2} & \dots & C_s \end{bmatrix}.$$

Пусть $J_j \subset \{1, \dots, m_i\}$ и $E_{ij} = \underbrace{(0, \dots, 1, \dots, 0)^*}_{j} \in R^{m_i}$ при $j \in J_i$, где $m_i \times m_i$ – размерность матрицы A_i ($i = 1, 2, \dots, s$).

Теорема 3.6.3. Предположим, что спектр матрицы A не пересекается с правой полуплоскостью. Тогда существует такое число ϵ_0 , что при коэффициенте усиления $0 < \epsilon < \epsilon_0$ управление $U = \epsilon(Y - S)$ решает задачу динамической наблюдаемости. Допустим теперь, что $C_i = E_{ij}^*$ и $B_i = E_{ij}$ ($i = 1, \dots, s; j \in J_i$). Тогда управление $U = \epsilon(Y - S)$ решает поставленную задачу при любом $\epsilon > 0$.

Доказательство. Обозначим $W(t) = X(t) - Z(t)$. Тогда $\dot{W} = (A - \epsilon BC)W$. Если B и C – произвольные неотрицательные матрицы, то устойчивость полученной линейной системы доказывается применением предложения 2.2.3 к каждой диагональной матрице $A_i - \epsilon B_i C_i$ ($i = 1, \dots, s$). Если матрицы B и C имеют специальный вид, то доказательство устойчивости полученной линейной системы состоит в повторении рассуждений теоремы 2.1.6, касающихся условия (2.1.12). Теорема доказана.

Полученный результат допускает перенесение на нелинейный случай (3.6.6).

Следствие 3.6.3.1. Предположим, что матрицы

$$A = D_x F_0(0, 0) = D_x F(0), \quad B = D_U F_0(0, 0) \quad \text{и} \quad C = D_x H(0)$$

удовлетворяют условиям теоремы 3.6.3. Тогда построенное там управление решает задачу динамической наблюдаемости для всех движений, лежащих в достаточно малой окрестности начала координат.

Доказательство очевидно.

Обсудим полученный результат. Прежде всего бросается в глаза, что при построении динамического наблюдения, в отличие от классического подхода [6, 34], наблюдаемость пары (A, C) не используется. Однако в том случае, когда линейная система неустойчива при нулевом управлении и нулевом возмущении, задачу динамического наблюдения можно решить при условии наблюдаемости этой пары и произвольности выбора матрицы B . Для исследования устойчивости системы, которой удовлетворяет ошибка наблюдения, полезными оказываются частотные критерии [5]. С их помощью проблема устойчивости решается путем специально сконструированных экспериментов с моделью.

Исследуем вопрос о влиянии помех на результаты наблюдения. Допустим, что наблюдению поддается искаженный сигнал $\hat{Y}(t) = Y(t) + \Delta(t)$, где $\Delta(t) \in \mathcal{L}^\infty$ – множество существенно или почти всюду ограниченных функций. Обозначим через \mathcal{C}^0 пространство непрерывных и ограниченных на бесконечном интервале функций скалярного аргумента. Будем считать, что $Z_1(t)$ и $Z_2(t)$ из этого пространства принадлежат одному классу эквивалентности, если $\|Z_1(t) - Z_2(t)\|_{\mathcal{C}} = \lim_{t \rightarrow \infty} \text{Sup} \|Z_1(t) - Z_2(t)\| = 0$. Пусть

$\bar{\mathcal{C}}^0$ – факторпространство пространства \mathcal{C}^0 относительно введенной эквивалентности. Это банахово пространство с нормой $\|\cdot\|_{\mathcal{C}}$. Построим отображение $G: Y \in \mathcal{L}^\infty \rightarrow \bar{Z} \in \bar{\mathcal{C}}^0$, где \bar{Z} – класс эквивалентности решения уравнения (3.6.6) или (3.6.7) при управлении, построенном в теореме 3.6.3.

Предложение 3.6.1. В линейном случае отображение G является непрерывным. В нелинейном случае оно определено в некоторой окрестности начала координат пространства \mathcal{L}^∞ и является липшицевым.

Доказательство. В линейном случае доказательство очевидно. В нелинейном случае $W(t) = Z_1(t) - Z_2(t)$ удовлетворяет уравнению

$$\begin{aligned} \dot{W} &= F(Z_1) - F(Z_2) - \epsilon BC(Z_1 - Z_2) + \epsilon(Y_1 - Y_2) = \\ &= [A - \epsilon BC]W + (D_X F(Z_1) - D_X F(0))W + o(Z_1)W + \epsilon(Y_1 - Y_2), \end{aligned}$$

или

$$\dot{W} = (A - \epsilon BC + A_0(t))W + \epsilon(Y_1 - Y_2).$$

Компоненты матрицы $A_0(t)$ можно считать сколь угодно малыми, так как при малой норме Y_i в пространстве \mathcal{L}^∞ малы $\|Z_i(t)\|$ для всех $t > 0$ и $i = 1, 2$. Поскольку при такой записи $W(t)$ является решением устойчивой линейной системы, доказательство очевидно.

На множестве липшицевых функций вида $G: \mathcal{L}^\infty \rightarrow \bar{\mathcal{C}}^0$ определим норму

$$\sup_{Y_1 \neq Y_2, Y_1, Y_2 \in \mathcal{L}^\infty} \frac{\|G(Y_1) - G(Y_2)\|_C}{\|Y_1 - Y_2\|}.$$

Оптимальным динамическим наблюдением будем называть наблюдение, построенное в теореме 3.6.3 и обладающее наименьшей нормой.

Следствие 3.6.3.2. Если выполнены условия теоремы 3.6.3 или следствия 3.6.3.1, то существует оптимальное динамическое наблюдение.

Доказательство состоит в повторении рассуждений теоремы 4 [30].

Проблема управления экологическими системами в предыдущих главах рассматривалась с концептуальных позиций. Были выделены фундаментальные свойства математических моделей экосистем и проведено их аналитическое исследование. Концептуальные модели обладают тем явным преимуществом, что они вскрывают структуру математической проблемы, возникающей при исследовании той или иной задачи. В то же время в реальных ситуациях для решения прикладных задач приходится конструировать модели, обладающие большой сложностью. При этом аналитические методы возможно использовать лишь на первом этапе, а для получения окончательного решения приходится обращаться к численным схемам. Это в свою очередь требует модификации самого подхода к процессу построения и исследования моделей. Если аналитические методы часто позволяют получать качественную картину поведения траекторий модели, исходя из качественных же предположений о характере основных зависимостей, то численное решение неизбежно требует выражения зависимостей в явной количественной форме. Поэтому оценка параметров уравнений (решение задачи идентификации) становится составной частью процесса построения модели.

Воспроизведение динамических свойств исследуемой системы с использованием численных методов и ЭВМ составляет основу метода имитационного моделирования. Возникший первоначально как метод статистических испытаний (метод Монте-Карло), имитационный подход быстро вышел за пределы этого частного способа имитации и превратился в мощный инструмент анализа сложных систем. В этой главе мы рассмотрим основные принципы имитационного моделирования. Вопросам математического и программного обеспечения задач моделирования будет посвящена следующая глава.

§ 4.1. Сущность метода

Существуют два способа получения новой информации о некотором объекте или системе. Это, во-первых, пассивное наблюдение и, во-вторых, наблюдение в условиях активного эксперимента. Последнее означает такую организацию эксперимента, при которой исследователь имеет возможность выбирать из некоторого множества входные управляющие воздействия, с тем чтобы увеличить информативность получаемых данных и, в част-

ности, оптимизировать выбранный им критерий. Активный эксперимент может проводиться либо с самим объектом, либо с его математической моделью, реализованной на ЭВМ. В этом последнем случае мы и говорим об имитационном эксперименте или об использовании метода имитационного моделирования. Таким образом, под имитационным моделированием понимают процесс построения модели сложной системы и проведения серий экспериментов с этой моделью, направленных либо на понимание специфики функционирования системы, либо на выработку стратегии управления, удовлетворяющей выбранным критериям [26]. В формулировке особенностей метода, приведенной в [7], говорится, что "имитационное моделирование — это процесс построения и исследования модели сложной системы, направленный на получение информации о самой моделируемой системе". Из этих определений можно вывести два следствия. Во-первых, мы говорим о необходимости применения метода имитационного моделирования в тех случаях, когда сложность системы превышает некоторый порог. Во-вторых, при этом условии конструирование модели является лишь первым шагом исследования. Не менее важно обеспечить рациональную организацию машинных экспериментов, позволяющую получать необходимую информацию при наименьших затратах машинного времени и других ресурсов ЭВМ. Поясним сказанное более подробно и прежде всего рассмотрим проблему сложности. К этому вопросу можно подходить с разных сторон. С точки зрения моделирования нас должна интересовать возможность воспроизведения на некотором абстрактном уровне основных свойств исследуемой системы, ее поведения как в реальных, так и в гипотетических ситуациях. Поэтому к трактовке понятия сложности в данном случае следует подходить с прагматических позиций, имея в виду именно эти цели.

Определение. Будем называть "сложной системой" такой объект реального мира, поведение которого невозможно предсказать с необходимой степенью детальности на основе учета обозримого набора ключевых параметров.

Таким образом, имея дело со сложной системой, мы должны исходить из того факта, что ее полное исследование невозможно и что следует заведомо ограничить себя лишь тем кругом задач, который позволяет получить частные ответы на ограниченный набор вопросов. Именно для решения той или иной конкретной задачи (или некоторого класса задач) и следует разрабатывать имитационную модель. Модель в данном случае может рассматриваться как "проекция" сложной конструкции в многомерном пространстве на подпространство значительно меньшей размерности. Поэтому для сложных систем и явлений в принципе может быть построен неограниченный набор моделей.

Из этого рассуждения следует и другой вывод. Коль скоро мы соглашаемся с тем, что для сложных систем построить полностью адекватные им модели невозможно, а следует конструировать лишь частные, но согласованные с характером решаемых задач описания, то возникает потребность в разработке специальной технологии имитационного моделирования. Эта технология должна охватывать все основные этапы построения, экспериментальной проверки и практического использования моделей. К вопросу о технологии моделирования мы еще вернемся в дальнейшем.

Обратимся к рассмотрению второй основной особенности метода имитационного моделирования, а именно к вопросу об организации машинных экспериментов с моделью. Имитационное моделирование в этом отношении аналогично экспериментированию с натурными объектами, хотя модель в этом отношении предоставляет гораздо больше возможностей. Как указано в [26], имитационная модель представляет собой "черный ящик", т.е. она обеспечивает задачу выходных сигналов, если на входы модели поступают те или иные внешние возмущения. Поэтому информацию о моделируемой системе в процессе машинного эксперимента можно получить в результате осуществления серий "прогонов" модели, а не путем ее "решения", как это делается при использовании аналитических методов. Иначе говоря, имитационные модели могут служить лишь средством анализа систем в тех условиях, которые задаются экспериментатором.

С этими важными обстоятельствами связана вторая особенность метода — необходимость рациональной организации и планирования машинного эксперимента. Очевидно, что для этих целей следовало бы воспользоваться всем тем арсеналом методов, который накоплен в математической теории планирования экспериментов [4], адаптировав их применительно к экспериментам, реализуемым на ЭВМ. Подробному рассмотрению этого вопроса будет посвящен заключительный параграф настоящей главы.

Несколько слов о терминологии. Как указано в [11], любая научная дисциплина имеет дело только с приближенными, "модельными" описаниями. При этом можно утверждать, что любая модель является имитационной, поскольку она имитирует реально происходящие процессы. Мы, однако, следуя установившейся традиции, под "имитационным моделированием" будем понимать тот частный метод имитации действительности, сущность которого описана выше и которому в литературе на английском языке соответствует термин "simulation modelling". С этой точки зрения можно утверждать, что хотя все модели являются имитационными, но некоторые из них следует считать более имитационными, чем другие. В [11] введено также понятие имитационной системы, под которой понимается "... совокупность моделей, имитирующих протекание изучаемого процесса, объединенная со специальной системой вспомогательных программ и информационной базой, позволяющих просто и оперативно реализовать вариантные расчеты". Термин "имитационная система" представляется нам весьма удачным. Он неоднократно будет эксплуатироваться в дальнейшем.

Перейдем к рассмотрению специфики имитационного исследования экосистем. Как уже отмечалось, при теоретическом анализе динамики экосистем приходится, как правило, абстрагироваться от описания конкретных ситуаций, выделяя и включая в модель некоторые обобщенные процессы, присущие целым классам реально существующих объектов. Это, разумеется, связано с теми целями, которые ставятся исследователем, — изучить по возможности наиболее широкий круг явлений, наиболее распространенные и типичные ситуации, встречающиеся в природе.

Цель построения имитационных моделей — принципиально иная. Назначение имитационных моделей в большинстве случаев сводится к постановке и решению важных прикладных задач — прогнозированию поведения реальных экосистем при тех или иных способах антропогенного воздей-

ствия и управления этими экосистемами. Поэтому при создании имитационной модели неизбежно происходит детализация и уточнение общих зависимостей, учет специфических механизмов, действующих в данной реальной ситуации. В этом смысле общая теория в экологии играет такую же роль, как и в других науках. Аналитические модели, подобно законам сохранения, определяют те запреты, которые нельзя нарушать при имитационном описании.

Мы уже упоминали во введении, что экосистема представляет собой термодинамически открытую систему, обменивающуюся энергией и веществом со своим окружением. Этот обмен происходит и внутри системы между ее отдельными частями. В организме или в абиотической среде накапливаются и расходуются отдельные органические и минеральные компоненты. Исходя из этого, в экосистеме выделяют "бункеры" или "емкости", запасы в которых и перетоки между которыми как раз и подлежат изучению. Такой балансовый подход уже намечался в общей теории. Так, например, при построении модели "микробное сообщество — среда" и "лес — вредитель" были введены коэффициенты утилизации, определяющие величину прироста биомассы потребителя на единицу потребленного субстрата или пищи. При построении имитационной модели этот балансовый подход доводится до своего логического конца. Выделяются переменные уровни, описывающие количества тех или иных веществ в различных частях системы, переменные темпы, определяющие скорости перетоков веществ из одной части системы в другую, управляющие и вспомогательные переменные, т.е. величины, от которых зависят скорости перетоков.

Выделение этих трех типов переменных стало традиционным при описании динамики любых экосистем (см., например, [33, 34]). Поясним введенные выше описания "емкости" или "бункера". При описании горизонтально однородного фитоценоза мы, например, выделяем различные части растений — живые и отмершие корни, стебли, листья, генеративные органы (колос). При этом в качестве переменных уровня в модель могут входить запасы углерода, азота, зольных элементов во всех этих частях. Более того, учитывая, что корни распределены по глубине почвенного профиля, обычно дополнительно выделяется несколько слоев корневой системы. Точно так же в наземном покрове может быть выделено несколько ярусов. Этот процесс стратификации означает замену реальной неоднородной по высоте и глубине распределенной системы ее многоемкостной моделью¹). Отметим, что одно и то же физическое пространство при этом может содержать несколько емкостей. Так, например, в слое почвы 0 — 10 см могут быть выделены емкости для воды, фосфора, различных форм минерального азота (аммиак, нитраты, нитриты), живых и отмерших корней, гумуса и т.д. Более того, число выделенных в почве слоев и их толщины для всех этих веществ могут быть, вообще говоря, различны. Таким образом, понятие "емкости" более тесно связано с методологией моделирования, чем с геометрическими характеристиками системы. Необходимость перехода от системы с распределенными параметрами к много-

¹) В литературе на английском языке такие модели обычно называются компартаментальными — compartment models (от слова compartment — отделение, купе, перегородка).

емкостной модели определяется тем, что имитационная модель реализуется на ЭВМ.

Любая реальная экосистема является достаточно сложным объектом. С этим обстоятельством связано выделение двух этапов при построении ее имитационной модели. На этапе анализа определяется структура модели, производится разбиение общей системы на блоки и дается математическое описание отдельных блоков. На этапе синтеза осуществляется стыковка блоков и общая сборка модели с применением ЭВМ. При этом, как отмечается многими авторами, на этапе сборки моделей целесообразно применение специальных языков имитационного моделирования.

§ 4.2. Структура имитационных моделей экосистем

Как отмечалось в предыдущем параграфе, имитационные модели экосистем конструируются как динамические балансовые структуры блочного типа. Динамический характер моделей связан с тем очевидным обстоятельством, что исследователя, как правило, интересует протекание процессов во времени. Это понятно, если идет речь о задачах научно-исследовательского плана: в экологии нас интересует не столько установившиеся состояния (хотя и они тоже), сколько те переходные процессы (так называемая сукцессия), которые возникают в ответ на то или иное антропогенное воздействие. Однако и в прикладных задачах наиболее часто приходится рассматривать динамику экосистем. Так, например, при выработке рациональных методов борьбы с насекомыми-вредителями лесного и сельского хозяйства следует учитывать динамику возрастной структуры вредителя, поскольку его вредоносность существенно зависит от фазы развития. Кроме того, необходимо описывать рост и развитие растений, создающих вредителю "кормовую базу" и совокупность процессов в окружающей среде, поскольку вспышки численности насекомых наступают только при вполне определенных сочетаниях многих условий.

Требование включения в модель балансовых соотношений по основным компонентам энерго- и массообмена связано с двумя обстоятельствами. Первое из них заключается в том, что существо процессов, описываемых с помощью компартментальных моделей, сводится к накоплению в отдельных компартментах, расходованию и перетокам энергии и вещества. Другое обстоятельство связано с чисто утилитарными целями. Так, в земледелии на стадии планирования ставится задача разработки энерго- и ресурсосберегающих технологий, обеспечивающих наиболее эффективное и экономичное использование всех антропогенных ресурсов — поливной воды, удобрений и т.п. При построении таких технологий одновременно решается задача охраны среды от загрязнений химическими мелиорантами. Эти же цели преследуются при решении задач оперативного управления. Построение балансовых моделей компартментального типа создает естественную основу для реализации такого подхода.

Наконец, блочная структура моделей связана с необходимостью описания взаимодействия процессов самой различной природы (физических, физико-химических, биологических, экологических) и обладающих самыми различными временами переходных процессов. Блочный принцип конструирования модели означает ее декомпозицию как на содержательном,

так и на формальном уровне описания, т.е. вывод и обоснование частных математических соотношений для отдельных блоков с их последующим объединением в комплексную модель. Очевидно, что очень важным при этом оказывается вопрос организации взаимодействия блоков при "сборке" модели.

В гл. 1 дано краткое описание структуры двух имитационных моделей. Ниже будет рассмотрен еще один пример — модель агрофитоценоза пшеницы. Подробное описание совокупности моделируемых процессов, структура модели в целом в силу ее достаточной сложности проясняют, как нам представляется, многие характерные особенности метода имитационного моделирования. В связи с этим помещенный ниже пример неоднократно используется далее как "базовый" для иллюстрации высказываемых утверждений.

4.2.1. Модель агрофитоценоза пшеницы. Агроэкосистемы обладают рядом существенных особенностей, которые отличают их от других встречающихся в природе экосистем. Основной из них является то, что агроэкосистема в принципе существует только за счет постоянного управляющего воздействия со стороны человека. С одной стороны, это обеспечивает высокую степень ее специализации и интенсивный уровень эксплуатации, с другой стороны, приводит к большим затратам энергии и материальных средств, поддерживающих эту специализацию и не дающих системе перейти в состояние устойчивого экологического равновесия, при котором преобладающими оказались бы дикие виды растительности, бесполезные для человека.

Биотическая часть агроэкосистемы содержит все звенья, характерные для любых наземных экосистем:

- продуценты, т.е. зеленые растения, связывающие под действием света углерод воздуха и превращающие его в первичные углеводы, а затем в аминокислоты, белки, целлюлозу и др.;

- консументы, т.е. травоядные и "хищники" различного трофического уровня, поедающие продуцентов или их отдельные части (к консументам в агроэкосистеме относятся также различные вредители и болезнетворные микроорганизмы);

- редуценты, т.е. виды, разлагающие остатки отмерших растений и животных и замыкающие тем самым цепь круговорота вещества в агроэкосистеме.

Абиотическая часть системы — среда обитания растений и животных — используется ими как "жизненное пространство" и как источник энергии и минеральных элементов.

Специализация агроэкосистемы способствует нарушению природного равновесия. Если в природной экосистеме все виды приспособлены друг к другу с точки зрения трофических, хорологических и иных связей так, что в целом они обеспечивают замкнутость циклов круговорота веществ, то в агроэкосистеме эти связи нарушены. Отсюда возникает необходимость подпитки агроэкосистемы значительными добавками элементов питания. Обеднение агроэкосистемы биологическими видами в любой момент времени может привести к разбалансированию многих экологических процессов, что заставляет компенсировать эту "экологическую бедность" введением севооборотов, т.е. обеспечивать вместо одновременного сосуществования на одном поле многих видов их "развертку" во времени и в пространстве. Таким образом, в отличие от природных экосистем,

существование агроэкосистемы естественным образом предполагает целенаправленную антропогенную деятельность.

Агроэкосистемы специфичны не только с точки зрения круговорота веществ. Энергетический цикл в них также складывается иначе, чем в природе. Если естественные энергетические процессы происходят в двух поясах — "зеленом" и "коричневом" — и обусловлены притоком солнечной энергии, то для агроэкосистем характерно значительное внешнее энергетическое воздействие со стороны обрабатывающих машин и агрегатов. Их использование в рамках технологии возделывания культуры по существу обеспечивает специализацию агроэкосистемы и рост ее продуктивности. По анализу Ю. Одума увеличение энергозатрат на гектар пашни от 0,7 до 7,5 кВт приводит к увеличению урожая от 2 до 9т/га. Из сказанного следует, что для исследования продукционного процесса агроэкосистем необходимо развитие системного подхода. Опишем структуру сезонной модели агрофитоценоза пшеницы. Детальное описание модели, включая систему дифференциальных уравнений и обоснование метода их интегрирования, содержится в [1, 18]. Посев пшеницы можно с достаточной степенью точности рассматривать как составленный растениями, равномерно размещенными на некоторой площади. Внешние условия наряду с воздействиями со стороны вредителей, болезней и сорняков определяют динамику роста и развития растений пшеницы и в конечном счете определяют уровень урожая и его качество. С другой стороны, сами растения в процессе вегетации изменяют среду своего обитания, оказывают влияние на динамику влаго- и теплообмена в почве и приземном воздухе, изменяют радиационный, газовый режим и т.д. Поэтому комплексная модель продуктивности посевов должна включать в себя взаимосвязанные описания процессов, происходящих как в самих растениях, так и в среде их обитания.

Будем считать, что площадь поля достаточно велика и что краевым эффектом вследствие этого можно пренебречь. Это позволяет описать агрофитоценоз как нестационарную одномерную систему с двумя независимыми переменными: вертикальной координатой x и временем t . Для построения модели нужно, следовательно, выделить некоторую единичную площадь поля. Состояние абиотической части системы в каждый момент времени характеризуется распределением по вертикали различных физических переменных: радиации, температуры и влажности воздуха в посевах, температуры и влажности почвы и т.п. (рис. 4.1). Точно так же биологическая часть системы характеризуется набором вертикально распределенных переменных: плотности ассимилирующей поверхности фитоэлементов $s(x)$ и поглощающей поверхности корней $\omega(x)$, плотности отдельных составляющих фитомассы (углеводы, аминокислоты, белки) и фитомассы в целом и др. В динамике все эти величины изменяются. Задачей моделирования как раз и является описание изменения этих переменных в пространстве (т.е. по вертикали) и во времени.

В модель включено описание процессов трех типов:

- 1) энерго- и массообмен, происходящий в среде обитания растений (в почве и приземном воздухе) и в самих растениях;
- 2) совокупность биофизических и физиологических процессов в растительном покрове (РП), определяющих прирост биомассы, рост и развитие отдельных органов растений и формирование конечного урожая;

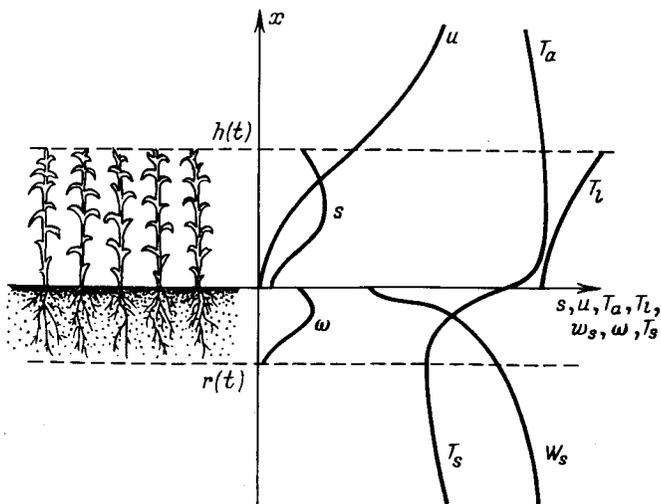


Рис. 4.1. Распределение фитоэлементов и параметров среды обитания в системе почва – растение – атмосфера; s – объемная плотность фотосинтезирующих органов, ω – объемная плотность поглощающей поверхности корней, T_a , T_l , T_s – температура соответственно воздуха, листьев и почвы, u – скорость ветра, w_s – объемная влажность почвы, h – высота посева, r – глубина проникновения корней

3) экологические взаимодействия культурных растений с сорняками, болезнетворными микроорганизмами и вредителями.

Сложность объекта – сельскохозяйственного поля и многообразие протекающих на нем процессов, которые необходимо учитывать при моделировании, неизбежно проводят к тому, что модель следует конструировать в виде некоторой блочной структуры. Обращение к каждому блоку модели осуществляется как при работе базовой модели в целом, так и при использовании ее специализированных вариантов. Особенностью функционирования такой блочной структуры является разделение информационных потоков на те, которые замыкаются внутри отдельных блоков, и на те, которые передаются из блока в блок. Рассматриваемые в целом, эти информационные потоки определяют характер и объем экспериментальных исследований, необходимых для идентификации модели.

Охарактеризуем те процессы, которые представлены в модели. Это прежде всего влагообмены в системе почва – растение – приземный воздух. Влага является важнейшим, а для ряда почвенно-климатических зон основным лимитирующим урожай фактором. Движение влаги в почве обеспечивает растения минеральным питанием, а влагосодержание растений посредством изменения устьичного сопротивления регулирует поступление углекислого газа в лист. Многие другие процессы жизнедеятельности растений также определяются содержанием и потоками воды в почве, растении и приземном воздухе. При моделировании процесса влагопереноса в системе важную роль играют такие характеристики архитектурно-посева, как распределение поглощающей поверхности корней по глубине почвы и распределение объемной плотности листьев по высоте растительного покрова.

Следующим процессом, который рассмотрен в модели, является теплоперенос. С температурным режимом связаны скорости протекания большинства биологических процессов в почве и в растении. Кроме того, задача турбулентного влагопереноса в посеве решается совместно с задачей теплопереноса в почве и в приземном воздухе.

В модели представлен радиационный режим посева, который тесно связан с температурным режимом и с архитектурой посева. Поглощаемая посевом фотосинтетически активная радиация определяет (совместно с потоками CO_2) интенсивность процесса фотосинтеза.

Охарактеризуем процессы, представленные в биологическом блоке. Это прежде всего процесс фотосинтеза. Он определяет включение углерода в метаболизм растения. Существенное влияние на фотосинтез оказывают температурный и радиационный режимы посева и степень открытия устьиц. Следующий процесс, описанный в физиологических блоках, — это процесс биосинтеза, в результате которого образуется структурная биомасса яруса. Перераспределение ассимилятов между различными органами связано с процессами транспорта. В модели описывается транспорт продуктов фотосинтеза и элементов минерального питания. Взаимодействие этих двух потоков вещества определяет распределение биомассы между наземной и подземной частями растения. Растение в физиологических блоках представлено в виде цепочки биохимических реакторов, объем которых пропорционален структурной биомассе соответствующих ярусов.

Совместно с процессами фотосинтеза в модели рассматриваются процессы дыхания, определяющие интенсивность накопления биомассы. Скорости протекания всех перечисленных процессов существенно зависят при этом от абиотических параметров посева.

Модель в целом с математической точки зрения представляет собой систему из нескольких уравнений в частных производных параболического типа и нескольких десятков обыкновенных дифференциальных уравнений. При переходе к численной схеме выбирается шаг интегрирования по координате x и по времени t . При этом с учетом требований, предъявляемых к точности решения, в почвенной части модели выделяется до тридцати узлов расчетной сетки, а в посеве — до десяти. Базовый временной шаг модели выбран равным одному часу, что позволяет с достаточной точностью осуществить имитацию суточного хода как абиотических (энергообмен), так и основных биотических (фотосинтез, метаболизм) процессов. Общая размерность пространства состояний модели после перехода к дискретному описанию превышает 1000.

При построении любой модели следует придерживаться некоторого принципа "равной точности" или равного уровня детализации блоков. Это означает, что не следует стремиться к излишней подробности описания одного из блоков, если при построении других приняты огрубляющие допущения. Хотя формализовать этот принцип затруднительно, обычно в процессе конструирования модели можно интуитивно оценить ту степень подробности в описании данного блока, которая согласована с уровнем детальности рассмотрения всех остальных.

Общая структурная схема модели представлена на рис. 4.2. В качестве входных переменных модели выступают контролируемые (агротехника) и неконтролируемые (погода) внешние воздействия. При этом динамика по-

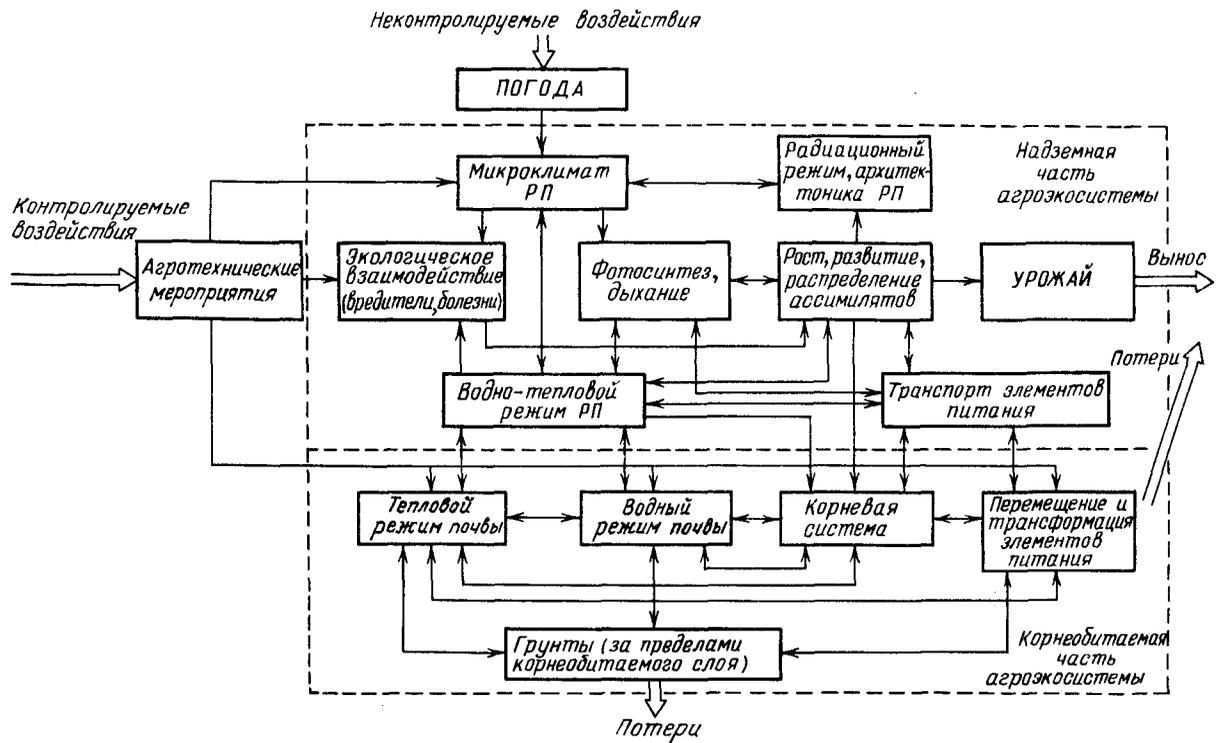


Рис. 4.2. Структурная схема модели продуктивности агроэкосистемы

годных условий представлена в модели реализациями многомерного случайного процесса.

а) Моделирование водного транспорта в системе почва — растение — атмосфера. Подмодель водного транспорта распадается на блок, моделирующий движение воды в почве, и блок, моделирующий транспорт воды в растениях. Движение воды в почве рассчитывается в соответствии с принятым в настоящее время представлением о законах влагопереноса в ненасыщенной зоне [12]. Влияние корневой системы растений на влагообмен поля сводится к дифференцированному поглощению влаги корнями из различных слоев почвенного горизонта. Проводимость стенок корней предполагается зависящей от температуры почвы, а поступление влаги в растение считается пропорциональным перепаду потенциала на границе почва — корень. Баланс влаги в расчетном слое почвы определяется интенсивностью транспирации растений, испарением с поверхности почвы, поступлением влаги в результате поливов или выпадения осадков и обменом на нижней границе. Изменение влажности различных почвенных слоев во времени протекает неодинаково. Это связано прежде всего с неравномерностью распределения поглощающей поверхности корней по глубине. В тех слоях почвы, где расположена основная масса корней, процессы изъятия влаги идут значительно быстрее, чем в других слоях почвы. Неодинаковая влажность различных почвенных слоев порождает потоки влаги между ними. Эти потоки переносят минеральные и органические вещества и изменяют профиль их концентрации.

На каждом временном шаге в модели рассчитываются профили влажности и водного потенциала почвы по глубине, потоки между слоями почвенного горизонта и количество свободной влаги на поверхности почвы, которая образуется при интенсивных осадках и поливах. При выпадении осадков часть влаги задерживается листовым пологом. В результате этого в растительном покрове образуется дисперсно распределенная влага, которая значительно влияет на процессы энерго- и массообмена и в конечном счете на фотосинтетическую активность посева. Поэтому процесс перехвата осадков также учитывается в модели.

Если преимущественное направление потоков воды в почвенном горизонте ориентировано снизу вверх, то идет иссушение верхних слоев и подпитка этих слоев из нижележащих. В результате этого процесса может повыситься засоленность почвогрунтов. На рис. 4.3 представлены результаты расчетов по модели и их сравнение с экспериментальными данными, полученные Б.Г. Заславским и А.Т. Нагиевым для засоленного хлопкового поля [6].

Блок, моделирующий движение воды в растении, служит для расчета скорости поглощения воды корнями и динамики функционирования устьичного аппарата. Поглощение воды корнями является в основном пассивным процессом, происходящим в результате внутреннего водного дефицита, вызванного транспирацией. Наиболее важными факторами, регулирующими поглощение воды, является доступная почвенная влага, аэрация почвы, температура почвы и градиент водного потенциала на границе почва — корень.

Ксилема играет роль распределительной системы, при посредстве которой ткани снабжаются водой. Значительно сопротивление вода претерпе-

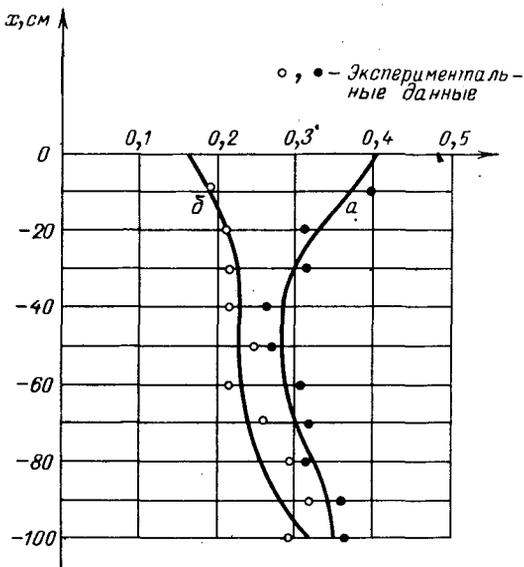


Рис. 4.3. Профили влажности почвы: а — на следующей день после полива; б — через десять дней после полива

вает при движении от окончаний сосудов в листьях к местам испарения, а также на границе лист — атмосфера. Испарение воды растением идет через устьица и кутикулу листьев и стеблей. Кутикулярная транспирация обычно составляет незначительную часть устьичного и лишь у молодых растений может давать почти половину общей потери воды. Поэтому основной вклад в регулирование потребления воды растением вносит функционирование устьичного аппарата. Регуляторный механизм устьиц служит для согласования двух противоположных тенденций. Открытие устьиц должно быть синхронизировано с потребностью тканей, ассимилирующих CO_2 , и в то же время не должно вести к чрезмерным потерям воды и пересыханию растения. В связи с этим апертура устьиц регулируется концентрацией CO_2 в листе и его оводненностью [5]. В определенных пределах влиянием температуры на регуляторный механизм устьиц можно пренебречь и, поскольку концентрация CO_2 в листе находится в прямой зависимости от облученности посева, иногда можно предполагать, что устьичное сопротивление определяется количеством поглощенной радиации.

Динамика раскрытия устьиц настолько сложно зависит от внутреннего состояния листа и внешних факторов, что даже при постоянных внешних условиях и стационарном освещении в водном транспорте растения могут возникнуть автоколебания, связанные с периодическим закрытием и раскрытием устьиц [5]. Этот процесс автоколебаний воспроизводится в модели. Правильный расчет устьичного сопротивления необходим прежде всего для высокой достоверности определения транспирационных и энергетических потоков, вычисляемых в блоке моделирования микроклимата, а также для определения количества накопленных ассимилятов, определяемого в блоке фотосинтеза и фотодыхания.

б) Моделирование энерго- и массообмена. Моделирование процессов энерго- и массообмена в системе почва – растение – атмосфера сводится к совместному решению следующих задач:

- расчет радиационного режима посева, включая рассмотрение коротковолновой и длинноволновой радиации;
- расчет аэродинамического режима посева;
- расчет турбулентного и тепло- и влагопереноса в посеве;
- расчет водного транспорта в почвенно-растительном комплексе;
- расчет теплового режима почв.

Моделирование динамики водного режима рассмотрено выше. Перейдем к рассмотрению всех остальных составляющих блока энерго- и массообмена.

Подробному рассмотрению вопросов расчета коротковолновой радиации в посеве посвящено большое число работ, подытоженных в монографиях [21, 23]. Учитывая "принцип равнозначности" для описания радиационного режима, включая интегральную радиацию (ИР) и фотоактивную радиацию (ФАР), можно принять полуэмпирические формулы, приведенные в [23]. Результатом расчета являются профили поглощаемой ИР и ФАР в зависимости от интенсивности приходящей радиации и архитектуры посева. Расчет длинноволновой (тепловой) радиации производится по формулам, приведенным в [23].

Для аэродинамического режима посева наиболее приемлемой является модель, развитая в ряде работ Г.В. Менжулина (см., например [9]). Подробный анализ этой модели показывает, что все расчеты могут быть выполнены предварительно и в общей модели представлены в виде таблиц или аппроксимационных зависимостей.

В настоящей работе принят второй путь. Профили скорости и коэффициент турбулентной вязкости в посеве задаются экспоненциальной зависимостью от интегрального листового индекса. Соответствующие коэффициенты аппроксимированы по данным [22] квадратичной регрессионной зависимостью от плотности фитозлементов и высота посева.

Для расчета турбулентного тепло- и влагопереноса в посеве используется диффузионная модель [1]. Два дифференциальных уравнения этой модели включают четыре переменных – температуру воздуха $T_a(x)$ и листьев $T_l(x)$, влажность воздуха в межлистовом пространстве $q_a(x)$ и в межклетнике $q_l(x)$. Для замыкания модели следует дополнительно использовать два конечных соотношения. Одно из них определяет связь влажности насыщения в межклетнике с температурой листьев. Другое соотношение – условие теплового баланса ярусов листьев. В результате задача расчета тепло- и влагопереноса в посеве сводится к необходимости совместного решения двух краевых задач, связанных нелинейными конечными соотношениями.

Наконец, теплоперенос в почве описывается обычным уравнением теплопроводности. Специфика этого уравнения для решаемой задачи заключается в том, что входящие в него коэффициенты (теплоемкость и теплопроводность почвы) зависят от влажности почвы в текущий момент времени.

Таким образом, моделирование процессов энерго- и массообмена в системе почва – растение – атмосфера сводится к необходимости совмест-

ного численного решения четырех краевых задач (влаго- и теплоперенос в посеве, влаго- и теплоперенос в почве), связанных условиями сшивания на границе почва — воздух и нелинейными конечными соотношениями в посеве.

Дадим краткое описание способа решения. Прежде всего, используя сеточный метод, необходимо выбрать узлы расчетной схемы и перейти к разностному аналогу дифференциальных уравнений. В модели принят переменный шаг по вертикальной координате и по времени. При этом в почве может быть выбрано до 30 узлов пространственной сетки, а в посеве — до 10 узлов. Базовый временной шаг модели принят равным одному часу. Расчет в целом производится следующим образом [1]. Задача расчета динамики водного транспорта в почве на время одного временного шага решается автономно по неявной схеме с линеаризацией. Для решения трех остальных задач используется итерационный метод: Первоначально задаются нулевые приближения для температуры поверхности почвы и температуры ярусов листьев. При этом могут не выполняться условия теплового баланса. Решая три автономных краевых задачи первого рода на нулевом шаге итерации, получаем в качестве результата профили температуры почвы, температуры и влажности воздуха в посеве. Линеаризуя уравнения теплового баланса, из условия их невязки, находим первые приближения для температуры поверхности почвы и температуры ярусов листьев, после чего процесс повторяется. Окончание итерационного процесса производится по достижении поправки по величинам температуры в пределах заданной точности. Как показали результаты расчетов, для завершения итерационного процесса требуется от 8 до 12 — 15 шагов.

Приведем некоторые результаты моделирования. На рис. 4.4 изображены профили температуры почвы, воздуха и листьев в ночное (3 часа) и дневное (15 часов) время суток. Расчет соответствует условиям безоблачного неба при скорости ветра 15 м/с и относительной влажности воздуха 63 — 70%. Кривая *a* соответствует суточному ходу температуры на уровне метеобудки при $T_{\min} = 12^{\circ}\text{C}$ и $T_{\max} = 22^{\circ}\text{C}$, а кривая *b* — соответственно $T_{\min} = 18^{\circ}\text{C}$ и $T_{\max} = 28^{\circ}\text{C}$. На рис. 4.5 приведен суточный ход температуры поверхности почвы и температуры почвы на глубинах 0, 10 и 20 см. Видно, что с увеличением глубины затухает амплитуда колебаний температуры и происходит сдвиг по фазе.

в) Моделирование процесса газообмена. Процесс фотосинтеза является основным процессом питания растений и определяется накоплением растениями органического вещества. Часть этого вещества расходуется листьями в процессе дыхания ночью и нефотосинтезирующими органами ночью и днем, а часть входит в состав клеток и тканей в качестве структурных и запасных веществ. Таким образом, величина биологического урожая определяется разностью между интенсивностью фотосинтеза и интенсивностью дыхания и фотодыхания. Чем больше эта разница, тем выше показатель чистой продуктивности фотосинтеза.

Основной задачей подблока газообмена в общей модели формирования урожая является формирование функции ассимиляции двуокиси углерода в зависимости от внешних факторов: температуры фотосинтезирующей поверхности, поглощенной ФАР, концентрации CO_2 у поверхности

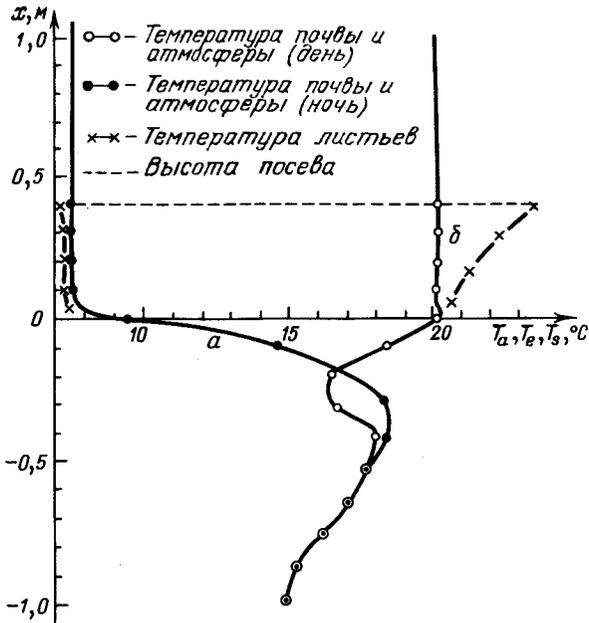


Рис. 4.4. Вертикальные профили температуры воздуха, листьев и почвы

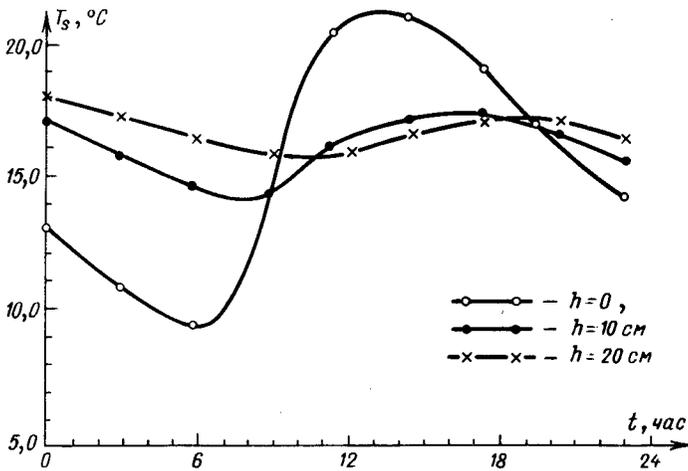


Рис. 4.5. Суточное изменение температуры почвы на разной глубине

листа, обеспеченности растения влагой и минеральными элементами, этапов органогенеза и физиологических характеристик моделируемого растения. Для формирования этой функции необходимо рассмотреть поток углекислого газа из межклеточного пространства к хлоропластам и его фиксацию на хлоропластах. Интенсивность этого потока определяется как физико-механическими свойствами листа (сопротивлением прилистового слоя, сопротивлением устьиц и кутикулы, сопротивлением мезофила, растворимостью CO_2 в жидкой фазе), так и его фото- и биохимическими свойствами.

В соответствии с этим модель газообмена записывается в виде системы балансовых уравнений с источниками и стоками, интенсивность которых определяется совместным воздействием внешних и внутренних факторов, зависящих от условий окружающей среды и морфофизиологических особенностей растения. Существенными отличиями рассматриваемого подблока газообмена от обычно применяемых в моделях продуктивности являются:

1) более полный учет дыхания структурной биомассы растения и выделения CO_2 в процессах метаболизма (эти потоки вычисляются в блоке роста и развития);

2) достаточно подробное описание фотосинтеза и фотодыхания, основанное на построении метода цикла Кальвина;

3) учет влияния температуры на растворимость CO_2 в процессе транспорта к хлоропластам;

4) учет изменения интенсивности фотосинтеза в процессе онтогенеза растений.

В результате была получена модель, позволяющая с необходимой степенью точности вычислять прирост биомассы растения в зависимости от его сортовых особенностей и изменения условий произрастания. Более подробное рассмотрение подблока газообмена приведено в работах [19, 20].

г) Моделирование процессов роста и распределения ассимилятов. В соответствии с общей идеей компартментального подхода в модели каждый ярус растения представляет собой единый биохимический реактор. Процесс биосинтеза описывается в каждом реакторе одним уравнением типа уравнения Михаэлиса – Ментен с некоторыми параметрами. Скорость синтеза структурной биомассы в первом ярусе зависит от количества растворимых углеводов, элементов минерального питания в данном ярусе и объема яруса, который считается пропорциональным количеству структурной биомассы. Предполагается, что зависимость скорости реакции от температуры яруса описывается законом Аррениуса. Так как от яруса к ярусу меняются температура и скорость поступления продуктов фотосинтеза, то меняется и скорость реакций.

Внутриклеточный транспорт связан с разобщением биосинтетических процессов во времени и пространстве. Поскольку он снижает уровень биосинтеза приблизительно только на 1%, внутриклеточный транспорт в модели, как правило, не учитывается. Помимо внутриклеточного движения в растении постоянно осуществляется перемещение субстратов из производящих органов в органы потребляющие, а также перераспределение метаболитов между частями целого растения. Предполагается, что

скорость транспорта растворимых углеводов пропорциональна разности их концентрации в соседних ярусах.

Минеральные вещества поступают в растение из почвы через корневую систему. Поглощаемые ионы пассивно передвигаются в "свободном пространстве" корня путем диффузии и с гидролитическим потоком жидкости. Проникновение минеральных веществ в клетки и передвижение по клеткам через плазмодесмы осуществляется за счет макроэргических связей АТФ. Распределение азота по растению регулируется потребностью растущих органов и зависит от соотношения минеральных веществ и углеводов в каждом ярусе.

Распад биомассы происходит как в функционирующих, так и в стареющих клетках. Удельная скорость распада зависит от температуры. Высвободившиеся в результате распада минеральные элементы частично могут реутилизироваться.

Под дыханием понимается процесс, связанный с распадом углеводов, в результате которого высвобождается энергия, обеспечивающая метаболизм и транспорт в растении. Так как кинетика метаболизма и транспорта уже описана, то из известных балансовых соотношений можно вычислить затраты субстрата на дыхание. Отметим, что при описании дыхания объединены две стадии преобразования химической энергии: стадии окисления субстрата, во время которой образуются макроэргические связи АТФ, и стадия использования энергии АТФ. Кроме того, в балансовом уравнении дыхания учитываются затраты углеводов на обеспечение энергией процесса биосинтеза и транспорта органических и неорганических веществ. В процессе дыхания выделяется углекислый газ, который частично используется в фотосинтезе. Его динамика описывается на основе балансовых соотношений.

Интенсивности превращений и транспорта отдельных субстратов определяются кинетическими уравнениями биосинтеза и транспорта, а динамика отдельных элементов биомассы определяется уравнениями баланса.

Таким образом, в блоке реализуется принцип "источников и стоков", позволяющий учесть как влияние абиотических факторов, так и адаптивные реакции растения.

д) Реализация модели на ЭВМ. В предыдущих пунктах рассмотрены особенности отдельных блоков модели. Очевидно, что каждый из блоков представляет собой описание той или иной группы однородных физических, биофизических, биохимических или физиологических процессов, протекающих в отдельных частях системы почва — растение — атмосфера. В то же время каждый из блоков решает вполне определенную математическую задачу.

Как уже указывалось, модель в целом представляет собой систему линейных и нелинейных уравнений в частных производных, а также нескольких десятков обыкновенных дифференциальных уравнений, как правило, нелинейных. Задача реализации численной схемы решения подобных систем уравнений на ЭВМ представляет собой самостоятельную и достаточно трудную задачу. В связи с этим возникает необходимость в разработке специального математического и программного обеспечения модели. Многие зарубежные модели реализованы с использованием специального

языка CSMP [32]. При разработке данной модели была использована специально созданная для этой цели система СИМОНА¹⁾ [14].

СИМОНА осуществляет сборку моделей из отдельных блоков, реализованных в виде подпрограмм на языке Фортран, передает управление от блока к блоку в ходе прогонки модели, управляет процессом пошагового заикливания и дроблением шага (в случае необходимости), управляет процессом общения с внешними устройствами ЭВМ и интерфейсом пользователя.

Для целей практического использования модели необходимо проведение расчетов для ансамбля неконтролируемых (случайных) погодных условий. Это влечет за собой необходимость статистической обработки результатов моделирования. Для автоматизации процесса накопления данных имитационного моделирования предусмотрена стыковка системы СИМОНА с разработанной ранее системой статистической обработки данных САФИСТ [8]. Организация имитационного эксперимента в целом осуществляется с помощью специального модуля — генератора факторных планов [15].

Процесс имитации сводится к последовательному выполнению следующих этапов:

- 1) сборка модели из отдельных блоков;
- 2) ввод начального состояния модели;
- 3) прогонка моделей с возможным прерыванием;
- 4) запись результатов моделирования на МД и (или) их распечатка.

В случае, если проводится машинный эксперимент, связанный с изучением влияния на конечный результат (например, урожай) различных входных управляющих воздействий, осуществляется многократная прогонка модели, для чего

- 1) вводится задание на генерацию плана эксперимента;
- 2) осуществляется многократная прогонка модели в соответствии с планом;
- 3) результаты расчета заносятся в базу данных системы САФИСТ;
- 4) производится статистическая обработка полученных данных и выдача результатов пользователю.

§ 4.3. Имитационные системы

В § 1.3 введения и в предыдущем пункте описана структура некоторых имитационных моделей, предназначенных для решения достаточно широкого круга задач как исследовательского, так и прикладного плана. Все они обладают рядом специфических особенностей, среди которых можно выделить:

- большую размерность пространства состояний;
- блочную структуру моделей;
- необходимость описания и последующего воспроизведения неконтролируемых входных воздействий;
- необходимость рационального планирования машинных экспериментов для целей изучения реакции модели на контролируемые входные воздействия.

¹⁾ СИМОНА — система имитационного моделирования нестационарных уравнений агроэкосистем.

При этом, как уже указывалось, сложность моделируемого объекта – экосистемы – приводит к необходимости разработки не одной какой-либо модели, а набора моделей, каждая из которых должна быть специализирована на решении задач определенного класса. Имитационная модель, в отличие от концептуальной, всегда конкретна, а ее структура определяется:

- целью моделирования;
- требуемой точностью воспроизведения свойств моделируемого объекта;
- характером доступной экспериментальной информации об объекте.

Говоря о наборе моделей, мы, следовательно, предполагаем, что с его помощью будет решаться некоторая совокупность взаимосвязанных задач. Учитывая, что набор моделей совместно с управляющими и обслуживающими программами образует некоторую систему взаимодействующих моделей, мы далее будем называть этот комплекс имитационной системой (термин, введенный в [11]).

4.3.1. Иерархическая структура имитационной системы. Конструируемые модели могут иметь различное функциональное назначение. Как отмечалось в [26], модели могут использоваться как

- средство осмысления действительности;
- средство общения специалистов определенной предметной области;
- средство обучения и тренажа;
- средство целенаправленного планирования эксперимента;
- инструмент для решения задач прогноза и управления.

В некоторых случаях, разумеется, построенная модель может оказаться полезной в достижении нескольких из перечисленных целей. В то же время, как очевидно, явная формулировка цели накладывает определенные ограничения на характер разрабатываемой модели.

Можно представить себе два альтернативных пути решения задач моделирования. Первый из них и, по-видимому, наиболее быстрый заключается в разработке простых, но не связанных между собой моделей, каждая из которых решает вполне определенную задачу. Однако, учитывая, что описание процессов при этом будет дублироваться и что каждая из моделей потребует создания специальной, не связанной с другими базы данных, такой путь вряд ли можно признать эффективным. Другой подход должен быть основан на создании системы взаимодействующих друг с другом моделей, в совокупности образующих согласованный по областям применения и структуре данных имитационный комплекс. Поясним эту мысль подробнее.

Из рассмотрения структуры существующих моделей следует, что разработка реалистичной модели является весьма трудоемким делом. Поэтому было бы расточительным строить модель для решения одной какой-либо задачи. Кроме того, в начале разработки, как правило, отсутствует достоверная информация о некоторых зависимостях или об их параметрах, т.е. модель строится при известной доле неопределенности в описании некоторых блоков. Все это наводит на мысль об организации "постоянного действующего" комплекса моделей, развиваемого по мере накопления знаний о моделируемой системе и пополняемого новыми специализированными моделями, расширяющими круг решаемых прикладных задач. Имен-

но такой комплекс моделей совместно с его математическим, программным и информационным обеспечением и следует считать имитационной системой. Рассмотрим ее структуру применительно к тем задачам, которым посвящена настоящая книга.

Центральной частью системы является базовая модель экосистемы (рис. 4.6). Базовая модель конструируется "на грани понимания" происходящих в системе процессов. Поэтому целью ее построения является систематизация всех существующих представлений о моделируемом объекте,

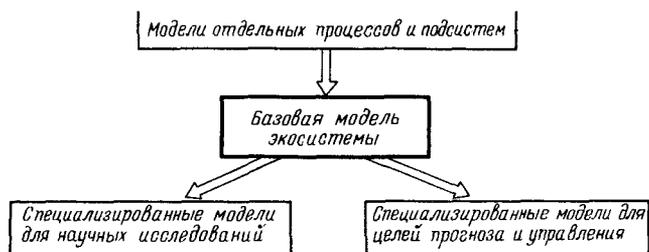


Рис. 4.6. Иерархия моделей

выявление "белых пятен", не позволяющих построить замкнутое описание, проверка гипотез о характере отдельных процессов, протекающих в системе, и т.д. Таким образом, базовую модель можно рассматривать как базу знаний о моделируемой системе. Очевидно, что эта модель не обязательно должна быть ориентирована на внешнего пользователя. Ее основным назначением является отработка основных принципов моделирования, а также оценка области применимости всех специализированных моделей. Базовая модель и поддерживающие ее программные средства относятся к внутрисистемному математическому и программному обеспечению имитационной системы (ИС).

Внешнюю часть системы составляют специализированные модели, ориентированные на внешнего пользователя и предназначенные для решения либо научно-исследовательских, либо производственных прикладных задач. Желательно, чтобы работа с этими внешними моделями допускала возможность организации диалогового режима с использованием языка и терминологии той предметной области, отображением которой являются эти модели.

Так, описанная в предыдущем параграфе динамическая модель агрофитоценоза может рассматриваться как базовая модель, поскольку при ее построении с достаточной степенью детальности учтены существующие представления о процессах, происходящих в растениях пшеницы и в среде их обитания. Специализация базовой модели может быть выполнена в двух направлениях.

Использование модели в научных исследованиях связано с организацией взаимодействия машинного и натурального эксперимента. А поскольку в агроэкологии такие эксперименты ставятся как в лабораторных (контролируемых), так и в полевых условиях, разрабатываемые для этих целей конкретные модели должны учитывать эту специфику, т.е. должны быть ориенти-

рованы на возможность воспроизведения свойств объекта в той реальной обстановке, в которой производится эксперимент. Другая ветвь специализации связана с прикладными задачами, возникающими в условиях производства. Здесь могут быть использованы во взаимодействиях с базовой моделью многочисленные полуэмпирические соотношения и регрессионные зависимости, полученные к настоящему времени и широко применяемые в агрометеорологии, почвоведении, агрохимии и т.д. При этом базовая модель может использоваться двояким образом. С одной стороны, она может служить для формирования границ применимости внешних полуэмпирических моделей, а с другой — для уточнения коэффициентов уравнений регрессии, прогнозных и иных полуэмпирических соотношений.

Обратимся теперь к вопросу о пополнении и развитии имитационной системы. Как уже отмечалось, механизмы отдельных протекающих в экосистеме процессов до конца не изучены. Часто даже не вполне ясно, в каком виде тот или иной процесс должен быть включен в общую модель. Поэтому в системе предусмотрена постановка "предмодельных" имитационных исследований (верхний уровень на рис. 4.6). На этом уровне изучаются отдельные блоки модели, при построении которых могут быть учтены как новые экспериментальные факты, так и усовершенствованные теоретические представления о тех или иных процессах. На этом же уровне может исследоваться чувствительность модели и определяться та степень детализации описания, которая согласуется с целями моделирования.

Отметим, что в данном изложении мы рассмотрели лишь иерархическую организацию самих моделей и не затронули пока вопросов из математического и программного обеспечения. В то же время описанная выше структура может считаться достаточно универсальной, так как она обеспечивает выполнение всех основных сформулированных выше целей моделирования и при этом не нарушается принцип конструирования каждой из моделей для решения вполне определенных конкретных задач.

4.3.2. Банк моделей. Рассмотренная выше структура имитационной системы позволяет перейти к обоснованию понятия банка моделей, которое мы будем связывать с информационным обеспечением задач моделирования. Действительно, как отмечалось в [11], для целей быстрой и оперативной организации серий имитационных экспериментов совокупность моделей должна быть объединена с некоторой информационной базой. Это требование следует признать обязательным во всех тех ситуациях, о которых говорилось выше. Поэтому специализированная база данных должна быть неотъемлемой частью любой имитационной системы.

Рассмотрим имитационную модель с этой точки зрения, т.е. как систему для переработки информационных потоков. Учитывая, что модель реализуется на ЭВМ, являющейся вычислителем дискретного действия, можно следующим образом схематически представить себе последовательность операций в процессе моделирования (рис. 4.7). На k -м временном шаге модели каждому из n ее блоков доступна информация, образующая вектор текущего состояния модели $x(k)$, вектор внешних воздействий и параметров модели a . Обращение к блокам производится поочередно в любой последовательности. При этом результатом вычислений, выполняемых в J -м блоке, является некоторый массив переменных $x_j(k+1)$, являющийся выходом этого блока в следующий момент времени (состояние J -го блока).

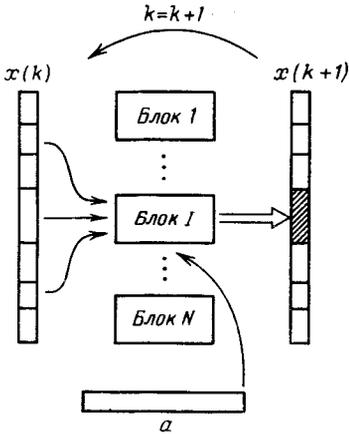


Рис. 4.7. Схема работы модели блочного типа

Состояние модели в целом (т.е. вектор x) является объединением векторов состояния всех блоков:

$$x = (x_1^*, x_2^*, \dots, x_n^*)^*$$

Поэтому результатом проработки всех блоков на данном шаге явится новое состояние модели $x(k+1)$. После вычисления этого нового состояния модельное время увеличивается на величину шага, массив $x(k+1)$ переписывается на место $x(k)$ и процесс вычисления повторяется.

Из этого схематического описания, следует, что в модели информационные потоки

разделяются на два типа. Часть из них замыкается внутри каждого блока, а другая часть передается от блока к блоку, образуя в целом "глобальное" состояние модели. Очевидно, что связь модели с внешним информационным обеспечением определяется именно этой второй частью информации (т.е. входными, выходными и межблочными потоками) и не зависит от переменных, замыкающихся в пределах каждого блока.

Поясним сказанное на описанном выше примере модели агрофитоценоза. Уравнения движения почвенной влаги являются одинаковыми практически для всех одномерных задач влагопереноса. Следовательно, реализованный в виде некоторой численной схемы блок влагообмена в почве может войти в качестве составной части в целый ряд моделей. С другой стороны, учитывая нелинейный характер этих уравнений, можно предложить несколько численных методов интегрирования исходной системы: метод обычной или потоковой прогонки с линеаризацией, метод конечных элементов, итерационный метод, использование схемы "предикатор — корректор" и др. Все эти схемы отличаются друг от друга по точности, затратам машинного времени, требованиям к размерам пространственной сетки. И при этом каждый из них обеспечивает расчет последующего состояния блока (т.е. массива влажности почвы в узлах расчетной сетки) в зависимости от входных переменных блока, заданных на предыдущем шаге. Аналогичным образом "работают" блок фотосинтеза, блок метаболизма и другие блоки модели. Следовательно, можно иметь несколько версий реализации одного и того же блока модели и производить выбор той или иной версии в процессе ее сборки, исходя из конкретного содержания решаемой задачи. Более того, многие блоки можно считать универсальными и использовать в самых разнообразных моделях. Мы уже упоминали, что блок водного режима почвы пригоден для широкого набора почв (исключение составляют так называемые "набухающие" почвы, изменяющие объемную массу при вариациях влагосодержания). Точно так же один и тот же блок микроклимата посева годится для всех сельскохозяйственных культур, посевы которых можно считать горизонтально однородными.

Учитывая изложенное, введем следующую терминологию. Б л о к о м модели будем называть программную единицу, реализующую числен-

ный метод решения некоторой математической задачи, описывающей группу функционально однородных (физических, физико-химических, биологических или экологических) процессов, происходящих в исследуемой системе. Вектор состояния блока — массив, определяющий совокупность выходных переменных данного блока в момент времени t_k . В этот массив входят переменные, необходимые для вычисления последующего состояния либо этого же, либо по крайней мере одного из других блоков, а также вспомогательные переменные, используемые несколькими блоками. Фиксация этих последних на каждом шаге модели, вообще говоря, не обязательна. Она производится для ускорения счета модели. Состояние модели — объединение векторов состояния всех блоков модели. Базовый шаг модели — интервал модельного времени между двумя последовательными состояниями модели. Отметим, что расчетный шаг модели может отличаться от базового и может меняться от блока к блоку. Это связано с возможностью дробления шага для обеспечения требуемой точности вычисления. Вектор параметров модели — набор констант, который должен быть задан для осуществления прогона модели. В отличие от переменных состояния, параметры не изменяются от шага к шагу. Начальное состояние — набор значений переменных состояния на момент начала счета. Входные воздействия — переменные, которые принудительно изменяются от шагу к шагу и являются функциями времени и, возможно, состояния модели. Входные воздействия разделяются на неконтролируемые (возмущения) и контролируемые (управления).

Нетрудно видеть, что введенные до сих пор понятия лишь незначительно отличаются от тех, которые обычно используются в теории дискретных динамических систем. Их пояснение было необходимо для того, чтобы другие термины, которые будут определены в дальнейшем, получили однозначное толкование.

Банк моделей — совокупность моделей, решающих задачи определенной предметной области и реализуемых в некоторой операционной среде на основе единого математического, программного, технического (тип и минимальная конфигурация ЭВМ) и информационного (база данных) обеспечения. Мы связываем, таким образом, понятие банка моделей с единством их информационной базы. Это представляется естественным в свете тех рассуждений о специфике имитационных систем, которые приведены в предыдущих пунктах.

Класс моделей — часть банка моделей, объединенная фиксированной структурой информационного обмена между различными блоками модели. Для каждого класса моделей может быть образована библиотека блоков. Как указывалось выше, для каждого блока в данном классе может существовать несколько версий (например, несколько подпрограмм для описания блока). Версии могут отличаться друг от друга детальностью описания внутривходовых процессов, численной схемой решения задачи или другими характеристиками. Однако все версии каждого блока одного и того же класса моделей имеют единый набор входных и выходных переменных и параметров. Модель — сужение класса моделей. Модель определяется перечислением конкретных имен блоков из библиотеки данного класса. При наличии нескольких версий в модель включа-

ется лишь одна версия для каждого блока. Для блоков модели, кроме того, могут быть указаны условия их выполнения. Реализация модели определяется заданием численных значений вектора параметров, вектора начального состояния и вектора входных воздействий. Очевидно, что непосредственная связь банка моделей с базой данных осуществляется именно на стадии реализации модели.

Проиллюстрируем введенные выше понятия на примерах имитационных моделей, рассмотренных выше. Описанная в п. 1.3.2 динамическая система образует модель зеленой дубовой листовертки. Класс моделей можно образовать, если рассматривать все основные виды насекомых-вредителей в данном лесном массиве, принадлежащих некоторому характерному типу. Объединяя эти виды по способу взаимодействия "лес – вредитель", можно получить, например, класс моделей, имитирующих динамику численности листогрызущих насекомых, насекомых-короедов и т.п. Наконец, включая в общую систему моделей блоки, учитывающие типы и разновидности почв, погодно-климатические данные и виды древостоя, и объединяя их с базой данных, мы получим банк моделей.

Описанную в п. 1.3.1 структуру ELM можно, напротив, уже считать банком моделей, поскольку она охватывает описание всех типовых разновидностей луговых биомов, а также содержит соответствующее информационное наполнение. При исследовании продукционного процесса в агроэкосистемах банк моделей составляют описания сезонной динамики возделываемых в зоне культур, совместно с моделями почвенных, погодно-климатических процессов и агротехнологий. К некоторому классу следует отнести модели культур с определенным типом морфогенеза (зерновые, травы, корнеклубнеплоды и т.д.). Описания динамики роста и развития растений конкретного вида, например пшеницы, образуют версии блока внутри одного класса. Следовательно, выбирая одну версию для каждого блока, мы можем получить, например, модель агрофитоценоза пшеницы, ячменя или другой культуры.

Очевидно, что введение понятий классов и банков моделей упорядочивает работу по созданию имитационных систем. В частности, при подобной организации блок микроклимата, например, может быть единым для всех культур сплошного сева. Это исключает необходимость дублирования при разработке программ. Однако введенные нами понятия имеют и более глубокий смысл. Опираясь на них, можно более эффективно и целенаправленно строить работу по созданию имитационных систем. Эти вопросы находят свое отражение в технологии моделирования, к описанию которой мы и переходим.

4.3.3. Технология имитационного моделирования. Необходимость в разработке специальной "технологии" имитационного моделирования возникает по двум причинам. Первая из них заключается в том, что при разработке моделей и в особенности при создании банка моделей необходимо не только решать задачи формализации описания тех или иных процессов, выбора численных методов и т.д., но и принимать обоснованные в той или иной степени решения относительно структуры данных, организации взаимодействия модели с другими программными модулями имитационной системы, обработки и представления результатов моделирования. Вторая сторона вопроса заключается в том, что разработка хорошей модели явля-

ется, как правило, коллективным делом и представляет собой длительный процесс, занимающий зачастую несколько лет. Ясно, что все этапы этого процесса должны быть согласованы.

Перейдем непосредственно к описанию технологии моделирования и прежде всего сформулируем основные этапы этого процесса, приняв за основу классификацию, предложенную в [26].

1. Конструирование модели:

1.1. Определение системы – установление границ подлежащей изучению реальной системы, ограничений и критериев эффективности ее функционирования.

1.2. Формулирование модели – переход от реальной системы к некоторой абстрактной логической схеме, определение целей моделирования и тех задач, которые должны быть решены в результате использования модели.

1.3. Обоснование структуры информационного обеспечения, анализ доступных для решения задачи данных, а также разработка внешней организации базы данных и ее концептуальной модели (т.е. глобальной логической организации базы данных).

1.4. "Трансляция" модели, т.е. описание модели на языке, воспринимаемом выбранной ЭВМ, – разработка общесистемного и прикладного математического обеспечения, разработка интерфейсных модулей для связи модели с базой данных.

2. Планирование и проведение машинных экспериментов:

2.1. Стратегическое планирование – обоснование метода последовательного осуществления серий экспериментов с целью изучения модели и оценки ее чувствительности.

2.2. Tактическое планирование – определение способа проведения очередной серии испытаний в рамках выработанной стратегии эксперимента.

2.3. Экспериментирование с моделью – проведение серий испытаний модели с целью получения желаемых данных.

3. Практическое использование модели:

3.1. Статистическая обработка результатов "прогонов" модели, качественная оценка ее динамических свойств и эксплуатационных характеристик.

3.2. Оценка адекватности модели, повышение до приемлемого уровня той степени уверенности, с которой исследователь может судить о корректности выводов, полученных на основе экспериментирования с моделью.

3.3. Интерпретация – построение выводов по данным имитационного эксперимента и их содержательное истолкование.

3.4. Реализация – практическое использование модели в той предметной области, для которой производилась разработка.

3.5. Документирование – регистрация хода разработки и его результатов, разработка описания, инструкций программисту и оператору, а также других документов, регламентированных ЕСПД.

Отметим, что содержание некоторых из перечисленных выше этапов очевидно и в комментариях не нуждается. Поэтому рассмотрим лишь те из них, которые выпадают из этой категории. Прежде всего заметим, что определение собственно модели, класса или банка моделей, а также состава

включаемых в них блоков производится на этапах 1.2 и 1.3. Имитационная модель, рассматриваемая лишь как "проекция" изучаемой сложной системы, призвана воспроизводить лишь часть из многообразных ее свойств. Структура модели определяется, исходя из целей моделирования, требуемой точности решения и доступной для использования в модели информации. В связи с этим иногда приходится отказываться от весьма привлекательного и имеющего ясную содержательную основу описания, существенно загроуляя его лишь потому, что отсутствуют экспериментальные данные, необходимые для идентификации модели. С другой стороны, даже и при наличии информации любое загроуление оправдано, если оно позволяет решить поставленную задачу с требуемой точностью. Как показывает опыт моделирования, многие разработки оказались бесполезными вследствие "переусложнения" модели. Модель в этих случаях оказывается неподкрепленной фактическим материалом, а ее эксплуатация требует неоправданно большого объема ресурсов ЭВМ.

Обратимся к этапу 1.4, основное содержание которого связано с разработкой программного обеспечения. Для сокращения сроков разработки было бы весьма желательно использовать на этом этапе стандартные средства и системы программирования существующих ЭВМ, включающие наиболее распространенные входные языки, такие, как ПЛ/1, Фортран и др. Однако, как показал опыт работы с достаточно сложными моделями, реализация имитационной системы на ЭВМ требует специальных программных средств, которые в подобных языках отсутствуют. Это приводит к необходимости разработки не только программных модулей для описания блоков модели, но и некоторого общесистемного математического и программного обеспечения. Эти вопросы подробно будут нами рассмотрены в следующей главе.

Этапы 2.1 и 2.2 являются весьма важными, поскольку от уровня решения возникающих здесь вопросов часто зависит успех разработки в целом. Мы уже говорили о том, что именно машинный эксперимент является основной чертой имитационного моделирования, отличающей его от других типов вычислений, производимых с помощью ЭВМ. Учитывая то значение, которое имеют эти этапы, их рассмотрению посвящен специальный параграф книги (§ 4.4). Здесь же лишь отметим, что разрабатываемая модель, помимо детерминированной части, может включать в себя описание тех или иных стохастических процессов, для имитации которых используются методы статистических испытаний. В этом случае на ЭВМ должны быть организованы серии "прогонов" модели и сформированы выборки, анализ которых должен производиться с использованием методов математической статистики [7]. Поэтому в имитационной системе должны быть предусмотрены процедуры генерации случайных величин с заданными законами распределения и случайных последовательностей с заданными корреляционными свойствами, а также программы для статистической обработки случайных величин и процессов [14, 15]. Эти соображения должны учитываться при организации машинных экспериментов.

Последним вопросом, заслуживающим специального внимания, является оценка адекватности модели (этап 3.2). При моделировании таких сложных объектов, какими являются экологические системы, постановка задачи оценки адекватности должна быть принципиально отличной от тех,

которые используются при моделировании, например, относительно простых по своей структуре технических систем. Многомерность, наличие сложной иерархической организации, адаптивность по отношению к внешним воздействиям и другие особенности экосистем, включая слабую изученность многих трофических связей и взаимодействий, не позволяют говорить об "адекватности" имитационной модели реальному объекту в общепринятом значении этого слова. Как же следует трактовать это понятие? Необходимо сформулировать достаточно ясный и однозначный ответ на этот вопрос, хотя бы потому, что выводы, сделанные с помощью модели, мы хотим применить к явлениям реального мира.

Адекватность модели будем оценивать точностью, которая может быть достигнута при решении той конкретной задачи (или класса задач), для которой (или для которых) разрабатывалась данная имитационная система. Так, например, если модель агрофитоценоза создается для целей управления водным режимом, то от нее в первую очередь требуется, чтобы она на разных стадиях вегетации растений правильно предсказывала их реакцию на условия засухи (водный стресс) и поливы, и именно в тех пределах вариаций погодных условий, которые характерны для данного региона. Оценку адекватности можно связать с той точностью, с которой производится расчет норм поливной воды и сроков полива. Действительно, было бы абсурдным требовать от модели, чтобы сроки полива определялись с точностью до часа и минуты. Задачу можно считать решенной, а модель адекватной реальности, если эти сроки будут определены с точностью до суток.

Этот пример показывает, что оценить адекватность модели достаточно просто в тех случаях, когда модель создается для решения задач управления.

Более затруднительной оказывается ситуация, когда модель конструируется для изучения или для уточнения плохо изученных механизмов некоторого явления. В этих случаях оценка адекватности модели всегда будет субъективной и, как указано выше, она будет отражать степень уверенности исследователя в относительной "истинности" модельных построений, которую ему удалось достичь в результате анализа серий машинных экспериментов. При этом, разумеется, следует оценить тот относительный вклад в общую картину поведения модели, который внесли ее составляющие, основанные как на вполне надежно установленных, так и на гипотетических механизмах.

В настоящем параграфе мы кратко охарактеризовали основные особенности технологии имитационного моделирования. Чтобы говорить о создании такой технологии в полном смысле этого слова, следовало бы сказать о программных средствах, обеспечивающих автоматизацию решения поставленных задач. Их описание содержится в гл. 5, в которой наряду с другими вопросами рассмотрена специфика технологической поддержки основных этапов (этапы 1.3 – 3.2) имитационного моделирования.

Однако прежде чем переходить к этим вопросам, следует рассмотреть особенности планирования имитационных экспериментов.

§ 4.4. Планирование имитационного эксперимента

Задачи, возникающие при исследовании сложных систем средствами имитационного моделирования, во многом аналогичны традиционным задачам общей теории динамических систем. Это исследование устойчивости, анализ чувствительности модели к вариациям ее параметров, идентификация и, наконец, решение разнообразных задач управления. В то же время особенности сложных систем, ряд из которых отмечался выше, приводят к необходимости разработки специальных методов исследования, учитывающих специфику процесса имитации. К этим особенностям относятся [24]:

- большая размерность пространства состояний, достигающая нескольких сотен и тысяч переменных;
- начальная неопределенность в описании отдельных процессов (блоков модели);
- наличие большого числа неопределенных параметров, подлежащих идентификации в процессе машинного эксперимента;
- наличие наряду с детерминированными многомерных случайных воздействий на входе системы и др.

Все это свидетельствует о том, что получение одной или нескольких машинных реализаций траектории модели вряд ли позволит выяснить характер изучаемой системы и особенности ее динамического поведения. Большая размерность системы не оставляет надежд на возможность использования традиционных методов синтеза систем, например методов оптимального управления. В связи с этим единственный реалистичный подход к задачам анализа и синтеза сложных систем должен быть основан на сочетании метода экспертных оценок с планированием и проведением машинных экспериментов [7, 15, 26]. При этом сам вычислительный процесс должен рассматриваться как некоторый факторный эксперимент с имитационной моделью, протекающий в условиях неопределенности, вызванной наличием случайных входных сигналов.

Под экспериментом принято понимать создание некоторого комплекса условий U , в результате которых могут наступать (или не наступать) события из некоторого заданного множества V [4]. Если на результаты эксперимента оказывают влияние случайные события той или иной вероятностной природы, то этот эксперимент называется *статистическим*.

Математическая теория планирования эксперимента, начиная с 30-х годов двадцатого века, развивалась применительно к изучению натуральных, реально существующих в природе систем и объектов. В последние годы в теории планирования эксперимента возникло новое направление, целью которого является изучение не самого объекта, а его математической (машинной) модели. Очевидно, что в этой области могут быть использованы все достижения общей теории. И все же следует учитывать то, что имитационный эксперимент во многом отличается от натурального. Так, при машинной имитации мы также часто воспроизводим случайные события. Однако механизм воспроизведения на ЭВМ программным путем "случайных" чисел не может быть, строго говоря, признан вероятностным, хотя последовательности этих чисел обладают многими свойствами, присущими случайным числам. Имитационный эксперимент имеет и другие особенности, о которых и пойдет речь в дальнейшем.

4.4.1. Понятие и цели планирования эксперимента. Под экспериментом понимается некоторый дискретный акт, который может многократно повторяться и в результате которого исследователь получает определенные сведения об интересующем его объекте. Последовательность действий экспериментатора можно условно разбить на три этапа: (1) определение условий проведения эксперимента; (2) реализация эксперимента, сопровождающаяся некоторыми измерениями характеристик объекта; (3) обработка полученных данных и анализ результатов.

Условия эксперимента обычно зависят от одного или нескольких факторов, которые исследователь может выбирать из некоторого множества по своему усмотрению. В таком случае эксперимент называется активным, а сами эти факторы – контролируемые переменными. Далее мы будем иметь дело только с активными экспериментами и поэтому слово "активный" часто будем опускать. Каждый отдельный фактор может иметь количественную или качественную природу. Примерами качественных факторов являются марки различных материалов, из которых изготовлена некоторая деталь, различные технологии и т.д. Для качественных факторов не существует понятия "расстояния", а введение метрики лишено смысла. Качественные факторы можно обычно рассматривать как принадлежащие некоторому метрическому пространству. В ранних работах предполагалось, что это пространство конечномерно. Однако в последние годы появились работы, в которых теория планирования эксперимента распространяется на тот случай, когда контролируемые переменные являются элементами функциональных пространств [4, 24, 25]. В дальнейшем мы будем обозначать область изменения контролируемых переменных через Σ , введя для элемента этой области обозначение x . В частности, в конечномерном случае Σ может ассоциироваться с частью евклидова пространства, $\Sigma \subset R^n$, а x – с некоторым n -мерным вектором, $x = (x_1, x_2, \dots, x_n)$. В каждой конкретной реализации эксперимента контролируемые переменные фиксируются на определенном уровне $x^{(i)}$ (т.е. располагаются в некоторых точках факторного пространства), так что условия эксперимента в целом характеризуются совокупностью величин $x^{(1)}, x^{(2)}, \dots, x^{(N)}$ и числом опытов N . При этом число опытов N может заранее не фиксироваться и a priori быть неограниченным, а величины $x^{(i)}$ не все могут быть различны.

Рассмотрим частный случай эксперимента, так называемый "факторный эксперимент". Пусть Σ – некоторая область евклидова пространства R^n . "Классическая" схема эксперимента caeteris paribus¹⁾ строится следующим образом. Все, кроме одного, факторы фиксируются на определенных уровнях, а один из них, например x_i , изменяется, принимая значения $x_i^{(1)}, x_i^{(2)}, \dots, x_i^{l_i}$, которые называются уровнями значений фактора. Затем поочередно варьируются все остальные факторы. В факторной схеме, напротив, осуществляется полный перебор, т.е. все уровни одного фактора комбинируются со всеми возможными уровнями всех остальных. Таким образом, постановка факторного эксперимента осуществляется в

$$N = l_1 l_2 \dots l_n$$

¹⁾Caeteris paribus (лат.) – "при прочих равных", означает поочередное изменение факторов.

точках пространства контролируемых переменных. Число и расположение точек в факторном пространстве как раз и называется "планом факторного эксперимента".

Отметим, что понятие факторного эксперимента естественным образом распространяется на случай качественных факторов. Только в этом случае мы говорим не об уровнях, а о градациях соответствующего фактора, предполагая, естественно, что число таких градаций конечно.

Обратимся ко второму этапу, т.е. к реализации эксперимента. В процессе реализации эксперимента производится измерение интересующих исследователя характеристик объекта, т.е. в каждой из N точек фиксируется некоторый результат $y^{(i)}$. Отметим, что переменная y может быть скалярной, векторной величиной, а может, так же как и контролируемые переменные, принадлежать пространству более сложной природы. В статистическом эксперименте значение $y^{(i)}$ является реализацией некоторой случайной величины или случайного процесса. Каковы источники этой случайности?

Известно, что любое измерение сопровождается ошибками. В простейшем случае ошибки измерений могут рассматриваться как аддитивные случайные величины с заданными корреляционными свойствами. С другой стороны, сам объект может иметь стохастическую природу. Так, например, при исследовании естественных экосистем мы производим наблюдения на фоне изменяющихся погодных условий, которые оказывают заметное влияние на состояние экосистемы. Если этот эксперимент повторяется в течение нескольких лет, то складывающиеся в каждом году метеорологические условия могут рассматриваться как реализации многомерного случайного процесса. При моделировании агроэкосистем "проигрывание" модели должно также производиться на фоне изменяющихся погодных условий. Очевидно, что в развитой модели погода должна имитироваться как некоторый случайный процесс с характерными для данного климатического региона вероятностными характеристиками.

Пусть ω — элементарное событие из некоторого вероятностного пространства Ω . Тогда результаты эксперимента в точке $x^{(i)}$ факторного пространства можно представить себе как реализацию случайной функции

$$y^{(i)} = A(x^{(i)}, \omega) \quad (i = \overline{1, N}), \quad (4.4.1)$$

где A — некоторый оператор, действующий из $\Sigma \times \Omega$ в Y .

Перейдем к рассмотрению содержания третьего этапа. В стохастическом эксперименте анализ и обработка данных должны производиться, естественно, с использованием статистических методов. При этом можно выделить два подхода. Первый из них основан на методах статистической проверки гипотез и дает качественный ответ при качественных предположениях о виде модели (4.4.1). Стандартный аппарат, который здесь используется, — дисперсионный анализ. Поэтому эксперименты такого типа можно назвать "дисперсионными". При втором подходе ставится задача количественной оценки входящих в модель неизвестных параметров на основании экспериментальных данных, полученных в условиях неопределенности. В общей теории динамических систем такие задачи ассоциируются с понятием параметрической идентификации моделей. Поскольку в этих случаях используются методы регрессионного анализа, эти эксперименты получили название регрессионных [25].

Задача дисперсионного эксперимента – получить ответ на вопрос о влиянии контролируемых переменных на "выход" объекта y . Точнее говоря, следует оценить значимость так называемых "главных эффектов" факторов и их "взаимодействий" [7]. При этом каждый из факторов в эксперименте может принимать конечное число градаций (уровней), а сами контролируемые переменные могут иметь как количественный, так и качественный характер.

При постановке регрессионного эксперимента предполагается, что выход объекта зависит, помимо контролируемой переменной x , от некоторого вектора параметров θ , значение которого заранее неизвестно. Основная модель регрессионного эксперимента имеет вид [25]

$$M\{y^{(i)} | x^{(i)}\} = \eta(x^{(i)}, \theta), \quad (4.4.2)$$

где $y^{(i)}$ – результат измерения в точке $x^{(i)}$, $\eta(x, \theta)$ – функционал, вид которого известен, а $\theta = (\theta_1, \theta_2, \dots, \theta_m)^*$ – вектор неизвестных параметров. Оператор $M\{\dots\}$ соответствует математическому ожиданию. Целью анализа в данном случае является определение оценок неизвестных параметров θ или оценки самой "поверхности отклика" $\eta(x, \theta)$ по результатам наблюдений (4.4.1) в некоторой области изменения контролируемых переменных Σ . Наиболее развиты и хорошо обоснованы методы получения линейных оценок для тех случаев, когда неизвестные параметры входят в модель линейно:

$$M\{y | x\} = \theta^* f(x), \quad (4.4.3)$$

где $f(x) = (f_1(x), f_2(x), \dots, f_m(x))^*$ – известные функции.

Сформулируем понятие плана эксперимента. Для этого напомним, что исследователь может по своему усмотрению назначать число и расположение контролируемых переменных $x^{(i)}$ в области планирования Σ . При этом в каждой точке $x^{(i)}$ эксперимент может проводиться несколько раз, образуя так называемые "повторности", число которых принято обозначать через r_i . П л а н о м эксперимента E будем называть совокупность величин

$$(x^{(1)}, r_1), (x^{(2)}, r_2), \dots, (x^{(l)}, r_l). \quad (4.4.4)$$

При этом общее число экспериментальных точек равно $N = \sum_{i=1}^l r_i$. Таким образом, план эксперимента характеризуется числом точек в области планирования N , их расположением $x^{(i)}$ и числом повторностей в каждой точке r_i .

Очевидно, что в активном эксперименте эти величины могут изменяться. Задача теории планирования эксперимента как раз и заключается в разработке методов, позволяющих определять рациональное размещение точек плана, исходя из некоторых критериев качества эксперимента.

Перейдем к обсуждению этих критериев. Постановка единичного (т.е. без повторностей) опыта в каждой точке множества Σ требует затрат определенных ресурсов, независимо от того, идет ли речь о натуральных или об имитационных экспериментах. С другой стороны, результат опыта в этой точке обладает той или иной информативностью и вносит определенный вклад в общую оценку модели. Это позволяет построить количественный критерий эффективности эксперимента. Например, в эксперименте по про-

верке статистических гипотез можно, задавая уровень значимости, принять в качестве такого критерия общие материальные затраты на проведение эксперимента в некотором стоимостном выражении. Эксперимент можно считать оптимальным, если план эксперимента (4.4.4) доставляет этой величине при заданном уровне доверительной вероятности минимальное значение. В машинном эксперименте, особенно при моделировании сложных систем, важным показателем является суммарное время вычислений. Поэтому здесь в качестве критерия может выступать показатель общего времени проведения эксперимента. В регрессионных экспериментах нас часто интересует качество оценок параметров, мерой чего может явиться дисперсионная матрица оценок. Учитывая это, можно сформулировать ряд критериев, являющихся выпуклыми функционалами от дисперсионной матрицы. С этими соображениями связано, например, введение понятий D -оптимальных и A -оптимальных планов [37].

Можно представить себе, что эффективность эксперимента оценивается критерием $J(e)$, где e — некоторый конкретный план, принадлежащий множеству планов (4.4.4), $e \in E$. Тогда оптимальный план e^* определяется как

$$e^* = \operatorname{argmin} J(e) \quad (4.4.5)$$

при некоторых ограничениях, наложенных на другие показатели эффективности эксперимента. В общем случае может быть сформулировано несколько критериев и выбор плана может осуществляться путем векторной оптимизации (планы, оптимальные по Парето; см. [4]). Таким образом, теория планирования эксперимента является некоторым новым междисциплинарным направлением, использующим методы как математической статистики, так и теории оптимального управления.

Ясно, что во многих случаях получить точное решение задачи оптимизации (4.4.5) затруднительно. В таких ситуациях при выборе плана эксперимента следует исходить из эвристических соображений. Иллюстрация этого подхода будет дана ниже.

Отметим в заключение основные особенности имитационных экспериментов, отличающие их от экспериментов, производимых с натурными объектами¹⁾. Это, во-первых, то обстоятельство, что имитационный эксперимент производится в ускоренном масштабе времени, так что в разумные сроки можно воспроизвести достаточно большое число "прогонов" модели. Фактор времени здесь является решающим, так как в противном случае, т.е. если эксперимент на ЭВМ занимает больше времени, чем в натуре, он вряд ли целесообразен. С этим связано также то, что число экспериментальных точек N может быть выбрано значительно большим, чем при исследовании исходной системы. А поскольку ясно, что выгоды, получаемые от оптимизации планирования эксперимента, должны быть тем больше, чем больше его объем, следует ожидать высокой эффективности от применения методов планирования имитационных экспериментов.

Второе обстоятельство, находящееся в некотором противоречии с первым, связано с большой размерностью модели и с большим числом подде-

¹⁾ Здесь не рассматривается тот очевидный случай, когда проведение активного натурального эксперимента невозможно, и следует ограничиться пассивными наблюдениями.

жащих варьированию факторов. Это говорит о том, что достаточно полное исследование модели в некотором однократном эксперименте вряд ли возможно. Следовательно, надо развивать методологию последовательного планирования эксперимента, уточняя в каждой серии полученные ранее оценки и дополняя их новыми. Обоснование такого подхода будет приведено в дальнейшем. Здесь реализуется то разделение имитационного процесса на "стратегическое" и "тактическое" планирование, о котором говорилось выше.

Можно подумать (и не без оснований), что методология последовательного планирования применима и к натурным экспериментам. Однако там она имеет ряд ограничений. Одно из них — это фактор времени. В машинном эксперименте легко зафиксировать в памяти результаты предыдущих серий, "доставив" план эксперимента в последующих реализациях и вовлекая в обработку на каждом шаге те результаты текущего и прошлых экспериментов, которые необходимо использовать для уточнения оценок. Другое обстоятельство и, быть может решающее связано с тем, что имитируемая в ЭВМ случайность основана на использовании так называемых "псевдослучайных" чисел (см. приложение 2). Возвращая начальное состояние генерирующей программы в исходное начальное состояние, можно получить в новой серии эксперимента ту же реализацию "случайного" процесса, что и в предыдущей. Это принципиально невозможно, например, в полевом опыте. Погодные условия год от года изменяются. И при экспериментальном исследовании природных экосистем (в том числе и агроэкосистем) необходимо сразу же ставить опыт в полном объеме, поскольку дожидаться даже "похожего" года вряд ли удастся.

Вернемся еще раз к вопросу о размерности. Хотя, как уже говорилось, размерность имитационных моделей велика, однако часто это лишь результат нашего незнания. Действительно, в имитационную модель агроэкосистемы, например, входит несколько десятков неизвестных параметров, характеризующих интенсивность энерго- и массообмена в системе, скорость фотосинтеза, кинетику биологических реакций в растениях и т.д. В то же время наш опыт говорит о том, что лишь небольшое число из них оказывает действительно решающее воздействие на продукционный процесс. Влияние большинства из этих параметров при их варьировании в разумных пределах проявляется на уровне "шумов", вызванных ошибками дискретизации модели и неучтенными факторами. Поэтому одной из задач имитационного моделирования является установление набора значимых факторов. Это приводит к необходимости постановки так называемых "отсеивающих" экспериментов.

Наконец, последняя особенность имитационного эксперимента связана с тем, что здесь мы имеем дело как правило, только с количественными факторами. Разумеется, можно представить себе эксперимент, связанный, например, с оценкой влияния на конечный результат версии того или иного блока (например, влияния способа дискретизации модели и численной схемы решения краевой задачи). Однако этот случай следует отнести, пожалуй, к экзотическим. Реально нас интересует влияние тех или иных величин на выход какого-либо блока модели или модели в целом. Вследствие этого в теории машинного эксперимента следует вести рассмотрение лишь количественных факторов.

Перечислим тот порядок, которого, по нашему мнению, следует придерживаться при организации машинного эксперимента.

1. Проведение отсеивающих экспериментов и выделение значимых факторов. Этот этап является, пожалуй, начальным. Необходимость его выполнения возникает сразу же после создания любой достаточно сложной модели. В теории динамических систем аналогичные вопросы решаются путем анализа чувствительности.

2. Постановка серий экспериментов по проверке статистических гипотез. На этой стадии уточняется влияние той совокупности факторов, которая выявлена на этапе 1.

3. Проведение регрессионных экспериментов по оценке влияния неизвестных параметров на отклик модели. В этой серии экспериментов, также дополняющей этап 1, продолжается изучение чувствительности модели. Эти результаты могут быть непосредственно использованы в задачах идентификации и управления.

4. Организация экстремальных экспериментов на ЭВМ (экспериментальное определение максимума или минимума функции отклика).

Об этом последнем этапе следует поговорить особо. В начале этого параграфа мы уже упоминали о том, что применение общих методов оптимизации при решении задач управления с использованием имитационных моделей практически невозможно. Это во многом связано с "проклятием размерности". "Сокращение" размерности может быть осуществлено двояким образом. Это либо построение сравнительно простых полуэмпирических моделей, либо так называемая экспериментальная оптимизация. В первом случае на основе серий регрессионных экспериментов (этап 3) строятся локальные аппроксимирующие модели, для которых указываются области применимости. Эти локальные модели, составляющие "внешнее" математическое обеспечение имитационной системы, используются уже в некоторой другой системе для выработки управленческих решений. Другой способ — решение задачи "в лоб" — формулировка цели управления в виде некоторой целевой функции и использование имитационной модели как "черного ящика", реакции которого на внешние управляющие воздействия могут изучаться экспериментально. В этом случае мы как раз и приходим к задаче планирования экстремальных экспериментов.

Далее в этом параграфе будут рассмотрены основные особенности планирования экспериментов в задачах имитационного моделирования. Рассмотрим первоначально дисперсионные эксперименты.

4.4.2. Планирование дисперсионных экспериментов. В дисперсионном эксперименте изучается вопрос о влиянии контролируемых переменных на выход модели. Точнее говоря, при условии правильно спланированного эксперимента мы можем получить ответ на вопрос о том, "значимы" ли влияния каждого из варьируемых в эксперименте факторов в отдельности и влияния комбинаций факторов. Основной метод анализа результатов эксперимента — дисперсионный анализ, который более точно следовало бы назвать "анализом дисперсий"¹⁾.

¹⁾ Точным переводом введенного Р.А. Фишером в 1925 году термина "variance analysis" следует считать "анализ дисперсий", однако в отечественной литературе прочно закрепился менее точно отражающий существо дела термин "дисперсионный анализ".

Основная модель дисперсионного анализа имеет вид [27]

$$y_i = x_{i1}\beta_1 + x_{i2}\beta_2 + \dots + x_{il}\beta_l + \epsilon_i \quad (i = \overline{1, N}), \quad (4.4.6)$$

где y_i — результат эксперимента в некоторой точке области планирования, ϵ_i — ошибки, а N — общее число наблюдений. Величины $\{\beta_j\}$ являются некоторыми идеализированными характеристиками изучаемой системы. Получение статистических выводов относительно этих величин, а также относительно ошибок $\{\epsilon_i\}$ и является целью дисперсионного анализа. Источниками случайных величин ϵ_i могут быть ошибки "измерения", а может быть, и стохастический характер самой исследуемой системы. Наименьшее предположение об этих величинах состоит в том, что их математические ожидания равны нулю:

$$M\{\epsilon_i\} = 0 \quad (i = \overline{1, N}), \quad (4.4.7)$$

дисперсии одинаковы, а корреляции отсутствуют:

$$M\{\epsilon_i\epsilon_j\} = \sigma^2 \delta_{ij}. \quad (4.4.8)$$

Кроме того, для получения более содержательных статистических выводов приходится ограничивать рассмотрение случаем нормального распределения, $\epsilon_i \subset N(0, \sigma)$. Следует оговориться, что в случае, если не удовлетворяются предположения о нормальности и о равенстве дисперсий, справедливость статистических выводов практически не нарушается [27]. В то же время не существует хорошей теории, учитывающей нарушение предположения (4.4.8) о независимости ϵ_i .

Как видно из (4.4.6), неизвестные эффекты β_j входят в модель аддитивно. Однако это — лишь следствие выбранного метода анализа, а не ограничение системы. Никаких предположений о линейности влияния контролируемых переменных на выход системы y при использовании дисперсионного анализа не делается, хотя знание вида исходной системы, разумеется, всегда полезно при содержательной трактовке полученных результатов. Наконец, отметим еще одно существенное обстоятельство. Модель дисперсионного анализа (4.4.6) внешне аналогична модели, используемой в регрессионном анализе. Эта аналогия широко используется. В то же время имеется и существенное отличие. В дисперсионном анализе коэффициенты x_{ij} — это определяемые планом эксперимента целые числа, равные обычно 0 или 1 (а в некоторых случаях — 1 и +1). Эти величины имеют смысл "переменных-указателей", значение которых свидетельствует о присутствии или отсутствии действия различных факторов β_j в эксперименте с номером i . Таким образом, матрица коэффициентов $X = (x_{ij})$ — это (0,1)-матрица или (-1,1)-матрица, имеющая N строк и l столбцов. Каждая из строк $(x_{i1}, x_{i2}, \dots, x_{il})$ матрицы X задает структуру i -го опыта в серии из N опытов, определяющих данный эксперимент в целом.

Рассматривая (4.4.6) как матричное соотношение

$$y = X\beta + \epsilon, \quad (4.4.9)$$

можно получить оценки наименьших квадратов неизвестных величин

$$\hat{\beta} = (X^*X)^{-1}X^*y \quad (4.4.10)$$

и оценку дисперсии ошибки (при $N > l$)

$$\hat{\sigma}^2 = \frac{S_{\min}^2}{N-l}, \quad (4.4.11)$$

где S_{\min}^2 — остаточная сумма квадратов:

$$S_{\min}^2 = \min_{\beta} \|y - X\beta\|^2. \quad (4.4.12)$$

Эти соотношения полезны, поскольку они указывают на аналогию дисперсионного и регрессионного анализа.

а) Трехфакторный анализ. Проиллюстрируем основные идеи метода на примере трехфакторного эксперимента. Хотя применение дисперсионного анализа распространяется в равной степени как на количественные, так и на качественные входные воздействия, мы здесь для определенности ограничимся случаем количественных переменных. Это ограничение несущественно и принято лишь потому, что ниже будет рассмотрена специфика планирования машинных экспериментов, где все факторы практически могут быть выражены в количественном виде.

Итак, мы имеем трехмерное пространство контролируемых переменных Σ . Эти переменные принято обозначать большими латинскими буквами, т.е., например, A , B и C . Будем предполагать, что для каждой из переменных зафиксировано конечное число уровней. Проводя через уровни каждой координатной оси плоскости, перпендикулярные двум другим осям, получим некоторую решетку. Узлы этой решетки, соответствующие комбинациям всех возможных уровней трех факторов, и являются теми потенциальными точками (или, как говорят, "ячейками") факторного пространства, в которых следует ставить эксперимент. Факторный эксперимент называется *полным*, если совокупность всех опытов занимает все возможные полученные таким образом ячейки.

Обозначим через I , J и K соответственно число уровней факторов A , B и C . Пусть M означает число повторений опыта в каждой из ячеек (M обычно называют числом реплик или повторностей опыта). В таком случае модель факторного эксперимента имеет вид

$$y_{ijkm} = \eta_{ijk} + \epsilon_{ijkm}, \quad (4.4.13)$$

где $i = \overline{1, I}$, $j = \overline{1, J}$, $k = \overline{1, K}$, $m = \overline{1, M}$, а

$$\eta_{ijk} = \mu + \alpha_i^A + \alpha_j^B + \alpha_k^C + \alpha_{ij}^{AB} + \alpha_{ik}^{AC} + \alpha_{jk}^{BC} + \alpha_{ijk}^{ABC} \quad (4.4.14)$$

Поясним смысл введенных обозначений. Величина η_{ijk} , как очевидно, представляет собой математическое ожидание отклика модели:

$$\eta_{ijk} = M\{y_{ijkm}\}.$$

Другие переменные носят названия "генерального среднего", "главных эффектов" и "взаимодействий". Для их определения воспользуемся следующим стандартным обозначением. Будем заменять точкой целочисленную переменную (индекс), по которой произведено осреднение, например,

$$\eta_{ij.} = \frac{1}{K} \sum_{k=1}^K \eta_{ijk},$$

$$\eta_{...} = \frac{1}{IJK} \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K \eta_{ijk}.$$

Тогда, в частности, имеют место равенства

$$\begin{aligned} \mu &= \eta_{\dots}, \quad \alpha_i^A = \eta_{i..} - \eta_{\dots} = \eta_{i..} - \mu, \\ \alpha_{ij}^{AB} &= \eta_{ij.} - \eta_{i..} - \eta_{.j.} + \eta_{\dots}, \\ \alpha_{ijk}^{ABC} &= \eta_{ijk} - \eta_{ij.} - \eta_{i.j} - \eta_{.jk} + \eta_{i..} + \eta_{.j.} + \eta_{..k} - \eta_{\dots}. \end{aligned} \quad (4.4.15)$$

При этом для всех значений i, j, k выполняются условия

$$\begin{aligned} \alpha_i^A &= \alpha_j^B = \alpha_k^C = 0, \\ \alpha_{i.}^{AB} &= \alpha_{.j}^{AB} = \alpha_{.j.}^{BC} = \alpha_{.k}^{BC} = \alpha_{i.}^{AC} = \alpha_{.k}^{AC} = 0, \\ \alpha_{ij.}^{ABC} &= \alpha_{i.k}^{ABC} = \alpha_{.jk}^{ABC} = 0. \end{aligned} \quad (4.4.16)$$

Смысл генерального среднего μ не нуждается в комментариях. Главный эффект i -го уровня A определяется как осредненное по всем j и k превышение отклика модели на этом уровне над генеральным средним. Из определения (4.4.15) следует, что суммы главных эффектов всех факторов равны нулю:

$$\sum_{i=1}^I \alpha_i^A = \sum_{j=1}^J \alpha_j^B = \sum_{k=1}^k \alpha_k^C = 0,$$

т.е. справедливость первого из соотношений (4.4.16). Если рассматривать главный эффект i -го уровня фактора A специально по отношению к j -му уровню фактора B , т.е. $\eta_{ij.} - \eta_{.j.}$, и вычислить отклонение этого эффекта от своего среднего $\eta_{i..} - \eta_{\dots}$, то мы получим парное взаимодействие i -го уровня A и j -го уровня B : $\alpha_{ij}^{AB} = (\eta_{ij.} - \eta_{.j.}) - (\eta_{i..} - \eta_{\dots})$, которое, как очевидно, симметрично относительно A и B . В парных взаимодействиях, так же как и в тройном взаимодействии α_{ijk}^{ABC} , проявляется возможная нелинейность исходной модели.

Существенно заметить, что вид модели (4.4.13) никак не связан со свойствами исследуемого объекта или его имитационной модели (если, разумеется, не учитывать того, что в линейном случае все взаимодействия факторов равны нулю). С этим обстоятельством связана универсальность метода, его применимость к широкому классу систем. При этом представление отклика в виде (4.4.13)–(4.4.15), как мы увидим в дальнейшем, позволяет сформулировать простые и удобные оценки влияния контролируемых переменных на выход модели. Оказывается, что введение новых переменных, а именно генерального среднего, главных эффектов и взаимодействий, приводит к разложению исходного пространства $\{\eta_{ijk}\}$ в прямую сумму ортогональных подпространств. Действительно, рассмотрим выражение

$$\sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K \sum_{m=1}^M (y_{ijkm} - \eta_{ijk})^2. \quad (4.4.17)$$

М.н.к.-оценки η_{ijk} , полученные минимизацией (4.4.17), равны

$$\hat{\eta}_{ijk} = y_{ijk}. \quad (4.4.18)$$

Отсюда в соответствии с теоремой Гаусса – Маркова следуют соотношения

$$\begin{aligned} \hat{\mu} &= y \dots, \\ \hat{\alpha}_i^A &= y_{i\dots} - y \dots, \\ \hat{\alpha}_{ij}^{AB} &= y_{ij\dots} - y_{i\dots} - y_{.j\dots} + y \dots, \\ \hat{\alpha}_{ijk}^{ABC} &= y_{ijk} - y_{ij\dots} - y_{i.j\dots} - y_{.jk\dots} + y_{i\dots} + y_{.j\dots} + y_{\dots k} - y \dots \end{aligned} \quad (4.4.19)$$

и аналогичные формулы для других взаимодействий, т.е. оценки, являющиеся линейными комбинациями исходных наблюдений y_{ijkm} .

Рассмотрим линейное векторное пространство, натянутое на векторы y_{ijkm} . Его размерность совпадает с общим числом наблюдений $N = IJKM$. Используя теорему Кокрана [39], можно показать, что в новых переменных – линейных оценках (4.4.19) это пространство распадается на девять ортогональных друг другу подпространств $L_\mu, L_A, L_B, \dots, L_{ABC}, L_e$, как это показано в табл. 4.1. Общая сумма квадратов (4.4.17) в силу ортогональности этих подпространств разбивается на ряд слагаемых

$$\begin{aligned} SS_{\text{общ}} &= \sum_{i,j,k,m} (y_{ijkm} - y \dots)^2 = \\ &= \sum_{i,j,k,m} y_{ijkm}^2 - IJKM y^2 \dots = SS_A + SS_B + SS_C + \\ &+ SS_{AB} + SS_{AC} + SS_{BC} + SS_{ABC} + SS_e, \end{aligned}$$

где, в частности,

$$\begin{aligned} SS_A &= JKM \sum_i (\hat{\alpha}_i^A)^2, \\ SS_{AB} &= KM \sum_{i,j} (\hat{\alpha}_{ij}^{AB})^2, \\ SS_{ABC} &= M \sum_{i,j,k} (\hat{\alpha}_{ijk}^{ABC})^2, \\ SS_e &= \sum_{i,j,k,m} (y_{ijkm} - y_{ijk})^2. \end{aligned} \quad (4.4.21)$$

Слагаемые в правой части (4.4.20) не все являются независимыми, поскольку они связаны этим соотношением. Однако любые два из них

Таблица 4.1

Ортогональные подпространства в трехфакторном анализе

Подпространство	Порождающие элементы	Размерность	Подпространство	Порождающие элементы	Размерность
L_μ	$\hat{\mu} = y \dots$	1	L_{BC}	$\hat{\alpha}_{11}^{BC}, \dots, \hat{\alpha}_{JK}^{BC}$	$(J-1)(K-1)$
L_A	$\hat{\alpha}_1^A, \dots, \hat{\alpha}_I^A$	$I-1$	L_{AC}	$\hat{\alpha}_{11}^{AC}, \dots, \hat{\alpha}_{IK}^{AC}$	$(I-1)(K-1)$
L_B	$\hat{\alpha}_1^B, \dots, \hat{\alpha}_J^B$	$J-1$	L_{ABC}	$\hat{\alpha}_{11}^{ABC}, \dots, \hat{\alpha}_{IJK}^{ABC}$	$(I-1)(J-1)(K-1)$
L_C	$\hat{\alpha}_1^C, \dots, \hat{\alpha}_K^C$	$K-1$	L_e	$\{y_{ijkm} - y_{ijk}\}$	$IJK(M-1)$
L_{AB}	$\hat{\alpha}_{11}^{AB}, \dots, \hat{\alpha}_{IJ}^{AB}$	$(I-1)(J-1)$			

могут быть выбраны для сравнения оценок дисперсии ошибки эксперимента σ^2 в совокупности с теми или иными эффектами. Так, поделив SS_e на соответствующее число степеней свободы (размерность подпространства L_e), равное $IJK(M-1)$, и применив к полученному выражению операцию математического ожидания, найдем

$$M\{ES_e\} = M\left\{\frac{SS_e}{IJK(M-1)}\right\} = \sigma^2. \quad (4.4.22)$$

Аналогично, для SS_A , например, получим

$$M\{ES_A\} = M\left\{\frac{SS_A}{I-1}\right\} = \sigma^2 + JK\sigma_A^2, \quad (4.4.23)$$

где через σ_A^2 обозначена сумма:

$$\sigma_A^2 = (I-1)^{-1} \sum_{i=1}^I (\hat{\alpha}_i^A)^2.$$

Обозначая далее

$$\sigma_{AB}^2 = (I-1)^{-1}(J-1)^{-1} \sum_{i=1}^I \sum_{j=1}^J (\hat{\alpha}_{ij}^{AB})^2, \quad (4.4.24)$$

$$\sigma_{ABC}^2 = (I-1)^{-1}(J-1)^{-1}(K-1)^{-1} \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K (\hat{\alpha}_{ijk}^{ABC})^2 \quad (4.4.25)$$

и вводя аналогичные обозначения для σ_{BC}^2 и σ_{AC}^2 , получим разложение общей дисперсии на составляющие, которые представлены в стандартной таблице трехфакторного эксперимента (табл. 4.2). Это позволяет использовать для проверки гипотез о значимости соответствующих эффектов стандартную статистическую процедуру, основанную на использовании F -критерия Фишера.

Пусть ошибки опыта ϵ_{ijkm} независимы и распределены нормально с нулевым средним и неизвестной дисперсией σ^2 ¹⁾. Пусть также справедлива гипотеза

$$H: \alpha_1^A = \alpha_2^A = \dots = \alpha_I^A = 0, \quad (4.4.26)$$

т.е. главный эффект фактора A отсутствует, а $\sigma_A^2 = 0$. Тогда в соответствии с табл. 4.2 мы имеем две независимые оценки дисперсии σ^2 :

$$\hat{\sigma}_I^2 = \frac{SS_A}{I-1} \quad (4.4.27)$$

и

$$\hat{\sigma}_{II}^2 = \frac{SS_e}{IJK(M-1)}. \quad (4.4.28)$$

Их отношение, т.е. $\hat{\sigma}_I^2/\hat{\sigma}_{II}^2$, подчиняется F -распределению Фишера с $(I-1)$ степенью свободы числителя и $IJK(M-1)$ степенью свободы знаменателя.

¹⁾ Отметим, что только здесь мы впервые используем предположение о нормальности ошибок. Все предыдущие выкладки справедливы при любом распределении величин ϵ_{ijkm} .

Таблица 4.2
Трехфакторный анализ с M повторностями

Источники дисперсии	Суммы квадратов	Степени свободы	$M\{ES\}$
Главные эффекты A	$SS_A = JKM \sum_i (\alpha_i^A)^2$	$I - 1$	$\sigma^2 + JKM \sigma_A^2$
Главные эффекты B	$SS_B = IKM \sum_j (\alpha_j^B)^2$	$J - 1$	$\sigma^2 + IKM \sigma_B^2$
Главные эффекты C	$SS_C = IJM \sum_k (\alpha_k^C)^2$	$K - 1$	$\sigma^2 + IJM \sigma_C^2$
Взаимодействия AB	$SS_{AB} = KM \sum_i \sum_j (\alpha_{ij}^{AB})^2$	$(I - 1)(J - 1)$	$\sigma^2 + KM \sigma_{AB}^2$
Взаимодействия BC	$SS_{BC} = IM \sum_j \sum_k (\alpha_{jk}^{BC})^2$	$(J - 1)(K - 1)$	$\sigma^2 + IM \sigma_{BC}^2$
Взаимодействия AC	$SS_{AC} = JM \sum_i \sum_k (\alpha_{ik}^{AC})^2$	$(I - 1)(K - 1)$	$\sigma^2 + JM \sigma_{AC}^2$
Взаимодействия ABC	$SS_{ABC} = M \sum_i \sum_j \sum_k (\alpha_{ijk}^{ABC})^2$	$(I - 1)(J - 1)(K - 1)$	$\sigma^2 + M \sigma_{ABC}^2$
Ошибки	$SS_e = \sum_i \sum_j \sum_k \sum_m (y_{ijkm} - y_{ijk.})^2$	$IJK(M - 1)$	σ^2
Общая сумма квадратов	$\sum_i \sum_j \sum_k \sum_m (y_{ijkm} - y_{....})^2$	$IJKM - 1$	

Выбирая определенный уровень значимости α (доверительную вероятность $P = 1 - \alpha$) и сравнивая экспериментально найденное значение $\hat{\sigma}_I^2 / \hat{\sigma}_{II}^2$ с критическим значением F -критерия $F_{(I-1), IJK(M-1)}^\alpha$, можно судить о справедливости гипотезы (4.4.26). Именно, гипотеза H отвергается тогда и только тогда, когда

$$\frac{\hat{\sigma}_I^2}{\hat{\sigma}_{II}^2} > F_{(I-1), IJK(M-1)}^\alpha \quad (4.4.29)$$

Таким образом, если гипотеза H верна, то вероятность отвергнуть ее при выполнении (4.4.29) равна α и не зависит от значений любых параметров, которые не определяются точно этой гипотезой (например, от σ^2). Проверка гипотез о значимости других главных эффектов и всех взаимодействий производится аналогично. Отметим, что если $M = 1$ (опыт во всех точках проведен без повторностей), то пространство L_e вырождается в нульмерное и оценка (4.4.28) не существует. В таких случаях необходимо иметь априорные сведения о некоторых взаимодействиях. Например, если мы знаем, что взаимодействие третьего порядка $A \times B \times C$ равно нулю, то в приведенной выше таблице SS_{ABC} может быть использовано вместо SS_e , т.е.

$$\hat{\sigma}_{II}^2 = \frac{SS_{ABC}}{(I-1)(J-1)(K-1)}$$

при соответствующем числе степеней знаменателя F -распределения.

Подводя итоги, можно сказать, что независимо от вида исходной системы применение дисперсного анализа позволяет выявить "значимость" определенных эффектов. Так, например, если в трехфакторном опыте с M повторностями будет выяснено, что оценки главного эффекта α_i^A , а также взаимодействия α_{ij}^{AB} , α_{ik}^{AC} , α_{ijk}^{ABC} с большой вероятностью равны нулю, то это означает, что первый фактор вообще не оказывает влияния на выход модели (или, точнее, его проявление вносит в измеряемую величину вклад, соизмеримый с уровнем шумов). Если все парные, тройные и т.д. взаимодействия оцениваются как незначимые, то можно считать факторы независимыми. Напротив, всегда можно улучшить оценки или сократить объем эксперимента, если, например, а priori известно, что модель линейна и, следовательно, взаимодействия отсутствуют. Подробно эти вопросы будут рассмотрены далее.

Рассмотрим в заключение этого пункта пример, иллюстрирующий специфику метода. На рис. 4.8,а и 4.8,б приведены результаты расчета динамики нарастания листового индекса в машинном эксперименте с описанной в п. 4.2.1 моделью агрофитоценоза пшеницы (листовой индекс равен отношению суммарной площади поверхности листьев всех растений, произрастающих на некоторой части поля, к величине площади этого поля). Эксперимент проводился на ЭВМ ЕС-1035. Все расчеты выполнены до наступления фазы колошения, поскольку известно, что при дальнейшем развитии посева листовый индекс начинает уменьшаться, а конечный урожай непосредственно связан с общей ассимилирующей поверхностью листьев. Нижние кривые на обоих графиках соответствуют засушливой погоде при отсутствии дождей и поливов. При этом посев формирует листовый индекс к фазе колошения $\sim 1,0 \div 1,1$. Все остальные кривые соот-

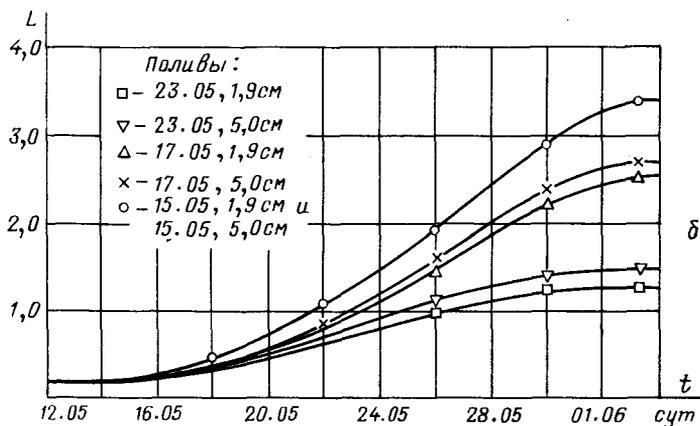
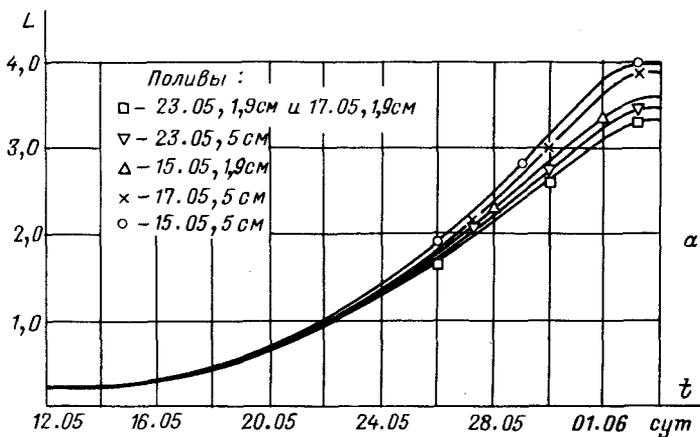


Рис. 4.8. Динамика роста листового индекса в имитационном эксперименте

ветствуют различным стратегиям полива в условиях засушливой погоды (см. рис. 4.8, б) или умеренно влажной погоды (см. рис. 4.8, а). Результаты расчета можно представить как трехфакторный эксперимент $2 \times 3 \times 2$, в котором испытаны две нормы полива ($1,9 \text{ см}^3/\text{см}^2$ и $5,0 \text{ см}^3/\text{см}^2$), три срока полива (15 мая, 17 мая и 23 мая) и две реализации погоды (дождливая и сухая). В соответствии с общепринятым подходом эти факторы обозначены через A , B и C . Результат эксперимента – листовый индекс в фазе колошения – представлен в табл. 4.3, а в табл. 4.4 сведены результаты обработки этих данных. Из этой таблицы следует, что значимыми (при доверительной вероятности $P = 0,95$) являются влияние факторов B (срок полива) и C (погода), а также их взаимодействие $B \times C$. Влияние остальных факторов незначительно. Физически это объяснимо, так как в фазе интенсивного роста именно срок полива определяет величину урожая. С другой стороны, погода резко меняет эту картину: как видно из рис. 4.8, а, выпадающие за этот период дожди приводят к уменьшению различий и сближению траекторий модели при всех стратегиях полива.

Таблица 4.3

Результат факторного эксперимента с моделью агроценоза

Сроки полива	Норма полива, т/га	Погода			
		умеренно дождливая		сухая	
		1,9	5,0	1,9	5,0
15 мая		3,5	4,0	3,4	3,4
17 мая		3,3	3,8	2,6	2,7
23 мая		3,3	3,4	1,3	1,5

Таблица 4.4

Результат обработки данных имитационного эксперимента

Сумма квадратов	Степени свободы	Величина эффекта	Значение F-критерия	Сумма квадратов	Степени свободы	Величина эффекта	Значение F-критерия
SS_A	1	0,159	7,57	SS_{BC}	2	1,31	89,5
SS_B	2	2,93	139,5	SS_{AC}	1	0,0510	2,43
SS_C	1	3,43	163,3	SS_{ABC}	2	0,0560	2,67
SS_{AB}	2	0,0246	1,17				

Вполне понятна также значимость взаимодействия $B \times C$ (т.е. стратегии полива и погоды). Отсюда следует важный практический вывод о необходимости корректировки сроков и норм полива с учетом складывающейся в сезоне метеорологической обстановки.

б) Планирование дисперсионных экспериментов в задачах имитационного моделирования. Мы установили, что существо дисперсионного анализа заключается в расщеплении исходного N -мерного пространства измерений $Y = \{y_{ijk...l}\}$ (где N — общее число экспериментальных точек) на ряд ортогональных друг другу подпространств, что позволяет получать независимые оценки для ряда эффектов (главных эффектов и взаимодействий различного порядка). Отметим, что получение самих этих оценок не обусловлено предположением о нормальности ошибок, поскольку оценки наименьших квадратов обладают (при известных допущениях) свойствами несмещенности, состоятельности и эффективности при любом распределении, имеющем ограниченный второй момент. В тех же случаях, когда ошибки подчинены нормальному (или близкому к нему; см. [27]) распределению, возможно получение более тонких статистических выводов.

Применимость дисперсионного анализа связана лишь с достаточно общими предположениями о свойствах изучаемого объекта, что явилось залогом

его широкого распространения в самых различных областях исследования. В то же время эта универсальность метода является и его недостатком, поскольку он не позволяет использовать имеющуюся (и часто очень дорогостоящую) дополнительную информацию об объекте. В связи с этим следует особо остановиться на специфике имитационного эксперимента. Можно рассчитывать на то, что учет конкретных свойств, которыми обладают машинные эксперименты, позволит сформулировать некоторую специальную технологию применения этого метода.

В машинных экспериментах приходится изучать влияние большого числа (десятков) контролируемых переменных (факторов). С другой стороны, эти эксперименты, как правило, проводятся в ускоренном (по сравнению с натурным экспериментом) масштабе времени, что позволяет одновременно (т.е. в одном эксперименте) получать и обрабатывать большой статистический материал. И тем не менее проблема размерности является одной из основных в случае как натуральных, так и машинных экспериментов. Приведем некоторые оценки. Расчет суточного цикла в имитационной модели агроценоза пшеницы занимает 6 с машинного времени на ЭВМ ЕС-1035. Следовательно, если мы будем изучать лишь один месяц вегетации при наличии десяти контролируемых переменных, то факторный эксперимент даже минимального объема ($2^{10} = 1024$ экспериментальных точек без повторностей) займет приблизительно 50 часов машинного времени. В то же время совершенно ясно, что излишняя подробность в изучении модели нецелесообразна. Вряд ли можно будет при этом найти физическую трактовку, например, для взаимодействия десятого порядка $A \times B \times \dots \times C$ или даже различных взаимодействий 3–4 порядков. Это означает, что полный факторный эксперимент избыточен — нет никакой необходимости в переборе всех возможных комбинаций факторов.

Сокращение числа экспериментальных точек (по отношению к полному перебору), естественно, уменьшает размерность пространства измерений Y . При этом в каждое из ортогональных подпространств попадают (в зависимости от вида плана) эффекты, количество которых оказывается большим, чем размерность подпространства. Это приводит к линейной зависимости определенных групп эффектов, а следовательно, и к невозможности независимого получения всех оценок. Эффекты, как говорят, "смешиваются" [7]. В приложении 3 приведена схема полного факторного эксперимента 2^3 и так называемая "дробная реплика" 2^{3-1} , порождаемая определяющим соотношением $I = ABC$. В данном эксперименте линейно зависимы (и оцениваются совместно) следующие четыре пары комбинаций: μ и α^{ABC} , α^A и α^{BC} , α^B и α^{AC} , α^C и α^{AB} . Отсутствие повторностей не позволяет также получить независимую оценку для дисперсии ошибки. Несмотря на такие "плохие" свойства плана, явление "смешивания" вполне допустимо, особенно в тех случаях, когда размерность пространства контролируемых переменных велика.

Учитывая эти обстоятельства, Бокс и Хантер (см., например, [7]) ввели следующие определения планов различного типа:

1. П л а н ы р а з р е ш е н и я III: ни один главный эффект не смешан с другим главным эффектом, но главные эффекты смешаны с двухфакторными взаимодействиями, которые сами могут быть смешаны друг с другом.

2. П л а н ы р а з р е ш е н и я IV: все главные эффекты оцениваются независимо; взаимодействия двух и более факторов смешаны друг с другом.

3. П л а н ы р а з р е ш е н и я V: все главные эффекты оцениваются независимо; все парные взаимодействия не смешаны с другими парными взаимодействиями, но эти последние могут быть смешаны с взаимодействиями трех факторов.

Методы построения планов с различной степенью разрешения изложены в [7]. Можно полагать, что планы разрешения IV уже приемлемы во многих ситуациях. И во всяком случае, вряд ли оправдано практическими соображениями использование планов разрешения, большего, чем V.

Далее мы будем рассматривать лишь двухуровневые планы, т.е. такие планы, в которых каждая из контролируемых переменных может принимать в эксперименте одно из двух фиксированных значений. Причина этого ограничения выяснится в дальнейшем. В двухуровневых планах можно изменить обозначения эффектов, введя вместо "амплитуд" главных эффектов и взаимодействий соответствующий "размах", например, для первого главного эффекта

$$\alpha^A = \eta_{2\dots} - \eta_{1\dots} = 2\alpha_2^A = -2\alpha_1^A$$

вместо

$$\alpha_1^A = \eta_{1\dots} - \eta_{\dots} \quad \text{и} \quad \alpha_2^A = \eta_{2\dots} - \eta_{\dots}$$

Это позволяет сократить число оцениваемых величин и избавиться от их линейной зависимости (типа, например, соотношений (4.4.16), из которых, в частности, следует, что $\alpha_1^A + \alpha_2^A = 0$). Значения переопределенных главных эффектов и взаимодействий равны, таким образом, с точностью до знаков удвоенным "старым" величинам (4.4.15). Как следует из приложения 3, общую модель двухуровневого плана можно представить в виде

$$y = X\beta + e, \tag{4.4.30}$$

где y и e – векторы наблюдений и ошибок, β – вектор оцениваемых эффектов (генеральное среднее, главные эффекты и взаимодействия различного порядка). Элементами матрицы X являются ± 1 . Число ее столбцов равно количеству оцениваемых параметров, т.е. 2^k , где k – число контролируемых переменных. Строки этой матрицы соответствуют отдельным экспериментальным точкам, т.е. каждая строка порождается определенной комбинацией контролируемых величин: если в опыте с номером j некоторая переменная принимает минимальное значение, то в j -й строке матрицы X этой переменной соответствует -1 ; в противном случае на это место помещается $+1$. Очевидно, что в полном факторном двухуровневом эксперименте число опытов $N = 2^k$ и матрица X квадратна.

Полный факторный эксперимент обладает одним дополнительным важным свойством – столбцы матрицы X ортогональны:

$$\langle x_j, x_l \rangle = \begin{cases} N & \text{при } j=l, \\ 0 & \text{при } j \neq l, \end{cases}$$

и, следовательно,

$$X^*X = NI, \tag{4.4.31}$$

где I – единичная матрица. Поэтому оценки наименьших квадратов легко определяются:

$$\hat{\beta} = (X^*X)^{-1}X^*y = \frac{1}{N}X^*y. \quad (4.4.32)$$

Ковариационная матрица этих оценок также диагональна:

$$\text{cov}\{\hat{\beta}\} = \sigma^2(X^*X)^{-1} = \frac{\sigma^2}{N}I, \quad (4.4.33)$$

и, следовательно, все оценки независимы. Отметим, что эта независимость оценок является следствием общей теории и связана с разложением N -мерного пространства измерений на N ортогональных друг другу одномерных подпространств. Очевидно, что ортогональность планов является их важным преимуществом и что это свойство желательно сохранить при планировании неполных факторных экспериментов.

Перейдем к обсуждению требований, предъявляемых к планам имитационного эксперимента. Некоторые общие требования сформулированы в [7]. Попытаемся конкретизировать их, исходя из особенностей машинной имитации. Отметим, что в настоящее время трудно подвести формальную основу под те соображения, которые высказаны ниже. Это лишь некоторая "канва", которой следует придерживаться при экспериментировании с имитационной моделью. Как говорится, "возможны варианты".

Итак, имитационные планы, по нашему мнению, должны обладать следующими особенностями:

- 1) включать в себя по возможности малое число опытов;
- 2) обеспечивать минимум дисперсии ошибки;
- 3) допускать возможность проверки адекватности модели;
- 4) обладать желаемой системой смешивания;
- 5) допускать возможность последовательного достраивания с уточнением некоторых оценок;
- 6) приводить к простой и допускающей содержательную интерпретацию обработке результатов эксперимента.

Как это обычно бывает при решении сложных задач, перечисленные выше требования противоречивы. Поэтому следует искать компромиссное решение. Попытаемся показать, что этот компромисс достигается, если принять за основу ортогональные двухуровневые планы [24].

Свойство ортогональности планов является весьма ценным, поскольку оно обеспечивает независимость получаемых оценок. Очевидно, что его надо сохранить, имея в виду последовательное планирование эксперимента (п. 5), т.е. проведение эксперимента в несколько этапов, когда планирование каждого последующего этапа уточняется с учетом результатов всех предыдущих. Ясно, что использование ортогональных планов позволяет не пересчитывать ранее полученные оценки. К этому вопросу мы еще вернемся в дальнейшем при рассмотрении "стратегии" и "тактики" эксперимента.

Кроме того, при получении оценок в данном случае обращение матрицы X^*X превращается просто в вычисление обратных величин, что позволяет избежать влияния машинных ошибок округления, возникающих при обращении матриц общего вида. Двухуровневые планы обладают наимень-

шим объемом среди всех прочих ортогональных планов (п. 1). Но это не единственное их преимущество. Более существенным является, пожалуй, то, что применение так называемого "метода перевала", основанного на теореме Бокса и Уилсона [7], позволяет путем удвоения числа экспериментальных точек получить из исходного плана, имеющего разрешение III, новый план разрешения IV. А.А. Шестиперовым (см. [26, 27] к гл. 5) предложен прием, позволяющий получать аналогичным путем планы, обладающие для отдельных переменных свойствами планов разрешения V. Эти обстоятельства позволяют обосновать регулярный метод последовательного планирования, удовлетворяющий требованию п.4.

Возможность проверки адекватности модели всегда связана с введением повторностей. Пусть а priori известно, что некоторые из эффектов отсутствуют в модели (например, из физических соображений может быть ясно, что в модель не могут входить взаимодействия третьего и более высоких порядков). Обозначим число оцениваемых эффектов через N_1 . Очевидно, что $N_1 \leq 2^k$, где k — число контролируемых переменных. Обозначим также через N_2 число экспериментальных точек. Критерии проверки адекватности модели (п. 3) можно построить лишь в случае, когда число опытов N превышает N_2 , т.е. когда $N > N_2 > N_1$. При этом хотя бы в некоторых точках опыт должен дублироваться.

Рассмотрим теперь требование п. 3. Дисперсионная матрица оценок $\hat{\beta}$ равна

$$D_{\beta} = \sigma^2 (X^* X)^{-1}. \quad (4.4.34)$$

Как уже указывалось, для получения "более хороших" оценок следовало бы так спланировать эксперимент, чтобы по возможности уменьшить эту величину. Иначе говоря, задача оптимизации плана сводится к максимизации какого-либо функционала от информационной матрицы

$$M_{\beta} = \sigma^{-2} X^* X. \quad (4.4.35)$$

Этот вопрос будет затронут в следующем пункте, где речь пойдет о регрессионных экспериментах. Здесь же заметим, что требование оптимальности плана может вступить в противоречие с условием его ортогональности. Тем не менее Боксом показано, что диагональные элементы в (4.4.34) минимизируются, если

$$X = N^{1/2} G,$$

а матрица G ортогональна, $G^* G = I$, и имеет первый столбец, состоящий из единиц. Таким образом, полный факторный эксперимент является в этом смысле оптимальным.

Наконец, несколько замечаний относительно оценок. Вид оценок вытекает из простого соотношения (4.4.32), а именно из выражения

$$\hat{\beta}_i = \frac{1}{N} \langle x_i, y \rangle,$$

где элементы i -го столбца матрицы X равны $+1$ или -1 . В каждом конкретном случае эти оценки легко находятся. Соответствующие формулы приведены, например, в [7]. Важно отметить, что значимость эффектов $\hat{\beta}_j$ в данном случае означает их отличие от нуля. При этом, учитывая, что чис-

ло степеней свободы для каждой величины $\hat{\beta}_i$ в ортогональном плане равно единице, здесь при проверке значимости одного эффекта можно использовать, помимо F -критерия, t -критерий Стьюдента (поскольку $F_{1,1} = t_1^2$).

в) Стратегическое и тактическое планирование эксперимента. В качестве основы планирования имитационного эксперимента мы выбрали ортогональные двухуровневые планы. Учет требований минимизации объема эксперимента не позволяет при большом числе факторов применять полные факторные планы. Это приводит к необходимости разбиения всего процесса экспериментального изучения модели на ряд последовательных этапов. Мы исходим из того, что дисперсионный эксперимент используется на первой стадии изучения модели, когда степень ее чувствительности к совокупности контролируемых переменных слабо изучена или неизвестна вообще. Поэтому надо сравнительно "быстро" и с малыми затратами получить первую, хотя бы достаточно грубую характеристику модели, а уже потом, на второй стадии уточнять необходимые детали.

Именно с этими обстоятельствами связано разбиение всех вопросов, возникающих при изучении модели, на "стратегические" и "тактические". Тактика планирования эксперимента обсуждена выше. Что же предстоит решить на уровне стратегии?

Здесь следует ответить на два вопроса.

1. Если на первом этапе выбран план относительно низкого разрешения (например, разрешения III), то каким образом, получив результат первого эксперимента, продолжить изучение объекта, с тем чтобы уточнить именно те оценки, которые выявлены как значимые в первом эксперименте?

2. Как перейти от "локального" двухуровневого плана к "глобальному"?

Первый из этих вопросов уже затрагивался выше. Его решение основано на использовании теоремы Бокса и Уилсона и ее обобщении, полученном А.А. Шестиперовым. Теорема Бокса и Уилсона формулируется следующим образом.

Пусть существует план (не обязательно ортогональный) для оценивания N_1 линейных эффектов с матрицей D_1 размера $N \times N_1$ ($N_1 < N = 2^k$); допустим также, что X — соответствующая $N \times (N_1 + 1)$ -матрица независимых переменных:

$$X = (\mathbf{1}, D_1),$$

где $\mathbf{1}$ — вектор, составленный из $+1$. Тогда

$$D_2 = \begin{pmatrix} X \\ -X \end{pmatrix}$$

будет матрицей плана для оценивания линейных эффектов $k + 1$ факторов. При этом, если план D_1 обеспечивает оценки главных эффектов, смешанные с парными взаимодействиями, оценки главных эффектов, получаемые на основании плана D_2 , независимы.

Таким образом, эта теорема позволяет по плану разрешения III методом перевала строить планы, имеющие разрешение IV. Как показано в [17], в тех случаях, когда N кратно четырем, можно построить план разрешения III для оценки $N_1 = N - 1$ эффектов при наличии N_1 факторов. В качестве примера в табл. 4.5 приведен план типа так называемых планов

План Плэкета – Бермана для 11 факторов

1	2	3	4	5	6	7	8	9	10	11
+	-	+	-	-	-	+	+	+	-	+
+	+	-	+	-	-	-	+	+	+	-
-	+	+	-	+	-	-	-	+	+	+
+	-	+	+	-	+	-	-	-	+	+
+	+	-	+	+	-	+	-	-	-	+
+	+	+	-	+	+	+	+	-	-	-
-	+	+	+	-	+	+	+	+	-	-
-	-	+	+	+	-	+	+	-	+	-
-	-	-	+	+	+	-	+	+	-	+
+	-	-	-	+	+	+	-	+	+	-
-	+	-	-	-	+	+	+	-	+	+
-	-	-	-	-	-	-	-	-	-	-

Плэкета – Бермана для $k = N_1 = 11$ и $N = 12$. С помощью метода перевала из него можно получить план разрешения IV для такого же количества факторов, но при удвоенном числе опытов, т.е. для $N = 2(k + 1) = 24$. Матрица D_2 для этого нового плана представлена в табл. 4.6. Для сравнения отметим, что для $k = 11$ схема полного факторного эксперимента включает в себя $2^{11} = 2048$ опытов.

Усовершенствование этого подхода предложено А.А. Шестиперовым. Существо дела сводится к тому, что метод перевала применяется лишь к одному или нескольким столбцам матрицы X , а остальные столбцы в нижней половине матрицы плана дублируют верхние. Если эту процедуру применить последовательно (по одному) к каждому из столбцов матрицы X , то полученный в результате эксперимент будет иметь разрешение V. Очевидно, что число опытов при этом существенно возрастет. Но это не имеет значения, поскольку подобную процедуру никогда не приходится доводить до конца. В результате первого шага большая часть эффектов, как правило, оказывается незначимой и, следовательно, уточнение оценок надо проводить лишь для некоторых факторов из общего числа исходных. Очевидно, что во многих практических случаях при этом будет получена определенная экономия по сравнению с таким порядком проведения эксперимента, при котором в качестве исходного плана сразу же выбирается план разрешения IV или V.

Обратимся к рассмотрению второго вопроса. Очевидно, что двухуровневые планы можно использовать без всяких оговорок в тех случаях, когда мы убеждены в том, что зависимость отклика от контролируемых переменных монотонна. Если же это неочевидно и в эксперименте необходимо проверить, обладает ли модель экстремумом по некоторым переменным или по их комбинациям, отвечающим некоторым комбинациям в факторном пространстве, то двухуровневая схема представляется необоснованной. В то же время ее преимущества, о которых говорилось выше, не позволяют так просто от нее отказаться. Выход из положения здесь можно

Таблица 4.6

План разрешения IV, полученный из плана Плэкета – Бермана методом перевала

1	2	3	4	5	6	7	8	9	10	11	12
+	+	-	+	-	-	-	+	+	+	-	+
+	+	+	-	+	-	-	-	+	+	+	-
+	-	+	+	-	+	-	-	-	+	+	+
+	+	-	+	+	-	+	-	-	-	+	+
+	+	+	-	+	+	-	+	-	-	-	+
+	+	+	+	-	+	+	-	+	-	-	-
+	-	+	+	+	-	+	+	-	+	-	-
+	-	-	+	+	+	-	+	+	-	+	-
+	-	-	-	+	+	+	-	+	+	-	+
+	+	-	-	-	+	+	+	-	+	+	-
+	-	+	-	-	-	+	+	+	-	+	+
+	-	-	-	-	-	-	-	-	-	-	-
+	-	-	-	-	-	-	-	-	-	-	-
-	-	+	-	+	+	+	-	-	-	+	-
-	-	-	+	-	+	+	+	-	-	-	+
-	+	-	-	+	-	+	+	+	-	-	-
-	-	+	-	-	+	-	+	+	+	-	-
-	-	-	+	-	-	-	+	+	+	+	-
-	-	-	+	-	-	-	+	+	+	+	-
-	+	-	-	-	+	-	-	+	-	+	+
-	+	+	-	-	-	+	-	-	+	-	+
-	+	+	+	-	-	-	+	-	-	+	-
-	-	+	+	+	-	-	-	+	-	-	+
-	+	-	+	+	+	-	-	-	+	-	-
-	+	+	+	+	+	+	+	+	+	+	+

искать на путях последовательной реализации "локального" или "глобального" подхода. Первый из них традиционен и заключается в том, что в области планирования выбирается некоторая "рабочая точка", вблизи которой локально исследуются свойства модели. При этом возникают трудности, связанные с последовательной сменой этой рабочей точки. Регулярные методы существуют лишь в теории экстремальных экспериментов, целью которых является экспериментальное нахождение максимума или минимума поверхности отклика.

Можно предложить другой — глобальный подход. Из физических соображений всегда можно оценить границы возможного изменения контролируемых переменных, что позволяет выбрать область планирования эксперимента в факторном пространстве в виде гиперпараллелепипеда (или, при подходящем выборе масштабов, в виде гиперкуба). В таком случае первоначальный план может быть поставлен в точках, являющихся подмножеством вершин куба. В случае, если после проведения эксперимента некоторые эффекты оцениваются как незначимые, но есть подозрение в том, что влияние этих факторов существует, можно осуществить дробление шага по тем или иным координатам. При этом новые эксперименталь-

ные точки следует, разумеется, размещать так, чтобы имела возможность рассматривать их в совокупности с ранее исследованными. Сказанное иллюстрирует рис. 4.9. На рис. 4.9,а крестиками отмечены точки проведения ортогонального плана разрешения III для $k = 3$ и $N = k + 1 = 4$, а на рис. 4.9,б кружками показаны те дополнительные точки, в которых следует поставить эксперимент при дроблении шага пространственной сетки. Очевидно, что в случае необходимости процедура дробления шага сетки может быть продолжена.

Таким образом, стратегическое планирование эксперимента имеет своей целью последовательный переход от "грубого" описания объекта к все более и более точному и детальному. Представляется, что эта последовательность действий является естественной для человека. Методы же планирования эксперимента в данном случае следует рассматривать лишь как инструмент, позволяющий в определенном смысле упорядочить эту деятельность.

В заключение отметим, что относительно простые алгоритмы генерации типов двухуровневых планов и их обработки позволяют автоматизировать этот вид деятельности. Об этом мы еще поговорим в следующей главе.

4.4.3. Планирование регрессионных экспериментов. В результате дисперсионного эксперимента с моделью выявляется набор значимых факторов, т.е. тех контролируемых переменных, влияние которых на выход модели превышает уровень "шумового" фона. Выявление количественных связей этих переменных с выходом модели является целью регрессионного эксперимента.

Можно представить себе два типа задач, возникающих на этом этапе. Первый из них связан с решением вопросов идентификации модели. Имитационная модель, реализованная в виде пакета программ, всегда включает в себя набор неизвестных констант, которые должны быть определены по экспериментальным данным для конкретной области применения модели. Задача определения этих констант (а в общем случае элементов более сложной природы, например функций) в физике называется обратной задачей, а в общей теории динамических систем — задачей идентификации.

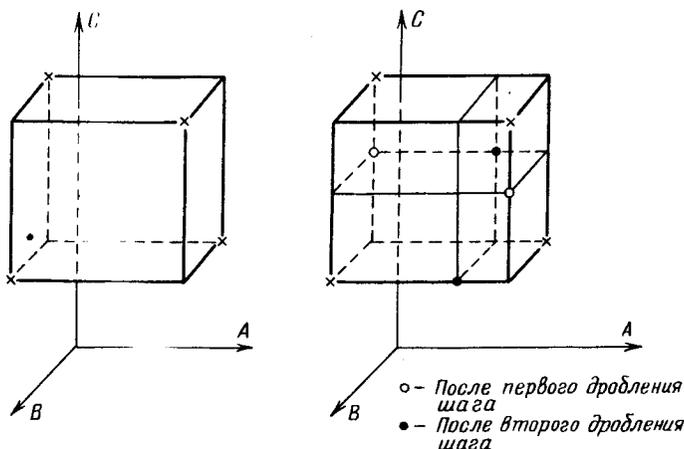


Рис. 4.9. Ортогональный план на кубе и дробление шага

Другого рода вопросы, требующие применения тех же методов, возникают при создании систем управления с использованием имитационных моделей. Учитывая большую сложность модели, значительное, как правило, время счета и т.п., можно показать, что использование процесса моделирования в реальном времени, т.е. непосредственно для выработки управленческих решений, вряд ли целесообразно, если вообще возможно. Поэтому наряду с базовой моделью (об этом уже говорилось) необходимо иметь набор достаточно простых полуэмпирических зависимостей, составляющих "внешнее" математическое обеспечение имитационной системы. Эти модели по существу можно рассматривать как некоторые аппроксимирующие выражения, приближенно описывающие зависимости "вход-выход" модели в той или иной ограниченной области изменения условий ее функционирования. Очевидно, что для нахождения таких аппроксимаций в полной мере могут быть использованы методы математической статистики и, в частности, аппарат регрессионного анализа.

Так, например, в агрометеорологии накоплен большой арсенал полуэмпирических зависимостей, связывающих продуктивность агроэкосистем с различными интегральными агрометеопараметрами — радиационным балансом, суммой осадков, суммой активных температур и т.д. Коэффициенты этих моделей определены экспериментально с использованием определенного фактического материала и, как и все зависимости такого рода, носят локальный характер. Очевидно, что использование имитационной модели в данном случае, с одной стороны, позволит уточнить эти константы, а с другой — обеспечит получение оценок границ применимости аппроксимирующих моделей.

Перейдем к описанию формальной постановки задачи. Пусть изучаемый выход аппроксимирующей модели может быть приближенно представлен как линейная комбинация известных функций от контролируемых переменных x некоторыми параметрами $\theta^* = (\theta_1, \theta_2, \dots, \theta_{i-1})$ (см. также (4.4.3)) :

$$\tilde{y} = \theta^* f(x). \quad (4.4.36)$$

Будем предполагать, что контролируемые переменные могут принимать значения из некоторой области планирования эксперимента Σ , являющейся частью евклидова пространства, $x \in \Sigma \subset R^n$. Отметим, что в некоторых случаях этого предположения недостаточно: область планирования может принадлежать функциональному пространству, например C^2 . Однако в задачах имитационного моделирования условие конечномерности представляется вполне уместным.

Предположение о "приближенной" зависимости (4.4.36) означает, что "истинный" выход модели в некоторой точке x^i отличается от \tilde{y}_i . Аппроксимирующая модель должна работать в различных ситуациях, которые часто отличаются друг от друга тем, что на исходную систему действуют те или иные случайные возмущения (для природных экосистем в качестве таких возмущений выступают случайные погодные факторы). Поэтому будем понимать приближенное равенство (4.4.36) в том смысле, что математическое ожидание выхода модели равно правой части этого равенства в каждой точке области изменения контролируемых переменных:

$$M\{y | x^i\} = \theta^* f(x^i), \quad (4.4.37)$$

а сам выход модели отличается от этого выражения на случайную величину ϵ_i , имеющую нулевое среднее значение и дисперсию σ_i^2 . Предполагается также, что случайные величины ϵ_i независимы. Задача заключается в нахождении оценок $\hat{\theta}$ неизвестных параметров θ на основе "измерения" выхода модели y .

Может показаться, что в некоторых ситуациях эти предположения искусственны. Однако общая теория развита именно для этого случая. Все другие требуют специального рассмотрения.

При сформулированных допущениях стандартным методом нахождения линейных оценок $\hat{\theta}$ является метод наименьших квадратов. Пусть в результате эксперимента, проведенного в $N > m$ точках x^1, x^2, \dots, x^N , зафиксирован ряд наблюдений y_1, y_2, \dots, y_N :

$$y_i = \theta^* f(x^i) + \epsilon \quad (i = \overline{1, N}). \quad (4.4.38)$$

Введя обозначения $y = (y_1, y_2, \dots, y_N)^*$, $\epsilon = (\epsilon_1, \epsilon_2, \dots, \epsilon_N)^*$,

$$F(f_{ij})_{N \times m} = \|f_j(x^i)\|,$$

перепишем (4.4.38) в виде

$$y = F\theta + \epsilon. \quad (4.4.39)$$

Пусть F — матрица полного ранга, $\text{rank } F = m$. Тогда наилучшие линейные оценки $\hat{\theta}$ определяются выражением

$$\hat{\theta} = (F^* \Sigma_\epsilon^{-1} F)^{-1} F^* \Sigma_\epsilon^{-1} y, \quad (4.4.40)$$

где через Σ_ϵ обозначена дисперсионная матрица ошибок, $\Sigma_\epsilon = \text{diag} \{ \sigma_1^2, \sigma_2^2, \dots, \sigma_N^2 \}$. Термин "наилучшие" здесь означает, что величины $\hat{\theta}$ в классе L всех линейных несмещенных оценок обращают в минимум дисперсионную матрицу оценок $D(\hat{\theta})^1$:

$$\hat{\theta} = \text{argmin } D(\theta). \quad (4.4.41)$$

При этом оказывается, что наилучшая линейная оценка (4.4.41) (НЛН-оценка) минимизирует сумму взвешенных дисперсий ошибок, т.е.

$$\hat{\theta} = \text{argmin } R(\theta), \quad (4.4.42)$$

где

$$R(\theta) = \sum_{i=1}^N \sigma_i^{-2} (y_i - f^*(x^i) \theta)^2.$$

Перечислим дополнительные свойства оценок.

1. Дисперсионная матрица НЛН-оценок равна

$$D(\hat{\theta}) = (F^* \Sigma_\epsilon^{-1} F)^{-1}. \quad (4.4.43)$$

¹⁾ Отметим, что класс знакопостоянных (знакоопределенных) матриц одного порядка частично упорядочен в том смысле, что для двух знакопеременных (знакоопределенных) матриц A_1 и A_2 можно записать $A_1 \leq A_2$ ($A_1 < A_2$), если $(x, Bx) \geq 0$ (соответственно $(x, Bx) > 0$), где $B = A_2 - A_1$. Поэтому операция минимизации в (4.4.41) имеет смысл.

2. Определитель дисперсионной матрицы НЛН-оценок является наименьшим среди всех линейных оценок:

$$|D(\hat{\theta})| < |D(\tilde{\theta})|, \quad (4.4.44)$$

где $\tilde{\theta}$ — произвольная линейная оценка, $\tilde{\theta} \neq \hat{\theta}$.

3. Наилучшей линейной оценкой поверхности отклика y в произвольной точке x является

$$\hat{y} = \hat{\theta}^* f(x); \quad (4.4.45)$$

дисперсия этой оценки равна $d_y(x) = f^*(x)D(\hat{\theta})f(x)$.

4. Пусть $\sigma_1^2 = \sigma_2^2 = \dots = \sigma_N^2 = \sigma^2$. Тогда несмещенной оценкой дисперсии σ^2 является величина

$$S^2 = \frac{R_0}{N - m},$$

где $R_0 = (y - F\hat{\theta})^*(y - F\hat{\theta})$ — наименьшая (остаточная) сумма квадратов.

Дисперсионная матрица НЛН-оценок, как видно из (4.4.43), является функцией тех точек, в которых выполнялся эксперимент. Следовательно, можно поставить задачу выбора такого числа и такого расположения точек в области планирования Σ , которое приводит к дополнительному улучшению свойств оцениваемых параметров и функции отклика модели. К этому вопросу мы теперь и переходим.

Однако прежде всего следует уточнить понятие эксперимента. Как уже упоминалось, регрессионный эксперимент ставится в N точках области планирования Σ . При этом не все из них могут быть различны. Как и в дисперсионном эксперименте, во всех или в некоторых точках измерения могут дублироваться.

Пусть, как и ранее, выбрано N экспериментальных точек, но в точке x^i эксперимент проводится в r_i повторностях. Общее число наблюдений становится при этом равным

$$\sum_{i=1}^N r_i.$$

Обозначим набор наблюдений функции отклика в i -й точке через $y_{i1}, y_{i2}, \dots, y_{ir_i}$. Тогда оказывается, что все выписанные выше соотношения сохраняют свою силу, если подставить в них $w_i = r_i \sigma_i^{-2}$ вместо σ_i^{-2} и

$$\bar{y}_i = \frac{1}{r_i} \sum_{j=1}^{r_i} y_{ij} \quad \text{вместо } y_i. \quad \text{Таким образом, наличие повторностей приводит к изменению весовых коэффициентов в минимизируемой сумме квадратов (4.4.42) и в остальных соотношениях, а также к замене измеренных значений функции отклика в каждой точке их осредненными величинами. Величины } \bar{y}_i \text{ поэтому являются достаточными статистиками метода наименьших квадратов. Учитывая все сказанное, введем, следуя [25], такое определение эксперимента.}$$

Определение. Экспериментом e_j называется совокупность величин

$$\begin{aligned} &x^{1j}, x^{2j}, \dots, x^{Nj}, \\ &w_{1j}, w_{2j}, \dots, w_{Nj}, \\ &y_1^j, y_2^j, \dots, y_N^j, \end{aligned} \quad (4.4.46)$$

где

$$w_{ij} = r_{ij} (\sigma_i^j)^{-2}, \quad \text{а} \quad y_i^j = \frac{1}{r_{ij}} \sum_{l=1}^{r_{ij}} y_{il}^j.$$

Будем считать эксперимент E_1 отличным от E_2 ($E_1 \neq E_2$), если хотя бы одна из входящих в определение эксперимента величин при $j = 1$ отличается от соответствующей величины при $j = 2$. Совокупность точек в области планирования $x^{1j}, x^{2j}, \dots, x^{Nj}$ будем называть спектром плана E_j , а набор величин

$$\begin{matrix} x^{1j}, & x^{2j}, & \dots, & x^{Nj}, \\ r_{1j}, & r_{2j}, & \dots, & r_{Nj} \end{matrix} \quad (4.4.47)$$

— планом эксперимента.

Как можно сравнивать различные эксперименты? Поскольку основная характеристика эксперимента — точность оценок — определяется дисперсионной матрицей, за критерий качества эксперимента можно принять или саму дисперсионную матрицу, или некоторый функционал от этой матрицы. Важную роль при этом играет так называемая информационная матрица Фишера

$$M = \sum_{i=1}^N w_i f(x^i) f^*(x^i) = \sum_{i=1}^N M(x^i), \quad (4.4.48)$$

поскольку, как это следует из (4.4.43),

$$D(\hat{\theta}) = M^{-1}. \quad (4.4.49)$$

При этом нетрудно видеть, что при заданной точности наблюдений $D(\hat{\theta})$, так же как и M , зависит только от плана эксперимента (4.4.47). Некоторые хорошо изученные критерии сравнения экспериментов приведены в приложении 4. Здесь же отметим, что наиболее распространенными в этой области являются понятия D -оптимальных планов, оптимизирующих определитель дисперсионной матрицы оценок:

$$|D(E)| \Rightarrow \min,$$

и A -оптимальных, минимизирующих след этой матрицы:

$$\text{Sp} D(E) \Rightarrow \min.$$

В заключение настоящего параграфа отметим, что ограничение, связанное с конечномерностью пространства контролируемых переменных Σ , можно ослабить и считать, что исследуемая модель может представляться в виде множества функционалов, заданных на некотором функциональном пространстве.

Необходимость в обобщении такого рода связана с тем, что рассматриваемые нами экологические процессы носят динамический характер, а их описание основано на использовании аппарата дифференциальных уравнений (как обыкновенных, так и в частных производных). Так, например, динамика тепло- и влагопереноса в почве и в приземном воздухе, переноса солей в почве, газообмена в посевах и ряда других процессов в агроэкосистемах описывается уравнениями в частных производных параболического

типа. При включении этих блоков в общую роль модель производится пространственно-временная дискретизация уравнений, так что в окончательном виде все соотношения представлены в модели в виде некоторых алгебраических соотношений. Это, однако, лишь завершающий этап конструирования модели. На предварительной стадии, которую мы назвали "предмодельной", должна быть выполнена параметризация исходного описания. А этот процесс связан с решением обратных задач математической физики — так называемых задач идентификации. Собственно, с этим этапом исследования, как уже говорилось в начале этого пункта, и связано применение методов планирования регрессионного эксперимента.

Приведем пример. Уравнение теплопереноса в почве в одномерном случае записывается в виде обычного уравнения теплопроводности, с тем отличием, что коэффициенты теплоемкости и теплопроводности зависят от величины объемной влажности почвы. Это уравнение имеет вид

$$c(w) \frac{\partial T_s}{\partial t} = \frac{\partial}{\partial x} \left(\lambda(w) \frac{\partial T_s}{\partial x} \right) \quad (4.4.50)$$

при соответствующих граничных и начальном условиях

$$\begin{aligned} T_s(t, 0) &= \theta_0(t), \quad T_s(t, r) = \theta_r, \quad t \geq 0, \\ T_s(0, x) &= \Theta(x), \quad x \in [0, r], \end{aligned} \quad (4.4.51)$$

где $T_s(t, x)$ — температура почвы, w — влажность почвы, $c(w)$ и $\lambda(w)$ — коэффициенты теплоемкости и теплопроводности. Экспериментальное исследование хода температуры производится путем установки на некоторых глубинах почвенного профиля x_1, x_2, \dots, x_l термодатчиков и непрерывной записи показаний многоканального термометра. Учитывая, что измерения происходят с определенной погрешностью, результат измерений в случае линейного измерительного устройства можно представить в виде

$$y(t, x_i) = \int_{t_0}^t \int_{x_i-a}^{x_i+a} h(t-\tau, x_i-\xi) T_s(\tau, \xi) d\tau d\xi + \epsilon_i(t) \quad (i = \overline{1, l}), \quad (4.4.52)$$

где $h(\tau, \xi)$ — весовая функция измерителя, а $\epsilon_i(t)$ — случайные ошибки измерения.

Содержательный смысл имеют обратные задачи двоякого рода:

1) Зная величину влажности почвы, определить по результатам измерения (4.4.52) зависимости $c(w)$ и $\lambda(w)$ при их подходящей параметризации, например,

$$c(w) = c_0 + c_1 w, \quad \lambda(w) = \lambda_0 + \lambda_1 w + \lambda_2 w^2. \quad (4.4.53)$$

2) Зная зависимости (4.4.53), использовать измерения температуры для оценки влажности почвы w .

Последняя задача является практически весьма важной, поскольку измерение температуры осуществляется гораздо проще, чем непосредственное измерение профиля влажности почвы. Планирование эксперимента в данном случае сводится к выбору числа и расположения по глубине почвенного профиля датчиков температуры, т.е. величин l, x_1, x_2, \dots, x_l . В случае дискретной по времени фиксации измерений (например, записи показаний термометра на магнитный носитель с целью последующей обра-

ботки результатов на ЭВМ) при планировании эксперимента следует определить также моменты времени фиксации измерений t_1, t_2, \dots, t_m . Обобщение задачи планирования эксперимента на этот и другие подобные случаи содержится в [24].

Наконец, в случае необходимости решения обратной задачи (задачи параметрической идентификации) целесообразно использовать метод эволюционного планирования эксперимента (ЭВОП). Пусть в качестве критерия близости траектории модели к данным натурального эксперимента выбрана некоторая выпуклая функция невязок. Задача идентификации сводится тогда к проблеме поиска минимума этого критерия по параметрам модели. Одним из методов поиска экстремума, требующих минимального возможного числа "прогонов" модели при определении направления движения к точке экстремума и сопровождающихся незначительными по объему вычислениями, является последовательный симплексный метод [36]. Отметим, что эффективность метода с ростом числа факторов повышается. Кроме того, в ходе проведения машинного эксперимента легко включать новые или исключать из рассмотрения старые переменные, а также учитывать ограничения на область изменения факторов.

Сложность структуры имитационной модели, необходимость создания банка моделей, объединенного с соответствующей базой данных и интерфейсом пользователя, — все это порождает особые требования, которые следует предъявить к математическому и программному обеспечению задач моделирования. Можно без преувеличения сказать, что хорошо продуманная организация имитационной системы является залогом ее жизнеспособности. Достаточно вспомнить, что появление простого языка DYNAMO позволило Дж. Форестеру развить новое направление в моделировании — системную динамику, а это обеспечило получение эффективных результатов для не очень сложных моделей [2, 29].

Тем очевиднее тот факт, что в значительно более сложной ситуации, с которой мы встречаемся в задачах имитационного моделирования экосистем, эффективность использования ЭВМ для получения содержательных выводов решающим образом зависит от качества программного обеспечения.

Основной целью, которая ставится при разработке программного продукта в данном случае, является автоматизация процесса конструирования модели и ее дальнейшей эксплуатации. С этой точки зрения имитационная система (ИС) может рассматриваться как пакет прикладных программ (ППП; см., например, [7]), ориентированный на применение в некоторой предметной области¹⁾: имитационные системы, как и пакеты программ, являются совокупностями программных модулей и средств управления их работой. Существует, однако, и ряд особенностей, отличающих ИС от ППП.

Если ППП разрабатываются для "решения" задач определенного класса, то "прогонка моделей" в рамках ИС — только одна из составных частей имитации. Включение в ИС программных средств, обеспечивающих автоматизацию планирования экспериментов, систем управления базой данных (СУБД) — вот те особенности, которые отличают имитационные системы от пакетов программ, если оставаться в рамках общепринятого понимания последнего термина [7, 23].

В настоящей главе излагаются основные принципы организации общесистемного и прикладного математического и программного обеспечения имитационных систем.

¹⁾ В нашем случае эту предметную область можно определить как "динамику экологических систем".

§ 5.1. Общая структура программного обеспечения имитационной системы

Рассматривая в гл. 4 особенности задач моделирования, мы выделили основные этапы технологии процесса имитации. Программные средства должны обеспечить возможность проведения на ЭВМ всего комплекса этих работ, начиная с этапа 1.4. Реализация достаточно сложной имитационной системы может происходить в течение длительного времени. Она может потребовать нескольких месяцев (и даже лет) работы. В связи с этим, например, этапы, связанные с отладкой программ, наполнением базы данных, прогонками модели и т.д., разделены во времени. В то же время для блочных моделей, содержащих десятки отдельных модулей, даже вопросы комплексной отладки программ представляют определенные трудности и этот процесс требует специальной организации. В настоящем параграфе мы рассмотрим общую схему комплекса программ имитационного моделирования, начав с формулировки общих требований, предъявляемых к пакетам такого рода.

5.1.1. Структура и состав программного комплекса. Образующий имитационную систему программный комплекс можно представить себе состоящим из двух основных частей: общесистемной части и прикладного "наполнения". Общесистемную часть образуют программы, организующие процесс имитации, т.е. комплексной отладки и "сборки" модели, управления базами данных, планирования машинного эксперимента, организации диалога пользователя с ЭВМ. Прикладная часть — это программные модули, реализующие собственно специфический счет для данной модели (класса или банка моделей), наполнение базы данных той или иной специальной информацией и т.д. Эти две части хотя и тесно связаны между собой, однако каждая из них имеет свою специфику, которая и будет рассмотрена в дальнейшем.

Требования к программному обеспечению имитационных систем следует формулировать, исходя из интересов и квалификации потенциальных пользователей этой системы. К таким пользователям могут относиться:

1) Пользователи — разработчики имитационных моделей, т.е. системные программисты и математики, принимающие участие в создании и совершенствовании как общесистемной части пакета программ, так и его прикладного наполнения. Как указывалось в гл. 4, для этих целей создается так называемая "базовая модель", с помощью которой отрабатываются и совершенствуются основные принципы моделирования и способы организации имитационного процесса в целом.

2) Инженеры — математики и прикладные программисты, т.е. специалисты, способные с использованием "готовой" общесистемной части разрабатывать и совершенствовать сами модели или их отдельные блоки. Специалисты такого рода должны иметь представление о структуре и возможностях общесистемной части ИС, используя ее для создания новых моделей, т.е. фактически для расширения области практического применения системы.

3) Конечные пользователи — специалисты из той предметной области, для которой выполняется разработка (биологи, экологи, агрометеорологи и другие специалисты). Конечные пользователи ИС имеют дело лишь с конкретным содержанием моделей, т.е. с тем цифровым материалом, который

образует наполнение базы данных, константы модели, ее начальные состояния. Они способны спланировать имитационный эксперимент, проанализировать и дать содержательную интерпретацию результатам моделирования. Очевидно, что от конечных пользователей не требуется знания "внутреннего" устройства модели, численных методов решения задач и конструкции самих программ. Как указано в [7], для них "математическое и программное обеспечение по принципу "черного ящика" является идеальным; соответствующие инструкции указывают его назначение, способ задания некоторых данных и форму представления результатов". Очевидно, что общение с ИС на уровне конечного пользователя (желательно в режиме диалога) должно производиться на языке, в максимальной степени использующем понятия той предметной области, для которой создается модель.

Мы уже упомянули о том, что все программное обеспечение можно разделить на две части — общесистемную и прикладную. Производя классификацию по другому признаку, его можно представить также состоящими из "внутреннего" и "внешнего" обеспечения. Внутреннее программное обеспечение призвано организовать вычислительный процесс при наиболее экономном использовании машинных ресурсов, в то время как целью внешнего является экономия ресурсов человека. Внешнее ПО также может включать в себя общесистемную и прикладную части. Так, очевидно, к общесистемной части относятся программы управления базой данных, диалоговые средства общения с моделью, программы управления машинным экспериментом. Прикладная часть внешнего ПО — это, например, простые модели регрессионного типа, те или иные модели принятия решений и т.д.

Приведем пример. В агрометеорологии разработан целый ряд простых регрессионных соотношений, позволяющих прогнозировать сроки сева яровых или озимых культур, сроки наступления фенологических фаз развития растений, производить оценку величины ожидаемого урожая и т.д. Эти расчеты основаны на использовании так называемых обобщенных агрометеорологических показателей, таких как гидротермический коэффициент (ГТК), биоклиматический показатель (БКП), коэффициент использования посевами фотосинтетически активной радиации и другие. Основанная на описании механизмов явления детальная имитационная модель может быть использована для уточнения оценок коэффициентов этих регрессионных уравнений, а также для определения границ их применимости. Учитывая, что работа с подобными относительно простыми моделями для агрометеоролога является делом привычным и вполне освоенным, они могут быть включены в качестве составной части в прикладное внешнее математическое и программное обеспечение имитационной системы.

Перейдем к формулировке общих требований, которые должно удовлетворять ПО имитационных систем. Совокупность всех этих требований можно разделить на две группы. Первая из них регламентирует те особенности ПО, которые должны быть присущи всем пакетам прикладных программ [7]; вторая относится к тем специальным условиям, которые отражают специфику имитационных систем. Перечислим эти требования.

1.1. Полнота отображения содержательных моделей, описывающих механизмы отдельных протекающих в системе процессов и их взаимодействия.

1.2. Эффективность реализованных алгоритмов, определяемая затратами ресурсов ЭВМ (машинного времени, памяти и др.) для воспроизведения свойств исследуемой системы с необходимой точностью.

1.3. Простота и удобство эксплуатации системы.

1.4. Надежность алгоритмов и программ, высокая степень отлаженности, наличие разнообразных тестовых задач, контроль входной информации и защита от получения абсурдных результатов.

1.5. Расширяемость системы, обеспечение возможности пополнения банка моделей — увеличения числа классов моделей и количества версий внутри каждого класса без существенной переделки программного обеспечения.

1.6. Адаптируемость ПО, т.е. его мобильность по отношению к изменению конфигурации вычислительной системы.

Кроме того, в имитационной системе должны быть предусмотрены:

2.1. Способы организации данных, обеспечивающие простоту и эффективность машинных экспериментов с моделью как на стадии их стратегического, так и тактического планирования и реализации.

2.2. Организация хранения в базе данных результатов прошлых машинных экспериментов с указанием условий их проведения и характеристик использованного плана эксперимента.

2.3. Возможность имитации стохастических процессов, т.е. генерации случайных величин и временных рядов с заданными вероятностными свойствами.

2.4. Наличие программ статистической обработки данных имитационного эксперимента, позволяющих связать в единый комплекс "внутреннюю" и "внешнюю" части программного обеспечения.

2.5. Программные средства, обеспечивающие трассировку модели, выдачу на экран видеотерминала, графопостроитель или АЦПУ промежуточных данных моделирования по запросу пользователя, а также возможность корректировки входной информации в режиме диалога.

Некоторые из перечисленных выше требований очевидны. Они вытекают из тех особенностей имитационных систем, о которых говорилось ранее. Так, например, необходимость той или иной визуализации результатов промежуточных вычислений следует из того, что процесс "прогонки" сложной модели может занимать несколько минут, а иногда несколько десятков минут машинного времени. Ясно, что в этих условиях всегда нужно иметь возможность своевременно убедиться в том, представляет ли данная реализация интерес для дальнейшей обработки или целесообразно прервать вычисления. Другие из перечисленных выше требований связаны с самим характером процесса имитации как длительного экспериментирования с моделью с целью изучения ее чувствительности, решения задач идентификации, постановки и решения разнообразных прикладных задач. Имитационная система всегда создается как "постоянно действующая" система, развиваемая и пополняемая по мере накопления новых экспериментальных фактов, развития наших представлений об особенностях функционирования моделируемого объекта, появления нового "социального заказа" на решение тех или иных прикладных проблем.

Перечислим основные компоненты ПО имитационной системы. К ним относятся:

— библиотека программных модулей, образующих содержательное (функциональное) наполнение системы;

- программные средства для организации процедуры последовательного планирования машинных экспериментов;
- средства управления процессом имитации, которые состоят из подсистемы управления "прогонкой" при получении одной траектории модели и набора таких траекторий;
- система управления базой данных;
- интерфейсные модули для связи модели с базой данных;
- интерфейс пользователя, предоставляющий пользователям различной квалификации возможность работать с имитационной системой на трех уровнях ее организации.

Отметим, что различные компоненты имитационной системы в разной степени связаны с ее прикладной направленностью. Наибольшую связь с предметной областью модели имеют функциональные блоки системы и ее внешнее программное обеспечение. Именно в этой части пакета реализуются как численные схемы, с помощью которых осуществляется численное интегрирование дифференциальных уравнений, описывающих динамику системы, так и специфические алгоритмы, связанные с решением прикладных задач (например, задач прогноза поведения системы при тех или иных внешних возмущениях, задач управления и др.). Ряд других модулей обладает свойствами некоторой универсальности и может использоваться при реализации динамических моделей широкого класса. Именно их мы и относим к общесистемному программному обеспечению. Это управляющие программы, средства организации эксперимента, СУБД. Промежуточное место занимают интерфейсные модули, обеспечивающие связь общесистемной и прикладной частей программного комплекса. Более того, можно утверждать, что выделение этих интерфейсных модулей в особую группу позволяет проектировать системные программные средства безотносительно к той предметной области, в которой будет использоваться данная имитационная система. В связи с этим далее мы можем рассматривать две части ИС, общесистемную и прикладную, в отдельности.

5.1.2. Прикладное математическое и программное обеспечение. Блочная структура моделей экосистем означает, что каждый из блоков, представляющих собой описание некоторой группы функционально однородных процессов, в то же время реализует численную схему решения той или иной математической задачи. В любой имитационной модели естественных экосистем должны быть представлены:

- группа блоков, описывающих процессы энерго- и массообмена в абиотической части системы;
- группа блоков, описывающих процессы первичного биосинтеза – накопления органического вещества фотосинтезирующими организмами;
- группа блоков, описывающих как внутрипопуляционные взаимодействия, так и экологические взаимодействия популяций различных видов (конкуренция, взаимодействие по типу "хищник – жертва" и т.д.).

Рассмотрим особенности возникающих при моделировании математических задач на примере агроэкологической системы. Процессы первой группы – это энерго- и массообмен в системе почва – растение – атмосфера, т.е. тепло- и влагоперенос в почвенно-растительном комплексе, включая транспорт воды по проводящей системе растений и ее испарение через

устычные отверстия и кутикулу, турбулентный тепло- и влагоперенос в приземном воздухе, радиационный режим посева (включая коротковолновую и длинноволновую составляющие радиации), газообмен посева, а также перенос различных минеральных веществ в почве.

Вторую группу процессов составляют фотосинтез, транспорт подвижных веществ по растению, биосинтез и распад аминокислот и белков, а также образование высокомолекулярных углеводов в различных органах растений, рост наземной части посева и корней, развитие растений и формирование конечного урожая. Наконец, процессами третьей группы являются экологические взаимодействия культурных растений с другими видами – сорняками, вредителями, болезнетворными организмами, а также экологические взаимодействия популяций различных видов животных, обитающих в почве.

Рассмотрим специфику математического описания каждой из групп процессов, перечисленных выше. При этом необходимо обосновать:

– выбор численной схемы решения математической задачи в каждом отдельном блоке;

– выбор метода "согласования" решений отдельных задач.

Обратимся первоначально к первой группе процессов. Моделирование энерго- и массообмена является необходимым звеном в решении любой задачи исследования наземных экосистем. Это и понятно, поскольку первичные продуценты – зеленые растения – получают энергию и "строительный материал" из почвы и приземного воздуха за счет переноса этих субстанций к корням и листьям из окружающей среды. Описание этих процессов традиционно относится к задачам математической физики. В агроэко-системах особенности задач этого рода лишь в малой степени связаны непосредственно с видом возделываемой культуры, а определяются в основном типом архитектуры посева. Так, для всех культур сплошного сева в случае горизонтально однородного посева мы приходим к одномерной задаче. С некоторым приближением к одномерной задаче могут быть сведены и описания процессов энерго- и массообмена для рядковых посевов (кукуруза, картофель). В других случаях, например при шпалерных посадках виноградника, следует остановиться на периодической двумерной структуре.

Итак, рассматриваемые задачи можно представить в виде

$$A\varphi = f, \quad (5.1.1)$$

где φ и f – функции одной, двух или трех переменных, а A – квазилинейный дифференциальный оператор. Решение $\varphi(t, x, y)$ ищется в одно- или двумерной области G , на границе которой g заданы краевые условия. Для нестационарного случая, кроме того, задается начальное условие.

Рассмотрим наиболее распространенный одномерный случай. Как показывает анализ возникающих при этом задач, общий вид дифференциального оператора $A(\varphi)$ можно представить в следующем виде:

$$A\varphi = A_1 \frac{\partial A_2 \varphi}{\partial t} - \frac{\partial}{\partial x} \left(A_3 \frac{\partial \varphi}{\partial x} - A_4 \varphi - A_5 \right) + A_6 \varphi, \quad (5.1.2)$$

причем в зависимости от вида коэффициентов A_i ($i = \overline{1, 6}$) мы имеем либо нестационарное линейное уравнение (A_i – заданные функции t и x), либо

нелинейную задачу (A_i или краевые условия зависят от решения $\varphi(t, x)$). Так, например, при описании влагопереноса в ненасыщенной зоне почвы $A_1 = \mu(\varphi)$ — дифференциальная влагоемкость почвы, $A_3 = A_5 = k(\varphi)$ — коэффициент теплопроводности, $A_2 = 1$, $A_4 = A_6 = 0$. При описании теплопереноса в почве $A_1 = C(t, x)$ — теплоемкость почвы, $A_3 = \lambda(t, x)$ — ее теплопроводность, $A_2 = 1$, $A_4 = A_5 = A_6 = 0$.

Интегрирование дифференциальных уравнений методами конечных разностей [22] или конечных элементов [13] приводит к вычислению приближенных значений функций φ_l в узлах сетки Q_l ($l = 1, 2, \dots, L$), где L — общее число узлов. Для этих целей уравнение (5.1.1) аппроксимируется системой линейных или нелинейных сеточных уравнений, коэффициенты которых определяются, исходя из значений коэффициентов исходных дифференциальных уравнений, расстояний между соседними входящими в шаблон узлами, а для околограничных узлов рассчитываются с учетом краевых условий.

Особенности реализации программного комплекса, обеспечивающего решение этих задач, определяются спецификой используемых сеточных методов, а именно:

- способами представления исходных данных;
- способом дискретизации задачи;
- методами решения сеточных уравнений;
- анализом решений на каждом временном шаге и принятием решений о дальнейших вычислениях.

Уравнения, описывающие процессы энерго- и массопереноса, образуют достаточно сложную систему, поскольку они связаны друг с другом либо посредством источник-стоковых членов, либо условиями "сшивания" решений на границах соответствующих областей. Так, уравнения влагопереноса в почве и в приземном воздухе связаны условием неразрывности потока на границе почва — воздух, а также условием равенства скорости интегрального поглощения воды корнями и испарения влаги растениями. В связи с этим под исходными данными задачи понимается тип системы дифференциальных уравнений, явный вид его коэффициентов, граничных и начальных условий, а также конечных соотношений, отражающих связь отдельных задач. Этап дискретизации задачи связан с переходом от области непрерывного изменения аргумента к области его дискретного изменения. Обоснование корректности такого перехода связано с доказательством устойчивости разностной схемы и сходимости решения разностной задачи к дифференциальной [13, 22], а также с получением оценок погрешности аппроксимации. Учитывая требование соблюдения баланса по основным переменным, целесообразно применять консервативные однородные разностные схемы, основанные на использовании интегро-интерполяционного метода.

Для нестационарного случая решение сеточных уравнений находится для последовательности временных интервалов. При этом могут использоваться прямые или итерационные методы, явные, неявные и смешанные схемы (например, схема Кранка — Николсона [13]). Для нестационарной задачи влагопереноса в почве выполнено подробное сравнение существующих методов [28]. Однако для решения общей задачи энерго- и массообмена такое исследование отсутствует. В то же время следует заметить, что

обоснование эффективного метода решения, обеспечивающего получение результата за минимальное время с приемлемой точностью, весьма способствовало бы прогрессу в этой области. В настоящее время решение задачи производится по "явно-неявной" схеме: неявная схема с линеаризацией или итерациями используется для каждого из блоков, а передача информации от одного блока к другому производится с запаздыванием на один временной шаг.

Описание двух других групп процессов (продуцирование биомассы и его перераспределение в экосистеме) не приводит к таким сложностям, хотя, быть может, это связано с их недостаточной изученностью в настоящее время. Основные трудности здесь возникают на этапе перехода от содержательного описания к формальной модели, а также на этапе идентификации, поскольку получение точных количественных соотношений о моделируемых процессах сопряжено с необходимостью постановки весьма тонких и дорогостоящих экспериментов. В то же время именно некоторая неопределенность в трактовке экспериментального материала как раз и позволяет выбрать простейшее описание из всех возможных. В связи с этими обстоятельствами в экологии, по-видимому, и появились, а затем стали традиционными так называемые компартментальные модели, о которых говорилось в предыдущих главах. Процессы, отвечающие этим блокам модели, описываются системами нелинейных (как правило) дифференциальных уравнений, численное интегрирование которых производится с использованием стандартных методов (например, одним из методов типа Рунге – Кутты).

Рассмотрим вопрос сопряжения отдельных блоков модели. Из анализа уравнений следует, что временные характеристики различных блоков могут различаться между собой на несколько порядков. Так, например, время переходного процесса турбулентного теплопереноса в посеве составляет доли минуты, постоянные времени процессов тепло- и влагопереноса в почве могут изменяться от одного до нескольких часов, а инерционность ростовых процессов достигает суток.

Известно, что временной шаг модели в целом должен определяться наиболее "быстрыми" составляющими решения. Это приводит к определенным трудностям при численном интегрировании общей системы. Системы дифференциальных уравнений, в которых можно выделить "быстрые" и "медленные" движения, принято относить к классу "жестких систем". Они обладают следующими основными свойствами [21]:

1. Жесткие уравнения, как правило, имеют два участка решения с существенно различным характером поведения, причем продолжительность первого участка значительно меньше второго.

2. Общее решение жесткого уравнения очень быстро стремится к частному "медленному" решению, независимо от начальных условий.

3. Даже при очень малых отклонениях от "медленного" решения производная резко возрастает по сравнению с ее значением на этом решении.

Эти обстоятельства нельзя не учитывать при выборе метода численного интегрирования уравнений модели. Как показывает анализ, в общей системе можно выделить три группы уравнений, обладающих различными временными свойствами.

К группе быстрых процессов относятся процессы влаго- и теплопереноса в посеве, процессы газопереноса в посеве, процессы диффузии CO_2 в меж-

клетнике и его растворения, а также процесс карбоксилирования. При выборе базового временного шага модели, равного одному часу или более, эти процессы следует рассматривать как квазистационарные, т.е. можно с достаточной точностью полагать, что переменные блока отслеживают все внешние изменения безынерционно.

Вторая часть блоков описывает процессы, протекающие в некотором "среднем" темпе; постоянные времени при этом равны нескольким часам. Уравнения необходимо интегрировать совместно при "замороженных" параметрах, характеризующих медленные изменения архитектуры растений, высоты посева, глубины проникновения корней и других аналогичных им переменных. Поэтому при реализации численной схемы можно полагать, что медленные накопления биомассы растений скачкообразно изменяются в процессе вегетации сравнительно небольшими ступенями, т.е. представляют собой кусочно-постоянные функции времени. Такие представления приводят к небольшим ошибкам в расчетах, но позволяют отказаться от пересчета коэффициентов уравнений средних движений, связанных с архитектурой, на каждом шаге их интегрирования.

Наконец, уравнения, описывающие медленные процессы, можно интегрировать с тем же или с большим шагом, что и уравнения для "средних" процессов. В последнем случае необходима специальная организация счета, связанная, например, с учетом накопления ассимилятов за несколько базовых временных шагов и с другими аналогичными кумулятивными процессами.

Таким образом, прикладной программный комплекс должен содержать модули, обеспечивающие решение следующих задач:

- выбор типа уравнений из некоторого класса и его параметризация;
- выбор типа и параметров пространственно-временной сетки;
- выбор конкретной процедуры численного интегрирования уравнения и задание ее параметров;
- обеспечение той или иной численной схемы согласования на каждом шаге решений отдельных уравнений или систем уравнений и пошаговое заикливание счета с возможной проверкой точности решения и автоматическим дроблением шага.

5.1.3. Общесистемное внутреннее программное обеспечение. Общесистемное внутреннее программное обеспечение предназначено для автоматизации процесса разработки и отладки модели, в то время как внешнее ПО призвано автоматизировать работу внешнего пользователя с моделью. Мы уже упоминали о том, что ЭВМ, являющиеся вычислителями дискретного действия, изначально не приспособлены к решению задач моделирования. Именно с этим обстоятельством и связана необходимость разработки достаточно сложных программных средств, призванных "рутинизировать" наиболее сложные и трудоемкие виды работ с моделью.

Как отмечается в литературе, при моделировании сложных систем, включающих сотни и тысячи переменных, недостаточно той элементарной организации данных, которая предусмотрена в универсальных языках, таких, например, как Фортран, Алгол, ПЛ/1 и др. [16]. Система имитационного моделирования должна содержать инструкции, облегчающие организацию исходных данных, позволяющие видоизменять состав и структуру модели, оперировать с массивами переменной структуры и размера, обеспе-

чивающие экспериментатору доступ к хранящейся в памяти машины информации и контроль за информационными потоками.

Имеются два пути реализации этих требований.

1. Разработка проблемно-ориентированного языка высокого уровня, так называемого SPL (Simulating Programming Language), и соответствующего транслятора. Очевидно, в этом случае все перечисленные выше требования могут быть реализованы с необходимой полнотой. Однако создание SPL – трудоемкий процесс и в настоящее время не существует языка моделирования, позволяющего конструировать и исследовать модели экосистем достаточно сложной структуры.

2. Создание пакета программ, оформленного на каком-либо из существующих языков (или использующего несколько языков). К подобным пакетам относятся GASP, FORSIM [34, 36, 37, 40–42], СИМФОР и др. Очевидно, что и в этом случае требуется разработка специального языка – входного языка пакета. Но в то же время эта разработка является значительно менее трудоемкой и требует привлечения меньшего числа программистов, чем создание SPL.

Как в том, так и в другом случаях система моделирования должна обеспечить выполнение ряда функций и в том числе [25]:

1. Предоставить в распоряжение пользователя, строящего модель, концептуальную основу для ясного и четкого мышления. Основные элементы языка системы моделирования должны быть применимы к широкому классу явлений. Применение языка заключается в том, чтобы отождествить компоненты исследуемой системы с соответствующими элементами языка, описать их на языке принятых понятий и установить соотношения между этими компонентами.

2. Предоставить в распоряжение пользователя средства для описания динамики поведения моделей.

3. Облегчить процесс программирования.

При этом две последние функции выполняются, как правило, на базе совместной работы общесистемного и прикладного ПО имитационной системы.

а) Существующие системы моделирования

Для некоторых систем моделирования характерно представление системы сразу в виде дифференциальных уравнений или уравнений в конечных разностях. Известным языком этого типа является язык DYNAMO. Обычно уравнения устанавливают соотношения между значениями функций в моменты времени t , $t + \Delta t$ и значениями их производных в момент времени $t + 0,5 \Delta t$. При этом каждое уравнение принадлежит одному из заданных классов, а число этих классов в системе моделирования ограничено.

При моделировании непрерывные изменения, происходящие в системе, представляются в виде некоторой совокупности отдельных событий – моделями дискретных событий.

Преимуществом рассматриваемого способа моделирования является то, что многие языки моделирования, используемые для описания дискретных событий, являются в то же время и мощными универсальными алгоритмическими языками, с помощью которых можно дать описание того, что же в действительности "происходит" в исследуемой системе, причем с любой желаемой степенью детализации.

Ниже дано краткое описание особенностей двух наиболее распространенных систем моделирования.

Система моделирования DYNAMO

Система моделирования DYNAMO [2, 29] предназначена для построения и анализа моделей процессов управления производственной деятельностью промышленных, торговых и сбытовых предприятий. Она позволяет понять, каким образом влияют на эффективность работы предприятия его производственная структура, технологический процесс производства и временные затраты, осуществляемые на выполнение тех или иных производственных операций.

На первом этапе развития системы моделирования DYNAMO были разработаны набор стандартных блоков и специальный язык, позволяющие описывать имитационные модели производственных и экономических объектов, которые можно описать в терминах системы рекуррентных уравнений.

Затем к этой системе был разработан специальный язык имитационного моделирования DYNAMO. По своему синтаксису этот язык очень похож на язык ФОРТРАН. Основные компоненты языка почти те же, что и в языке ФОРТРАН. Язык DYNAMO позволяет реализовать на ЭВМ рекуррентный процесс численного решения системы разностных уравнений, с помощью которых описывается модель.

Процесс разработки модели в системе моделирования DYNAMO может быть представлен в виде последовательности следующих основных этапов;

- формализация разрабатываемой модели в терминах языка DYNAMO;
- задание начального состояния модели;
- определение значений входных переменных;
- задание констант модели;
- определение спецификации вывода;
- определение шифра программы модели;
- определение числа итераций проведения эксперимента на модели.

Начальное состояние моделируемой системы задается уравнениями, записанными на языке DYNAMO. На каждой итерации вычисляются значения переменных модели в конце очередного интервала системного времени. Все интервалы имеют одинаковую длину. Полный период имитации поведения системы равен целому числу элементарных интервалов. Значения переменных модели могут выводиться в конце каждой итерации в табличной или графической форме. Организация вывода и продолжительность процесса моделирования определяется в программе модели специальными инструкциями языка.

Цикл работы программы, написанной на языке DYNAMO, может быть условно представлен в виде последовательной реализации следующих основных этапов:

- запись программы модели в ЭВМ и формирование таблиц, которые будут использоваться при выполнении программы;
- перевод описания модели с языка DYNAMO на внутренний язык машины;
- решение уравнений и вычисление значений переменных модели, а также подготовка данных для вывода необходимой информации;

- вывод результатов счета;
- изменение констант модели и подготовка новой серии расчетов.

Применение системы моделирования DYNAMO на практике существенно облегчает задачу пользователя, который может исследовать поведение модели системы или моделируемого процесса на ЭВМ, не зная ни численных методов решения, ни теории дифференциальных уравнений (следует отметить, что для решения дифференциальных уравнений в системе моделирования DYNAMO используется метод Эйлера). Пользователь сравнительно легко может записать свою модель на языке DYNAMO и проанализировать результаты моделирования на ЭВМ, получив их в удобной для анализа форме.

Система имитационного моделирования СИМФОР/РАФОС

Система имитационного моделирования СИМФОР написана на языке Фортран-IV и работает с операционной системой реального времени с разделением функций (РАФОС) СМ ЭВМ [18].

При исследовании систем с использованием средств СИМФОР под моделированием понимается создание и исследование моделей реальных систем программными средствами СИМФОР. При этом программа модели имитирует процесс функционирования системы, воспроизводя как непрерывно, так и дискретно изменяющиеся параметры системы.

При разработке модели система рассматривается как множество объектов, каждый из которых определяется некоторой совокупностью параметров, существенных с точки зрения описания процесса функционирования системы. Множество значений параметров объектов системы считается упорядоченным и называется состоянием системы.

Моделирование функционирования дискретной системы рассматривается как последовательная смена состояний системы в дискретные моменты времени. Каждое изменение состояния системы связано с изменениями значений параметров ее объектов либо с изменением связей между ними. Таким образом, описание модели исследуемой системы определяется заданием алгоритмов изменения состояния системы, условий порождения и уничтожения объектов системы. В СИМФОР это осуществляется с помощью "событий", "постоянных" и "временных" объектов модели. При описании моделей непрерывных систем полагается, что параметры модели изменяются непрерывно, а связи между ее компонентами и алгоритм функционирования практически не изменяются. В СИМФОР описание непрерывных процессов осуществляется с использованием "структурных" переменных, позволяющих задавать взаимосвязи на языке, близком по форме к дифференциальным или разностным уравнениям.

Возможность описания моделей непрерывно-дискретных систем обеспечивается в СИМФОР наличием совместимых описаний дискретных и непрерывных систем.

СИМФОР предоставляет пользователю необходимые средства для описания модели исследуемой системы. Описание модели осуществляется в соответствии с принятыми в СИМФОР соглашениями на языке ФОРТРАН-IV.

Использование СИМФОР существенно упрощает процесс построения модели, ее программирования и отладки.

При разработке модели с использованием средств СИМФОР необходимо:

- выполнить анализ исследуемой системы с использованием набора понятий, входящих в СИМФОР, т.е. построить описание модели;

- описать логику работы моделей и алгоритмов ее функционирования на языке ФОРТРАН-IV в соответствии с принятыми соглашениями;
- подготовить исходные числовые данные в виде "управляющих перфокарт";
- осуществить отладку программы модели с использованием развитой системы диагностики ошибок.

СИМФОР содержит развитый набор средств управления процессом моделирования:

- запуск модели;
- прохождение модели с использованием трассировки и сбор статистической информации об объектах модели;
- окончание реализации процесса моделирования по условию, определенному программистом;
- обработку статистической информации;
- вывод таблиц, графиков, гистограмм и собранных статистик.

Из приведенного выше краткого описания некоторых из существующих систем следует, что эти пакеты не обеспечивают выполнения всех функций имитационных систем в полном объеме. Язык ДУНАМО в связи с жесткой конструкцией имеет весьма ограниченную область применения. Система имитационного моделирования СИМФОР является более гибкой. Ее основным недостатком для моделирования сложных систем заключается в том, что конструкция языка не допускает блочной структуры модели – вектор правых частей дифференциальных уравнений должен быть описан целиком с самого начала. В связи с этим в следующем пункте будет рассмотрена более совершенная, хотя и более узко специализированная система СИМОНА [19].

б) Система имитационного моделирования динамики агроэкосистем – СИМОНА

Система имитационного моделирования нестационарных процессов в агроэкосистемах СИМОНА [19] предназначена для исследования сложных динамических систем, в частности агроэкосистем, путем построения и анализа имитационных моделей, воспроизводящих функционирование процессов в растительном покрове и в окружающей среде. Динамические модели этих процессов представляются в виде систем дифференциальных уравнений как в обыкновенных, так и в частных производных. Реализуемая в системе СИМОНА имитационная модель имеет блочную структуру. При этом каждый из блоков описывает ту или иную группу функционально однородных процессов и в то же время решает определенную математическую задачу. Выходной информацией, вырабатываемой блоками на каждом временном шаге модели, является его вектор состояния $x^i(t_{k+1})$, где t_{k+1} – время окончания k -го шага. Входными данными блоков служат: (1) полный вектор состояния модели в начале шага, т.е.

$$x(t_k) = \begin{bmatrix} x^1(t_k) \\ x^2(t_k) \\ \vdots \\ x^n(t_k) \end{bmatrix},$$

где n – число блоков; (2) вектор параметров модели $p = (p_1, p_2, \dots, p_m)^*$,

представляющий собой набор констант модели; (3) вектор-функции входных управляющих воздействий $u(t_k)$ ($k = 1, 2, \dots$) и внешних возмущений $w(t_k)$ ($k = 1, 2, \dots$); (4) набор величин, определяющих параметры расчетной схемы – число и расположение узлов пространственной сетки, базовый шаг интегрирования ST , требуемая точность решения дифференциальных уравнений и т.д.

Обращение к блокам модели осуществляется поочередно в произвольном порядке. После проработки всех блоков вычисляется новое состояние модели в момент времени $t_{k+1} = t_k + ST$, после чего система переходит к следующему временному шагу. В настоящее время существуют две версии системы – СИМОНА-ЕС, работающая в пакетном режиме, и СИМОНА-СМ, работающая в режиме диалога.

Как указывалось в гл. 4, при исследовании достаточно сложных систем, к каковому, бесспорно, относится и экосистеме, необходимо создание состоящего из нескольких классов банка моделей. В соответствии с этим достаточно развитое программное обеспечение ИС должно допускать работу с одной моделью из данного класса, работу с классами и с банком моделей. Рассмотрим, как эти возможности реализуются в системе СИМОНА.

Понятие банка моделей мы связали с существованием единой, обслуживающей эти модели базы данных. В соответствии с этим на уровне банка моделей необходимо произвести согласование глобальной логической организации данных (концептуальной модели данных) и ее внешней организации с областью данных банка моделей. Именно с этими требованиями связано выделение вектора состояний модели (в том числе вектора начального состояния), вектора параметров, двух типов входных воздействий (возмущающих воздействий и управлений) а также выхода модели. Последний представляет собой совокупность тех интегральных величин, которые характеризуют процесс моделирования в целом и должны использоваться при выработке тех или иных решений. Отметим, что существующие СУБД (например, СУБД "Спектр" для ЕС-ЭВМ или "Квант-М" для СМ-ЭВМ) весьма неудобны для совместной работы с "большими" пользовательскими программами. Поэтому необходима разработка специальных интерфейсных модулей, обеспечивающих "перекачку" информации из базы данных в область данных модели и обратно.

Дальнейшая конкретизация структуры модели производится на уровне класса моделей, который мы определили в гл. 4 как совокупность моделей, характеризующихся единством обмена информации между блоками. Очевидно, что целесообразной организацией работы с классами является создание библиотеки блоков с указанием их версий. В рамках системы СИМОНА это осуществляется с использованием специальных модулей CATLG и TYPES.

Работа с моделью из данного класса начинается с выбора из общего числа содержащихся в библиотеке блоков по одной версии для каждого блока, а также указания условий выполнения каждого блока. Последнее необходимо, например, потому что не все блоки должны выполняться на каждом шаге модели. Более того, включение некоторых блоков обусловлено совокупностью ряда событий, которые в каждой реализации модели заранее не определены (например, включение в работу блока формирования генеративных органов зерновых культур связано с завершением фазы выхода в

трубку, прохождение которой в свою очередь определяется погодными условиями и агротехникой). Отметим, что автоматизация процесса сборки модели из некоторого заранее не фиксированного набора блоков существенно облегчает работу с моделью на стадиях ее конструирования, отладки и идентификации. Эту работу выполняет модуль BEGIN.

Собственно процесс выполнения модели начинается с присваивания определенных числовых значений параметрам модели, ее начальному вектору и входным воздействиям. В рамках системы СИМОНА эти функции выполняет модуль INPUT. Наконец, организация процесса вычисления траектории модели с возможными прерываниями производится с помощью работ BEGIN и GOING.

Задание на сборку и функционирование модели пишется на языке управления пакетом и в варианте СИМОНА-ЕС набивается на перфокартах или вводится с пульта видеотерминала.

Система имитационного моделирования вводит это задание, запоминает его, формирует служебные поля и выполняет перечисленные в задании работы.

Опишем основные функции языка пакета.

1. $\langle \text{задание на работу} \rangle ::= \langle \text{работа} \rangle | \langle \text{задание на работу} \rangle \langle \text{работа} \rangle$
2. $\langle \text{работа} \rangle ::= \langle \text{имя работы} \rangle \langle \text{параметры работы} \rangle$
3. $\langle \text{имя работы} \rangle ::= \text{CATLG} | \text{TYPEC} | \text{INPUT} | \text{GOING} | \text{BEGIN} | \text{ENDWR}$

В системе имитационного моделирования предусматривается пять видов работ и ключевое слово ENDWR, означающее, что список работ завершен.

Идентификатор CATLG означает, что необходимо составить каталог переменных модели (под переменной модели мы имеем в виду элемент массива состояния модели). Каждая переменная функционирует в модели и доступна любому блоку по ее номеру. Номером переменной является ее порядковый номер в массиве состояния модели, но поскольку пользователю удобнее работать не с номерами переменных, а с их мнемоническими именами, в системе составляется каталог, устанавливающий взаимно однозначное соответствие между мнемоническим именем и номером переменной. Это позволяет при последующих распечатках и статистической обработке использовать собственные имена переменных. Работа CATLG и означает работу по вводу мнемонических имен с перфокарт, составление каталога и запись этого каталога на МД или МЛ. Кроме того, эта же работа может выполнять и другую функцию — формирование библиотеки системы.

Идентификатор TYPEC означает, что необходима распечатка каталога переменных модели или библиотеки системы. Идентификатор INPUT означает, что необходимо провести работу по вводу с перфокарт, МД или МЛ начальных данных модели и записать эти данные в поле состояния модели. Идентификатор GOING означает, что модель в данной форме уже функционировала, что на МД хранятся записи пошаговых состояний, задание на работу и массив последнего состояния модели. Эта работа заключается в том, чтобы считать с МД необходимые данные и продолжать выполнение модели в соответствии с заданием. Все вновь вырабатываемые состояния модели должны записываться на МД вслед за выработанным ранее. Идентификатор BEGIN означает, что модель в данной форме только начи-

нает функционировать и что никакой информации о предшествующей работе (кроме каталога переменных и библиотеки системы) на МД не имеется.

Естественно, что работы GOING и ENDWR не имеют никаких параметров. Для их выполнения достаточно подложить в колоду перфокарту, в первых пяти колонках которой набито имя работы GOING или ENDWR. Что касается остальных работ, то для их функционирования необходимы дополнительные сведения (параметры работы).

4. $\langle \text{параметры работы} \rangle ::= \langle \text{параметры каталогизации} \rangle | \langle \text{параметры начальных данных} \rangle | \langle \text{параметры модели} \rangle \langle \langle \text{Параметры каталогизации} \rangle \rangle$ используются с работами CATLG и TYPES.

5. $\langle \text{параметры каталогизации} \rangle ::= \langle \text{виды работы} \rangle [\langle \text{список описаний} \rangle]$

6. $\langle \text{вид работы} \rangle ::= \text{КАТЛГ} | \text{БИБЛИ}$

Параметр «вид работы» используется в связи с тем, что программы, выполняющие работы CATLG и TYPES, могут формировать (распечатывать) либо каталог переменных, либо библиотеку переменных, либо и то и другое вместе.

В работе TYPES используется только вид работы. И задание на распечатку может выглядеть так:

БИБЛИ

КАТЛГ

Каждый вид работы набивается в первых пяти колонках отдельной перфокарты. Если должны выполняться обе работы, то их порядок фиксирован: сначала библиотека, а потом каталог.

При выполнении работы формирования модели необходимо подложить в колоду перфокарт «список описаний».

7. $\langle \text{список описаний} \rangle ::= \langle \text{описание} \rangle | \langle \text{список описаний} \rangle$
 $\langle \text{описание} \rangle$

8. $\langle \text{описание} \rangle ::= \langle \text{описание имени} \rangle \langle \text{разделитель описаний} \rangle$

9. $\langle \text{описание имени} \rangle ::= \langle \text{описание имени блока} \rangle | \langle \text{описание имени переменной} \rangle$

10. $\langle \text{описание имени блока} \rangle ::= \langle \text{имя блока} \rangle \langle \text{начало интервала} \rangle \langle \text{конец интервала} \rangle$

11. $\langle \text{имя блока} \rangle ::=$ последовательность из пяти символов

12. $\langle \text{начало интервала} \rangle ::= \langle \text{номер переменной} \rangle$

13. $\langle \text{конец интервала} \rangle ::= \langle \text{номер переменной} \rangle$

14. $\langle \text{номер переменной} \rangle ::=$ последовательность из четырех цифр

15. $\langle \text{описание имени переменной} \rangle ::= \langle \text{имя переменной, номер переменной} \rangle$

16. $\langle \text{имя переменной} \rangle ::=$ последовательность из пятнадцати символов

17. $\langle \text{разделитель описаний} \rangle ::= , | K$.

При выполнении работы CATLG в качестве параметров, кроме вида работы, используется список описаний. Каждое описание набивается на отдельной перфокарте. На этой же перфокарте набивается разделитель описаний. Разделитель описаний может принимать два значения: запятая (,) или (K). Разделитель указывает, что список описаний завершен.

С видом работы БИБЛИ употребляются описания блока, а с видом работы КАТЛГ употребляется описание имен переменных. Каждое описание состоит из имени блока и двух номеров переменных: начало и конец интервала. Наличие в описании границ интервала вызвано

тем, что в системе каждому содержательному блоку отводится некоторое число глобальных переменных модели, расположенных подряд в COMMON-области. Эти интервалы в данном классе моделей закрепляются за каждым блоком раз и навсегда и представляют собой векторы состояний блоков. Номер переменной начала интервала — это номер первой переменной вектора состояния блока. Номер переменной конца интервала — это номер последней переменной вектора состояния. С видом работы КАТЛГ употребляется описание имени переменной.

Работа по вводу начальных данных также нуждается в списке параметров. Блоку, выполняющему эти функции, необходимо получать информацию о значении переменных и номерах переменных, которым эти значения следует присвоить. Параметром работы ВВОД является список данных.

18. $\langle \text{параметры начальных данных} \rangle ::= \langle \text{список данных} \rangle$

19. $\langle \text{список данных} \rangle ::= \langle \text{данное} \rangle | \langle \text{список данных} \rangle \langle \text{данное} \rangle$

20. $\langle \text{данное} \rangle ::= \langle \text{номер переменной} \rangle \langle \text{константа} \rangle \langle \text{разделитель описаний} \rangle$

21. $\langle \text{константа} \rangle ::=$ последовательность из восьми цифр с десятичной точкой

Работа BEGIN представляет собой задание на сборку и выполнение модели. Параметрами этой работы являются "параметры модели".

22. $\langle \text{параметры модели} \rangle ::= \langle \text{дата} \rangle \langle \text{шаг} \rangle \langle \text{имя контрольной функции} \rangle \langle \text{признак печати модели} \rangle \langle \text{признак печати протокола} \rangle \langle \text{список предложений} \rangle$

23. $\langle \text{дата} \rangle ::= \langle \text{номер секунды} \rangle \langle \text{номер минуты} \rangle \langle \text{номер часа} \rangle \langle \text{номер числа} \rangle \langle \text{номер месяца} \rangle$

24. $\langle \text{номер секунды} \rangle \langle \text{номер минуты} \rangle \langle \text{номер часа} \rangle \langle \text{номер числа} \rangle \langle \text{номер месяца} \rangle ::=$ двузначное целое число

Первым параметром работы является дата начала работы. Эта информация необходима для того, чтобы содержательные блоки могли рассчитывать потоки радиации, температуру и влажность воздуха и другие переменные, зависящие от времени суток. Дата начала работы набивается на отдельной перфокарте и состоит из пяти двузначных цифр. Например, 0125071504

Эта дата означает, что модель должна начать функционировать с 1-й секунды 25-минуты 7-го часа 15-го апреля.

25. $\langle \text{шаг} \rangle ::= \langle \text{константа} \rangle$

Шаг модели — это разность времени между двумя состояниями модели, выраженная в долях суток. В системе шаг модели может меняться, но с перфокарты вводится так называемый "эталонный шаг модели". Как правило, после этого числа состояние модели будет записываться на МД. Значение шага модели набивается на отдельной перфокарте.

26. $\langle \text{имя контрольной функции} \rangle ::=$ последовательность из пяти символов

Имя контрольной функции — это информация, которая необходима для осуществления контроля модели по точности. В системе предусматривается использование нескольких таких контрольных функций: оценка точности по интервалу, по абсолютной величине интервала и т.д. Имя контрольной функции набивается на отдельной перфокарте.

27. $\langle \text{признак печати модели} \rangle ::= \text{ПЕЧАТ} \mid \text{НЕТПЧ}$

Признак печати модели может принимать два значения: ПЕЧАТ или НЕТПЧ. Если этот признак имеет значения ПЕЧАТ, то после ввода, синтаксического контроля и запоминания модели (список предложений) вся информация будет распечатана на АЦПУ.

28. $\langle \text{признак печати протокола} \rangle ::= \text{NO} \underline{\text{ }} \underline{\text{ }} \mid \text{ALL} \underline{\text{ }} \underline{\text{ }} \mid \text{PART} \underline{\text{ }} \underline{\text{ }}$

Система предоставляет пользователю возможность печатать на АЦПУ протокол о выполнении модели. Если этот признак имеет значение PART $\underline{\text{ }} \underline{\text{ }}$, то в конце каждого шага модели (эталонного) на АЦПУ будет напечатан номер шага и текущее время модели.

Если значение признака ALL $\underline{\text{ }} \underline{\text{ }}$, то на АЦПУ будет печататься не только номер числа и текущее время, но и имена всех выполняемых содержательных блоков внутри каждого шага. Этот режим предусматривается для облегчения отладки модели. Если значение признака NO $\underline{\text{ }} \underline{\text{ }}$, то протокол печататься не будет. Признак печати протокола набивается на отдельной перфокарте в первых пяти колонках. Вслед за этими параметрами следует список предложений – собственно выполняемых операторов.

29. $\langle \text{список предложений} \rangle ::= \langle \text{предложение} \rangle \mid \langle \text{список предложений} \rangle \langle \text{предложение} \rangle$

30. $\langle \text{предложение} \rangle ::= \langle \text{имя операции} \rangle \langle \text{список параметров} \rangle$

Каждое новое предложение начинает пробиваться с новой перфокарты и обязательно содержит имя операции и необходимый для ее выполнения список параметров.

31. $\langle \text{имя операции} \rangle ::= \langle \text{имя содержательного блока} \rangle \langle \text{имя оператора} \rangle$

32. $\langle \text{имя содержательного блока} \rangle ::= \text{последовательность из пяти символов}$

33. $\langle \text{имя оператора} \rangle ::= \text{NOTES} \mid \text{MEMOR} \mid \text{CHECK} \mid \text{EINIS}$

Имя операции набивается на отдельной перфокарте в первых пяти колонках и может быть либо именем содержательного блока (именем оттранслированной, отлаженной, собранной программы, которая хранится в библиотеке системы), либо именем оператора.

В первую версию языка сборки, кроме шести видов работ, перечисленных выше, включены четыре оператора языка, и пятым оператором, в сущности, является имя содержательного блока с условием его выполнения. Если в предложении упоминается имя содержательного блока, это означает, что программа с этим именем будет найдена на МД, загружена в оперативную память и на нее будет передаваться управление всякий раз, как имя этого блока встретится в предложении задания (если при этом выполняются некоторые условия).

Оператор языка CHECK означает, что после окончания очередного шага модели необходимо проверить соответствие рассчитанного состояния модели предыдущему. Если такого соответствия нет, то ряд блоков, перечисленных в параметрах оператора CHECK, будет рассчитан заново с дробленным шагом. Шаг будет дробиться до тех пор, пока между двумя состояниями модели не будет требуемого соответствия. После этого блоки (из числа перечисленных в операторе) будут пересчитываться до тех пор, пока вся модель не выйдет на время, заданное шагом (эталонным) модели. При этом во всех блоках, не принимающих участия в расчетах, вектор их состояния будет интерполироваться.

Оператор языка NOTES означает, что после окончания полного шага модели новое состояние модели должно быть записано на МД. Оператор языка MEMOR означает, что необходимо проверить, не существует ли сигнала оператора (перфокарты или ПЛ) о прерывании работы. При наличии сигнала по этому оператору на МД будет записана COMMON-область и выполнение работы прекращено. Оператор языка FINIS означает, что весь список предложений закончен, что необходимо (в зависимости от условия) либо перейти к выполнению следующего шага модели, либо закончить ее выполнение и перейти к следующей работе.

Из смысла этих трех операторов языка вытекает, что их порядок в списке предложений фиксирован: CHECK, NOTES, MEMOR и FINIS. Операторы CHECK, NOTES и MEMOR могут отсутствовать, но оператор FINIS присутствует всегда. Из функциональных свойств перечисленных выше операторов следует, что у них должны быть различные (по смыслу и по форме) списки параметров.

34. $\langle \text{список параметров} \rangle ::= \langle \text{список условий} \rangle | \langle \text{список проверок} \rangle | \langle \text{список номеров переменных} \rangle$

Всего в первой версии языка имеется три вида списков: список условий, который употребляется с именем содержательных блоков и оператором FINIS, список проверок для оператора CHECK и список номеров переменных для операторов NOTES и MEMOR.

35. $\langle \text{список условий} \rangle ::= \langle \text{условие} \rangle | \langle \text{список условий} \rangle \langle \text{условие} \rangle$

36. $\langle \text{условие} \rangle ::= \langle \text{номер переменной} \rangle \langle \text{знак} \rangle \langle \text{граница} \rangle \langle \text{разделитель параметров} \rangle$

37. $\langle \text{знак} \rangle ::= = | \neq$

38. $\langle \text{граница} \rangle ::= \langle \text{константа} \rangle$

Каждый содержательный блок (или оператор FINIS) будет выполняться (т.е. на них будет передаваться управление) только в том случае, если выполняются все условия, перечисленные в списке условий.

39. $\langle \text{список номеров переменных} \rangle ::= \langle \text{номер переменной} \rangle \langle \text{разделитель номеров} \rangle | \langle \text{список номеров переменных} \rangle \langle \text{номер переменной} \rangle \langle \text{разделитель номеров} \rangle$

40. $\langle \text{разделитель номеров} \rangle ::= =, | - | K$

Список номеров переменных является списком параметров для операторов NOTES и MEMOR. Наличие этого списка параметров для оператора NOTES объясняется следующими соображениями.

Пользователю могут не понадобиться все параметры модели, включенные в массив состояния, так как некоторые из них не нужны для последующей статистической обработки или оценки результатов функционирования модели. Наличие списка номеров позволяет запомнить на МД только необходимые заказчику переменные.

Параметром оператора MEMOR является только одно число в списке номеров. Это число представляет собой число шагов, через которые следует запрашивать оператора о необходимости прерывания работы и запоминания модели на МД. Например,

MEMOR
0100K

означает, что запросы оператору будут производиться через каждые 100 шагов. В первом варианте операционной системы ответом оператора служит

перфокарта, в первых пяти колонках которой располагается слово ЗАПМН | НЕЗАП.

41. $\langle \text{список проверок} \rangle ::= \langle \text{проверка} \rangle | \langle \text{список проверок} \rangle \langle \text{проверка} \rangle$
42. $\langle \text{проверка} \rangle ::= \langle \text{имя содержательного блока} \rangle \langle \text{список номеров переменных} \rangle \langle \text{граница} \rangle \langle \text{разделитель имен} \rangle$
43. $\langle \text{граница} \rangle ::= \langle \text{константа} \rangle$

Список проверок является списком параметров в оператор СHECK. Каждая проверка начинает набиваться с новой перфокарты. На первой перфокарте набивается имя содержательного блока (в первых пяти колонках). Далее следует группа перфокарт со списком номеров переменных, которые набиваются по правилам, описанным выше. И, наконец, на последней перфокарте проверки набивается граница и разделитель списка параметров, указывающий на продолжение или завершение списка проверок.

Наличие в списке предложений оператора СHECK вызовет следующие действия системы: по переменным, перечисленным в списке номеров переменных, будут рассчитаны значения некоторой функции для k -го и $(k + 1)$ -го состояний модели. Если разность этих функций будет меньше или равной значению границы, то это значит, что контроль прошел и что дробление шага в этом блоке производить не следует. В противном случае шаг будет раздроблен и все блоки, не прошедшие контроля, будут пересчитаны. Имя функции, по которой производится контроль, вводится в начале работы вместе с датой и шагом. В системе существует некоторый набор таких функций. Такой контроль будет производиться для всех блоков, указанных в проверках.

Система является открытой для пополнения ее новыми видами функций контроля, которые могут появиться в процессе эксплуатации.

Таким образом, с помощью системы СИМОНА, используя язык управления заданиями, возможно собирать модели из отдельных блоков, осуществлять формирование и ввод начальных данных, выполнять счет по модели и фиксировать его результаты. При этом подготовка входной информации может осуществляться как в автономном режиме, так и с использованием базы данных. Точно так же результат моделирования может либо непосредственно отображаться на устройстве визуализации, либо включаться в состав базы данных. Организация серий машинных экспериментов осуществляется с помощью других модулей имитационной системы.

5.1.4. Автоматизация машинного эксперимента. В последнее время сформировалось и нашло широкое применение понятие "вычислительный эксперимент" [5, 22]. Вычислительный эксперимент характеризуется тем, что для каждой математической модели с целью установления ее адекватности физическому объекту необходимо решать большое число вариантов задачи (варьируя определяющие ее параметры) и, кроме того, частично видоизменять (корректировать, уточнять) саму математическую модель. Именно эти особенности (многомодельность и многовариантность) являются наиболее существенными особенностями этого метода. Кроме того, обычно подчеркивается, что отдельные решения задачи, получаемые в результате очередной реализации вычислительного эксперимента, не отбрасываются, а используются впоследствии для сопоставления с другими расчетами.

С этой точки зрения любой имитационный эксперимент является в то же время вычислительным. Однако понятие имитационного эксперимента является несколько более общим, так что не всякий вычислительный эксперимент можно считать имитационным.

Последнее утверждение, вообще говоря, вытекает из тех особенностей имитационных систем, о которых уже говорилось в гл. 4. Тем не менее для более четкого понимания специфики процесса имитации следует еще раз остановиться на этом вопросе. Цель вычислительного эксперимента, как правило, однозначна: установить соответствие между физическим экспериментом и его машинным аналогом, будь то непосредственный просчет вариантов или решение обратной задачи математической физики. Эта цель ставится и при имитационном моделировании. В гл. 4 мы отнесли ее к этапу предмодельных исследований, т.е. к этапу изучения отдельных физических процессов с целью получения количественной основы для описания блока комплексной модели (например, блока динамики почвенной влаги, блока фотосинтеза и других в модели агроэкосистемы).

В то же время комплексная имитационная модель обладает значительно большей сложностью. Наличие сотен переменных состояния и десятков неопределенных параметров существенно усложняет исследование. Благодаря большой размерности в некоторых случаях сама модель не может быть исследована до конца и даже после многих ее "прогонов" остается в некотором смысле "черным ящиком". Если идет речь о базовой модели, то машинный эксперимент не преследует цели достижения ее полной адекватности реальному объекту. Он призван лишь очертить границы, в которых данное описание приемлемо. Напротив, прикладная модель разрабатывается для вполне конкретных целей, а ценность такой модели определяется тем кругом практических задач, которые можно решить с ее помощью. Таким образом, можно утверждать, что имитационный эксперимент является многоплановым, а его цели видоизменяются на различных этапах построения, исследования и практического использования модели.

Специфика имитационного эксперимента приближает его к натурному, и эти два понятия оказываются значительно ближе друг к другу, чем к понятию "вычислительный эксперимент".

Перечислим те требования, которым должно удовлетворить программное обеспечение задач планирования эксперимента.

1. Учет многоцелевого характера имитационных исследований, обеспечение достижения следующих основных целей, которые преследуются при работе с моделью:

- предварительное исследование модели, достижение соответствия динамики поведения объекта и модели на качественном уровне (характер протекания и взаимодействия основных процессов, характер реакции модели на внешние воздействия);

- исследование чувствительности модели путем проведения первоначально отсеивающих дисперсионных, а затем регрессионных экспериментов;

- решение задачи идентификации (обратной задачи); для прикладных моделей под идентификацией понимается получение количественного соответствия результатов натурального и машинного экспериментов по тем пара-

метрам, которые будут в дальнейшем использоваться для принятия решений;

– собственно решение прикладных задач прогноза и управления (оценка областей достижимости программных траекторий, постановка экстремальных экспериментов).

2. Стохастический характер имитационного эксперимента, что требует наличия программ генерации случайных величин и последовательностей, а также программ статистической обработки результатов эксперимента, согласованных с видом генерируемых планов.

3. Учет того обстоятельства, что модель работает в тесном взаимодействии с базой данных. Поэтому общая организация программных средств должна быть ориентирована на то, что изменяемые в процессе проведения эксперимента входные параметры модели должны быть согласованы с той информацией, которая хранится в базе данных. С другой стороны, результат эксперимента, также должен фиксироваться в базе данных и быть доступным для обработки совместно с результатами последующих экспериментов.

Необходимость получения ансамбля траекторий модели в каждом имитационном эксперименте, а также связанная с этим проблема получения устойчивых статистических оценок параметров сигналов на выходе модели естественным образом приводят к задаче автоматизации всего комплекса вычислений, выполняемых ЭВМ в процессе эксперимента. В этой задаче можно, по-видимому, выделить три основных этапа.

1. Автоматизация процесса вычисления одной траектории модели. В существующих системах она решается с использованием тех или иных языков моделирования.

2. Автоматизация процесса выработки плана эксперимента (см. гл. 4), т.е. выбора количества и расположения точек в факторном пространстве, числа градаций факторов, числа повторностей и т.д. в связи с той конкретной задачей исследования модели, которая ставится в данном машинном эксперименте.

3. Автоматизация процесса накопления полученных в эксперименте данных, их сохранения на длительный срок, поиска и статистической обработки с последующим отображением полученной информации.

Желательно, чтобы программное обеспечение имитационной системы, предназначенное для решения задач второго и третьего этапов, допускало работу в режиме диалога.

Хотя в настоящее время существуют пакеты генерации имитационных планов и обработки данных, они не обеспечивают выполнения всех перечисленных выше требований. К их основным недостаткам можно отнести:

– отсутствие программ ввода с внешних носителей или терминала и программ вывода на эти устройства, что приводит к необходимости в каждом конкретном случае разрабатывать такие программы;

– отсутствие общей базы данных у программ большинства готовых пакетов, что делает необходимым создание специальных программ, передающих информацию из одной программы пакета в другую, отсутствие средств связи с той или иной базой данных;

– ряд математических алгоритмов пакетов не рассчитан на непрерывную обработку данных и допускает остановки в работе без достаточно идентификации причины;

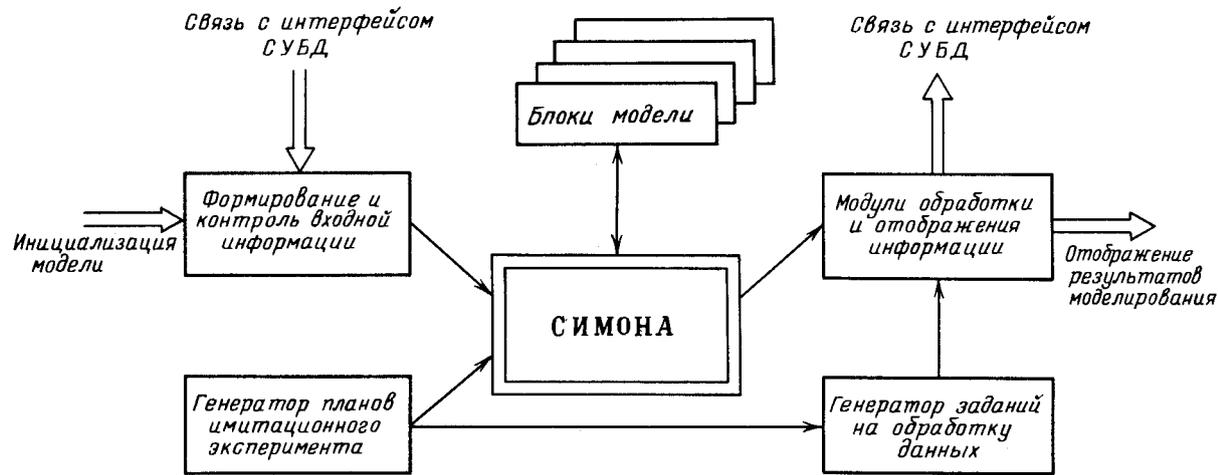


Рис. 5.1. Структура имитационной системы

– в процессе получения спектра плана эксперимента и обработки данных, особенно при последовательном планировании, приходится оперировать с достаточно большими массивами, которые одновременно не помещаются в оперативную память, а программы пакетов не рассчитаны на обработку массивов "по частям".

Самый же главный их недостаток, пожалуй, состоит в том, что при разработке подобных пакетов лишь в малой степени учитывалась та общая методология имитационных исследований, о которой говорилось выше.

В связи с изложенным была предпринята попытка разработать специальный комплекс программ генерации имитационных планов и обработки результатов машинного эксперимента. На рис. 5.1 приведена общая блок-схема, показывающая взаимодействие системы планирования эксперимента с описанной в п. 5.1.3 имитационной системой СИМОНА [20, 26, 27]. Комплекс планирования эксперимента состоит из модуля генерации имитационных планов и модуля статистической обработки информации. В первую версию пакета вошли программы, осуществляющие построение шести типов двухуровневых планов.

1. Планы полного факторного эксперимента.
2. Ортогональные планы разрешения III, предназначенные для оценки линейных эффектов.
3. Ортогональные планы разрешения IV (оценка линейных эффектов, не смешанных с эффектами парных взаимодействий).
4. Ортогональные планы разрешения V.
5. Ортогональные планы, занимающие промежуточное положение между IV и V.
6. Неортогональные насыщенные дробные реплики.

Планы 3 – 5 носят эволюционный характер, так как строятся последовательно на основе использования планов группы 2 (см. выше п. 4.4.2).

В целом комплекс состоит из основных и вспомогательных программ. К основным относятся программы генерации планов 1–6 и обработки их результатов. Вспомогательными являются программы связи с системой СИМОНА, перекодировки информации, а также программы построения матриц Адамара, являющихся основой для построения плана эксперимента разрешения III. Пакет реализован на ЕС ЭВМ в рамках операционной системы ОС ЕС.

§ 5.2. Внешнее программное обеспечение имитационной системы

В предыдущем параграфе описана структура программных средств, которые мы назвали "внутренним" программным обеспечением имитационных систем. Оно предназначается для автоматизации процесса конструирования модели и доведения ее до такого уровня, на котором результаты моделирования становятся доступными пользователю-непрограммисту. Именно для обеспечения возможности работы с моделью такого внешнего (конечного) пользователя и разрабатывается другая часть программных средств, а именно внешнее программное обеспечение. В чем заключается его особенности? С точки зрения сервиса это использование для общения с моделью лишь общепотребительных понятий разговорного языка и специальной терминологии, которая принята в той предметной

области, на которую ориентирована модель, а также наличие простых и понятных инструкций оператору. С содержательной стороны — это обеспечение наиболее полного использования для решения прикладных задач тех возможностей, которые предоставляются имитационной системой. Среди этих последних отметим необходимость обеспечения доступа к входной информации, экспериментирования с моделью, получения простых аппроксимационных зависимостей (типа уравнений регрессии), обоснования и выбора решений в процессе управления системой.

Часть этих задач, разумеется, решается с помощью тех средств, которые описаны выше. К ним, например, относятся программы, обеспечивающие планирование дисперсионных и регрессионных экспериментов. Аналогично этому, современные СУБД предоставляют для общения с базой данных программные средства различного уровня. СУБД "Спектр", в частности, содержит в своем составе базовый язык и макрогенератор (СПМАКРО), предназначенные для взаимодействия программистов с базами данных, а также простой интерактивный язык запросов СПИНТЕР, структура которого ориентирована специально на конечного пользователя-непрограммиста. Поэтому разделение математического и программного обеспечения на "внутреннее" и "внешнее" условно, как и любая классификация. Сделано это для того, чтобы при разработке программного комплекса с возможной полнотой учесть те специальные требования, которые будут способствовать широкому и эффективному использованию имитационной системы. Поэтому ниже мы рассмотрим лишь эту дополнительную специфику задачи моделирования и наметим контуры ее решения.

5.2.1. Управление базой данных. Существующая в настоящее время обширная литература по конструированию и применению баз данных (см., например, [12]) делает излишними подробное изложение этого вопроса. В связи с этим здесь мы лишь коротко рассмотрим два его аспекта: специфику использования баз данных в составе имитационных систем вообще и при моделировании экологических систем в частности. Сразу же отметим, что существующие СУБД плохо приспособлены для решения задач имитационного моделирования. Это и понятно, поскольку они создавались для управления некоторой стационарной структурой иерархического, сетевого или реляционного типа. Прикладные программы пользователя и язык управления рассматривались при этом как вспомогательное средство преобразования этой структуры. В то же время схема информационных потоков в имитационной модели принципиально иная. Вычисление траектории модели, а тем более проведение серий спланированных машинных экспериментов порождают динамическую организацию данных, которая системами управления БД не поддерживается.

Имеется и еще одна причина. Программный комплекс имитационной системы сам по себе является весьма большим по числу команд и имеет сложную и определенным образом структурированную собственную организацию, не совпадающую с организацией СУБД. Поэтому было бы наивным предполагать, что можно "замкнуть" модель на БД с помощью, например, такого простого средства, как обращение к ее содержимому.

Трудности, встречающиеся на этом пути, могли бы быть преодолены путем создания специальной СУБД, ориентированной на работу в составе имитационной системы и поддерживающей динамические структуры дан-

ных, появляющихся в задачах моделирования. Однако в настоящее время такие разработки отсутствуют.

В связи с этим приходится прибегать к паллиативным решениям. Основной путь организации взаимодействия ИС и БД в настоящее время заключается в разработке специальных интерфейсных модулей, обеспечивающих обмен данными между БД и моделью. Такие интерфейсные модули должны быть двух типов. Первые из них на основе хранящейся в базе данных информации формируют файл параметров модели, файл начального состояния и файлы параметров входных воздействий (включая управляющие параметры генератора имитационных планов). Интерфейс второго типа осуществляет фиксацию результатов моделирования совместно с необходимыми дискрипторами, идентифицирующими условия проведения машинного эксперимента, и осуществляет запись этой информации в соответствующий раздел базы данных. Обстоятельством, несколько облегчающим эту деятельность, является то, что работа с входной и выходной информацией и собственно процесс имитации разделены во времени. Поэтому осуществляющие эти различные виды работ программы могут быть оформлены в виде отдельных загрузочных модулей. Так, например, в рамках СУБД "Спектр" подготовка исходных для моделирования данных может осуществляться с использованием программных средств, предоставляемых макрогенератором системы СПМАКРО, хотя здесь и имеются некоторые неудобства, связанные с тем, что функционирование макрогенератора обеспечивается лишь во взаимодействии пользователя с администратором базы данных.

Перейдем к рассмотрению особенностей информационного обеспечения задач моделирования экологических систем. Наиболее полная структура такого обеспечения непосредственно связана с тем комплексом проблем, которые решаются в рамках так называемого "экологического мониторинга" [10, 11], ориентированного на выявление, прогнозирование и оценку антропогенных изменений состояния отдельных природных экосистем и биосферы в целом. Разработка экологического мониторинга связана с решением важных прикладных задач — выявление допустимой антропогенной нагрузки на существующие в природе наземные и водные экологические системы, прогнозирование хода протекающих в них процессов в связи с все увеличивающимися темпами выброса загрязняющих веществ в атмосферу и вымывания химических мелиорантов из почвы. В наиболее полной постановке возникающие здесь проблемы следует рассматривать как задачи управления динамической экосистем, т.е. выработки таких внешних воздействий, которые вызывают экзогенную сукцессию и сдвиг равновесия в желательном для человека направлении. Аналогичные задачи в земледелии и растениеводстве решаются в рамках агромониторинга [9] — информационной системы, обеспечивающей решение проблемы повышения продуктивности агроэкосистем при одновременном удовлетворении требований охраны окружающей среды.

Поскольку цель разработки систем экологического мониторинга и агромониторинга сводится к решению прикладных задач прогноза и управления, целесообразно (и это, вероятно, единственно правильный путь) связать их с соответствующими имитационными моделями. А учитывая многообразие этих задач, было бы желательно дать их классификацию.

Традиционно принято выделять три уровня, на которых формулируются цели и разрабатываются методы регулирования состояния окружающей среды. Это локальный (отдельные биогеоценозы или акваценозы), региональный и глобальный уровни [10, 11]. Каждому из них отвечает специфический набор моделей, требующий вполне определенного информационного обеспечения, т.е. требуется разработка специальной информационной базы.

В гл. 4 понятие банка моделей мы связали с определенной базой данных. Таким образом, при решении в рамках экологического мониторинга задач прогноза поведения экосистем, управления их динамикой, охраны окружающей среды можно говорить о банках локальных, региональных и глобальных моделей.

Приведенная выше классификация задач и ассоциированных с ними банков моделей по "территориальному масштабу" не является единственно возможной. Другую классификацию можно связать с "временным масштабом", а точнее, с длительностью того периода времени, на протяжении которого сказываются последствия принимаемых решений. Учитывая естественную сезонную периодичность протекающих в природе процессов, в [8] выделены три типа хозяйственных решений — многолетние, сезонные и оперативные. Соответственно этому можно говорить о трех типах имитационных моделей.

Эта классификация наиболее четко прослеживается в задачах управления продуктивностью агроэкосистем. Решения первого временного уровня (многолетние) касаются мелиорации и окультуривания земель и реализуются в виде проектов мелиоративного землеустройства. Последствия этих решений, которые можно назвать перспективными или проектными, сказываются на протяжении 10 – 15 и более лет. Соответствующие модели многолетнего типа, которые используются при проведении вариантных предпроектных расчетов, названы моделями плодородия сельскохозяйственных полей. Вторая группа решений — те, которые принимаются на текущий вегетационный период. Это, как очевидно, традиционно выполняемые плановые расчеты. Основой плановых решений должна явиться модель сезонной динамики агроэкосистем — так называемая модель продуктивности. Наконец, оперативные решения, необходимость которых связана с изменчивостью погодных условий, принимаются уже в ходе вегетации посевов. Количественная основа таких решений — динамическая модель формирования урожая. Ясно, что каждый из перечисленных типов решений требует, вообще говоря, своего информационного обеспечения. В то же время из следующего пункта будет ясно, что плановые и оперативные задачи необходимо решать совместно. В связи с этим можно говорить о двух банках имитационных моделей, каждому из которых соответствует свое информационное наполнение:

- 1) банк моделей плодородия почв;
- 2) банк моделей продуктивности агроэкосистем и формирования урожая.

Подведем некоторые итоги. Решение разнообразных задач, связанных с прогнозом поведения экосистем и управления их динамикой в рамках тех или иных имитационных систем, требует создания баз данных, специализированных по типу моделей и характеру информационного наполнения. Принятая нами классификация, учитывающая пространственно-временные

масштабы, на которые распространяются принимаемые решения, приводит к появлению девяти возможных комбинаций. Однако, как показывает более подробный анализ, не все они имеют практическое значение. Так, вряд ли имеет смысл говорить об "оперативных" решениях, касающихся биосферы в целом. Сказанное иллюстрирует табл. 5.1, в которой дана краткая формулировка задач того или иного уровня, которые приводят к необходимости создания специальных банков моделей при соответствующем математическом, программном и информационном обеспечении.

5.2.2. Управление экосистемами. В предыдущем пункте, обсуждая особенности базы данных, мы уже имели в виду прикладную направленность имитационных систем, создаваемых для решения задач управления. Ориентация на эти задачи в определенной мере связана с тематической направленностью этой книги. Отметим, что специфика задач управления экосистемами уже обсуждалась в гл. 3, в которой установлен ряд фундаментальных результатов, касающихся их общей математической структуры. Особенности же решения задач управления в рамках имитационных систем до сих пор не рассматривались. Настоящий пункт как раз и будет посвящен изложению этого вопроса.

Как всегда, мы начнем с трудностей, встречающихся на этом пути. И прежде всего отметим слабую изученность многих процессов в экосистемах. Ранее уже упоминалось, что наиболее изученными являются процессы энерго- и массообмена: поглощение и использование фотосинтетически активной радиации растениями, динамика водно-теплового режима, газообмена, миграция подвижных соединений в почвах и т.д. На другом полюсе с точки зрения изученности находятся действующие на различных уровнях экосистемы процессы саморегуляции, обеспечивающие определенную стабильность биоценоза, адаптивность его реакций в ответ на внешние воздействия. В то же время ясно, что именно эти последние процессы необходимо учитывать в первую очередь при разработке методов принятия решений, т.е. при решении задач управления экосистемами. Промежуточное положение занимают трофические взаимодействия в экосистемах. Хотя в этой области исследования создан большой теоретический задел и, более того, многие механизмы трофического взаимодействия популяций различных видов хорошо изучены как на модельном уровне, так и непосредственно в экспериментах, тем не менее получить надежные количественные оценки параметров взаимодействия часто оказывается невозможным. И во всяком случае, для каждой конкретной ситуации приходится проводить специальную исследовательскую работу по анализу как литературных источников, так и фактических данных, если они имеются.

Итак, трудности, встречающиеся при разработке методов управления экосистемами на основе использования имитационных моделей, связаны со следующими обстоятельствами:

- слабая изученность ряда процессов, что приводит к начальной неопределенности при разработке модели и алгоритмов управления;
- многомерность объекта, наличие большого числа переменных состояний и неопределенных параметров, подлежащих оценке в процессе идентификации;
- адаптивные свойства объекта управления и, в частности, изменение его реакций на внешние воздействия при их многократном повторении.

Таблица 5.1

Масштаб времени	Уровни принятия решений		
	локальный	региональный	глобальный
$T < 1$ года	Принятие оперативных решений с использованием моделей отдельных экосистем	—	—
$T = 1$ году	Принятие плановых решений по эффективному использованию ресурсосберегающих и природоохраных технологий	Разработка планов природопользования	—
$T > 1$ года	Принятие проектных решений о комплексах мелиоративных, почвозащитных и природоохраных мероприятий	Проекты регионального природопользования, размещение новых предприятий на территории с учетом требований охраны среды	Прогноз антропогенного воздействия на биосферу

Рассмотрим пути преодоления этих трудностей. О начальной неопределенности мы уже говорили в гл. 4. Именно с этим связана необходимость создания базовой модели блочного типа и обеспечение условий ее изменения и пополнения по мере накопления новых данных без коренной "переделки" имитационной системы.

Другая сторона этого же вопроса заключается в следующем. Если необходимо тем или иным способом получить "замкнутое" описание, а сведения о механизмах отсутствуют, то каким образом это можно сделать? Очевидно, единственный путь — это включение в общую модель отдельных соотношений, полученных путем статистической обработки некоторого фактического материала. Включение в модель ряда регрессионных блоков, разумеется, ограничивает область ее применимости и делает ее менее универсальной. Но тем не менее оно всегда оправдано, если управление, которое в дальнейшем предполагается использовать, не затрагивает (или затрагивает в малой степени) те процессы, которые описаны на уровне регрессионных соотношений. Точнее говоря, если с помощью уравнения регрессии описан некоторый блок модели или взаимодействие блоков, то в это уравнение в качестве аргументов (так называемых "предикторов") должны быть включены все переменные, которые подвержены изменению под влиянием управляющих воздействий. Это, разумеется, предъявляет дополнительные требования к исходному экспериментальному материалу.

Обратимся к проблеме размерности модели, которая также частично обсуждалась ранее. Именно большая размерность модели и связанные с ней значительные затраты машинного времени не позволяют в настоящее время использовать классические методы решения задач управления и, в частности, методы оптимального управления. Выход из этого положения следует искать на пути использования, во-первых, методов "экспериментальной оптимизации" и, во-вторых, декомпозиции задачи, т.е. создания упрощенных локальных моделей значительно меньшей размерности. Оба эти пути решения задачи уже обсуждались. Здесь мы только отметим, что первый из них связан с применением идей и методов планирования экстремальных машинных экспериментов, а второй — с созданием "внешнего" математического и программного обеспечения имитационных систем.

Рассмотрим, наконец, адаптивные свойства экологической системы как объекта управления. Эти свойства экосистемы проявляются в том, что эффективность управляющего воздействия, приложенного в некоторый момент времени, зависит от всей предыстории системы, т.е., в частности, от того, какие управления действовали в прошлом. В гл. 3 изучалась управляемость компартментальных систем и были выведены соотношения, позволяющие оценить области достижимости тех или иных состояний системы. Адаптивность экосистемы с точки зрения достижимости означает, что эти области не остаются постоянными, а становятся зависимыми от характера прошлых воздействий. Изменчивость реакций в ответ на многократное повторение одних и тех же воздействий вообще характерна для живых систем вне зависимости от уровня их организации. Она проявляется и на уровне организма животного или растения и на уровнях популяции или биоценоза. Учитывая, что целью управления всегда является перевод системы в некоторую заданную область пространства состояний, можно в данном случае разделить управляющие воздействия на две группы — те, которые трансформируют в желательном направлении область достижи-

мости системы (например, расширяют ее), и те, которые изменяют состояние системы, перемещая его в фазовом пространстве в заданном направлении. Пример, иллюстрирующий эту особенность задач управления экосистемами, будет приведен ниже.

Таким образом, мы приходим к выводу, что программное обеспечение задач управления должно включать в себя:

- средства, обеспечивающие планирование экстремальных экспериментов и обработку получаемых в процессе таких экспериментов данных;
- средства взаимодействия исходной имитационной модели с набором простых локальных моделей, составляющих внешнее математическое обеспечение имитационной системы;
- собственно программы, реализующие те или иные алгоритмы принятия решений;
- средства анализа последствий прошлых решений, принятых и реализованных до текущего момента времени.

Рассмотрим принципы реализации перечисленных выше особенностей проблемы управления экосистемами на примере задачи управления продукционным процессом в агроэкосистемах. Эта задача сводится к необходимости выработки решений о нормах и сроках поливов, дозах и сроках внесения минеральных подкормок растений, проведении тепломелиоративных мероприятий (например, мелкодисперсное распыление влаги над посевом или пескование тяжелых почв) и т.д. Задача решается на двух уровнях. На стадии планирования необходимо распределить по отдельным полям имеющиеся и, как правило, ограниченные ресурсы. Непосредственно в ходе вегетации растений возникает задача оперативного управления. Принятие решений в обоих случаях производится в условиях неопределенности, связанных с вероятностным характером погодных условий.

Задачи планирования решаются в статической постановке. Они хорошо изучены, и многочисленные методы их решения описаны в литературе. В наиболее полной постановке они сводятся к нелинейным стохастическим моделям, а точнее говоря, к моделям стохастического программирования. Имитационная же модель, описывающая динамику продукционного процесса, создает основу для выработки решений на стадии оперативного управления.

Отметим, что решения задач обоих уровней взаимосвязаны. Действительно, при выработке планового (на сезон вегетации) решения необходимо учитывать возможность его практической реализации. Это означает, в частности, что должно существовать по крайней мере одно управление, реализующее плановую траекторию. При решении задач планирования это достигается за счет того, что используемые здесь так называемые "производственные функции" получаются путем статистической обработки данных машинного эксперимента, т.е. они входят в состав внешнего математического обеспечения имитационной системы, которое согласовано с ее внутренним математическим обеспечением¹⁾). Разумеется, при этом

¹⁾ Отметим, что производственные функции, которые следует рассматривать как некоторые аппроксимирующие соотношения, носят интерполяционный характер. И, следовательно, любому выбранному решению, расположенному в области определения этих функций, соответствует по крайней мере одна траектория модели.

нет взаимно однозначного соответствия, поскольку, например, один и тот же уровень урожая может быть получен при некоторой влагообеспеченности вне зависимости от того, достигается ли это за счет естественно выпадающих осадков или поливов. Тем не менее можно выделить некоторую типичную для данного региона погодную реализацию и соответствующую ей технологию, которую как раз и можно считать плановой. Эта задача существенно облегчается, если хранить в базе данных не только внешнюю по отношению к модели информацию, но и результаты машинных экспериментов, на основе которых осуществлено построение производственных функций, положенных в основу решения задачи планирования. Траекторию (или набор траекторий), отвечающую принятому плану, будем называть "эталонной" или "программной": траекторией ("эталонным" или "программным" набором).

Перейдем к рассмотрению задачи оперативного управления. Ее характерной особенностью является то, что решения о проведении тех или иных технологических мероприятий принимаются в дискретные моменты времени, связанные с определенными фазами развития растений (с так называемыми фенофазами). В ходе вегетации реализация погодных условий всегда отклоняется от той, которая принята при выборе плана. Таким образом, если в некоторый момент времени будет зафиксировано определенное состояние растений, то учет реально складывающейся метеобстановки приведет к тому, что фактическая и программная траектория разойдутся. Поскольку расчет программной траектории на стадии планирования выполнен, исходя из условия согласования с заданиями высшего уровня, в процессе оперативного управления желательно действовать так, чтобы фактическая траектория не слишком уклонялась от программной. Совершенно очевидно, что если в точности реализуется плановая агротехника и та метеобстановка, которая была заложена в расчет плана, то будет получена именно программная траектория и план будет выполнен. Однако вероятность этого события равна нулю. Поэтому в качестве критерия в задаче управления следует принять максимум вероятности выполнения плана. При этом наличие набора программных траекторий позволяет выбрать ту из них, которая ближе в смысле некоторой нормы к фактической, т.е., как очевидно, соответствует наиболее "похожему" по метеоусловиям году из тех лет, информация о которых хранится в базе данных.

Выскажем в заключение одно замечание. Пусть в результате решения задачи оперативного управления на некотором временном шаге обнаружилось, что вероятность достижения программной траектории мала. Очевидно, что в этом случае следует пересмотреть плановое решение, т.е. вернуться к повторному рассмотрению задачи планирования на верхнем уровне иерархии в целом. Поскольку это может привести к пересмотру многих показателей смежных отраслей (например, на уровне хозяйства при пересмотре плана по растениеводству может понадобиться изменение плана по животноводству), этот возврат к пересчету плана следует производить лишь в крайних случаях. Поэтому оперативные решения должны удовлетворять еще одному критерию: они по возможности не должны приводить к необратимым изменениям управляемой системы, т.е. должны оставлять достаточно широкой "область достижимости" программной траектории для последующих воздействий на систему.

5.2.3. Интерфейс пользователя. В рамках имитационной системы решаются разнообразные задачи — проводятся предмодельные исследования, осуществляется построение и исследование отдельных блоков модели, ее комплексная отладка, осуществляется оценка чувствительности и идентификация параметров модели и, наконец, в той или иной предметной области решается комплекс прикладных задач. Многообразие функций системы предполагает в свою очередь наличие многих потенциальных пользователей, отличающихся друг от друга как по своим интересам, так и по уровню профессиональной подготовки. Выше в п. 5.1.1 были выделены три типа пользователей: разработчиков модели, прикладных программистов, способных совершенствовать модель в рамках заданной структуры, и конечных пользователей-непрограммистов. Очевидно, что эта классификация еще не является полной и возможно введение дополнительных градаций в каждой группе. Так, например, в группе конечных пользователей существуют специалисты, использующие имитационную модель и машинные эксперименты в своих научных исследованиях, и специалисты, решающие прикладные задачи чисто производственного плана (например, задачи управления производством). Очевидно, что для первых наибольший интерес представляет внутреннее математическое обеспечение и средства автоматизации машинного эксперимента. Пользователи второго типа, как правило, должны обращаться к внешнему математическому обеспечению имитационной системы.

Если еще учесть, что работа с имитационной системой может выполняться как в интерактивном, так и в пакетном режимах, то станет ясным, насколько сложными оказываются требования к программам, осуществляющим взаимодействие человека с машиной. В связи с этим разработка интерфейсных модулей превращается в самостоятельную и достаточно сложную задачу.

Табл. 5.2 дает представления о видах работ, проводимых в процессе создания и эксплуатации имитационной системы, и о распределении этих работ по типам пользователей. Конечные пользователи имеют дело с "готовой" системой. Их роль сводится к постановке и проведению машинных экспериментов по оценке чувствительности модели, ее идентификации, решению разнообразных прикладных задач.

В противоположность этому разработчики системы начинают работу "с нулевого цикла" и должны довести ее до этапа комплексной отладки модели. Промежуточное место в этом комплексе работ занимают прикладные программисты. Их основной целью является разработка версий блоков модели и связь банка моделей с внешним математическим и программным обеспечением.

Из всего, что было сказано, следует, что разработка имитационной системы должна сопровождаться созданием по крайней мере двух видов языков, один из которых должен быть ориентирован на пользователей-программистов, а другой — на конечных пользователей-непрограммистов.

Принципиальное отличие функций этих двух языков заключается в том, что пользователь-программист, даже если он не является разработчиком модели, должен иметь возможность трансформировать блоки модели, включить в нее новые версии блоков и т.д. Пользователь же непрограммист имеет дело с оттранслированной и собранной в загрузочный модуль сис-

Таблица 5.2

Вид работы	Типы пользователей		
	Разработчики системы	Пользователи-программисты	Конечные пользователи
Разработка структуры и создание банка моделей	+		
Разработка структуры классов моделей и создание классов моделей	+		
Разработка версий блоков модели в рамках созданных классов	+	+	
Комплексная отладка моделей	+	+	
Разработка внешнего математического обеспечения системы		+	
Проведение машинных экспериментов по оценке чувствительности моделей		+	+
Идентификация моделей		+	+
Решение прикладных научно-исследовательских задач в рамках имитационной системы			+
Решение прикладных производственных задач			+

темой и может изменять только условия проведения машинного эксперимента и ограниченные определенными рамками исходные данные.

Диалоговый режим должен обеспечить при решении каждой из задач моделирования следующие виды работ: ввод, коррекция и контроль исходной информации, подготовка к вычислениям, вычисление траектории модели, вывод результатов. Следует отметить, что вычисление траектории модели зачастую представляет собой весьма длительный процесс и пользователь должен иметь возможность контролировать промежуточные результаты вычисления, прерывая в случае необходимости решение задачи или корректируя входную информацию. Поскольку с точки зрения пользователя именно эти цели являются основными, при разработке диалоговых систем из общей иерархии целей [6] следует выделить ясность, легкость обучения пользователя и надежность (защиту от ошибок).

В свете изложенного можно полагать, что диалог в системе должен строиться по иерархическому принципу. На высшем уровне целесообразно использовать диалог типа "меню", в котором пользователю предлагается указать один из допустимых в данном состоянии режимов работы (например, коррекция параметров, коррекция начального состояния, коррекция параметров вычислительной схемы, счет). При работе на этом уровне пользователь должен иметь возможность вывода на экран терминала справочной информации, поясняющей содержание каждого из режимов. На втором уровне, как правило, осуществляется ввод, коррекция или контроль числовой информации. Диалог этого уровня строится по типу "запрос с указанием синтаксиса ответа". Третий уровень должен дать

возможность пользователю контролировать прохождение задачи и вмешиваться в процесс счета.

Все функции диалога необходимо организовать таким образом, чтобы

- 1) дать пользователю возможность контролировать свои действия;
- 2) защитить его программу от бессмысленной информации;
- 3) работа с системой не вызывала раздражение пользователя (в частности, система не должна выдавать на дисплей "лишнюю" информацию).

С целью облегчить создание прикладных диалоговых систем в рамках имитационной системы СИМОНА-СМ разработан пакет стандартных подпрограмм, обеспечивающих простую реализацию диалога описанного типа. В пакет входят четыре модуля: ELECT, IENT, RENT и КСНП.

Процедура ELECT реализует диалог типа "меню". Она выполняет следующие функции: выводит на экран терминала текстовую строку запроса и массив допустимых ответов (если в обращении задан файл справочной информации, то дополнительно на дисплей выводится ответ HELP), принимает с дисплея ответ пользователя (при ответе HELP выводит на экран файл справочной информации и повторяет запрос), определяет, какому из допустимых ответов соответствует ответ пользователя, и передает этот номер в виде целой константы в программу пользователя. Если в массиве ответов первый элемент непуст, то он является ответом, принимаемым по умолчанию. Выход из программы невозможен до тех пор, пока пользователь не введет один из допустимых ответов. После ответа, не прошедшего синтаксический контроль, выводится сообщение об ошибке и запрос повторяется. Процедуру рекомендуется использовать как узловую точку программы, к которой программа возвращается после прохождения каждого из режимов. После процедуры обычно ставится оператор вычисления условного перехода.

Процедуры IENT и RENT предназначены для ввода с терминала соответственно целого или вещественного числа. Принципы работы программ идентичны. Они выполняют следующие функции: выводят на дисплей строку запроса, значение числа (целое или вещественное), принимаемое по умолчанию; минимальное и максимальное допустимые значения, если они заданы; принимают с дисплея вводимое число в символьной форме, допустимой в языке ФОРТРАН; проверяют синтаксис введенной строки и, если это необходимо, нахождение введенного числа в допустимых границах; передают значение полученного числа программе пользователя; в случае ошибки выводят на дисплей ее диагностику и повторяют запрос. Выход из программ невозможен до тех пор, пока пользователь не введет один из допустимых ответов. Процедуры рекомендуется использовать на втором уровне диалога в режимах ввода или коррекции числовой информации. Наличие пустого ответа (значение по умолчанию) позволяет использовать их для контроля введенной ранее или имеющейся в пакете информации.

Процедура КСНП реализована как программа-функция целого типа; при отсутствии символов в буфере терминала она не вызывает остановки программы ввода. При наличии символов в буфере первых из них вводятся в младший байт слова, а остальные символы в буфере до первого появления символов (VK) игнорируются. Процедуру рекомендуется использовать в режимах счета, занимающего значительное время для организации на терминал текущей информации (текущее время модели, номер итера-

ции, текущее состояние модели и т.п.), для организации режимов экстренной печати и для организации прерываний.

Созданный пакет программ облегчает адаптацию человека к вычислительной технике при разработке и отладке имитационных моделей, а также при постановке машинных экспериментов в рамках имитационных систем.

В заключение отметим, что любая диалоговая система подобного типа должна обладать свойствами гибкости, ясности поведения, простоты использования и простоты обучения [6]. Под гибкостью диалоговой системы следует понимать такое ее поведение, которое учитывает потребности и уровень квалификации пользователя. Ясность поведения означает, что пользователь в процессе работы с системой способен построить ее непротиворечивую модель, с помощью которой поведение системы становится "прозрачным" для пользователя. Простота, кроме всего прочего, означает наличие хорошей структуры диалога, отсутствие лишней информации, возможность запросить дополнительные сведения в случае необходимости. Пользователю следует облегчить привыкание к системе, представляя ему средства помощи в любой возможной ситуации.

Приложение 1

ВСПОМОГАТЕЛЬНЫЕ ЛЕММЫ

Техническая лемма 1. Пусть Ω – метрическое пространство, E_1 и E_2 – банаховы пространства и $G: \Omega \times E_1 \rightarrow E_2$ – непрерывная функция. Предположим, что при всех $X \in \Omega$ уравнение $G(X, Y) = 0$ имеет единственное решение. Таким образом, определена функция $H: \Omega \rightarrow E_1$, для которой $G(X, H(X)) = 0$. Предположим, что замыкание образа некоторой замкнутой окрестности $\bar{U} \subset \Omega$ точки X компактно. Тогда функция $H(\cdot)$ непрерывна в точке X .

Доказательство. Пусть $X_n \in \bar{U}$, $X_n \rightarrow X_0 \in \bar{U}$ и $Y_n = H(X_n) \not\rightarrow Y_0 = H(X_0)$. Пользуясь компактностью, будем считать, что $Y_n \rightarrow Y \neq H(X_0)$. Тогда по непрерывности имеем $G(X_0, Y) = 0$, что противоречит единственности решения Y_0 . Лемма доказана.

Техническая лемма 2. Если квазимонотонный положительно однородный оператор $G = (g_1, \dots, g_m)$ таков, что каждая его i -я компонента $g_i(x_1, \dots, x_m)$ ($i = 1, \dots, m$) как функция x_i удовлетворяет условию Липшица в некоторой окрестности нуля, т.е. $|g_i(x_1, \dots, x_i + h, \dots, x_m) - g_i(x_1, \dots, x_m)| < P|h_i|$, то существует достаточно большое $N > 0$, при котором $F + NI$ – монотонный положительный оператор.

Доказательство. Пусть $X \geq Y$. Тогда

$$(G(X) - G(Y))_i = [g_i(X) - g_i(x_1, \dots, y_i, \dots, x_m)] + [g_i(x_1, \dots, y_i, \dots, x_m) - g_i(Y)].$$

Вторая скобка неотрицательна. Положим

$$N = \text{Sup} \{ |g_i(X) - g_i(x_1, \dots, y_i, \dots, x_m)| / |x_i - y_i| \mid \|X\| = 1, 0 \leq y_i \leq x_i \}.$$

В силу однородности и квазимонотонности оператора G имеем монотонность $G + NI$. Лемма доказана.

Техническая лемма 3. Пусть матрица A имеет неотрицательные недиагональные элементы и ее собственное значение с максимальной вещественной частью неположительно. Обозначим через

$$B(\lambda) = I\lambda^{n-1} + B_1\lambda^{n-2} + \dots + B_{n-1} = \det\lambda I - A^{-1}$$

присоединенную матрицу. Тогда $B_i \geq 0$ ($i = 1, \dots, n-1$) и в случае неразложимости A имеет место строгое неравенство $B_{n-1} > 0$.

Доказательство. Введем в рассмотрение матрицу

$$H = \left[\begin{array}{c|c} A & Y \\ \hline -X^* & 0 \end{array} \right], \text{ где векторы } X \geq 0 \text{ и } Y \geq 0.$$

Тогда $\det[\lambda I - H] = \sum_{k=0}^m s_k^H \lambda^{m-k+1}$. Здесь и далее s_k^H (s_k^A) – сумма главных миноров порядка k матрицы $-H$ ($-A$). Отсюда $s_k^H = s_k^A + t_k$, где t_k – сумма главных миноров, опирающихся на последние строку и столбец. Известно, что $s_k^A \geq 0$ при

$\lambda(A) < 0$. Рассмотрим минор

$$\Delta(i_1, \dots, i_k, m+1) = \left[\begin{array}{c|c} -\bar{A} & \bar{Y} \\ \hline -\bar{X}^* & 0 \end{array} \right],$$

$$\bar{X} = (x_{i_1}, \dots, x_{i_{k-1}}), \quad \bar{Y} = (y_{i_1}, \dots, y_{i_{k-1}}),$$

$$\text{где } \bar{A} = \left[\begin{array}{ccc} a_{i_1 i_1} & \dots & a_{i_1 i_{k-1}} \\ a_{i_{k-1} i_1} & \dots & a_{i_{k-1} i_{k-1}} \end{array} \right].$$

Пользуясь формулой окаймленного определителя, получаем $\bar{\Delta}(i_1, \dots, i_k, m+1) = \bar{X}^* B_{\bar{A}}^{-1}(0) \bar{Y}$, где $B_{\bar{A}}^{-1}(0)$ – присоединенная матрица для матрицы \bar{A} при $\lambda = 0$. Известно, что $\lambda(\bar{A}) < \lambda(A)$, и если A неразложима, то неравенство строгое. Следовательно, $\lambda(\bar{A}) < 0$. Поскольку \bar{A} имеет неотрицательные недиагональные элементы, $B_{\bar{A}}^{-1}(0) \geq 0$, и поэтому $\bar{X}^* B_{\bar{A}}^{-1}(0) \bar{Y} \geq 0$. По формуле окаймленного определителя имеем

$$\begin{aligned} \sum_{k=0}^m (s_k^A + t_k) \lambda^{m-k+1} &= \lambda \det(\lambda I - A) + X^* B(\lambda) Y = \\ &= \lambda \sum_{k=0}^m \lambda^{m-k} s_k^A + X^* B(\lambda) Y. \end{aligned}$$

Отсюда при произвольных $X \geq 0$ и $Y \geq 0$ получаем

$$t_1 = 0, \quad X^* Y = t_2 = \sum_{i=1}^m \left[\begin{array}{c|c} -c_{ii} & y_i \\ \hline -x_i & 0 \end{array} \right] \geq 0,$$

$$X^* B_k Y = t_k = \sum_{1 \leq i_1 < \dots < i_k \leq n} \bar{\Delta}(i_1, \dots, i_k, n+1) \geq 0.$$

Следовательно, $B_k \geq 0$. Если матрица A неразложима, то $B(\lambda) > 0$ при $\lambda(A) < \lambda$. Поэтому $B_{n-1} = B(0) > 0$ при $\lambda(A) < 0$. Лемма доказана.

Следствие. Если матрица A неразложима, то характеристический полином матрицы $A - YX^*$ имеет положительные коэффициенты ($X \geq 0, Y \geq 0$).

Доказательство. Легко видеть, что искомым характеристический полином равняется определителю матрицы $\left[\begin{array}{c|c} \lambda I - A & Y \\ \hline -X^* & 1 \end{array} \right]$. Воспользовавшись формулой окаймленного определителя, положительностью собственных главных миноров матрицы A при $\lambda(A) < 0$ и технической леммой 3, получаем доказательство следствия.

Приложение 2

ГЕНЕРАТОРЫ ПСЕВДОСЛУЧАЙНЫХ ЧИСЕЛ И ПОСЛЕДОВАТЕЛЬНОСТЕЙ

В гл. 2 рассмотрены условия, при которых детерминированные системы с дискретным временем обладают так называемым "хаотическим" поведением, визуально напоминающим случайный процесс. Как установлено во многих работах последних лет, "хаос" может возникать уже в достаточно простых случаях, например при отображении отрезка $[0, 1]$ в себя:

$$x_{n+1} = f(x_n), \quad x_0, x_n \in [0, 1]. \quad (\text{П}2.1)$$

Хотя машинные алгоритмы генерации псевдослучайных последовательностей появились независимо от доказательства этого факта, все они, как теперь ясно, порождают хаотические траектории.

При моделировании экосистем основным фактором случайности является метеорологическая обстановка: погодные условия в модели следует рассматривать как реализации некоторого случайного процесса.

Как показывает исследование вероятностных характеристик основных метеорологических параметров, с достаточной для практических целей точностью погоду можно рассматривать как случайный процесс вида

$$w(t) = P(t)a + \varphi(t), \quad (\text{П}2.2)$$

где $P(t)$ – матрица заданных периодических (с периодом 24 часа) функций времени, a – случайный вектор, $\varphi(t)$ – стационарный случайный процесс с нулевым математическим ожиданием и заданной корреляционной матрицей.

Учитывая, что имитационная модель реализуется с некоторым временным шагом, вместо (П2.2) следует записать

$$w(t_k) = P(t_k)a + \varphi(t_k) \quad (k = 1, 2, \dots) \quad (\text{П}2.3)$$

и рассматривать входное воздействие модели в виде некоторой случайной последовательности.

Таким образом, для решения вопроса воспроизведения в модели воздействий вида (П2.3) нужно рассмотреть:

- 1) методы генерации независимых случайных чисел и случайных векторов с заданной ковариационной матрицей;
- 2) методы генерации случайных последовательностей с заданной корреляционной функцией.

Генераторы псевдослучайных чисел. В настоящее время существует множество методов машинной имитации случайных величин с различными законами распределения. При этом величины с произвольным распределением обычно формируются с помощью тех или иных преобразований одного или нескольких независимых значений случайной величины ξ , равномерно распределенной на отрезке $[0, 1]$. Поэтому задача моделирования распадается на два этапа.

1. Получение реализации случайных чисел, равномерно распределенных на отрезке $[0, 1]$ с помощью физического устройства или специальной программы "датчика случайных чисел".

2. Получение с помощью этих чисел реализаций случайной величины с более сложным законом распределения.

Узловой проблемой является, таким образом, моделирование равномерно распределенных случайных чисел. Для этого, как правило, используются рекуррентные формулы, с помощью которых производятся последовательные вычисления псевдослучайных чисел. Задание одной или нескольких входных величин полностью определяет такую последовательность. В то же время многие ее свойства аналогичны свойствам случайных чисел. Наиболее распространенными способами рекуррентного типа являются соотношения, использующие числа Фибоначчи и так называемый "множительно-конгруэнтный метод" (метод вычетов). Реализация последнего производится, например, на основе соотношений

$$u_0 = 1, \quad u_n = u_{n-1} M \{ \text{mod } 2^m \}, \quad \xi_n = u_n 2^{-m}, \quad (П2.4)$$

где M – достаточно большое число, m – число двоичных разрядов в мантиссе ячейки ЭВМ. Для определенности выбирают $M = 5^{2p+1}$, где p – целое, причем, как очевидно, должно быть выполнено неравенство $5^{2p+1} \leq 2^m$. Запишем (П2.4) в виде

$$\xi_n = \{ M \xi_{n-1} \} = \{ 5^{(2p+1)n} \xi_0 \}, \quad \xi_0 = 2^{-m}. \quad (П2.5)$$

Это позволяет описать получение последовательности ξ_1, ξ_2, \dots следующим образом. Запишем 2^{-m} в системе исчисления с основанием 5:

$$2^{-m} = \sum_{k=1}^{\infty} a_k 5^{-k} = 0, a_1 a_2 a_3 \dots, \quad (П2.6)$$

где a_k принимают одно из пяти значений: 0, 1, 2, 3, 4. Соотношения (П2.5) означают, что запятая в (П2.6) на каждом шаге переносится на $2p+1$ позиций вправо и целая часть отбрасывается. Отсюда ясно, что при малых $2p+1$ величины ξ_1, ξ_2, \dots будут зависимы. Последовательность ξ_n является периодической. Показано, что ее период равен 2^{m-2} , поскольку величина 5^{2p+1} в двоичном виде оканчивается на 01, а все остальные $m-2$ разрядов чисел ξ_n пробегает все возможные комбинации. В ЕС ЭВМ, например, для записи чисел отводится 32 двоичных разряда, из которых один служит для задания знака числа. Следовательно, $m = 31$ и период $L = 2^{29} = 536869912$. При этом величину множителя M можно принять равной 1220703125. Соответствующая подпрограмма для нахождения последовательности псевдослучайных чисел может иметь вид:

```
SUBROUTINE RAND (NR, RN)
  N1 = 1220703125 * NR
  IF (N1) 1, 2, 2
  1  N1 = N1 + 214783647 + 1
  2  NR = N1
  RN = NR * 0.4656613 E - 9
  RETURN
END
```

Описанный генератор подвергался всесторонней экспериментальной проверке и проявил достаточно хорошие статистические свойства. Чтобы построить с его помощью несколько последовательностей случайных чисел, в машину надо ввести различные начальные значения NR 1, NR 2, ..., $0 < \text{NRI} < 2^{31} - 1$. Если же требуется повторить начальный отрезок какой-либо последовательности, достаточно внутри вызывающей программы присвоить переменной NR ее исходное значение.

Рассмотрим некоторые способы имитации случайных чисел с другими законами распределения. Существуют два основных способа преобразования равномерно распределенных чисел в нормально распределенные. На основе центральной предельной теоремы можно утверждать, что случайная величина

$$\eta^{(l)} = \sqrt{\frac{12}{l}} \sum_{i=1}^l \left(\xi_i - \frac{1}{2} \right)$$

распределена асимптотически нормально с параметрами $(0, 1)$. В частности, при $l = 12$

$$\eta^{(12)} = \sum_{i=1}^{12} \left(\xi_i - \frac{1}{2} \right) \quad (\text{П2.7})$$

имеет распределение приблизительно $N(0, 1)$.

Другой алгоритм из двух равномерно распределенных независимых случайных чисел ξ_1 и ξ_2 образует пару независимых нормально распределенных чисел с параметрами $(0, 1)$:

$$\eta_1 = \sqrt{-2 \ln \xi_1} \sin 2\pi \xi_2, \quad \eta_2 = \sqrt{-2 \ln \xi_1} \cos 2\pi \xi_2. \quad (\text{П2.8})$$

Для моделирования случайных величин с экспоненциальным распределением используется метод обратных функций. Пусть $\varphi(x)$ — строго монотонная функция на интервале $(0, 1)$, а случайная величина η определяется зависимостью $\eta = \varphi(\xi)$, причем для η задана плотность распределения $f(y)$ ($a < y < b$, где границы a, b могут быть бесконечными). В случае, если $\varphi(x)$ монотонно возрастает,

$$F_\eta(x) = \int_a^x f(t) dt = P(\varphi(\xi) < x) = P(\xi < \varphi^{-1}(x)) = \varphi^{-1}(x),$$

где $\varphi^{-1}(x)$ — обратная к $\varphi(x)$ функция. Отсюда

$$\varphi(\xi) = F_\eta^{-1}(\xi).$$

С другой стороны,

$$P(F^{-1}(\xi) < x) = P(\xi > F(x)) = F(x).$$

Поэтому для моделирования η получаем

$$\eta = F^{-1}(\xi). \quad (\text{П2.9})$$

Для монотонно убывающей последовательности формула несколько видоизменяется:

$$\eta = F^{-1}(1 - \xi), \quad (\text{П2.10})$$

что, однако, эквивалентно (П2.9), поскольку ξ и $1 - \xi$ распределены одинаково.

Пусть случайная величина y имеет экспоненциальное распределение, $f(y) = \alpha e^{-\alpha y}$, $y > 0$. В этом случае $F(y) = 1 - e^{-\alpha y}$ и

$$\eta = \alpha^{-1} \ln \xi. \quad (\text{П2.11})$$

Пусть, наконец, требуется воспроизвести реализации n -мерных случайных векторов с заданным математическим ожиданием m и ковариационной матрицей $R = R^* = (\rho_{ij})_{n \times n}$, где

$$\rho_{ij} = M\{(\eta_i - m_i)(\eta_j - m_j)\}. \quad (\text{П2.12})$$

Вектор η с такими моментными характеристиками образуется из совокупности независимых случайных величин ξ_j ($i = 1, n$):

$$m_{\xi_i} = 0, \quad \text{cov}\{\xi_i \xi_j\} = 0 \quad \text{при } i \neq j$$

с помощью преобразования

$$\eta = A \xi + m,$$

где

$$R = AA^*. \quad (\text{П2.13})$$

Если в качестве матрицы A выбрать треугольную:

$$A = \begin{bmatrix} a_{11} & 0 & \dots & 0 \\ a_{21} & a_{22} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix},$$

то для ее коэффициентов получим

$$a_{ij} = \frac{\rho_{ij} - \sum_{k=1}^{j-1} a_{ik} a_{jk}}{\sqrt{\rho_{ij} - \sum_{k=1}^{j-1} a_{jk}^2}},$$

$$\sum_{k=i}^n a_{ik} a_{jk} = 0 \quad (1 < j < i < n).$$

Моделирование псевдослучайных последовательностей. Имитация на ЭВМ стационарных случайных последовательностей с рациональным спектром основана на том известном факте, что любая последовательность такого вида может быть получена путем пропускания дискретного белого шума через стационарную линейную систему. Пусть $\nu(k)$ ($k = 1, 2, \dots$) – последовательность независимых случайных векторных величин размерности m с нулевым математическим ожиданием и единичной ковариационной матрицей:

$$M\{\nu(k)\} = m_{\nu} = 0, \quad (П2.14)$$

$$M\{\nu^*(k)\nu(l)\} = R_{\nu}(l-k) = I_m \delta_{l,k}.$$

В случае, если эта последовательность подается на вход линейной стационарной устойчивой системы вида

$$\xi(k) = \sum_{i=0}^p A^i \nu(k-i) - \sum_{j=1}^q B^j \xi(k-j), \quad (П2.15)$$

то на ее выходе в установившемся режиме будет получена стационарная случайная последовательность с матрицей спектральной плотности

$$\Phi_{\xi}(z) = K^*(z^{-1})K(z) \quad (П2.16)$$

или с корреляционной матрицей

$$R_{\xi}(k) = \frac{1}{2\pi j} \oint_{\Gamma} \Phi_{\xi}(z) z^k \frac{dz}{z}, \quad (П2.17)$$

где

$$K(z) = (I + B^1 z^{-1} + \dots + B^q z^{-q})^{-1} (A^0 + A_1 z^{-1} + \dots + A_p z^{-p}),$$

а через Γ обозначена окружность единичного радиуса на плоскости z .

При моделировании решается обратная задача: корреляционная матрица (П2.17) или матрица спектральной плотности (П2.16) считается заданной, а целью является нахождение уравнения устойчивого формирующего фильтра (П2.15).

Вычисление передаточной функции формирующего фильтра $K(z)$ по матрице спектральной плотности $\Phi_{\xi}(z)$ с использованием соотношения (П2.16) называется ее факторизацией. Эта операция является реализуемой в силу симметричного расположения нулей и полюсов $\Phi_{\xi}(z)$ относительно единичной окружности:

$$\Phi_{\xi}(z) = \Phi_{\xi}^*(z^{-1}). \quad (П2.18)$$

Эта операция не единственна, поскольку требование устойчивости накладывает ограничение лишь на расположение полюсов передаточной матрицы формирующего фильтра $K(z)$; расположение же ее нулей произвольно. В литературе описан ряд численных алгоритмов факторизации рациональных матриц. В частности, в работах Х. Сингха и Е. Денмана, а также К. Блэка и Е. Денмана предложены алгоритмы факторизации относительно мнимой оси скалярных и матричных полиномов¹⁾. Этот

¹⁾ Singh H.N., Denman E.D. A computational algorithm for spectral factorization with applications to filter synthesis // J. Int. Electronics. – 1977. – V. 42, N 6. – P. 559–568. Black K.E., Denman E.D. Spectral factorization of proper matrix polynomials // J. Int. Electronics. – 1977. – V. 42, N 6. – P. 569–579.

алгоритм легко может быть трансформирован для дискретного случая, т.е. для случая факторизации матричных полиномов относительно единичной окружности. При программной реализации уравнения формирующего фильтра (П2.15) следует учесть, что вывод сгенерированной реализации случайного процесса и ее дальнейшее использование в модели могут производиться лишь после затухания переходного процесса, вызванного свободной составляющей на выходе формирующего фильтра.

Отметим в заключение следующее обстоятельство. Если исходный случайный процесс является непрерывным, а для его моделирования используется цифровая техника, то возникает проблема выбора периода квантования сигнала по времени. Действительно, спектральная плотность физически реализуемых непрерывных процессов с ростом частоты асимптотически затухает. В то же время функция спектральной плотности дискретного (квантованного по времени) процесса $\Phi_x(e^{j\omega T})$ является периодической функцией частоты ω с периодом, равным $2\pi/T$, где T – период квантования сигнала. Основным требованием, которое должно быть при этом выполнено, сводится к приближительному воспроизведению спектра исходного процесса в основной полосе частот $[-\Omega, \Omega]$. В таком случае на выбор периода квантования сигнала должно быть наложено следующее условие, вытекающее из теории Котельникова: $T \ll \pi/\Omega$.

Приложение 3

ФАКТОРНЫЙ ЭКСПЕРИМЕНТ 2^k И ЕГО АНАЛИЗ

Рассмотрим первоначально случай $k = 3$. В факторном эксперименте 2^3 существуют восемь комбинаций, образующих "ячейки" факторного пространства. Каждую ячейку можно представить себе как точку трехмерного пространства с координатами $(\pm 1, \pm 1, \pm 1)$. Поэтому эксперимент 2^3 включает в себя восемь отдельных опытов, каждый из которых соответствует одной из вершин куба, определяемого этими точками. Схема двухуровневого трехфакторного эксперимента приведена в табл. ПЗ.1, последний столбец которой соответствует отклику модели. В соответствии с общепринятыми обозначениями факторы обозначены через A, B и C , а отклик — буквами a, b или c , если соответствующий фактор находится на верхнем уровне. Введем так называемую матрицу плана

$$D = (d_{ij}) \quad (i = 1, 2, \dots, 8; j = 1, 2, 3),$$

где d_{ij} равно -1 , если j -й фактор в i -м опыте находится на нижнем уровне, и равно $+1$ в противном случае. Очевидно, что три средних столбца табл. ПЗ.1 как раз и образуют эту матрицу. В общем случае k факторов матрица D имеет размеры $N \times k$, где $N = 2^k$.

Пользуясь обычным подходом, представим модель трехфакторного эксперимента в виде

$$y_{ijlm} = \eta_{ijl} + \epsilon_{ijlm}, \quad (ПЗ.1)$$
$$\eta_{ijl} = \mu + \alpha_i^A + \alpha_j^B + \alpha_l^C + \alpha_{ij}^{AB} + \alpha_{il}^{AC} + \alpha_{jl}^{BC} + \alpha_{ijl}^{ABC},$$

где i, j, l принимают значения 1 или 2, $m = \overline{1, r}$ (r — число повторностей или реплик). Как следует из общей теории, данное представление означает разложение исходного пространства наблюдений на восемь ортогональных подпространств, каждое из которых в данном случае одномерно¹⁾.

Преобразуем эту модель следующим образом. Введем первоначально матрицу независимых переменных X размера $N \times N$ (табл. ПЗ.2). Ее первый столбец образу-

¹⁾ Преобразование (ПЗ.1) можно рассматривать как изменение базиса линейного векторного пространства. Это преобразование неособенное, что вытекает из факта существования обратного линейного отображения:

$$\mu = \eta_{\dots}, \quad \alpha_i^A = \eta_{i..} - \eta_{\dots},$$
$$\alpha_{ij}^{AB} = \eta_{ij.} - \eta_{i..} - \eta_{.j.} + \eta_{\dots},$$
$$\alpha_{ijl}^{ABC} = \eta_{ijl} - \eta_{ij.} - \eta_{i..} - \eta_{.jl} + \eta_{i..} + \eta_{.j.} + \eta_{..l} - \eta_{\dots},$$

где (\dots) означает осреднение по соответствующему индексу. Смысл введенных соотношений сводится к тому, что в новом базисе соответствующие подпространства оказываются взаимно ортогональными. Разумеется, существует бесконечное число способов перехода к ортогональному базису. Тот, который выбран, обладает тем преимуществом, что он естественным образом связан с планом эксперимента.

Таблица П3.1
Эксперимент 2³ без повторностей

Номер опыта	Факторы			Отклик
	A	B	C	
1	-1	-1	-1	1
2	+1	-1	-1	a
3	-1	+1	-1	b
4	+1	+1	-1	ab
5	-1	-1	+1	c
6	+1	-1	+1	ac
7	-1	+1	+1	bc
8	+1	+1	+1	abc

Таблица П3.2
Матрица независимых переменных эксперимента 2³

Опыт	Матрица X								Отклик у.
	1	Матрица плана эксперимента			x ₁ x ₂	x ₁ x ₃	x ₂ x ₃	x ₁ x ₂ x ₃	
		x ₁	x ₂	x ₃					
1	+	-	-	-	+	+	+	-	Y ₁₁₁
2	+	+	-	-	-	-	+	+	Y ₂₁₁
3	+	-	+	-	-	+	-	+	Y ₁₂₁
4	+	+	+	-	+	-	-	-	Y ₂₂₁
5	+	-	-	+	+	-	-	+	Y ₁₁₂
6	+	+	-	+	-	+	-	-	Y ₂₁₂
7	+	-	+	+	-	-	+	-	Y ₁₂₂
8	+	+	+	+	+	+	+	+	Y ₂₂₂

ет вектор $1 = (1, 1, \dots, 1)^*$. Следующие k столбцов ($k = 3$) составлены из матрицы плана эксперимента $D(x_1, x_2, x_3)$. Наконец, компоненты оставшихся столбцов определяются как поэлементные произведения соответствующих столбцов матрицы D , например, $x_1 x_3 = (+1, -1, +1, -1, -1, +1, -1, +1)^*$. Главные эффекты и взаимодействия не являются независимыми, поскольку, в частности

$$\alpha_1^A + \alpha_2^A = 0, \quad \alpha_{i1}^{AB} + \alpha_{i2}^{AB} = 0 \quad (i = 1, 2),$$

$$\alpha_{ij1}^{ABC} + \alpha_{ij2}^{ABC} = 0 \quad (i, j = 1, 2).$$

Поэтому следует исключить "лишние" неизвестные, обозначив через β_0 общее среднее μ , через β_s — главный эффект s -го фактора (т.е. например, $\beta_1 = \alpha_2^A = -\alpha_1^A$), через β_{jl} — взаимодействие двух факторов ($\beta_{12} = \alpha_{11}^{AB} = -\alpha_{12}^{AB} = -\alpha_{21}^{AB} = \alpha_{22}^{AB}$) и т.д. После этого исходную модель (П3.1) можно переписать в виде

$$y_i = \beta_0 + \sum_{j=1}^k d_{ij} \beta_j + \sum_{j=1}^{k-1} \sum_{l=j+1}^k (d_{ij} d_{il}) \beta_{jl} + \dots + (d_{i1} d_{i2} \dots d_{ik}) \beta_{12\dots k} + \epsilon_i. \quad (\text{П3.2})$$

где d_{jj} – элементы матрицы D , равные +1, если фактор j находится в опыте на верхнем уровне, и равные -1 в противном случае. Здесь для отклика модели y и ошибки ϵ принята "сквозная" нумерация, соответствующая номеру строки в табл. ПЗ.2. Используя введенную выше матрицу независимых переменных X , запишем окончательное выражение для оценки неизвестных эффектов в виде следующей регрессионной модели:

$$y = Xz + l, \quad (ПЗ.3)$$

где

$$y = (y_1, y_2, \dots, y_N)^*, \quad z = (z_1, z_2, \dots, z_N)^* =$$

$$= (\beta_0, \beta_1, \dots, \beta_k, \beta_{12}, \dots, \beta_{12\dots k}), \quad l = (\epsilon_1, \epsilon_2, \dots, \epsilon_N)^*.$$

В этих соотношениях z_1 соответствует генеральному среднему μ , z_2, \dots, z_{k+1} – главным эффектам, а z_{k+2}, \dots, z_N – взаимодействиям соответствующего порядка (в частности, z_N представляет собой взаимодействие порядка k). В двухуровневых экспериментах принято определять эффекты несколько иначе, чем в общей модели. Так, вместо α_1^A и α_2^A обычно вводится

$$\alpha^A = \eta_{2..} - \eta_{1..} = 2\alpha_2^A = -2\alpha_1^A. \quad (ПЗ.4)$$

Аналогично этому, вместо взаимодействий α_{ij}^{AB} ($i, j = 1, 2$) новое взаимодействие вводится по формуле

$$\alpha^{AB} = \frac{1}{2} [(\eta_{22.} - \eta_{12.}) - (\eta_{21.} - \eta_{11.})] = 2\alpha_{11}^{AB} = -2\alpha_{11}^{AB} \quad (ПЗ.5)$$

и т.д. Таким образом, величины z_i , начиная с $i = 2$, означают половины переопределенных главных эффектов и взаимодействий. Введение новых величин связано с естественным желанием избавиться от "лишних" переменных и перейти к совокупности линейно независимых величин.

Действуя стандартным образом, на основании регрессионной модели (ПЗ.3) можно получить уравнения для оценок неизвестных эффектов:

$$(X^*X)\hat{z} = X^*\bar{y}, \quad (ПЗ.6)$$

где через \bar{y} обозначены либо непосредственно измеренные в эксперименте величины, либо величины, осредненные по повторностям:

$$\bar{y}_{ijl} = y_{ijl}$$

При этом предполагается, что число повторностей в каждой из ячеек одинаково.

Легко заметить, что матрица X ортогональна: скалярные произведения ее различных столбцов равны нулю:

$$\sum_{s=1}^N x_{si} x_{sj} = \begin{cases} N & \text{при } i = j, \\ 0 & \text{при } i \neq j. \end{cases} \quad (ПЗ.7)$$

В связи с этим

$$(X^*X)^{-1} = \frac{1}{N} I \quad (ПЗ.8)$$

и решение нормальных уравнений метода наименьших квадратов (ПЗ.6) принимает простой вид

$$\hat{z} = \frac{1}{N} X^* \bar{y}. \quad (ПЗ.9)$$

Последнее соотношение легко интерпретируется. Так, оценка генерального среднего модели есть не что иное, как выборочное среднее:

$$\hat{z}_1 = \hat{\mu} = \frac{1}{N} \sum_{i=1}^N \bar{y}_i = \frac{1}{2N} \sum_{i=1}^N \sum_{m=1}^r y_{im}, \quad (ПЗ.10)$$

где r – число повторностей.

Поскольку каждый фактор в половине опытов находится на верхнем и в половине на нижнем уровнях, оценка любого главного эффекта равна удвоенной разности средних откликов модели на этих двух уровнях, например,

$$\hat{\alpha}^A = 2\hat{z}_2 = \frac{2}{N} \sum_{s=1}^N x_{s1} \bar{y}_s = \frac{\sum_{p=1}^{N/2} \bar{y}_p}{N/2} - \frac{\sum_{q=1}^{N/2} \bar{y}_q}{N/2}, \quad (\text{П3.11})$$

причем индекс p относится к откликам тех комбинаций факторов, при которых фактор A принимает максимальное значение, а индексу q соответствует минимальное значение этого фактора.

Имея оценки эффектов, можно найти их дисперсии и ковариации. Ковариационная матрица оценок находится обычным образом:

$$\hat{\Sigma}_z = \sigma^2 (X^* X)^{-1} = \frac{\sigma^2}{N} I. \quad (\text{П3.12})$$

Отсюда следует, что все полученные оценки некоррелированы. Поскольку все главные эффекты и взаимодействия (исключая общее среднее) определены как удвоенные величины z_j , дисперсии всех оценок одинаковы и равны $4\sigma^2/N$, а дисперсия оценки среднего равна σ^2/N .

Наконец, можно для этого частного случая вывести соотношения, необходимые для проверки гипотез о значимости влияния выделенных эффектов. Если исключить генеральное среднее, то в эксперименте 2^k имеется $2^k - 1$ оцениваемых эффектов — главных эффектов и взаимодействий всех возможных порядков. Использование общих формул дисперсионного анализа позволяет записать для соответствующих сумм квадратов следующее выражение:

$$SS_j = r 2^{k-2} \hat{\alpha}_j^2 \quad (j = 1, (2^k - 1)), \quad (\text{П3.13})$$

где $\hat{\alpha}_j = 2\hat{z}_{j+1}$. Эти величины входят в числитель F -критерия при числе его степеней свободы, равном единице. При $r > 1$ в качестве знаменателя F -статистики принимается сумма квадратов ошибок SS_I , имеющая $2^k(r-1)$ степеней свободы. Учитывая, что

$$SS_{\text{общ}} = SS_I + \sum_{j=1}^{N-1} SS_j,$$

а также, что

$$SS_{\text{общ}} = \sum_{i=1}^N \left(\bar{y}_i - \frac{1}{N} \left(\sum_{j=1}^N \bar{y}_j \right) \right)^2,$$

можно записать

$$SS_I = SS_{\text{общ}} - \sum_{j=1}^{N-1} SS_j. \quad (\text{П3.14})$$

Таким образом, при выполнении гипотезы об отсутствии влияния j -го эффекта на выход модели можно получить две независимые оценки дисперсии ошибки эксперимента σ^2 : $(\hat{\sigma}_1^2)_j = SS_j$ и $\hat{\sigma}_{II}^2 = SS_I / 2^k (r - 1)$ при числах степеней свободы соответственно 1 и $2^k (r - 1)$. При $r = 1$ проверку гипотезы о значимости каких-либо эффектов можно производить только в тех случаях, если имеется априорная информация о том, что влияние некоторых других эффектов пренебрежимо мало. Так, например, если заранее известно, что влияние эффектов с номерами $m + 1, m + 2, \dots, N$ незначимо, то в качестве оценки дисперсии σ^2 можно принять

$$\hat{\sigma}_{II}^2 = \frac{SS_{m+1} + \dots + SS_N}{N - m},$$

что позволяет осуществить проверку гипотезы о значимости влияния всех остальных эффектов от первого до m -го включительно.

Т а б л и ц а П 3.3
Дробная реплика $I = ABC$

Опыт	1	x_1	x_2	x_3	$x_1 x_2$	$x_1 x_3$	$x_2 x_3$	$x_1 x_2 x_3$	у
2	+	+	-	-	-	-	+	+	$y_2 = y_{211}$
3	+	-	+	-	-	+	-	+	$y_3 = y_{121}$
5	+	-	-	+	+	-	-	+	$y_5 = y_{112}$
8	+	+	+	+	+	+	+	+	$y_8 = y_{222}$

С ростом числа факторов k количество отдельных опытов в эксперименте возрастает. При этом приходится отказываться не только от повторностей (реплик), но и от полного перебора всех комбинаций факторов.

Наиболее естественный путь, который мог бы привести к сокращению числа экспериментальных точек, как очевидно, должен быть связан с априорными сведениями о наличии или отсутствии тех или иных эффектов. Однако такая информация, как правило, отсутствует. Кроме того, в настоящее время неизвестно общее решение задачи построения минимального плана, обеспечивающего возможность получения независимых оценок ненулевых эффектов, если известно, что часть эффектов отсутствует. В связи с этим часто применяют регулярные способы сокращения числа опытов с последующим анализом свойств полученного плана. Наиболее простой способ — построение д р о б н ы х р е п л и к.

Рассмотрим вновь табл. П3.2. Пусть план эксперимента включает только половину всех опытов, а именно те из них, для которых составляющие вектора x_1, x_2, x_3 равны +1. Четыре опыта, удовлетворяющие этому условию, сведены в табл. П3.3.

Поскольку мы здесь имеем только четыре опыта, порождаемое наблюдениями пространство четырехмерно. Оно в данном случае может быть представлено в виде прямой суммы не более четырех ортогональных друг другу подпространств. Из вида матрицы X следует, что в ней попарно линейно зависимыми (и даже совпадающими) являются столбцы 1 и $x_1 x_2 x_3$, x_1 и $x_2 x_3$, x_2 и $x_1 x_3$, x_3 и $x_1 x_2$, которые образуют четыре группы ортогональных векторов. Следовательно, в данном случае, например, оценки μ и α^{ABC} попадают в одно и то же подпространство, имеющее размерность 1, и, следовательно, не могут быть оценены независимо друг от друга. Как говорят, оценки μ и α^{ABC} образуют смешанную группу. Точно так же попарно смешаны эффекты α^A и α^{BC} , α^B и α^{AC} , α^C и α^{AB} . Разумеется, если все взаимодействия равны нулю, то на основании исследуемого плана можно получить несмещенные оценки генерального среднего и трех главных эффектов. В других же случаях получение всех оценок невозможно.

Приложение 4

КРИТЕРИИ СРАВНЕНИЯ ПЛАНОВ РЕГРЕССИОННОГО ЭКСПЕРИМЕНТА

Рассмотрим регрессионный эксперимент, целью которого является оценка параметров модели

$$M\{y | x\} = \eta(x, \theta) = \theta^* f(x), \quad (\text{П4.1})$$

где $\theta = (\theta_1, \theta_2, \dots, \theta_m)^*$ – m -мерный вектор неизвестных параметров, $f(x) = (f_1(x), f_2(x), \dots, f_m(x))^*$ – известные функции, x – конечномерный вектор. Предполагается, что эксперимент, связанный с измерением или вычислением y , может быть выполнен в некоторых (заранее не фиксированных) точках некоторой области планирования $\Sigma \subset R^n$. Пусть измерительный или имитационный эксперимент выполнен в точках x_1, x_2, \dots, x_N , $x_i \in \Sigma$, причем в результате эксперимента получен набор независимых скалярных случайных величин y_1, y_2, \dots, y_N с дисперсиями $\sigma_1^2, \sigma_2^2, \dots, \sigma_N^2$.

Известно, что наилучшими линейными оценками неизвестных параметров θ являются величины

$$\hat{\theta} = M^{-1}z, \quad (\text{П4.2})$$

где матрица M и вектор z равны

$$M = \sum_{i=1}^N N w_i f(x_i) f^*(x_i), \quad (\text{П4.3})$$

$$z = \sum_{i=1}^N w_i y_i f(x_i). \quad (\text{П4.4})$$

При этом предполагается, что матрица M невырождена, а $w_i = \sigma_i^{-2}$. Дисперсионная (ковариационная) матрица оценок $\hat{\theta}$ равна

$$D(\hat{\theta}) = M^{-1}. \quad (\text{П4.5})$$

В случае, если в каждой из N точек эксперимент повторен r_i раз, то формулы (П4.2) – (П4.4) сохраняются, если подставить в них вместо σ_i^{-2} величины $w_i = r_i \sigma_i^{-2}$, а вместо

y_i – величины $r_i^{-1} \sum_{j=1}^{r_i} y_{ij}$, осредненные для каждой из точек x_i . Матрица M , называемая

информационной матрицей Фишера, играет важную роль при анализе эффективности планов эксперимента, поскольку именно этой матрицей определяется точность оценок (см. (П4.5)). Нетрудно видеть, что отдельные экспериментальные точки вносят в эту матрицу аддитивные вклады:

$$M = \sum_{i=1}^N M(x_i), \quad (\text{П4.6})$$

где $M(x_i) = w_i f(x_i) f^*(x_i)$.

При сделанных выше предположениях наилучшей линейной оценкой поверхности отклика $\eta(x, \theta)$ в произвольной точке x является

$$\hat{\eta} = \hat{\theta}^* f(x), \quad (П4.7)$$

а дисперсия этой оценки равна

$$d(x) = f^*(x) D(\hat{\theta}) f(x). \quad (П4.8)$$

Существенно отметить, что дисперсионная матрица оценок $\hat{\theta}$ и дисперсия оценки поверхности отклика $\hat{\eta}$ в точке x зависят от всего набора точек x_1, x_2, \dots, x_N и величин r_1, r_2, \dots, r_N , образующих план эксперимента. Поэтому можно улучшить "наилучшие" оценки за счет выбора числа и расположения этих точек в области планирования Σ , а также повторностей опыта в каждой из точек x_i .

Можно сформулировать общие критерии эффективности планов эксперимента, в которые будет входить наряду с другими характеристиками точность получаемых оценок. Однако здесь мы этого делать не будем. В настоящем приложении приведем лишь известные из литературы (см., например, [4] из перечня литературы к гл. 4) критерии сравнения экспериментов, исходя из требований по точности оценок. Эти оценки полезны, если необходимо выбрать тот или иной план эксперимента из некоторого класса.

Эксперимент E_1 предпочтительнее эксперимента E_2 в смысле точности оценок параметров ($E_1 > E_2$), если

1) $D(E_1) < D(E_2)$, т.е. если разность двух матриц $D(E_2) - D(E_1)$ является положительно определенной;

$$2) |D(E_1)| < |D(E_2)|;$$

$$3) \text{Sp } D(E_1) < \text{Sp } D(E_2);$$

$$4) \max_i D_{ii}(E_1) < \max_i D_{ii}(E_2);$$

$$5) D(\vartheta, E_1) < D(\vartheta, E_2), \text{ где } \vartheta = T\theta.$$

Смысл большинства критериев очевиден. Способ сравнения 5), как очевидно, переходит в 1), если $T = I_m$. С другой стороны, если нас интересует точность не всех параметров, а только части, например,

$$\theta_1, \theta_2, \dots, \theta_l \quad (l < m),$$

то матрицу T в 5) следует выбирать равной

$$T = \| I_l, 0 \|,$$

где 0 – нулевая $l \times (m - l)$ -матрица.

Рассмотрим способы сравнения экспериментов в смысле точности оценки поверхности отклика $\hat{\eta}(x, \theta)$. Как следует из (П4.8), дисперсия этой оценки, помимо плана эксперимента, зависит от самого аргумента x . Поэтому, выбирая скалярный критерий сравнения, следует перейти к некоторым осредненным оценкам.

Будем считать, что эксперимент E_1 предпочтительнее эксперимента E_2 , если

$$6) \max_x d(x, E_1) < \max_x d(x, E_2);$$

$$7) \int_Z d(x, \epsilon_1) dx < \int_Z d(x, E_2) dx.$$

При этом Z -область интегрирования может совпадать с областью планирования Σ , являясь частью ее или располагаться вне Σ . Последний случай встречается в задаче экстраполяции. В частном случае область Z может вырождаться в точку:

$$d(z, E_1) < d(z, E_2),$$

где z может принадлежать области Σ или находиться вне этой области

СПИСОК ЛИТЕРАТУРЫ

К главе 1

1. *Бельговский М.А.* Устойчивость листовертки против низких температур//Сообщ. Института леса АН СССР. – 1959. – Вып. 12. – С. 75–83.
2. Биотехнология/Под ред. А.А. Баева. – М.: Наука. – 1984. – 309 с.
3. *Будыко М.И.* Глобальная экология. – М.: Мысль. – 1977. – 327 с.
4. *Васильев Б.Р., Звонцова Н.А., Савинов Н.П., Шмидт В.М.* Математический анализ роста деревьев//Ботанический журнал. – 1973. – Т. 58, № 9. – С. 1294–1301.
5. *Викторов Г.А.* К вопросу о причинах размножения насекомых//Зоологический журнал. – 1955. – Т. 34, вып. 2. – С. 48–51.
6. *Вольпе П.* Биохимия клеточного цикла. – М.: Мир. – 1979. – 95 с.
7. *Воронов А.А.* Устойчивость, управляемость, наблюдаемость. – М.: Наука. – 1979. – 335 с.
8. *Воронцов А.И.* Итоги и перспективы лесозащиты//Лесоведение. – 1977, № 5. – С. 37–50.
9. *Герасименко Л.М., Пущева М.А.* Получение синхронной культуры *Chlorella vulgaris*//Микробиология. – 1968. – Т. 37, вып. 1. – С. 70–74.
10. *Гильманов Т.Г.* Математическое моделирование биохимических циклов в травяных экосистемах. – М.: Изд-во МГУ. – 1978.
11. *Дибров Б.Ф., Жаботинский А.М., Холоденко Б.Н.* Динамическая устойчивость метаболической цепи с одной петлей обратной связи//Биофизика. – 1981. – Т. 26, вып. 4. – С. 590–595.
12. *Дибров Б.Ф., Жаботинский А.М., Холоденко Б.Н.* Динамическая устойчивость и параметрическая стабилизация стационарных состояний неразветвленных метаболических путей//Биофизика. – 1981. – Т. 26, вып. 5. – С. 790–795.
13. *Егоров Н.Н., Рубцова Н.Н., Соложенкина Т.Н.* Дубовая листовертка в Воронежской области//Зоологический журнал. – 1961. – Т. 40, вып. 8. – С. 1172–1183.
14. *Енькова Е.И.* Влияние температуры воздуха на набухание и раскрытие листовых почек черешчатого дуба//Научные записки ВЛТИ. – 1960. – Т. 21. – С. 71–84.
15. *Епифанова О.И.* Гормоны и размножение клеток. – М.: Наука. – 1965. – 243 с.
16. *Жаботинский А.К.* Концентрационные автоколебания. – М.: Наука. – 1974. – 178 с.
17. *Заславский Б.Г.* Стохастическая модель роста клеточной популяции//Проблемы кибернетики. – 1972. – Вып. 25. – С. 139–151.
18. *Заславский Б.Г., Полуэктов Р.А.* Исследование динамики импульсной системы управления культивированием микроорганизмов//Проблемы кибернетики. – 1973. – Вып. 27. – С. 187–194.
19. *Заславский Б.Г.* Колебательность и устойчивость режимов в культивационных системах//Динамика систем. Межвузовский сборник. – Горький: ГГУ, 1976. – Вып. 9. – С. 142–155.
20. *Иерусалимский Н.Д., Неронова Н.М.* Количественная зависимость между концентрацией продуктов обмена и скоростью роста микроорганизмов//ДАН СССР. – 1965. – Т. 161, № 6. – С. 1437–1440.

21. *Исаев А.С., Хлебопрос Р.Г.* Анализ динамика численности лесных насекомых на основе принципа стабильности подвижных экологических систем//Журнал общей биологии. – 1974. – Т. 35, № 5. – С. 737–745.
22. *Ли Э., Маркус Л.* Основы теории оптимального управления. – М.: Наука. – 1967. – 574 с.
23. *Малек И., Фенцль З.* Непрерывное культивирование микроорганизмов. – М.: Мир. – 1968. – 564 с.
24. *Марри Дж.* Нелинейные дифференциальные уравнения в биологии. Лекции о моделях. – М.: Мир. – 1983. – 397 с.
25. *Назаренко В.Г., Сельков Е.Е.* Механизм контактного угнетения как возможный источник сверхнизкочастотных биологических ритмов//Колебательные процессы в биологических и химических системах. – Пушкино-на-Оке. – 1967. – Т. 2. – С. 145–148.
26. *Нейлор Т.* Машинные имитационные эксперименты с моделями экономических систем. – М.: Мир. – 1975. – 500 с.
27. *Николаев П.И., Соколов Д.П.* Кинетическая зависимость процессов культивирования микроорганизмов//Прикладная биохимия и микробиология. – 1968. – Т. 4. – С. 365 – 372.
28. *Николаев П.И., Соколов Д.П.* Определение коэффициентов уравнений, описывающих культивирование микроорганизмов//Прикладная биохимия и микробиология. – 1968. – Т. 4. – С. 562–569.
29. *Николис Г., Пригожин И.* Самоорганизация в неравновесных системах. – М.: Мир. – 1979. – 512 с.
30. *Одум Ю.* Основы экологии. – М.: Мир. – 1975. – 740 с.
31. *Перт С.Дж.* Основы культивирования микроорганизмов и клеток. – М.: Мир. – 1978. – 331 с.
32. *Полуэктов Р.А., Пых Ю.А., Швытов И.А.* Динамические модели экологических систем. – Л.: Гидрометеиздат. – 1980. – 288 с.
33. *Полуэктов Р.А., Вол И.А., Заславский Б.Г., Пых Ю.А., Финтушал С.М.* Базовая модель продуктивности агроэкосистем//Проблемы экологического мониторинга моделирование экосистем. – Л.: Гидрометеиздат. – 1983. – Т. 6, – С. 189–201.
34. *Полуэктов Р.А., Вол И.А., Заславский Б.Г., Малкина И.Г., Пых Ю.А., Финтушал С.М.* Имитационная модель развития агроценоза. Препринт. – М.: ВНИИ системных исследований. – 1984. – 84 с.
35. *Романовский Ю.М., Степанова Н.В., Чернавский Д.С.* Математическая биофизика. – М.: Наука. – 1984. – 304 с.
36. *Рубин А.Б., Пытьева Н.Ф., Ризниченко Г.Ю.* Кинетика биологических процессов. – М.: Изд-во МГУ. – 1977. – 328 с.
37. *Рубцов В.В.* Математическое описание фазы личинки у листогрызущих насекомых//Научно-техн. бюллетень по агрономической физике. – Л.: Агрофизический институт. – 1977. – № 31. – С. 9 – 14.
38. *Рубцов В.В., Швытов И.А.* Модель динамики численности листогрызущих насекомых в лесной экосистеме//Сб. трудов по агрономической физике. – Л.: Агрофизический институт. – Вып. 38. – 1976. – С. 85–95.
39. *Рубцов В.В., Рубцов Н.Н.* Анализ взаимодействия листогрызущих насекомых с дубом. – М.: Наука. – 1984. – 182 с.
40. *Рубцова Н.Н., Рубцов В.В.* Об учете яйцекладок зеленой дубовой листовертки//Лесоведение. – 1978. – № 2. – С. 31–38.
41. *Сельков Е.Е.* Исследование условий возникновения периодических колебаний в системах ферментативных реакций с обратной связью//Колебательные процессы в биологических и химических системах. – М.: Наука. – 1967. – Т. I. – С. 81–92.
42. *Тимофеев-Ресовский Н.В.* Популяции, биогеоценозы и биосфера Земли//Математическое моделирование в биологии. – М.: Наука. – 1975. – С. 19 – 29.
43. *Тимофеев-Ресовский Н.В., Тюрюканов А.Н.* Об элементарных биохорологических подразделениях биосферы//Бюлл. МОИП. Сер. биол. – 1966. – Т. 21, № 1. – С. 123–132.
44. *Фабф Г.* Промышленные микроорганизмы//Промышленная микробиология и успехи генетической инженерии. – М.: Мир. – 1984. – 172 с.
45. *Хакен Г.* Синергетика. – М.: Мир. – 1980. – 404 с.

46. Шеннон Р. Имитационное моделирование систем – искусство и наука. – М.: Мир. – 1978. – 418 с.
47. Эбелинг В. Образование структур при необратимых процессах. – М.: Мир. – 1979. – 279 с.
48. Allwright D.J. The Hopf Bifurcation in Some Biochemical Control Loops//Journal of Mathematical Biology. – 1981. – V. 11, № 1. – P. 85–93.
49. Clark C.W. Mathematical Models of the Renewable Resources//SIAM Review. – 1979. – V. 21, № 1. – P. 81–99.
50. Dai L.S. On the existence, uniqueness and global asymptotic stability of the periodic solution of the modified Michaelis-Menten mechanism//Journal of differential equations. – 1979. – V. 31. – P. 392–417.
51. Dibrov B.F., Zhabotinsky A.M., Kholodenko B.N. Dynamik Stability of Steady State and Static Stabilization in Unbranched Metabolic Parthways//Journal of Mathematical Biology. – 1982. – V. 15, № 1. – P. 51–64.
52. Hasting S., Tyson J., Webster D. Existence of Periodic Solutions for Nenative Feedback Cellular Control Systems//Journal of Differential Equations. – 1977. – V. 25. – P. 39–64.
53. Holling C.S., Dantzig G.B., Baskerville C., Jones D.D., Clark W.C. A case study of forest ecosystem pest management//Luxemburg: IASSA. – 1975. – 46 p.
54. Jones D.D. Biology of budworm model. – Luxemburg. – IASSA. – 1974. – 42 p.
55. Lee K.Y., Berr R.U., Gage S.H., Kharker A.N. Formulation of mathematical model for insect pest ecosystems – the cereal leaf hettle problem//J. Theor. Biol. – 1971. – V. 59, № 1. – P. 13–21.
56. Levine D.S. Periodic or unbounded solutions for a class of three-dimensional ode systems with biological applications//Trends in the theory and practice of non-linear analysis/Ed. V. Lakshimkantham. – North-Holland: Elsevier Science publishers. B.V. – 1985. – P. 263–274.
57. Morris R.F. (ed.). The dynamics of epidemic spruce budworm populations//Mem. Entomol. Soc. Can. – 1963. – № 31. – P. 1–332.
58. Mees A.I., Rapp P.E. Periodic Metabolic Systems Oscillations in Maltiple-Loop Negative Feedback Biochemical Control Network//Journal of Mathematical Biology. – 1978. – V. 5. P. 99–114.
59. Nihei T., Sasa T., Miyashi S., Suzuki K., Tamiya M. Change of Photosynthetic Activity of Clorella Cells During the Course of Their Normal Life Cycle//Archive fur Mikrobiologia. – 1954. – V. 21. – P. 156–166.
60. Regan D.L., Roper G.H., Moss F.J. Response of Continuous Culture to Stimuli in Glucose Feed Rate and Dilution Rate//Biotechnology and Bioengineering. – 1971. – V. 13, № 6.
61. Selgrade J.E. Mathematical Analysis of Cellular Process with Positive Feedback//SIAM Journal on Applied Mathematic. – 1979. – V. 36, № 2. – P. 219–229.
62. Selgrade J.E. A Hopf Bifurcation in Single-Loop Positive Feedback Systems//Quarterly Applied Mathematics. – 1982. – V. 40, № 3. – P. 347–351.
63. Schutte F. Untersuchungen über die Populationsdynamik des Fischenwichtlers//Zeitschrift fur angew. Entomol. – 1957. – B. 40, № 13.
64. Schwerdtfeger F. Über die Ursache des Massenwechsels der Insekten//Zeitschrift fur angew. Entomol. – 1941. – B. 28, № 2.
65. Schwerdtfeger F. Die Waldkrankheiten, II Auflage. – Hamburg-Berlin. – 1957. – S. 21–30.
66. Schwerdtfeger F. Forstinsekten in Ur- und Nuterwald, Algem. Forstzeitschrift. – 1975. – B. 43, № 2. – S. 481–485.
67. Swick K.E. Some reducible models of age dependent dynamics//SIAM Journ. on Applied Mathematics. – 1985. – V. 45, № 2. – P. 256–267.
68. Van Dyne G.M. Grassland management, research and training viewed in a system context //Range Sci. Dep. Sci. Fort Collins: Colorado State Univ. – 1969. Ser. № 3. – 50 p.
69. Van Dyne G.M. (ed.). The ecosystem concept in natural resource management. – New York: Acad. Press. – 1969. – 383 p.
70. Vrana D. Daughter Cells as an important factor in determining the physiological state of Yeast populations//Biotechnology and bioengineering. – 1976. – V. 18, № 3–4. – P. 297–309.

71. *Vrana D.* Specific rate of growth of buds related to the genealogical age of the parent cell in *Canadia utili*//*Folia Microbiologica*. – 1983. – V. 28 (2). – P. 138–140.
72. *Weiss P., Kavanau A.* Model of growth and growth control in mathematical terms//*Journ. of Genetic Physiology*. – 1957. – V. 41. – P. 1–47.
73. *Winkler C.* An optimization techniq for the budworm forest pest model//*Research memorandum RM 75-11*. – Luxemburg: IASSA. – 1975.

К главе 2

1. *Боуэн Р.* Методы символической динамики. – М.: Мир. – 1979. – 245 с.
2. *Баснакьян И.А., Бирюков В.В., Крылов Ю.М.* Математическое описание основных кинетических закономерностей процесса культивирования микроорганизмов//*Микробиология. Итоги науки и техники. Управляемое и непрерывное культивирование микроорганизмов*. – М.: ВИНТИ. – 1976. – Т. 5. – С. 5–75.
3. *Гантмахер Ф.Р.* Теория матриц. – М.: Наука. – 1967. – 575 с.
4. *Гиллдерман Ю.И., Куорина К.И., Полетаев И.А.* Модели л-систем (системы с лимитирующими факторами)//*Исследования по кибернетике/Под ред. А.А. Ляпунова*. – М.: Советское радио. – 1970. – 240 с.
5. *Гилмор Р.* Прикладная теория катастроф, т. 2. – М.: Мир. – 1984. – 285 с.
6. *Заславский Б.Г.* Исследование динамических характеристик комплексной культуры с помощью математической модели//*Микробиологическая промышленность*. – 1973. – № 7 (103). – С. 25–28.
7. *Заславский Б.Г.* Математические основы анализа возрастной структуры популяции, культивируемой в многокомпонентной среде//*Теоретическая и экспериментальная биофизика. Межвузовский сборник*. – Калининград. – 1975. – С. 148–156.
8. *Заславский Б.Г.* Колебательность и устойчивость режимов в культивационных системах//*Динамика систем*. – 1976. – Вып. 9. – С. 142–155.
9. *Заславский Б.Г.* Хаос в популяции//*ДАН СССР*. – 1981. – Т. 258, № 3. – С. 533–536.
10. *Заславский Б.Г.* Исследование квазигомоклинической структуры, порождаемой полугруппой операторов в банаховом пространстве//*Сибирский математический журнал*. – 1982. – Т. 23, № 6. – С. 80–90.
11. *Заславский Б.Г.* Динамика численности управляемых популяций//*Автоматика и телемеханика*. – 1983. – № 2. – С. 71–80.
12. *Заславский Г.М.* Стохастичность динамических систем. – М.: Наука. – 1984. – 271 с.
13. *Иерусалимский Н.Д.* Биохимические основы регуляции скорости роста микроорганизмов//*Изв. АН СССР. Сер. биол.* – 1967. – С. 339–350.
14. *Иерусалимский Н.Д., Чернавский Д.С.* К вопросу об определяющем звене в системе ферментативных реакций//*Изв. АН СССР. Сер. биол.* – 1965. – Т. 5. – С. 665–672.
15. *Красносельский М.А.* Оператор сдвига по траекториям дифференциальных уравнений. – М.: Наука. – 1966. – 331 с.
16. *Крейн С.Г.* Функциональный анализ. – М.: Наука. – 1972. – 544 с.
17. *Кутателадзе С.С., Рубинов А.М.* Двойственность Минковского и ее приложения. – Новосибирск: Наука. – 1976. – 254 с.
18. *Лифшиц С.* Устойчивость нелинейных систем автоматического управления. – М.: Мир. – 1967. – 1983 с.
19. *Лурье А.И.* Некоторые нелинейные задачи теории автоматического регулирования. – М., Л.: Гостехиздат. – 1951. – 216 с.
20. *Макин И.Г.* Теория устойчивости движения. – М.: Наука. – 1966. – 530 с.
21. *Марсден Дж., Мак-Кракен М.* Бифуркация рождения цикла и ее приложения. – М.: Мир. – 1980. – 368 с.
22. *Моришима М.* Равновесие, устойчивость, рост. – М.: Мир. – 1972. – 297 с.
23. *Немыцкий В.В., Степанов В.В.* Качественная теория дифференциальных уравнений. – М., Л.: Гостехиздат. – 1947. – 448 с.
24. *Никайдо Х.* Выпуклые структуры и математическая экономика. – М.: Мир. – 1972. – 517 с.
25. *Одум Ю.* Основы экологии. – М.: Мир. – 1975. – 740 с.

26. *Опоицев В.И.* Динамика коллективного поведения // Автоматика и телемеханика. – 1975. – № 1. – С. 129–138.
27. *Перт С.Дж.* Основы культивирования микроорганизмов и клеток. – М.: Мир. – 1978. – 331 с.
28. *Печуркин Н.С., Терсков И.А.* Анализ кинетики роста и эволюции микробных популяций. – Новосибирск: Наука. – 1975. – 275 с.
29. *Плисс В.А.* Нелокальные проблемы теории колебаний. – М., Л.: Наука. – 1964. – 367 с.
30. *Позмогова Н.И.* Синхронные культуры // Итоги науки и техники. Сер. микробиол. Культивирование микроорганизмов. – 1981. – Т. II. – С. 118–151.
31. *Полужтов Р.А., Пых Ю.А., Швытов И.А.* Динамические модели экологических систем. – Л.: Гидрометеоздат. – 1980. – 288 с.
32. *Романовский Ю.М., Степанова Н.В., Чернавский Д.С.* Математическая биофизика. – М.: Наука. – 1984. – 304 с.
33. *Руш Н., Абетс П., Лалуа М.* Прямой метод Ляпунова в теории устойчивости. – М.: Мир. – 1980. – 300 с.
34. *Фишберн П.* Теория полезности для принятия решений. – М.: Наука. – 1978. – 352 с.
35. *Хмель И.А., Иерусалимский Н.Д.* Влияние азрации на рост *Azotobacter vinelandii* в проточной культуре // Микробиология. – 1967. – Т. 36. – С. 632–639.
36. *Шефер Х.* Топологические векторные пространства. – М.: Мир. – 1971. – 359 с.
37. *Шкидченко А.Н.* Влияние концентрации растворенного кислорода на рост и потребление субстрата *Torulopsis latvica* // Микробиология. – 1973. – Т. 42. – С. 816–821.
38. *Alekseev V.V., Kornilovsky A.N.* Ecosystems stochasticity model // Ecological Modeling. – 1985. – V. 28. – P. 217–229.
39. *Bunow B.* How Haotic is Haos? Haotic and Other "Noisy" Dynamics in Frequency Domain // Mathematical Bioscience. – 1979. – V. 47, № 3–4. – P. 221–237.
40. *Busenberg S.N., Travis C.C.* On the Use of Reducible Functional Differential Equations in Biological Models // Journal of Mathematical Analysis and Applications. – 1982. – V. 89, № 1. – P. 46–66.
41. *Chow S.N., Hale J.K.* Periodic Solution of Autonomous Equations // Journal of Mathematical Analysis and Applications. – 1978. – V. 66. – P. 495–506.
42. *Demetrius L.* Multiplicative Processes I // Mathematical Biosciences. – 1971. – V. 12. – P. 261–272.
43. *Fenichel N.* Asymptotic Stability with Rate Conditions // Indiana University Mathematical Journal. – 1974. – V. 23, № 12. – P. 1109–1137.
44. *Kaplan J.L., Yorke J.A.* Chaotic Behavior of Multidimensional Differential Equations // Lecture Notes in Mathematics. – 1979. – V. 730. – P. 204–227.
45. *Lewowicz J.* Invariant Manifolds for Regular Points // Pacific Journal of Mathematics. – 1981. – V. 96, № 1. – P. 163–174.
46. *Marotto F.R.* Snap–Back Repeller Imply Chaos in R^n // Journal of Mathematical Analysis and Applications. – 1978. – V. 63, № 1. – P. 199–233.
47. *Marotto F.R.* Perturbation of Stable and Chaotic Differential Equations // Journal of Mathematical Analysis and Applications. – 1979. – V. 72, № 2. – P. 716–729.
48. *Mor J.R., Fiecher A.* Continuous Cultivation of *Saccharomyces Cerevisiae* I. Growth on Ethanol Under Steady State Conditions // Biotechnology and Bioengineering. – 1968. – V. 10, № 2. – P. 159–176.
49. *Nussbaum R.D.* Periodic Solutions of Some Nonlinear Autonomous Differential Equations // Annali di Matematica Pura ed Applicata. – 1974. – V. 101. – P. 263–306.
50. *Nussbaum R.D.* A global bifurcation Theorem with Applications to Functional Analysis. – 1975. – V. 19. – P. 319–339.
51. *Shiraiwa K., Kurata M.* Generalization of Theorem of Marotto // Nagoya Mathematical Journal. – 1981. – V. 82. – P. 83–97.
52. *Vrana D.* Daughter Cells as an Important Factor in Determining the Physiological State of Yeast Populations // Biotechnology and Bioengineering. – 1976. – v. 18, № 3–4. – P. 297–309.
53. *Vrana D.* Specific Rate of Growth of Buds Related to the Genealogical Age of the Parent Cell in *Candida Utilis* // Folia Microbiologica. – 1983. – V. 28(2). – P. 138–140.

1. *Аткинсон Б.* Биохимические реакторы. – М.: Пищевая промышленность. – 1979. – 280 с.
2. *Барбашин Е.А.* Функции Ляпунова. – М.: Мир. – 1970. – 240 с.
3. *Барбашин Е.А.* Методы сечений в теории динамических систем. – Минск: Наука и техника. – 1979. – 120 с.
4. *Болтянский В.Г.* Математические методы оптимального управления. – М.: Наука. – 1969 – 408 с.
5. *Воронов А.А.* Устойчивость, управляемость, наблюдаемость. – М.: Наука. – 1979. – 336 с.
6. *Воронов А.А.* Введение в динамику сложных управляемых систем. – М.: Наука. – 1985. – 352 с.
7. *Гантмахер Ф.Р.* Теория матриц. – М.: Наука. – 1967. – 575 с.
8. *Красовский Н.Н.* Теория управления движением. – М.: Наука. – 1968. – 476 с.
9. *Красносельский М.А.* Положительные решения операторных уравнений. – М.: Физматгиз. – 1962.
10. *Красносельский М.А.* Оператор сдвига по траекториям дифференциальных уравнений. – М.: Наука. – 1966. – 331 с.
11. *Красносельский М.А., Вайникко Г.М., Забрейко П.П., Рутицкий Я.Б., Стеценко В.Я.* Приближенное решение операторных уравнений. – М.: Наука. – 1969. – 445 с.
12. *Крейн С.Г.* Функциональный анализ. – М.: Наука. – 1972. – 544 с.
13. *Лефшец С.* Устойчивость нелинейных систем автоматического управления. – М.: Мир. – 1967. – 183 с.
14. *Ли Э., Маркус Л.* Основы теории оптимального управления. – М.: Наука. – 1972. – 574 с.
15. *Малек И., Фенцль З.* Непрерывное культивирование микроорганизмов. – М.: Пищевая промышленность. – 1968. – 564 с.
16. *Немыцкий В.В., Степанов В.В.* Качественная теория дифференциальных уравнений. – М., Л.: Гостехиздат. – 1947. – 448 с.
17. *Никайдо Х.* Выпуклые структуры и математическая экономика. – М.: Мир. – 1972. – 517 с.
18. *Опойцев В.И., Хуродзе Т.А.* Нелинейные операторы в пространствах с конусом. – Тбилиси: Изд-во ТГУ. – 1984. – 269 с.
19. *Перт С.Дж.* Основы культивирования микроорганизмов и клеток. – М.: Мир. – 1978. – 331 с.
20. *Печуркин Н.С., Терсков И.А.* Анализ кинетики роста и эволюции микробных популяций. – Новосибирск: Наука. – 1975. – 275 с.
21. *Руш Н., Абетс П., Лалуа М.* Прямой метод Ляпунова в теории устойчивости. – М.: Мир. – 1980. – 300 с.
22. *Терсков И.А., Гительзон И.И.* Применение плотностатного процесса управления для управляемого культивирования микроорганизмов // Непрерывное и управляемое культивирование микроорганизмов. – М.: Наука. – 1967. – С. 3–13.
23. *Уткин В.И.* Скользящие режимы в задачах оптимизации и управления. – М.: Наука. – 1981. – 367 с.
24. *Филиппов А.Ф.* Дифференциальные уравнения с разрывной правой частью. – М.: Наука. – 1985. – 224 с.
25. *Шефер Х.* Топологические векторные пространства. – М.: Мир. – 1971. – 359 с.
26. *Brammer R.F.* Controllability in Linear Autonomous Systems with Positive Controller// *SIAM Journal on Control.* – 1972. – V. 10, № 2. – P. 339–353.
27. *Kelly A.* The Stable, Center-Stable, Center, Center-Unstable manifolds//*Journal of Differential Equations.* – 1967. – V. 3, № 4. – P. 540–570.
28. *Martin R.H., Sachs E.* Positive Control for a Quasimonotone Systems of Differential Equations//*Mathematical Analysis and Applications.* – 1981. – V. 84, № 2. – P. 584–595.
29. *Ohta Y., Maeda H., Kodama S.* Reachability, Observability and Realizability of Continuous Time Positive Systems//*SIAM Journal on Control and Optimization.* – 1984. – V. 22, № 2. – P. 171–180.
30. *Rolewicz S.* On Optimal Observability of Lipschitz Systems//*Lecture Notes in Economics and Mathematical Systems.* – 1984. – № 226. – P. 152–158.

31. *Saperstone S.H., Yorke J.A.* Controllability of Linear Oscillatory Systems Using Positive Control//SIAM Journal on Control. – 1971. – V. 9, № 2. – P. 253–262.
32. *Victory H.D.* On nonnegative solutions of matrix equations//SIAM Journal on Algebraic and Discrete methods. – 1985. – V. 6, № 3. – P. 406–412.
33. *Walter G.G.* Eigenvalues and Structure of Compartmental models//Mathematical Biosciences. – 1984. – V. 71. – P. 181–199.
34. *Wonham W.M.* Linear Multivariable Control. A Geometric Approach//Lecture Notes in Economics and Mathematical Systems. – 1974. – № 101. – P. 344.

К главе 4

1. *Бондаренко Н.Ф. и др.* Моделирование продуктивности агроэкосистемы. – М.: Гидрометеиздат. – 1982. – 262 с.
2. *Глобус А.М.* Экспериментальная гидрофизика почв. Методы определения потенциала и коэффициента переноса почвенной влаги. – Л.: Гидрометеиздат. – 1969. – 335 с.
3. Динамическая теория биологических популяций / Под ред. Р.А. Полуэктова. – М.: Наука. – 1974.
4. *Ермаков С.М.* (ред.) Математическая теория планирования эксперимента. – М.: Наука. – 1983. – 390 с.
5. *Заславский Б.Г., Бриш А.Б.* Модельное исследование колебательных и устойчивых режимов функционирования устьичного аппарата растений // Доклады ВАСХНИЛ. – 1981. – № 3. – С. 26–28.
6. *Заславский Б.Г., Нагиев А.Т.* Моделирование движения влаги и солей в почвенной толще хлопкового поля // Доклады АН АзССР. – 1982. – № 10. – С. 13–17.
7. *Клейнен Дж.* Статистические методы в имитационном моделировании. – М.: Статистика. – 1978. – Вып. 1. – 220 с; Вып. 2. – 333 с.
8. *Кутиков Л.М. и др.* Система программ для статистической обработки информации "САФИСТ" // Экономика и мат. методы. – 1975. – Т. I, вып. 1.
9. *Менжулин Г.В.* К методике расчета метеорологического режима в растительном сообществе // Метеор. и гидрол. – 1970. – № 2. – С. 92–99.
10. *Моисеев Н.Н.* Математика ставит эксперимент. – М.: Наука. – 1979. – 223 с.
11. *Моисеев Н.Н.* Математические задачи системного анализа. – М.: Наука. – 1981. – 487 с.
12. *Нерлин С.В., Саноян М.Г., Аракелян А.Н.* О способе учета поглощения воды корнями воды при моделировании влагообмена на сельскохозяйственном поле // Докл. ВАСХНИЛ. – 1976. – № 9. – С. 40–43.
13. *Полуэктов Р.А., Жуковский Е.Е.* Имитационное моделирование агроэкосистемы и его информационная база // Проблемы экологического мониторинга и моделирования экосистем. – Л.: Гидрометеиздат. Т. III. – 1980.
14. *Полуэктов Р.А., Смысловский А.И., Финтушал С.М.* "СИМОНА" – система имитационного моделирования и язык управления моделями // Теоретические основы и количественные методы программирования урожаев. – Л.: Тр. АФИ. – 1979. – С. 101–112.
15. *Полуэктов Р.А., Шестиперов А.А.* Автоматизация имитационного эксперимента при моделировании динамических систем // Изв. ЛЭТИ им. В.И. Ульянова (Ленина). – 1981. – Вып. 293. – С. 26–30.
16. *Полуэктов Р.А.* Имитационная модель продуктивности агроэкосистемы / Tag. Ver. Akad. Landwirtsch. – Wiss. – Berlin. – 1981. – V. 188. – S. 19–27.
17. *Полуэктов Р.А. и др.* Базовая имитационная модель продуктивности пшеницы // Wiss. Zeitschrift der Humboldt-Universität zu Berlin. Math.-Nat. R. XXXIII. – 1984. – S. 339–342.
18. *Полуэктов Р.А. и др.* Имитационная модель развития агроценоза (препринт). – М.: ВНИИСИ. – 1984. – 83 с.
19. *Пых Ю.А.* Математические модели фотосинтеза листа в общей модели продуктивности агроэкосистем // Сб. трудов по агроном. физике. – 1976. – Вып. 38. – С. 30–46.
20. *Пых Ю.А.* Подмодель фотосинтеза и фотодыхания растений // Теоретические основы и количественные методы программирования урожаев. – 1976. – С. 39–50.

21. Росс Ю.К. Радиационный режим и архитектоника растительного покрова. – Л.: Гидрометеиздат. – 1975. – 342 с.
22. Сиротенко О.Д., Бойко А.П. Динамическая модель агроценоза // Тр. ИЭМ.– 1976. – Вып. 8 (67). – С. 12–36.
23. Тооминг Х.Г. Солнечная радиация и формирование урожая. – Л.: Гидрометеоздат. – 1977. – 199 с.
24. Успенский А.Б. Обратные задачи математической физики – анализ и планирование эксперимента // Математические методы планирования экспериментов. – Новосибирск: Наука. – 1981. – С. 199–242.
25. Федоров В.В. Теория оптимального эксперимента. – М.: Наука. – 1971. – 311 с.
26. Шеннон Р. Имитационное моделирование систем – искусство и наука. – М.: Мир. – 1978. – 381 с.
27. Шеффе Г. Дисперсионный анализ. – М.: Физматгиз. – 1963. – 625 с.
28. Goudriaan J. Crop meteorology: a simulation Study. – Wageningen: Pudoc. – 1977. – 246 p.
29. Hall A.E., Björkman O. Model of leaves photosynthesis and respiration // Perspectives of biophysical ecology. – 1975. – № 12. – P. 55–72.
30. Holling C.S., Dantzig G.B., Baskerville C., Jones D.D., Clark W.C. A case study of forest ecosystem pest management // Luxembourg: IIASA. – 1975. – 46 p.
31. Penning de Vries F.W.T., Brunsting A.H.M., van Laar H.H. Products requirements and efficiency of biosynthesis // Journ. Theor. Biol. – 1974. – V. 7. – P. 377–379.
32. Penning de Vries F.W.T., van Laar H.H. (eds.). Simulation of plant growth and crop production. – Wageningen: Centre for agricultural publications. – 1982. – 306 p.
33. Van Dyne G.M. Grassland management, research and training viewed in a system context // Range Sci. Dep. Sci. Ser. № 3. – Fort Collins: Colorado State Univ. – 1969. – 50 p.
34. Van Dyne G.M. (ed.). The ecosystem concept in natural resource management. – New York: Academic Press. – 1969. – 383 p.

К главе 5

1. Адлер Ю.П., Маркова Е.Е., Грановский Ю.В. Планирование эксперимента при поиске оптимальных условий. – М.: Наука. – 1976. – 280 с.
2. Беркович Р.П., Корявов П.П., Павловский Ю.Н., Сушков Б.Г. DYNAMO – язык математического моделирования (формальное описание). – М.: ВЦ АН СССР. – 1972. – 37 с.
3. Бродский В.З. Введение в факторное планирование эксперимента. – М.: Наука. – 1979. – 223 с.
4. Будико М.И. Климат и жизнь. – Л.: Гидрометеоздат. – 1972. – 472 с.
5. Горбунов-Пасадов М.М. и др. Пакет САФРА. Программное обеспечение вычислительного эксперимента // Пакеты прикладных программ. – М.: Наука. – 1983. – С. 12–50.
6. Денинг В., Эссинг Т., Маас С. Диалоговые системы человек – ЭВМ. Адаптация к требованиям пользователя. – М.: Мир. – 1984. – 110 с.
7. Ершов А.П., Ильин В.П. Пакеты программ как методология решения задач // Пакеты прикладных программ. – М.: Наука. – 1982. – С. 4–18.
8. Жуковский Е.Е., Нерпин С.В., Полуэктов Р.А. Модели продуктивности растительного покрова и управление формированием урожая // Принципы управления продукционными процессами в агроэкосистемах. – М.: Наука. – 1976. – С. 86–96.
9. Жуковский Е.Е., Нерпин С.В., Полуэктов Р.А. Агромониторинг и принципы его реализации // Теоретические основы и количественные методы программирования урожая. – Л. – 1979. – С. 3–13 (Труды Агрофизического НИИ).
10. Израэль Ю.А. Глобальная система // Метеорология и гидрология. – 1974. – № 7. – С. 3–8.
11. Израэль Ю.А. и др. Экологический мониторинг и регулирование состояния окружающей среды // Проблемы экологического мониторинга и моделирования систем, т. IV. – Л.: Гидрометеоздат. – 1981. – С. 6–19.
12. Мартин Дж. Организация баз данных в вычислительных системах. – М.: – 1980. – 662 с.

13. Марчук Г.И. Методы вычислительной математики. – М.: Наука. – 1980. – 450 с.
14. Налимов В.В. Теория эксперимента. – М.: Наука. – 1971. – 280 с.
15. Налимов В.В., Голикова Г.И. Логические основания планирования эксперимента. – М.: Металлургия. – 1976. – 128 с.
16. Нейлор Т и др. Машинные имитационные эксперименты с моделями экономических систем. – М.: Мир. – 1975. – 500 с.
17. Полуэктов Р.А., Заславский Б.Г. Оперативное управление производственным процессом // Моделирование и управление процессами в агроэкосистемах. – Л.: Агрофизический НИИ. – 1985 (Труды АФИ).
18. Полуэктов Р.А., Романцев В.В., Козлов О.С. Моделирование систем и автоматизация машинного эксперимента. – Л.: ЛЭТИ им. В.И. Ульянова (Ленина). – 1982. – 63 с.
19. Полуэктов Р.А., Смысловский А.И., Финушал С.М. "СИМОНА" – система имитационного моделирования и язык управления моделями // Теоретические основы и количественные методы программирования урожая. – Л. – 1979. – С. 105–112 (Труды АФИ).
20. Полуэктов Р.А., Шестиперов А.А. Автоматизация имитационного эксперимента при моделировании динамики агроэкосистем // Научно-техн. бюллетень по агрофизике. – Л. – 1980. – № 42. – С. 56–58.
21. Ракитский Ю.В., Устинов С.М., Черноруцкий И.Т. Численные методы решения жестких систем дифференциальных уравнений. – Л.: ЛПИ им. М.И. Калинина. – 1977. – 82 с.
22. Самарский А.А. Введение в теорию разностных схем. – М.: Наука. – 1971. – 552 с.
23. Самарский А.А. Пакеты прикладных программ как средство обеспечения сложных физических расчетов // Перспективы системного и теоретического программирования. – Новосибирск: СО АН СССР. – 1979. – С. 5–14.
24. Свиревев Ю.М., Тарко А.М. Опыт математического анализа глобального экологического мониторинга // Теоретические основы и опыт экологического мониторинга / Под ред. В.Е. Соколова и Н.И. Базилевич. – М.: Наука. – 1983. – С. 138–145.
25. Шеннон Р. Имитационное моделирование систем – искусство и наука. – М.: Мир. – 1978. – 418 с.
26. Шестиперов А.А. Планирование зоотехнического эксперимента // Бюллетень ВНИИ генетики и разведения сельскохозяйственных животных, вып. 58. – Л. – 1982. – С. 13–19.
27. Шестиперов А.А. Последовательное планирование зоотехнических экспериментов // Бюллетень ВНИИ генетики и разведения сельскохозяйственных животных. – Л. – 1982. – Вып. 59. – С. 41–44.
28. Щербаков Р.А., Пачепский Я.А., Кузнецов М.Я. Водная миграция ионов и химических соединений в почвах. Движение влаги (материалы по математическому обеспечению ЭВМ). – Пушкино: Научный центр биологических исследований АН СССР. – 1981. – 44 с.
29. Форрестер Дж. Динамика развития города. – М.: Прогресс. – 1974. – 285 с.
30. Blunden G.P., Krasnow H.S. The process concept as a basis for simulation modeling // Simulation. – 1967. – № 9. – P. 89–94.
31. Vox G.E.P., Cox D.R. An analysis of transformations // J. Roy Statistical Soc. Ser. B. – 1964. – V. 26.
32. Vox G.E.P., Wilson K.B. On the experimental attainment of optimum conditions // J. Roy. Statistical Soc. Ser. B. – 1951. – V. 13, № 1.
33. Brennan R.D. Continuous system modeling programs: State-of-the-art and prospectus for development. – Oslo: IFIP Working Conference on Simulation Languages. – 1967.
34. Buxton J.N., Laski J.G. Control and simulation language // Computing J. – 1962. – V. 5. – P. 373–380.
35. Connor W.S., Young S. Fractional factorial designs of experiments with factors at two and three levels. Washington: National Bureau of Standards, Applied Mathematics Series: U.S. Government Printing Offices. D.C. – 1961. – № 58.
36. Dahl O.J., Nygard K. SIMULA – an ALGOL-based simulation language // New York. Commun. ACM. – 1966. – V. 9. – P. 671.
37. Fatolari E. FORSIM IV user's guide, SR-99. – The Mitre Corp. – New York. – 1964.

38. *Fishman G.S., Kiviat P.J.* The analysis of simulation generated time series // *Management Sci.* – New York. – 1971. – V. 18. – P. 527–557.
39. *Hill W.J., Hunter W.G.* A review of response surface methodology: A literature survey // *Technometrics.* – 1966. – V. 8, № 4.
40. *Kiviat P.J.* Development of discrete digital simulation languages // *Simulation.* – 1967. – V. 8, № 2. – P. 17–23.
41. *Kiviat P.J.* GASP – a general activity simulation program. Project № 90. 17-019(2). Manroeville. Pennsylvania: Appl. Res. Lab. U.S. Steel Corp. – 1963. – 283 p.
42. *Pritsker A.A.B., Kiviat P.J.* Simulation with GASP II: a FORTRAN-based simulation language. – New York: Englewood Cliffs. – 1969.
43. *Rechtschaffner R.L.* Saturated fractions of 2 and 3 factorial designs // *Technometrics.* – 1967. – V. 9. – P. 569–575.

СПИСОК ОСНОВНЫХ ОБОЗНАЧЕНИЙ

$R_+^m = \{x \in R^m \mid (X)_i \geq 0\}$ – положительный конус евклидова пространства

x_i – $(X)_i$ -я компонента вектора евклидова пространства

$X \geq Y$ обозначает $\{(X)_i \geq (Y)_i \mid i = 1, \dots, m\}$, $X, Y \in R^m$

$X \geq \neq Y$ обозначает $X \geq Y$ и $X \neq Y$

$X > Y$ обозначает $\{(X)_i > (Y)_i \mid i = 1, \dots, m\}$, $X, Y \in R^m$

$[X, Z] = \{Y \mid X < Y < Z\}$ – порядковый интервал упорядоченного векторного пространства

$F: R_+^m \rightarrow R^p$ ($F: R_+^m \rightarrow R_+^p$) – оператор (соответственно неотрицательный оператор $F \geq 0$)

A – внутренность множества

\bar{A} – замыкание множества

$F: \overset{\circ}{R}_+^m \rightarrow \overset{\circ}{R}_+^p$ – положительный оператор $F > 0$

B^* – транспонированная матрица B (сопряженный оператор)

$\det B$ – определитель матрицы B

$\langle X, Y \rangle$ – скалярное произведение в гильбертовом пространстве (значение функционала Y в точке X банахова пространства). Если $X, Y \in R^m$, то $\langle X, Y \rangle =$

$$= \sum_{i=1}^m (X)_i (Y)_i, \rho(X) = \sum_{i=1}^m x_i \text{ при } X = (x_i)_{i=1}^m \in R^m$$

I – единичная матрица

$F|_{E'}$ – ограничение оператора $F: E \rightarrow R$ на подмножество $E' \subset E$

$\text{Pr}_{E'}$ – проекция векторного пространства E на подпространство $E' \subset E$

id – тождественное отображение

\dot{x} – производная функции $x(t)$ по времени

$D_X F(X_0)$ – дифференциал (производная) Фреше функции F по X , вычисленный в точке X_0