Министерство образования и науки Российской Федерации

Владивостокский государственный университет
экономики и сервиса

В.И. ДУЛЕПОВ
О.А. ЛЕСКОВА

ЭКОСИСТЕМНЫЙ АНАЛИЗ

Учебное пособие

Владивосток
Издательство ВГУЭС
2006

ББК 65.01
Д 79

Дулепов В.И., Лескова О.А.
Д 79 ЭКОСИСТЕМНЫЙ АНАЛИЗ: Учеб. пособие. – Владивосток: Изд-во ВГУЭС, 2006. – 248 с.

Учебное пособие составлено в соответствии с учебной программой курса, а также требованиями образовательного стандарта России к учебной дисциплине «Экосистемный анализ». Является дополнением к уже изданному учебному пособию «Системная экология» (Владивосток: Изд-во ВГУЭС, 2004) и способствует более глубокому дополнительному и самостоятельному изучению предмета. Призвано значительно расширить представления студентов о методах моделирования и статистической обработки экологических данных и применении этих методов для анализа экосистем.

Предназначено студентам всех форм обучения.

ББК 65.01

Печатается по решению РИСО ВГУЭС.

ВВЕДЕНИЕ

В настоящее время экологические системы рассматриваются как структурные единицы («ячейки») биосферы, которая является единым целым, а ее части (экосистемы), обладая известной автономией, тесно взаимосвязаны друг с другом. Биогеопотоки, переносящие вещество и энергию из одних экосистем в другие, исключают возможность их изолированного существования и создают своеобразную непрерывность (континуум) всей биосферы. Экосистемы характеризуются определенным уровнем структурной и функциональной организации. Их структурированность определяется особенностями пространственного распределения взаимосвязанных между собой косных и живых компонентов и градиентностью термодинамических характеристик по горизонтали и вертикали на суше, в гидросфере и в атмосфере. Функциональная организация экосистем проявляется в согласованности процессов, обеспечивающих круговорот веществ, протекание биогеохимических циклов. В результате функционирования экосистем происходят непрерывная миграция атомов, осуществляющаяся в форме биогеохимических циклов, и новообразование органических веществ из минеральных, в основном за счет трансформации и аккумуляции в экосистемах солнечной энергии и в меньшей степени за счет хемосинтеза органических веществ микроорганизмами.

В процессе функционирования экосистем возникают предпосылки их преобразования, так как неполная нейтрализация воздействия на среду одних популяций другими ведет к изменению свойств биотопа и обусловливает адаптивную перестройку сообщества. Точно так же функционирование измененного сообщества оказывается причиной его дальнейшего изменения.

В свою очередь 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, т.е. она изучает вопросы о том, как сообщества живых организмов, используя органические, минеральные и энергетические ресурсы, создают новое органическое вещество, как это вещество распределяется между другими организмами в системе и каким образом оно в конце концов вновь распадается на свои минеральные компоненты. При этом необходимо понять, за счет каких механизмов создается и поддерживается удивительная стабильность природных комплексов, что обеспечивает сохранение и изменение их пространственно-временной структуры, какими факторами определяются их развитие и эволюция. Все многообразие этих проблем от изучения надорганизменного уровня функционирования биосистем до исследования структуры связей между организмами и их средой невозможно решить без привлечения методов экосистемного анализа.

Подход к изучению экосистем на системном уровне включает комплексное изучение всех ее элементов с качественной точки зрения, различные количественные методы изучения, такие как статистические и математические методы, метод моделирования. И здесь главная проблема, которая встанет перед исследователем, это получение достоверных количественных данных на основе исследований по всем живым и косным компонентам с учетом динамики их изменения в сезонных и межгодовых аспектах.

Глава 1
МЕТОДОЛОГИЧЕСКИЕ АСПЕКТЫ ЭКОСИСТЕМНОГО АНАЛИЗА

1.1. Предмет и задачи экосистемного анализа

Под экосистемным анализом мы понимаем исследование структуры и функциональных особенностей экологических систем с целью установления экологических закономерностей на экосистемном уровне и возможности прогнозирования их развития, а также динамики изменения основных компонентов экологических систем. Мерами структуры и функций экосистем являются число видов, число трофических уровней, численность и биомасса видов, скорости создания первичной и вторичной продукции, потоки энергии и круговорот питательных веществ и т.п. Они отражают экологические взаимодействия между популяциями, а также между особями и окружающей их физической средой. Способность сообществ, входящих в экосистемы, сопротивляется экологическим пертурбациям, обусловлена гомеостатическими механизмами особей и ростовыми реакциями популяций. Но помимо них обычно исследуются такие черты организации, которые способствуют сохранению стабильности сообществ и экосистем и эффективности их функций.

Итак, основными направлениями экосистемного анализа являются изучение структуры экосистем и их функционирования. Отсюда возникает многообразие возникающих в экосистемном анализе задач, которые можно сгруппировать по следующим направлениям исследований:

1. Изучение структуры экосистем при различных физико-химических условиях обитания биоты (в разных биотопах). Здесь рассматриваются широтные изменения структуры экосистем, пространственные по площади и вертикали, а также в зависимости от различных физико-химических свойств окружающей среды. Как правило, изучается видовой состав экосистем, относительное обилие видов по численности и биомассе, связь этих показателей с абиотическими факторами. Делается количественная оценка видовой структуры по трофическим уровням, и выделяются руководящие виды и группы.

2. Изучение функционирования экосистем проводится также при разных абиотических условиях и пространственном распределении. Функционирование изучается с точки зрения создания первичной и вторичной продукции, оценки ее на различных трофических уровнях. Исследуются схемы потоков вещества и энергии в экосистеме, круговорот питательных веществ, углерода, азота, воды и т.п. Исследуются экологические взаимодействия между популяциями и группами животных, динамики их численности и биомассы с учетом изменения абиотических и биотических (хищничество, конкуренция, комменсализм) факторов.

3. При изучении функционирования важное место уделяется оценке устойчивости экосистем, в которую можно вкладывать разный смысл. Например, устойчивой может называться экосистема, отличающаяся постоянством своего состояния, отсутствием изменений. Другой критерий – инерция, устойчивость к нарушениям, третий – «эластичность», способность возвращаться к исходному состоянию после нарушения внешними возмущающими силами. Или в качестве критерия можно выбрать амплитуду изменений определенных характеристик экосистемы в ответ на внешние воздействия, при которых система может возвращаться в исходное состояние.

4. Между живыми и косными компонентами экосистем происходит непрерывный обмен различными элементами, обуславливаемый взаимодействием биологических и геохимических процессов. Для их изучения исследуются биогеохимические циклы, где особое внимание уделяется круговороту азота, фосфора, серы, углерода, железа, марганца, а также исследуются процессы новообразования органического вещества как одного из фундаментальных свойств экосистем. Подавляющее большинство органики синтезируется из минеральных веществ в процессе фотосинтеза, т.е. за счет утилизации солнечной энергии. В неизмеримо меньшем количестве органическое вещество образуется хемосинтезирующими бактериями с использованием энергии экзотермических процессов окисления некоторых восстановленных соединений. Следует отметить, что вещества, окисляемые хемосинтетиками, в большинстве случаев представляют собой продукты анаэробного распада органических соединений, создаваемых фотосинтезирующими растениями. Таким образом, исследование функционирования экосистем в этом направлении связано с решением задач новообразования органического вещества.

1.2. Цели и методы исследования экосистем

Цели проведения экосистемного анализа для конкретных биосистем в той или иной степени связаны с возможностью прогнозирования динамики экосистем и понимания экологических процессов, происходящих на экосистемном уровне. Хотя с точки зрения биологического подхода экологический процесс это довольно широкое понятие (не конкретное), так как экологические процессы бывают разные. Если включаем в анализ функционирования экосистем действия человека, то это уже эколого-экономические процессы, которым тоже следует уделять внимание, так как человечество существует в биосфере Земли и тесно с ней связано. В последнее время у всех появилась озабоченность в связи с быстрым разрушением различных экологических систем. И, по сути, для решения этих проблем мы должны рассматривать и анализировать эколого-экономические системы.

Цели экосистемного анализа можно подразделить следующим образом:

1. Прогноз динамики развития экосистем.
2. Разработка рекомендаций рационального управления природными ресурсами.
3. Оценка возможности и планирования устойчивого развития общества с учетом возникающего загрязнения окружающей среды и истощения ресурсов.
4. Прогноз и предотвращение сукцессий экосистем.
5. Развитие теории экосистем и протекания в них различных экологических процессов.
6. Определение цикличности экологических процессов и характеристик экосистем.
7. Прогнозирование динамики изменения различных характеристик эколого-экономических систем.

Методы, которые применяются в экосистемном анализе, также подразделяются на несколько групп. Они, как правило, связаны с методологией системного анализа и исследованием сложных систем. К специфическим для экологии можно отнести экспериментальные методы анализа искусственных биосистем, создаваемых в аквариумах различных объемов, прудах для разведения рыб и беспозвоночных, хозяйствах аквакультуры. Здесь можно проводить экспериментальные исследования при различных биотических и абиотических факторах, ставить различные опыты и управлять продукцией биосистемы. К другой группе методов можно отнести метод моделирования, где для анализа динамики экосистемы строится математическая модель и затем она исследуется. Математические модели имеют свою классификацию, и здесь мы упомянем только то, что можно разрабатывать и исследовать модели теоретические, например описываемые с помощью дифференциальных уравнений, и получать качественное решение. Можно разрабатывать имитационные модели, в которых все блоки модели описываются с помощью различных функциональных зависимостей с конкретными коэффициентами или статистическими функциями распределения коэффициентов, а затем исследуется поведение модели при различных условиях. Особняком стоят экспертные методы, которые позволяют привлекать высококвалифицированных специалистов из различных областей знаний и в результате их опросов, качественных заключений о поведении биосистем, специальных методов анализа таких данных делать заключения о поведении и развитии биосистем. Эти методы имеют хорошую перспективу при анализе эколого-экономических систем и экосистем, где имеются лишь качественные данные, мало количественных наблюдений и требуются оценки экспертов.

При наличии количественных данных об эколого-экономической системе можно использовать оптимизационные методы, основанные на построении целевой функции с определенными ограничениями и поиском максимума или минимума целевого функционала. Эти методы полезны и при поиске оптимальных решений в рациональном природопользовании.

Наиболее часто используемыми методами анализа больших систем являются статистические методы, они обычно подразумевают наличие больших объемов разнообразных количественных данных по структуре и функциональным особенностям экосистемы, и их можно разбить на несколько групп:

- дескриптивные методы анализа данных по одной или нескольким переменным;
- линейный и нелинейный регрессионный анализ, простые и множественные регрессии;
- дисперсионный анализ для сравнения одной или нескольких переменных, выявления влияния различных факторов;
- многомерные методы: главные компоненты, факторный, канонический, дискриминантный и кластерный анализы.

В последнее время для целей прогнозирования и классификации применяется метод нейтронных сетей. В настоящее время нейтронные сети имеются в некоторых статистических пакетах и применение этого метода в экосистемном анализе имеет большую перспективу.

1.3. Роль моделирования в экосистемном анализе

Основные этапы экосистемного анализа представлены на рис. 1.1, верхняя часть которого отражает деятельность специалиста, занимающегося тем или иным классом экосистем (Страшкраба, Гнаук, 1989). Проводимые им исследования реальной действительности благодаря индуктивной (как правило, статистической) и дедуктивной (теоретической) абстракции приводят к идеализации системы, определяемой как вербальная модель. Конкретная форма модели обычно обусловлена результатом объединения теоретических концепций с экспериментальными (эмпирическими) данными. При этом любая вербальная модель базируется на более или менее развитой теории моделируемой системы.

Вербальная модель формализуется посредством математических методов (нижняя половина рис. 1.1), в результате чего мы получаем математическую модель. Модель обеспечивает обозримость объекта путем сокращения его избыточности и, следовательно, является звеном между теоретическим и эмпирическим познанием. Модели используются для замены реальных процессов и систем, однако их поведение должно быть сравнимо с поведением реальных объектов. Вычислительный алгоритм модели, реализуемый на ЭВМ, делает возможной имитацию явления с помощью модели. Полученные в результате имитации данные можно использовать для сравнения с результатами наблюдений, а также с имеющейся информацией о поведении реальных систем.

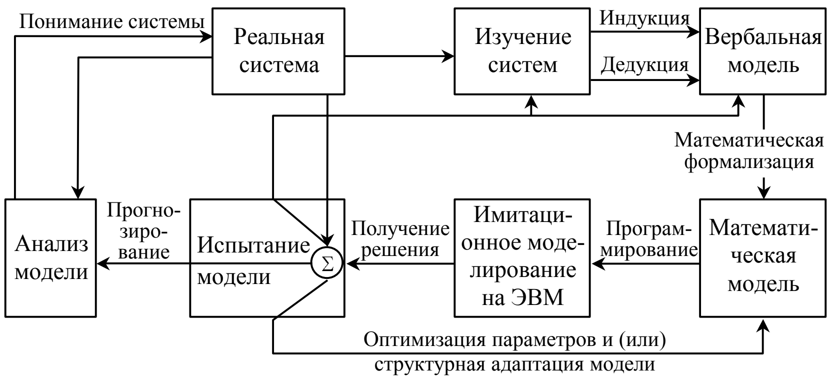


Рис. 1.1. Основные этапы экосистемного анализа

Отношения между основными переменными состояния данной модели обычно не согласуются полностью с отношениями между переменными состояния данной системы. Таким образом, действительность искажается (погрешность оценки отношений). Сравнение модели и объекта с целью оценки погрешности дает возможность улучшить согласование модели с реальной системой. Эта процедура может выполняться многократно. По завершении испытания математической модели ее можно применять для изучения других состояний этой же системы и даже иных систем. Такой подход получил название прогнозирования. Использование моделей с целью увеличения теоретических и экспериментальных знаний называется анализом модели. Модель применяется для проведения имитационных экспериментов, которые позволяют находить ответы на заданные вопросы (например реакция системы на различные внешние эффекты).

В основе системного анализа лежит абстрагирование, позволяющее выявить главные свойства математической модели. Оно означает предельное упрощение реальности. Модель никогда не идентична реальности, а скорее является продуктом нашего собственного воображения. Левинс (Lewins, 1966) предложил классификацию моделей, в которой принимаются во внимание лишь два из трех основных свойств модели: общность, реальность и точность. В первый класс входят модели, в которых общность приносится в жертву другим свойствам – реальности и точности. Второй класс составляют модели, являющиеся реалистичными и имеющими общий характер, однако менее точные. В третьем классе моделей приоритет принадлежит таким свойствам, как общность и точность.

Создание и практическое использование моделей являются главными проблемами в системном анализе, где необходимо уделять внимание следующему: модель строится на основе гибкой гипотезы с тем, чтобы информацию можно было извлекать даже из ошибок. Модели, представляющие собой искусственные построения, а не проверяемые гипотезы, не способствуют поступательному процессу познания. С другой стороны, модель, основанная на гибкой гипотезе, способствует поступательному движению, даже если в данный момент времени не имеется математического решения. Такой подход получил название гипотетико-дедуктивной методологии.

Процесс получения новых знаний делится на следующие этапы:

1. Создание предварительной модели на основе данных наблюдений и выводов исследователя, а также оценки соответствующей литературы.
2. Преобразование предварительной модели в формализованную гипотезу.
3. Получение из гипотезы проверяемых выводов.
4. Правильность выводов следует проверять с помощью данных, не участвующих в формировании гипотезы.
5. Если выводы оказываются правильными, то познавательный процесс может быть продолжен, начиная с третьего этапа, с тем, чтобы получить очередной вывод. Если это не удается, следует начать процедуру заново с первого этапа.

Гипотетико-дедуктивная методология отличается от описательной научной методологии тем, что она допускает существование неточностей и риск ошибочной интерпретации результатов. Любая модель, позволяющая лучше понять суть проблемы, играет не менее важную роль, чем факты и наблюдения.

Контрольные вопросы

1. Основные направления экосистемного анализа.
2. Задачи экосистемного анализа.
3. Цели исследования экосистем.
4. Методы экосистемного анализа.
5. Этапы экосистемного анализа.

Глава 2
ЭКОСИСТЕМЫ И ИХ ХАРАКТЕРИСТИКИ

2.1. Системы и виды систем, их иерархическая структура

Система – это часть объективной реальности, ограниченная рядом условий. Она состоит из более мелких единиц – элементов, связанных между собой различными отношениями (связями). Все, что находится за пределами системы, называется окружающей средой. Система и окружающая среда разделены оболочкой – границей системы (рис. 2.1). Например, все популяции в озере, связанные друг с другом, а также с «сопутствующими» факторами (например свет, кислород), составляют одну экосистему, в то время как суша будет являться для этой системы окружающей средой. Граница системы проходит по береговой линии и водной поверхности.

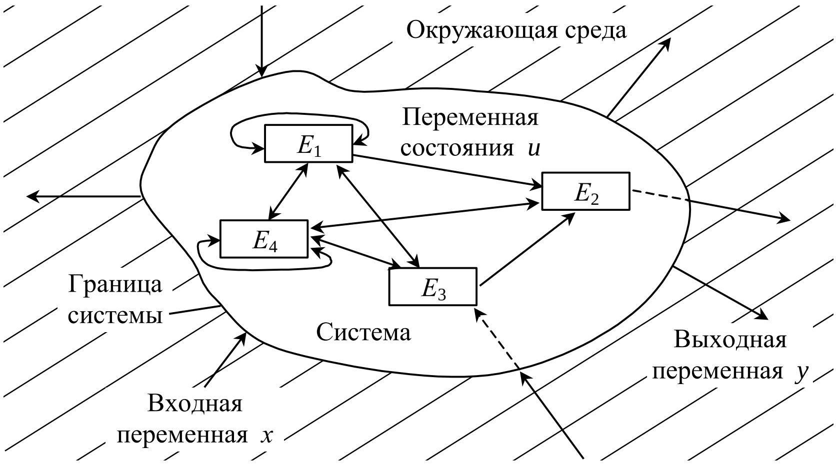


Рис. 2.1. Выделение (кибернетической) системы из окружающей среды: E_i – элемент системы

Определение границ системы носит условный характер и строится на ряде предположений. Один и тот же элемент может быть частью системы или окружающей среды в зависимости от того, в каком аспекте он рассматривается. Однако это не значит, что мы не располагаем критериями, которые позволяют относить элементы к системе или окружающей ее среде. Поскольку связи между элементами системы представляют собой важную характеристику, граница системы обычно проводится таким образом, чтобы большая часть взаимосвязей, в особенности обратных связей, находилась внутри системы и лишь некоторые из них выходили за ее пределы. Можно выделить два вида наиболее важных связей: прямую и обратную (рис. 2.2).

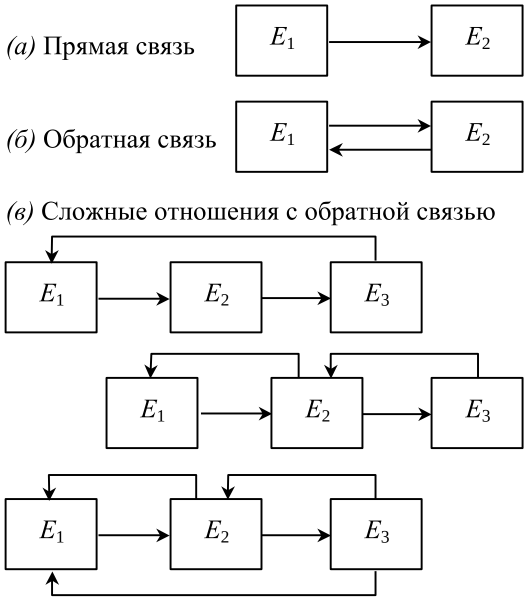


Рис. 2.2. Отношения (связи) между элементами системы:
а) прямая связь: элемент E_1 воздействует на элемент E_2, но не испытывает на себе влияния E_2;
б) обратная связь: E_1 действует на E_2 и одновременно испытывает обратное влияние E_2;
в) примеры сложных структур, характеризующихся наличием обратной связи между несколькими элементами

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.

Рассмотрим структуру экосистем более подробно. На рис. 2.3 приведена упрощенная схема взаимодействия различных элементов в природной экосистеме. С термодинамической точки зрения любая экосистема является открытой: для своего функционирования она нуждается в постоянном снабжении энергией и веществом. На рис. 2.3 источники энергии и вещества представлены в виде совокупности материальных и энергетических ресурсов экосистемы. Для естественных биогеоценозов эти ресурсы находятся под воздействием климатических управляющих факторов и часто – под воздействием человека.

Биологическая часть экосистемы с точки зрения трофических отношений делится на два компонента:

1) автотрофный компонент, к которому относятся фото- и хемосинтезирующие организмы (продуценты);

2) гетеротрофный компонент, 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.

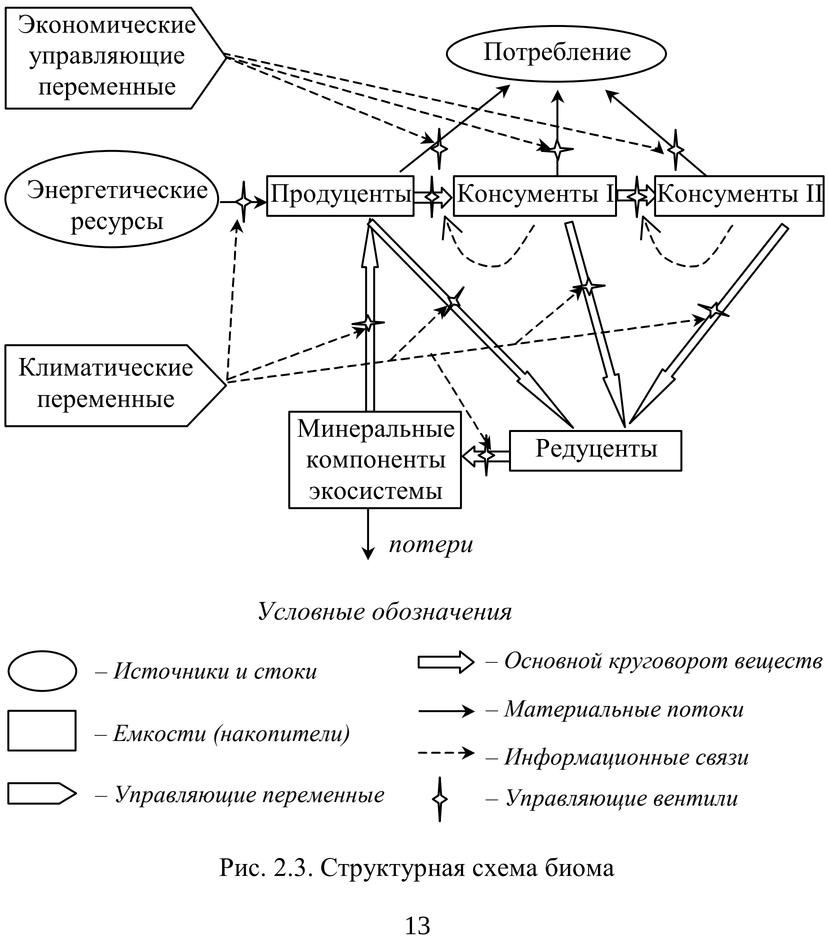


Рис. 2.3. Структурная схема биома

На автотрофном уровне происходит фиксация световой энергии и превращение простых неорганических веществ в сложные неорганические соединения. К гетеротрофной части биоценоза относятся макроконсументы – 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 и высвобождают неорганические питательные вещества, которые могут быть вновь использованы продуцентами. Таким образом, замыкается круговорот веществ в биосфере.

Природные экосистемы в значительной степени находятся под воздействием климатических и погодных факторов. Основными климатическими факторами являются суммарная солнечная радиация, температурный режим, 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. В совокупности с населяющими их животными, грибами и бактериями они составляют различные биомы, различающиеся по структуре и видовому составу, по характеру и скоростям процессов создания и утилизации органического вещества.

Все вышеизложенное позволяет отнести экосистемы к разряду систем, именуемых крупномасштабными или сложными, имеющими следующие характеристики.

1. Сложность: количество и разнообразие видов связей между элементами системы, а также между системой и окружающей средой, очень велико.

2. Целостность: система имеет свойства, которые становятся явными только в результате взаимодействия ее отдельных элементов.

3. Многомерная устойчивость: нелинейные и нестационарные системы могут иметь несколько неустойчивых областей, число которых определяется количеством бифуркационных точек системы.

4. Управляемость: система может переходить из одного состояния в другое в течение определенного промежутка времени. Система называется управляемой, если на нее можно оказывать целенаправленное воздействие.

5. Наблюдаемость: информацию о предыдущем состоянии системы можно получить исходя из ее нынешнего состояния.

6. Буферность и способность к сохранению: переход от одного состояния к другому в результате вмешательства каких-то факторов не носит взрывной характер, а характеризуется постепенным развитием.

7. Обработка информации и ее хранение: экосистемы могут преобразовывать получаемую информацию в соответствии со своей специфи-

кой. Они также способны соединять эту информацию с другими данными, хранящимися в самой системе, с тем, чтобы выдать новую информацию.

8. Качественные различия между элементами системы.

9. Структуры в экосистемах характеризуются физическими условиями (например делением пространства), световыми и энергетическими условиями, химическими условиями, которые складываются из количественного распределения органических и неорганических веществ, биологическими условиями (например трофические уровни, экологический спектр и разнообразие видов) и временной структурой (например биологической сукцессией, эволюцией системы).

10. Функции экосистемы обусловлены характером циркуляции вещества и энергии, взаимодействием элементов, входящих в систему, а также взаимосвязью между экосистемой и окружающей средой. Между взаимодействующими популяциями на различных трофических уровнях имеются особые экологические функции (например соперничество, отношения хищник-жертва и др.).

В целом все системы имеют следующие переменные (рис. 2.4).

Рис. 2.4. Характеристики системы с помощью входных, выходных и возмущающих переменных, а также переменных состояния

Входные переменные x_i, через которые осуществляется воздействие окружающей среды на систему (например биологические, химические, гидрометеорологические условия, географическое местоположение экосистемы и т.д.). Выходные переменные y_i, характеризующие реакцию системы на воздействие окружающей среды. Переменные состояния u_i, характеризующие способность системы к сохранению и осуществлению обратной связи, а также ее реакцию на влияние случайных факторов. В экосистемах на эти переменные обычно накладываются возмущающие переменные d_i.

Действие этих переменных в пространстве и времени обычно описывается при помощи детерминированных или стохастических функций в зависимости от того, можно ли связать определенное значение или вероятностное распределение возможных значений зависимых переменных с каждым значением независимых переменных.

Обычно системы считаются изолированными, открытыми или замкнутыми. В замкнутых системах не происходит обмена энергией, веществом и информацией с окружающей средой, и в большинстве случаев они представляют собой идеализацию реальной действительности. Имея дело с естественными системами, в лучшем случае надо говорить о замкнутости лишь в отношении вещества и энергии. Например, пресноводные экосистемы с точки зрения обмена веществ могут быть замкнутыми (озера, пруды с незначительным притоком и оттоком воды) или открытыми (запруженные водоемы с высоким уровнем притока и оттока воды, потоки).

2.2. Кибернетическое описание экосистем и классификация моделей

2.2.1. Управление в экологических системах

Любой процесс управления заранее предполагает существование управляемых объектов и системы, управляющей этими объектами. В этом смысле управление означает целенаправленное воздействие управляющей системы на один или несколько управляемых элементов. Регулирование – это процесс управления с обратной связью (частный случай управления), в котором ведется непрерывное измерение управляемой переменной и ее сравнение с заданной переменной. Цель управления формализуется посредством целевой функции или целевого функционала, причем целью в первом случае служат фиксированные величины, а во втором – переменные. Выражение «экстремальное управление» используется в тех случаях, когда речь идет о минимуме или максимуме целевого функционала по управляющим переменным. Если же речь идет только об одной управляющей переменной, то получение минимума или максимума целевой функции определяется как оптимальное управление. При наличии нескольких управляющих переменных между различными целями управления может существовать лишь компромиссное решение. Все процессы управления как в технологических, так и в нетехнологических системах (а значит и в экосистемах) носят информативный характер.

Процесс управления предполагает наличие управляющего элемента и управляемой системы (управляемого элемента). Управляющий элемент производит сбор и обработку информации, получаемой из окружающей среды или из самой системы, и оказывает необходимое воздействие на управляемую систему. Свойства любой управляемой системы зависят от ряда факторов, таких как параметры системы, структура системы, возмущения, алгоритмы управления и т.д.

В экологических процессах и системах трудно провести линию раздела между управляющей и управляемой системами, поскольку, как правило, участвующие процессы ввиду их сложности физически неразделимы. И если иногда подобная дифференциация проводится, несмотря на существующие трудности, то она выполняется в виде умозрительного эксперимента, чтобы тем самым подкрепить наши собственные доводы и сделать возможным использование кибернетических методов, первоначально предназначавшихся для изучения систем с независимой физической материализацией управляющей и управляемой систем.

В экологических системах различают три группы управления (рис. 2.5) (Страшкраба, Гнаук, 1989):

1) естественное внутреннее управление (саморегуляция или самооптимизация);

2) естественное внешнее управление через возмущающие переменные окружающей среды;

3) внешнее управление, предполагающее воздействие человека на экосистему и охрану окружающей среды.

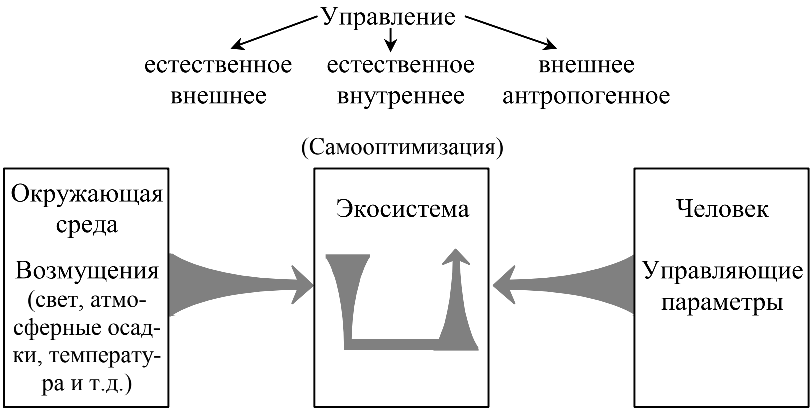


Рис. 2.5. Управление в экологических системах

Информационное содержание экосистемы тесно связано с управлением. Вся информация о процессах, проходящих в подсистемах, а также данные о качественном составе и состоянии отдельных частей системы и окружающей среды, служат информационными источниками для управления системой. Например, температура играет информативную роль для биологического блока модели, поскольку она дает информацию о важных характеристиках окружающей среды и управления интенсивностью различных процессов.

Паттен (1961) принимает, что потоки энергии и массы в экосистемах служат для переноса информации. Меншуткин (1969, 1971) определяет информацию как совокупность механизмов с обратной связью, где каждый отдельный перенос энергии и массы вызывает встречный поток информации.

Концепция саморегуляции, целевой функции, управляющего элемента и управляемой системы в экосистеме представлена на рис. 2.6. Целевая функция Q характеризует генетически фиксированную способность организмов к выживанию. Управляющий элемент является «мозгом», действующим на двух уровнях. Информация обрабатывается при помощи возмущающих переменных и подсистем, сопоставляется с целевой функцией и затем передается в виде сигнала w. Этот сигнал определяет, какая активность микроорганизмов является оптимальной для данного отрезка времени. Данный уровень получил название оптимизатора – термин, заимствованный из технической кибернетики. Управляемая система располагает рядом возможностей для реализации сигнала w. Одна из них связана с выбором соответствующих значений параметров. Этот уровень определяется как управляющий элемент (контроллер), так как он оказывает влияние на управляемую систему сигналом w*.

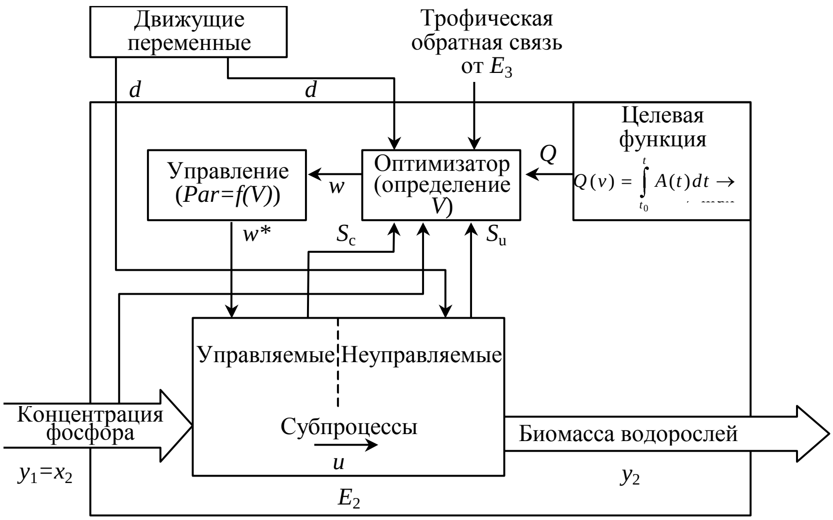


Рис. 2.6. Кибернетическая идеализация экологической системы на примере фитопланктонной подсистемы

2.2.2. Иерархическая структура экосистем

Теоретические принципы иерархических структур разработаны Месаровичем и др. (1970) при проведении экономических исследований. Биологическая значимость описания иерархических систем была установлена им двумя годами ранее. Он пришел к выводу, что введение понятия «иерархия» в биологии будет играть такую же роль, как введение концепции обратной связи. Это в особенности относится к системам «хищник-жертва», где существование иерархической структуры давно считается общепризнанным фактом. Иерархическую организацию системы можно рассматривать под тремя различными углами зрения:

1. Иерархия биологической организации считается классической характеристикой биологических систем: различные уровни организации (экосистема, популяция, особь, надклеточный уровень физиологических процессов, клеточный и подклеточный уровни) являются типичными примерами организационной иерархии. Координация в этом случае направлена от верхнего организационного слоя к параллельным единицам нижележащего слоя. Например, физиологические процессы, происходящие в особи, координированы друг с другом.

2. Трофическая иерархия – классический экологический термин, который по смыслу согласуется с термином «эшелон», введенным Месаровичем применительно к общим системам: на рис. 2.7-а показана схема многоуровневой системы. Информационные связи в ней действуют в обоих направлениях. Информация поступает из нижних эшелонов в верхние в ходе процесса. И именно сверху осуществляется координация параллельных единиц на нижнем уровне.

3. Иерархия управления. На рис. 2.7-б процессы прямого управления, адаптации и выбора разделены. В экологии прямое управление означает непосредственное, всегда фиксированное воздействие переменной на процесс (например, воздействие экзогенных возмущающих переменных в соответствии с фиксированным уравнением). Адаптация – это такое явление, при котором действующие параметры меняются в соответствии с поступающей информацией. Под выбором понимается отбор нескольких возможных путей реализации (например, из-за присутствия в системе нескольких различных видов) путем изменения структуры системы посредством включения или выключения отдельных связей. Этот процесс получил название самоорганизации системы. В тех случаях, когда генетический код настолько сильно меняется, что это служит началом биологической эволюции, применяется термин «самоэволюция». Целевая функция, или стратегия системы и ее элементов, также подвергается изменению в ходе этого процесса.

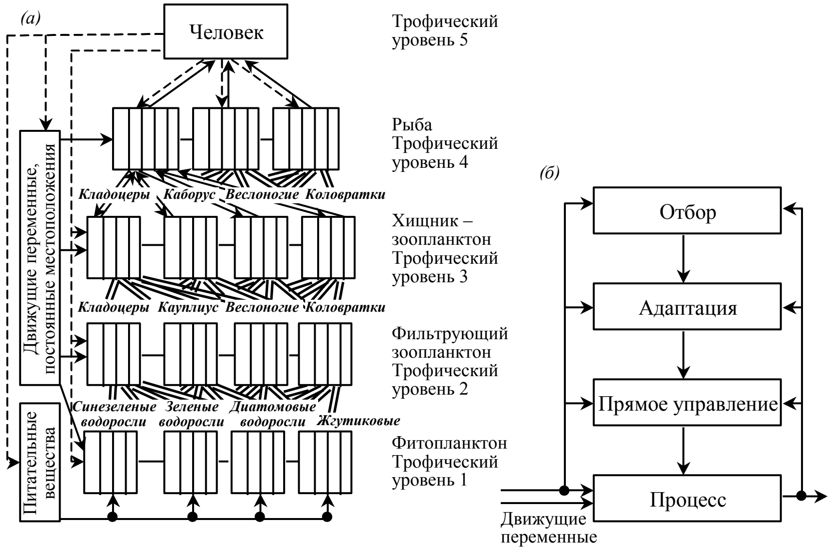


Рис. 2.7. Иерархическая структура экосистем:
а) трофическая иерархия. На каждом трофическом уровне (эшелоне) имеются различные группы организмов. Прямоугольники означают виды, а прерывистые линии – возможности антропогенного управления;
б) иерархия управления

2.2.3. Кибернетическая классификация моделей экосистем

В прикладной кибернетике основой для классификации систем служат различные виды управления. Это связано с разнообразными техническими реализациями управления. При рассмотрении экосистем природа самих систем остается неизменной, в то время как наши представления и модели меняются. Тем не менее подобная классификация применима к этим моделям в соответствии с теми видами управления, которым мы отдаем предпочтение. На рис. 2.8 показано разделение между классами тех моделей экосистем, которые характеризуются самым высоким типом управления. Иерархические модели вводятся здесь как новый класс моделей.

В процессе развития экологии пресноводных водоемов модели различных процессов или подсистем претерпели многократные изменения. К примеру, влияние зоопланктона на фитопланктон первоначально сводилось лишь к выеданию фитопланктона зоопланктоном (прямое управление). Райли и др. (1949) ввели в экологическое моделирование механизмы обратной связи и учли влияние зоопланктона на развитие фитопланктона.

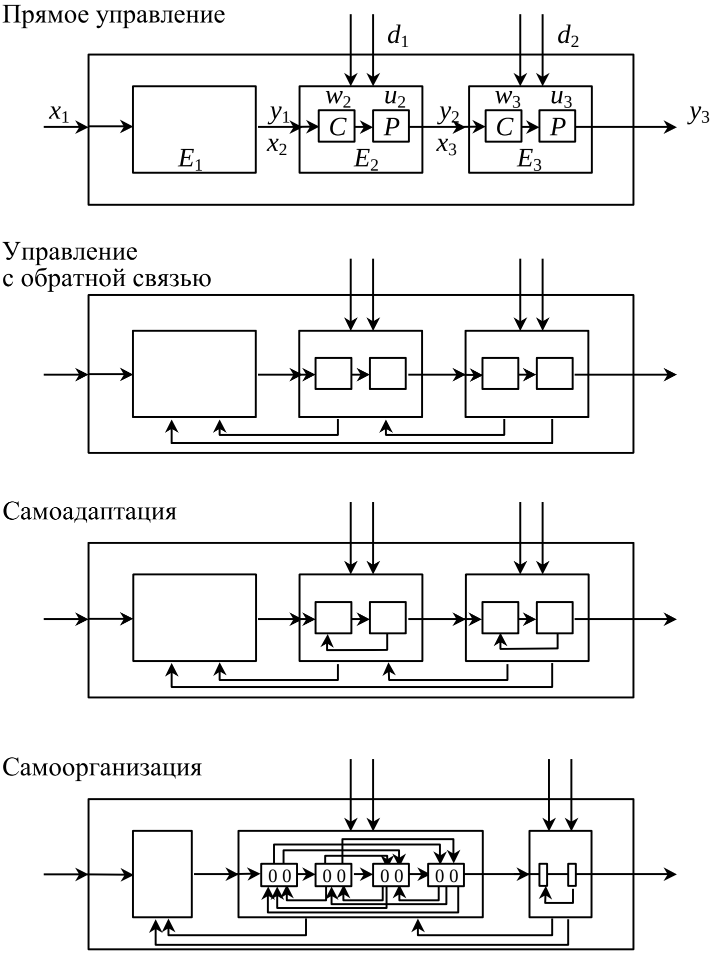


Рис. 2.8. Классификация моделей экосистем
на основе иерархии управления:

Символами обозначены только модели с прямым управлением. E_1 – критическое питательное вещество, E_2 – фитопланктон, E_3 – зоопланктон, x_i – входные переменные, y_i – выходные переменные, d_i – возмущающие переменные, C – управляющий элемент, P – система управления, w_i – потоки вещества, u_i – информационные потоки

Появление идеи биологического сообщества как адаптивной системы явилось следующим шагом в этом направлении. Любое изменение скорости фильтрации зоопланктона в ответ на изменение содержания фитопланктона можно рассматривать как адаптивное управление взаимодействиями в системе фитопланктон – зоопланктон. Любое изменение в составе зоопланктона, влияющее на его доступность для хищников, является выражением принципа самоорганизации.

2.2.4. Динамика состояния и управления в экосистемах

Изучение динамики экосистем средствами моделирования имеет своей конечной целью выработку методов прогнозирования влияния на экосистемы антропогенных воздействий, решение задач рационального природопользования, 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. В относительно простых ситуациях, возникающих в рамках биотехнологий при культивировании экосистем в искусственных условиях, описание их динамики может быть выполнено весьма подробно. В то же время при управлении природными комплексами соответствующие модели оказываются гораздо более сложными, и получение тех или иных рекомендаций оказывается возможным только на основе анализа результатов машинных экспериментов. Тем не менее, в обоих случаях структура управления обладает рядом общих свойств, отличающих ее от задачи управления техническими объектами.

Существование различных временных масштабов, а также способов управления экосистемами привело к выделению различных порядков в динамике экосистем (Straskraba, 1980). Эти порядки связаны с меняющимися свойствами системы (состояние, функция, структура, длительная биологическая эволюция элементов).

Динамикой системы называется изменение ее состояния во времени. В экосистемах происходит непрерывная циркуляция вещества, которая не связана с изменением скорости и называется динамикой состояния системы. В модели последняя обозначается с помощью фиксированной структуры S и вектора фиксированного параметра p (табл. 2.1, уравнение (1)). Этот порядок динамики относится к классу моделей с прямым и обратным управлением.

Второй порядок динамики связан с самоадаптивным управлением при условии, что параметры зависят от состояния системы: $p=p(S)$, уравнение (2). Поскольку функция системы изменяется, этот процесс называется динамикой функции. В этом случае система не срабатывает так, как того требуют независимые от времени функциональные отношения. Прилагая к этому явлению часто употребляемое понятие «параметр, зависимый от времени», отметим крупный недостаток: он не дает никакой информации о происхождении временной зависимости параметров. Для экосистем мы констатируем, что временная зависимость параметров является следствием зависимости параметра только от состояния системы. Это типичный случай биологической адаптации (например тепловой) организмов в экосистеме. Температурная зависимость описывается той или иной функцией, где параметры меняются в зависимости от действия температуры. Алгоритмическое функционирование такой системы осуществляется с помощью модели, параметры которой меняются, а структура остается неизменной.

Таблица 2.1

Порядки динамики систем

Порядок динамики системы	$\frac{dS}{dt}$	Стратегия моделирования	Высшая степень управления
1 порядок: динамика циркуляции вещества (кинематика)	$f(S,p,S_0,t)$ (1)	Постоянные параметры, фиксированная структура	Прямое управление или управление с обратной связью
2 порядок: динамика функционирования системы	$f(S,p(S),u,S_0,t)$ (2)	Переменные параметры, фиксированная структура; параметры меняются вместе с состоянием системы	Самоадаптация
3 порядок: динамика структуры системы	$f(S(u),p(S),u,S_0,t)$ (3)	Меняющаяся структура; взаимодействие между элементами меняется вместе с состоянием системы	Самоорганизация
4 порядок: динамика эволюции системы	$f(S(u),p(S),u(s),S_0,t)$ (4)	Целевая функция меняется вместе со структурой системы	Самоэволюция

S – фиксированные переменные состояния; S_0 – начальные значения S (исходные условия); p – вектор фиксированных параметров скорости; u – фиксированный вектор управления; $S(u)$ – переменные состояния, зависящие от управления, $p(S)$ – вектор параметров, зависящих от управления; $u(S)$ – вектор управления, зависящий от состояния; t – время.

Третий порядок динамики относится к управлению с помощью самоорганизации. Структура системы зависит от ее состояния и управления: $S=S(u)$, уравнение (3). Для экосистемы типичным примером является существование различных видов при различных условиях окружающей среды и управления: из широкого разнообразия потенциально имеющихся видов выбирается какой-то отдельный вид. Этот порядок называется динамикой структуры системы. Алгоритмическая реализация системы выполняется на модели, имеющей переменную структуру.

Интерпретация дарвиновской эволюции видов как длительного динамического процесса делает необходимым введение термина «динамика эволюции системы». В этом случае целевую функцию $u=u(S)$ можно рассматривать как функцию состояния системы, уравнение (4).

Порядки динамики расположены в той же иерархической последовательности, что и типы управления, так как низкие порядки всегда входят в более высокие. Временной масштаб всех порядков возрастает при движении снизу вверх. Поскольку циркуляция вещества есть непрерывный процесс, продолжительность периода физиологической адаптации может варьировать от нескольких часов (например адаптация водорослей к свету) до нескольких недель (например адаптация высших организмов к изменениям температуры). Отбор различных видов может проходить на протяжении длительного периода, начиная от нескольких дней и кончая месяцами или годами в зависимости от продолжительности жизни изучаемых организмов (экологические временные масштабы).

Различные трофические уровни, популяции, обитающие в пределах этих уровней, а также отдельные особи и физиологические процессы в одной и той же экосистеме развиваются в соответствии с различными порядками динамики, действующими одновременно. Можно представить себе, что в каждой подсистеме кто-то принимает решения, в результате которых происходит отбор тех или иных видов управления (рис. 2.9), причем сохраняется возможность управления одним и тем же процессом с помощью различных способов в разные периоды времени.

Характерная особенность экосистем как объектов управления состоит в том, что, как правило, допустимыми являются только однонаправленные входные воздействия. Это отбор или отлов особей из популяции, добавление субстрата и изъятие приросшей биомассы при культивировании, это внесение удобрений, поливы, укосы, прореживания в процессе формирования урожая и т.п. Имеется и еще обстоятельство, отличающее задачу управления экосистемами от классической. В данном случае фазовые переменные, означающие численность, концентрацию, плотность популяции, сухой вес массы растений и животных и т.п., могут принимать лишь неотрицательные значения и допустимые управления должны приводить к сохранению этого свойства. Отсюда возникает необходимость привлечения новых методов, ориентированных на процессы управления объектами и имеющих такие ограничения на допустимые управления, при которых нулевые значения управляющих воздействий лежат на границе области их допустимых значений.

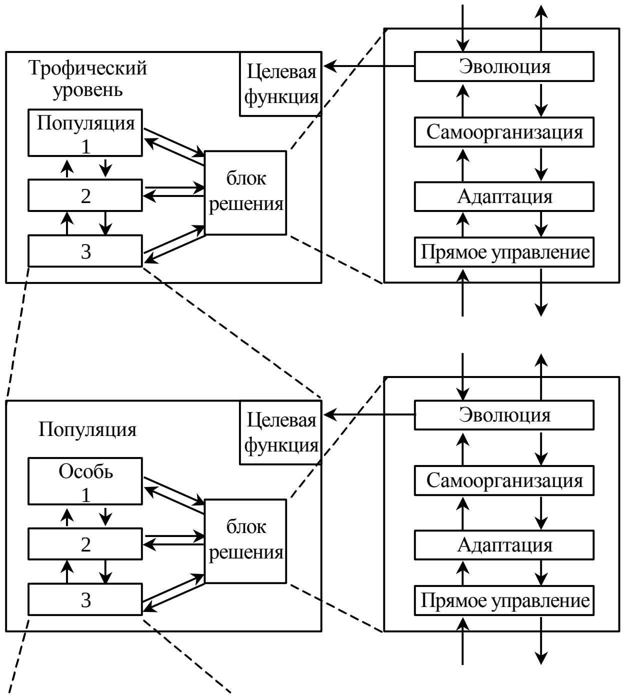


Рис. 2.9. Объединенная иерархия организации и управления в пресноводной экосистеме. Блоки решения играют роль переключателей между различными видами управления

Перечисленные вопросы возникают в первую очередь при рассмотрении задач управления искусственно культивируемыми сообществами микроорганизмов. Ситуация значительно усложняется при управлении природными экосистемами. 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. Значительно сложнее 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 трансформации 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, влаги в почвах, растениях и приземном воздухе, трансформация и миграция соединений азота, фосфора, калия и других химических элементов в почвах и 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.

Использование метода моделирования как мощного средства изучения динамики экосистем приносит ощутимую пользу уже на ранних стадиях, связанных с выбором структуры и уровня детализации модели. Существенное преимущество имитационного эксперимента по сравнению с натурным состоит в быстроте, 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. В то же время нельзя отрицать и использование аналитических моделей. Опыт работы с имитационными моделями говорит о том, что качественное исследование обобщенных аналитических моделей совершенно необходимо для эффективного использования и осмысливания результатов как имитационного, так и натурного эксперимента. Аналитические модели носят концептуальный характер и позволяют отказаться от слепого перебора бесконечного числа вариантов, показать, какие задачи разрешимы, а какие принципиально не имеют решения.

2.3. Условия существования популяций в экосистемах

Популяции, входящие в состав экосистем, независимо от их сложности, механизма управления в них, многообразия функциональных связей, существуют при определенных условиях, основными из которых являются условия среды обитания, приспособленности организмов к условиям среды, взаимоотношений со средой.

Емкость среды как регулятор численности популяций. Число особей в популяции ограничивается наличными ресурсами. Например, число пар синиц в лесу не может превысить число имеющихся мест для гнездования, численность хищников не может повыситься так сильно, чтобы численность их жертвы опустилась ниже уровня, необходимого для поддержания популяции самих хищников. Если плотность популяции перейдет на тот уровень, который не в состоянии обеспечить среда, смертность превысит рождаемость и численность популяции сократится. Если плотность популяции низка по сравнению с наличными ресурсами, то рождаемость превысит смертность и численность популяции начнет возрастать. Для каждого набора условий среды существует некая плотность популяции, при которой рождаемость и смертность точно уравновешивают друг друга и не происходит ни роста, ни сокращения численности популяции. Это равновесное состояние популяции соответствует емкости среды (*К*) (рис. 2.10). Как только численность популяции становится выше или ниже *К*, в ней возникают процессы, вновь приводящие ее величину в соответствие с емкостью среды. При низких плотностях популяции (т. *B*) рост популяции положителен, при высоких плотностях (т. *A*) – отрицателен. Точка *К* соответствует устойчивому равновесию.

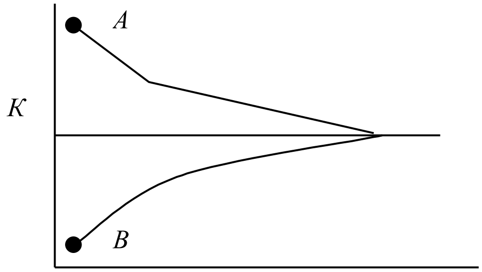


Рис. 2.10. Приближение численности популяции к равновесию: по вертикали – плотность популяции, по горизонтали – время. Приближение численности популяции к равновесному состоянию с течением времени. *A* и *B* – плотности двух популяций, рождаемость и смертность которых находятся в равновесии

Способность среды поддерживать вид изменяется в зависимости от климата и наличия ресурсов двух типов. К первому типу относятся невозобновляемые ресурсы, такие как пространство или места гнездования, которые могут быть полностью использованы популяцией, что создает верхний предел численности популяций. Ко второму типу относятся возобновляемые ресурсы, такие как пища, вода и свет, которыми популяция снабжается непрерывно. Потребности обширной популяции могут быть так велики, что количество возобновляющихся ресурсов понизится до очень низкого уровня, их станет трудно находить или ассимилировать и они не будут обеспечивать дальнейший рост популяции; однако эти ресурсы никогда не истощаются полностью. Возобновляющиеся ресурсы поддерживаются на некотором равновесном уровне благодаря сбалансированности между их эксплуатацией и продукцией. Когда популяция достигает численности, соответствующей емкости среды, потребности ее в ресурсах становятся равны скорости их возобновления. Если численность популяции превышает емкость среды, то эксплуатация превышает продукцию, ресурсы истощаются, и популяция начинает сокращаться и наоборот.

Закон минимума Либиха и закон толерантности Шелфорда. Любое условие, принадлежащее к пределу толерантности (приспособленности, выносливости) или превышающее его, называется лимитирующим фактором. При стационарном состоянии лимитирующим будет то вещество, доступные количества которого наиболее близки к необходимому. Идея о том, что выносливость организма определяется слабым звеном в цепи его экологических потребностей, была показана Либихом в 1840 г. Он установил, что урожай культур часто лимитируется не теми элементами питания, которые требуются в больших количествах, а теми, которые требуются в ничтожных количествах, но которых очень мало. Вывод Либиха о том, что «рост организмов зависит от того элемента питания, который присутствует в минимальном количестве», стал известен как закон минимума Либиха. Для успешного применения на практике данной концепции надо к ней добавить два вспомогательных принципа. Первый ограничительный: закон Либиха строго применим только в условиях стационарного состояния, т.е. когда приток и отток энергии и вещества сбалансированы. Второй важный вспомогательный принцип касается взаимодействия факторов. Доступность (высокая концентрация) одного вещества или действие какого-то фактора влияет на скорость потребления элемента питания, содержащегося в минимальном количестве.

Лимитирующим фактором может быть не только недостаток, но и избыток некоторых факторов, таких, например, как свет, тепло и вода. Следовательно, организмы характеризуются экологическим минимумом и экологическим максимумом: диапазон между этими двумя величинами называется пределом толерантности. Представление о лимитирующем влиянии максимума наравне с минимумом ввел Шелфорд, сформулировавший закон толерантности. Существует ряд вспомогательных принципов, которые дополняют закон толерантности:

1) организмы могут иметь высокий диапазон толерантности в отношении одного фактора и узкий – в отношении другого фактора;

2) организмы с широким диапазоном толерантности по всем факторам обычно наиболее широко распространены;

3) если условия по одному экологическому фактору не оптимальны для вида, то может сузиться и диапазон толерантности к другим экологическим факторам. Например, при лимитирующем содержании азота снижается засухоустойчивость злаков;

4) период размножения обычно является критическим, в этот период многие факторы среды часто становятся лимитирующими.

Степень толерантности в экологии выражают с помощью приставок «стено» – узкий и «эври» – широкий. Например, эвритермный, т.е. организм имеет высокую выносливость при широком диапазоне температуры.

Организмы в условиях лимитирующих факторов приспосабливаются к ним, а также изменяют условия так, чтобы ослабить действие лимитирующих факторов. Такая компенсация факторов особенно эффективна на уровне сообщества, но возможна и на уровне вида.

Биологический оптимум. Для каждого организма имеется некоторое определенное сочетание условий среды, оптимальное для его роста, существования и размножения. По обе стороны от этого оптимума биологическая активность постепенно снижается, пока, наконец, условия не станут такими, в которых организм вообще не может существовать. Диапазон условий среды, необходимых для сохранения популяции, гораздо уже, чем диапазон, обеспечивающий выживание отдельной особи. Для процветания популяции необходимы, помимо поддержания существования отдельных особей, ресурсы, обеспечивающие рост и размножение.

В общем взаимоотношения организма со средой можно представить на графике (рис. 2.11), изображающем зависимость уровня биологической активности (измеряемой в любых выбранных единицах) от изменений условий среды.

На графике кривая симметрична, хотя на практике соответствующие кривые не обязательно могут быть симметричными, скорее они будут асимметричными. Независимо от формы кривой активности экологическое распределение вида по градиенту определяется тремя уровнями толерантности. Во-первых, экстремальные условия могут вызвать полное нарушение важнейших биологических функций, что быстро приведет к гибели. Такие условия наступают за пределами точек *с* и *с'*. Во-вторых, для того чтобы организмы могли в течение длительного времени оставаться в стационарном состоянии, они должны сохранять известный уровень активности. В пределах отрезка *b*-*b'* градиента среды организм может существовать в течение неопределенно долгого времени. За пределами этого отрезка уровень активности организма слишком низок для того, чтобы все его функции поддерживались на соответствующем уровне, так что он может существовать в этих условиях очень недолго. В-третьих, популяция может сохранять свою численность лишь в том случае, если размножение компенсирует гибель. Размножение требует известных ресурсов, а, следовательно, биологическая активность должна при этом превышать уровень, необходимый для поддержания существования особи. Поэтому диапазон условий, в котором может существовать популяция уже, чем пределы толерантности отдельной особи. Особи могут жить в средах, не подходящих для поддержания популяции, но при этом их численность будет сохраняться только за счет иммиграции из популяций, живущих в более подходящих местообитаниях, где размножение превышает гибель.

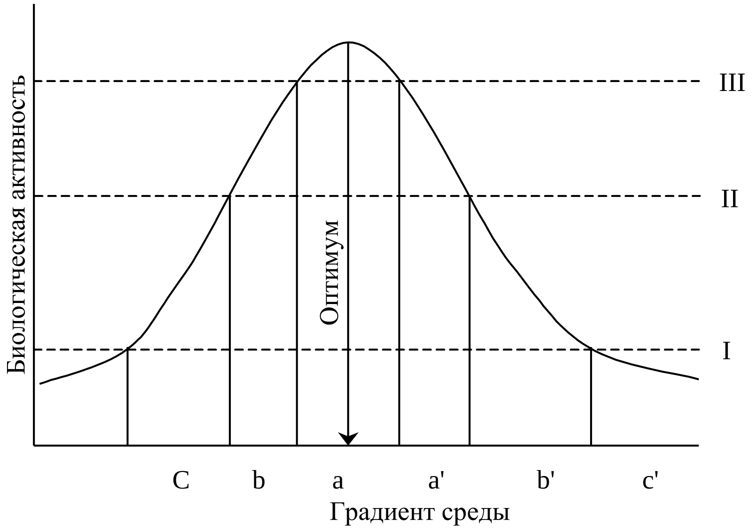


Рис. 2.11. Зависимость биологической активности от градиента условий среды

Уровни активности, необходимые для поддержания жизненно важных биологических функций (I), существования особи (II), существования популяции (III), определяют экстремальные летальные условия (с и с'), пределы выносливости особей (b,b') и популяции (а и а')

Энергетическая эффективность. Для передвижения**,** роста и поддержания разнообразных функций организмам необходима энергия. Энергия поступает в экосистему в форме лучистой энергии солнечного света, которую растения в процессе фотосинтеза превращают в химическую энергию. Скорость, с которой растения ассимилируют энергию солнечного света, называется первичной продуктивностью. Первичная продукция лежит в основе всей трофической структуры сообщества. Все механизмы экосистемы приводятся в движение энергией, поставляемой фотосинтезом. Естественно возникает вопрос, насколько эффективно экосистема использует эту энергию? Количество энергии, получаемой в процессе метаболизма на каждом трофическом уровне, уменьшается по мере переноса энергии с одного уровня на другой по пищевой цепи. Продуктивность каждого трофического уровня ограничивается продуктивностью уровня, непосредственно ему предшествующего. Продуктивность каждого последующего трофического уровня обычно составляет не более 5–20% предыдущего. Относительное количество энергии, передающейся от одного трофического уровня к другому, называется *экологической эффективностью* или *эффективностью пищевой цепи*. Отсюда возникают пирамиды энергии или продукции, которая всегда закономерно сужается к вершине, в отличие от пирамиды биомасс или численностей. Движение энергии через сообщество зависит от эффективности, с которой организмы потребляют свои пищевые ресурсы и превращают их в биомассу. Эта эффективность также называется эффективностью пищевой цепи (*ЭПЦ*) или экологической эффективностью (*ЭЭ*). Экологическая эффективность определяется как внутренними, физиологическими, характеристиками организмов, так и их внешними, экологическими, взаимоотношениями с окружающей средой. Разберем отдельное звено пищевой цепи (рис. 2.12).

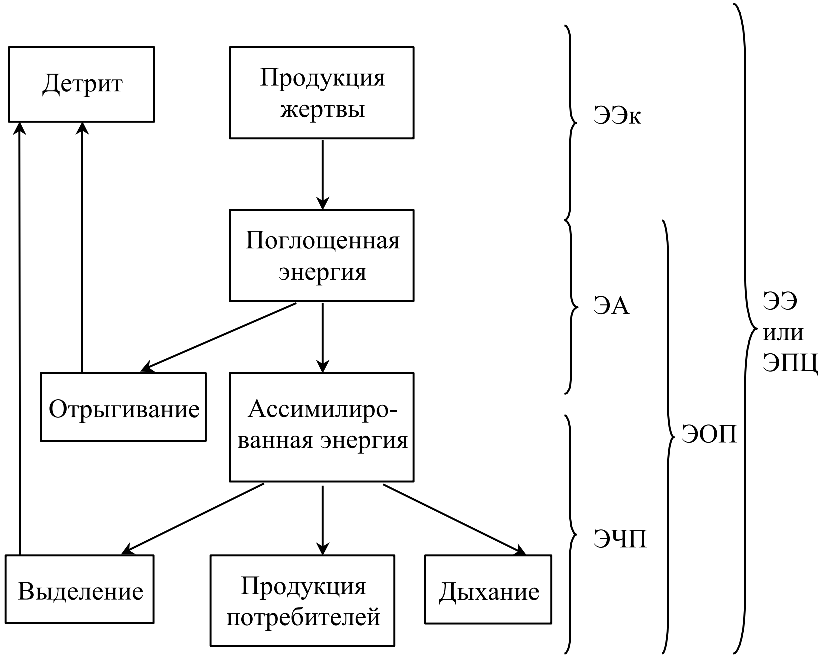


Рис. 2.12. Схема распределения энергии в пределах одного звена пищевой цепи и энергетические эффективности, связанные с каждым метаболическим этапом в организме

Детрит, продуцируемый в результате выбрасывания и выделения неперевариваемых остатков пищи, остается на трофическом уровне жертвы и может потребляться другими организмами.

Экологическая эффективность (*ЭЭ*) зависит от эффективностей трех главных ступеней в потоке энергии: эксплуатации (*ЭЭк*), ассимиляции (ЭА) и эффективности чистой продукции (*ЭЧП*) (табл. 2.2). Произведение *ЭА* и *ЭЧП* дает *ЭОП* (эффективность общей продукции) – процент съеденной пищи, превращенной в биомассу консумента. Произведение *ЭЭк* и *ЭОП* дает *ЭПЦ* (пищевой цепи) или *ЭЭ* – процент содержащейся в жертве «пищевой энергии», превращенной в биомассу консумента.

Таблица 2.2

Определение энергетических эффективностей

Эффективность	Формула
1. Эффективность эксплуатации – ЭЭк	$\dfrac{\text{Çàãëàòûâàíèå ïèùè}}{\text{ïðîäóêöèÿ æåðòâû}}$
2. Эффективность ассимиляции – ЭА	$\dfrac{\text{Àññèìèëÿöèÿ}}{\text{çàãëàòûâàíèå}}$
3. Эффективность чистой продукции – ЭЧП	$\dfrac{\text{Ïðîäóêöèÿ ïîòðåáèòåëåé}}{\text{àññèìèëÿöèÿ}}$
4. Эффективность общей продукции – ЭОП	$2 \cdot 3 = \dfrac{\text{ïðîäóêöèÿ ïîòðåáèòåëåé}}{\text{çàãëàòûâàíèå}}$
5. Экологическая эффективность – ЭЭ	$1 \cdot 2 \cdot 3 = \dfrac{\text{ïðîäóêöèÿ êîíñóìåíòà}}{\text{ïðîäóêöèÿ æåðòâû}}$

При исследовании эффективности функционирования сообщества возникает вопрос: куда отнести образовавшийся детрит? Если отнести его к трофическому уровню жертвы, то его придется рассматривать как неиспользованную энергию и включать в эффективность использования. Если же отнести детрит к трофическому уровню консумента, то его придется включать в эффективность общей продукции, что приведет к увеличению общей ЭЭ. Большинство экологов выделяют детрит в самостоятельную пищевую категорию, не принадлежащую ни к одному из трофических уровней.

Приход и расход энергии в сообщество должны быть сбалансированы, это можно выразить следующей схемой:

Приход	Расход
Ассимилированная энергия	Дыхание + рост
Внесение в систему	Вынесение из системы

Таким образом, вся ассимилированная энергия или энергия, внесенная в сообщество, либо рассеивается в процессе дыхания и идет на рост организмов, либо выносится из него.

Качество пищи и распределение энергии для обеспечения различных функций организмов определяет характер потока энергии через сообщество. Наиболее сильные различия в этом отношении существуют между водными и наземными средами, что обусловлено глубокими отличиями в приспособленности организмов к каждой из этих сред. В водных экосистемах энергия течет быстро и весьма эффективно переносится от одного трофического уровня к другому, что создает возможность для образования длинных пищевых цепей. В наземных экосистемах некоторая часть энергии быстро рассеивается, а тем самым перенос энергии от одного уровня к другому относительно неэффективен; остальная же энергия в течение длительного времени хранится в растениях в виде опорных тканей, и в почве в виде органического детрита.

В наземных местообитаниях экологические эффективности обычно ниже, и плотоядные в наземных сообществах в среднем могут питаться не выше, чем на третьем трофическом уровне, тогда как водные плотоядные могут добывать пищу на четвертом и пятом трофических уровнях. Это не значит, что число звеньев в наземных цепях питания не может быть больше трех; некоторая часть энергии может пройти через десяток звеньев, прежде чем она рассеивается в процессе дыхания. Возможно, однако, что эти высокие трофические уровни не содержат достаточного количества энергии, чтобы полностью обеспечить существование популяции, например, какого-либо хищника.

Контрольные вопросы

1. Характеристики сложных систем.
2. Понятие прямой и обратной связи в системах.
3. Выражением какого принципа является декомпозиция систем до элементарных подсистем?
4. Что характеризуют переменные состояния системы?
5. Признаки замкнутых и открытых систем.
6. Группы управления в экологических системах.
7. Понятие оптимального управления в экосистемах.
8. Уровни иерархической организации экосистем.
9. Порядки динамики экосистем.
10. Роль моделирования при исследовании динамики экосистем.

Глава 3
СТРУКТУРА СООБЩЕСТВ, ВИДОВОЕ РАЗНООБРАЗИЕ И ЭКОЛОГИЧЕСКАЯ НИША

3.1. Структура сообществ и видовое разнообразие

3.1.1. Сообщество

В самых общих чертах термин «сообщество» означает группу совместно обитающих видовых популяций, например, к каком-либо пруду или лесу. Сообщество организмов следует рассматривать, прежде всего, как некое организованное целое, и любое определение должно отражать взаимодействия между входящими в его состав популяциями всех трофических уровней, которые имеются в данном местообитании. Действительно, виды адаптируются к присутствию других видов, а поэтому так же, как и популяции, обладают свойствами, выходящими за пределы свойств составляющих их особей; сообщество – это нечто большее, чем простая сумма отдельных популяций и их взаимодействий. В настоящее время самым точным можно считать определение Уиттейкера (1975), который описывает сообщество как сочетание популяций растений, животных и микроорганизмов, взаимодействующих друг с другом в пределах данной среды и образующих тем самым особую живую систему со своим собственным составом, структурой, взаимоотношениями с окружающей средой, развитием и функциями. Концепция сообщества нередко оказывается некоей абстракцией, и поэтому при конкретном определении его возникают определенные трудности. В действительности сообщества – это открытые системы и обычно непрерывно переходят одно в другое вдоль тех или иных градиентов среды, а не занимают четко ограниченные зоны (Джиллер, 1988).

Сообщества как живые системы взаимодействующих видовых популяций имеют определенную организацию и развиваются по определенным закономерностям. Один из возможных способов исследования организации сообщества – изучение его на уровне отдельных составляющих, когда поведение и динамика популяций индивидуальных видов рассматриваются с точки зрения взаимодействий между популяциями и в пределах каждой из них. Такой метод, возникший при популяционных исследованиях отдельных видов или пар видов, трудно распространить на многовидовые системы, поэтому к таким системам применяют альтернативный подход, при котором упор делают на общие особенности структуры сообщества. В подобного рода исследованиях полезной оказывается концепция гильдии, т.е. группы видов, использующих определенный ресурс или совокупность ресурсов функционально сходным образом. Члены таких гильдий сильно взаимодействуют друг с другом и слабо – с остальным сообществом. Так, можно говорить о гильдии насекомоядных птиц или о гильдии ящериц со сходным местообитанием. Гильдии служат аренами наиболее интенсивных межвидовых взаимодействий.

3.1.2. Структура сообщества

К структуре сообщества относятся:

– всевозможные способы связей и взаимодействий между отдельными членами сообщества (например, типы распределения ресурсов и пространственное и временно́е обилие видов в данном сообществе);

– проявляющиеся на уровне сообщества свойства, обусловленные этими взаимоотношениями (трофические уровни, скорости и эффективности связывания энергии и ее переноса, сукцессия, круговорот питательных веществ и т.п.).

Трудно рассматривать все стороны этой сложной и обширной проблемы. Можно, тем не менее, проанализировать структуру сообщества с помощью двух важных показателей его организации: число составляющих его видов и их относительное обилие. Исследование этих показателей позволяет экологу ответить на следующие вопросы:

1. Как виды приспосабливаются друг к другу, образуя сообщество?
2. Чем определяется число видов, составляющих разные сообщества?
3. Как могли бы взаимодействия между видовыми популяциями ограничивать это число сверху?
4. С чем связаны различия в относительном обилии видов в сообществе?

Располагая такими данными, можно выявить общие закономерности в организации природных сообществ разных географических областей, подтвердить наличие у них определенной структуры и приступить к выявлению некоторых основных правил их организации. Биотические сообщества резко различаются по числу входящих в них видов животных и растений (т.е. по видовому богатству), причем при большом видовом богатстве между сообществами существуют различия по относительному обилию составляющих видов.

3.1.3. Видовое разнообразие

Наряду с такими простыми показателями, как число видов, нередко в качестве более наглядной характеристики видового богатства сообщества используют видовое разнообразие, поскольку в этот показатель входит как число видов, так и их относительное обилие. Выбор одного из всей массы предложенных математических индексов зависит от возможности обнаружить и определить все имеющиеся в сообществе виды, а также от трудности оценки их обилия (Дулепов и др., 2004). Для многих целей число имеющихся видов служит простейшей и самой полезной мерой локального или регионального разнообразия.

Один из методов оценки числа видов, встречающихся в отдельных областях, состоит в следующем: карты обширных территорий разбивают на квадраты одинаковой величины и накладывают на них карты географических ареалов отдельных видов. Эти и другие исследования выявили хорошо известные широтные градиенты видового богатства, т.е. заметное увеличение числа видов в большинстве групп организмов по направлению к экватору. Ярким примером служит сравнение разнообразия древесных пород в большинстве дождевых тропических лесов и однородных древостоев северных районов. Типичный широтный градиент наблюдается у гнездящихся птиц, муравьев, кораллов, оболочников, равноногих раков, голожаберных и брюхоногих моллюсков. Один из недостатков подобных исследований связан с тем, что число местообитаний в данном квадрате или в данной области зависит от рельефа местности. Другой недостаток – большее разнообразие местообитаний в низких широтах (от тропических до бореальных), чем в высоких (где разнообразие постепенно сокращается).

В исследованиях меньшего масштаба сравнивается видовое богатство по многим различным местообитаниям в пределах широтных поясов. При этом обычно выявляются различия между смежными местообитаниями, даже если между ними нет физических преград, препятствующих переходу видов из одного местообитания в другое.

Широтные градиенты присущи не всем организмам. Они не обнаружены в таких группах илоядных морских беспозвоночных, как офиуры и голотурии, которые вообще не отличаются большим разнообразием. У австралийских позвоночных широтные градиенты практически отсутствуют. Эти исключения из правила заслуживают дальнейшего исследования, и выявление их причин важно для понимания структуры сообществ.

Повторяющиеся закономерности в распределении видового богатства позволяют считать, что для них можно найти общее объяснение. Если сообщество представляет собой некий структурированный комплекс организмов, то его структура должна определяться взаимодействиями между составляющими его организмами. Экологическая ниша – это отражение места, занимаемого организмом или видом в сообществе, причем в это понятие входят помимо устойчивости к физическим факторам среды также взаимодействия с другими организмами. Поскольку отдельные виды занимают различные ниши, возникает вопрос: «Какие факторы регулируют число ниш в данной области?» В результате поисков ответов на этот вопрос появились такие понятия, как упаковка видов, перекрывание ниш и ширина ниши. Объединение этих понятий создало специальную дисциплину – теорию ниши.

3.2. Экологическая ниша, конкуренция и хищничество

3.2.1. Теория ниши

Гриннелл (1917) ввел термин «ниша» для обозначения самой мелкой единицы распространения вида. Он подразумевал, что ниши разных видов не перекрываются, и определял потенциальный характер распространения отдельного вида в отсутствие взаимодействия с другими видами. Определение ниши, данное Элтоном (1927), охватывает главным образом функциональный ее аспект, описывая нишу как место данного организма в биотической среде в смысле его пищевых связей и взаимодействий с врагами. Определение Элтона касалось в сущности реального, а не потенциального места данного вида в природе.

Одновременно с концепцией ниши развивалась связанная с ней концепция конкурентного исключения. Согласно этой концепции, два вида с идентичной экологией не могут сосуществовать в одном и том же месте. Поначалу эта концепция не вызвала у экологов большого интереса, однако проведенные математические расчеты (уравнения Лотки-Вольтерра) и контролируемые лабораторные эксперименты (знаменитые работы Гаузе и Парка) показали, что при установлении равновесия между популяциями двух видов часто имеет место конкурентное исключение. С тех пор принцип конкурентного исключения, гласящий, что «полные конкуренты не могут сосуществовать бесконечно», стал одним из официальных догматов теоретической экологии. Важно следствие, вытекающее из этого принципа. Если два вида сосуществуют, то между ними должно быть какое-то экологическое различие, а это означает, что каждый из них занимает свою особую нишу.

Хатчинсоном (1958) было дано формальное и потенциально количественное определение ниши. Он считал, что нишу следует определять с учетом всего диапазона физических, химических и биотических переменных среды, к которым должен быть адаптирован данный вид и под действием которых видовая популяция живет и возобновляется бесконечно долгое время. В идеале каждую такую переменную можно рассматривать как некий градиент, на котором у каждого вида имеется свой диапазон активности или устойчивости. Примером может служить освещенность в лесу, которая убывает логарифмически от вершин деревьев к почве, по мере того, как растения преграждают путь к свету. Эволюция видов идет в соответствии с этим градиентом, и каждый из них адаптируется к разным диапазонам освещенности.

Градиент каждого фактора среды можно представить себе как некое измерение пространства. Если данная ниша имеет *n* значимых измерений, то ее можно описать в терминах *n*-мерного пространства (гиперпространства). Получить гиперпространство можно последовательно добавляя по одному измерению. На рис. 3.1 изображена реакция некоего вида на градиент одного из факторов среды, где некоторая мера приспособленности дает нормальное распределение относительно точки оптимума на градиенте.

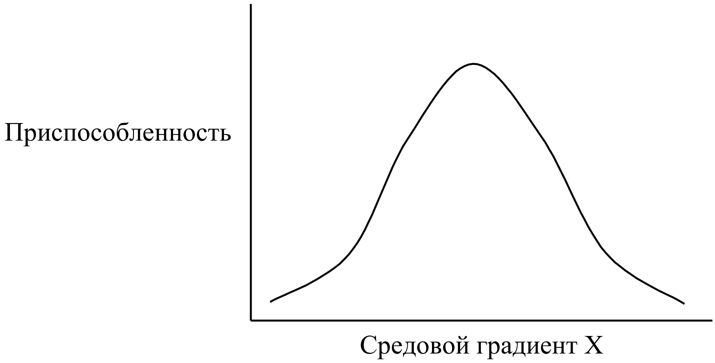
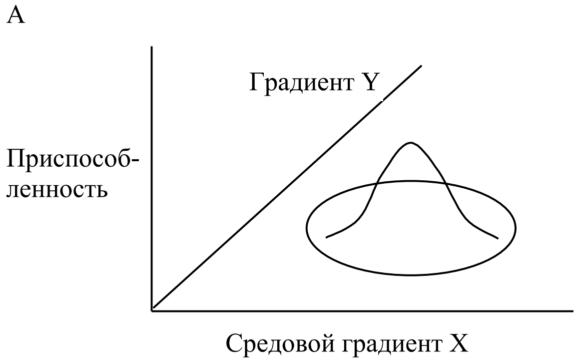


Рис. 3.1. Реакция вида на градиент одного из факторов среды. Приспособленность оценивается по успеху размножения, величине популяции и выживаемости

Можно также графически представить реакции вида одновременно на два или три ресурса (рис. 3.2) и распространить эту операцию на любое число координатных осей, используя принципы *n*-мерной геометрии; в результате получится очень сложное гиперпространство, отражающее реакции данной видовой популяции на все факторы среды. Предполагается, что при этом учтены все значимые переменные и что они независимы друг от друга.

Хатчинсон различает два состояния видовой ниши: фундаментальную нишу, охватывающую все множество оптимальных условий, в которых данный вид может обитать в отсутствие врагов, и реализованную нишу, или тот фактический комплекс условий, в которых этот вид обычно существует. Реализованная ниша меньше фундаментальной или равна ей.

Такой многомерный подход дает возможность представить себе взаимоотношения между видами и тем самым способствует нашему пониманию организации сообщества. Грубо говоря, общее пространство ниш данного местообитания можно представить себе в виде «*n*-стороннего» ящика, в который все ниши данного сообщества упакованы подобно «*n*-сторонним» шарам. Если ниши всегда дискретны (т.е. между фундаментальными нишами нет перекрывания), то видовое богатство сообщества зависит от общего занимаемого ими пространства (переменная местообитания) и от среднего размера каждой ниши (видовая переменная). При данном разнообразии ресурсов широкие ниши приведут к меньшему видовому богатству, чем узкие.



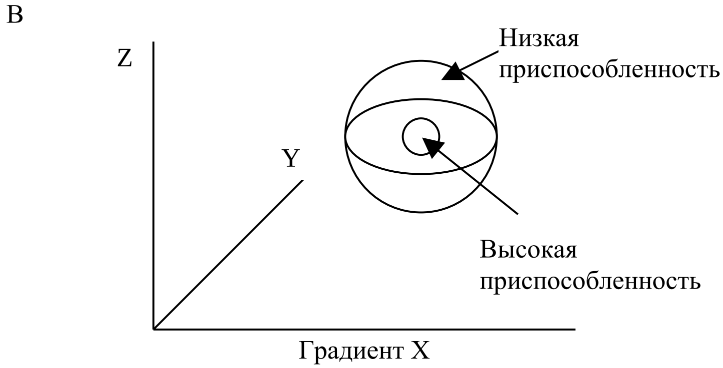


Рис. 3.2. А – одновременная реакция вида на два градиента (X,Y), В – на три градиента (X,Y,Z) факторов среды с учетом приспособленности вида

3.2.2. Ширина ниши

Ширина (размер) ниши имеет важнейшее значение для теории ниши, приведет к более глубокому пониманию причин видового разнообразия. При многомерном подходе ширина ниши определяется как общая сумма всего разнообразия ресурсов, используемых популяцией вида. Для того чтобы измерить ее на этом уровне, необходимо описать все относящиеся к ней параметры и непрерывно оценивать соотношение одновременно используемых ресурсов, что представляется совершенно невыполнимым. Поэтому экологическую нишу все чаще стали отождествлять с распределением активности вида всего лишь по одной или нескольким наиболее важным осям ниши. Таким образом, нишу для каждого вида определяют с помощью функции использования одного из ресурсов (распределения активности вида) вдоль его градиента. При условии действительной независимости выбранных измерений ниши общее многомерное ее использование можно представить как произведение отдельных одномерных функций использования. Самые важные из описанных таким образом характеристик ниши – это высота (максимальная интенсивность использования ресурсов или уровень активности) и область, охватываемая кривой использования. Последняя может служить несколько ограниченной, но доступной мерой ширины ниши.

Ширина ниши данного вида слагается из двух отдельных компонент. Внутрифенотипическая компонента (ВФК) описывает уровень изменчивости в использовании ресурса отдельными особями, а межфенотипическая компонента (МФК) – изменчивость по этому параметру среди особей всей видовой популяции. Общая ширина ниши В=ВФК+МФК. Если ширина ниши на 100% определяется МФК, вид полиморфен и состоит из специализированных форм; если же ширина ниши на 100% определяется ВФК, вид мономорфен и состоит только из одних неспециализированных форм. Совершенно очевидно, что реальные популяции занимают в этом смысле промежуточное положение.

Существуют два основных метода для оценки использования ресурса. Первый сводится к простому описанию потребления видом некоего непрерывного ресурса с точки зрения среднего значения (d) и ширины (w – одно стандартное отклонение) кривой использования этого ресурса вдоль его градиента. Большое значение w характеризует широкую нишу. Если значение w слева и справа от среднего его значения различны, то возможны подгонки. Эту меру ширины обычно применяют для оценки использования ресурса на основе морфологической изменчивости какого-либо признака, связанного с его потреблением (например пищедобывательных структур). Второй метод не требует непрерывного распределения ресурса; он основан на определении доли использования различных состояний ресурса (например разных видов жертв). Для видов с широкой нишей характерно потребление ресурсов, пропорциональное их доступности, тогда как виды с узкой нишей обычно усиленно потребляют лишь некоторые их составляющие.

3.2.3. Перекрывание ниш

Большинство организмов не обитает в своей потенциальной фундаментальной нише, а вследствие взаимодействий с другими организмами занимает меньшую по размерам реализованную нишу. Главными взаимодействиями обычно считаются хищничество и конкуренция; последняя оказалась связанной с теорией ниши через концепцию перекрывания ниш. Ниши видов, входящих в сообщество, не представляют собой дискретные, не взаимодействующие друг с другом единицы; напротив, виды склонны частично разделять между собой свои фундаментальные ниши, в результате чего на один и тот же ресурс одновременно претендуют две или более видовые популяции. Пользуясь терминологией Хатчинсона, гиперпространства ниш одних видов включают в себя части гиперпространств других видов, т.е. перекрываются с ними.

Если перекрывание очень незначительно или если ресурсы сверхобильны, то виды с перекрывающимися нишами могут сосуществовать в практически раздельных и почти фундаментальных нишах. Если ниши перекрываются в большей степени и имеющиеся в зоне перекрывания ресурсы не в состоянии удовлетворить потребности видов, то обилие менее приспособленного вида будет ограничиваться в результате его взаимодействия с более приспособленным. В конечном счете в перекрывающихся частях любых двух ниш может произойти конкурентное исключение. Приняв такое допущение, рассмотрим гипотетические последствия перекрывания между нишами двух видов, выраженного в разной степени.

1. Две фундаментальные ниши идентичны. В такой весьма маловероятной ситуации вид с более высокой конкурентоспособностью полностью вытеснит другой вид.

2. Фундаментальная ниша одного вида целиком включена в более обширную нишу другого вида. В этом случае первый вид, если он менее конкурентоспособен, будет вытеснен; если же он более конкурентоспособен, то он вытеснит другой вид из области, за которую происходит конкуренция (рис. 3.3-А).

3. Фундаментальные ниши частично перекрываются. При этом более конкурентоспособный вид занимает область перекрывания и у каждого вида есть собственное, никем не оспариваемое жизненное пространство (рис. 3.3-Б). Поэтому существование теоретически возможно, но оно зависит от того, какую степень перекрывания может выдержать менее конкурентоспособный вид.

4. Ниши непосредственно примыкают одна к другой (рис. 3.3-В). Прямое конкурентное исключение при этом невозможно, но такое расположение ниш может отражать его результат.

5. Ниши полностью разобщены, так что каждый вид занимает свою фундаментальную нишу (рис. 3.3-Г).

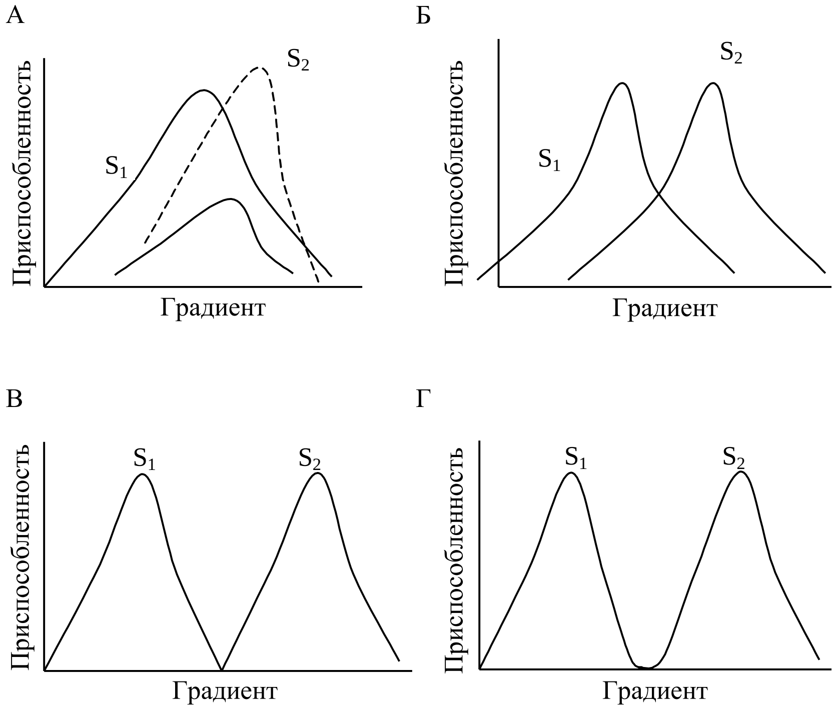


Рис. 3.3. Возможные взаимоотношения между нишами двух видов вдоль основного градиента среды:

А – включение одной ниши в другую; Б – перекрывающиеся ниши; В – соприкасающиеся ниши; Г – разобщенные ниши

Например, леса в большинстве случаев содержат не 5–7 видов растений, а гораздо большее их число. Эти добавочные виды также будут использовать градиент освещенности, вклиниваясь между центральными частями популяций основных видов. Исходя из приведенных рассуждений, включение дополнительных видов на этот градиент должно уменьшать ширину ниш уже присутствующих видов и приводить к все более плотной упаковке возрастающего числа видов растений по тому же градиенту освещенности.

Перекрывание ниш обычно измеряют на основании данных об использовании таких ресурсов, как пища и микроэкотоп (сочетание важных и легко измеряемых факторов). Таким образом, перекрывание ниш описывают как перекрывание в использовании данного ресурса между двумя соседними видами на его градиенте.

Самые простые измерения основаны на разделении функций использования ресурсов; при этом перекрывание описывается следующим отношением, характеризующим разделение ресурсов:

$$\rho_{ij} = d_{ij} / w_{ij}, \qquad (3.1)$$

где d_{ij} – разность средних значений использования ресурсов для видов i и j, а w_{ij} – общая ширина кривой использования (одно стандартное отклонение), вычисляемая по формуле

$$w_{ij} = (w_i^2 + w_j^2 / 2)^{1/2}.$$

Если ρ_{ij} меньше трех, то между двумя видами теоретически должно иметь место некоторое взаимодействие. Должен также существовать некий минимальный уровень разделения ресурсов, ниже которого действует принцип конкурентного исключения.

Более сложные измерения основаны на ряде методов, включающих использование процента сходства, критерия χ^2 и теории информации. В настоящее время применяется по крайней мере восемь различных индексов перекрывания ниш. Картину перекрывания ниш между всеми членами данной гильдии или сообщества можно получить с помощью матрицы ресурсов при перекрывании ниш: строят матрицу $m \cdot n$, указывающую количество каждого из m состояний ресурса, используемых n разных видов, и на ее основе – матрицу $n \cdot n$ перекрывания между всеми парами видов.

Можно также оценить совокупное перекрывание по двум или более измерениям ресурсов, с тем, чтобы получить некую меру общего перекрывания между видами. В случае независимых ресурсов используется произведение значений перекрывания по отдельным ресурсам; если же ресурсы связаны между собой, то следует суммировать эти значения. Метода, который учитывал бы различные степени независимости между измерениями ресурсов, до сих пор не создано.

При измерении перекрывания между нишами необходимо учитывать, что непрерывные спектры ресурсов не обеспечивают равноценных экологических возможностей на всем своем протяжении. Так, например, мелкая жертва, по всей вероятности, будет более обильной, чем крупная. Возможны также существенные различия в использовании ресурсов между разными весовыми, размерными и возрастными классами данного вида. Такие внутривидовые различия необходимо принимать во внимание при сравнении перекрывания ниш нескольких видов.

Перекрывание ниш и конкуренция. Значения перекрывания ниш часто приравнивают к коэффициенту конкуренции (α) классических уравнений межвидовой конкуренции Лотки-Вольтерра. Однако такие сравнения чреваты биологическими затруднениями, и истинное соотношение между перекрыванием ниш и конкуренцией остается неясным.

Как показано на рис. 3.3, одно лишь перекрывание кривых использования ресурсов не обязательно ведет к конкуренции. Подобным же образом интенсивность конкуренции не обязательно должна зависеть от степени перекрывания ниш. Чем обильнее данный ресурс, тем менее вероятно, что его совместное использование приведет к конкуренции; поэтому не приходится ожидать конкуренции в случае совместного использования какого-нибудь неограниченного ресурса, предоставляемого видам некоторыми местобитаниями (самым ярким примером служит кислород в большинстве наземных систем). Таким образом, зависимость между перекрыванием ниш и конкуренцией обусловливается в первую очередь соотношением между потребностью в данном ресурсе и его количеством в среде, или степенью насыщения среды данным ресурсом.

Существует мнение о существовании обратной зависимости между конкуренцией и перекрыванием ниш. Согласно этому мнению, максимально допустимое перекрывание в условиях интенсивной конкуренции должно быть меньше, чем в средах с более низким отношением между потребностями в ресурсе и его наличием.

3.2.4. Диффузная конкуренция

Анализ перекрывания ниш привел к другому аспекту теории ниши – диффузной конкуренции. Ниша данного вида обычно перекрывается лишь с ограниченным числом соседних ниш вдоль градиента одного ресурса; однако по мере того, как мы начинаем изучать все большее и большее число факторов среды, число потенциальных соседей возрастает. Поэтому, хотя попарное перекрывание ниш может быть невелико, суммарный эффект такой диффузной конкуренции может сильно сокращать величину реализованной ниши, иногда даже до такой степени, что она становится слишком мала для поддержания жизнеспособной популяции (рис. 3.4). В результате вид может оказаться «выпихнутым» группой других видов.

Число потенциальных ниш в данном сообществе можно, таким образом, рассматривать в зависимости от того, в какой степени развитие этого сообщества приводит к разделению частично перекрывающихся ниш при данном экологическом режиме. В терминах простой модели «шаров и ящика» это означает, что объем каждого шара можно уменьшить, втискивая в ящик все новые и новые шары. Число шаров, которые можно упаковать таким образом в имеющееся пространство, зависит от упругости или податливости.

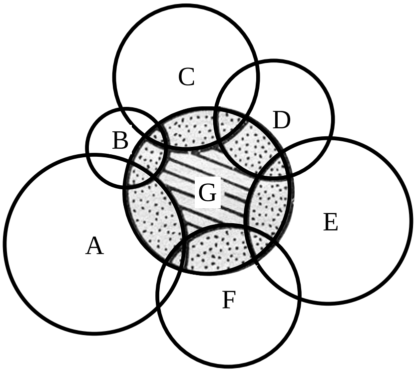


Рис. 3.4. Диффузная конкуренция может сократить фундаментальную нишу вида G (покрытая точками и заштрихованная область) до его реализованной ниши (только заштрихованная область)

3.2.5. Динамика ниши

Ниши имеют свойство изменяться как во времени, так и в пространстве вместе с изменением физической и биотической среды (например, ниша может изменять свое местоположение во всем пространстве ниш).

Временные изменения могут быть краткосрочными, т.е. укладываться в период жизни одной особи или нескольких поколений (экологическая шкала времени). Особенно характерны такие изменения для организмов, претерпевающих в процессе развития ту или иную форму метаморфоза: насекомых с полным превращением, планктонных и прикрепленных ракообразных, большинства водных насекомых, амфибий и т.п. Такие организмы занимают в разные периоды своего жизненного цикла совершенно разобщенные ниши. Для других организмов, например насекомых с неполным превращением или хищников, переходящих по мере своего роста на питание более крупной жертвой, характерно более постепенное и непрерывное изменение ниши. Изменения ниши, несомненно, происходили также в эволюционных масштабах времени, когда возникновение новых адаптивных зон открывало возможности для их освоения путем эволюции и адаптивной радиации.

В меньших масштабах величина или ширина реализованной ниши может изменяться в результате реакций данного вида или его конкурентов на изменения запасов ресурсов или на активность этих ресурсов. Такие изменения предсказывает теория оптимального фуражирования, в основе которой лежит мысль о том, что отдельные потребители должны максимизировать свою индивидуальную приспособленность (обычно путем максимизации чистого энергетического выигрыша) своим поведением, связанным с добыванием пищи. Наконец, весьма вероятны суточные и сезонные ритмы смещений пространства ниш данного сообщества, так что взаимосвязи каждого входящего в сообщество вида с остальными видами также должны непрерывно изменяться. Возвращаясь к упоминавшейся выше простой модели сообщества, изменяться могут не только форма и величина «шаров» и их расположение в «ящике», но также емкость и форма самого «ящика». Именно эти динамические свойства ниши настолько затрудняют ее измерение, что в лучшем случае можно получить лишь разрозненные данные, указывающие на относительную ширину и перекрывание ниш в пределах сообществ, а затем с их помощью разобраться в сложностях организации сообществ в целом.

Ниши и организмы образуют комплементарные пары. Согласно одной точке зрения, подчеркивающей роль самого организма в создании и определении собственной среды обитания, ниша порождается тем, кто ее занимает (Джиллер, 1988). Подобный взгляд частично приемлем, если иметь в виду постройку таких сооружений, как гнезда, термитники, плотины бобров и т.п., поскольку эти сооружения изменяют среду и образуют часть ниши организма, хотя и не всю ее. Согласно альтернативной точке зрения ниша представляет собой свойство данного сообщества и вне этого сообщества лишена смысла. Это подразумевает, что ниши создаются абиотическими и биотическими компонентами экосистемы, т.е. предсуществуют и заполняются в результате адаптации видов на протяжении некоторого периода эволюционного изменения. Поэтому следует ожидать, что в экосистемах со сходными условиями среды сообщества должны быть построены сходным образом и содержать одну или несколько в основном идентичных ниш. Адаптации популяций, занимающих такие ниши в этих независимо сложившихся сообществах, также должны быть сходными, даже если сами виды совершенно неродственны между собой. Это явление носит название экологической эквивалентности или конвергентной эволюции, и его существование служит доводом в пользу того, что ниша порождается сообществом.

Для того чтобы два неродственных друг другу вида в процессе эволюции достигли почти полной идентичности, ниши, к которым они адаптировались, также должны быть почти одинаковыми. Если считать нишу свойством видовой популяции, то такая идентичность невозможна.

Однако среди относительно крупных гильдий или сходных гильдий с неравным числом видов таких точных соответствий, по-видимому, обнаружить не удается; у них наблюдается замещение трех видов двумя, пяти – тремя или еще более сложные отношения. Это позволяет предполагать, что в подобных ситуациях наблюдаемые ниши являются скорее свойством занимающих ее видов.

Следует ли считать нишу свойством вида или сообщества? Создается впечатление, что определенное экологическое пространство ниш создается физическими и биотическими компонентами экосистемы, т.е. это свойство сообщества в целом. Это пространство в двух сходных насыщенных экосистемах может быть поделено между входящими в них видами по совершенно одинаковой схеме, что приведет к экологической эквивалентности – на этом уровне ниша представляется свойством сообщества. Сложные соотношения при замещениях видов в двух сходных экосистемах могут быть обусловлены историческими факторами, таксономическими преградами, препятствующими конвергенции, или различиями в состоянии ресурсов. Это должно оказывать влияние на число, свойства имеющихся видов и на вероятность их эквивалентности. Поэтому ниши, наблюдаемые в одном или в обоих таких сообществах, могут в большей степени представлять собой свойство входящих в данные экосистемы видов.

3.2.6. Конкуренция и ниша; лимитирующее сходство и дифференциальное перекрывание ниш

Конкуренция оказывает влияние как на размеры ниши, так и на степень перекрывания между нишами разных видов. Какой степени может достигать перекрывание ниш, не вызывая конкурентного исключения? В этом вопросе заключена проблема лимитирующего сходства.

Лимитирующее сходство. Для количественного выражения того уровня, до которого может доходить сходство между сосуществующими видами, исследователями было предложено множество моделей. В основе многих из них лежит модель конкуренции Лотки-Вольтерра и разделение видов по одномерному спектру ресурсов. Каждый вид, входящий в данную систему, использует данный ресурс определенным образом, описываемым соответствующей функцией использования ($f(x)$), и степень разделения ниш характеризуется отношением (3.1), которое велико для хорошо разделенных и мало для сильно перекрывающихся ниш.

Первые модели исходили из допущения определенной и постоянной формы функции использования, тогда как более поздние допускали наличие в уравнениях конкуренции случайно флуктуирующих компонентов (например флуктуации емкости популяции). Во всех случаях необходимым условием для совместного существования примыкающих друг к другу видов является отношение $d/w>1$, при котором на принимаемые моделью допущения не накладывается слишком больших ограничений. Другой подход допускает «эволюцию» функции использования, в результате чего входящие в систему виды достигают максимальной приспособленности. Эти модели предсказывают, что система стремится к достижению ситуации, при которой $d/w>1$ для видов, стоящих рядом на градиенте данного ресурса.

Альтернативный метод теоретически рассматривает шансы на выживание у вида, который пытается втиснуть свою нишу между нишами двух уже имеющихся видов. Мак-Артур и Левинс (1967) установили, что лимитирующее отношение d/w для успешной инвазии равно 1,56.

Имеется известный предел перекрывания ниш и сосуществование видов, использующих один и тот же ресурс, возможно лишь в том случае, когда среднее различие между ними превышает типичную внутривидовую изменчивость, т.е. $d>w$. Обычно отношение $d/w\leq 1$ указывает на потенциально сильную конкуренцию за данный ресурс, а $d/w>3$ позволяет считать, что между видами нет взаимодействия, т.е. нет перекрывания ниш.

Для получения надежной оценки действительного потребления ресурсов и сравнения сообществ в различных местообитаниях часто необходимо оценивать различия в потребляемых ресурсах, используя для этого видовой признак, определяющий положение данного вида на какой-либо оси ресурса. В этих целях очень часто используют морфологические признаки, поскольку их легко измерить. При этом измеряют среднее расстояние между ближайшими соседями в морфологическом пространстве (например, между видами, самыми близкими по длине тела), с тем чтобы получить определенную меру степени упаковки видов, а тем самым и лимитирующего сходства. В основе этого метода лежит допущение, что морфологическое расстояние тесно связано с экологическим сходством. В большинстве подобного рода исследований использовались размеры трофических структур, размеры или вес тела консумента как показатели размеров потребляемой пищи. Это приемлемо, когда размеры консумента и его пищи коррелируют, т.е. чем крупнее данный признак, тем крупнее потребляемый пищевой ресурс.

Дифференциальное перекрывание ниш. Сходные виды могут сосуществовать, если их ниши, сильно перекрываясь по одному измерению ресурса, в значительной мере разделены по другому его измерению. Это явление известно под названием дифференциальное перекрывание ниш (рис. 3.5). В зависимости от цели исследования, можно считать, что ниши двух видов сильно перекрываются, следовательно, виды конкурируют между собой или, напротив, что эти виды занимают совершенно изолированные ниши. В настоящее время собрано большое количество данных о том, как два сходных вида могут сосуществовать и тем самым избегать конкуренции в результате дифференциального перекрывания ниш по двум или более градиентам среды.

Конкуренция за ограниченные ресурсы – главный фактор, определяющий упаковку видов, а тем самым и их разнообразие. Интенсивная эксплуатационная конкуренция за лимитирующие ресурсы может привести к вытеснению одного вида другим, а интерференционная конкуренция – закрыть доступ к ресурсам, имеющимся в достатке, что также ведет к исключению одного или нескольких конкурентов. Каждый вид, очевидно, имеет свои собственные, слегка отличные от наблюдающихся у других видов, предпочтения в отношении ресурсов и способности к их использованию. При нехватке ресурсов преобладания достигают лишь наиболее эффективные потребители, что обеспечивает максимальное использование ресурсов сообществом в состоянии равновесия. В масштабах короткого экологического времени это может привести к уменьшению равномерности в распределении видов, а в ряде случаев – к уменьшению их числа. В течение более длительных промежутков времени возрастающая межвидовая конкуренция нейтрализуется отбором, направленным на специализацию и дифференциальное перекрывание ниш. Это делает возможным устойчивое сосуществование возрастающего числа видов в результате поддерживаемой конкуренцией дифференциации ниш вплоть до какого-то теоретического предела, зависящего либо от числа наличных дискретных ресурсов, либо от максимально приемлемого перекрывания ниш, либо от того и другого вместе.

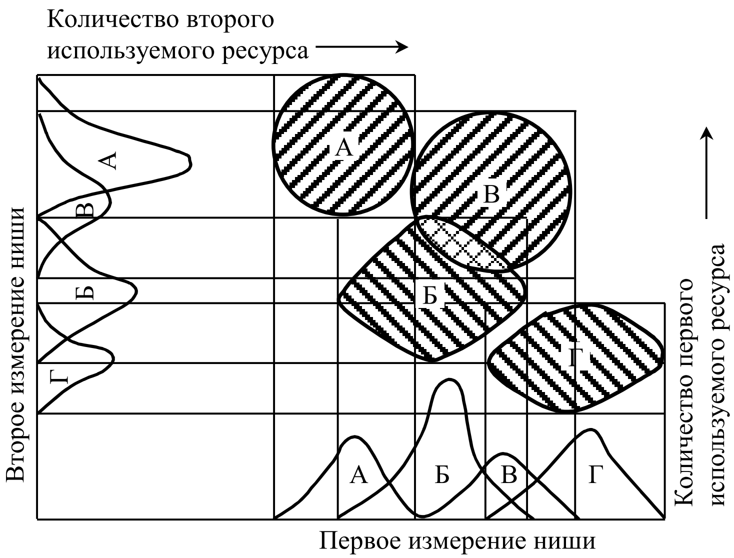


Рис. 3.5. Дифференциальное перекрывание ниш четырех видов (А–Г). Пары видов со значительным перекрыванием по одной оси ниши могут ослабить конкуренцию путем расхождения по другой ее оси. Заштрихованные участки – пространство ниши, занимаемое в результате каждым видом

Обычно ниши разделены по многим измерениям, и с увеличением числа конкурирующих видов им приходится расходиться по все большему и большему числу измерений, с тем чтобы сохранилось минимальное перекрывание ресурсов, необходимое для смягчения конкурентных взаимодействий в насыщенном сообществе. Таким образом, совпадение по одному измерению обычно, хотя и не всегда, коррелирует с обособленностью по другому, комплементарному, измерению. По-видимому, можно выявить два или три важных комплементарных измерения, по которым виды, входящие в данную гильдию, могут разделяться в пространстве ниш.

Величину многомерного перекрывания можно получить разными способами. Мак-Артур и Левинс (1967) предсказывают, что для сосуществования без конкуренции необходимо, чтобы перекрывание между соседними видами не превышало 54%; сходные величины приводят и другие авторы. В большинстве исследований, рассматривающих конкуренцию как главный механизм, определяющий структуру гильдии, у большей части пар видов суммарное перекрывание не достигает этой цифры. Анализ главных компонент дает лучшее приближение к сравнению ниш разных видов в терминах хатчинсоновского *n*-мерного пространства. В таких исследованиях выявляется более четкое разделение видов, что служит дополнительным подтверждением того, что комплементарность измерений ниш направлена на поддержание минимального перекрывания ресурсов и снижает конкурентные взаимодействия.

3.2.7. Хищничество и видовое разнообразие

Важным механизмом создания структуры сообщества, альтернативным механизму, основанному на разделении ресурсов, опосредованном конкуренцией, является хищничество. Если какой-либо фактор смертности оказывает очень сильное воздействие на самый конкурентоспособный или самый многочисленный вид (компенсаторная смертность), то конкурентное исключение может предотвращаться в течение неопределенно долгого времени. При этом становится возможным более сильное перекрывание ниш, а следовательно, и увеличение числа сосуществующих видов. В роли такого фактора смертности может выступить хищничество, которое, таким образом, влияет на характер локального видового разнообразия.

Введение зависящего от плотности жертвы хищничества в модель минимального разделения ниш привело к следующим результатам:

– минимальное расстояние, разделяющее ниши в присутствии хищника, никогда не бывает больше, чем в его отсутствие;

– с повышением эффективности хищничества минимальное разделение ниш уменьшается.

Точно так же выбор жертвы, зависящий от ее плотности, может привести к устойчивому равновесию в теоретических моделях двух конкурирующих видов жертвы, где прежде никакого равновесия не существовало. Для этого хищник должен обладать способностью к функциональным и численным реакциям на изменения плотности жертвы. Возможно, что переключение (непропорционально частое нападение на наиболее обильную жертву) будет иметь при этом более важное значение. Установлено, что переключение оказывает стабилизирующее влияние в системах «один хищник – *n* жертв» и представляет собой единственный механизм, способный стабилизировать взаимодействия в тех случаях, когда ниши жертв полностью перекрываются. Такую роль могут играть неспециализированные хищники. Предпочтение более специализированными хищниками доминантного конкурента действует таким же образом, как переключение хищника, и может стабилизировать теоретические взаимодействия в моделях, в которых прежде не существовало равновесия между видами жертвы, при условии, что их ниши в какой-то степени разделены.

Согласно гипотезе о воздействии хищничества избирательное потребление доминантного или самого обильного вида может поддерживать относительно высокое видовое разнообразие в масштабах экологического времени. Здесь возможна также система с обратной связью, при которой новые внедряющиеся виды хищников существуют за счет вновь появляющихся видов жертвы. Видообразование в масштабах эволюционного времени также возможно на основе таких процессов, как отбор, зависящий от плотности, когда давление хищничества стимулирует разнообразие жертв, давая преимущество более редким формам над более обычными. Хищничество может оказывать воздействие также на динамику и пространственное распределение популяции жертвы, что в свою очередь влияет на структуру и функции сообщества.

3.2.8. Взаимодействия хищник – жертва и видовое разнообразие

Имеется немало данных, свидетельствующих о способности хищников сократить популяцию жертвы до уровня, который ниже емкости среды. Подобное же действие могут оказывать некоторые паразиты. Хищник изымает из сообщества особей жертвы и тем самым высвобождает ресурсы, делая их доступными для других видов. Паразит может оказывать косвенное воздействие, понижая способность зараженных животных успешно конкурировать за имеющиеся ресурсы. Таким же образом могут действовать и паразиты растений.

Большая часть данных, свидетельствующих в пользу гипотезы о роли хищничества в возникновении видового разнообразия у животных, получена в результате изучения зоопланктона и сообществ приливно-отливной зоны. Видовое разнообразие экспериментальных сообществ, содержащих планктонных ракообразных, питающихся фитопланктоном, резко изменяется при введении в них рыб (Нейл, 1975). Избирательное уничтожение хищными рыбами доминирующих растительноядных рачков *Ceriodaphnia* привело к повышению численности популяций большинства постоянных членов сообщества, особенно мелких ветвистоусых рачков, и создало возможность для успешной инвазии молоди трех видов-«колонистов». Выедание рыбами и хвостатыми амфибиями крупных зоопланктонных видов обычно ведет к увеличению числа более мелких видов. К аналогичным результатам приводит выедание крупных планктонных организмов такими хищными ракообразными, как мизиды. Таким образом, избирательное хищничество снижает конкуренцию на нижних трофических уровнях, препятствуя доминантным конкурентам монополизировать пространство и поддерживая высокое видовое разнообразие. Хищников, которые действуют таким образом, называют ключевыми хищниками.

Хищничество не всегда ведет к повышению видового богатства на нижних трофических уровнях. Хотя хищники могут понизить плотность популяции жертвы, теоретически это не обязательно должно привести к меньшему потреблению жертвой данного ресурса – условие, необходимое для повышения видового разнообразия. В самом деле, ослабление внутривидовой конкуренции может стимулировать активность вида и его размножение, что повлечет за собой повышенное использование ресурса. Обилие консументов, предпочтение ими определенных видов растений и различная встречаемость последних также могут влиять на результат выедания. Например, там, где доминантный вид несъедобен, выедание просто усиливает его доминирование и еще больше снижает разнообразие.

Низкое давление хищников не оказывает заметного влияния на разнообразие сообщества. Точно так же, когда предпочитаемая жертва субдоминантна, удаление хищника не оказывает существенного воздействия на сообщество. Таких хищников называют «слабо взаимодействующими».

Теоретически в моделях «один хищник – две жертвы» эквивалентное выедание (отсутствие предпочтения того или иного вида жертвы) может повлиять на конкурентное сосуществование видов-жертв лишь в тех местах, где уже существует потенциальное равновесие. Разнообразие может возрасти только в таких условиях, когда у видов с меньшей конкурентоспособностью скорость роста популяции выше, чем у доминантов.

Хищничество на одном трофическом уровне может привести к «каскадным» эффектам на других уровнях, в результате чего разнообразие сообщества в целом понизится.

Наконец, хищничество оказывает влияние на разнообразие сообщества только при сильных конкурентных взаимодействиях на нижних трофических уровнях. Для того чтобы хищничество могло привести к высокому разнообразию сообщества, необходим целый ряд следующих предпосылок:

1. Наличие у хищника определенных предпочтений.

2. Наличие сильных и ассиметричных перекрестных конкурентных взаимосвязей на нижних трофических уровнях пищевой сети. Влияние верховных хищников в сообществах приливно-отливной зоны обусловлено тем, что хищник играет ключевую роль в регуляции численности своей предпочитаемой жертвы, которая, в свою очередь, доминирует в межвидовой конкуренции на нижнем трофическом уровне (т.е. оба способны к сильным взаимодействиям). Эти перекрестные связи между видами-жертвами приводят к тому, что изменения обилия хищников отражаются на перераспределении лимитирующего ресурса – пространства – и обусловливают резко выраженный каскад последствий, возникающих при вмешательстве в сообщество.

3. Ресурсы высвобождаются под действием хищника и могут обеспечить существование новых видов или более многочисленных популяций уже присутствующих видов.

4. Высокая интенсивность хищничества. Только в том случае, когда обратная связь между хищником и предпочитаемой им доминантной жертвой достаточно сильна, хищничество может поддерживать разнообразие сообщества. Низкая интенсивность хищничества обычно оказывает действие только в сочетании с другими формами стресса. Так, сочетания поражения нематодами и межвидовой конкуренции достаточно для понижения конкурентного преимущества овса и возможного сосуществования с ним ячменя. Однако интенсивное хищничество может вести к уменьшению видового разнообразия: разнообразие видов достигает максимума при умеренном выедании (т.е. кривая зависимости между интенсивностью выедания и видовым богатством имеет горб).

5. Для поддержания высокого видового разнообразия контролирующий фактор должен не только стать главным регулятором численности доминантного вида-жертвы, но и сам должен регулироваться наличием жертвы и не лимитироваться собственными хищниками. Если же такое лимитирование происходит, то хищничество на этом высоком уровне приведет к уменьшению разнообразия на более низких уровнях в результате каскадных эффектов. Вырисовывается закономерность: четное число сильно взаимодействующих между собой уровней ведет к большему разнообразию и сложности, а нечетное – к доминированию и простоте.

3.2.9. Взаимодействия между конкуренцией и хищничеством

Коннелл (1980) полагает, что в очень суровых условиях среды численность популяций падает ниже тех уровней, при которых они конкурируют, в результате изреживания под действием физических факторов, а при благоприятных условиях – под действием хищников. В промежуточных умеренно суровых условиях, когда смертность от прямых стрессовых воздействий физических факторов и от природных врагов должна понижаться, вероятность того, что численность популяций достигнет уровней, при которых они будут конкурировать, возрастает, и, таким образом, конкуренция становится главной организующей силой (рис. 3.6). Существует сходное объяснение (Грайм, 1973) для разнообразия растений в тех случаях, когда усиление стрессовых воздействий (физических факторов или растительноядных животных) понижает жизнеспособность видов-доминантов, повышая в результате видовое разнообразие.

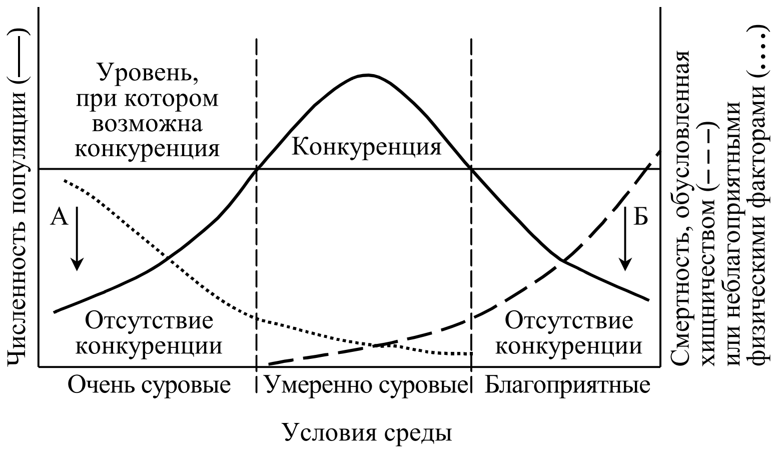


Рис. 3.6. Классификация механизмов организации сообщества, основанная на воздействии внешних факторов:
А – популяции, численность которых ограничивается неблагоприятными физическими факторами; Б – популяции, численность которых ограничивается интенсивным хищничеством

В этой схеме не учитывается то обстоятельство, что в благоприятных средах межвидовая конкуренция, вероятно, возрастает, а это может ослабить их влияние на популяции жертв. С энергетической точки зрения большее разнообразие видов-жертв может обеспечить существование большего числа хищников. Это могло бы означать, что относительное влияние хищничества на доминантные по конкурентоспособности виды не больше, чем на любые другие виды-жертвы. Само хищничество также может стать умеренным в сложных средах, например в тропиках. В простых средах хищникам легко найти жертву, поэтому они, несомненно, могли бы контролировать видовое разнообразие. При увеличении сложности среды находить виды-жертвы становится все труднее в связи с обилием убежищ, и трофическая эффективность хищника, по-видимому, понижается.

Более удовлетворительный подход к роли хищничества и конкуренции в видовом разнообразии основан на трофическом статусе и общих размерах. Конкуренция может регулировать число видов в гильдии только в том случае, если размер популяции соответствует емкости среды или близок к ней. Особенно это касается животных, имеющих высокий трофический статус, которые мало зависят от флуктуаций среды. В отличие от этого для гильдий, имеющих более низкий трофический статус, более вероятна регуляция видового разнообразия хищниками.

Существует также точка зрения, согласно которой в сообществах с небольшим числом трофических уровней конкуренция должна быть самым важным всеобщим организующим фактором, однако по мере увеличения числа трофических уровней и числа видов начинает относительно возрастать роль хищничества. В качестве общего правила это положение гораздо менее приемлемо потому, что хищничество может увеличить или уменьшить разнообразие сообщества в зависимости от числа взаимодействующих трофических уровней.

Несмотря на то, что разногласия по вопросу об относительном значении конкуренции и хищничества в ассамблеях животных, очевидно, остаются неразрешенными, среди экологов растений достигнуто некоторое единство мнений. Растительноядные беспозвоночные в некоторых случаях могут определять физиономические черты растительности, но влияние большинства растительноядных, в особенности насекомых, скорее всего ограничивается их воздействием на популяцию какого-либо одного вида. Поэтому для большинства растительных систем лимитирующим фактором служат ресурсы. Таким образом, межвидовая конкуренция представляется главной движущей силой, определяющей распространение и обилие растений, но направление, в котором она действует, может до некоторой степени зависеть от растительноядного животного. Хищники увеличивают доступность ресурсов, снижая активность доминантного консумента. Если такое регулируемое хищниками сообщество еще не насыщено видами, оно может стать насыщенным особями. То есть многие ресурсы, высвободившиеся в результате сокращения численности доминантного конкурента, будут потребляться другими видами сообщества. Таким образом, несмотря на активность хищников, ресурсы будут использоваться в полной мере, конкуренция между сосуществующими видами будет продолжаться, и именно эта конкуренция останется главным фактором, контролирующим на данной стадии разнообразие видов, их относительное обилие и использование ими ресурсов. При некоторых обстоятельствах хищничество может изменить число сосуществующих видов, но такие сообщества открыты для новых видов не постоянно: достигается некий предел, когда ресурсы снова используются полностью. Даже если хищники удерживают численность всех видов на низком уровне, быстро переключаясь с одной жертвы на другую, высвобождающиеся ресурсы дают возможность сообществу пополняться новыми видами только до тех пор, пока потребление ресурсов не достигло своего предела. При этом использование ресурсов и обилие видов снова будут регулироваться конкуренцией.

Конкуренцию принято считать главным организующим фактором сообщества как по причине ее воздействия на размеры ниши, так и на основе определения ниши, данного Хатчинсоном. Здесь возникает противоречие, поскольку альтернативным механизмом может быть и хищничество. Однако ему нельзя отводить главную роль ни в непосредственном определении видового состава, ни в предотвращении конкурентного исключения. Возможно, что в некоторых случаях хищничество служит фактором, непосредственно регулирующим видовое богатство, но конечным механизмом, создающим структуру сообщества, все же остается конкуренция.

3.2.10. Насыщение сообществ

Увеличение числа измерений ниши, дифференциальное перекрывание ниш, а иногда и хищничество – все это ведет к снижению конкуренции, а, следовательно, может обеспечить сосуществование большего числа видов, чем было бы возможно в иных условиях. Новые виды, пополняющие сообщество, сами становятся ресурсами и расширяют спектр ресурсов для других видов. Прибавление новых видов к сообществу представляет самонарастающий эволюционный процесс – разнообразие порождает разнообразие. Уровень видового разнообразия обладает положительной обратной связью: чем больше видов, тем ниже скорость вымирания и тем быстрее возрастает разнообразие. При этом должен существовать какой-то предел ввиду ограниченности мировых запасов энергии и питательных веществ, а также потому, что каждый вид должен состоять, по крайней мере, из одной особи. Если большинство видов достаточно сильно конкурирует или конкурировало в прошлом со сходными видами или существуют пределы, ограничивающие суммарное пространство ниши, ширину ниши и допустимое перекрывание, то можно предсказать, что число видов, существование которых может обеспечить данная область, ограниченно. Поскольку, в конечном счете, видовое богатство контролируется конкуренцией, можно ожидать, что существуют сообщества, достигшие такого насыщения и стабильного разнообразия. Например, видовое богатство возрастает с увеличением обследуемой площади, однако при этом наблюдается не простая линейная зависимость, а угасающее возрастание числа видов, и после того, как эта площадь достигает известного предела, число обнаруживаемых новых видов невелико. Этот предел часто рассматривают как минимальную площадь, занимаемую сообществом. Зависимость между числом видов (S) и площадью (A) описывается степенной функцией

$$S=CA^{Z}, \qquad (3.2)$$

где C – константа, а Z – тангенс угла наклона линии регрессии на графике $logS/logA$; значения Z чаще всего лежат в пределах от 0,12 до 0,17. Это связано с тем, что небольшие участки материка содержат примерно столько же видов, сколько и крупные. Такие небольшие континентальные области непрерывно обмениваются видами с окружающими областями, поэтому им не угрожает низкий уровень иммиграции, характерный для изолированного острова. Малые значения Z могут также объясняться низкой скоростью вымирания.

В подобных ситуациях возникает проблема масштабов и различных способов количественной оценки разнообразия видов. Небольшое по площади местообитание обеспечивает существование некоторого числа видов, которое называется α или внутриэкотопным разнообразием. При сравнениях местообитаний друг с другом или вдоль градиентов факторов среды обнаруживаются изменения видового состава. Степень или скорость изменения называется β или межэкотопным разнообразием. Значение β можно получить из уравнения

$$\beta = \frac{S_c}{\overline{S}}, \qquad (3.3)$$

где S_c – общее число видов, встречающихся на данной трансекте или в ряде выборок (каждый вид учитывается один раз), а $\overline{S}$ – среднее число видов на одну выборку. Для одной выборки β=1, а для двух выборок, не содержащих ни одного общего вида, β=2. Зависимость число видов – площадь на материках и асимптотическое приближение ее к пределу указывают на существование некоего максимума видового богатства.

Существует несколько моделей насыщения материков. Одна из них принадлежит Мак-Артуру (1969). В ней полагается, что:

- скорость видообразования будет возрастать, но возрастать все медленнее по мере роста разнообразия (чем больше видов, тем больше возможностей для появления новых видов, но тем у них меньше шансов достигнуть успеха);

• при этом скорость вымирания будет расти все быстрее (в результате конкуренции и снижения величины популяций по мере возрастания упаковки сообщества).

Эти предпосылки, по мнению автора, должны приводить к стабильному разнообразию (рис. 3.7-а). Альтернативная теория (Майр, 1963) утверждает, что видообразование происходит быстрее при наличии большего количества незанятых ниш, приводя к адаптивной радиации. В большинстве случаев географически изолированные популяции не дают начала новым видам, потому что им не удается найти незанятые ниши. Если это так, то рис. 3.7-а не соответствует действительности, и возможно, что с увеличением разнообразия скорость видообразования снижается (рис. 3.7-б). Эти различия можно объяснить, допустив два вида видообразования:

а) конкурентное видообразование, идущее с большей скоростью при истощении биоты, поскольку при этом имеется много возможностей для видообразования путем конкурентного взаимодействия, но замедляющееся при росте разнообразия;

б) географическое видообразование: чем больше видов, тем больше популяций, изолированных географически, и тем быстрее формируются новые виды, т.е. скорость возрастает при росте разнообразия. В любом случае предсказывается достижение стабильного разнообразия, а не бесконечное его увеличение.

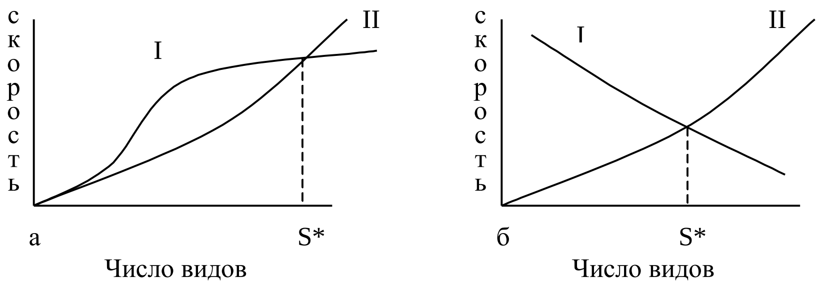


Рис. 3.7. Две модели насыщения видами материков:
а – модель Мак-Артура; б – альтернативная модель, основанная на взглядах Майра; I – кривая видообразования; II – кривая вымирания; S* – равновесное число видов

Таким образом, число видов, сосуществующих в любой точке пространства (α-разнообразие), ограничено в экологическом смысле, а во многих сообществах – на протяжении длительных периодов эволюционного времени. Если изучать видовое разнообразие в более широких масштабах, такое ограничение становится менее очевидным вследствие горизонтального замещения видов (β-разнообразие). Это может обеспечить возможность сосуществования в пределах данной области большего числа видов, чем то, которое способно реально разделять ресурсы какой-то общей точки в пространстве. Такое межэкотопное разнообразие может непрерывно возрастать в течение эволюционного времени, однако смена видов по горизонтали внутри данной области возможна лишь до известного предела (минимальная площадь, способная обеспечить существование сообщества, или величина популяции, необходимая для выживания вида), о чем свидетельствует характер зависимостей, изображенных на рис. 3.7. Расширив обследуемую площадь еще больше, чтобы в нее вошли сходные местообитания, в различных зоогеографических областях обнаруживается дальнейшая смена видов, обусловленная историческими причинами (γ-разнообразие). Здесь тоже должен существовать какой-то предел, но вследствие продолжающихся изменений, связанных с дрейфом континентов и эволюцией, установить этот предел трудно.

Внутри отдельных местообитаний существует насыщение видами, обусловленное двумя механизмами. Во-первых, механизмом, который не связан с взаимодействиями, а зависит только от свойств, внутренне присущих данному местообитанию и данному виду. Во-вторых, механизмом, связанным с взаимодействием, при котором видовое разнообразие ограничивается прямой или диффузной конкуренцией внутри местообитания. На протяжении эволюционного времени первый тип равновесия может постепенно сменяться вторым.

Если сообщества могут достигать насыщения, то различия в видовом разнообразии просто зависят от возраста сообщества и от уровня насыщения, т.е. от того, до какой степени все имеющиеся ресурсы используются максимально возможным числом видов. Однако насыщенные сообщества также могут различаться по видовому разнообразию. Мак-Артур вывел простое уравнение для приближенного описания видового разнообразия животных (D_s), использующих какой-либо ресурс:

$$D_s = \frac{D_r}{D_u}(1 + C\bar{\alpha}), \tag{3.4}$$

где D_r – все разнообразие ресурсов, используемых сообществом в целом; D_u – ширина ниши каждого вида (предположительно одинаковая); C – число потенциальных конкурентов или соседей в пространстве ниш – выражение размерности данного местообитания; $\bar{\alpha}$ – средний коэффициент конкуренции или среднее перекрывание ниш.

Эта формула предсказывает, что насыщенные сообщества могут различаться по видовому разнообразию под влиянием трех потенциально взаимодействующих факторов:

1. Разнообразие имеющихся ресурсов (по всему гиперпространству ниш). Область с более широким спектром доступных ресурсов (большим D_r и C) может содержать больше ниш, а, следовательно, обеспечивать существование большего числа видов, чем область с более узким спектром ресурсов.

2. Ширина ниш или степень специализации обитающих в ней видов (D_u). Если виды используют небольшие части общего гиперпространства ниш (т.е. занимают небольшие ниши), то число сосуществующих видов может увеличиваться.

3. Средняя степень перекрывания между нишами разных видов ($\bar{\alpha}$). Более сильное перекрывание (уменьшение размера ниши, доступной исключительно данному виду) ведет к большему разнообразию, даже если два сообщества сходны по пунктам 1 и 2.

В терминах использованной ранее простой модели сообщества в виде «шара и ящика» видовое богатство может быть увеличено, когда имеется более объемистый «ящик» (1) или когда «шары» мельче (2) или мягче (3).

3.3. Структура сообщества – закономерности и правила

3.3.1. Закономерности в структуре сообщества

В организации сообществ установлен ряд закономерностей:

1. Межвидовые взаимодействия (конкуренция и хищничество) обычно сужают нишу каждого вида по сравнению с размерами, допускаемыми его физиологией или морфологией.

2. Внутривидовая конкуренция ведет к расширению ниши, а межвидовая к ее сужению.

3. Если виды очень сходны и если при сокращении ресурсов конкуренция усиливается, то возможно конкурентное исключение видов из областей перекрывания ниш.

4. В пределах отдельных видов и между видами, относящимися к одной гильдии, могут существовать экологически значимые различия в размерах, если они разделены по градиенту одного из главных ресурсов.

5. В гильдиях, обедненных видами, происходит компенсация плотностью или экологическое высвобождение.

6. У сосуществующих видов ниши разделены: у животных – по трем главным измерениям (временному, пространственному, трофическому), у растений – обычно по двум (временному и пространственному).

7. Хищничество оказывает более сильное воздействие на ассамблеи растений, чем на ассамблеи животных, однако его влияние на видовое

разнообразие зависит от числа сильно взаимодействующих трофических уровней.

8. Сообщества достигают насыщения видами на протяжении экологического и эволюционного времени.

9. На макрогеографическом уровне виды замещают друг друга горизонтально, при переходе от одного местообитания к другому.

10. В большинстве групп организмов наблюдаются изменения видового богатства по градиенту географической широты.

11. В обширных областях обитает больше видов, чем в малых.

12. В местообитаниях со сложной средой видов больше, чем в простых местообитаниях.

13. В устойчивых местообитаниях видов больше и они занимают более узкие ниши, чем в кратковременных.

14. Распределение по обилию в больших совокупностях таксономически близких видов обычно соответствует каноническому логарифмически нормальному распределению.

15. В небольших совокупностях близких видов или в простых сообществах относительное обилие видов обычно контролируется одним доминирующим фактором или ресурсом.

16. Лишь немногие виды встречаются во многих местообитаниях и притом редко. Для большинства видов характерно среднее обилие в среднем числе сообществ, высокое обилие в нескольких или низкое во многих сообществах.

Теория конкуренции позволяет дать наиболее правдоподобное объяснение большинству из этих закономерностей. Однако она исходит из того, что в основе структуры сообществ лежат взаимодействия между видами.

3.3.2. Правила организации сообществ

В организации сообществ вырисовывается несколько «основных правил»:

1. Видовое разнообразие сообществ не может возрастать бесконечно. В равновесных сообществах максимальное число видов определяется общим пространством ниш и лимитирующим сходством между видами, зависящими от конкуренции. Степень лимитирующего сходства может быть выявлена только в тех случаях, когда виды дифференцируются лишь по какой-либо одной оси ресурса. Виды животных, потребляющие сходные пищевые ресурсы, должны различаться в 1,3 раза по общим размерам или в 2 раза по массе. Если разделение видов в пространстве ниш происходит по более чем одному измерению, то лимитирующее сходство сохраняется, но его трудно выразить количественно. Дробление пространства ниш может идти в пределах, устанавливаемых максимальным или оптимальным уровнем использования видами

имеющихся ресурсов. В неравновесных сообществах насыщение видами и лимитирующее сходство выявить трудно. В любом случае предельное число видов будет зависеть от соотношения между наличием ресурсов и потребностью в них, а также от того, является ли состояние неравновесия постоянной или преходящей чертой данного сообщества.

2. Увеличение разнообразия ресурсов происходит в условиях ослабленной межвидовой конкуренции (например в обедненных биотах).

3. Как правило, разделение ниш между парами видов увеличивается при возрастании видового разнообразия. Это сопровождается уменьшением размеров ниши и видовой популяции вплоть до некоторого предела, определяемого минимальными потребностями в ресурсах и (или) минимальной величиной популяции. Однако диффузная конкуренция и суммарное перекрывание между видами с увеличением видового разнообразия действительно возрастают. Увеличение разделения между парами видов с увеличением разнообразия, возможно, уравновешивает общий уровень конкуренции, испытываемой любым видом, что снова подразумевает существование лимитирующего сходства между видами.

4. О насыщении сообщества свидетельствует также зависимость число видов – площадь, а постоянство параметра Z (наклон кривой, описывающей эту зависимость) в различных ситуациях можно считать еще одним правилом, которому подчиняется организация сообществ.

5. При рассмотрении числа видов в сообществе очень большое значение имеют число ниш, их ширина и перекрывание. Поэтому любые факторы, влияющие на эти параметры, следует считать важными факторами организации сообщества. Неоднородность среды позволяет в значительной степени объяснить видовое богатство просто потому, что от нее зависит число потенциальных ниш, а также степень дробления их пространства.

Контрольные вопросы

1. Показатели, характеризующие структуру сообщества.
2. Количественное определение ниши. Фундаментальная и реализованная ниша.
3. Ширина ниши. Признаки полиморфных и мономорфных видов.
4. Возможные перекрывания фундаментальных ниш двух видов.
5. Последствия частичного перекрывания ниш двух видов.
6. Механизм диффузной конкуренции.
7. Дифференциальное перекрывание ниш.
8. Влияние хищничества и конкуренции на видовую структуру сообществ.
9. При каких условиях среды обитания в сообществах не наблюдается конкуренции?
10. Внутри- и межэкотопное видовое разнообразие.

Глава 4
МЕТОДЫ АНАЛИЗА И МОДЕЛИРОВАНИЯ СООБЩЕСТВ И ЭКОСИСТЕМ

4.1. Методы моделирования

В настоящее время в моделировании можно выделить три независимых направления:

1. *Стохастическое моделирование* (построение экспериментальной статистической модели). Этот вид моделирования применяется при неизвестных параметрах системы, причем входные параметры и переменные состояния системы изменяются во времени. Большое внимание уделяется именно временным и пространственным изменениям входных переменных и переменных состояния, а также параметрам системы и их первоначальной неопределенности. Для получения решения требуются такие подходы, которые были бы совместимы со стохастической природой входных переменных и уравнений состояния. До тех пор, пока процессы, происходящие внутри системы (подсистемы) и описывающие их дифференциальные уравнения, остаются неизвестными, в математическом моделировании можно использовать статистические методы.

2. *Аналитическое моделирование* (построение теоретической модели). В данном случае предполагается, что входные переменные и уравнения состояния известны.

3. *Анализ адаптивных систем и моделирование* (применение самообучающихся алгоритмов к информации, получаемой в процессе построения теоретических и экспериментальных моделей). Использование адаптивных методов позволяет производить анализ систем, при котором степень первоначальной неопределенности уменьшается благодаря информации, поступающей в ходе процесса.

На практике часто имеют место комбинации этих направлений, когда, например, в аналитических моделях некоторые коэффициенты могут рассматриваться как стохастические.

Способ, выбираемый для анализа систем, зависит от цели моделирования. Модели делятся на две категории: дескриптивные модели и модели поведения.

Дескриптивная модель позволяет получить информацию о взаимосвязях между наиболее важными переменными экосистемы. Входные переменные модели определяются на основе причинных связей или экспериментальных данных. Большинство дескриптивных моделей состоит из линейных и нелинейных регрессий.

Модели поведения используются для описания поведения систем во время переходного состояния. При этом осуществляется запись стационарного и нестационарного состояний системы (поведение при переходе от одного состояния к другому). Описать переходное поведение можно различными способами:

– путем исследования структуры сигналов на входе и выходе системы (отношение «вход-выход»). Входные переменные влияют на выходные переменные через неизвестные переменные состояния. Изменчивость выходных переменных называется эффектом системы. Указанный метод находит практическое применение только тогда, когда речь идет о стохастических сигналах;

– путем изучения реакции системы на особые проверочные сигналы, подаваемые на вход. Такой подход к описанию системы считается оптимальным, однако возмущение системы, не имеющей обратной связи, с помощью проверочных сигналов является очень сложным;

– с помощью внутренней структуры, поведения системы в начале поступления сигнала и описания этой структуры (например дифференциальными уравнениями). Этот метод, будучи наиболее приспособленным для получения информации, в частности о сложных линейных системах, существенно зависит от числа энергетических источников, числа обратных связей в системе и других факторов.

Поведение линейных систем под воздействием проверочных сигналов можно определить статическими и динамическими характеристиками. Статическое поведение описывает выровненное устойчивое состояние под воздействием постоянных возмущающих переменных, тогда как динамические характеристики описывают поведение системы в переходном состоянии, т.е. под воздействием постоянно меняющихся возмущающих переменных.

4.1.1. Стохастическое моделирование

Стохастическое моделирование основано на методах теории вероятностей и математической статистики и включает статистические и динамические методы. Различие между двумя методами зависит от того, включена ли в анализ временна́я переменная.

Статистические методы (не учитывающие время в качестве переменной) включают в себя простую и множественную линейную и нелинейную корреляцию и регрессию, дисперсионный, дискриминантный и факторный анализы, методы оценки параметров (методы прямой оценки, такие как оценки Маркова, оценки максимального правдоподобия и оценки Байеса).

Динамические методы (с учетом временно́й переменной; анализ временных рядов) включают анализ Фурье, корреляционный и спектральный анализы, весовые и передаточные функции. Эти методы известны под общим названием динамическая статистика.

Статистические операции. Часто при изучении водной экосистемы возникает вопрос: существует ли взаимосвязь между различными переменными такой экосистемы. Ответ можно получить посредством корреляционного и регрессионного анализов в зависимости от цели и задачи исследования. Вид связи между переменными выражается регрессионным уравнением, а ее интенсивность с помощью корреляции. Подробно статистические методы изложены в учебном пособии «Системная экология» (Дулепов и др., 2004).

Регрессия и корреляция. Регрессионный анализ необходим для решения задач, в которых стохастические зависимости описываются функциями с одной или несколькими переменными, определяемыми как независимые.

Множественная регрессия имеет одну зависимую и несколько независимых переменных. Наиболее часто используются линейные уравнения. Во многих ситуациях для описания экологических процессов используются полиномы *n*-го порядка (полиномиальные регрессии): $y = \sum_{i=0}^{n} a_i x^i$, где n – порядок полинома.

В результате замены переменных x функциями синуса и косинуса в простейшем случае получаем уравнение

$$y = a + b_1 \sin x + b_2 \cos x\,. \qquad (4.1)$$

Оно представляет собой простой вид *периодической регрессии,* или так называемый полином Фурье. В более широком смысле этот метод называется анализом Фурье. Данный вид регрессии часто используется для определения периодической направленности изменений.

Корреляционный анализ применяется с целью установления степени корреляции между двумя или бóльшим числом стохастических переменных, а также для определения степени стохастической зависимости, существующей между ними. Последняя может быть описана с помощью коэффициентов корреляции. Двусторонние и многосторонние зависимости описываются простыми коэффициентами, а также коэффициентами частной и множественной корреляции. Коэффициент частной корреляции также необходим при выборе переменных, имеющих связь с составляющими элементами экосистемы. Если предполагается многомерное нормальное распределение, оно будет служить мерой линейной зависимости двух случайных переменных x_j, x_k при устранении влияния всех остальных случайных переменных $x_1,\ldots,x_{j-1},\ x_{j+1},\ldots,x_{k-1},x_{k+1},\ldots,x_n$. Квадрат коэффициента корреляции называется показателем эффективности или коэффициентом детерминации. Обычно его выражают в %, умножая его значение на 100. Это мера дисперсии, которая обусловлена за счет регрессии, и чем она больше, тем теснее связь.

Дисперсионный и ковариационный анализы. Дисперсионный анализ представляет собой способ качественного и количественного изучения влияния одной или нескольких переменных на результаты эксперимента. В тех моделях, где это влияние имеет фиксированный характер, обычно сравниваются лишь средние значения нескольких случайных выборок. Однако в моделях, учитывающих случайные эффекты, сами влияющие факторы рассматриваются как случайные выборки из множества возможных проявлений этих факторов. Такая же картина может иметь место при рассмотрении непрерывно контролируемых качественных параметров, например искусственно созданной аквасистемы.

Ковариационный анализ можно использовать при количественном изучении различной степени воздействия одной или нескольких переменных на экспериментальные данные, и при этом обязательно учитывается влияние дополнительных случайных переменных. По существу этот метод позволяет объединить дисперсионный и регрессионный анализы, каждый из которых относится к моделям с фиксированными воздействиями.

Дискриминантный и факторный анализы. Дискриминантный анализ применяется для разделения или классификации объектов и их связи с двумя или более совокупностями (группами, популяциями). Фактически он представляет собой способ разделения. Последнее выполняется на основе анализа количественных характеристик и учета дискриминантной (разделительной) функции, с которой связано принятие решения о проведении классификации. Дискриминантный анализ позволяет, например, классифицировать водные экосистемы в зависимости от количества кислорода или фосфора.

Факторный анализ используется для изучения соотношений между случайными переменными, обусловленных общими причинами, или факторами, а также с целью вывода этих соотношений. При этом особое внимание уделяется точечному оцениванию параметров.

Методы дискриминантного и факторного анализов были разработаны в общей экологии с целью типизации пространственно-временных отношений различных видов и переменных окружающей среды. В результате информация о различных биологических видах или переменных окружающей среды, представляющих собой кластеры в пространстве и времени, поступает одновременно. Это обстоятельство является отправной точкой, от которой начинается использование *кластерного анализа* для обнаружения, например, источников загрязнения или взаимосвязи между переменными трофического состояния.

Методы оценки параметров. Обычно очень трудно получить параметры модели из реальной системы. Поэтому они оцениваются исходя из данных наблюдений за различными переменными и их выборочных статистических функций. Это могут быть точечные оценки или оценки на интервалах, в зависимости от того, определяется ли сам параметр или интервал, в котором он находится. Подобные оценки удобно получать с помощью статистических компьютерных пакетов программ.

Методы оценки параметров, в основе которых лежит метод наименьших квадратов, оказались наиболее эффективными в экологии. Существуют прямые и косвенные методы.

Прямые методы (регрессия, марковская оценка, оценка Байеса, оценка максимального правдоподобия) позволяют оценивать параметры в течение одного шага исходя из массива данных измерений на входе и выходе сигнала из системы. При использовании этих методов могут возникнуть проблемы их применимости на практике ввиду необходимости иметь априорную информацию о системе, играющей роль исходной модели. Косвенные методы делятся на рекурсивные и нерекурсивные. Рекурсивные методы позволяют подстраивать модель с данной структурой к системе пошаговым способом, при котором каждая новая группа данных измерений дает возможность сделать очередной шаг в подстройке модели. Все параметры должны подвергаться переоценке на протяжении всего процесса.

Динамические методы. Эти методы обычно фигурируют в научной литературе под названием «анализ временных рядов». Это стандартные методы статистической оценки результатов, полученных на основе измерений и наблюдений переменных в определенном временном интервале.

4.1.2. Аналитическое моделирование

Основу аналитического моделирования составляют дифференциальные уравнения, описывающие причинно-следственные связи в экосистеме (пространственные модели состояния). Вывод таких уравнений является первым этапом аналитического моделирования, за которым следуют оценка параметров, имитация и испытание модели.

4.1.2.1. Разностные и дифференциальные уравнения

Дифференциальные уравнения содержат производные от переменных состояния по независимым переменным (например времени, глубине, расстоянию). Разностные уравнения описывают изменения переменных состояния, вызываемые незначительными различиями независимых переменных. Разностные уравнения используют также при описании дискретных процессов.

Разностное уравнение превращается в дифференциальное уравнение, если скорость роста измеряется как мгновенная скорость ($\Delta t \to 0$). Уравнения считаются параметризованными, если коэффициентам, которые раньше были представлены в рассматриваемых уравнениях в виде символов, придаются действительные значения. Решения (интегралы) таких уравнений определяются как аналитические. Решение представляет собой алгебраическое уравнение, позволяющее получить значения переменных состояния в момент времени *t*.

Прежде чем приступить к обсуждению более сложных моделей динамики численности, остановимся кратко на основных сведениях о дифференциальных уравнениях.

Дифференциальное уравнение вида

$$\frac{dx}{dt} = f(x), \qquad x \in R \tag{4.2}$$

правая часть которого не зависит от *t*, называется *автономным.* Дифференцируемая функция $x(t)$ называется *решением* уравнения (4.2), если при подстановке в это уравнение $x(t)$ и $x'(t)$ получается истинное тождество, т.е. $x'(t)=f(x(t))$ при всех значениях *t*. Решение $x(t)$ представляется геометрически графиком функции $x(t)$. Этот график определяет интегральную кривую уравнения (4.2) на плоскости *t*, *x*. Свойства решений $x(t)$ в большей степени определяются видом функции $f(x)$.

Если функция $f(x)$ обращается в ноль при некотором значении *x*, например при $x=\alpha$, то постоянная функция $x(t)\equiv\alpha$ является решением уравнения (4.2). Действительно, $x'(t)\equiv 0$ и $f(x(t))\equiv f(\alpha)=0$. Такое решение называется равновесным решением или стационарной точкой (положением равновесия, равновесной точкой, неподвижной точкой, особой точкой) уравнения (4.2). Процесс, начавшийся в состоянии *a*, всегда в нем остается. Если функция $f(x)\neq 0$, то решения $x(t)$ либо возрастают ($f(x)>0$), либо убывают ($f(x)<0$) с ростом *t*.

Эти свойства решений удобнее изображать на оси *x*, чем на плоскости *t*, *x*. Если $f(x)\neq 0$, для $x\in(a,b)$, то на этом интервале рисуется стрелка, показывающая направление изменения *x*. Если $f(a)=0$, то стационарное решение $x(t)\equiv a$ изображается точкой $x = a$. Такое геометрическое изображение качественного поведения решений уравнения (4.2) называется фазовым портретом. Ось *x* называется при этом фазовой прямой, а точка $x(t)$ – фазовой точкой.

Если решение *x* нестационарное, то оно должно быть либо возрастающим, либо убывающим; таким образом, если число неподвижных точек конечно, то может существовать только конечное число «различных» фазовых портретов. Под словом «различные» подразумевается «отличающиеся набором областей, в которых *x* возрастает или убывает». Рассмотрим случай одной неподвижной точки x = а (рис. 4.1).

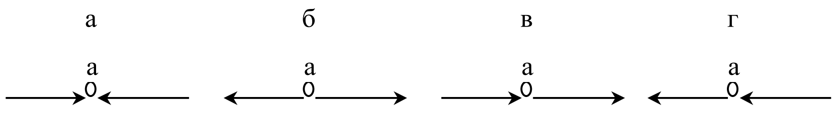


Рис. 4.1. Четыре возможных фазовых портрета
в случае одной изолированной неподвижной точки (пояснения в тексте)

На каждой из получаемых полупрямых $(x<a,\ x>a)$ функция f может быть либо положительной, либо отрицательной. Следовательно, фазовый портрет должен соответствовать одному из четырех случаев, изображенных на рис. 4.1. Рассмотрим каждый из этих случаев.

а) Функция $f(x)$ положительна при $x<a$ и отрицательна при $x>a$. Следовательно, нестационарное решение $x(t)$ возрастает при $x<a$ и убывает при $x>a$. Неподвижная точка $x = a$ в этом случае является устойчивой (при малом отклонении начального условия от равновесного фазовая точка с течением времени приближается к положению равновесия – неподвижной точке) и называется аттрактором.

б) Функция $f\ (x)$ отрицательна при $x<a$ и положительна при $x>a$. Следовательно, нестационарное решение $x(t)$ убывает при $x<a$ и возрастает при $x>a$. Неподвижная точка x = а в этом случае является неустойчивой (при малом отклонении начального условия от равновесного фазовая точка с течением времени удаляется от положения равновесия – неподвижной точки) и называется репеллером.

в) Функция $f(x)$ положительна при $x<a$ и при $x>a$. Следовательно, нестационарное решение $x(t)$ возрастает и при $x<a$, и при $x>a$. Неподвижная точка $x = a$ в этом случае является полуустойчивой (при малом отклонении начального условия от равновесного в одну сторону фазовая точка с течением времени приближается к неподвижной точке, а при малом отклонении в другую сторону – удаляется от нее) и называется шунтом.

г) Функция $f(x)$ отрицательна при $x<a$ и при $x>a$. Следовательно, нестационарное решение $x(t)$ убывает при $x<a$ и при $x>a$. Неподвижная точка $x = a$ в этом случае также является полуустойчивой – шунтом.

Заметим, что соображения, которые использовались для случая одной неподвижной точки, сохраняют свою силу, если точка $x=a$ одна из многих неподвижных точек на фазовом портрете. Другими словами, качественное поведение x в окрестности любой неподвижной точки должно быть таким же, как в одном из случаев, изображенных на рис. 4.1. Говорят, что это поведение определяет характер (вид, природу) неподвижной точки, и для его описания применяют термины, которые были приведены выше.

Из сказанного следует вывод, что фазовый портрет любого автономного уравнения полностью определяется видом его неподвижных точек. Говорят, что два дифференциальных уравнения вида $x' = f\ (x)$ качественно эквивалентны, если они имеют равное число неподвижных точек одинакового характера, расположенных в одинаковом порядке на фазовой прямой.

Для того чтобы описать качественные свойства решений $x(t)$ (интегральных кривых) автономного уравнения (4.2), поступают обычно следующим образом. На одном рисунке изображают три различных графика:

1) график функции $f(x)$ (причем за ось x принимают вертикальную прямую, направленную вверх, а за ось $f(x)$ – горизонтальную прямую, направленную справа налево),

2) фазовый портрет, фазовой прямой которого служит ось x,

3) график функции $x(t)$, вид которой (в нормальных координатах) качественно определяется по первым двум графикам.

В качестве примера рассмотрим построение этих графиков для уравнения нормального размножения $x' = \varepsilon x$ (рис. 4.2).

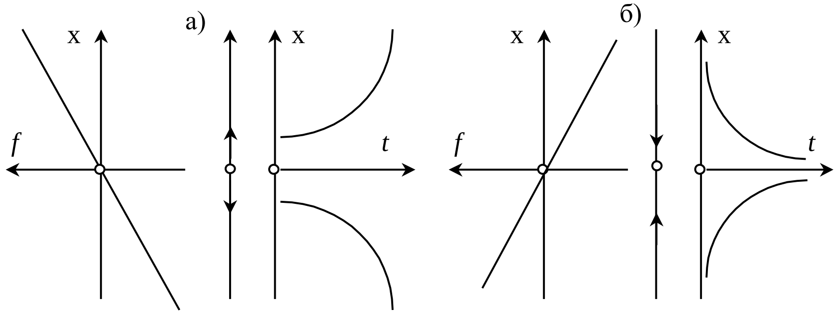


Рис. 4.2. Графики функции $f(x)$: фазовые портреты и интегральные кривые уравнения $x' = \varepsilon x$; а) $\varepsilon>0$; *б)* $\varepsilon<0$

Для любых $\varepsilon \neq 0$ имеется только одна неподвижная точка $x = 0$. Если $\varepsilon>0$, то эта точка неустойчива – репеллер; любая интегральная кривая, для которой $x_0 \neq 0$, неограниченно удаляется от прямой $x=0$, причем со скоростью, увеличивающейся пропорционально расстоянию до этой прямой. Если $\varepsilon<0$, то точка $x=0$ устойчива – аттрактор; любая интегральная кривая, для которой $x_o \neq 0$, стремится к прямой $x=0$, причем со скоростью, уменьшающейся пропорционально расстоянию. Для особого случая $\varepsilon=0$ все точки оси x являются неподвижными и интегральными кривыми оказываются прямые $x(t) \equiv x_0$.

Аналогично можно провести качественное исследование любого автономного уравнения и для нелинейной функции $f(x)$. Для этого надо найти все нули этой функции (т.е. все стационарные решения уравнения $x' = f(x)$), исследовать, к какому из указанных трех типов принадлежит каждое из этих решений, а затем построить фазовый портрет и на его основе исследовать поведение интегральных кривых на плоскости t, $x(t)$ в каждой из полос, ограниченной двумя ближайшими стационарными решениями.

Качественное исследование позволяет получить весьма полезную информацию о характере динамики системы, описываемой уравнением (4.2).

4.1.2.2. Шаги аналитического моделирования

Аналитическое моделирование включает четыре шага (рис. 4.3). Все эти шаги в простой модели сливаются воедино, а в сложной требуют применения специальных методов.

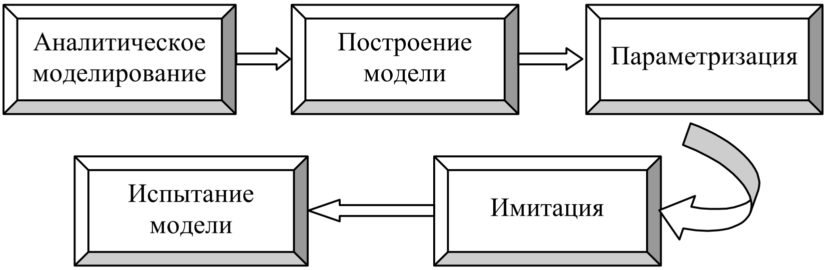


Рис. 4.3. Этапы аналитического моделирования

Формирование концепции модели и ее построение. На рис. 4.4 показана схема построения аналитической модели: строится концепция модели, затем составляются уравнения, описывающие поведение системы.

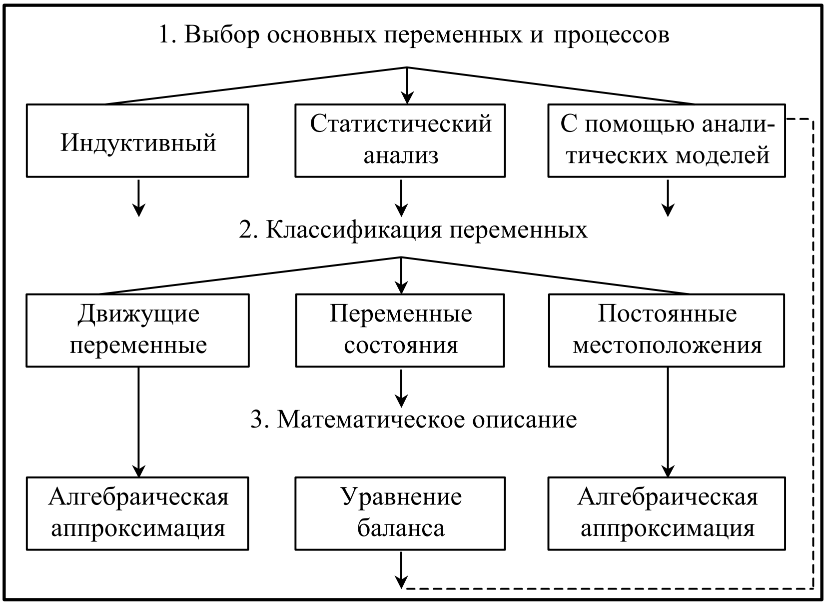


Рис. 4.4. Этапы построения аналитической модели

Математическая модель, как и любая другая модель, означает упрощение объективной реальности. Она должна отражать наиболее существенные свойства реальной системы, а вовсе не все ее детали.

Переменные, характеризующие систему и фигурирующие в модели экосистемы, делятся на три группы: *переменные состояния, движущие переменные, постоянные местоположения.*

Любая переменная состояния подчиняется закону сохранения массы и энергии, а различные изменения описываются с помощью дифференциальных уравнений (уравнения баланса).

Движущие переменные вызывают внешние возмущения, приводящие к изменениям переменных состояния системы, хотя сами они не зависят от этих переменных состояния. Их часто называют входными переменными. Температура, суммарная радиация, продолжительность солнечного освещения и входящие потоки – типичные движущие переменные экосистем.

Постоянные местоположения характеризуют физические границы, в пределах которых происходят изучаемые процессы. В водных экосистемах таковыми являются внешняя форма и внутренняя структура водного бассейна. Объем, глубина и форма водоема, расположение входящих и выходящих потоков – вот некоторые из факторов, влияющих на формирование внешней конфигурации бассейна. Моделируемому водоему придается упрощенная геометрическая форма. Внутренняя структура состоит из определенного вида вертикальных слоев и (или) горизонтальных разрезов, рассматриваемых в модели как однородные.

Параметризация. Этот термин означает определение количественных значений параметров. Для этого существуют три способа.

1. Предварительная оценка получается на основе данных лабораторных и полевых наблюдений за ходом процессов и влиянием различных факторов с помощью корреляционного анализа или различных методов оценки параметров.

2. Комбинации параметров, отвечающих моделируемой ситуации, можно получить исходя из оценки, базирующейся на методах оптимизации параметров. Первая задача параметрической оптимизации состоит в правильном определении целевой функции, в соответствии с которой минимизируется разность между наблюдаемыми и прогнозируемыми значениями. Вторая задача связана с нахождением глобального минимума, при этом в процессе поиска обнаруживаются локальные минимумы. Следующая проблема состоит в том, что для экологических моделей часто оказываются одинаково оптимальными несколько сочетаний параметров, однако до сих пор не найден объективный способ определения наилучшего из них. В некоторых случаях можно наложить ограничения на значения параметров и (или) переменных состояния, а также составляющие баланса. Выход за пределы ожидаемых или физически

возможных диапазонов параметров и получение отрицательных значений переменных состояния можно рассматривать как явный признак неудачного сочетания параметров. В этом случае можно применить методы оптимизации с ограничениями, но они являются более сложными. Оптимизация параметров сложной системы – задача неразрешимая, в частности, из-за недостаточности объема и качества реальных данных. Практическая оптимизация параметров имитационных моделей реальна лишь для ограниченного числа параметров (меньше 10), однако для этого необходимо иметь качественные начальные оценки. Поэтому такая операция может иметь ценность лишь при окончательной доводке модели, которая сама по себе уже является качественной.

3. Оценить роль тех или иных параметров имитационной модели можно с помощью *анализа чувствительности*. Цель его – определить, как модель реагирует на изменение значений параметров, что в свою очередь позволяет сделать вывод о правильности модели и оценок параметров. Существуют три классических метода анализа чувствительности:

А) Анализ чувствительности к изменениям основных параметров. Произвольно изменяются значения некоторых параметров в предполагаемых пределах их значимости и исследуется отклик системы.

Б) Экспериментальный (или конечный) анализ чувствительности. Значения параметров p_i изменяются на некоторую конечную величину Δp, модель заново проходится и сопоставляются значения выходных переменных модели для номинального и измененного параметрических векторов. Результат зависит от Δp. Когда значения Δp очень малы ($< 10^{-1}\%$), получаются те же результаты, что и при использовании аналитического метода анализа чувствительности.

В) Аналитический (дифференциальный) анализ чувствительности. Проводится расчет функций чувствительности, являющихся частными производными переменных состояния по параметрам $S(p_i) = \dfrac{\partial u_j}{\partial p_i}$.

Оценивая чувствительность к внешним параметрам, т.е. к параметрам, связанным с движущими переменными или постоянными местоположения, мы видим, как может себя повести данная экологическая система при различных условиях (т.е. в иных географических условиях, при иной трофической структуре и т.д.). Особая проблема связана с чувствительностью модели к величине интервала между выборками и видом аппроксимации движущих переменных.

Для внутренних параметров, характеризующих переменные состояния, цель заключается в определении важности параметров для последующего приближения модели к реальной действительности. Следует обращать особое внимание на те параметры, к которым модель наиболее чувствительна.

Производится оценка действительных неопределенностей входных данных, которые затем суммируются с тем, чтобы получить и оценить неопределенность модельного прогноза. Трудности такого подхода связаны с необходимостью определять меру изменения итогового выхода моделирования, а также доказывать то, что точности входов могут быть определены с достаточной достоверностью.

4.1.2.3. Имитационное моделирование

Основу метода имитационного моделирования составляет воспроизведение динамических свойств исследуемой системы с использованием численных методов и ЭВМ. В настоящее время имитационный подход превратился в мощный инструмент анализа сложных систем.

Сущность метода. Существуют два способа получения новой информации о некотором объекте или системе. Это, во-первых, пассивное наблюдение и, во-вторых, наблюдение в условиях 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 частности, оптимизировать выбранный им критерий. Активный эксперимент может проводиться либо с самим объектом, либо с его моделью, реализованной на ЭВМ. Под имитационным моделированием понимают процесс построения сложной системы и проведение серий экспериментов с этой моделью, направленных либо на понимание специфики функционирования системы, либо на выработку стратегии управления, удовлетворяющей выбранным критериям.

Сложная система – это такой объект реального мира, поведение которого невозможно предсказать с необходимой степенью детализации на основе обозримого набора ключевых параметров. Имея дело со сложной системой, необходимо исходить из того факта, что ее полное исследование невозможно и что следует ограничиться тем кругом задач, который позволяет получить ответы на частные вопросы. Именно для решения той или иной конкретной задачи и следует разрабатывать имитационную модель. Следовательно, для сложной системы может быть построен неограниченный набор моделей.

Имитационное моделирование аналогично экспериментированию с натурными объектами, хотя модель представляет в этом отношении гораздо больше возможностей. 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 серии «прогонов» модели, а не путем ее решения, как это делается при использовании аналитических методов. Имитационные модели могут служить лишь средством анализа систем в тех условиях, которые задаются экспериментатором.

В чем же заключается специфика имитационного исследования экосистем? При теоретическом исследовании динамики экосистем приходится абстрагироваться от описания конкретных ситуаций, выделяя и включая в модель некоторые обобщенные процессы, присущие всем классам реально существующих объектов. Это связано с целями, которые ставит исследователь: изучить, по возможности, наиболее широкий круг явлений, наиболее распространенные 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 воздействия и управления этими экосистемами. Поэтому при построении имитационных моделей обязательно происходит детализация и уточнение общих зависимостей, учет специфических механизмов, действующих в данной реальной ситуации.

Построение имитационной модели состоит из двух этапов. На первом этапе, этапе анализа, определяется структура модели, производится разбиение общей системы на блоки и дается математическое описание отдельных блоков. На втором этапе, этапе синтеза, осуществляется стыковка блоков и общая сборка модели с применением ЭВМ. На этапе сборки целесообразно использование специальных языков имитационного моделирования.

Структура имитационных моделей экосистем. Имитационные модели экосистем конструируются как динамические балансовые структуры блочного типа. Динамический характер моделей связан с тем, что процессы, происходящие в экосистемах, протекают во времени. В экологии исследователя, как правило, интересуют не столько установившиеся состояния, сколько те переходные процессы, которые возникают в ответ на то или иное антропогенное воздействие.

Требование включения в модель балансовых соотношений по основным компонентам энерго- и массообмена связано с двумя обстоятельствами. Первое заключается в том, что процессы, протекающие в экосистемах, связаны с перетоком и расходованием энергии и вещества. Второе обстоятельство связано чисто с целями рационального природопользования, например, при разработке ресурсосберегающих технологий одновременно решаются задачи охраны среды от загрязнения и эффективного и экономичного использования всех ресурсов.

Блочная структура модели связана с необходимостью описания взаимодействия процессов самой различной природы 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. Очень важным при этом оказывается вопрос организации взаимодействия блоков при «сборке» модели.

Имитационные системы. Практически все имитационные модели обладают рядом специфических свойств, важными из которых являются:

- большая размерность пространства состояний;
- блочная структура моделей;
- необходимость описания и последующего воспроизведения неконтролируемых входных воздействий;
- необходимость рационального планирования машинного эксперимента для целей изучения реакции модели на контролируемые входные воздействия.

При этом, как уже указывалось выше, сложность моделируемой 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 всегда конкретна, ее структура определяется целью моделирования, требуемой точностью воспроизведения свойств моделируемого объекта и характером доступной информации об объекте.

Набор моделей совместно с управляющими и обслуживающими программами образует некоторую систему взаимодействующих моделей, которая называется имитационной системой и имеет иерархическую структуру.

Центральной частью системы является базовая модель экосистемы (рис. 4.5). Целью построения базовой модели является систематизация всех существующих представлений о моделируемом объекте, выявление «белых» пятен, не позволяющих построить замкнутое описание, проверка гипотез о характере некоторых процессов, протекающих в системе. 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.

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. 4.5). 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.

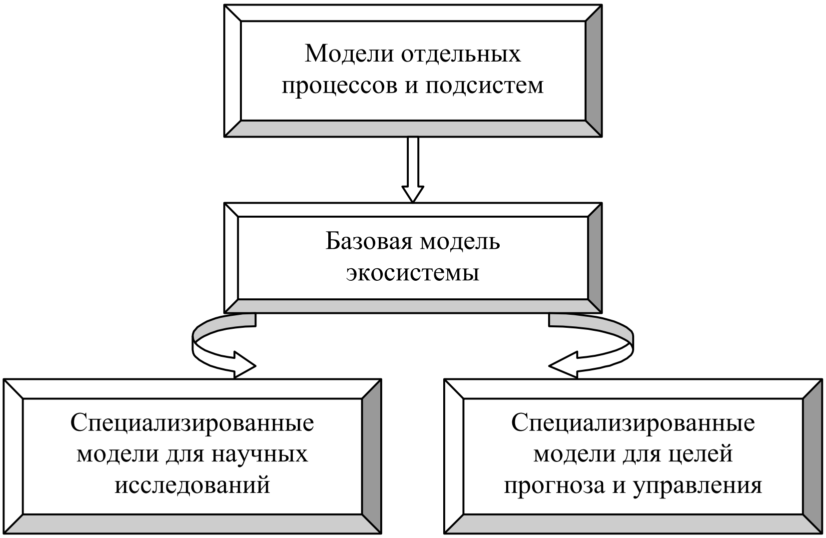


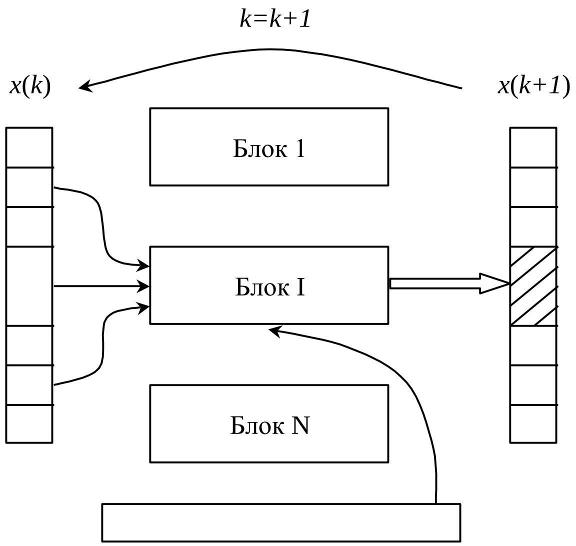
Рис. 4.5. Иерархия моделей

Описанная структура может считаться достаточно универсальной, т.к. она обеспечивает выполнение всех сформулированных выше целей моделирования и при этом не нарушается принцип конструирования каждой из моделей для решения вполне определенных конкретных задач.

Неотъемлемой частью любой имитационной системы должна быть специализированная база данных, т.к. имитационная модель предназначена для переработки информационных потоков. Учитывая, что модель реализуется на ЭВМ, являющейся вычислителем дискретного действия, можно представить схематически последовательность операций в процессе моделирования (рис. 4.6).

На k-м временном шаге модели каждому из ее n 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). Состояние модели в целом (т.е. вектор x) является объединением векторов состояния всех блоков:

$$x = (x_1^*, x_2^*, ..., x_n^*)^* .$$



а

Рис. 4.6. Схема работы модели блочного типа

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 схематичного 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. 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.

В имитационных системах используются следующие термины. Блоком модели называется программная единица, реализующая численный метод решения некоторой математической задачи, описывающей группу функционально однородных 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. Состояние модели – объединение векторов состояния всех блоков модели. Базовый шаг модели – интервал модельного времени между двумя последовательными состояниями модели. 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 шага к шагу и являются функциями времени и, возможно, состояния модели. Входные воздействия разделяются на неконтролируемые (возмущения) и контролируемые (управление). Банк моделей – совокупность моделей, решающих задачи определенной предметной области и реализуемых в некоторой операционной среде на основе единого математического, программного, технического и информационного обеспечения. Понятие банка моделей тесно связано с единством их информационой базы. Реализация модели определяется заданием численных значений вектора параметров, вектора начального состояния, вектора входных воздействий. Непосредственная связь банка моделей с базой данных осуществляется именно на стадии реализации модели. Приведенные термины помогают более эффективно строить работу по созданию имитационных систем, с использованием технологии имитационного моделирования.

Необходимость разработки специальной технологии имитационного моделирования возникла по двум причинам. Первая заключается в том, что при разработке моделей и при создании банка моделей необходимо решать задачи не только формализации описания тех или иных процессов, выбора численных методов, но и принимать обоснованные решения относительно структуры данных, организации взаимодействия модели с другими программными модулями имитационной системы, обработки и представления результатов моделирования. Вторая причина заключается в том, что разработка хорошей модели является, как правило, коллективным делом и представляет длительный процесс, все этапы которого должны быть согласованы.

Технология моделирования заключается в выполнении следующих этапов:

1. Конструирование модели:

Определение системы – установление границ подлежащей изучению реальной системы, ограничений и критериев эффективности ее функционирования.

Формулирование модели – переход от реальной системы к некоторой абстрактной логической схеме, определение целей моделирования и тех задач, которые должны быть решены в процессе использования модели.

Обоснование структуры информационного обеспечения, анализ доступных для решения задачи данных, а также разработка внешней организации базы данных и ее концептуальной модели.

«Трансляция» модели, т.е. описание модели на языке ЭВМ, разработка математического обеспечения, интерфейсных модулей для связи с базой данных.

2. Планирование и проведение машинных экспериментов:

Стратегическое планирование – обоснование метода последовательного осуществления серий экспериментов с целью изучения модели и оценки ее чувствительности.

Тактическое планирование – определение способа проведения очередной серии испытаний в рамках выбранной стратегии эксперимента.

Экспериментирование с моделью – проведение серии испытаний модели с целью получения желаемых результатов.

3. Практическое использование модели:

Статистическая обработка результатов «прогонов» модели, качественная оценка ее свойств и эксплуатационных характеристик.

Оценка адекватности модели.

Интерпретация – построение выводов по данным имитационного эксперимента и их содержательное истолкование.

Реализация – практическое использование модели.

Документирование – регистрация хода разработки и его результатов.

Из всех этапов специального внимания заслуживает этап 3.2 – оценка адекватности модели. Многомерность, наличие сложной иерархической организации, адаптивность по отношению к внешним воздействиям и другие особенности экосистем, например, слабая изученность трофических связей и взаимодействий не позволяют говорить об адекватности имитационной модели реальному объекту в общепринятом значении этого слова. Поэтому адекватность модели будет оцениваться той точностью, которая может быть достигнута при решении той или иной конкретной задачи, для которой и разрабатывалась имитационная система. Если модель создается для решения задач управления, то оценить адекватность модели бывает достаточно просто. Напротив, в тех ситуациях, когда модель конструируется для изучения плохо изученных механизмов, оценка адекватности модели зачастую носит субъективный характер, т.к. она отражает степень уверенности автора в относительной истинности модельных построений. В этом случае следует оценить тот относительный вклад в общую картину поведения модели, который внесли ее составляющие, основанные как на вполне надежных, так и на гипотетических представлениях.

4.1.3. Адаптивное моделирование

Адаптивное моделирование является подразделом стохастического моделирования. Схема адаптивного моделирования показана на рис. 4.7, адаптация модели к системе происходит автоматически с помощью поиска на ЭВМ. Для динамических систем требуется динамическая самоадаптация модели. При этом возможна автоматическая адаптация как параметров, так и структуры. Задача адаптации – добиться оптимального поведения модели, как это определено целевой функцией. Эти свойства системы, которые до начала адаптации были мало, а то и вовсе неизвестны, в ходе адаптации должны обнаружиться.

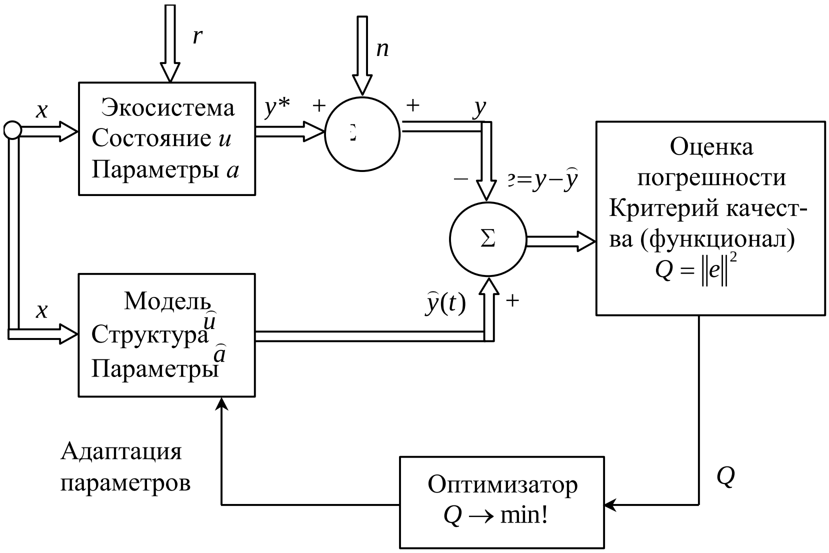


Рис. 4.7. Блок-схема адаптивного моделирования:
Информация, поступающая в систему, используется для построения модели. Качество модели непрерывно оценивается путем оценки погрешностей, поэтому управляющие решения можно принимать уже в процессе адаптации. Параметры модели $\hat{a}$ оцениваются путем итераций и использования новых данных наблюдений до тех пор, пока функционал Q не достигнет минимума или не станет ниже установленного уровня

В саморегулирующихся моделях (системах) адаптация основывается на модификации параметров, но при этом модель называется самоорганизующейся, если причиной адаптации служит изменение структуры.

4.2. Методы анализа сообществ и экосистем

В данной главе рассматриваются методы обработки и анализа экологических данных в пакете прикладных программ «Statistica». Подробное руководство по пользованию данным пакетом изложено в учебном пособии В.П. Боровиков, И.П. Боровиков «Statistica. Статистический анализ и обработка данных в среде Windows» (М.: Филинъ, 1997).

4.2.1. Методы анализа одной и множества переменных

4.2.1.1. Разведочный анализ экологических данных

Любой статистический анализ целесообразно начинать с предварительной обработки данных, которая включает вычисление всевозможных описательных статистик, позволяет определить зависимости между данными, разбить их различными способами, просмотреть эти группы визуально и т.д. К описательным (дескриптивным) статистикам относятся среднее, медиана, дисперсия, стандартное отклонение, ассиметрия, эксцесс, квантили и т.д.

Разведочный анализ в ППП «Statistica» проводится в модуле «Basic Statistics/Tables» – Основные статистики/таблицы. Для описания основных элементов диалога в данном модуле рассмотрим конкретный пример.

Пример. Имеются данные об уровнях первичного продуцирования, распределения биогенных элементов и температурном режиме по сезонам в различных районах залива Петра Великого (табл. 4.1, данные, используемые в примерах данной главы, носят условный характер, ссылка на них недопустима).

Предварительное исследование данных заключается в анализе основных характеристик переменных (температура, биогены, продукция) и сравнении их средних в различных участках залива Петра Великого.

С помощью переключателя модуля в системе «Statistica» необходимо открыть модуль «Basic Statistics/Tables» – Основные статистики/таблицы. В появившейся стартовой панели этого модуля необходимо выбрать метод «Descriptive statistics» – описательные статистики, в котором можно:

- выбрать переменные для анализа (для нашего примера это переменные ТЕМПЕРАТУРА, НИТРАТЫ, ФОСФАТЫ, ПРОДУКЦИЯ);
- вычислить различные описательные статистики;
- оценить близость распределения к нормальному закону;
- построить гистограммы;

• сгруппировать данные, разбив их определенным образом на классы и т.д.

Таблица 4.1

Исходные данные для проведения разведочного анализа

№ п/п	Район	Месяц	Сезон	Температура	Нитраты, мг/м^3	Фосфаты, мг/м^3	Продукция, мгС/м^3
1	2	3	4	5	6	7	8
1	Амурский	март	весна	5	30.5	16.8	2.92
2	Амурский	апрель	весна	6	19.5	11.8	1.95
3	Амурский	май	весна	4	25	14	1.12
4	Амурский	июнь	лето	12	26	8.1	18.1
5	Амурский	июль	лето	19	19.4	4.3	19.9
6	Амурский	август	лето	20	17.6	6.4	21.7
7	Амурский	сентябрь	осень	14	17.5	9.7	41.5
8	Амурский	октябрь	осень	10	15.2	8.6	33
9	Амурский	ноябрь	осень	10	20.5	11.7	20.5
10	Уссурийский	март	весна	1.5	18.1	14.8	1.01
11	Уссурийский	апрель	весна	1.1	19	12.1	1.21
12	Уссурийский	май	весна	5	26.3	13.5	2.7
13	Уссурийский	июнь	лето	11	21.4	12.8	17.2
14	Уссурийский	июль	лето	17	19.8	11.2	29.2
15	Уссурийский	август	лето	19.5	18.6	9.3	12.9
16	Уссурийский	сентябрь	осень	13.5	25.3	11.6	44.8
17	Уссурийский	октябрь	осень	7.3	21	9.8	36.7

Окончание табл. 4.1

1	2	3	4	5	6	7	8
18	Уссурий-ский	ноябрь	осень	8	22.7	10.4	7.46
19	Восток	март	весна	4	19.7	16.1	3.71
20	Восток	апрель	весна	3.1	18	15.3	2.8
21	Восток	май	весна	6	26	11.7	3.16
22	Восток	июнь	лето	12	21.5	5.1	34.8
23	Восток	июль	лето	19	18.7	4.7	20.8
24	Восток	август	лето	20	29	7.6	34.6
25	Восток	сентябрь	осень	14	32	8.3	47.3
26	Восток	октябрь	осень	9.6	29	10.6	39.2
27	Восток	ноябрь	осень	10	25	14.1	25.2
28	Посьета	март	весна	5	28.7	14.3	1.95
29	Посьета	апрель	весна	4.2	31.4	15.5	2.05
30	Посьета	май	весна	6	21.7	11.6	2.46
31	Посьета	июнь	лето	12	17.3	8.6	25.2
32	Посьета	июль	лето	19	16.7	7.5	26.8
33	Посьета	август	лето	20	19.6	8.8	27.6
34	Посьета	сентябрь	осень	14	24.6	16.2	53.8
35	Посьета	октябрь	осень	8.9	15.1	14.4	44.1
36	Посьета	ноябрь	осень	10	17.5	12.8	19.7

В табл. 4.2 представлены результаты расчета основных описательных статистик (среднее, минимум, максимум, дисперсия и стандартное отклонение).

Простой анализ этих результатов мало информативный. Гораздо интереснее сравнить данные отдельно по участкам залива в разные сезоны года. Для этого можно построить категоризованную гистограмму, где по категориям РАЙОН и СЕЗОН будут представлены данные, например, о температуре (рис. 4.8). Для построения категоризованной гистограммы необходимо на панели инструментов нажать на пиктограмму «Галерея графиков» и в появившемся окне выбрать опцию «Категоризованные графики» – «Гистограмма», а затем выбрать в списке переменных переменные для построения графика.

Таблица 4.2

Результаты расчета основных статистик
(по всему зал. Петра Великого)

Переменная	Средняя (mean)	Минимум (minimum)	Максимум (maximum)	Дисперсия (variance)	Стандартное отклонение (Std. dev)
Температура	10,575	1,1	20,0	33,83	5,82
Нитраты	22,08	15,1	32,0	22,55	4,75
Фосфаты	11,11	4,3	16,8	11,42	3,38
Продукция	20,25	0,01	53,8	258,44	16,08

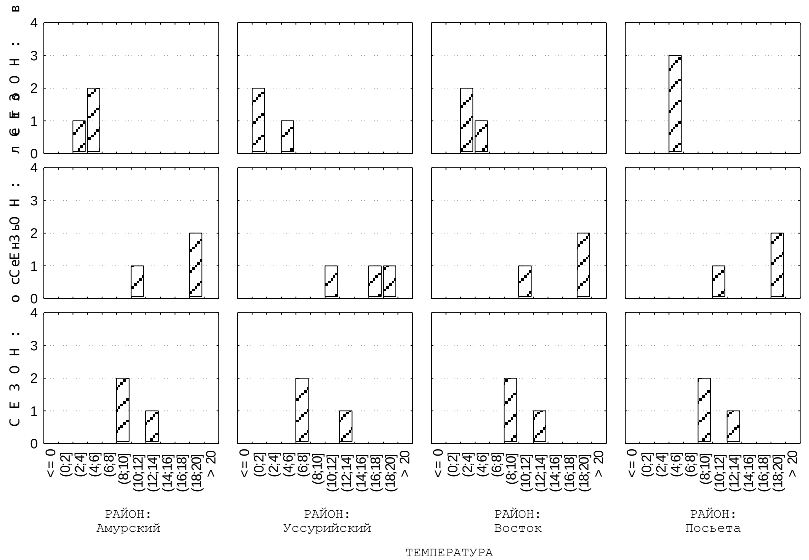


Рис. 4.8. Температурный режим зал. Петра Великого (по районам и сезонам)

Предположим, что у вас возникла необходимость построить таблицу частот распределения интересующих вас переменных в целом по заливу или по отдельным участкам. Для этого необходимо в стартовой панели модуля «Основные статистики и таблицы» выбрать опцию «Frequency tables» – «Таблица частот». В появившемся окне нужно выбрать переменные для анализа, например НИТРАТЫ и ФОСФАТЫ, и задать шаг и начальное значение переменных. Запустив вычислительную процедуру, получаем следующую таблицу (табл. 4.3).

Таблица 4.3

Таблица частот распределения фосфатов и нитратов в зал. Петра Великого (шаг выборки равен 3, начиная с минимального значения)

Фосфаты			Нитраты		
Интервалы частот	Кол-во	%	Интервалы частот	Кол-во	%
4.3<=x<7.3	4	11.1111	15.1<=x<18.1	8	22.222
7.3<=x<10.3	10	27.7778	18.1<=x<21.1	11	30.555
10.3<=x<13.3	11	30.5555	21.1<=x<24.1	4	11.111
13.3<=x<16.3	10	27.7778	24.1<=x<27.1	7	19.444
16.3<=x<19.3	1	2.7778	27.1<=x<30.1	3	8.333
			30.1<=x<33.1	3	8.333

В таблице указаны интервалы концентрации биогенных элементов, частота попадания в данный интервал и частота, выраженная в процентах. Так, анализируя полученные результаты, можно сделать вывод, что, например, в заливе Петра Великого частота встречаемости концентрации нитратов, лежащей в пределах от 18,1 до 21,1 мкг/л, равна 11.

Часто бывает необходимо построить таблицу частот не для всей выборки, а для отдельных случаев (в нашем примере, допустим, для конкретных районов или сезонов). Для этого в том же окне «Таблицы частот» нужно воспользоваться кнопкой «select cases» – «выбор случаев», и в выпавшем окне задать условие выбора случаев. В окне «Таблицы частот» можно также построить гистограммы значений переменных, на которые можно при желании наложить плотность нормального распределения, проверить согласие данных с помощью критериев Колмогорова-Смирнова, Лилиерфорса, вычислить статистику Шапиро-Уилкса.

4.2.1.2. Корреляционный анализ в системе «Statistica»

Корреляция, или, точнее, коэффициент корреляции, является мерой зависимости двух величин или мерой взаимной согласованности в изменчивости двух или нескольких признаков, явлений. Коэффициент корреляции является безразмерной величиной и изменяется в пределах от -1 до +1. Если наблюдается тенденция возрастания одной величины при росте другой, то имеет место положительная корреляционная зависимость, если при возрастании одной величины другая уменьшается, то говорят об отрицательной коррелированности величин. Чем ближе коэффициент корреляции к +1 или -1, тем ближе корреляционная зависимость к линейной.

В модуле «Основные статистики/таблицы» легко можно вычислить и проанализировать корреляционную матрицу данных выбранных переменных. Для этого в стартовом окне модуля необходимо выбрать опцию «correlation matrices – корреляционные матрицы». После этого на экране появляется окно «Pearson Product-Moment Correlation – Корреляция Пирсона». С помощью двух кнопок вверху окна задаются переменные для двух возможных типов корреляционных матриц: квадратной и прямоугольной.

Вернемся к примеру данной главы (табл. 4.1) и построим квадратную корреляционную матрицу для проверки зависимости между переменными НИТРАТЫ, ФОСФАТЫ, ТЕМПЕРАТУРА и ПРОДУКЦИЯ. Выберем в списке переменных в окне «Корреляция Пирсона» нужные переменные и запустим вычислительную процедуру. В результате расчетов получаем следующую корреляционную матрицу (табл. 4.4).

Таблица 4.4

Корреляционная матрица для переменных ТЕМПЕРАТУРА – ПРОДУКЦИЯ

Объем выборки N=36				
Переменная	ТЕМПЕРАТУРА	НИТРАТЫ	ФОСФАТЫ	ПРОДУКЦИЯ
ТЕМПЕРАТУРА	1	-0.180716	-0.717063*	0.589436*
НИТРАТЫ	-0.180716	1	0.244628	-0.047261
ФОСФАТЫ	-0.717063*	0.244628	1	-0.357222*
ПРОДУКЦИЯ	0.589436*	-0.047261	-0.357222*	1

Примечание. Звездочками помечены корреляции, значимые на уровне $p < 0,05$.

Как видно из таблицы, обнаружена достоверная корреляционная связь между следующими переменными: ТЕМПЕРАТУРА – ФОСФАТЫ, ТЕМПЕРАТУРА – ПРОДУКЦИЯ, ПРОДУКЦИЯ – ФОСФАТЫ.

Причем между переменными ТЕМПЕРАТУРА – ФОСФАТЫ эта зависимость отрицательная и по всей вероятности близка к линейной, коэффициент корреляции равен -0,72. Это означает, что с повышением температуры концентрация фосфатов в водах залива уменьшается. Между переменными ТЕМПЕРАТУРА – ПРОДУКЦИЯ коэффициент корреляции равен 0,59, что свидетельствует о значимой положительной связи между этими переменными. Что касается зависимости между переменными ПРОДУКЦИЯ – ФОСФАТЫ, то достоверный коэффициент корреляции, равный -0,36, дает основание сделать вывод об отрицательной нелинейной корреляционной связи, т.е. концентрация фосфатов уменьшается по мере возрастания уровня первичного продуцирования.

Необходимо заметить, что коэффициент корреляции имеет ошибку репрезентативности и необходимо в обязательном порядке проверять его на достоверность (Дулепов и др., 2004). Но, вычисляя корреляционную матрицу в ППП «Statistica», исследователь избавлен от необходимости делать такую проверку, т.к. это делается автоматически при вычислении коэффициентов корреляции, и все значимые коэффициенты в таблице помечаются звездочками.

При проведении первичного анализа экологических данных вычисление коэффициентов корреляции считается очень целесообразным, т.к. обнаружение связи между переменными служит отправной точкой для дальнейших исследований, например, для построения регрессионной модели.

4.2.1.3. Регрессионный анализ в системе «Statistica»

Линейный регрессионный анализ. Под регрессией подразумевается зависимость изменения одного признака от изменений другого или нескольких признаков. Построение регрессионной модели заключается в составлении и решении уравнений регрессии, т.е. моделировании зависимости между переменными путем подбора соответствующей функции (линейной или нелинейной).

Необходимо напомнить, что строить регрессионную модель имеет смысл только тогда, когда между изучаемыми признаками обнаружена достоверная корреляционная связь. Как уже отмечалось, между переменными температура и фосфаты обнаружена сильная отрицательная корреляционная связь (табл. 4.4), причем, мы сделали предположение, что эта связь близка к линейной. Попробуем проверить это, построив линейную регрессионную модель.

Итак, мы имеем значения пары переменных x – ТЕМПЕРАТУРА и y – ФОСФАТЫ. Переменная x носит название независимой переменной, или предиктора, переменная y называется зависимой переменной или откликом. Мы хотим определить зависимость переменной y от x или предсказать, какими будут значения y при данных значениях x. Точное описание линейной регрессионной модели дано в учебном пособии

«Системная экология» (Дулепов и др., 2004), кратко напомним лишь вид этой модели. Наблюдаемые величины связаны между собой регрессионной зависимостью вида

$$y\,(i)=b_1 \cdot x(i)+b_0+e_i,$$

где b_1, b_0 – неизвестные константы, e_i – ненаблюдаемые случайные величины с нулевым средним и неизвестной дисперсией, не меняющейся от опыта к опыту, или ошибки наблюдения. Предполагается, что ошибки некоррелированы в разных опытах и распределены по нормальному закону. Если независимых переменных несколько, то мы имеем дело с множественной линейной регрессионной моделью.

Линейная модель может быть исследована в модуле «Множественная регрессия – Multiple regression». Общая задача состоит в том, чтобы по данным наблюдений оценить параметры b_1, b_0 модели наилучшим способом, проверить гипотезу о значимости регрессии, оценить адекватность модели.

Из переключателя модулей Statistica необходимо открыть модуль «Множественная регрессия» и в стартовой панели модуля выбрать переменные для анализа: зависимая – ФОСФАТЫ, независимая – ТЕМПЕРАТУРА. После запуска вычислительной процедуры получаем таблицу с результатами анализа. Окно результатов имеет следующую структуру: верх окна информационный. Он состоит из двух частей: в первой части содержится основная информация о результатах оценивания, во второй высвечиваются значимые регрессионные коэффициенты. Внизу окна находятся функциональные кнопки, позволяющие всесторонне рассмотреть результаты анализа. В табл. 4.5 приведены результаты регрессионного анализа.

Таблица 4.5

Результаты регрессионного анализа
для зависимой переменной ФОСФАТЫ

R=.71706308 RI=0,51417946 Adjusted RI=0,49989063						
F(1,34)=35.985 p<0,00000 Стандартная ошибка оценки: 2.3900						
Переменные	BETA	St. Err. of BETA	B	St. Err. of B	t(34)	p-level
Intercept			15.52	0.84	18.57	2.24^{-19}
ТЕМПЕРАТУРА	-0.72*	0.12	-0.42	0.07	-5.999	8.65^{-7}

В данной таблице:

R – коэффициент множественной корреляции;

RI – квадрат коэффициента корреляции, называемый обычно коэффициентом детерминации и показывающий долю общего разброса данных (относительно выборочного среднего), которую объясняет построенная регрессионная прямая;

Adjusted RI – скорректированный коэффициент детерминации, определяемый как

$$Adjusted\ RI=1–(1–RI)\cdot(n/(n–p),$$

где n – число наблюдений в модели, p – число параметров модели (число независимых переменных плюс 1, т.к. в модель включен свободный член);

стандартная ошибка оценки является мерой рассеяния наблюдаемых значений относительно регрессионной прямой;

Intercept (оценка свободного члена регрессии) – значение коэффициента b_0 в уравнении регрессии;

St. Err. of B (стандартная ошибка оценки свободного члена);

t(34) и p-level (значение t-критерия и уровень значимости p). Т-критерий используется для проверки гипотезы о равенстве 0 свободного члена регрессии;

F и p – значение F-критерия для оценки адекватности модели и значение уровня значимости;

BETA – стандартизованный коэффициент регрессии (звездочкой помечен значимый коэффициент регрессии), т.е. коэффициент при независимой переменной;

B – точечные оценки параметров модели.

Значения коэффициента детерминации лежат в интервале от 0 до 1. В нашем примере RI=0,51417946. Это неплохой показатель, показывающий, что построенная регрессия объясняет более 51% разброса значений переменной ФОСФАТЫ относительно среднего.

При проведении любого статистического анализа вывод о принятии нулевой гипотезы или достоверности рассчитанных коэффициентов делается на сравнении соответствующих уровней значимости для того или иного критерия с общепринятыми для биологических и экологических исследований. Уровень значимости – это вероятность неправильного отвержения гипотезы, когда она верна. Обычно задают небольшие уровни значимости: 0,1, 0,05 и т.д. В экологических и биологических исследованиях общепринят уровень значимости, равный 5%, это означает, что в 5 случаях из 100 мы можем ошибиться, отвергая нулевую гипотезу. Иными словами, если значение рассчитанного уровня значимости меньше 0,05, то нулевая гипотеза отвергается и принимается альтернативная. Так, в нашем примере были вычислены значение t-критерия для проверки предположения о равенстве нулю свободного члена регрессии (нулевая гипотеза) и соответствующее ему значение уровня значимости ($p=2{,}24^{-19}$). Поскольку уровень значимости значительно меньше 0,05, то нулевая гипотеза отвергается, и рассчитанное значение свободного члена в регрессионной модели признается достоверным.

Для оценки адекватности построенной модели также необходимо проанализировать значение F-критерия и рассчитанный для этой статистики уровень значимости p. Из таблицы видно, что значение F-критерия при числе степеней свободы 1 и 34 равен $F(1,34)=35.985$ на уровне значимости $p<0{,}00000$. Эти значения дают основание утверждать, что построенная регрессия высоко значима, т.е. нулевая гипотеза о том, что между переменными ФОСФАТЫ и ТЕМПЕРАТУРА нет линейной зависимости ($b_1=0$) отвергается.

В столбце B даны коэффициенты линейной модели – свободный член и коэффициент при независимой переменной. Из таблицы видно, что коэффициент при независимой переменной также высоко значим. Итак, построенная линейная регрессионная модель имеет вид

$$\text{\textit{ФОСФАТЫ}} = 15{,}52 - 0{,}42 \cdot \text{\textit{ТЕМПЕРАТУРА}}.$$

По этой модели легко спрогнозировать значение концентрации фосфатов при любой температуре в заливе.

Исходные данные с подогнанной прямой имеют вид, представленный на рис. 4.9.

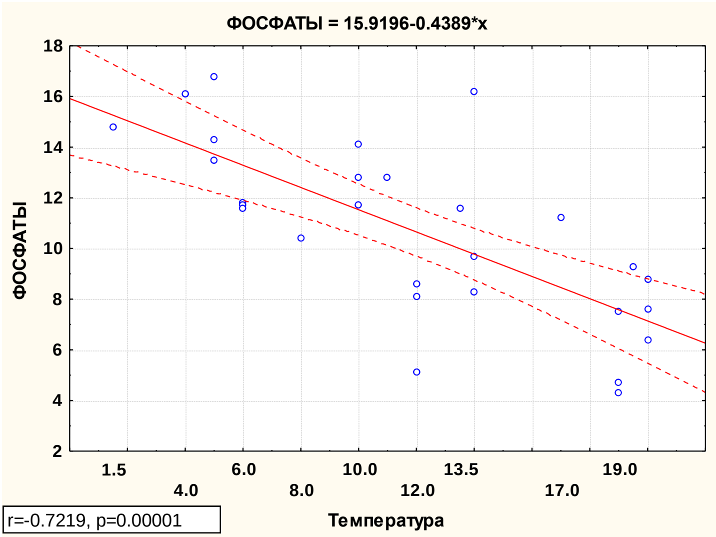


Рис. 4.9. Линейная регрессия для данных ТЕМПЕРАТУРА и ФОСФАТЫ

Упражнение для самостоятельного анализа

По данным табл. 4.6 записать уравнение линейной множественной регрессии между зависимой переменной ПРОДУКЦИЯ и тремя независимыми переменными ФОСФАТЫ, НИТРАТЫ, ТЕМПЕРАТУРА. Сделать выводы об адекватности модели, коэффициентах регрессии. Какова величина коэффициента детерминации?

Таблица 4.6

Результаты регрессионного анализа для зависимой переменной ПРОДУКЦИЯ

R=0,59857037 RI=0,35828648 Adjusted RI=0,29812584						
F(3,32)=5.9555 p<0,00240 Std.Error of estimate: 13.468						
	BETA	St. Err. of BETA	B	St. Err. of B	t(32)	p-level
Intercept			-9.83256	18.18262	-0.54077	0.592413
ТЕМПЕРАТУРА	0.686533	0.203175	1.897601	0.561584	3.379019	0.001928
НИТРАТЫ	0.046551	0.146052	0.157588	0.494427	0.318729	0.752003
ФОСФАТЫ	0.123678	0.206092	0.588324	0.980356	0.600113	0.552657

Нелинейное оценивание. Построение нелинейной регрессионной модели в пакете «Statistica» производится в модуле «Nonlinear Estimation – Нелинейное оценивание». В этом модуле собраны процедуры, позволяющие оценить нелинейные зависимости между данными. Здесь можно выбрать различные модели зависимостей либо задать собственную функцию, можно выбрать различные методы оценивания неизвестных параметров и самим задавать функцию потерь.

Пусть имеется N независимых переменных $x(1)$, $x(2),\ldots,x(\mathrm{n})$ и одна зависимая переменная y. Предполагается, что зависимая переменная связана с независимыми функциональной зависимостью $y=f(x(1),\ldots,x(n))$.

Функция неизвестна или известен ее вид в общих чертах и по имеющимся данным необходимо оценить эту функциональную зависимость, т.е. оценить неизвестные параметры по результатам измерений.

В стартовой панели модуля можно выбрать различные методы обработки. Система предлагает на выбор следующие методы:

User-specified regression – Определенная пользователем регрессия;
Logistic regression – Логистическая регрессия;
Exponential growth regression – Регрессия экспоненциального роста;
Piecewise linear regression – Кусочно-линейная регрессия.

Логистическая регрессия является обобщением логит-регрессии. В логит-модели отклик принимает лишь два значения – 0 или 1, независимая переменная может принимать несколько значений. В данной работе мы не будем подробно останавливаться на логистической регрессии, из данных табл. 4.1 видно, что ни одна из переменных не принимает значения 0 и 1. Для ознакомления с логит-моделью рекомендуем обратиться к книге В.П. Боровиков, И.П. Боровиков «Statistica. Статистический анализ и обработка данных в среде Windows».

Попытаемся построить нелинейную регрессионную модель экспоненциального типа для зависимой переменной ПРОДУКЦИЯ и тремя предикторами ФОСФАТЫ, НИТРАТЫ, ТЕМПЕРАТУРА. Математически эта зависимость записывается следующим образом:

$$y=C+exp(b_0+b_1{\cdot}x_1+b_2{\cdot}x_2+\ldots+b_n{\cdot}x_n)+\varepsilon,$$

где последнее слагаемое ε обозначает ошибку наблюдений.

Параметры C, b_0,b_1, b_2, b_n в этой модели неизвестны и их нужно оценить. Результаты вычислительной процедуры представлены в табл. 4.7.

Таблица 4.7

Результаты оценивания неизвестных параметров для экспоненциальной регрессии

<table>
<tr><td colspan="6">Модель экспоненциального роста ($y=c+exp(b_0+b_1{\cdot}x_1+b_2{\cdot}x_2....)$)</td></tr>
<tr><td colspan="6">Зависимая переменная: ПРОДУКЦИЯ</td></tr>
<tr><td colspan="6">Функция потерь: 5853.6269546 R=0,59402. Дисперсия за счет регрессии: 35.286%</td></tr>
<tr><td rowspan="2">Оценка</td><td>C</td><td>B_0</td><td>ТЕМПЕРАТУРА</td><td>НИТРАТЫ</td><td>ФОСФАТЫ</td></tr>
<tr><td>-337.548</td><td>5.79478</td><td>0.005238</td><td>0.000492</td><td>0.001694</td></tr>
</table>

На пересечении строки ОЦЕНКА и соответствующего столбца с именем независимой переменной получаем нужное значение коэффициента при этой переменной.

Подставляя оценки параметров в уравнение экспоненциальной функции, получаем следующую модель:

$$\textit{ПРОДУКЦИЯ}=337.548+exp(5.79478+0.0052{\cdot}\textit{ТЕМПЕРАТУРА}+$$
$$+0.0005{\cdot}\textit{НИТРАТЫ}+0.001694{\cdot}\textit{ФОСФАТЫ}$$

Как видно из таблицы, коэффициент множественной корреляции равен 0,59402, а коэффициент детерминации – 35,286%, что является не очень высоким результатом. Оценить адекватность модели в модуле «Нелинейное оценивание» можно только графически, с помощью анализа остатков. Остатки представляют собой разности между наблюдаемыми величинами и прогнозируемыми с помощью построенной модели. Если модель подобрана правильно, то остатки будут вести себя достаточно хаотично, в них не будет систематической составляющей, резких выбросов, в чередовании их знаков не будет никаких закономерностей, остатки будут независимы друг от друга, т.е. остатки должны быть похожи на независимые нормальные величины. Для проверки этого предположения необходимо построить график остатков, например, на нормальной вероятностной бумаге (рис. 4.10). Для построения графика остатков в функциональной части окна «Результаты анализа» модуля «Нелинейное оценивание» имеются специальные кнопки – «Distribution of residuals – Гистограмма распределения остатков», «Normal probability plot of residuals – График остатков на нормальной вероятностной бумаге» и др. Если бы остатки были строго нормальными, они хорошо бы легли на прямую нормального распределения.

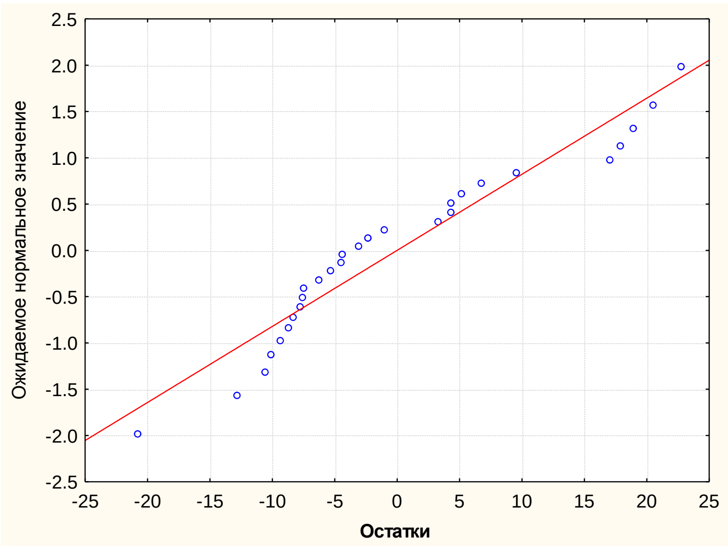


Рис. 4.10. График остатков на нормальной вероятностной бумаге

Из рисунка видно, что остатки довольно плохо ложатся на прямую, что дает нам основание сомневаться в адекватности построенной экспоненциальной модели и начать поиск другой нелинейной функции, удовлетворительно описывающей зависимость между искомыми переменными.

4.2.1.4. Анализ данных, сравнение двух выборок

Первичный анализ данных в системе «Statistica» позволяет не только оценить основные описательные статистики, но и сделать некоторые выводы о зависимости переменных, об их влиянии друг на друга. Предположим, вас интересует, существуют ли значимые различия в средних значениях температуры по сезонам года. Для проверки нулевой гипотезы (предположения) о том, что значимых различий в температуре по сезонам нет, достаточно в модуле «Основные статистики/таблицы» вызвать вычислительную процедуру «t-test for independent samples – t-критерий для независимых выборок». В появившемся окне задать в строке «input file» – «one record per case (use a grouping variable)» – «одна запись на случай (используя группирующую переменную)» группирующую переменную (в нашем примере – СЕЗОН) и зависимую переменную – ТЕМПЕРАТУРА. После запуска вычислительной процедуры на экране появится окно с результатами анализа (табл. 4.8).

Таблица 4.8

Результаты тестирования

Группирующая: СЕЗОН Группа 1: весна Группа 2: лето					
Переменная	Mean весна	Mean Лето	t-value	df	p
ТЕМПЕРАТУРА	4.241667	16.70833	-10.537	22	4.68317E-10

Mean – среднее значение, t-value – значение критерия, df – число степеней свободы, p – уровень значимости.

Самое важное в этой таблице – уровень значимости *p* – тот минимальный уровень, на котором можно отвергнуть нулевую гипотезу и принять альтернативную, заключающуюся в том, что различия в температурном режиме весной и летом значимы.

В нашем примере уровень значимости очень маленький, p=4.68317^{-10}. Это дает нам право отвергнуть нулевую гипотезу о равенстве средних значений температуры весной и летом и принять альтернативную, что средние значения температуры летом и весной очень сильно отличаются. Этот же вывод можно сделать, анализируя график «box&whisker plot» – «график ящик с усами» (рис. 4.11). При построении графика обязательно нужно задать тип графика: Mean/SE/SD (Среднее/стандартная ошибка/стандартное отклонение).

Смысл этих графиков очень простой: точки в центре прямоугольников (ящиков) соответствуют средним значениям переменной ТЕМПЕРАТУРА летом и весной. От этих значений берется положительное и отрицательное стандартные отклонения, положительная и отрицательная стандартные ошибки, получаются «усы» и «ящики». Из графика хорошо видно, что среднее значение температуры летом увеличилось, интервалы стандартных ошибок не перекрываются. Следовательно, и без анализа уровня значимости мы можем сделать вывод о том, что нулевая гипотеза о равенстве средних значений температуры может быть отклонена.

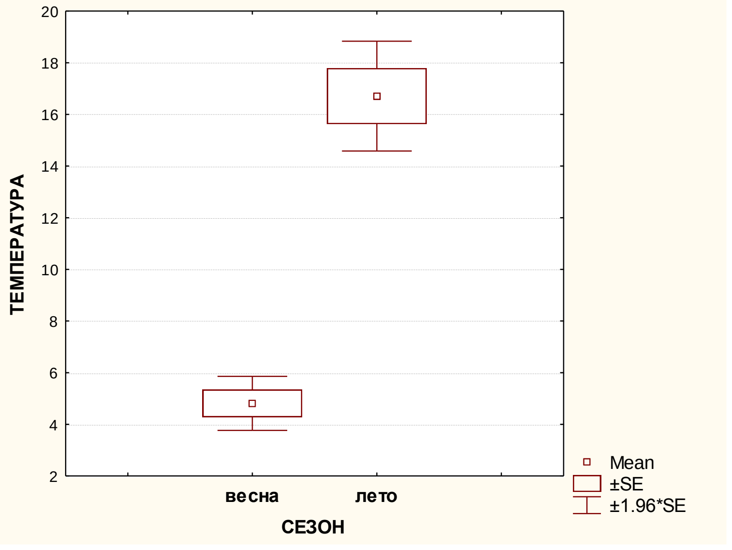


Рис. 4.11. График «ящики с усами»
для переменной ТЕМПЕРАТУРА весной и летом

Судя по данным табл. 4.1, значения нитратов и фосфатов в заливе в течение года сильно колеблются. Возникает естественный вопрос: «Равномерно ли изменяются концентрации биогенов в заливе?» Это равносильно принятию нулевой гипотезы о том, что средние значения концентрации нитратов и фосфатов одинаковы. Для проверки этой гипотезы достаточно рассчитать критерий Стьюдента в модуле «Основные статистики и таблицы». Открыв этот модуль, необходимо в стартовом окне модуля выбрать t-test for dependent samples (t-критерий для зависимых выборок), затем выбрать переменные для анализа (НИТРАТЫ, ФОСФАТЫ) и запустить вычислительную процедуру. Для начала полезно просмотреть результаты анализа графически, для чего необходимо построить график «ящик с усами» (рис. 4.12).

На графике хорошо видно, что среднее значение нитратов гораздо больше среднего значения фосфатов, интервалы стандартных ошибок не перекрываются, следовательно, можно предположить, что гипотеза о равномерном изменении биогенов в заливе может быть отклонена и мы должны принять альтернативную гипотезу о том, что средние значения нитратов и фосфатов в заливе сильно отличаются друг от друга. Проверим это предположение с помощью результатов тестирования (табл. 4.9).

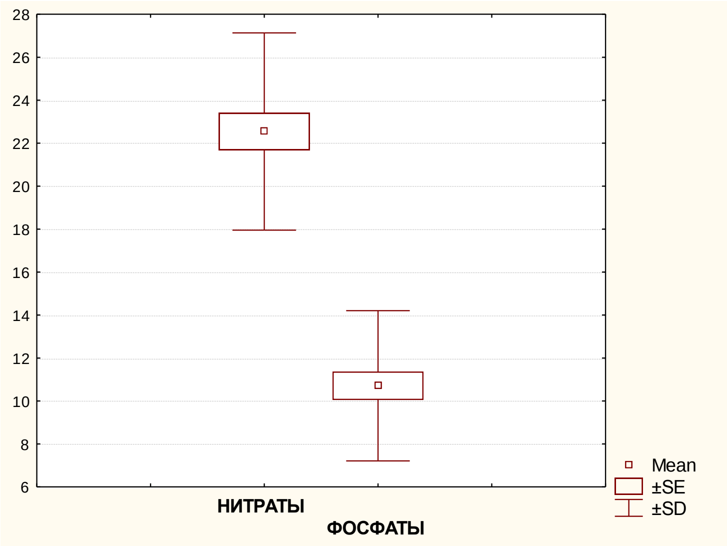


Рис. 4.12. График «ящик с усами»
для переменных НИТРАТЫ и ФОСФАТЫ

Таблица 4.9

Таблица результатов t-критерия

Выделенные курсивом различия значимы на уровне p <0,05000							
	Mean	Std.Dv.	N	Diff.	t	df	p
НИТРАТЫ	22.08	4.75					
ФОСФАТЫ	11.11	3.38	36	10.96667	12.87462	35	7.77E-15

В столбцах таблицы последовательно даны: средние значения величин НИТРАТЫ и ФОСФАТЫ, стандартные отклонения, число наблюдений, разность между средними значениями переменных, значение статистики t-критерия (критерия Стьюдента), число степеней свободы, уровень значимости. В этой таблице нас интересует последний столбец – уровень значимости для рассчитанного значения t-критерия при числе степеней свободы df=35. Как уже отмечалось ранее, нулевая гипотеза отвергается, если уровень значимости для рассчитанного критерия меньше 0,05. В нашем случае результат высоко значим на уровне значительно меньше 0,001, следовательно, мы с высокой степенью уверенности заключаем, что средние значения концентраций биогенов в заливе различны.

Упражнение для самостоятельного анализа

На основании данных табл. 4.10 с результатами тестирования и графика (рис. 4.13) принять или отвергнуть нулевую гипотезу о равенстве средних значений концентрации нитратов в заливах Посьета и Восток.

Таблица 4.10

Результаты тестирования

Группирующая: РАЙОН					
Группа 1: Посьета Группа 2: Восток					
Переменная	Mean Посьета	Mean Восток	t-value	df	p
НИТРАТЫ	21.4	24.32222	-1.146734	16	0.268351537

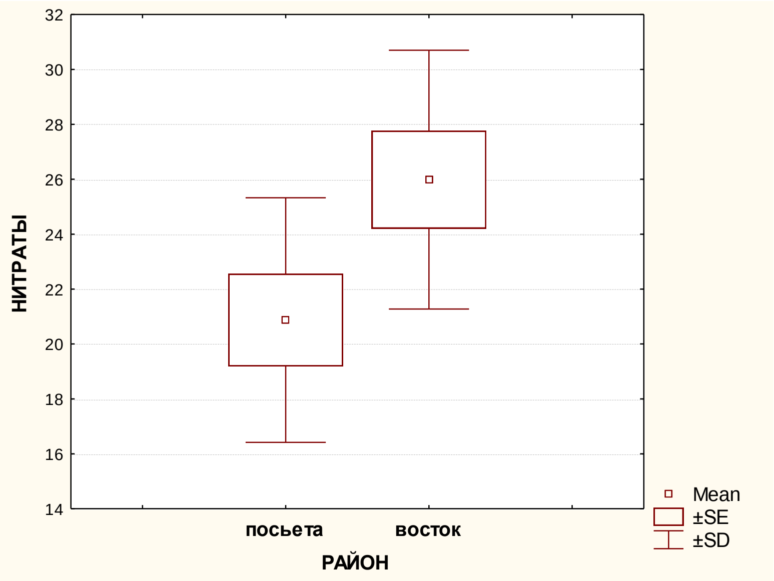


Рис. 4.13. График «ящики с усами»
для переменной НИТРАТЫ в заливах Посьета и Восток

4.2.2. Дисперсионный анализ

Суть дисперсионного анализа состоит в разложении общей дисперсии признака на составляющие: дисперсию, обусловленную влиянием конкретных факторов, их суммарным влиянием, комбинацией различных факторов, а также остаточную или случайную дисперсию. Термин «остаточная» означает, что эта дисперсия может быть вызвана случайными, не включенными в анализ факторами или ошибками измерений.

Дисперсионный анализ в пакете «Statistica» проводится в модуле ANOVA/MANOVA. Основываясь на данных табл. 4.1, попробуем с помощью однофакторного дисперсионного анализа решить задачу: «Носит ли процесс продуцирования сезонный характер?», т.е., выражаясь языком математической статистики, мы должны принять или отвергнуть нулевую гипотезу, состоящую в том, что средние значения продукции в различные сезоны года одинаковы. В данном случае вся дисперсия исходной переменной (ПРОДУКЦИЯ) будет разложена на дисперсию, обусловленную влиянием фактора СЕЗОН (межгрупповая дисперсия – MS effect) и остаточную (внутригрупповую) дисперсию – MS error. После активизации данного модуля и выбора переменных для анализа (зависимая – ПРОДУКЦИЯ, группирующая или фактор – СЕЗОН) запускаем процедуру вычислений дисперсий (табл. 4.11).

Таблица 4.11

Таблица дисперсионного анализа

Фактор 1-СЕЗОН					
df Effect	MS Effect	df Error	MS Error	F	p-level
2	3238.534	33	77.82746	41.61172	9.51E-10

В столбцах таблицы последовательно даны число степеней свободы для межгрупповой дисперсии, межгрупповая дисперсия, число степеней свободы для остаточной дисперсии, остаточная дисперсия, значение критерия Фишера и уровень значимости. Влияние фактора на изменчивость исходного признака признается значимым, если для значения критерия Фишера, рассчитываемого как отношение межгрупповой дисперсии к остаточной, уровень значимости меньше общепринятого в экологии – 0,05. В нашем случае уровень значимости значительно меньше и можно смело отвергнуть нулевую гипотезу и принять альтернативную – процесс продуцирования носит четко выраженный сезонный характер, т.е. средние значения продукции в различные сезоны года сильно различаются (рис. 4.14).

Как уже говорилось выше, с помощью дисперсионного анализа можно проверить влияние на результативный признак нескольких факторов, а также их совместное действие. Любого исследователя может заинтересовать вопрос: «Как процесс продуцирования зависит от таких факторов, как глубина и сезон?». В данном случае речь идет о двухфакторном дисперсионном комплексе, в котором вся дисперсия исходной переменной ПРОДУКЦИЯ будет разложена на четыре составляющие: дисперсия, обусловленная влиянием фактора СЕЗОН, дисперсия, обусловленная влиянием фактора ГЛУБИНА, дисперсия, обусловленная совместным влиянием этих факторов, и остаточная или случайная дисперсия. Результаты вычислений двухфакторного комплекса представлены в табл. 4.12.

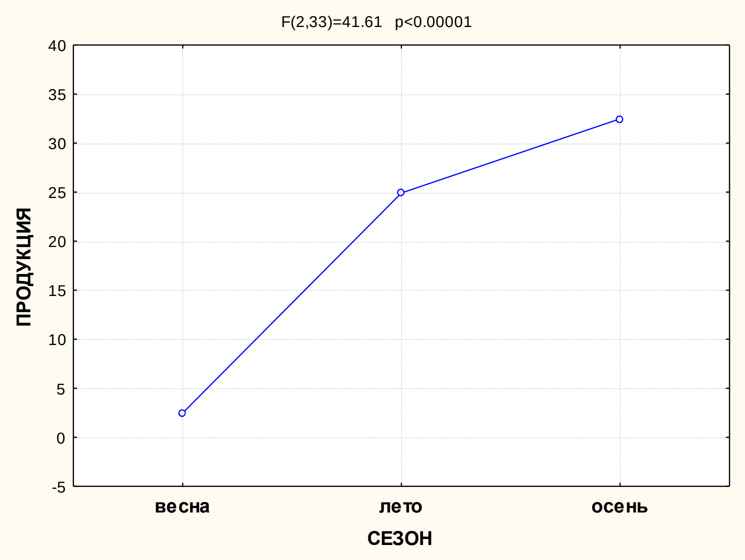


Рис. 4.14. Средние значения продукции по сезонам

Таблица 4.12

Результаты двухфакторного комплекса

Факторы	df Effect	MS Effect	df Error	MS Error	F	p-level
1 – сезон	2	3238.534	27	31.12871	104.0369	2.05E-13
2 – глубина	2	280.379			9.007087	0.001007
12 – взаимодействие	4	291.7683			9.372967	6.95E-05

Примечание. Обозначения в столбцах таблицы такие же, как в табл. 4.11.

В строках таблицы даны: 1 – дисперсия, обусловленная влиянием фактора СЕЗОН, 2 – дисперсия, обусловленная влиянием фактора ГЛУБИНА, 12 – дисперсия, обусловленная совместным влиянием этих факторов. Остаточная дисперсия указана в пятом столбце. Из таблицы видно, что уровень значимости для всех факторов и их совместного влияния очень маленький, ниже, чем 0,05, следовательно, процесс образования первичной продукции достоверно зависит от таких факторов, как СЕЗОН, ГЛУБИНА. Совместное их влияние также достоверно. На основании полученных данных можно также рассчитать степень влияния факторов на изменчивость переменной ПРОДУКЦИЯ, и узнать, сколько процентов дисперсии выпадает на долю случайных факторов. Степень влияния факторов (табл. 4.13) рассчитывается как частное от деления суммы квадратов отклонений для соответствующего компонента дисперсии на общую сумму квадратов отклонений исходного признака, в сумме они должны составлять 100%.

Таблица 4.13

Степень влияния факторов

Компоненты дисперсии	Число степеней свободы, df	Дисперсия, MS	Сумма квадратов=df·MS	Степень влияния факторов
Сезон	2	3238.534	6477.069	71.60641
Глубина	2	280.379	560.7579	6.199388
Взаимодействие	4	291.7683	1167.073	12.90243
Остаточная	27	31.12871	840.4751	9.291766
Общая			9045.375	100

Из таблицы видно, что почти 72% вариабельности исходной переменной ПРОДУКЦИЯ зависит от влияния такого фактора, как СЕЗОН, чуть больше 6% изменчивости обеспечивает фактор ГЛУБИНА, от их совместного влияния зависит почти 13% дисперсии результирующего признака, и 9% дисперсии зависит от случайных, не включенных в анализ факторов.

В общем случае для многофакторного дисперсионного комплекса дисперсия исходной переменной раскладывается на 2^p составляющих, где p – число факторов, включенных в анализ.

Упражнение для самостоятельного анализа

Изучалось влияние на вес иглокожих таких факторов, как грунт, вид иглокожих и условия обитания. Для этого был построен трехфакторный дисперсионный комплекс. На основании данных табл. 4.14 с результатами расчета дисперсионного комплекса сделать вывод о том, какие факторы и их комбинации влияют достоверно на вес иглокожих, влияние каких факторов незначительно и сколько процентов изменчивости исходного признака приходится на случайные факторы. Обозначения в столбцах таблицы такие же, как в предыдущих примерах.

Таблица 4.14

Результаты трехфакторного дисперсионного комплекса

Факторы	df Effect	MS Effect	df Error	MS Error	F	p-level
1 – условия	1	3582.022	24	1200.414	2.983989	0.096941
2 – вид	1	9399.303	24	1200.414	7.83005	0.00997
3 – грунт	2	4596.688	24	1200.414	3.829252	0.036029
12 – взаимодействие	1	101.0025	24	1200.414	0.08414	0.774255
13 – взаимодействие	2	217.5758	24	1200.414	0.181251	0.835358
23 – взаимодействие	2	1275.441	24	1200.414	1.062501	0.361293
123 – взаимодействие	2	1020.136	24	1200.414	0.84982	0.439957

4.2.3. Многомерные методы

4.2.3.1. Дискриминантный анализ

Цель дискриминантного анализа состоит в том, чтобы на основе измерения различных характеристик (признаков, параметров) объекта классифицировать его, т.е. отнести к одной из нескольких групп некоторым оптимальным способом. Число групп (классов) заранее известно, также заранее известно, к какой группе принадлежит каждый объект. Нужно построить правило, позволяющее по результатам измерений параметров указать группу, к которой он относится.

В случае двух групп объектов дискриминантный анализ эквивалентен множественной регрессии (зависимой переменной является номер группы). Независимые переменные с наибольшими стандартизованными коэффициентами регрессии дают наибольший вклад в предсказание принадлежности объекта к группе.

В модуле «Discriminant analysis/Дискриминантный анализ» ППП «Statistica» реализовано два общих метода этого анализа: стандартный и пошаговый (включения и исключения). В пошаговом методе модель строится последовательно по шагам: для метода включения на каждом шаге оценивается вклад в функцию дискриминации не включенных в модель переменных. Переменная, дающая больший вклад, включается в модель, далее система переходит к следующему шагу. В пошаговом методе исключения вначале в модель включаются все переменные, затем производится их последовательное исключение.

Приведем классический пример проведения дискриминантного анализа, заимствованный из книги Боровикова В.П. «Популярное введение в программу STATISTICA» (1998). Им является пример Фишера – анализ цветков ириса. Задача состоит в том, чтобы по результатам измерения длины и ширины чашелистиков и лепестков цветков ириса отнести ирис к одному из трех типов: SETOSA, VERSICOL, VIRGINIC. Данные для этого примера имеются в директории Example в файле с названием *irisdat*. В нем содержатся результаты измерения 150 цветков ириса, по 50 каждого типа (в табл. 4.15 приведены данные для первых десяти цветков).

Таблица 4.15

Структура исходных данных для проведения дискриминантного анализа

№ п/п	Длина листика	Шир. листика	Длина пестика	Шир. пестика	Тип ириса
1	5	3.3	1.4	0.2	SETOSA
2	6.4	2.8	5.6	2.2	VIRGINIC
3	6.5	2.8	4.6	1.5	VERSICOL
4	6.7	3.1	5.6	2.4	VIRGINIC
5	6.3	2.8	5.1	1.5	VIRGINIC
6	4.6	3.4	1.4	0.3	SETOSA
7	6.9	3.1	5.1	2.3	VIRGINIC
8	6.2	2.2	4.5	1.5	VERSICOL
9	5.9	3.2	4.8	1.8	VERSICOL
10	4.6	3.6	1	0.2	SETOSA

Для того чтобы начать дискриминантный анализ, необходимо выбрать переменные для анализа: независимыми переменными будут первые четыре, группирующей – тип ириса. Кроме этого, необходимо выбрать метод анализа, реализующий, например, пошаговый метод включения. После запуска вычислительной процедуры получаем следующую таблицу с результатами анализа (табл. 4.16).

В столбцах таблицы последовательно даны: значения статистики лямбда Уилкса (вероятность ошибки) для каждой включенной в модель переменной, значения F-критерия, связанного с лямбдой Уилкса, уровень значимости, значения толерантности.

Таблица 4.16

Результаты анализа данных

Независимые переменные	Wilks' Lambda	F-remove (2,144)	p-level	Toler.
Длина листика	0.035025	35.59018	2.76E-13	0.365126
Ширина листика	0.03058	21.93593	4.83E-09	0.608859
Длина пестика	0.031546	24.90433	5.14E-10	0.649314
Ширина пестика	0.024976	4.721151	0.010329	0.347993

Из таблицы видно, что в модель были включены все четыре переменные. Что означают приведенные статистики? Значения статистики лямбда Уилкса изменяются в интервале от 0 до 1. Если ее значение близко к 0, то это свидетельствует о хорошей дискриминации, и, соответственно, если значение лямбда Уилкса приближается к 1, то дискриминация была проведена плохо. Иными словами, от значения этой статистики зависит мощность дискриминации, которая вычисляется следующим образом:

Мощность=1 – вероятность ошибки (лямбда Уилкса).

Следовательно, если лямбда Уилкса близка к 0, то мощность дискриминации близка к 1, и наоборот.

Толерантность есть 1 минус квадрат множественной корреляции этой переменной с другими переменными модели. Переменные с малой толерантностью могут привести к ошибкам, т.к. переменная с малой толерантностью несет мало дополнительной информации и включать в модель ее нецелесообразно. Поэтому на начальном этапе работы в данном модуле необходимо в опции ТОЛЕРАНТНОСТЬ задать малое число (например *epsilon*), чтобы исключить из модели переменные с толерантностью меньше заданной. В рассматриваемом примере все переменные имеют достаточную толерантность.

Результаты анализа можно посмотреть графически. Для этого нужно вывести диаграмму рассеяния канонических значений (рис. 4.15).

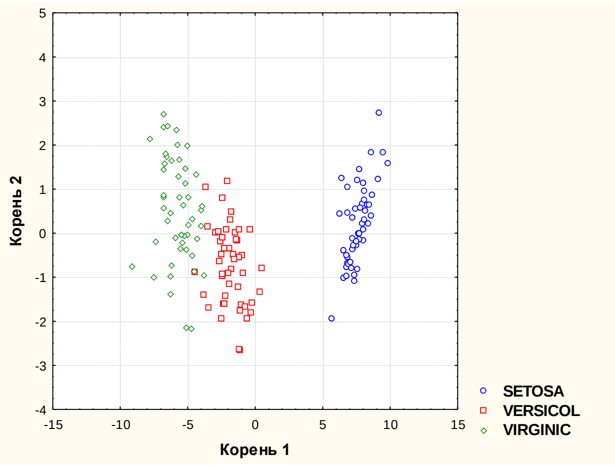


Рис. 4.15. Разделение трех типов ириса

На графике четко видно, что все цветки разделены на три группы, хотя имеется некоторое перекрывание у типов Versicol и Virginic, т.е. некоторые объекты были классифицированы неверно.

Начинающий исследователь, может задаться вопросом: «Для чего нужно было проводить дискриминацию, если мы и так знаем, какой цветок относится к какому типу?». Ответ на этот вопрос скрывается за целью анализа, т.е. мы на основании имеющихся данных должны построить такое правило, по которому сможем отнести уже неизвестный объект к одной из известных групп. Это можно сделать одним из трех способов.

Первый – вычисление классификационных меток с помощью функций классификации (табл. 4.17).

Таблица 4.17

Функции классификации, построенные пошаговым методом вперед

Переменные	SETOSA	VERSICOL	VIRGINIC
Длина листика	-16.4306	5.211451	12.76655
Шир. листика	23.58787	7.07251	3.68528
Длина пестика	-17.3984	6.434229	21.07911
Шир. пестика	23.54417	15.69821	12.44585
Константа	-86.3085	-72.8526	-104.368

Для новых цветков классификационные метки будут вычисляться по следующим формулам (табличные данные округлены):

$$SETOSA = -16.43 \cdot \text{Длина листика} + 23.59 \cdot \text{Шир. листика} - 17.40 \cdot \text{Длина пестика} + 23.54 \cdot \text{Шир. пестика} - 86.31$$

$$VERSICOL = 5.21 \cdot \text{Длина листика} + 7.07 \cdot \text{Шир. листика} + 6.43 \cdot \text{Длина пестика} + 15.7 \cdot \text{Шир. пестика} - 72.85$$

$$VIRGINIC = 12.77 \cdot \text{Длина листика} + 3.69 \cdot \text{Шир. листика} + 21.08 \cdot \text{Длина пестика} + 12.445 \cdot \text{Шир. пестика} - 104.37$$

Если мы получили новый цветок со значениями длины и ширины листика и пестика, то эти значения нужно подставить в формулы классификационных меток и вычислить значения меток для всех трех типов ириса. Новый цветок относится к тому классу, для которого классификационное значение максимально.

Второй способ классификации – использование расстояния Махаланобиса (квадрата расстояния Махаланобиса/Squared Mahalanobis distance) от каждого объекта до центров групп. В табл. 4.18 указаны квадраты расстояний Махаланобиса для первых десяти цветков. Цветок относится к группе, до которой расстояние Махаланобиса минимально.

Таблица 4.18

Квадраты расстояний Махаланобиса до центров групп

№ п/п	Наблюдаемые	SETOSA	VERSICOL	VIRGINIC
1	SETOSA	0.241888	90.66022	181.5587
2	VIRGINIC	208.5713	27.31885	1.894429
3	VERSICOL	105.2663	2.232916	13.072
4	VIRGINIC	207.918	31.74924	4.450617
*5	VIRGINIC	133.0668	5.252888	7.235931
6	SETOSA	1.333689	84.0118	170.0569
7	VIRGINIC	173.1838	26.56203	11.04843
8	VERSICOL	131.6617	8.430669	14.76468
*9	VERSICOL	130.8624	8.669699	6.506762
10	SETOSA	2.286424	113.6509	210.0239

Примечание. Неправильные классификации помечены *.

Каким образом работать с этой таблицей? Мы видим, что первый цветок относится к типу SETOSA, расстояние до центра группы SETOSA равно 0,24, и оно является минимальным. Данный цветок был классифицирован правильно, чего не скажешь о цветках под номерами 5 и 9.

Третий способ классификации – использование апостериорных вероятностей (Posterior probabilities). До анализа мы задаем для каждого случая вероятность, с которой он принадлежит к какому-то типу. После того, как анализ выполнен, можно пересчитать эти вероятности и получить апостериорные вероятности классификации. Объект относится к группе с максимальной апостериорной вероятностью. В табл. 4.19 приведены значения апостериорных вероятностей для первых десяти цветков.

Таблица 4.19

Апостериорные вероятности

№ п/п	Наблюдаемые	SETOSA	VERSICOL	VIRGINIC
1	SETOSA	1	2.32E-20	4.24E-40
2	VIRGINIC	1.4E-45	3.01E-06	0.999997
3	VERSICOL	4.21E-23	0.99559	0.00441
4	VIRGINIC	7.01E-45	1.18E-06	0.999999
*5	VIRGINIC	1.28E-28	0.729388	0.270612
6	SETOSA	1	1.11E-18	2.3E-37
7	VIRGINIC	6.2E-36	0.000428	0.999572
8	VERSICOL	1.67E-27	0.959574	0.040427
*9	VERSICOL	7.41E-28	0.253228	0.746772
10	SETOSA	1	6.57E-25	1.4E-45

Примечание. Неправильные классификации помечены *.

Интерпретация этой таблицы очень проста. В первом столбце указан тип ириса для каждого случая, в последующих – апостериорные вероятности отнесения каждого цветка к определенному типу. Например, известно, что четвертый цветок относится к типу VIRGINIC, именно к этому типу он и будет отнесен по результатам анализа, т.к. апостериорная вероятность, с какой он относится к данному типу, максимальна.

Добавим в исходную таблицу несколько пустых строк и внесем в них данные о цветках, тип которых нам неизвестен (табл. 4.20).

Таблица 4.20

Новые наблюдения о цветках ириса

№ п/п	Длина листика	Шир. листика	Длина лепестка	Шир. лепестка	Тип ириса
151	7	2	4	1	неизвестен
152	4.9	2.6	5.3	0.8	неизвестен
153	5.7	3	4.2	0.5	неизвестен

Вычисление классификационных меток довольно громоздко, поэтому попробуем классифицировать новые объекты с помощью квадратов расстояния Махаланобиса (табл. 4.21).

Таблица 4.21

Квадраты расстояний Махаланобиса от новых наблюдений до центров групп

№ п/п	Наблюдаемые	SETOSA	VERSICOL	VIRGINIC
151	неизвестен	103.9348	32.12595	64.50883
152	неизвестен	198.314	66.63412	75.63646
153	неизвестен	74.5899	30.58805	72.9121

Из таблицы видно, что все новые цветы относятся к типу VERSICOL, т.к. расстояние Махаланобиса до центра именно этой группы минимально. Проверим сделанный вывод с помощью классификации с использованием апостериорных вероятностей (табл. 4.22).

Таблица 4.22

Апостериорные вероятности для новых объектов

№ п/п	Наблюдаемые	SETOSA	VERSICOL	VIRGINIC
151	---	2.55E-16	1	9.29E-08
152	---	2.52E-29	0.989026	0.010974
153	---	2.79E-10	1	6.45E-10

Анализ этой таблицы подтверждает наш вывод о том, что все три новых цветка относятся к типу VERSICOL.

4.2.3.2. Кластерный анализ

Кластерный анализ объединяет различные процедуры, используемые для проведения классификации. В результате проведения кластерного анализа вся исходная совокупность объектов разделяется на кластеры (группы) схожих между собой объектов. Кластер – это группа объектов, обладающая свойством плотности (плотность объектов внутри кластера выше, чем вне его), дисперсией, отделимостью от других кластеров, формой, размером. Сложность задачи кластерного анализа состоит в том, что реальные объекты являются многомерными, описываемыми несколькими параметрами, и объединение объектов проводится в пространстве многих измерений.

Методы кластеризации делятся на агломеративные (объединительные) и итеративные дивизивные (разделительные). В агломеративных методах происходит последовательное объединение наиболее близких объектов в один кластер. Процесс такого последовательного объединения можно показать на графике в виде дендрограммы (дерева объединений).

Исходными данными для анализа могут быть объекты и их параметры или матрица расстояний между объектами. Если расстояния даны не сразу, то объединительные алгоритмы начинаются с вычисления расстояний между объектами. Расстояние между объектами – одна из мер сходства. Чем меньше расстояние между объектами, тем больше они схожи. Для измерения расстояния между объектами часто используют евклидову метрику, т.е. если объект описывается двумя параметрами, то его можно представить точкой на плоскости, тогда расстояние между объектами – это расстояние между точками, вычисленное по теореме Пифагора. Но в некоторых задачах евклидова метрика может оказаться неподходящей, т.к. не все объекты можно представить точкой на плоскости. В таких задачах используется понятие меры сходства (расстояние – одна из мер сходства). Важной мерой сходства являются статистические коэффициенты корреляции. Для бинарных объектов часто просто вычисляют количество параметров, которые совпадают у объектов, затем это число делят на число параметров и получают меру сходства. Этот коэффициент носит название простого коэффициента совстречаемости.

В программе Statistica доступны следующие меры сходства объектов: евклидова метрика, квадрат евклидовой метрики, манхэттенское расстояние (расстояние городских кварталов), метрика Чебышева, метрика Минковского, пирсоновский коэффициент корреляции, коэффициент совстречаемости.

В пакете реализованы следующие методы кластеризации – агломеративные: joining (tree clustering)/построение дендрограммы, two way joining/объединение по переменным и случаям одновременно; итеративный – *k*-means clustering/метод *k*-средних.

Правил иерархического объединения кластеров также несколько. Из них наиболее часто применяемыми являются метод одиночной связи, метод полной связи и метод средней связи. В методе одиночной связи на первом шаге объединяются два объекта, имеющие между собой максимальную меру сходства. На следующем шаге к ним присоединяется объект с максимальной мерой сходства с одним из объектов кластера и т.д. Для включения объекта в кластер требуется максимальное сходство лишь с одним из объектов кластера. Отсюда и название метода одиночной связи, нужна лишь одна связь, чтобы присоединить объект к кластеру. Недостатком этого метода является то, что в результате образовываются слишком большие кластеры.

В методе полных связей мера сходства между объектами – кандидатом на включение в кластер и всеми членами кластера – не может быть меньше некоторого порогового значения. В методе средней связи мера сходства между кандидатом и членами кластера усредняется, например, берется просто среднее арифметическое мер сходства.

Напомним, что все агломеративные методы дают картину последовательного объединения объектов в один кластер, анализ дендрограммы позволяет визуально оценить, на какое количество кластеров целесообразно разбивать исходную выборку объектов.

Итеративный метод *k*-средних работает непосредственно с объектами, которые в результате классификации будут разделены на определенное число кластеров. В данном методе объект относится к тому классу, расстояние (евклидово) до которого минимально. Метод *k*-средних работает следующим образом:

1) вначале задается некоторое разбиение данных на кластеры, вычисляются центры тяжести кластеров;

2) происходит перемещение точек: каждая точка помещается в ближайший к ней кластер;

3) вычисляются центры тяжести новых кластеров;

4) шаги 2, 3 повторяются, пока не будет найдена стабильная конфигурация (т.е. кластеры перестанут изменяться) или число итераций не превысит заданное пользователем. Итоговая конфигурация и является искомой.

Рассмотрим конкретный пример. Имеются данные о питании горбуши, выловленной в Беринговом море в разных районах с 1993 по 2003 г. (Дулепова и др., 2005). Исходные данные представлены в табл. 4.23.

Как видно из таблицы, пищевой спектр горбуши имеет одиннадцать составляющих (данные указаны в процентах). Горбуша представлена пятью размерными единицами и выловлена в четырех районах Берингова моря в различные годы (всего имеется 17 выборок). С помощью кластерного анализа попробуем ответить на два вопроса:

1. Все ли гидробионты, входящие в состав пищевого спектра горбуши, играют одинаковую роль в питании?

2. Различаются ли спектры питания у рыб, выловленных в разных районах в различные годы и имеющих разный размер?

Таблица 4.23

	Эвфаузиевые	Амфиподы	Копеподы	Декаподы	Птероподы	Ойкоплевры	Гребневики	Медузы	Сагитты	Нектон	Прочие
8/45/03	4.4	0	1.1	0	0	0	0	0	0	95.5	0
12/35/03	0	14	7.1	0	0	0	0	0	0	78.9	0
12/45/03	12.7	14	25.2	18.8	8.3	0	0	0	1.9	19.1	0
12/55/03	11.7	9.7	7.2	54.5	15.1	0	0	0	0.2	1.6	0
8/45/93	16.9	3.1	1.9	0	6.4	0	0	0	0	71.7	0
12/35/93	6.9	6.7	38.4	0	12.2	0	0	0	0	35.7	0.1
12/45/93	31.1	10.1	7	0.5	13.7	0	0	0	0	37.3	0.3
12/55/93	36.5	2.5	6.5	0.3	23	0	0	0	0.1	31.1	0
8/45/95	44.6	0.1	2	0.5	7	0	0	0	0	40.5	5.3
8/55/95	45	25	0	0	21	0.1	0	0	0	7.8	1.1
11/45/95	87	0	0	0	6	5	0	0	0	1.5	0.5
11/55/95	10	20.2	0	0	9	10	0	0	19	30.2	1.6
12/45/95	48.1	12.9	16.7	0.7	6.1	0.3	0	0	1.8	13.3	0.1
12/55/95	10	20.2	7.6	0	9	0	0	0	19	29.8	4.4
8/20/02	49.6	27	23.4	0	0	0	0	0	0	0	0
8/25/02	54.3	34.9	8.5	0	0	0	0		2.2	0	0.1
12/20/02	17.6	61.2	12.6	4.4	0	0	0	0	0	4	0.2

Примечание. В первом столбце первая цифра означает номер района вылова, вторая – размер горбуши (см), третья – год вылова.

Начать исследование целесообразно с построения дерева объединения, для чего необходимо выбрать из предлагаемых методов кластеризации метод «joining (tree clustering)». На рис. 4.16-а,б представлены дендрограммы, показывающие процесс объединения составляющих пищевого спектра горбуши в кластер по методу одиночной и полной связи.

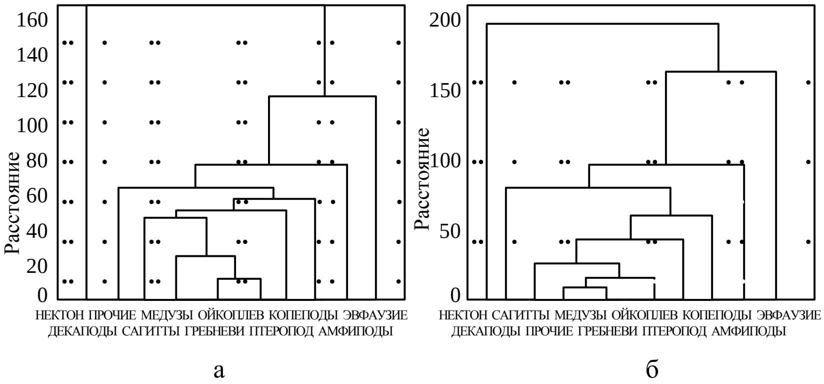


Рис. 4.16. Дерево объединения составляющих
пищевого спектра горбуши в кластер
(а – по методу одиночной связи, б – по методу полной связи)

Из рисунков видно, что выбор метода объединения практически не повлиял на результат, на дендрограммах четко выделяются по три кластера, причем два из них, видимо, будут содержать по одной переменной. Для проверки этого предположения воспользуемся итеративным методом «*k*-средних». Так как мы будем разбивать выборку на группы по всем параметрам пищевого спектра, то кластеризацию нужно проводить по переменным (т.е. по столбцам). В поле «число кластеров» зададим число групп, на которые мы хотим разбить исходную выборку – 3. В результате проведенного анализа получились три кластера (табл. 4.24, 4.25, 4.26).

Таблица 4.24

Первый кластер

<table>
<tr><td colspan="2">Члены кластера 1 и расстояния от центра кластера</td></tr>
<tr><td colspan="2">Кластер содержит 1 переменную</td></tr>
<tr><td></td><td>Нектон</td></tr>
<tr><td>Расстояние</td><td>0</td></tr>
</table>

Таблица 4.25

Второй кластер

Члены кластера 2 и расстояния от центра кластера					
Кластер содержит 9 переменных					
	Амфиподы	Копеподы	Декаподы	Птероподы	
Расстояние	17.42615	10.73309	11.91805	7.755975	
	Ойкоплевры	Гребневики	Медузы	Сагитты	Прочие
Расстояние	5.265337	5.398764	5.398764	6.190197	5.016996

Таблица 4.26

Третий кластер

Члены кластера 3 и расстояния от центра кластера	
Кластер содержит 1 переменную	
	Эвфаузиевые
Расстояние	0

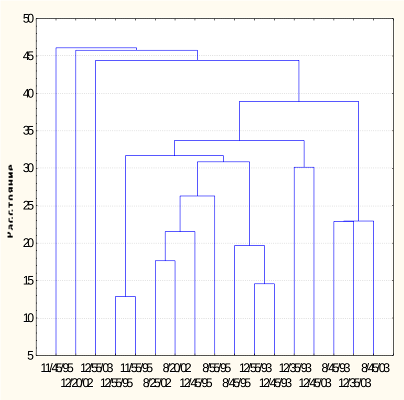
Теперь можно ответить на первый поставленный вопрос о роли различных гидробионтов в питании горбуши. В первый и третий кластеры попало по одной составляющей пищевого спектра, а именно нектон и эвфаузиевые, во второй – все остальные гидробионты. Молодь горбуши в основном питается эвфаузиевыми, соответственно, этот вид гидробионтов занимает особое место в питания горбуши. Остальные девять групп гидробионтов из второго кластера играют равноценную роль в питании горбуши и в отсутствии какой-либо группы могут быть заменены другими. Таким образом, с помощью кластерного анализа удалось ответить на первый вопрос.

Ответить на второй вопрос предлагается самостоятельно, исследовав дендрограммы объединения объектов в кластер (рис. 4.17-а,б) и табл. 4.27–4.29 с информацией о данных каждого кластера.

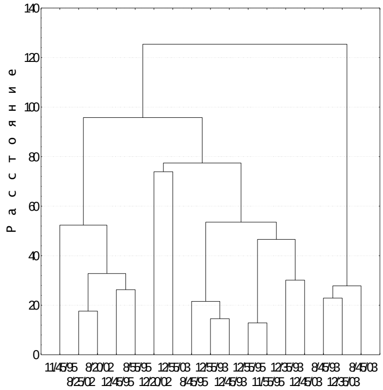
Таблица 4.27

Первый кластер

Члены кластера 1 и расстояния от центра кластера					
Кластер содержит 9 объектов					
	8/45/03	12/35/03	8/45/93	12/35/93	
Расстояние	14.96606	10.78433	7.147223	10.81935	
	12/45/93	12/55/93	8/45/95	11/55/95	12/55/95
Расстояние	5.923859	9.370584	9.315372	9.306107	8.724468



а



б

Рис. 4.17. Дерево объединения горбуши разных размеров
по сходству спектров питания в кластер
(а – по методу одиночной связи, б – по методу полной связи)

Таблица 4.28

Второй кластер

Члены кластера 2 и расстояния от центра кластера			
Кластер содержит 3 объекта			
	12/45/03	12/55/03	12/20/02
Расстояние	6.608595	10.98592	12.22014

Таблица 4.29

Третий кластер

Члены кластера 3 и расстояния от центра кластера					
Кластер содержит 5 объектов					
	8/55/95	11/45/95	12/45/95	8/20/02	8/25/02
Расстояние	6.59773	11.40363	4.801538	5.678944	5.208476

4.2.3.3. Канонический корреляционный анализ

Канонический корреляционный анализ позволяет находить максимальные связи между двумя группами переменных с совместным распределением. Напоминаем, что цель простого корреляционного анализа – нахождение взаимосвязей между двумя переменными. Подробно методология канонического корреляционного анализа изложена в учебном пособии «Системная экология» (Дулепов, Лескова, Майоров, 2004), в данном параграфе будет описан процесс проведения анализа с применением ППП «Statistica».

Поскольку цель канонического анализа – обнаружение взаимосвязей между двумя группами переменных, а ППП «Statistica» работает только с одной исходной матрицей данных, то все исходные данные будут представлены в одной рабочей книге (таблице). Задача исследователя будет состоять в том, чтобы разбить эту исходную матрицу данных, состоящую из p переменных, на два подмножества из r и q переменных так, что $p=r+q$. Корреляционная матрица, вычисляемая по основной матрице данных, будет состоять из блоков:

$$S = \begin{bmatrix} A_{rr} & C_{rq} \\ C'_{qr} & B_{qq} \end{bmatrix},$$

где A_{rr} – корреляционная матрица между членами множества r, B_{qq} – корреляционная матрица между членами множества q, C_{rq} – корреляционная матрица между членами двух множеств, C'_{rq} – транспонированная матрица.

На первом шаге отыскиваются линейные комбинации величин из каждого множества, имеющих максимальную корреляцию между собой. Эти линейные комбинации будут являться первыми осями новых координатных систем. Затем рассматриваются следующие линейные комбинации с наибольшей корреляцией между собой, но не коррелированные с первой линейной комбинацией и т.д. для всех возможных пар признаков. Найденные линейные комбинации есть не что иное, как канонические корреляции. Канонические корреляции и векторы, определяющие их веса, задаются собственными числами и собственными векторами двух матриц: $\left|C' A^{-1} C - B\right|$ и $\left|C B^{-1} C' - A\right|$. Обе матрицы имеют одинаковые собственные числа λ_i, а их собственные векторы дают коэффициенты линейных комбинаций для обеих групп переменных. Для собственных значений больших 0 рассчитываются коэффициенты канонических корреляций $R_{кан} = \sqrt{\lambda_i}$.

Рассмотрим канонический анализ на конкретном примере. Пусть имеются данные о количественном составе фито- и зоопланктона – численность инфузорий, диатомовых, перидиней и мелких жгутиковых, а также численность личиночных форм бентоса (табл. 4.30).

Таблица 4.30

Исходные данные для канонического анализа

Численность								
клеток фитопланктона, млн кл/м^3				личинок зоопланктона, экз./м^3				
Инфузории	Диатомовые	Перидинеи	Мелк. жгутиковые	Двустворчатые	Брюхоногие	Полихеты	Иглокожие	Усоногие
65	8100	500	100	13	59	142	24	25
41	303485	450	5915	5	5	432	29	118
22	259875	630	15000	98	38	6412	34	283
770	76200	10000	17000	770	2022	981	56	483
3480	80250	980	216000	5871	5844	431	1218	722
225	111650	450	97875	2051	920	670	307	175
485	108640	1120	149800	8250	868	1567	372	682
1350	523980	21780	40500	1452	88	506	129	109
4572	443872	4992	37856	500	160	500	15	1498
1040	508620	1260	29190	588	18	150	12	47
280	382900	1050	4200	169	14	218	8	30

С помощью канонического анализа попытаемся установить зависимость между численностями различных групп фитопланктона и численностями личинок бентосных животных. Для этого необходимо разбить исходную матрицу данных на два подмножества: 1) численность клеток фитопланктона, 2) численность личинок бентосных беспозвоночных.

Как видно из таблицы, численность планктона подвержена сильным колебаниям. Связаны ли увеличения численностей личинок бентоса с увеличением численностей клеток фитопланктона? В табл. 4.31 приведены коэффициенты корреляции между четырьмя переменными, характеризующими численность фитопланктона, и численностями пяти видов зоопланктона. Эта таблица соответствует матрице C_{rq} с четырьмя строками и пятью столбцами. Простое рассмотрение корреляционной матрицы малоинформативно: можно сделать только выводы о сильной или слабой корреляционной зависимости между переменными, а также трудоемко, если количество переменных, включенных в анализ, велико. По этой матрице, а также по корреляционным матрицам для каждого множества переменных вычисляются собственные числа и соответствующие им канонические корреляции.

Таблица 4.31

Коэффициенты корреляции между численностями различных видов фито- и зоопланктона

	Зоопланктон				
Фитопланктон	Двустворчатые	Брюхоногие	Полихеты	Иглокожие	Усоногие
Инфузории	0.199	0.456	-0.25	0.437	0.83
Диатомовые	-0.39	-0.483	-0.078	-0.41	-0.01
Перидинеи	-0.10	-0.08	-0.14	-0.13	0.02
Мелк. жгутиковые	0.896	0.79	-0.09	0.93	0.39

По грубому эмпирическому правилу канонические корреляции, меньшие 0,3 вряд ли окажутся значимыми, но иногда рассматривают и корреляции, близкие к этому порогу, чтобы выяснить, не поддаются ли они содержательной интерпретации. Число собственных значений объединенной корреляционной матрицы равно числу переменных в малом подмножестве, в нашем примере будут вычисляться четыре канонических корня, значения которых приведены на рис. 4.18 и в табл. 4.32.

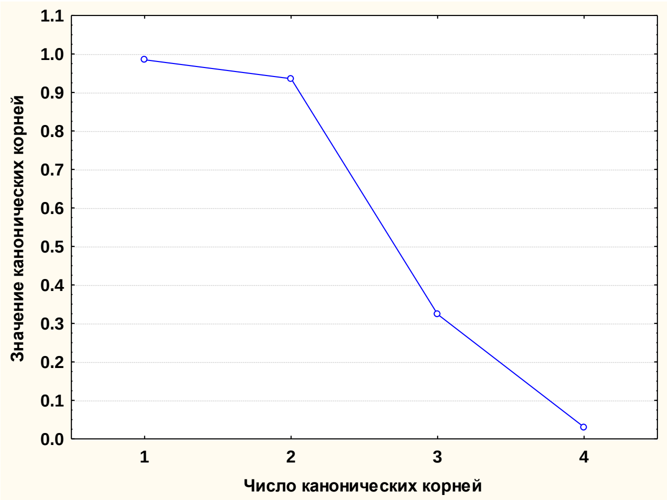


Рис. 4.18. График канонических корней

Из графика видно, что значимыми оказываются первые три канонические корреляции, поскольку их значения больше 0,3.

Таблица 4.32

Собственные числа и канонические корреляции

№ п/п	Собственное число	Каноническая корреляция	Доля дисперсии	
			Фитопланктон	Личинки бентоса
1	0.990668	0.995323	0.362102	0.497041
2	0.938285	0.968651	0.175519	0.156499
3	0.329839	0.574316	0.170906	0.03963
4	0.025469	0.159592	0.291473	0.158492

Первые три канонические корреляции отражают 71% вариабельности численности фитопланктона и около 69% вариабельности численности личинок бентосных животных. Анализ канонических корреляций дает непосредственную меру того, какая часть вариабельности двух множеств обусловлена взаимосвязями между ними, в нашем примере степень этой взаимосвязи очень велика. Веса канонических переменных для обоих подмножеств приведены в табл. 4.33 и 4.34 соответственно.

Таблица 4.33

Собственные векторы для численностей фитопланктона

Вид	Векторы для корреляций		
	1	2	3
Инфузории	0.522695	1.093272	0.206515
Диатомовые	-0.15503	-0.25424	-1.22922
Перидинеи	-0.00989	-0.09093	0.482325
Мелкие жгутиковые	0.629681	-0.98382	-0.50949

Среди численностей фитопланктона первая корреляция приписывает наибольший вес численности мелких жгутиковых и чуть меньше – численности инфузорий, а среди личинок зоопланктона – численности иглокожих. Следовательно, увеличение численности иглокожих прямо пропорционально увеличению численностей инфузорий и мелких жгутиковых. Вторая корреляция противопоставляет численность инфузорий численности мелких жгутиковых в первом подмножестве и численность усоногих численности двустворчатых во втором. Причем эти комбинации тесно взаимосвязаны друг с другом. И, наконец, третья каническая корреляция связывает противоположности между численностями брюхоногих и иглокожих с отсутствием диатомовых.

Таблица 4.34

Собственные векторы для численностей личинок бентоса

Вид	Векторы для корреляций		
	1	2	3
Двустворчатые	0.000947	-0.90758	0.713
Брюхоногие	-0.20844	0.091356	3.034912
Полихеты	-0.09728	-0.21099	0.069771
Иглокожие	0.925259	-0.0212	-3.13228
Усоногие	0.486677	0.836421	-0.30988

На рис. 4.19 (а, б, в) показаны диаграммы рассеивания значений переменных по каждому каноническому корню. Смысл этих диаграмм прост, если значения найденных канонических переменных вытянуты вдоль воображаемой оси эллипса, то каноническая корреляция высоко значима. Анализируя графики, этот вывод можно сделать относительно первого и второго канонического корня. Разброс значений по третьему корню довольно велик, тем не менее, эта каноническая корреляция была включена в анализ, т.к. она объясняет более 17% дисперсии численности фитопланктона, а ее собственное число больше 0,3.

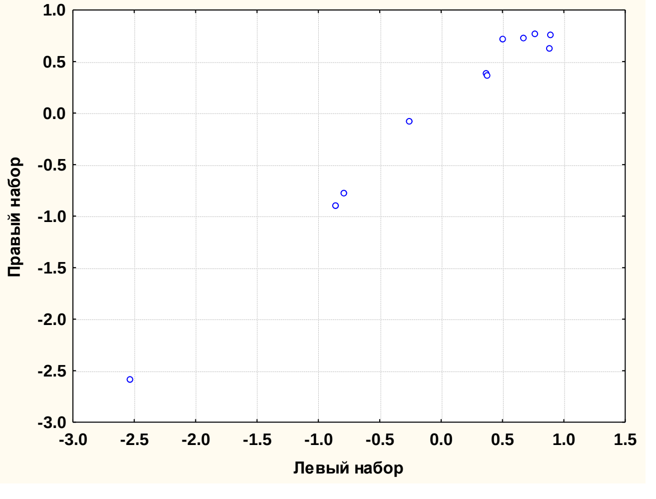


Рис. 4.19-а. Диаграмма рассеивания по первому каноническому корню

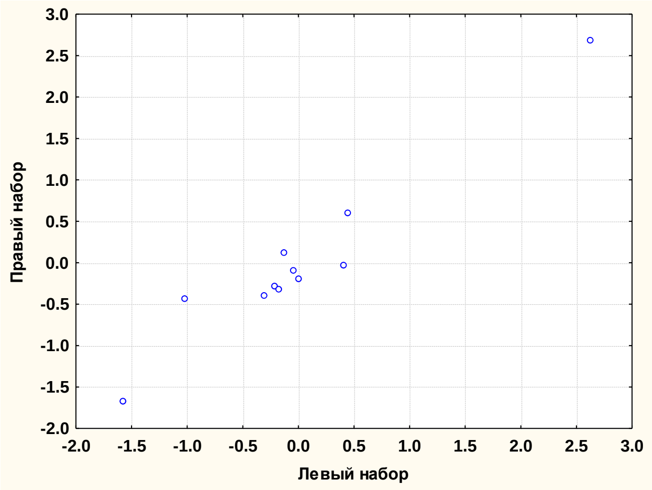


Рис. 4.19-б. Диаграмма рассеивания по второму каноническому корню

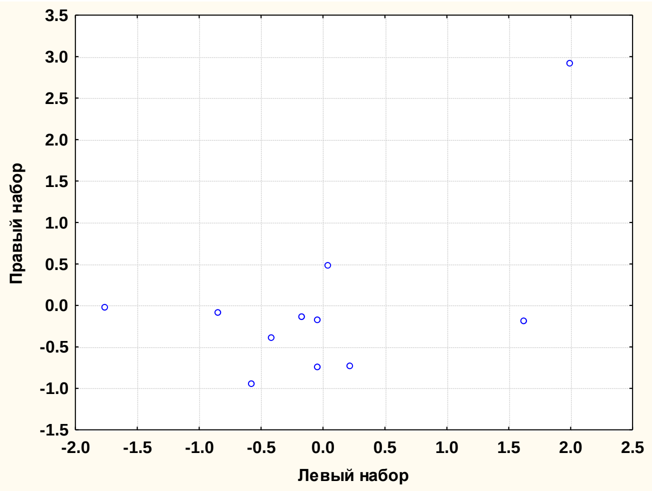


Рис. 4.19-в. Диаграмма рассеивания по третьему каноническому корню

Многомерная модель канонических корреляций представляет собой мощный инструмент исследования и обобщения сложных взаимосвязей между двумя множествами переменных, но для получения достоверных оценок весов канонических переменных рекомендуется использовать как минимум в 20 раз больше наблюдений, чем число переменных, включенных в анализ. В приведенном же примере объем выборки очень мал, поэтому распространять сделанные выводы о взаимосвязи численностей фито- и зоопланктона на реальные водные экосистемы нецелесообразно.

4.3. Математические модели процессов в биосистемах

Биологические системы – классы больших систем, обладающих рядом специфических особенностей, характеризующих жизнь: обменом веществ, способностью расти и размножаться, реагировать на внешние воздействия. Биологические системы являются открытыми системами в том смысле, что они получают извне вещества и энергию и создают из них сложные структуры. Биологические системы могут существовать только благодаря развитию подсистем управления, регулирующих реакции обмена веществ и жизнедеятельности организма. Они обладают способностью воспринимать и обрабатывать информацию управляющих сигналов.

4.3.1. Иерархия процессов в биосфере

По степени сложности внутренних связей и взаимодействия с окружающей средой принято выделять пять основных уровней биологических систем (Дулов, Цибаров, 2001).

Одноклеточные (вирусы, бактерии, простейшие микроорганизмы) могут образовывать колонии; уровень их организации невысок, но по количеству обрабатываемой информации они значительно превосходят технические системы. Именно с этим обстоятельством связаны попытки создания нового поколения компьютеров на бактериальной основе.

Многоклеточные – животные и растительные организмы, тела которых состоят из многих специализированных клеток, объединенных в ткани и органы; образуют уже единый организм, а не колонию, как одноклеточные. По отношению к многоклеточным можно говорить о следующем, более высоком уровне организации биосистем.

Популяция – совокупность особей одного вида, объединенных общностью места и времени проживания (в результате чего эти особи имеют возможность общаться между собой), которая воспроизводит себя в течение большого числа поколений, длительно занимает определенное пространство. Это еще более высокий уровень организации биосистемы, обладающий дополнительными, более сложными, связями, чем многоклеточные. В определенном смысле популяцию можно считать системой, составленной из организмов.

Биогеоценоз – сообщество живых организмов (не обязательно одного вида) вместе со средой обитания, ограниченное некоторыми природными границами и имеющее одно или несколько устойчивых состояний.

В историческом плане биогеоценоз – это сообщество организмов, приспособившихся друг к другу и к окружающей их среде в процессе эволюции. Они связаны друг с другом и со средой обитания разнообразнейшими каналами, по которым циркулируют потоки массы, энергии и информации. Взаимоотношение организмов – следствие борьбы за существование и естественного отбора при заданных условиях среды.

Биосфера – область активной жизни, охватывающей нижнюю часть атмосферы, гидросферу и верхнюю часть горных массивов; в биосфере все живые организмы и вся среда их обитания образуют целостную динамическую систему. В эту область попадают и результаты деятельности человека.

Иногда говорят, что биосфера – объединение экосистем мира. Некоторые ученые считают экосистему тождественной биогеоценозу. Однако по современным представлениям они не совсем совпадают. Экосистема подчеркивает лишь наличие каналов связи, по которым циркулируют потоки массы, энергии и информации. У экосистем нет пространственных ограничений.

4.3.2. Модель Вольтерра для однородной популяции

Модель Вольтерра по степени сложности и детальности описания относится к однородным детерминистским моделям.

Уравнение Вольтерра. Рассмотрим популяцию некоторых организмов, численность которой ограничена сверху каким-либо лимитирующим фактором (ограниченный ареал, ограниченное количество пищи и т.п.). Численность популяции в момент времени t описывается объемной (числовой) плотностью $N(t)$. В соответствии с этим динамика ее численности описывается уравнением

$$\frac{dN}{dt} = F^{+} - F^{-}, \tag{4.3}$$

где F – заданная функция или функционал.

Предполагается, что естественный прирост особей пропорционален их численности. Тогда

$$F^{+} = \alpha N. \tag{4.4}$$

Коэффициент пропорциональности $\alpha>0$ носит название коэффициента естественного прироста особей.

В процессе конкуренции за пространство, пищу в популяции преобладают парные взаимодействия. В этом случае убыль особей пропорциональна произведению численностей, т.е.

$$F^{-} = \gamma N^{2}. \tag{4.5}$$

Коэффициент $\gamma>0$ учитывает конкуренцию внутри популяции, т.е. учитывает убыль при естественном отборе. Поэтому его называют коэффициентом внутривидовой конкуренции особей.

При сделанных предположениях динамика численности популяции описывается уравнением Вольтерра, полученным им в 1931 г.:

$$\frac{dN}{dt} = (\alpha - \gamma N)N \,. \tag{4.6}$$

Если вынести за скобки коэффициент α и сделать замену $N = (\alpha / \gamma)n$, то уравнение (4.6) примет вид

$$\frac{dn}{dt} = \alpha(1-n)n \,. \tag{4.7}$$

Это – так называемое логистическое уравнение. Оно относится к классу нелинейных обыкновенных дифференциальных уравнений и в результате решения преобразуется к форме

$$n = \frac{n_0 \exp(\alpha t)}{1 - n_0 + n_0 \exp(\alpha t)} \,. \tag{4.8}$$

При $t\to\infty$ значение $n\to1$. Уравнение (4.7) имеет два стационарных решения: $n=1$, $n=0$. Второе решение – точка минимума, первое – точка максимума значений безразмерной численности $n(t)$.

Численность популяции описывается логистическим законом (4.8). В общем случае он обладает следующими особенностями. На начальном участке кривой $n(t)$ лимитирующий фактор γ не играет особой роли. На этом участке плохое влияние популяции мало сказывается на среде обитания и на предельной численности популяции. Картина меняется, когда популяция выходит на участок кривой, где скорость роста достигает своего максимального значения. Здесь влияние лимитирующего фактора начинает сказываться. Влияние популяции на среду своего обитания и обратное влияние среды на популяцию становятся существенными. Результатом «вредного» влияния популяции на среду обитания является, как правило, снижение предельно возможной численности самой популяции.

4.3.3. Непрерывный процесс культивирования микроорганизмов

В реальной действительности часто приходится изучать процессы при наличии нескольких видов популяций. Мы рассмотрим N видов микроорганизмов, каждый из которых имеет концентрацию N_i. Для их питания нужно M субстратов. Концентрация субстратов обозначается через c_j. Эволюция биосистемы описывается набором величин $N_i\ (i = \overline{1, N})$, $\tilde{n}_j\ (j = \overline{1, M})$.

Считаем, что микроорганизмы и субстраты заключены в некотором объеме V. В него извне подается жидкость вместе с питательными субстратами, имеющими в момент подачи концентрации c_j^0. Интенсивность подачи питательного раствора (его расход) равна Q. Чтобы не было переполнения, питательный раствор выводится из объема с такой же интенсивностью. Вместе с ним выводятся субстраты и микроорганизмы.

Через V_{ij} обозначим скорость, с которой I-й организм выедает j-й субстрат. Эта величина носит название функции выедания. В результате поедания субстратов питательной среды возрастает биомасса какого-либо вида микроорганизмов. Обозначим через k_{ij} долю новой массы I-го вида, которая возникла в результате поедания j-го субстрата.

Таким образом, перечислен некий набор характеристик, имеющих биологическую природу: концентрация, скорость выедания субстратов и доля прироста биомассы.

При такой постановке задачи отсутствует межвидовая борьба. Возможна лишь конкуренция из-за ограниченности объема и пищи. Следовательно, условия соответствуют модели Вольтерра.

В кинетике любых процессов принято пользоваться не скоростями реакций V_{ij}, а константами таких реакций $\overline{V_{ij}}$. Выразим скорость выедания субстратов через ее константу. Будем считать, что особи придерживаются раздельного питания. Это утверждение соответствует гипотезе парного взаимодействия. Тогда

$$V_{ij} = \overline{V_{ij}} c_j . \qquad (4.9)$$

Выведем систему дифференциальных уравнений для концентраций N_i и c_j. Искомые параметры удовлетворяют системе матричных уравнений

$$\frac{d\hat{O}}{dt} = F^+ - F^- , \qquad (4.10)$$

где принято обозначение $\hat{O} = \begin{pmatrix} N_i \\ c_j \end{pmatrix}$.

Будем считать, что микроорганизмы не поступают в объем V через область ввода питательной жидкости. В этом случае прибыль микроорганизмов связана с приростом за счет съедания субстратов со скоростью выедания (V_{ij}) и пропорциональна доле прироста биомассы (k_{ij}). Прибыль субстрата обусловлена поступлением питательного раствора и пропорциональна расходу Q. Поэтому с учетом (4.9) имеем

$$F^{+} = \begin{pmatrix} \sum_j k_{ij} \overline{V}_{ij} c_j N_i \\ c_j^0 \, Q \end{pmatrix}.$$

Убыль микроорганизмов связана только с их выносом из объема питательным раствором и пропорциональна расходу Q, т.е. равна QN_i. Убыль субстрата обусловлена выходом его при вытекании раствора и съеданием всеми микроорганизмами. Следовательно

$$F^{-} = \begin{pmatrix} QN_i \\ c_j Q + \sum_i \overline{V}_{ij} c_j N_i \end{pmatrix}.$$

После подстановки в (4.10) полученных значений F^{+}и F^{-} будем иметь следующую систему уравнений:

$$\frac{dN_i}{dt} = \sum_{j=1}^{M} k_{ij} \overline{V}_{ij} c_j N_i - QN_i \,, \tag{4.11}$$

$$\frac{dc_j}{dt} = (c_j^0 - c_j)Q - \sum_{i=1}^{N} \overline{V}_{ij} c_j N_i \,. \tag{4.12}$$

Эта система описывает непрерывный процесс культивирования микроорганизмов. Она замкнута, если заданы величины $Q, k_{ij}, \overline{V}_{ij}$ и c^0_j при всех $i = \overline{1, N}$ и $j = \overline{1, M}$. Ее стационарное решение находится из алгебраической системы

$$\sum_{j=1}^{M} k_{ij} \overline{V}_{ij} c_j = Q,$$

$$(\sum_{i=1}^{N} \overline{V}_{ij} N_i + Q) c_j \quad = c_j^0 Q.$$

Простейший вид системы (4.11), (4.12) соответствует однородной популяции (n=1, N_1=N) при одном субстрате (m=1, c_1=c), т.е. системе

$$\frac{dN}{dt} = (ck_{11} \overline{V}_{11} - Q)N \,, \tag{4.13}$$

$$\frac{dc}{dt} = (c_0 - c)Q - cV_{11}N. \tag{4.14}$$

Ее стационарным решением будут выражения

$$c = q/(k_{11}\overline{V}_{11}), \qquad N = c_0 k_{11} - Q/\overline{V}_{11}.$$

Рассмотренная здесь модель непрерывного процесса культивирования микроорганизмов является обобщением модели хемостата для микроорганизмов.

4.3.4. Модели «хищник – жертва»

Для биосистем с большей степенью организации, чем у одноклеточных, описанная выше модель далека от реальности, т.к. не учитывает взаимодействия между особями. Кроме того, она не учитывает физиологической ограниченности возможности потребления пищи особями и ограниченности самой пищи из-за ограниченности скорости ее воспроизведения и ареала обитания.

Реально у элементов и видов высокоорганизованных структур всегда возникает противоречие «интересов». Следовательно, для сохранения биогеоценоза нужно некоторое динамическое равновесие, то или иное устойчивое состояние.

Одной из проблем, возникающих при описании биогеоценозов, является размерность пространства независимых переменных и число искомых функций. Действительно, реальные биогеоценозы насчитывают сотни видов, которые, в свою очередь, подразделяются на отличающиеся друг от друга группы. Если в качестве фазовых переменных брать только численности групп, то и тогда динамическая система будет огромна. Но структура биогеоценоза иерархична. Их динамика определяется одним или несколькими видами, называемыми доминирующими видами. Поэтому в качестве фазовых переменных можно взять численности одного, двух или нескольких видов, рассматривая их взаимодействие с остальными видами и окружающей средой, подразделенной на биотическую и абиотическую.

В данном параграфе рассматриваются модели сосуществования только двух видов, из которых один («жертва») служит пищей для другого («хищника»).

4.3.4.1. Модель неограниченного потребления

Пусть N_1 и N_2 – численности жертв и хищников соответственно. Задача моделирования решается при следующих предположениях: пища жертв не лимитирована; хищник питается только жертвой; размножение жертв таково, что прирост их численности за малый промежуток времени пропорционален самой численности (т.е. прирост жертв равен Q_1/N_1); убыль жертв пропорциональна произведению численности жертв и хищников (т.е. убыль жертв равна $\gamma_{12}N_2N_1$); прирост хищников пропорционален численности жертв и хищников (прирост хищников равен $\gamma_{21}N_1N_2$).

Константа скорости гибели жертв (γ_{12}) носит название коэффициента конкурентоспособности жертв или коэффициента защиты, константа скорости прироста хищников (γ_{21}) – коэффициента конкурентоспособности хищников или коэффициента прожорливости, Q_1 и Q_2 – соответственно коэффициента размножения жертв и коэффициента смертности хищников, а величина $V_{21}=\gamma_{21}N_1$ – скорости выедания.

В соответствии с принятой постановкой задачи

$$\hat{O}=\begin{pmatrix} N_1 \\ N_2 \end{pmatrix}, \quad F^{+}=\begin{pmatrix} Q_1 N_1 \\ \gamma_{21}N_1N_2 \end{pmatrix}, \quad F^{-}=\begin{pmatrix} \gamma_{12}N_2N_1 \\ Q_2N_2 \end{pmatrix}.$$

Подстановка этих выражений в (4.10) приводит к системе из двух нелинейных обыкновенных дифференциальных уравнений:

$$\frac{dN_1}{dt}=(Q_1-\gamma_{12}N_2)N_1,$$
$$\frac{dN_2}{dt}=(\gamma_{21}N_1-Q_2)N_2.$$

Замена $N_1=\dfrac{Q_2}{\gamma_{21}}n_1, \quad N_2=\dfrac{Q_1}{\gamma_{12}}n_2$ преобразует эту систему к виду

$$\frac{dn_1}{dt}=Q_1(1-n_2)n_1\,, \tag{4.15}$$

$$\frac{dn_2}{dt}=Q_2(n_1-1)n_2\,. \tag{4.16}$$

Это модель ненасытного хищника, так как скорость выедания линейно зависит от численности жертв.

Качественное исследование системы. Рассмотрим фазовую плоскость (n_1, n_2). Для исследования траекторий на ней и особых точек системы (4.15) и (4.16) имеем уравнение

$$\frac{dn_2}{dn_1}=-\alpha\frac{n_2}{n_1}\frac{n_1-1}{n_2-1}\,, \tag{4.17}$$

$$\alpha=Q_2/Q_1.$$

Особая точка (0,0) являющаяся стационарной точкой исследуемой системы, интереса не представляет. Особая точка (1,1) – это точка типа центр; она является также стационарной точкой нашей системы, соответствующей равновесию.

Первый интеграл системы уравнений (4.15) и (4.16) совпадает с решением уравнения (4.17). Этот интеграл записывается в виде

$$\left(\frac{n_2}{\exp(n_2)}\right) = C\left(\frac{n_1}{\exp(n_1)}\right)^{-\alpha}, \quad C > 0. \tag{4.18}$$

На фазовой плоскости (n_1, n_2) он изображается замкнутыми кривыми с центром в точке (1,1) (рис. 4.20). Их называют траекториями на фазовой плоскости. Учитывая форму траекторий, решение (4.18) при заданных значениях C называют овалами Вольтерра.

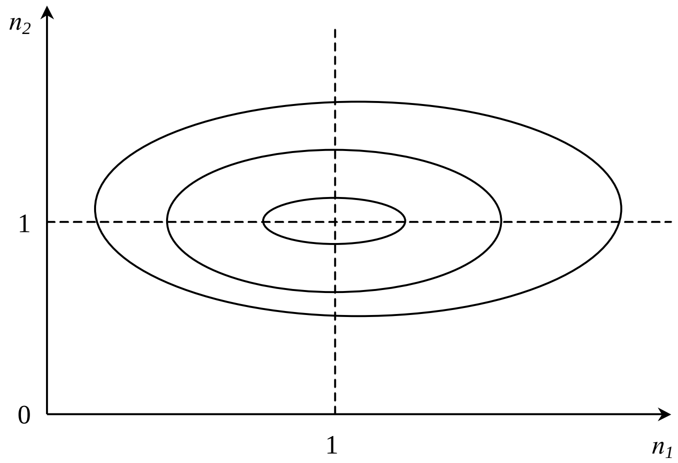


Рис. 4.20. Овалы Вольтерра

При малых колебаниях вблизи равновесного состояния (1,1) (рис. 4.21) их период равен

$$T = \frac{2\pi}{\ln 2}\sqrt{t_1 t_2}, \qquad t_i = \frac{\ln 2}{Q_i} \quad (i = \overline{1,2}).$$

Здесь t_1 – время, необходимое для удвоения численности особей первого вида, t_2 – время, необходимое для уменьшения вдвое численности особей второго вида.

В результате математического исследования рассматриваемой модели Вольтерра сформулировал три биологических закона.

Закон периодичности цикла: колебания двух видов являются периодическими и их период зависит от Q_1, Q_2 и C, т.е. от коэффициентов размножения и смертности, а также от начальных условий для численности обоих видов.

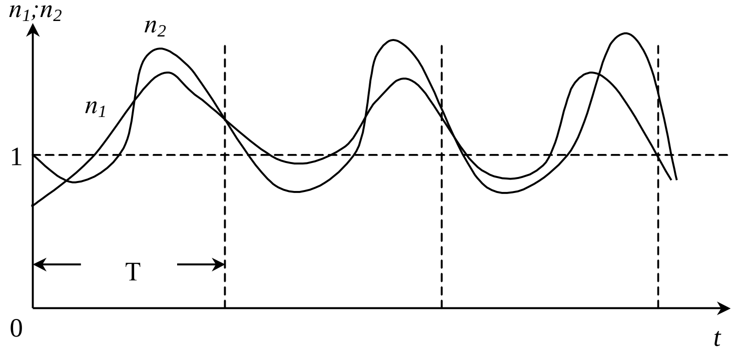


Рис. 4.21. Изменение численности особей

Закон сохранения средних значений: средние значения численности особей обоих видов постоянны, каковы бы ни были начальные значения численностей особей обоих видов, если сохраняются постоянными коэффициенты размножения и смертности обоих видов, а также условия нападения и защиты (т.е. сохраняются Q_1, Q_2, γ_{12}, γ_{21}).

Закон смещения средних значений: если истребление одинаково и пропорционально численности обоих видов, то наблюдается увеличение среднего значения численности пожираемого вида и уменьшение такового для пожирающего; усиленное покровительство (удобрение, подкормка, защита от вредителей и т.д.) над пожираемым видом увеличивает средние значения численностей.

Последний закон справедлив только при $Q_1>0$. При $Q_1 \leq 0$ оба вида в конце концов погибают.

4.3.4.2. Модели ограниченного потребления

Модель предыдущего раздела не учитывает такие существенные свойства реальных биогеоценозов, как ограниченность прожорливости хищника, скорости размножения жертв и численности последних.

Ни хищники в целом, ни их отдельные особи не могут потреблять бесконечное количество пищи. Поэтому необходимы ограничения на скорость выедания. В модели ограниченной прожорливости предполагается, что на начальном участке скорость выедания, как и в модели неограниченного потребления, пропорциональна численности жертв. После потребления некоторого их количества (N_1^*) скорость выедания перестает зависеть от численности жертв, и ее кривая выходит на плато (горизонтальный участок). Это означает, что

$$V_{21} = \begin{cases} \gamma_{21} N_1 & \text{ïðè} \quad N_1 \leq N_1^* \\ \gamma_{21} N_1^* & \text{ïðè} \quad N_1 \geq N_1^* \end{cases}.$$

В силу сделанных предположений относительно V_{21} модель биогеоценоза с ограниченной прожорливостью особей хищников будет состоять из двух групп систем:

$$\frac{dn_1}{dt} = Q_1(1 - n_2)n_1,$$
$$\frac{dn_2}{dt} = Q_2(n_1 - 1)n_2$$

при $n_1 \leq n_1^*$ и

$$\frac{dn_1}{dt} = Q_1(1 - n_2)n_1,$$
$$\frac{dn_2}{dt} = Q_2(n^*{}_1 - 1)n_2$$

при $n_1 \geq n_1^*$.

Реальная численность жертв всегда ограничена сверху. Более того, при некотором уровне численности, превышающем N_1^*, в популяции начинаются эпизоотии или другие неблагоприятные для данного вида явления.

Модель биогеоценоза с ограниченной численностью жертв описывается такой же системой уравнений (4.15) и (4.16), но с ограничением $n_1 \leq n_1^*$. Эта система, выйдя на ограничение $n_1 = n_1^*$, ведет себя как некоторая популяция (в нашем случае хищников) с экспоненциальным ростом или убыванием в зависимости от значений параметров Q_2, γ_{21} и N_1^*.

4.3.4.3. Модели с ограниченной скоростью размножения

Посмотрим, как будет выглядеть система уравнений, описывающих эволюцию биогеоценоза, при лимитированной рождаемости жертв. В рамках этой модели скорость размножения жертв задается выражением:

$$Q_1 N_1 = \begin{cases} Q_1 N_1, \\ Q_1 N_1^* \end{cases}.$$

Следовательно, когда численность жертв превышает определенный предел, скорость их размножения становится постоянной (не зависящей от N_1). Сама модель биогеоценоза с ограниченной рождаемостью жертв будет состоять из двух групп систем:

$$\frac{dn_1}{dt} = Q_1(1 - n_2)n_1,$$
$$\frac{dn_2}{dt} = Q_2(n_1 - 1)n_2$$

при $n_1 \leq n_1^*$ и

$$\frac{dn_1}{dt} = Q_1(n_1^* - n_2 n_1),$$
$$\frac{dn_2}{dt} = Q_2(n_1 - 1)n_2$$

при $n_1 \geq n_1^*$.

В зависимости от начальных данных модель биогеоценоза с ограниченной скоростью размножения жертв имеет два режима: если в начальный момент $n^0_1 < n^*_1$, то система полностью аналогична модели Вольтерра для неограниченного потребления; если $n^0_1 \geq n^*_1$, то система имеет квазипредельный цикл, для которого $n_1 = n^*_1$, и траектория касается овала Вольтерра.

4.3.5. Обобщение модели «хищник – жертва»

Модели «хищник – жертва», описанные в предыдущих разделах, не учитывают наличия различных видов хищников и жертв, а также внутривидовой конкуренции. Модель Костицына устраняет указанные недостатки.

Модель Костицына. Обозначим численность хищников i-го вида через N_i ($i = \overline{1, N}$), а численность жертв j-го вида через c_j ($j = \overline{1, M}$). Предполагается парность взаимодействий, включая внутривидовую конкуренцию и принцип раздельной диеты.

Пусть Q_i – коэффициент смертности хищников вида i, R_j – коэффициент размножения жертв вида j, γ_{ij} – константа скорости прироста хищников вида I за счет выедания жертв вида j, $\tilde{\gamma}_{ji}$ – коэффициент защиты жертв вида j от хищников вида i. Если внутривидовая конкуренция не существенна, то обобщенная модель «хищник – жертва» аналогична модели для процесса непрерывного культивирования микроорганизмов:

$$\frac{dN_i}{dt} = \left(\sum_{j=1}^{M} \gamma_{ij} c_j - Q_i \right) N_i,$$
$$\frac{dc_j}{dt} = \left(R_j - \sum_{i=1}^{N} \tilde{\gamma}_{ji} N_i \right) c_j.$$

Жертвы фактически играют роль питательных субстратов.

Чтобы учесть внутривидовую конкуренцию, введем в рассмотрение еще две группы коэффициентов, которые характеризуют убыль популяции (вида): δ_{kj} – коэффициент внутривидовой конкуренции хищников и

α_{ik} – коэффициент внутривидовой конкуренции жертв. Модифицированная модель хищник – жертва имеет следующий вид:

$$\frac{dN_i}{dt} = \left(\sum_{j=1}^{M} \gamma_{ij} c_j - Q_i \right) N_i - \sum_{k=1}^{N} \delta_{ik} N_k N_i, \tag{4.19}$$

$$\frac{dc_j}{dt} = \left(R_j - \sum_{i=1}^{N} \tilde{\gamma}_{ji} N_i \right) c_j - \sum_{k=1}^{M} \alpha_{jk} c_k c_j \ . \tag{4.20}$$

Это и есть модель Костицына. В этой модели все коэффициенты положительны.

Если i-й вид питается только (i-1)-м, являясь пищей (i+1)-му, а последний (N-й) вид никем не поедается, то биогеоценоз описывается системой

$$\frac{dN_i}{dt} = \left(Q_i - \sum_{j=1}^{N} \gamma_{ij} N_j \right) N_i \tag{4.21}$$

с соответствующей интерпретацией N_i и коэффициентов. Здесь

$$Q_1>0,\ Q_i<0 \text{ при } i>1;\ \gamma_{ij}=-\gamma_{ji}\ (i\neq j);$$

$$\gamma_{ii}\neq 0,\ \gamma_{i,i+1}\neq 0,\ \gamma_{i,i-1}\neq 0, \text{ прочие } \gamma_{ij}=0.$$

Качественные исследования системы (4.21) показали, что стационарная точка такого биоценоза всегда асимптотически устойчива. Решения системы представляют либо затухающие колебания около стационарной точки, либо монотонно стремятся к ней.

При N=M=1 система уравнений (4.19) и (4.20) имеет предельную устойчивую точку (узел) со строго положительными координатами, если выполнены условия

$$R_1\gamma_{11}>Q_1\alpha_{11},$$

$$\{R_1\delta_1(\alpha_{11}-\gamma_{11})+Q_1\alpha_{11}(\tilde{\gamma}_{11}+\delta_{11})\}^2\geq 4\gamma_{11}\tilde{\gamma}_{11}(R_1\delta_{11}+Q_1\tilde{\gamma}_{11})(R_1\gamma_{11}-Q_1\alpha_{11}).$$

Приведенные результаты исследования модели Костицына указывают на возможность стационарных режимов эксплуатации биосистем, что важно в хозяйственном аспекте. Такие режимы обеспечивают сохранность биоценоза в течение длительного времени, поддерживая при этом ритмичность и постоянство качества самого процесса эксплуатации.

Экстремальные свойства. Запишем систему уравнений (4.19) и (4.20) единообразно. Для этого введем следующие обозначения:

$$x_k = \begin{cases} c_k & \text{ïðè} \quad k\leq M, \\ N_i & \text{ïðè} \quad k=M+i, \end{cases} \qquad \varepsilon_k = \begin{cases} R_k & \text{ïðè} \quad k\leq M, \\ -Q_i & \text{ïðè} \quad k=M+i. \end{cases}$$

Кроме того, введем матрицу конкуренции:

$$\gamma_{ki} = \begin{cases} \tilde{\gamma}_{kl} & \text{ïðè} \quad k \le M, \quad M < l \le M+N, \\ \alpha_{kl} & \text{ïðè} \quad k,l \le M, \\ -\gamma_{kl} & \text{ïðè} \quad M < k \le M+N, \quad l \le M, \\ -\delta_{kl} & \text{ïðè} \quad M \le k, \quad l = M+N, \\ 0 & \text{ïðè} \;\; \text{ïðî÷èõ} \quad k,l. \end{cases}$$

Тогда наша система преобразуется к форме

$$\dot{x}_k = \frac{dx_k}{dt} = x_k(\varepsilon_k - \sum_l \gamma_{kl} x_l). \tag{4.22}$$

Потребуем симметрии матрицы конкуренции (т.е. свойства $\gamma_{kl}=\gamma_{lk}$) и перейдем от переменных $x_k, \dot{x}_k$ к переменным $\xi_k, \dot{\xi}_k$ по формулам

$$2\sqrt{x_k} = \xi_k, \qquad \frac{dx_k}{dt} = \frac{\xi_k}{2}\frac{d\omega_k}{dt}. \tag{4.23}$$

Введем функцию

$$W = U_r - \frac{1}{2}\sum_{k,l} \gamma_{kl} x_l x_k, \tag{4.24}$$

$$U_R = \sum_k \varepsilon_k x_k$$

или в новых переменных

$$W = \frac{1}{4}\sum_k \varepsilon_k \xi_k^2 - \frac{1}{32}\sum_{k,l} \gamma_{kl} \xi_k^2 \xi_l^2. \tag{4.25}$$

Непосредственной подстановкой (4.23) в (4.22) и дифференцированием (4.25) по ξ_k можно убедиться в справедливости соотношений

$$\frac{d\xi_k}{dt} = \frac{\partial W}{\partial \xi_k}, \tag{4.26}$$

$$k = \overline{1 + M + N}.$$

Учитывая (4.26), получаем

$$\frac{dW}{dt} = \sum_k \frac{\partial W}{\partial \xi_k}\frac{d\xi_k}{dt} = \sum_k \left(\frac{\partial W}{\partial \xi_k}\right)^2 \ge 0. \tag{4.27}$$

Отсюда следует непрерывность возрастания $W(t)$ вдоль траектории развития биологического сообщества.

Выражения (4.22), (4.23), (4.26) и (4.27) позволяют сделать два важных вывода:

• стационарная точка x^* динамической системы (4.22) является стационарной точкой функции W, и наоборот;

• стационарная точка $W^* = W(x^*)$ является одновременно точкой экстремума функции W, и наоборот.

Таким образом, можно утверждать, что система устойчива по Ляпунову, если функция W обладает максимумом W^*, т.к. $L = W^* - W$ представляет собой функцию Ляпунова для системы (4.22).

Сумма U_R носит название репродуктивного потенциала. Он характеризует прирост биомассы сообщества за счет внутреннего резерва в случае, когда конкуренция и лимитирование ресурсов отсутствуют. Другими словами, U_R определяет скорость прироста биомассы только за счет репродуктивных физиологических возможностей организмов. Более того, как следует из (4.24) и (4.27), сообщество в процессе эволюции стремится максимизировать свой репродуктивный потенциал.

Двойная сумма в (4.24) характеризует усредненные затраты на конкуренцию. Следовательно, в процессе развития сообщество стремится их минимизировать. Таким образом, эволюция биологической системы происходит по траектории, увеличивающей репродуктивный потенциал и снижающей потери из-за конкуренции.

Учет миграции в модели биогеоценоза. Приведенные однородные модели биогеоценоза не учитывают изменение численности за счет эмиграции или иммиграции особей относительно границ ареала. Если численности особей любых видов очень велики, то в обычных условиях процесс миграции не существенен, поскольку числом особей, пересекающих границу ареала, можно пренебречь по сравнению с обитающими внутри него. Однако в каких-то экстремальных условиях (например, при пожарах или сильных наводнениях) миграцией пренебречь нельзя. Эмиграцию или иммиграцию малочисленных (особенно ценных) видов необходимо учитывать и в условиях, типичных для данного места обитания.

В рамках однородных моделей миграция особей учитывается только с помощью дополнительных членов типа источников. Тогда весьма многочисленный класс биологических систем (биогеоценозы, популяции) может быть описан системой динамических уравнений вида

$$\frac{dx}{dt} = \alpha_i + \sum_{j=1}^{n} \varepsilon_{ij} x_j - \sum_{j=1}^{n} \gamma_{ij} x_j x_i, \qquad i = \overline{1, n}. \tag{4.28}$$

Здесь $\{x_i\}$ – набор фазовых переменных динамической системы (4.28). Ими могут быть численности видов, размеры возрастных групп популяций. В моделях биогеоценоза под x, как правило, понимают численность особей группы i, обладающей дискретным набором некоторых индивидуальных признаков (вид, возрастная группа, пол и т.п.). Параметр α_i характеризует иммиграцию (α_i>0) или эмиграцию (α_i<0) *i*-й группы между соседними биоценозами; $\varepsilon_{ii}=\varepsilon_i$ – коэффициент естественного прироста или убыли *i*-й группы (размножения или смертности); ε_{ij} – коэффициент при $i\neq j$ характеризует взаимные превращения особей и связанные с этим переходы между группами (например переход из одной группы в другую). С помощью коэффициентов γ_{ii} учитывают внутривидовые лимитирующие факторы (ограниченность ареала, внутривидовая борьба и т.д.), а с помощью параметров γ_{ij} при $i\neq j$ – взаимодействие между особями разных групп (межвидовая борьба, симбиоз, паразитизм и т.п.).

Биогеоценоз характеризуется общностью территории и, как правило, замкнутость системы, что позволяет пренебречь миграционными процессами. Межвидовые «браки», приводящие к рождению особей, отсутствуют. Поэтому

$$\alpha_i \cong 0,\ \varepsilon_{ij} = 0 \text{ при } (i \neq j),\ \varepsilon_{ii} = \varepsilon_i .$$

Если нас не интересует разделение особей одного вида по признакам индивидуальности (по возрасту, полу, размерам и т.д.), тогда система (4.28) примет вид модели Вольтерра-Костицына, которую можно считать простейшей динамической моделью биогеоценоза.

4.3.6. Модель непрерывного распределения численности по индивидуальным признакам

При постановке оптимальных задач или при исследовании биосферы не всегда удается ограничиться моделями первого уровня описания. Так, например, особи определенного возраста или размера оцениваются по-разному. Следовательно, в этом случае нужно прогнозировать динамику роста и накопления массы. Решение проблемы сохранности видов в биосфере также требует исследования динамики численности видов с учетом распределения по ряду индивидуальных признаков. Для таких целей требуется построение дискретно-непрерывных математических моделей динамики биогеоценозов.

Пусть t – астрономическое время; τ – индивидуальное время (возраст); $x=\{x_i\}$ – набор прочих, за исключением возраста, признаков индивидуальности элемента системы, т.е. особи. Система, описывающая эволюцию одной особи, представляется как

$$\dot{x} = dx/dt = v(x,\tau,t),$$

$$\dot{\tau} = d\tau/dt = 1$$

где v – заданный вектор).

Эту же систему можно представить как систему динамических уравнений

$$\dot{x}_i = v_i(x,\tau,t),$$
$$\dot{\tau} = 1, \qquad i = \overline{1,n}.$$

Требуется построить математическую модель эволюции всего сообщества.

Обозначим через N_α числовую плотность особей вида α в момент времени t в возрасте τ с прочими признаками индивидуальности, заданными вектором $x \in R^n$. Тогда очевидно, что $N_\alpha = N_\alpha(t,\tau,x)$. Будем считать численность особей столь большой, что для $\forall t$ встречаются особи всех $\tau \geq 0$ и x. Это означает непрерывность распределения по τ и x.

Если ω и $d\omega$ – соответственно объем и элементарный объем в пространстве индивидуальных признаков $x \in R^n$, а Ω и $d\Omega = d\omega\, d\tau$ – в расширенном пространстве $(x,\tau) \in R^{n+1}$, то (по смыслу N_α) количество особей вида α в фазовом объеме $d\Omega$ равно

$$dN_\alpha = N_\alpha d\Omega\,.$$

Общее число особей вида α

$$\hat{O}_\alpha(t) = \int\limits_\Omega N_\alpha)(x,\tau,t)d\Omega\,.$$

Так как $\tau \geq 0$, то Ω – верхняя часть полупространства $\{x,\tau\}$.

Поскольку в природе существует естественный отбор, включая смертность особей, то нужно ввести коэффициент смертности $D_\alpha(x,\tau,t)$, имеющий смысл вероятности гибели особи вида α, обладающей индивидуальными признаками x, в возрасте τ и в момент времени t. Смертность особей приводит к их убыли. Тогда суммарная убыль особей вида α

$$F_\alpha^- = \int\limits_\Omega D_\alpha N_\alpha d\Omega\,.$$

В общем случае коэффициент смертности может зависеть и от численностей особей. Ограничимся рассмотрением только тех популяций, у которых среднее время между рождениями значительно превосходит физически бесконечно малый интервал времени dt. Тогда рождением особей можно пренебречь.

Для биосистем важен следующий динамический закон:

Закон эволюции биосистем: изменение с течением времени численности биологических особей вида α с любыми допустимыми индивидуальными характеристиками обусловлено естественным отбором.

Его математическая формулировка имеет вид

$$\frac{d}{dt}\int_{\Omega} N_{\alpha} d\Omega = -\int_{\Omega} D_{\alpha} N_{\alpha} d\Omega . \tag{4.29}$$

Если из индивидуальных признаков нас интересует только возрастной состав изучаемых видов, то мы можем воспользоваться уравнениями

$$\frac{\partial \overline{N_{\alpha}}}{\partial t} + \frac{\partial \overline{N_{\alpha}}}{\partial \tau} + \overline{D_{\alpha}}\,\overline{N_{\alpha}} = 0 , \tag{4.30}$$

$$\overline{N_{\alpha}} = \int_{\omega} N_{\alpha} dx, \quad \overline{D}_{\alpha}\,\overline{N}_{\alpha} = \int_{\omega} D_{\alpha} N_{\alpha} dx .$$

Это – система уравнений математической экологии, описывающих возрастной состав популяции.

4.4. Модели динамики численности локальной популяции

Построение математических моделей динамики численности популяции является необходимым при разработке стратегии оптимального управления популяциями промысловых видов. Основная цель управления заключается в том, чтобы довести численность популяции до оптимального уровня, определяемого емкостью среды обитания, из года в год сохранять ее размер на этом уровне и за счет наиболее рационального соотношения особей разного пола и возраста добиться наивысшего ежегодного прироста численности промышленной части популяции (Фрисман, 1996).

Построение моделей, как правило, связано с максимальной схематизацией природных процессов и большим количеством ограничений. Процесс моделирования предполагает описание простейших ситуаций и анализ очень небольшого числа факторов, существенно влияющих на изучаемое явление.

Общепризнано, что основоположником в области моделирования популяционной динамики является Томас Мальтус.

4.4.1. Модель Мальтуса

Рассмотрим популяцию организмов, обитающую в условиях неограниченных ресурсов питания. Предположим, что популяция не подавляется никаким другим видом. В силу размножения и смертности число живых особей в популяции будет меняться с течением времени.

Пусть $x(t)$ – число живых организмов в момент времени $t+\Delta t$. Тогда разность

$$x(t+\Delta t) - x(t) = \Delta x$$

даст приращение функции $x(t)$ за промежуток времени от t до $t+\Delta t$.

Это приращение складывается следующим образом: за время Δt все взрослые особи или часть их произведут потомство; часть особей может погибнуть. Таким образом,

$$\Delta x = B - D, \tag{4.31}$$

где B – число родившихся от t до $t+\Delta t$, D – число погибших за это же время.

Величина B, разумеется, зависит от длины промежутка Δt: чем больше Δt, тем больше и B. Кроме этого, величина B зависит от количества «родителей». Чем больше взрослых особей, тем больше потомство. Таким образом,

$$B = \Phi(x, \Delta t), \tag{4.32}$$

где функция Φ растет с ростом x или Δt и равна нулю, если равна нулю одна из этих переменных.

Что касается переменной Δt, то самые простые эксперименты показывают, что она должна входить линейно: если промежуток наблюдения Δt увеличить, например, в два раза, то и прирост потомства микроорганизмов увеличится в два раза. Таким образом,

$$\Phi(x, \Delta t) = f(x)\,\Delta t. \tag{4.33}$$

Вопрос о характере функции $f(x)$ сложнее. Пока мы знаем только то, что она монотонно растет с ростом x и равна нулю при $x=0$. Каков этот рост? Он существенно зависит от биологических особенностей исследуемого вида. Мы ограничимся простейшим случаем, когда численность потомства пропорциональна количеству «родителей»:

$$f(x) = ax,\ a = const.$$

Этот случай реализуется, например, при делении клеток.

Итак,

$$B(x, \Delta t) = \alpha x \Delta t. \tag{4.34}$$

Величина α называется коэффициентом рождаемости или мгновенной рождаемостью.

Аналогичные соображения применимы и к D. Таким образом, положим,

$$D(x, \Delta t) = \beta x\,\Delta t, \tag{4.35}$$

где β – коэффициент смертности или мгновенная смертность.

Следовательно,

$$\Delta x=\alpha x\,\Delta t-\beta x\,\Delta t. \qquad (4.36)$$

Разделим обе части равенства (4.36) на Δt и перейдем к пределу при $\Delta t \to 0$. В результате получим

$$\frac{dx}{dt}=\alpha x-\beta x, \qquad (4.37)$$

или

$$\frac{dx}{dt}=\varepsilon x, \qquad (4.38)$$

где $\varepsilon=\alpha-\beta$.

Таким образом, мы получили обыкновенное дифференциальное уравнение первого порядка, которое называется *уравнением нормального размножения,* или *уравнением Мальтуса.* Величина ε в этом уравнении имеет важный биологический смысл. Она характеризует прирост популяции в единицу времени и называется *мальтузианским параметром* популяции.

Это уравнение нетрудно проинтегрировать. Можно рассудить, например, таким образом. Пусть $x(t)$ – некоторое решение уравнения (4.38). Подставив это решение в (4.38), мы получим тождество $x'(t)\equiv\varepsilon x(t)$, или $\frac{1}{x(t)}x'(t)\equiv\varepsilon$, или $\frac{d[\ln x(t)]}{dt}\equiv\varepsilon$, или, наконец, $\frac{d}{dt}[\ln x(t)-\varepsilon t]\equiv 0$.

Итак, если $x(t)$ – решение уравнения (4.38), то производная функции $\ln x(t)-\varepsilon t$ равна тождественно нулю. Это означает, как известно, что сама функция равна постоянной $\ln x(t)-\varepsilon t=\alpha$.

Отсюда после потенцирования $x(t)=e^{\varepsilon t+\alpha}$ или, обозначив $e^{\alpha}=C$, получаем:

$$x(t)=Ce^{\varepsilon t}. \qquad (4.39)$$

Заметим на будущее, что, действуя формально, решение (4.39) можно получить, если «разделить переменные» в уравнении (4.38), т.е. записать его в виде $\frac{dx}{x}=\varepsilon dt$, а затем каждую часть этого уравнения проинтегрировать по своей переменной.

Итак, любое решение уравнения (4.38) должно быть равно $Ce^{\varepsilon T}$ с некоторым C. Зная это, легко найти решение уравнения (4.38), удовлетворяющее условию $x(t_0)=x_0$. Подставив в (4.39) начальные данные $x(t_0)=x_0$, найдем C, соответствующее этим данным:

$$x_0=Ce^{\varepsilon t_0},\ т.е.\ C_0=x_0\cdot e^{-\varepsilon t_0}.$$

Таким образом, искомое решение имеет вид

$$x(t) = x_0 e^{\varepsilon(t-t_0)}. \tag{4.40}$$

Здесь t_0 – момент времени, с которого мы начали наше наблюдение за популяцией, а $x(t_0)$ – количество живых организмов в популяции в этот момент. Таким образом, функция (4.40) дает искомый закон изменения $x(t)$ с течением времени (рис. 4.22).

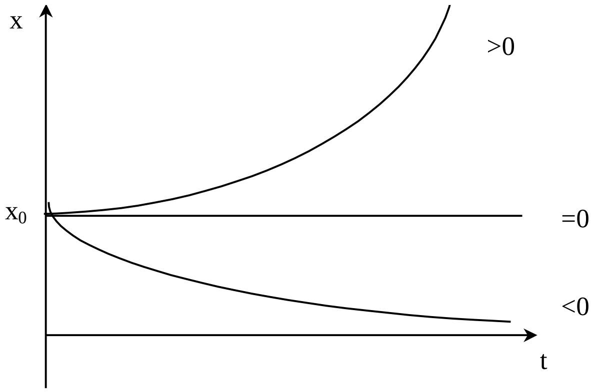


Рис. 4.22. Интегральные кривые уравнения нормального роста (4.38) при различных значениях мальтузианского параметра ε

Полученное решение (4.40) уравнения (4.38) имеет ряд важных свойств. Во-первых, непосредственно из вида решения следует, что при $\varepsilon>0$, т.е. в случае, когда коэффициент рождаемости (α) больше коэффициента смертности (β), популяция из любого начального состояния, большего нуля, неограниченно возрастает, стремясь к бесконечности, а при $\varepsilon<0$, т.е. в случае, когда коэффициент рождаемости меньше коэффициента смертности, популяция из любого начального состояния убывает до нуля. При $\varepsilon=0$ численность популяции не изменяется (рис. 4.21). Во-вторых, как бесконечно большое значение численности при $\varepsilon>0$, так и ноль при $\varepsilon<0$ могут быть достигнуты при ограниченном $x_0>0$ только за бесконечно большое время. Наконец, в-третьих, время, требуемое популяции для удвоения при $\varepsilon>0$ или «уполовинивания» при $\varepsilon<0$, не зависит от текущего состояния популяции. Действительно, обозначим τ время удвоения популяции при $\varepsilon>0$, тогда на основании (4.40) имеем:

$$x(t+\tau) = x_0 e^{\varepsilon(\tau-t_0)} = 2x(t) = 2x_0 e^{\varepsilon(t-t_0)} = x_0 e^{\varepsilon(t-t_0)+\ln 2}$$

или

$$\varepsilon(t+\tau-t_0)=\varepsilon(t-t_0)+\ln 2, \text{ откуда } \varepsilon\tau=\ln 2, \text{ или } \tau = \frac{\ln 2}{\varepsilon}.$$

Найденный нами закон изменения численности популяции, описываемый уравнениями (4.38) и (4.40), носит пока предположительный гипотетический характер. Его вывод основывается на нескольких предположениях, одни из которых более близки к реальности, другие – менее. Таковы предположения о неограниченных ресурсах питания, об отсутствии питания других видов, о структуре функции $\Phi(x,\Delta t)$ и т.д. Совершенно очевидно, что в природе в чистом виде не существуют популяции с такими свойствами, мы сознательно пошли на упрощения и рассмотрели не реальный процесс, а его упрощенную копию, его модель. Вопрос о том, насколько эта модель соответствует реальности, решает экспериментальная проверка.

Из формулы (4.40) следует, что при $\varepsilon>0$ численность поголовья растет с ростом τ неограниченно как экспонента. Разумеется, ни в одной реально существующей популяции такой рост не наблюдается. Это и понятно. Те предположения, на основе которых мы вывели уравнение (4.38) – изолированность популяции, неограниченность ресурсов питания и т.п. – в реальных природных условиях не выполняются. Таким образом, уравнение (4.38) имеет смысл либо в теоретическом аспекте (оно показывает, как развивалась бы популяция, если бы ей не мешали и неограниченно подкармливали), либо описывает динамику искусственно созданной и поддерживаемой популяции (например, популяции грибков, выделяющих пенициллин) на начальных этапах ее роста.

Уравнение (4.38) впервые получил и исследовал Томас Мальтус в 1302 г. Заблуждение Мальтуса заключалось в том, что это уравнение, справедливое для очень узкого класса популяций, он считал универсальным законом не только для всей природы, но и для человеческого общества. Вместе с тем не следует забывать, что работа Мальтуса послужила толчком для многих фундаментальных исследований и, в том числе, легла в основу логических построений Чарльза Дарвина при создании теории естественного отбора.

4.4.2. Модель популяционного взрыва

При выводе уравнения Мальтуса предполагалось, что коэффициенты рождаемости (α) и смертности (β) постоянны и не зависят от уровня численности. Однако для популяций организмов, размножающихся половым путем (без самооплодотворения), величина рождаемости (B) скорее всего пропорциональна не численности популяции, а числу встреч разнополых особей. При постоянном соотношении полов в популяции число встреч разнополых особей пропорционально квадрату численности и поэтому вместо (4.34) можно записать: $B(x,\Delta \mathrm{t}) = \alpha x^2 \Delta t$. В свою очередь смертность (D) во многих случаях также зависит от числа встреч особей, особенно если она определяется конкуренцией за какой-либо ресурс жизнедеятельности, поэтому будем считать, что $D = вx^2\Delta t$, и тогда получим для Δx

$$\Delta x = B - D = \alpha x^2 \Delta t - \hat{a} x^2 \Delta t = cx^2 \Delta t.$$

Отсюда получаем уравнение динамики

$$\frac{dx}{dt} = cx^2. \tag{4.41}$$

В этом случае при малых x прирост численности идет гораздо медленнее нормального, а при больших – гораздо быстрее. Уравнение (4.41) при малых x описывает динамику популяции малочисленного вида. Так, китам некоторых видов в настоящее время весьма трудно найти себе пару для размножения, поэтому динамика их популяций описывается уравнением (4.41), причем x мало. Другим примером динамики популяции, описываемой уравнением, близким к (4.41) при больших x, является изменение численности народонаселения (т.е. числа людей на земле). Ферстер с сотрудниками определяли по данным о численности населения земного шара методом наименьших квадратов значения коэффициентов уравнения динамики $dx/dt=cx^{\rho}$ и получили следующую оценку для ρ: $\rho=2{,}01$.

Покажем, что при $c>O$ и при $x_0>O$ интегральные кривые уравнения взрыва монотонно неограниченно возрастают с увеличением t, но ведут себя принципиально иначе, чем интегральные кривые уравнения Мальтуса.

Разделяя переменные в (4.41), находим решения уравнения взрыва:

$$\int \frac{dx}{x^2} = \int c\,dt \text{ , откуда } -\frac{1}{x} = ct + C \text{ .}$$

Подставляя в последнее уравнение начальные условия $x(t_0)=x_0$, находим $C=-ct_0-1/x_0$,. Учитывая это, получаем

$$x = \frac{1}{c(\frac{1}{cx_0} + t_0 - t)} \text{, или } x = \frac{1}{c(\tau(t_0 x_0) - t)} \text{ при } t < \tau \text{ ,}$$

где $\tau(t_0,,x_0) = t_0+1/(cx_0)$. Из этих формул следует, что интегральные кривые уравнения взрыва представляют собой половины гипербол, которые имеют вертикальную асимптоту $t = \tau$ (рис. 4.23).

Это означает, что в случае, когда прирост населения пропорционален числу пар, численность популяции становится бесконечно большой за конечное время. Физически такой вывод соответствует взрывообразному характеру процесса. (Для уравнения Мальтуса бесконечный рост популяции происходит за бесконечное время).

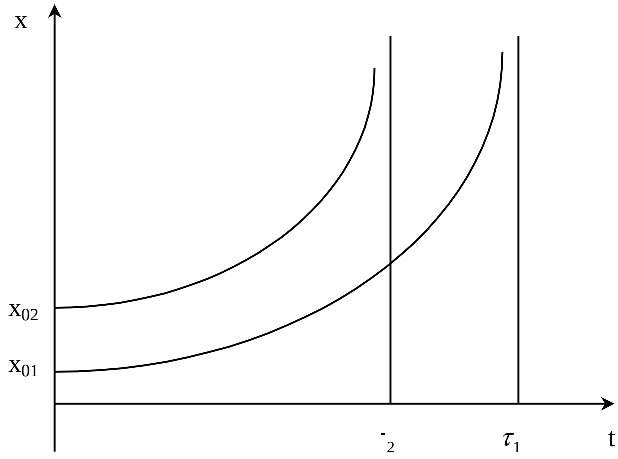


Рис. 4.23. Интегральные кривые уравнения популяционного взрыва (4.41) при различных начальных условиях

Для численности народонаселения земного шара Ферстер с сотрудниками получили следующую оценку для τ: τ=2026 год н.э. Разумеется, при *t* близких к τ, идеализация, принятая при описании процесса дифференциальным уравнением взрыва, неприменима, так что реальное количество населения за конечное время бесконечных значений не достигнет. Более того, скорость прироста реальных популяций не может ни оставаться постоянной при росте численности, как в модели Мальтуса, ни расти неограниченно, как в модели популяционного взрыва. Поэтому были разработаны модели, учитывающие замедление популяционного роста при больших численностях. Рассмотрению таких моделей посвящены следующие параграфы.

4.4.3. Модель Ферхюльста

В своих рассуждениях Мальтус предполагал, что коэффициенты рождаемости *(α)* и смертности *(β)* постоянны и не зависят от уровня численности. Это привело его к утверждению о неограниченности роста численности, которое, очевидно, маловероятно для природных популяций. Следущий шаг был сделан в 1945 г. П.Ф. Ферхюльстом. Отказавшись от требований неограниченности ресурсов, Ферхюльст учел влияние излишней тесноты, излишней плотности организмов. Это влияние в модели Ферхюльста заключается в том, что коэффициенты рождаемости и смертности не были постоянными величинами, а зависели от численности, причем рождаемость *α(x)* убывала, а смертность *β(x)* росла с ростом *x*. При конкретизации вида функций *α(x)* и *β(x)* Ферхюльст фактически ограничился предположением об их линейности (рис. 4.24), т.е. полагал, что

$$\alpha(\text{x})=\alpha_1 - \beta_1\text{x} \text{ и } \beta(\text{x})=\alpha_2+ \beta_2\text{x} \qquad (4.42)$$

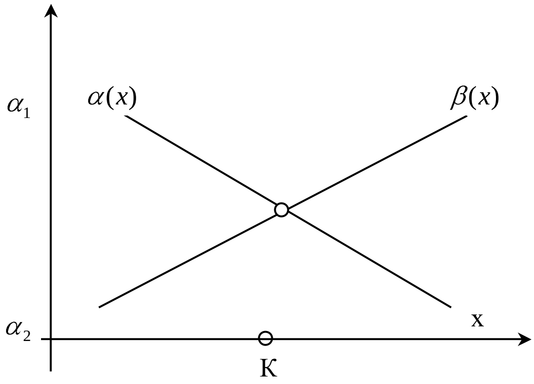


Рис. 4.24. Графики функций *a(x)* и *β(x)* в модели Ферхюльста

Подставив значения для *α(x)* и *β(x)* в уравнение (4.36), получаем

$$\Delta x=(\alpha_1 - \beta_1 x)x\Delta t - (\alpha_2+\beta_2 x)x\Delta t = (\alpha_1 - \alpha_2)x\Delta t - (\beta_1+\beta_2)x^2\Delta t) \qquad (4.43)$$

или

$$\Delta x=rx\Delta t - sx^2\Delta t, \qquad (4.44)$$

где $r= \alpha_1 - \alpha_2$, $s= \beta_1+\beta_2$.

Это равенство позволяет обосновать определенную корректность предположения о линейности функций *α(x)* и *β(x).* Действительно, второе слагаемое (точнее, вычитаемое) в (4.44) отражает снижение скорости роста популяции из-за внутривидовой конкуренции. Но конкуренция тем выше, чем больше количество встреч между особями, а количество встреч, как уже указывалось, пропорционально произведению $x{\cdot}x$, т.е. x^2. Разделив равенство (4.44) на Δt и переходя к пределу при $\Delta t \to 0$, получим

$$\frac{dx}{dt} = rx - sx^2, \qquad r > 0, \quad \delta > 0. \qquad (4.45)$$

Это и есть уравнение Ферхюльста. Оно отличается от модели Мальтуса тем, что величина мальтузианского параметра ε в этом случае не является постоянной, а есть линейная функция численности: $\varepsilon=r-sx$. Коэффициент r в этом уравнении является показателем специфической скорости роста популяции. Он фактически равен максимальной потенциальной скорости роста, которую достигла бы популяция в отсутствии лимитирующих факторов, т.е. при неограниченном запасе ресурсов жизнедеятельности. Коэффициент *s* мы вправе назвать коэффициентом самолимитирования, или коэффициентом внутривидовой конкуренции.

Уравнение Ферхюльста часто записывают в ином виде. Вынесем за скобки *rx*. Тогда

$$\frac{dx}{dt} = rx(1 - \frac{sx}{r})$$

или

$$\frac{dx}{dt} = rx(\frac{\frac{r}{s} - x}{\frac{r}{s}}).$$

Обозначив $r/s=K$, получим окончательно

$$\frac{dx}{dt} = rx(\frac{K - x}{K}). \tag{4.46}$$

Решение этого уравнения с начальным условием $x(t_o)=x_0$ нетрудно найти. Но многие выводы о свойствах решения можно получить и без интегрирования. Проведем его качественное исследование. Найдем стационарные решения уравнения (4.46). Для этого надо приравнять к нулю его правую часть:

$$rx(\frac{K - x}{K}) = 0.$$

Решая это уравнение, найдем два стационарных решения: $x=0$ и $x=K$. Построим далее график функции

$$f(x) = rx(\frac{K - x}{K}) \tag{4.47}$$

и фазовый портрет уравнения (4.46) рис. 4.25, слева.

Качественное исследование позволяет заключить, что любое решение (любая интегральная кривая) с начальным условием $0<x_0<K$ с ростом t монотонно возрастает и стремится к величине K, но не превосходит эту величину.

Можно пойти еще дальше и выяснить дополнительные подробности об этом возрастании. Продифференцировав это уравнение (точнее говоря, мы дифференцируем тождество, которое получится, если в уравнение подставить его решение $x(t)$), будем иметь

$$\frac{d^2x}{dt^2} = r\frac{dx}{dt} - 2r\frac{x}{K}\frac{dx}{dt} = r(1 - 2\frac{x}{K})\frac{dx}{dt}\,. \tag{4.48}$$

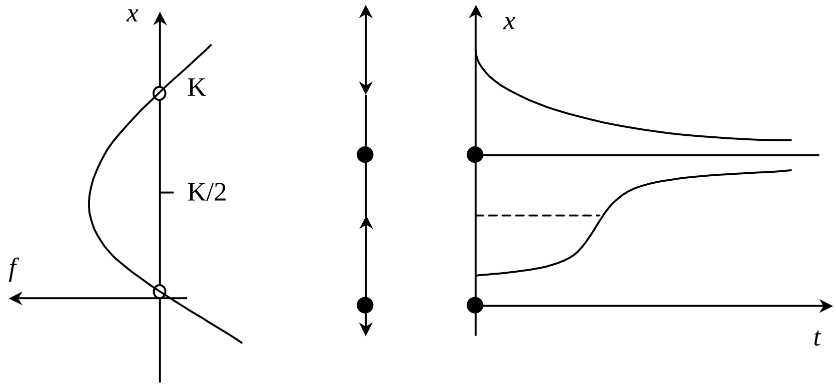


Рис. 4.25. График функции (4.47), фазовый портрет
и интегральная кривая уравнения (4.43)

Так как $x'(t)>0$, то знак правой части определяется знаком разности $1–2x/K$. Отсюда следует, что пока $x<K/2$, вторая производная d^2x/dt^2 положительна и, следовательно, график функции $x(t)$ вогнутый. Если же $x>K/2$, то график функции выпуклый. Действуя подобными методами, можно показать, что при $t\to\infty$ график функции $x(t)$ прижимается снизу к прямой $x=K$, т.е. $x(t)$ асимптотически стремится к пределу K, никогда не превосходя этой величины.

Совершенно аналогично исследуется случай, когда $x_o= x(t_o)>K$. В этом случае решение $x(t)$ при $t\to+\infty$ стремится к K, монотонно уменьшаясь.

Вместе с тем уравнение (4.46) не так уж трудно и проинтегрировать. Разделив переменные, получим

$$\frac{Kdx}{x(K-x)} = rdt \text{ или } \left(\frac{1}{x}+\frac{1}{K-x}\right)dx = rdt.$$

Считая $x<K$, после интегрирования будем иметь

$$\mathrm{Ln}x-\ln(K-x) = rt - \ln C$$

или

$$\frac{x}{K-x} = Ce^{rt}. \tag{4.49}$$

Пусть для простоты $t_0=0$ и $x(0)=x_0<K$. Подставив это значение C в (4.49), получим

$$\frac{x}{K-x} = \frac{x_0}{K-x_0}e^{rt}.$$

Отсюда

$$x(t) = \frac{x_0 K e^{rt}}{K - x_0 + x_0 e^{rt}}, \tag{4.50}$$

или

$$x(t) = \frac{x_0 K}{x_0 + (K - x_0)e^{-rt}}.$$

Анализируя эту функцию, можно также прийти к тем выводам о ее свойствах, которые мы получали, исходя из дифференциального уравнения. Найдем абсциссу точки перегиба интегральной кривой. Подставим в (4.48) выражение x из (4.50), получим

$$\frac{d^2x}{dt^2} = r\left(\frac{K - x_0 - x_0 e^{rt}}{K - x_0 + x_0 e^{rt}}\right)\frac{dx}{dt}.$$

Отсюда видно, что при $K\text{-}x_o\text{-}x_o e^{rt} > 0$ производная $x''(t)>0$, и, следовательно, функция $x(t)$ вогнута; при $K\text{-}x_o\text{-}x_o e^{rt} < 0$ производная $x''(t)<0$, и, следовательно, функция $x(t)$ выпукла. Абсцисса точки перегиба удовлетворяет уравнению

$$K\text{-}x_0\text{-}x_o e^{rt} = 0,$$

откуда

$$t = \tau = \frac{1}{r}\ln\frac{h - x_0}{x_0}.$$

Точка (τ, $K/2$) является точкой перегиба. Так как производная $x'(t)$ для всех t больше нуля, то это значит, что наша кривая нигде не имеет экстремумов. Наконец, из формулы самого решения (4.50) видно, что $x(t)\to K$ снизу при $t\to+\infty$, а из начальных данных следует, что при $t = 0$ $x(0) = x_0$. Всего этого достаточно, чтобы представить себе вид кривой (рис. 4.24, справа). Из чертежа видно, что если в начальный момент популяция была небольшая $(x<K/2)$, то развитие популяции идет по выпуклой кривой до точки (τ, $K/2$). В этой точке кривая перегибается и асимптотически (т.е. при $t\to+\infty$) стремится к прямой $x(t)=K$, никогда не достигая ее. Поэтому величину K называют максимальной численностью популяции (теоретически), возможной в данных условиях. Величина K является мерой емкости экологической ниши популяции.

График $x(t)$ напоминает вытянутую букву S. Его называют логистической кривой роста популяции (иногда S-образной кривой). Для многих естественных популяций эта кривая хорошо совпадает с экспериментальными данными. Иначе говоря, построенная модель достаточно точно отражает особенности роста популяции в условиях ограниченности ресурсов жизнедеятельности. Пользуясь функцией (4.50), мы можем не только прогнозировать численность популяции в любой момент времени, но и предсказывать максимальную численность, теоретически возможную в данных условиях.

В дальнейшем различными авторами предложено много модификаций основного логистического уравнения

$$dx/dt = x\varepsilon(x),$$

заключающихся в уточнении вида зависимости коэффициента прироста от численности, т.е. вида функции $\varepsilon(x)$. Оказалось, что для многих конкретных популяций в силу каких-либо физических или онтогенетических соображений естественно считать функцию $\varepsilon(x)$ отличной от линейной. В большинстве случаев, однако, эта функция так же, как и в модели Ферхюльста, положительна при малых значениях x, отрицательна при больших x, монотонно убывает с ростом численности для всех значений x и пересекает ось Ox только в одной точке $x=K$. Ясно, что все эти модели качественно эквивалентны уравнению (4.46), т.е. имеют такой же фазовый портрет, как и приведенный на рис. 4.23. Поэтому все они предсказывают, что из любого начального состояния популяция монотонно стремится к некоторому фиксированному уровню численности. При этом получаются решения уравнений динамики, которые весьма напоминают S-образную логистическую кривую, но значительно лучше соответствуют экспериментальным данным для каждого конкретного случая. Рассмотрим несколько примеров таких моделей, применяемых при описании динамики популяций промысловых видов рыб и беспозвоночных.

4.4.4. Модели Пелла-Томлинсона и Фокса

Модель Пелла-Томлинсона. Для этой модели можно представить мальтузианский параметр в виде мальтузианской функции

$$\varepsilon = r(1 - (x/K)^{\rho})$$

и получить следующее уравнение динамики:

$$\frac{dx}{dt} = rx\left(1 - \left(\frac{x}{K}\right)^{\rho}\right). \tag{4.51}$$

Очевидно, что при $\rho=1$ это уравнение совпадает с моделью Ферхюльста. При $\rho<1$ график функции $\varepsilon/r = 1-(x/K)^{\rho}$ – вогнутая кривая, а при $\rho>1$ – выпуклая (рис. 4.26, слева).

Это означает, что в первом случае $(\rho<1)$ плотностная регуляция уровня численности популяции существенно проявляется уже при низких значениях численности, в то время как во втором случае $(\rho>1)$ – при значениях численности, близких к емкости экологической ниши K.

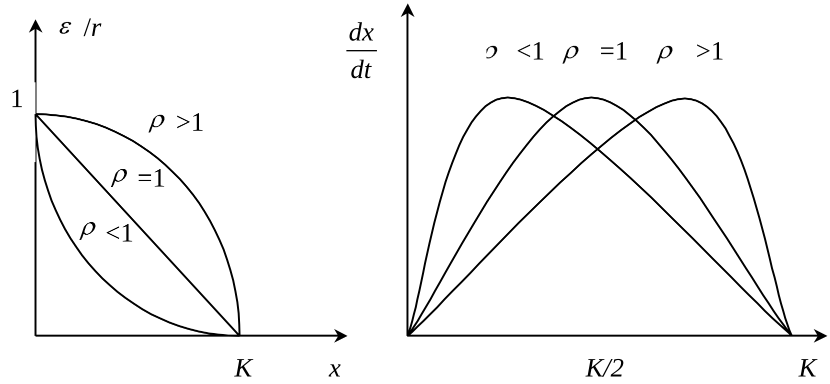


Рис. 4.26. Графики функций ε/r = $1-(x/K)^{\rho}$ и $dx/dt = rx(1-(x/K)^{\rho})$ при различных значениях коэффициента ρ

Функция $f(x) = rx(1-(x/K)^{\rho})$ достигает максимума в точке $x_м = K(1+\rho)^{-1/\rho}$, которая лежит левее $K/2$ ($x_м<K/2$) при $\rho<1$ и правее $K/2$ ($x_м>K/2$) при $\rho>1$ (рис. 4.26, справа). Следовательно, при $\rho<1$ точка перегиба «квазилогистической» интегральной кривой уравнения (4.46) лежит ниже точки перегиба логистической кривой, изображенной на рис. 4.23, а при $\rho>1$ – выше этой точки.

Модель Фокса. Эта модель является по существу аппроксимацией предельного случая предыдущей модели при $\rho\to 0$. Мальтузианская функция модели Фокса имеет вид

$$\varepsilon(x) = r(1-(\ln x/\ln K),$$

т.е. является линейной функцией от $\ln x$. Ей соответствует следующее уравнение динамики:

$$\frac{dx}{dt} = rx\left(1-\frac{\ln x}{\ln K}\right). \qquad (4.53)$$

Решение этого уравнения, удовлетворяющее начальным условиям $x(0) = x_0$, имеет вид

$$x(t) = K\left(\frac{x_0}{K}\right)^{e^{-rt}}, \qquad (4.54)$$

а его график называется кривой Гомперца. Функция

$$f(x) = rx(1-(\ln x/\ln K))$$

достигает максимума в точке $x_м = K/e$. Следовательно, точка перегиба кривой Гомперца и соответственно максимум скорости роста популяции, динамика которой описывается моделью Фокса, приходятся именно на это значение численности.

4.4.5. Принцип Олли. Модель Базыкина

До сих пор мы рассматривали модели динамики численности, в которых мальтузианская функция $\varepsilon(x)$ монотонно убывала с ростом численности при любых ее значениях. Однако для популяций некоторых видов высших организмов $\varepsilon(x)$ убывает с ростом x лишь при больших значениях численности. При малых же значениях численности $\varepsilon(x)$ увеличивается при возрастании x. Такой тип зависимости вызван тем, что в популяциях организмов с ярко выраженным групповым поведением и стремлением к агрегации сказывается эффект группы, заключающийся в существенно большей плодовитости особей в агрегированной группе при некоторых средних значениях численности, нежели в условиях недозаселенности. Описанный групповой эффект носит название принципа Олли.

Построим модель динамики численности популяции особей, обладающих таким групповым поведением. Будем, как и прежде, считать, что коэффициент смертности $\beta(x)$ является монотонно возрастающей функцией; это подтверждается оценками смертности для подавляющего большинства популяций. Возрастание смертности с ростом x объясняется ростом конкуренции за ограниченный ресурс (пищу, пространство и т.п.). Ограничимся линейным представлением функции $\beta(x)$:

$$\beta(x) = \alpha_2 + \beta_2 x, \tag{4.55}$$

которое было обосновано при выводе модели Ферхюльста. Что же касается коэффициента рождаемости $\alpha(x)$, то для многих видов животных, способных мигрировать, достаточно свободно и просторно заселяющих свой ареал, т.е. характеризующихся эффектом Олли, предположение о его линейности оказывается не совсем верно. Дело в том, что при малых плотностях численности размножение определяется, скорее, вероятностью встречи брачных партнеров, чем физиологической плодовитостью. Величина рождаемости в популяции таких видов оказывается очень низкой при малых значениях численности и растет с ростом численности. Однако при больших значениях численности достигается физиологический предел рождаемости и рост $\alpha(x)$ прекращается. Моделируя этот эффект, А.Д. Базыкин в 1969 г. предложил следующий вид зависимости $\alpha(x)$:

$$\alpha(x) = \frac{cx}{M + x}. \tag{4.56}$$

При малых численностях популяции ($x \ll M$) величиной x в знаменателе можно пренебречь и функция $\alpha(x)$ оказывается пропорциональна x, как в модели популяционного взрыва. При больших же численностях *(x>>M)* в знаменателе можно пренебречь уже величиной *M* и *α(x)* оказывается постоянной (равной *c* – физиологическому пределу коэффициента рождаемости), как в модели нормального роста. Величине *M,* используемой в выражении (4.56), можно дать ясную биологическую интерпретацию: *M* – это такое значение численности, при котором коэффициент рождаемости в популяции оказывается вдвое меньше возможного физиологического максимума этой величины. Действительно, при $x = M$ имеем: $\alpha(x) = \alpha(M) = c/2$.

Для того чтобы учесть падение рождаемости, вызванное перенаселением и конкуренцией при слишком больших значениях численности, можно вычесть из правой части выражения (4.56) величину $\beta_1 x$ аналогично тому, как это сделано при выводе модели Ферхюльста:

$$\alpha(x) = \frac{cx}{M+x} - \beta_1 x\,. \tag{4.57}$$

Подставляя в уравнение (4.37) вместо α и β выражения для $\alpha(x)$ и $\beta(x)$ из (4.57) и (4.55), получаем, наконец, модель, предложенную Базыкиным для описания динамики численности популяции животных, для которых справедлив принцип Олли:

$$\frac{dx}{dt} = \frac{cx^2}{M+x} - ax - bx^2, \tag{4.58}$$

где $a=\alpha_2$, $b=\beta_1+\beta_2$.

Графики функций $\alpha(x)$ и $\beta(x)$ имеют две точки пересечения при $x=K$ и $x=k$ (рис. 4.27, слева). При этих значениях численности рождаемость оказывается равна смертности.

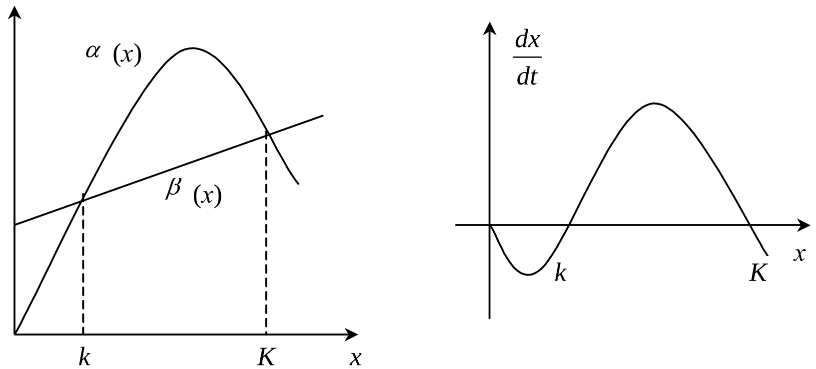


Рис. 4.27. Графики функций $\alpha(x)$, $\beta(x)$ и $\dfrac{dx}{dt} = \dfrac{cx^2}{M+x} - \alpha x - bx^2$

Таким образом, уравнение (4.58) имеет три стационарные точки x=0, x=K, x=k (рис. 4.27, справа) и может быть записано в виде

$$\frac{dx}{dt} = \frac{\mu x(K - x)(x - k)}{M + x}. \qquad (4.59)$$

На основе графика этой функции легко получить фазовый портрет и представить примерное поведение интегральных кривых (рис. 4.28).

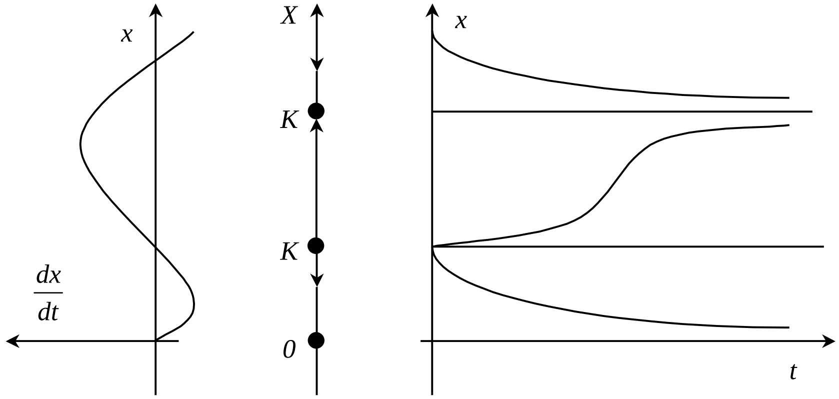


Рис. 4.28. График функции $\mu x(K - x)\ (x - k)/(M+x)$, фазовый портрет и интегральная кривая уравнения (4.59)

Итак, в случае, когда наблюдается эффект *Олли,* в популяции возможны два ненулевых равновесных значения численности K и k $(K>k)$, причем K – устойчивое равновесие, а k – неустойчивое. Это означает, что в популяции существует критический порог уровня численности, равный k. Если начальное значение численности x_o окажется больше k, то с течением времени численность популяции приближается к равновесному значению K. Если же x_0 меньше k, то численность популяции монотонно убывает до нуля и популяция вымирает. Таким образом, любое (в том числе антропогенное) снижение численности такой популяции ниже критического уровня чревато ее вырождением. По-видимому, в состоянии, близком к критическому, находятся сейчас популяции ряда видов китов, популяции уссурийского тигра, дальневосточного леопарда и ряд других.

Отметим, наконец, одно весьма существенное свойство всех рассмотренных моделей динамики численности одновидовых изолированных популяций, процессы рождения и гибели в которых идут непрерывно во времени, а сами модели являются обыкновенными автономными дифференциальными уравнениями. Во всех этих моделях отсутствуют периодические ("колебательные") режимы динамики. Действительно, все множество начальных значений численности разбивается на области притяжения к устойчивым стационарным точкам. Возможны только два типа динамического поведения: практически неосуществимый случай неустойчивого равновесия, если начальная численность оказывается как раз на границе зон притяжения, и реальный случай, заключающийся в монотонном переходе к соответствующему устойчивому равновесию, если начальная численность лежит внутри какой-либо из таких зон. Конкретный характер перехода к равновесному состоянию (вид зависимости $x(t)$) для каждой из моделей имеет свои особенности. Тем не менее, все одновидовые модели с непрерывным временем, в которых рассматривается динамика только общей численности популяции и не учитывается фактор запаздывания, не дают ничего, кроме монотонного перехода к равновесному состоянию. Поэтому к 30-м годам прошлого столетия твердо установилась точка зрения о том, что плотностно зависимые факторы регулирования численности непременно должны приводитъ к устойчивости численности популяции. Однако результаты выполненных за последние годы исследований одновидовых моделей динамики численности популяций с неперекрывающимися поколениями требуют, по-видимому, пересмотра этой точки зрения.

В заключение этого раздела отметим, что динамические модели с непрерывным временем до сих пор весьма успешно применяются для описания изменения численности взаимодействующих популяций разных видов. Начиная с работ А. Лотки и В. Вольтерры, в которых была показана возможность эндогенных колебаний численности в системе двух популяций, взаимодействующих по принципу хищник – жертва или паразит – хозяин, это направление развилось в стройную эколого-математическую школу, в рамках которой анализируются наиболее фундаментальные проблемы современной теоретической экологии.

4.5. Математическое моделирование в системе человек – окружающая среда

Современная эпоха выдвинула проблему взаимоотношения человеческого общества с природой. Воздействие человека на биосферу возросло многократно и продолжает быстро возрастать. Анализ глобальной экологической обстановки приводит к огромной проблеме – изучению всего комплекса взаимосвязанных динамических процессов, протекающих в биосфере. Это обстоятельство привело к созданию нового направления исследований, получившего название глобального моделирования. Методологической базой комплексного изучения наиболее важных сторон развития человеческого общества является системный анализ.

4.5.1. Модели Форрестера и Медоуза

С конца 60-х годов 20-го столетия в странах Запада возник ряд исследовательских центров и организаций, чья теоретическая и практическая деятельность направлена на решение проблемы развития цивилизации в условиях ограниченности земных ресурсов. Важную роль среди них играет Римский клуб – неправительственная и некоммерческая организация.

В 70-м году на очередной годичной сессии Римского клуба одному из ведущих специалистов в области теории управления – Дж. Форрестеру, в течение ряда лет занимавшемуся исследованием сложных динамических систем, было предложено разработать глобальную модель развития мира. Им была разработана компьютерная модель «Мир – 2», называемая теперь моделью Форрестера (Дулов, Цибаров, 2001). Это была первая попытка формализовать глобальное описание экологических процессов. Предложенный вариант модели содержал лишь два экологических параметра: численность населения и загрязнение окружающей среды. Он позволял оценить взаимное влияние этих параметров, с одной стороны, и темпов экономического развития – с другой. Впервые была продемонстрирована принципиальная возможность объединить производственные, социальные и экологические процессы одним формализмом.

Дальнейшие исследования были поручены группе под руководством Д. Медоуза. Модель Форрестера была несколько видоизменена вследствие увеличения числа уравнений и задаваемых функций. Ее принято называть моделью Медоуза. Иногда обе модели объединяют в одну, называя ее моделью Форрестера-Медоуза.

Рассмотрим модель Форрестера, т.к. она достаточно полно отражает сущность глобальных моделей Мира.

Пусть p – численность населения, U – капиталовложения в промышленность и сельское хозяйство всей Земли, S – доля всего капитала, вложенного в сельское хозяйство, R – невозобновляемые природные ресурсы Земли, Z – загрязнение (общее количество на Земле). Единицей времени считается год, единицей капитала – капитал, приходящийся на душу населения в 1970 г. Величина S – безразмерная. Единицей ресурсов считается годовое потребление ресурсов, приходившееся на человека в 1970 г. Начальные данные для этих переменных относятся к 1900 г. Предполагается, что для введения интегральных характеристик могут быть записаны уравнения на уровне однородных систем.

Эту систему уравнений в матричной форме можно записать в виде одного уравнения:

$$\frac{d\hat{O}}{dt} = F^{+} - F^{-}. \tag{4.60}$$

Индексами плюс и минус отмечены величины воздействия, которые способствуют количественному росту или убыванию основных параметров соответственно. Матрицы – столбцы *Ф*, F^+, F^- заданы выражениями

$$\hat{O} = \begin{pmatrix} p \\ U \\ S \\ Z \\ R \end{pmatrix}; \qquad F^+ = \begin{pmatrix} p^+ \\ U^+ \\ S^+ \\ Z^+ \\ 0 \end{pmatrix}; \qquad F^- = \begin{pmatrix} p^- \\ U^- \\ S^- \\ Z^- \\ R^- \end{pmatrix}.$$

Поскольку в правой части последнего уравнения стоят только отрицательные величины, можно сформулировать оптимизационную задачу о наиболее рациональном расходовании невозобновляемых ресурсов за время *T*, т.е. о минимизации R^-.

Предполагается, что в правых частях системы (4.60) могут быть учтены зависимости между переменными, известные из каких-либо априорных соображений, а также экспериментальные взаимосвязи. Эти зависимости могут быть нелинейными. В частности, соответствующие функции могут быть заданы таблично. Рассмотренная модель Форрестера состоит из пяти дифференциальных уравнений и двадцати таблично заданных зависимостей.

Основные выводы из прогнозов Форрестера сводятся к следующему. При сохранении современных тенденций развития общества неизбежен серьезный кризис во взаимодействиях человека и среды уже в следующем столетии. Этот кризис автор модели объясняет противоречием между ограниченностью природных ресурсов и ростом капитала, вкладываемого в промышленное и сельскохозяйственное производство. Именно этот рост вместе с ростом населения и темпов его потребления приводит к быстрому загрязнению среды, истощению ресурсов, росту смертности, а в конечном итоге к упадку производства. Несмотря на то, что было рассмотрено несколько сценариев прогноза, отличающихся исходными данными, результат был одинаково пессимистичным: угроза катастрофы не может быть отодвинута за пределы 2100 г.

Аналогичные выводы были сделаны и группой Медоуза. По мнению авторов модели, для предотвращения мировой катастрофы необходимо немедленно принять меры по созданию экономической и экологической стабильности, следствием которой должно стать состояние глобального равновесия.

Вообще, ни разработки Форрестера, ни исследования Медоуза не являются прогнозами в традиционном понимании, это направление научных исследований пытается лишь определить альтернативы будущего развития.

4.5.2. Глобальная модель биосферы

В работах российских ученых был предложен иной подход для анализа глобальных процессов. Основной принцип их метода заключается в том, что человек и вся его деятельность – это составная часть общих процессов в биосфере. Чтобы представить долговременные тенденции в экологии, нужно изучить протекание биосферных процессов и уже в рамках этих закономерностей рассматривать возможные сценарии развития человеческой цивилизации.

Описание модели. Глобальная модель биосферы, разработанная российскими учеными, представляет собой попытку совместного описания изменения некоторых характеристик биосферы при различных вариантах социального и экономического развития, а главное – нахождение тех путей, выявление тех критериев и принципов, которые могли бы обеспечить возможность совместного стабильного развития человеческого рода и биосферы при достаточно высоком уровне развития общества.

Модель биосферы состоит из взаимодействующих блоков, разрабатываемых относительно автономно. Такой модульный принцип позволяет наращивать степень детализации каждого блока без необходимости менять структуру других блоков.

Длительность отражаемых в модели процессов составляет десятки, максимум сотни лет. В качестве временного шага модели выбран один год.

При описании биосферы выделяются три пространственных блока: атмосфера, океан и регионы суши. Если разделение суши на регионы связано с природными, экономическими и политическими границами, то атмосфера и океан в принятом описании считаются едиными. Последнее предположение оправдано еще и тем, что процессы перемешивания в атмосфере и океане протекают гораздо быстрее, чем на суше.

Состояние каждого блока модели определяется набором некоторых переменных, которые в совокупности и составляют вектор основных фазовых переменных модели. Основными фазовыми переменными блока «Атмосфера» являются солнечная радиация, пыль, пары воды, температура; блока «Океан» – фитопланктон, нектон, биогенные элементы, загрязнение; блока «Регионы суши» – население, загрязнение, лес, травяная растительность, сельскохозяйственная растительность, животные, гумус, минеральные ресурсы, энергетические ресурсы.

В целом модель описывается задачей Коши для системы обыкновенных дифференциальных нелинейных уравнений, отражающих все связи между рассматриваемыми компонентами регионов биосферы. В модели имеется 400 коэффициентов, требующих количественного определения, и около 200 связей, нуждающихся в математическом описании.

Основные блоки модели

Блок «Атмосфера». Основу биогеоценотического процесса составляет энергообмен, описываемый уравнением сохранения энергии. В модели учитываются только два источника энергии: излучение Солнца (главный источник) и сгорание топлива биогенного происхождения. При этом в уравнении присутствуют члены, описывающие энергию излучения Солнца, поступающую в атмосферу; энергию, отражаемую в космос; долю солнечной энергии, идущей на нагревание атмосферы; энергию Солнца, вступающую в контакт с фотосинтезирующими элементами биосферы; площадь ледников; запыленность атмосферы и структуру растительного покрова. В блоке учитывается круговорот кислорода, углекислого газа и азота. Особое место в ней занимает модель климата. От климата зависит сама возможность жизни на Земле и существование цивилизации. Однако в настоящее время удовлетворительных способов прогнозирования и оценки основных параметров климата нет. Для проведения методических расчетов применяются простейшие зависимости, позволяющие оценить лишь один параметр – среднюю температуру атмосферы как функцию количества углекислого газа, а также значений энергии солнечной радиации и запыленности атмосферы.

Блок «Океан» представляет модель идеального перемешивания, содержащую следующие переменные: количество фито-, зоопланктона, нектона и биогенных питательных элементов; уровень загрязненности океана; содержание углекислого газа в верхнем слое перемешивания и в глубинах океана.

Блок «Регионы суши». Растительный покров суши отличается большим разнообразием видов, широким диапазоном показателей продуктивности и интенсивности газообмена. Для учета многообразия форм растительности вводится понятие региона суши, что позволяет приближенно учесть зональный характер растительного покрова Земли.

Продукция фотосинтеза потребляется населяющими Землю животными и человеком. В модели она рассматривается как единая фазовая переменная биосферы, характеризующаяся некоторыми средними скоростями роста и отмирания. Биопродукция животных также потребляется человеком.

Большую роль в глобальных циклах вещества и энергии играют почвообразовательные процессы, которые сопровождаются биохимической деятельностью организмов. В рамках модели все многочисленные этапы образования почвенного гумуса, который является последним звеном в цепи биохимического преобразования органических веществ, объединены для каждого региона в один компонент – гумус.

Модуль «Население». Он наиболее важен и представляет наибольшую трудность при моделировании глобального влияния людей на биосферу. Демографические характеристики не только определяют структуру резерва рабочей силы, но и влияют на все стороны экономики и биосферу в целом. В основе демографического описания лежит «закон сохранения» вида:

$$\frac{dl}{dt} = (\alpha - \beta)l ,$$

где l – численность населения; α – коэффициент рождаемости; β – коэффициент смертности. Подобное соотношение можно записать в дифференциальной или разностной форме с учетом миграции, возраста и т.п. Но все эти модификации и усложнения не изменяют существенного смысла этого уравнения, которое является одним из законов сохранения. Наиболее сложные вопросы связаны с описанием процессов рождаемости и смертности. В модели учитываются различные показатели смертности для зрелых и детских возрастов, которые отличаются для разных регионов. Характер зависимости коэффициента рождаемости очень сильно меняется от региона к региону.

В данной главе была рассмотрена модель глобальных процессов в биосфере, построенная на основе использования системного подхода. Задача подобных моделей – давать оценки изменения тенденций развития в результате тех или иных решений и выявлять возможность опасных экологических ситуаций. Для повышения адекватности модели необходимо ввести в нее дополнительные блоки, отражающие экономические и политические факторы и последствия научно-технической революции, а также точнее учитывать взаимосвязи с уже имеющимися блоками.

Контрольные вопросы

1. От чего зависит способ, выбираемый для анализа экосистем?
2. Основные этапы аналитического моделирования.
3. Статическое поведение системы. Какое состояние системы возникает под действием постоянных возмущающих переменных?
4. Свойства имитационных моделей экосистем.
5. Возможно ли использование результатов имитационной модели без оценки ее адекватности?
6. Пределы изменения коэффициентов корреляции и детерминации.
7. Всегда ли регрессионная модель позволяет прогнозировать значения признака в зоне экстраполяции?
8. В чем смысл нулевой гипотезы об отсутствии линейной регрессии?
9. Основная идея дисперсионного анализа.
10. Чем определяется эффективность дискриминации? Способы отнесения нового объекта к одной из нескольких заранее известных групп.
11. Правила иерархического объединения кластеров.
12. Модель Лотки-Вольтерра. При каких условиях численности видов «хищника» и «жертвы» остаются в среднем постоянными величинами?
13. Какой биологический смысл имеет мальтузианский параметр популяции?
14. В чем отличие уравнения Ферхюльста от модели Мальтуса?
15. Принцип Олли. При каких значениях численности достигается физиологический предел рождаемости?

Глава 5
ПРОГНОЗ ДИНАМИКИ ПОКАЗАТЕЛЕЙ СИСТЕМ

5.1. Методы прогнозирования

Прогноз – это результат экстраполяции прошлого в будущее. Прогнозы строятся на основе некоторых объективных правил, которые определяют совокупность вычислений и действий, необходимых для получения прогноза. Необходимо различать понятия прогноз и предсказание. Предсказание – это субъективная оценка будущего, в то время как прогноз – несмещенная оценка будущих значений. Если субъективные предсказания достаточно убедительно указывают на то, что несмещенные оценки будущего вряд ли возможны, прогнозы необходимо модифицировать, т.е. подправлять (Льюис, 1986).

Применяемые методы прогнозирования зависят от путей исследования. Если мы пытаемся найти факторы, определяющие причинно-следственный механизм, т.е. факторы, определяющие поведение прогнозируемого показателя, прогноз по которым либо известен, либо найти нетрудно – это первый путь. Этот путь приводит к математическому моделированию, построению модели поведения эколого-экономического объекта. Второй путь – не вдаваясь в механизм движения, попытаться предсказать будущее положение, анализируя временный ряд показателя изолированно, рассматривая только динамику его во времени или других сопряженных характеристиках.

Методы прогнозирования существенно различаются в зависимости от того, является ли прогнозирование краткосрочным или среднесрочным. В первом случае прогноз строится на один-два момента времени вперед (квартал, месяц, неделю и т.п.) и, как правило, оперативен и непрерывен. В большинстве случаев краткосрочного прогнозирования данные берутся либо за месяц, либо за неделю: соответственно прогноз – на один-два месяца или неделю. При среднесрочном прогнозировании данные, как правило, ежегодные, а прогноз строится на 5–10 лет вперед.

Указанные различия между задачами краткосрочного и среднесрочного прогнозирования приводят к необходимости решать их разными методами. В первом случае это методы, основанные на идее экспоненциального сглаживания, предложенные Брауном (1962), а во втором – методы выравнивания и экстраполяции трендов. Указанные методы имеют ряд преимуществ, которые сделали их весьма популярными:

– сравнительная простота;
– экономичность вычислений;
– возможность автоматического построения прогнозов;
– наличие хорошего математического обеспечения.

5.1.1. Краткосрочное прогнозирование (методы экспоненциального сглаживания)

Методы краткосрочного прогнозирования применяются в случаях, когда:

1) частота данных за рассматриваемый период не более года (недельные, месячные, квартальные данные);

2) прогноз делается для конкретного объекта отдельно и последовательно на каждый следующий момент времени;

3) прогнозы строятся для большого числа объектов или ресурсов.

В качестве примера записи прогноза рассмотрим следующую схему: «прогноз спроса на биоресурс в следующем месяце равен спросу на него в текущем месяце». Это утверждение в символах математики выглядит так:

$$f_{t+1}=d_t, \qquad (5.1)$$

где f – прогноз, d – спрос, t – промежуток времени (например месяц).

Запись прогноза следующего месяца, равного среднеарифметическому шести предыдущих месяцев, запишется в следующем виде:

$$f_{t+1} = \frac{1}{6}\sum_{i=t}^{t-5} d_i \ , \qquad (5.2)$$

где $\sum_{i=t}^{t-5} d_i$ – сумма значений спроса от t-го месяца до t – 5 месяца.

5.1.1.1. Прогнозирование стационарных показателей

Стационарными называются такие показатели, индивидуальные значения которых меняются со временем и не изменяют своего среднего на достаточно продолжительном отрезке времени. На рис. 5.1 представлена картина стационарного ряда, на котором среднее значение биомассы в течение месяца приблизительно равно 100 единицам в течение года: отдельные показатели колеблются вверх и вниз, тогда как среднее значение биомассы достаточно устойчиво.

Поскольку прогнозы строятся на информации о поведении объекта в прошлом, они всегда будут иметь ошибку. Поэтому большинство прогностических схем и алгоритмов основываются на идее минимизации таких ошибок, причем как положительных, так и отрицательных (прогнозируемое значение может быть как больше, так и меньше реального значения показателя). Очевидно, что обычная сумма этих ошибок не может служить удовлетворительным критерием их малости, поскольку вне зависимости от применяемого метода прогнозирования эта сумма будет стремиться к нулю. Более адекватная мера качества прогноза – сумма квадратов ошибок, поскольку квадраты всегда неотрицательны независимо от того, была первоначально ошибка положительной или отрицательной. Таким образом, каждая ошибка входит в сумму квадратов. Поскольку любой прогноз несет на себе определенную степень ошибки, имеется лишь возможность прогнозирования показателя в среднем, т.е. возможны случайные отклонения от реального значения влево и вправо. Такие отклонения обычно предполагаются распределенными нормально, т.е. их распределение совпадает с распределением Гаусса. Эта гипотеза верна до тех пор, пока индивидуальные значения наблюдений не будут резко отличаться от своего среднего, таким образом, мы исключаем возможность выбросов как вниз, так и вверх относительно среднего. Допущение о нормальности ошибок прогноза требует меры разброса или рассеяния ошибок вокруг среднего. Обычной мерой разброса служит стандартное отклонение σ.

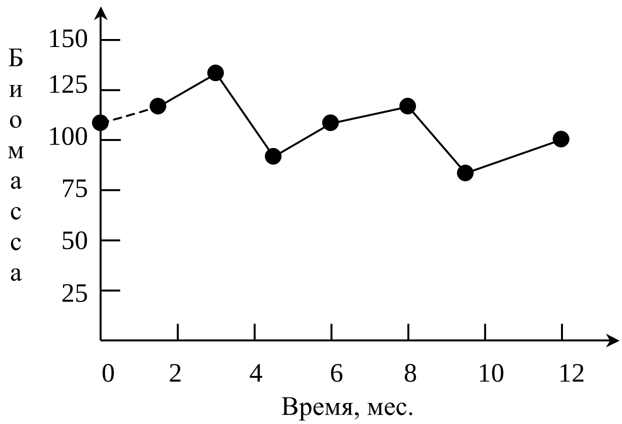


Рис. 5.1. Типичная картина стационарного ряда

Итак, каждый прогноз будет характеризоваться двумя основными показателями. Первый – значение прогнозируемого показателя на будущий момент времени, т.е. сам прогноз. Второй – стандартное отклонение прогноза, которое характеризует разброс прогнозируемого значения вокруг реального.

В краткосрочных системах прогнозирования итоговое значение показателя рассматривается как одно наблюдение, например биомасса за день, за неделю и т.д. Увеличивая длину наблюдаемого периода, мы увеличиваем объем выборки и уменьшаем влияние отдельных колебаний прогнозируемого показателя за период, а, следовательно, получаем возможность построения такого прогноза. В то же время с увеличением длины периода уменьшается скорость приспособления метода прогнозирования к фактическим колебаниям данных. Баланс между этими двумя эффектами достигается выбором подходящего интервала наблюдения за показателем или базой прогноза.

Для удовлетворительного краткосрочного прогнозирования необходимо, чтобы в течение двух единиц длины наблюдаемого периода имелось в наличии хотя бы одно наблюдение, т.е. чтобы с 50%-й вероятностью в каждой единице периода лежало хотя бы одно наблюдение. Это условие минимально. Число единиц времени, на которое делается прогноз, называется горизонтом прогнозирования.

Скользящее среднее. Традиционным методом прогнозирования будущего значения показателя является усреднение n его прошлых значений, т.е. получение так называемого скользящего среднего. Формально его можно определить как

$$m_t = \frac{1}{n}\sum_{i=t}^{t-n+1} d_i \tag{5.3}$$

или

$$m_t = m_{t-1} + \frac{1}{n}(d_t - d_{t-n})\,. \tag{5.4}$$

Из (5.2) следует, что текущее значение скользящего среднего отличается от предыдущего на величину $1/n$ разности текущего значения показателя и его значения на момент времени, сдвинутого на n единиц назад. Вычисленное значение m_t в случае стационарного ряда полагается равным прогнозу ожидаемого значения показателя в будущем не только на период прогноза, но и на период, следующий за ним, и далее. Последнее не означает, что если прогноз делается на шесть месяцев вперед, то прогноз на остальные пять месяцев не может быть модифицирован по истечении первого месяца. Подобную поправку осуществить не всегда просто, однако эта идея весьма близка к идеям текущего планирования, когда лишь на первые два месяца составляются конкретные планы, а на третий и четвертый лишь приблизительные. На следующие два месяца по истечении двух предыдущих планы могут быть скорректированы с учетом дополнительной информации.

Скользящее среднее имеет ряд особенностей.

1. Для того чтобы начать процесс скользящего среднего, необходимо иметь в запасе n-1 прошлых значений наблюдений. Прогноз не может быть построен раньше чем через n моментов времени.

2. Данным, включенным в процесс скользящего среднего, присваивается одинаковый вес, всем остальным данным присваивается нулевой вес. Вес отдельного наблюдения указывает на долю вклада его значения в значение среднего, и в случае скользящего среднего эта доля равна $1/n$ для наблюдений, входящих в среднее, и нулю для всех наблюдений, отсутствующих в нем. При этом более свежие данные имеют тот же вес,

что и более старые, вместе с тем понятно, что более свежие данные имеют более важное значение и поэтому должны иметь и больший вес. Для устранения этого недостатка существует процедура усреднения с разными весами. Уравнения (5.5) и (5.6) представляют два из подобных способов усреднения: первый основан на дробных, а второй на десятичных весах. В обоих случаях сумма весов равна единице, это условие необходимо для того, чтобы соответствующие величины были средними значениями:

$$m_t = \frac{1}{2}d_t + \frac{1}{4}d_{t-1} + \frac{3}{16}d_{t-2} + \frac{1}{16}d_{t-3} \tag{5.5}$$

или

$$m_t = 0{,}4d_t + 0{,}3d_{t-1} + 0{,}2d_{t-2} + 0{,}1d_{t-3}\,. \tag{5.6}$$

3. Если не учитываются более старые данные, то для скользящего среднего это может оказаться слишком расточительным. На практике, для того чтобы скользящее среднее было не столь чувствительным, используют усреднение по двадцатиточечному периоду.

4. Чувствительность скользящего среднего обратно пропорциональна n – числу точек, входящих в среднее, поэтому без изменения n чувствительность изменить невозможно.

Большинство из перечисленных особенностей скользящего среднего устраняется при использовании системы весов экспоненциально взвешенного среднего.

Экспоненциально взвешенное среднее. Вместо одной из рассмотренных выше систем весов рассмотрим целый ряд весов, убывающих во времени по экспоненциальному закону. Этот ряд определим следующим образом:

$$\alpha + \alpha(1-\alpha) + \alpha(1-\alpha)^2 + \alpha(1-\alpha)^3 + \ldots + \alpha(1-\alpha)^n, \tag{5.7}$$

где α – константа экспоненциального сглаживания.

Для истинного среднего сумма ряда должна стремиться к единице[1].

С помощью экспоненциально взвешенного ряда весов экспоненциально взвешенное среднее запишем как

$$u_t = \alpha d_t + \alpha(1-\alpha)d_{t-1} + \alpha(1-\alpha)^2 d_{t-2} + \alpha(1-\alpha)^3 d_{t-3} + \ldots\,. \tag{5.8}$$

---

[1] Имеется в виду, что сумма этого ряда стремится к единице при неограниченном увеличении числа слагаемых. Если вынести за скобку α, то в скобках имеем сумму бесконечной геометрической прогрессии со знаменателем $0<1-\alpha<1$. По формуле суммы бесконечно убывающей геометрической прогрессии $1+(1-\alpha)+(1-\alpha)^2+\ldots+=1/(1-(1-\alpha))=1/\alpha$, поэтому сумма весов равна $\alpha\cdot 1/\alpha=1$, а, значит, экспоненциально взвешенная величина удовлетворяет условию среднего.

Это равенство можно переписать в эквивалентной форме:

$$u_t = \alpha d_t + (1-\alpha)[\alpha d_{t-1} + \alpha(1-\alpha)\ d_{t-2} + \alpha(1-\alpha)^2 d_{t-3} + ...]. \qquad (5.9)$$

С помощью (5.8) выразим u_t через другие члены последовательности, получим

$$u_{t-1} = \alpha d_{t-1} + \alpha(1-\alpha)d_{t-2} + \alpha(1-\alpha)^2 d_{t-3} + \alpha(1-\alpha)^3 d_{t-4} + ... , \qquad (5.10)$$

т.е. сумма членов в квадратных скобках (5.9) есть не что иное, как u_{t-1}. Поэтому, подставляя u_{t-1} в уравнение (5.9), получаем окончательно

$$u_t = \alpha d_t + \alpha(1-\alpha)u_{t-1}. \qquad (5.11)$$

Это основное уравнение, определяющее простое экспоненциально взвешенное среднее. На его основе строятся другие модели экспоненциального сглаживания.

Экспоненциально взвешенное среднее имеет ряд преимуществ перед традиционным скользящим средним.

1. Для построения прогноза по экспоненциально взвешенному среднему необходимо задать лишь начальную точку прогноза; дальнейшее прогнозирование возможно по поступлении свежих данных, т.е. нет необходимости заново строить процедуру прогноза.

2. В экспоненциально взвешенном среднем значения весов убывают со временем. Поэтому здесь нет точки, на которой веса обрываются, т.е. зануляются.

3. Для вычисления экспоненциально взвешенного среднего u_t требуются всего два значения: прошлое значение среднего u_{t-1} и текущее значение d_t.

При условии равенства «среднего значения степени старения данных» (или чувствительности прогноза)

$$\frac{n-1}{2} = \frac{1-\alpha}{\alpha}. \qquad (5.12)$$

Типичные значения α, используемые в области прогнозирования эколого-экономических систем, лежат в пределах от 0,05 до 0,3. Это означает, что длина усреднения в скользящем среднем с точки зрения чувствительности прогноза может быть найдена в соответствии с n по табл. 5.1.

Таблица 5.1

Соответствие значений α и *n*

α	0,05	0,1	0,2	0,3
n	39	19	9	5,66 (≈6)

Значит, если выбранное значение α равно 0,1 (наиболее популярное), то для применения скользящего среднего необходимо запомнить 18 прошлых значений показателя, что очень непрактично.

4. Чувствительность экспоненциально взвешенного среднего в целях повышения адекватности прогностической системы может быть в любой момент времени изменена путем изменения величины α. Чем выше α, тем выше чувствительность среднего; чем ниже α, тем устойчивее становится экспоненциально взвешенное среднее. На практике не рекомендуется брать значения ниже 0,05, так же как и выше 0,3. И, наконец, если более подходящими оказываются более высокие значения α, то это указывает на нарушение условий стационарности, т.е. простое экспоненциально взвешенное среднее становится неприемлемым.

Вычисление прогноза по методу простого экспоненциального сглаживания. При наличии прошлого прогноза вычислить текущий прогноз можно легко по формуле (5.9) или (5.11). Но иногда нагляднее использовать специальные номограммы (рис. 5.2).

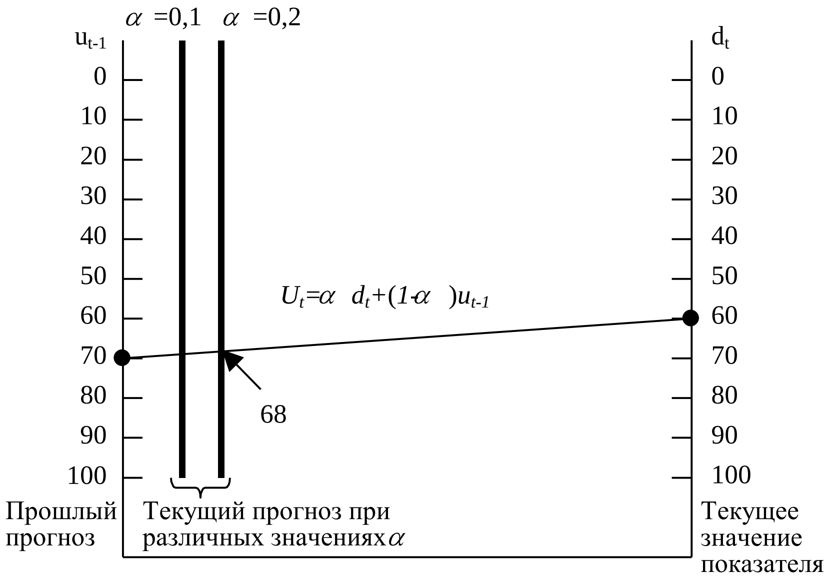


Рис. 5.2. Номограмма вычисления прогноза
по схеме простого экспоненциально взвешенного ряда

Номограмма состоит из трех идентичных линейных шкал. На левой откладывается значение прошлого прогноза (u_{t-1}), на правой текущее значение (d_t). Между двумя этими шкалами располагается шкала текущего прогноза (u_t) для различных значений константы α в методе экспоненциально взвешенного среднего. Ее положение определяется следующим образом. Например, если α=0,01, то она должна отстоять от левой шкалы на расстоянии 1/10 расстояния между левой и правой шкалами (расстояние между шкалой u_t и правой шкалой d_t равно 9/10 расстояния между шкалами u_{t-1} и d_t). Нетрудно найти эти расстояния и при других значениях α; некоторые из них показаны в табл. 5.2.

Таблица 5.2

Расположение шкалы u_t

α	Расстояние до u_{t-1}	Расстояние до d_t	α	Расстояние до u_{t-1}	Расстояние до d_t
0,1	$\frac{1}{10}$	$\frac{9}{10}$	0,33	$\frac{1}{3}$	$\frac{2}{3}$
0,15	$\frac{3}{20}$	$\frac{17}{10}$	0,5	$\frac{1}{2}$	$\frac{1}{2}$
0,2	$\frac{1}{5}$	$\frac{4}{5}$			

Текущий прогноз u_t с помощью номограммы строится следующим образом. Точки на шкале прошлого прогноза (u_{t-1}) и текущего значения прогнозируемого показателя (d_t) соединяются прямой линией. Искомой точкой будет точка пересечения этой прямой со шкалой прогноза (u_t) с выбранным значением α.

Например, если d_t=60, а u_{t-1}=70 и значение α=0,2, то u_t приблизительно будет равно 68 (рис. 5.2), что может быть найдено и по формуле (5.11):

$$u_t=(0,2\cdot 60)+(0,8\cdot 70)=12+56=68.$$

Вычисления, необходимые для построения прогноза по методу экспоненциального взвешенного среднего по формуле (5.11), достаточно просты. Однако возможна другая, более компактная запись уравнения (5.11). Поскольку разность $d_t - u_{t-1}$ представляет собой не что иное, как текущее значение ошибки прогноза e_t, то уравнение (5.11) можно записать в виде

$$u_t = u_{t-1} + \alpha(d_t - u_{t-1}) \qquad (5.13)$$

или

$$u_t = u_{t-1} + \alpha e_{t-1}. \qquad (5.14)$$

Последнее уравнение описывает поведение простейшего самонастраивающегося механизма с пропорциональным запаздыванием.

При прогнозировании эколого-экономических показателей часто необходимо получить целочисленные значения прогнозов, что достигается путем их округления.

5.1.1.2. Прогнозирование нестационарных показателей – линейный рост и сезонность

В данной главе будут рассмотрены методы прогнозирования, работающие в нестационарных условиях, т.е. когда среднее не остается постоянным, а изменяется со временем. Изменяющееся среднее называется трендом. Тренды различаются по характеру и типу.

Характер тренда

1. Линейный тренд

Линейным трендом называется такой закон изменения среднего, при котором среднее возрастает или убывает со временем по линейной зависимости. Линейный тренд может быть возрастающим или убывающим.

2. Сезонные тренды

Тренд называется сезонным, если среднее меняется циклически в соответствии с некоторым циклом. В большинстве случаев на практике этот тренд не меняется в течение года, причем среднее за каждый месяц по сравнению со средним за весь год может и падать и подниматься. Например, динамика первичного продуцирования имеет сезонный характер.

3. Смешанные сезонно-линейные тренды

Этот тип тренда представляет собой комбинацию из двух вышерассмотренных. Примером такого тренда является выпуск продукции аквахозяйств. Увеличение доли выпускаемой продукции в долгосрочном аспекте определяет линейный тренд, а сезонные колебания связаны с ограничением производства морепродуктов при неблагоприятных климатических условиях.

Типы трендов

1) Аддитивные тренды

В аддитивных трендах фактические значения отклоняются от среднего в положительную или отрицательную сторону приблизительно на одинаковую величину. Например, средний прирост величины спроса на продукцию аквахозяйства за месяц может составлять десять единиц измерения.

2) *Мультипликативные тренды или тренды отношений*

В мультипликативных трендах увеличение или уменьшение фактического значения составляет приблизительно одинаковый процент относительно среднего, определяемого характером тренда. Например, спрос на биопродукцию с увеличивающимся линейно-мультипликативным трендом будет увеличиваться на 2% за месяц.

3) *Комбинация линейных и мультипликативных трендов*

Этот тип тренда является соединением первых двух. Его изучение достаточно сложно, поэтому он употребляется довольно редко.

При описании тренда любого показателя необходимо задать характер тренда и его тип. Характер тренда определяет его среднее, а тип – отклонение от среднего. Рассмотрим наиболее распространенные варианты (виды) трендов и соответствующие им прогностические модели.

1А. Линейно-аддитивный тренд

Показатель с таким видом имеет среднее, которое увеличивается (или убывает) приблизительно на одинаковую величину с каждым моментом времени (рис. 5.3). Среднее, как и в случае линейно-мультипликативного тренда (рис. 5.4), является возрастающей функцией времени. Но в случае линейно-аддитивного тренда разброс отклонений фактических значений вокруг тренда приблизительно постоянен, тогда как в случае линейно-мультипликативного тренда этот разброс увеличивается со временем.

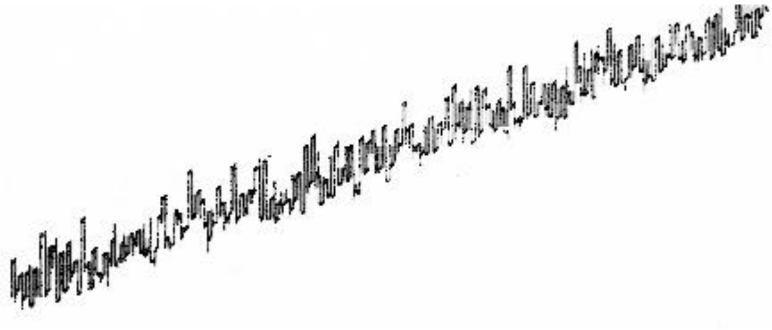


Рис. 5.3. Динамика спроса с линейно-аддитивным трендом (*1А*)

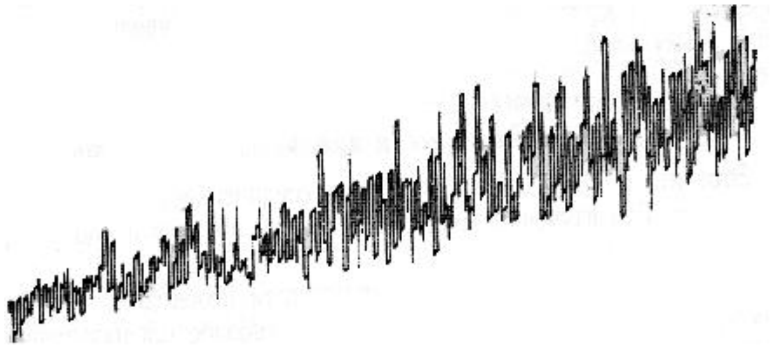


Рис. 5.4. Динамика спроса с линейно-мультипликативным трендом (*1В*)

1В. Линейно-мультипликативный тренд

Значение показателя при таком виде превосходит (или будет меньше) предыдущее значение приблизительно на один и тот же процент на всем рассматриваемом промежутке времени. На рис. 5.4 показана подобная ситуация. При этом со временем увеличивается не только среднее, но и разброс индивидуальных значений вокруг среднего (тренда).

3А. Комбинация линейного и сезонно-аддитивного тренда

Этот тип тренда также может описывать ситуацию чисто сезонного тренда без линейного элемента. Однако в общем случае для модели этого типа характерно присутствие сезонного тренда, который в свою очередь может линейно расти. Линейный и сезонно-аддитивный тренды изображены на рис. 5.5. Как видно из рисунка, из года в год повторяются два пика.

Рис. 5.5. Динамика спроса с линейным
и сезонно-аддитивным трендом (*3А*)

3В. Комбинация линейного и сезонно-мультипликативного тренда

Эта комбинация описывает еще и случай чисто сезонно-мультипликативного тренда без линейного роста (рис. 5.6). Как и для комбинации линейного и сезонно-аддитивного трендов, аналитическое исследование этого типа тренда предполагает возможность линейного роста.

На описанных трендах типов 1А, 1В, 3А, 3В основаны соответствующие прогностические модели. Так, на основе линейно-аддитивного тренда (1А) разработан ряд моделей, в которых полагается, что среднее прогнозируемого показателя d_t изменяется по линейной функции от времени:

$$d_t=\mu+\lambda t+\varepsilon_t, \tag{5.15}$$

где μ – среднее процесса, λ – скорость его роста, ε_t – случайная ошибка с нулевым средним.

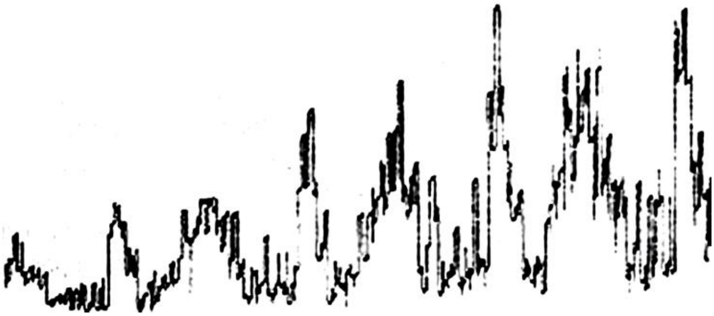


Рис. 5.6. Динамика спроса с сезонно-мультипликативным трендом (*3В*)

К линейно-аддитивным прогностическим моделям относятся модели, основанные на методах Холта, Холта с модификациями Муира, двойного сглаживания Брауна, адаптивного сглаживания Брауна. Предполагая, что среднее прогнозируемого показателя изменяется линейно, все перечисленные методы нацелены на вычисление скорости линейного роста (или падения) показателя во времени. Отличие моделей друг от друга и состоит в методах оценки показателя роста (падения). Так, в методе Холта (1957) фактор роста оценивается по разности между экспоненциально взвешенной средней процесса u_t и их предыдущими значениями u_{t-1}. Муир (1958) доказал, что показатель роста совпадает с оценкой коэффициента линейного тренда по методу наименьших квадратов. Метод Брауна (1962) основан на идее дисконтированной взвешенной регрессии. В конечном итоге все эти модели с большей или меньшей степенью точности приводят к вычислению прогноза по формуле

$$f_{t+\tau}=u_t+b_t\tau, \tag{5.16}$$

где $f_{t+\tau}$ – прогноз на τ моментов времени, b_t – показатель роста, τ – количество моментов времени прогнозирования, u_t – оценка среднего текущего значения. Подробно все эти модели описаны в монографии «Методы прогнозирования экономических показателей» (Льюис, 1986).

На основе линейно-мультипликативного тренда (1В) Муиром была разработана линейно-мультипликативная модель.

Иногда изменение среднего процесса (уравнение (5.15)) зависит от времени не линейно, а пропорционально самому значению среднего μ (т.е. линейно в логарифмах). Тогда более подходящей будет мультипликативная модель, описываемая уравнением

$$d_t = (d_{t-1} - \varepsilon_{t-1})\rho + \varepsilon_t, \tag{5.17}$$

где ρ – мультипликативный коэффициент тренда.

Прогноз на момент времени t+τ найдем из уравнения

$$f_{t+\tau} = v_t\, r_t^{\tau}, \tag{5.18}$$

где r – несмещенная оценка ρ (мультипликативный коэффициент тренда процесса d_t), v_t – сглаживающая функция. О способах вычисления сглаживающей функции и несмещенной оценки можно также узнать из работы Льюиса (1986).

Методы прогнозирования, основанные на мультипликативных моделях трендов, не получили широкого распространения, хотя Муир показал, что для некоторых типов данных такие модели дают лучшие по сравнению с моделями линейных трендов прогнозы. Заметим, что мультипликативные тренды сводятся к линейным заменой фактических наблюдений их логарифмами.

Для случая комбинации линейных и сезонно-аддитивных трендов Холтом и Винтером была разработана сезонно-декомпозиционная прогностическая модель, а Браун для таких трендов смог применить предложенный им для линейно-аддитивных трендов метод сглаживания.

В прогностической модели предполагается, что характеристики движения ряда показателя, а именно стационарность, линейность и сезонность, могут быть разделены, изучены и оценены отдельно одни от других. Окончательный прогноз будет осуществляться сведением прогнозов различных элементов в один.

При прогнозировании сезонного ряда необходимо определить, как изменение значения переменной в данный момент времени связано с изменением значения этой переменной, отстоящей на сезонный цикл. А поскольку каждый момент времени принадлежит одному циклу, задача заключается в установлении формы сезонной зависимости. Сезонные колебания численно описываются так называемыми коэффициентами сезонности. Они представляют собой отношение сезонного значения показателя на некоторый момент времени к среднему значению этого показателя, соответствующего моментам времени, лежащим внутри цикла. При прогнозировании сезонных рядов необходимо помнить последние L коэффициентов сезонности (L – длина сезонного цикла, т.е. число единиц времени, содержащихся в цикле). Иногда с целью облегчения расчетов показатели, сходные по сезонным характеристикам, объединяются в группы, для каждой из которых предусмотрен один общий коэффициент сезонности. Таким образом, уменьшаются затраты, связанные с хранением информации, и повышается устойчивость и репрезентативность набора коэффициентов сезонности.

Сумма коэффициентов сезонности за год близка к 12 (т.е. среднее коэффициентов равно 1). Это необходимое условие для несмещенности прогнозов. Многие методы декомпозиции предполагают в какой-либо форме наличие линейного тренда, вследствие чего при построении прогноза учитывают связанный с этим линейный рост. Сезонный анализ данных без выделения и оценивания линейного тренда привел бы к смещению коэффициентов сезонности, т.е. к заметному отличию суммы этих коэффициентов за год от 12.

Сезонно-декомпозиционная прогностическая модель Холта-Винтера основана на применении метода экспоненциально взвешенного среднего. Оценка стационарно-линейного и сезонного факторов для нее производится по следующей схеме. В первую очередь оценивается стационарный фактор (т.е. производится оценка среднеежемесячного значения независимо от времени года). Процедура оценки отличается от ранее рассмотренного метода Холта тем, что ряд текущих значений d_t очищен от сезонности делением его на величину коэффициента сезонности (F_{t-L}). Например, если текущее значение d_t в январе 1991 г. было равно 1600 единицам, а коэффициент сезонности равен 0,8, то значение, очищенное от сезонности (декомпозиционное), будет равно 1600/0,8=2000 единицам.

Кроме того, вычисляются оценка линейного роста (на основе модели роста Холта) и оценка сезонного фактора. Коэффициент сезонности (F_z) представляет собой отношение значения текущего наблюдения к среднестационарному значению, т.е. этот коэффициент в момент времени t равен d_t/u_t. При изолированной оценке трех факторов, определяющих движение процесса, прогноз на τ моментов времени вперед ($f_{t+\tau}$) строится из трех элементов: суммируются оценка линейного роста b_t и оценка стационарного фактора u_t, и результат с учетом сезонности домножается на соответствующее значение коэффициента сезонности $F_{t-L+\tau}$:

$$f_{t+\tau} = (u_t + b_t\tau)F_{t-L+\tau}\,. \tag{5.19}$$

Модель Холта-Винтера в практике прогнозирования сезонных временных рядов встречается чаще всего. Ее прогностическая точность не уступает точности других еще более сложных моделей поведения сезонно изменяющихся временных рядов (среднеабсолютная процентная ошибка по этой модели в большинстве случаев меньше 50%).

Обобщенный сглаживающий метод Брауна почти полностью совпадает с взвешенным методом наименьших квадратов. Разница заключается в более сложном выборе модели, на основе которой строится взвешенная регрессия. Харрисон (1964) показал, что в условиях применения адаптивного сглаживания Брауна к сезонным моделям значение коэффициента дисконтирования регрессии должно быть достаточно велико, чтобы придать значимый вес, по крайней мере, последним десяти точкам наблюдения, и в то же время должно достаточно малым, чтобы выполнялось условие локальной адекватности модели. Такая противоречивая рекомендация часто делает модель Брауна непригодной для использования, но не в случаях линейного тренда.

Модель, основанная на трендах типа 3В, предполагает генерирование процесса линейно-мультипликативным трендом. Вычисление показателя линейного роста и экспоненциально-взвешенного среднего производится при тех же аргументах, что и в случае линейно-аддитивной модели, но с учетом коэффициента сезонности (Льюис, 1986).

5.1.1.3. Меры точности прогноза

Стандартное отклонение. Разброс или рассеяние значений некоторой переменной вокруг среднего, как правило, измеряется стандартным отклонением. Стандартное отклонение вычисляется как квадратный корень из дисперсии, которая в свою очередь определяется как «среднее квадратов ошибок».

Для большинства прогнозов сумма ошибок стремится к нулю, т.е. положительные и отрицательные ошибки компенсируют друг друга. Вот почему сумма ошибок не может служить удовлетворительной мерой разброса. Существует и другой способ сделать ошибки неотрицательными независимо от того, были ли первоначально они положительными или отрицательными. Этот способ основан на использовании абсолютного значения ошибки ($e_t = |d_t - f_t|$), и мерой разброса служит здесь среднее абсолютное значение отклонения ошибки (MAD_t – Mean Absolute Deviation):

$$MAD_t = \alpha|e_t| + (1-\alpha)MAD_{t-1}. \qquad (5.20)$$

Среднее абсолютное отклонение связано со стандартным отклонением (σ_t). Для довольно большого класса статистических распределений значение σ_t несколько больше MAD_t и строго пропорционально ему. Константа пропорциональности для различных распределений колеблется между 1,2 и 1,3 (для нормального распределения это значение равно $\sqrt{\pi/2} = 1{,}2533$). В качестве компромисса можно взять 1,25, поэтому $\sigma_t = 1{,}25\ MAD_t$.

Среднеабсолютная процентная ошибка (*MAPE* – Mean Absolute Percentage Error) есть среднее абсолютных значений ошибок прогноза, выраженных в процентах относительно фактических значений показателя:

$$MAPE = \frac{1}{n}\sum_{t=0}^{n-1}\frac{|e_t|}{d_t}100. \qquad (5.21)$$

Показатель *MAPE*, как правило, используется при сравнении точности прогнозов разнородных объектов прогнозирования, поскольку этот показатель характеризует относительную точность прогноза. Типичные значения *MAPE* и их интерпретация показаны в табл. 5.3.

Таблица 5.3

Интерпретация типичных значений *MAPE*

MAPE, %	Интерпретация
<10	Высокая точность
10–20	Хорошая точность
20–50	Удовлетворительная точность
>50	Неудовлетворительная точность

При d_t=0 *MAPE* становится бесконечной, поэтому данные не могут принимать нулевые значения. Если d_t=0, целесообразно пропускать вычисления.

Средняя процентная ошибка (*MPE* – Mean Percentage Error) и *средняя ошибка* (*ME* – Mean Error) – показатели смещенности прогноза. При условии, что потери при прогнозировании, связанные с завышением будущего значения, уравновешиваются занижением, идеальный прогноз должен быть несмещенным, и обе меры должны стремиться к нулю. Наиболее популярный относительный показатель смещения *MPE* определяется как

$$MPE = \frac{1}{n}\sum_{t=0}^{n-1}\frac{e_t}{d_t}100\,. \tag{5.22}$$

Он не должен превышать 5% и не определяется для нулевых данных.

Средняя ошибка уже не является относительным показателем, а характеризует степень смещения прогноза и рассчитывается по формуле

$$ME = \frac{1}{n}\sum_{t=0}^{n-1} e_t\,. \tag{5.23}$$

Средний квадрат ошибки (*MSE*) определяется формулой

$$MSE = \frac{1}{n}\sum_{t=0}^{n-1} e_t^2\,, \tag{5.24}$$

а *сумма квадратов* (*SSE*) –

$$SSE = \sum_{t=0}^{n-1} e_t^2\,. \tag{5.25}$$

MSE и *SSE* чаще всего используются при выборе оптимальных моделей прогнозирования. В большинстве пакетов прикладных программ по прогнозированию именно эти два показателя принимаются в качестве критерия при оптимальном выборе параметров модели.

5.1.1.4. Адаптивное прогнозирование

Под адаптивным прогнозированием понимают методы прогнозирования, основанные на адаптации к данным другой информации, на базе которой строится прогноз. Основное свойство такого метода: при поступлении новых данных значение прогноза меняется, адаптируясь к вновь поступившей информации, и становится, таким образом, более чувствительным к ней. При небольшом изменении значений данных прогноз также будет мало изменяться. Более чувствительный прогноз приведет и к меньшей разнице между прогнозируемыми и фактическими значениями, а значит, и точность будет выше. Требование к малочувствительности прогноза в условиях устойчивости не так очевидно, поскольку в этом случае и высоко- и низкочувствительный прогноз приведет к одним и тем же значениям. Необходимость в низкочувствительном прогнозе возникает в случае, когда движение стационарно изменяющегося ряда нарушается в один из моментов времени скачком (рис. 5.7-а).

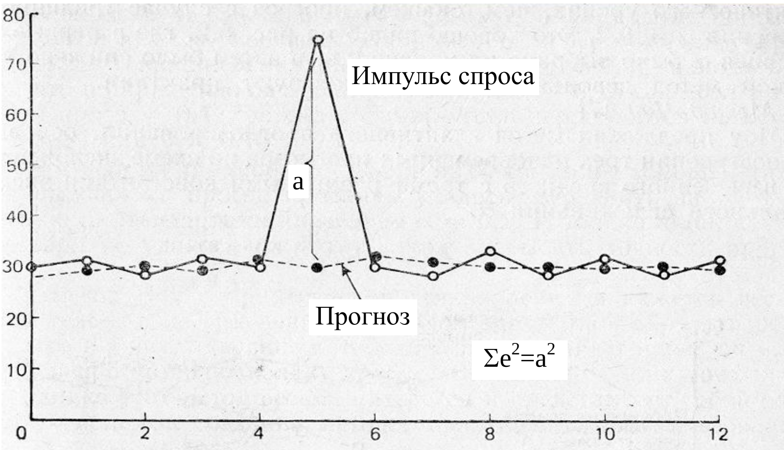


Рис. 5.7-а. Реакция низкочувствительного прогноза
на единовременный импульс в величине спроса

В подобной ситуации низкочувствительный прогноз мало изменит свое значение, и единственной ошибкой прогноза будет ошибка, связанная с моментом импульса. Рассмотрим теперь случай использования высокочувствительного прогноза и проанализируем его реакцию на резкое изменение значения прогнозируемого показателя (рис. 5.7-б). Поскольку все прогнозы так или иначе учитывают поступающую информацию, момент времени, следующий за моментом наступления скачка, также будет отражать его действие. Далее при низкочувствительном прогнозе движение показателя станет опять стабильным и выйдет на прежний, нормальный средний уровень, в то время как высокочувствительному прогнозу для этого потребуется более длительный промежуток времени. В каждый момент времени на этом промежутке значение ошибки прогноза будет достаточно велико, так как прогноз на нем будет превосходить фактическое значение показателя. И хотя эти ошибки будут противоположны по знаку ошибке в момент скачка в сумме квадратов $a^2+b^2+c^2+d^2+e^2$, они приведут к значению, намного большему, чем для низкочувствительного прогноза, для которого эта сумма будет равна лишь a^2.

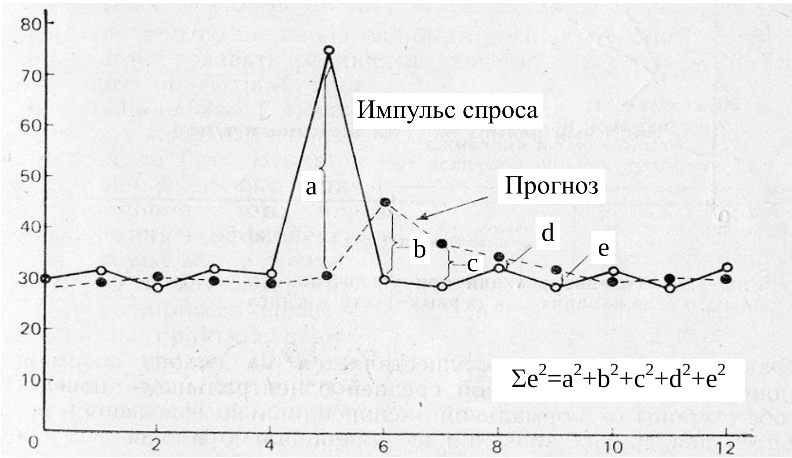


Рис. 5.7-б. Реакция высокочувствительного прогноза на единовременный импульс в величине спроса

Мы видим необходимость игнорирования прогностическими системами единовременных импульсов с тем, чтобы их действие не распространялось на следующие моменты времени.

Подобная картина резкого повышения спроса может встретиться и на практике, однако прогнозирование рядов, подверженных с достаточной частотой подобным иррегулярным колебаниям вследствие применения адаптивных методов, вряд ли целесообразно. С точки зрения адаптивных методов было бы предпочтительнее относительно плавное повышение спроса, эффект которого в будущем был бы смягчен и распространился после подъема на достаточно большой промежуток времени.

Впервые метод адаптивного прогнозирования был предложен Чоу (1965). Этот метод основан на построении трех одновременных прогнозов по схеме экспоненциально взвешенного среднего с тремя различными константами экспоненциального сглаживания α.

Фактический прогноз осуществляется на основе «нормальной» экспоненциально взвешенной средней с центральным значением α. По обе стороны от экспоненциально взвешенной средней строятся два других значения экспоненциально взвешенной средней со значением α на 0,05 меньше и больше центрального значения. Таким образом, если исходный прогноз строится при значении α=0,1, то низкочувствительный прогноз соответствует значению α=0,05, а высокочувствительный прогноз соответствует α=0,15.

Произвольным образом выбранные в начальном периоде константы сглаживания с течением времени автоматически меняются по следующему правилу. Если по критерию ошибок прогноза один из крайних прогнозов станет лучше нормального, в следующие моменты времени за «нормальный» прогноз берется прогноз, соответствующий лучшему крайнему прогнозу. Для нового прогноза строятся свои крайние прогнозы.

По своей концепции очень схож с методом Чоу метод Тамара. Он основан на обобщенной модели Холта-Винтера и включает в себя прогнозирование тренда и сезонных факторов. В соответствии со степенью ошибки прогностической системы Тамара предлагает задавать различные значения коэффициенту экспоненциального сглаживания.

Кроме вышеуказанных методов адаптивного прогнозирования существует и ряд других, например метод Тригга-Лича, метод Шоуна и др. Подробно с этими методами можно ознакомиться в уже упоминавшейся монографии Льюиса.

5.1.1.5. Анализ временных рядов с помощью автокорреляций

Большинство данных, имеющих экологическую природу, измеряется последовательно во времени (зависят от времени дискретно или непрерывно), т.е. исследователь имеет дело с так называемым временным рядом, который может содержать естественные случайные флуктуации или «шум», которые искажают исходные данные. Нужен метод или инструмент, с помощью которого можно было бы убрать эффект воздействия шума, после чего оценить характеристики ряда, необходимые для построения соответствующей прогностической модели.

Для этой цели используется несколько статистических методов, однако чаще всего в прогностических моделях при идентификации тренда применяется так называемый метод анализа автокорреляции.

Автокорреляция. С помощью корреляции измеряется степень тесноты связи между двумя переменными. Автокорреляция – это степень зависимости внутри самой переменной, т.е. «переменной на себя». Это достигается путем сопоставления фактического ряда данных с тем же рядом, но сдвинутым на некоторое число отрезков времени назад, т.е. наблюдению d_{t-i} ставится в соответствие наблюдение d_{t-i-K}, где K – лаг, выраженный отрезком времени. Математическое определение коэффициента автокорреляции r_K с лагом K выглядит так:

$$r_K = \frac{\sum_{t=0}^{n-1}(d_t - \bar{d}_t)(d_{t-K} - \bar{d}_{t-K})}{\sigma_{d_t}\sigma_{d_{t-K}}}, \quad (5.26)$$

где $$\sigma_{d_t} = \sqrt{\frac{\sum_{t=0}^{n-1}(d_t - \bar{d}_t)^2}{n-1}}, \; \sigma_{d_{t-K}} = \sqrt{\frac{\sum_{t=0}^{n-1}(d_{t-K} - \bar{d}_{t-K})^2}{n-1}}, \quad (5.27)$$

d_t – ряд фактических данных;

d_{t-K} – ряд сдвинутых данных;

$\bar{d}_t$ – среднее фактических значений;

$\bar{d}_{t-K}$ – среднее сдвинутых значений;

n – число рассматриваемых пар значений, уменьшающееся с увеличением лага K.

Коэффициент автокорреляции изменяется от -1 до 1. Значение коэффициента, близкое к единице, указывает на сильную положительную зависимость между фактическим рядом данных и рядом, сдвинутым на K единиц времени. В этом случае пары наблюдений будут близки друг к другу. Если окажется, что большее наблюдение составляет пару с меньшим, то, как правило, коэффициент автокорреляции будет отрицательным и близким к -1. Можно утверждать, что на практике коэффициент автокорреляции приближенно с 95%-й уверенностью будет значим, если значение r_K с лагом будет (по абсолютной величине) превышать $2/\sqrt{n}$, т.е. приблизительно 0,6, 0,4 0,3 и 0,3 соответственно для n, равных 12, 24, 36 и 48.

Анализ автокорреляций – очень популярный инструмент статистического анализа, который стал доступен и легок в применении с развитием пакетов статистических программ для ЭВМ. Например, в ППП «Statistica» в модуль «Time Series/ Forecasting- Анализ временных рядов и прогнозирование» встроена специальная процедура ARIMA для вычисления автокорреляций и авторегрессий.

Первые разности. Если ряд наблюдений содержит линейный аддитивный тренд, то избавиться от тренда можно посредством перехода к «первым разностям». Переход к первым разностям порождает новый ряд, представляющий собой разность соседних значений фактического ряда наблюдений (табл. 5.4). Очевидно, что среднее значение первых разностей есть не что иное, как скорость роста фактического ряда наблюдений.

Таблица 5.4

Поведение первых разностей
при выделении линейно-аддитивного тренда

Фактический ряд, d_t	6		10		15		21		24		30		35		41
Первые разности $d_t^* = d_t - d_{t-1}$		4		5		6		3		6		5		6	

Аппарат первых разностей и анализ автокорреляций может применяться для идентификации временных рядов любых типов.

На рис. 5.8-б показана автокоррелограмма (т.е. зависимость автокорреляции от временного лага) для стационарного ряда. Этот ряд (рис. 5.8-а) представляет собой значения спроса на продукцию аквахозяйства с января (J) по декабрь (D) 1976 и 1977 гг. (Льюис, 1986). Точками на рисунке обозначены две параллельные прямые, определяющие 95%-е доверительные границы значимости коэффициента автокорреляции (±две стандартные ошибки, т.е. $\pm 2/\sqrt{n}$ =0,2, где n=24). Поскольку ни один из коэффициентов автокорреляции не лежит за этими границами, а в изменении значений коэффициентов отсутствует определенная закономерность, можно заключить, что автокоррелограмма указывает на то, что в данном случае ряд показателей является стационарным.

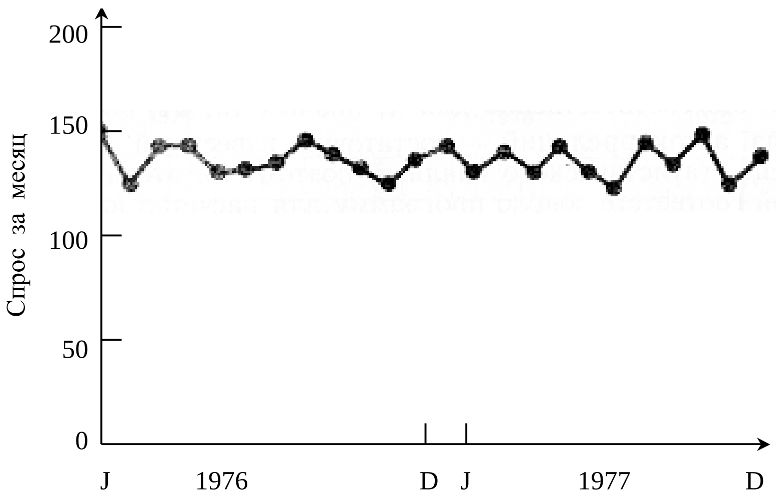


Рис. 5.8-а. Типичная картина стационарных рядов

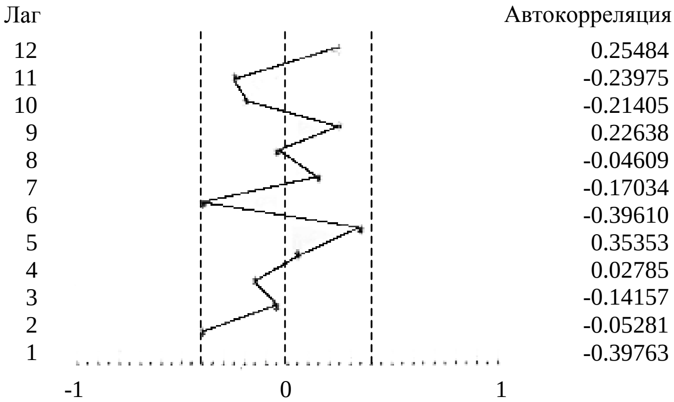


Рис. 5.8-б. Автокоррелограмма типичного стационарного ряда

На рис. 5.9-а показан слабо выраженный линейно-аддитивный тренд на протяжении двух лет. Рост в данном случае довольно мал, и исследование автокоррелограммы для этих данных (рис. 5.9-б) указывает на едва заметную зависимость значений коэффициентов автокорреляции от величины лага. Коэффициенты автокорреляции в среднем уменьшаются с увеличением лага, максимальное значение соответствует лагу, равному единице, и равно 0,43234; минимальное значение коэффициента автокорреляции соответствует сдвигу на 12 месяцев и равно -0,18765. Такая сильная линейная зависимость свидетельствует о наличии линейно-аддитивного тренда.

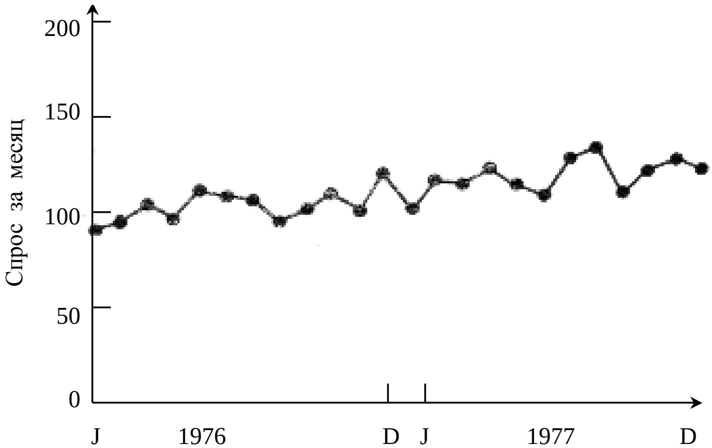
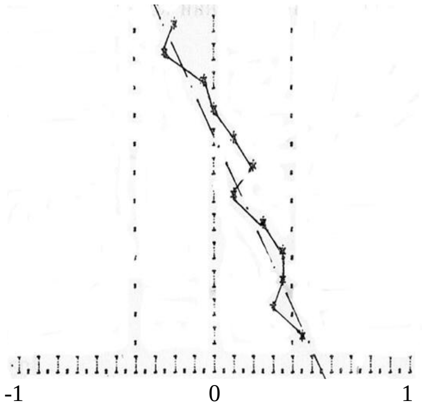


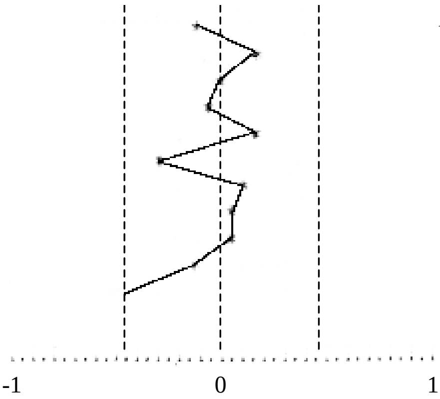
Рис. 5.9-а. Данные с линейно-аддитивным трендом



Лаг	Автокорреляция
12	-0.18765
11	-0.24736
10	-0.04134
9	-0.02209
8	0.08208
7	0.22355
6	0.10402
5	0.26142
4	0.33646
3	0.34499
2	0.30991
1	0.43234

Рис. 5.9-б. Автокоррелограмма в случае линейно-аддитивного тренда (нулевые разности)

Построим для исходного ряда данных с линейно-аддитивным трендом ряд первых разностей и соответствующую ему автокоррелограмму (рис. 5.9-в). Исходный ряд можно рассматривать как ряд с нулевыми разностями.



Лаг	Автокорреляция
11	-0.09563
10	0.13011
9	-0.01027
8	-0.03326
7	0.12878
6	-0.22717
5	0.08382
4	0.05458
3	0.05137
2	-0.09813
1	-0.47480

Рис. 5.9-в. Автокоррелограмма в случае линейно-аддитивного тренда (первые разности)

Автокоррелограмма первых разностей после того, как линейно-аддитивный тренд переходом к разностям был исключен, отчетливо показывает, что первые разности можно считать случайно разбросанными, а рис. 5.8-б и 5.9-в схожи своей хаотичностью. Итак, если автокорреляционный анализ указывает на то, что в значениях коэффициентов автокорреляций нулевых разностей заметна строгая линейная зависимость, а переход к первым разностям исключает ее, то исходный ряд содержит линейно-аддитивный тренд.

На рис. 5.10-а изображен ряд сезонно колеблющегося показателя с подъемами и спадами. Только для сезонных трендов можно ожидать сильной корреляционной зависимости между показателями, соответствующими данному месяцу, в том же месяце прошлого года (рис. 5.10-б). Таким образом, максимальные значения коэффициента автокорреляции, наблюдаемые при лаге 12 и 24 месяца, равны соответственно 0,54027 и 0,34421, причем оба эти коэффициента значимы (т.е. лежат за границами 95% доверительной полосы, равной в данном случае ±0,3). Это указывает на сильную зависимость между наблюдениями за один и тот же месяц, но разных лет. Наоборот, если лаг равен 6 или 18 месяцам, т.е. наблюдение, соответствующее подъему, сравнивается с наблюдением, соответствующим спаду, коэффициент автокорреляции должен быть отрицательным. Это полностью подтверждается автокоррелограммой, где минимальные значения коэффициентов автокорреляции соответствуют лагу в 6 и 18 месяцев и равны -0,67416 и -0,42020 соответственно. Показателем чисто сезонного ряда без линейного тренда служит автокоррелограмма с большим числом значимых максимальных и минимальных значений коэффициентов автокорреляции. Поскольку на рис. 5.10-б не обнаруживается линейная зависимость величины автокорреляции от величины лага, то исходный ряд не имеет линейного тренда, и переход к первым разностям здесь нецелесообразен.

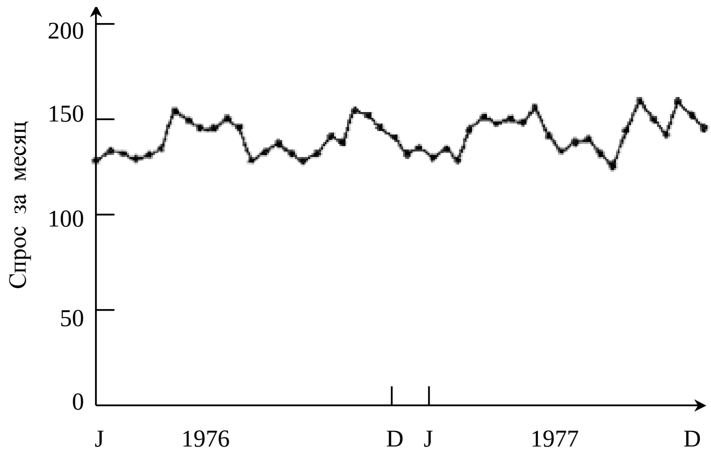


Рис. 5.10-а. Данные с сезонным трендом

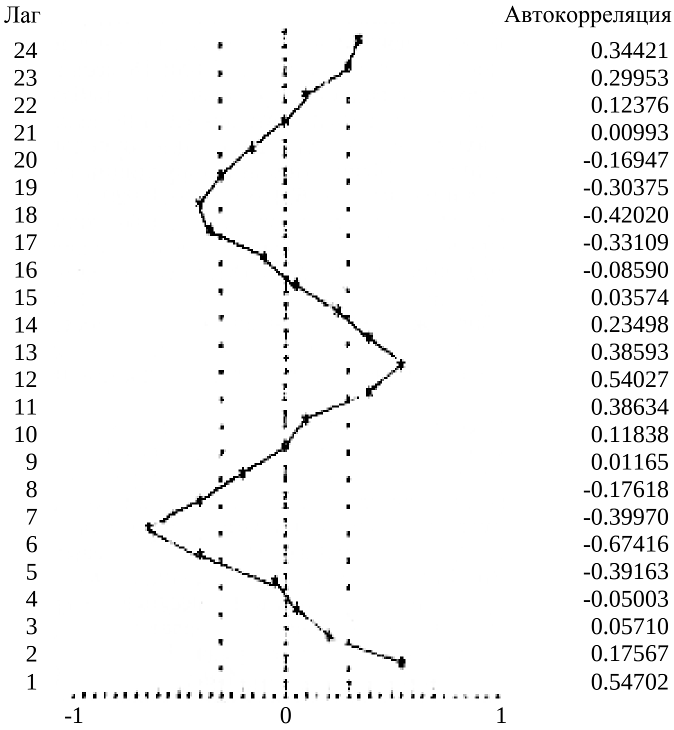


Рис. 5.10-б. Автокоррелограмма данных с сезонным трендом (нулевые разности)

5.2. Методы среднесрочного прогнозирования (регрессия и криволинейное выравнивание)

Среднесрочное прогнозирование, как правило, целесообразно в случаях, когда:

1) имеются ежегодные данные и их можно взять из официальных источников;

2) прогнозы являются одноразовыми, т.е. не повторяются и не подправляются (адаптируются) с поступлением новых данных;

3) прогнозы осуществляются для временных рядов относительно малой длины;

4) прогнозируется, например, динамика не отдельного объекта, а процесса, имеющего более общую природу.

В подобных случаях прогностические модели должны быть более сложными, чем модели краткосрочного прогнозирования. Метод линейной регрессии оказывается достаточно надежным, а соответствующая статистическая модель достаточно обоснованной. Этот метод линейных и нелинейных регрессий одновременно сочетает в себе не только относительную простоту вычислений, связанных с применением метода, но и возможность описания достаточно широкого класса процессов. Регрессионное и криволинейное выравнивание чуть ли не единственный способ построения среднесрочных прогнозов.

Прогностические модели, основанные на методах линейной регрессии, обладают рядом характерных особенностей:

– для применения этих методов ряды данных должны быть длиннее, чем для методов экспоненциального сглаживания;

– они не допускают адаптации: с добавлением новых данных процедура построения прогноза должна быть повторена заново;

– эти методы непригодны для сезонного прогнозирования.

При использовании регрессионных методов в анализе временных рядов предполагается, что каждому моменту времени t соответствует одно наблюдение, а всего имеется n наблюдений зависимой переменной y. Как и в случае краткосрочного прогнозирования период упреждения прогноза обозначается через τ, что соответствует числу точек, на которые строится прогноз. Техника регрессионного анализа изложена в учебном пособии «Системная экология» (Дулепов и др., 2004). В следующем параграфе подробно будет рассмотрен механизм криволинейного выравнивания, которое является одним из методов среднесрочного прогнозирования.

5.2.1. Криволинейное выравнивание (подбор кривых, сводящихся к линейному тренду)

Метод наименьших квадратов и процедуры подбора прямой регрессии полностью переносятся и на случай, когда уравнение кривой может быть после некоторых преобразований сведено к линейному тренду:

$$\hat{y}_t = a + bt,$$

где $\hat{y}_t$ – выравненное значение y_t, соответствующее моменту времени t; a и b – коэффициенты.

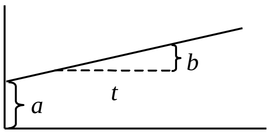
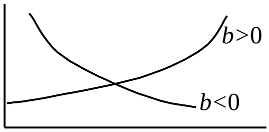
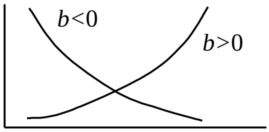
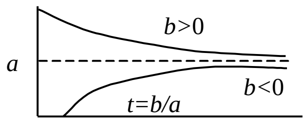
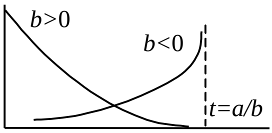
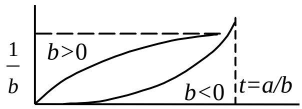
Для того чтобы найти прогноз, необходимо сначала оценить параметры линейного тренда, подставить их в исходное уравнение кривой, а затем вычислить прогноз.

В практике криволинейного выравнивания широко распространены два вида преобразований: натуральный логарифм (ln) и обратное преобразование (1/t). При этом возможно преобразование как зависимой переменной y, так и независимой t или одновременно и той и другой. Если преобразовывается только одна переменная, то название этого преобразования получает приставку «семи», например, «семи-логарифмы»; если преобразовываются обе переменные, то к названию преобразования добавляется слово «двойное» (Льюис, 1986).

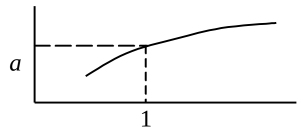
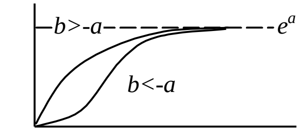
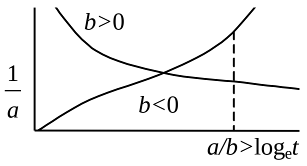
На основе логарифмирования и обратного преобразования разработаны восемь возможных преобразований кривых, полученных комбинацией из индивидуальных преобразований зависимой и независимой переменной. Эти кривые, их преобразования и вид представлены в табл. 5.5.

Таблица 5.5

Кривые, сводящиеся к прямой некоторым преобразованием

Название кривой	Уравнение	Преобразование	Вид кривой
1	2	3	4
Линейная	$\hat{y}_t = a + bt$	–	
Экспоненциальная (простая)	$\hat{y}_t = ae^{bt}$	$y_t = \ln y_t$	
Степенная	$\hat{y}_t = at^b$	$y_t = \ln y_t$ $T = \ln t$	
Гиперболическая I типа	$\hat{y}_t = a + b/t$	$T = 1/t$	
Гиперболическая II типа	$\hat{y}_t = 1/(a + bt)$	$y_t = 1/y_t$	
Гиперболическая III типа, или простая рациональная	$\hat{y}_t = t/(a + bt)$	$y_t = 1/y_t$ $T = 1/t$	

Окончание табл. 5.5

1	2	3	4
Логарифмическая	$\widehat{y}_t = a + b\ln t$	$T = \ln t$	
S-образная	$\widehat{y}_t = e^{a+b/t}$	$y_t = \ln t$ $T = 1/t$	
Обратнологариф-мическая	$\widehat{y}_t = \dfrac{1}{a + b\ln t}$	$y_t = 1/y_t$ $T = \ln t$	

Простая экспоненциальная кривая

Простая экспоненциальная кривая (экспонента) определяется уравнением

$$\widehat{y}_t = ae^{bt}, \tag{5.28}$$

От обеих частей уравнения (5.28) возьмем натуральный логарифм. Получим

$$\ln \widehat{y}_t = \ln a + bt \ln e. \tag{5.29}$$

Обозначим $Y_t = \ln y_t$, тогда

$$\widehat{Y}_t = a' + bt, \tag{5.30}$$

где $a'=\ln a$. a' и b могут быть найдены с помощью стандартной процедуры регрессионного анализа.

На рис. 5.11 показана простая экспоненциальная кривая, выравнивающая исходный ряд данных (ежегодные затраты на природоохранные мероприятия). Важное свойство простой экспоненты: темп роста (при $b>0$) для этой кривой постоянен для любого момента времени. Экспоненциальная модель широко применяется в экологических и экономических моделях.

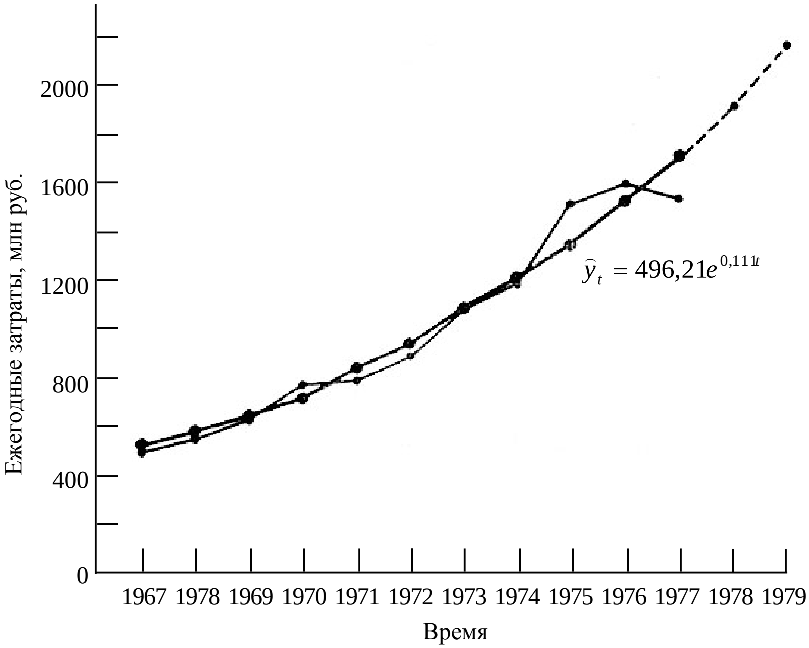


Рис. 5.11. Экспоненциальная кривая и прогноз
(прогноз обозначен пунктиром)

Степенная кривая

Уравнение степенной зависимости имеет вид

$$\hat{y}_t = at^b . \tag{5.32}$$

Возьмем натуральный логарифм от обеих частей уравнения (5.32)

$$\ln \hat{y}_t = \ln a + b \ln t \tag{5.33}$$

и обозначим логарифм зависимой переменной как $\hat{Y}_t = \ln \hat{y}_t$ и независимой t как T=ln t. Тогда уравнение (5.33) примет вид

$$\hat{Y}_t = a' + bT . \tag{5.34}$$

Для вычисления значений a=antilna' и b применяется стандартная процедура регрессии. На рис. 5.12 показан пример степенной кривой. Степенная кривая хорошо аппроксимирует показатели, непрерывно возрастающие со временем при положительном b или убывающие при b отрицательном.

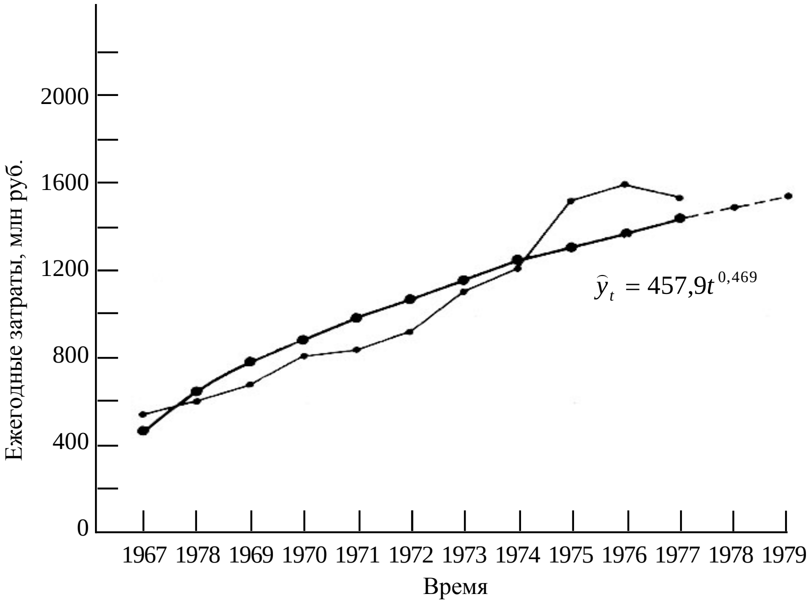


Рис. 5.12. Степенная кривая и прогноз

Гиперболическая кривая I типа

Обычная гипербола задается уравнением

$$\hat{y}_t = a + b/t\ . \tag{5.35}$$

Она может быть преобразована в линейную зависимость простым переобозначением независимой переменной. Произведем обратное преобразование переменной t. Тогда $T=1/t$ и уравнение (5.35) перепишется как

$$\hat{y}_t = a + bT\ . \tag{5.36}$$

Значения a и b находятся стандартными методами регрессионного анализа.

На рис. 5.13 показан пример гиперболической кривой I типа. Для гиперболы такого типа при $b>0$ значение $\hat{y}_t$ уменьшается с ростом t и асимптотически приближается к a. Подобного рода кривая может применяться для выравнивания и прогнозирования показателя, который с течением времени падает до некоторого отличного от нуля уровня. При $b<0$ значение $\hat{y}_t$ положительно, только если $t>b/a$; увеличение t приводит в этом случае и к увеличению $\hat{y}_t$ с асимптотической границей, равной a.

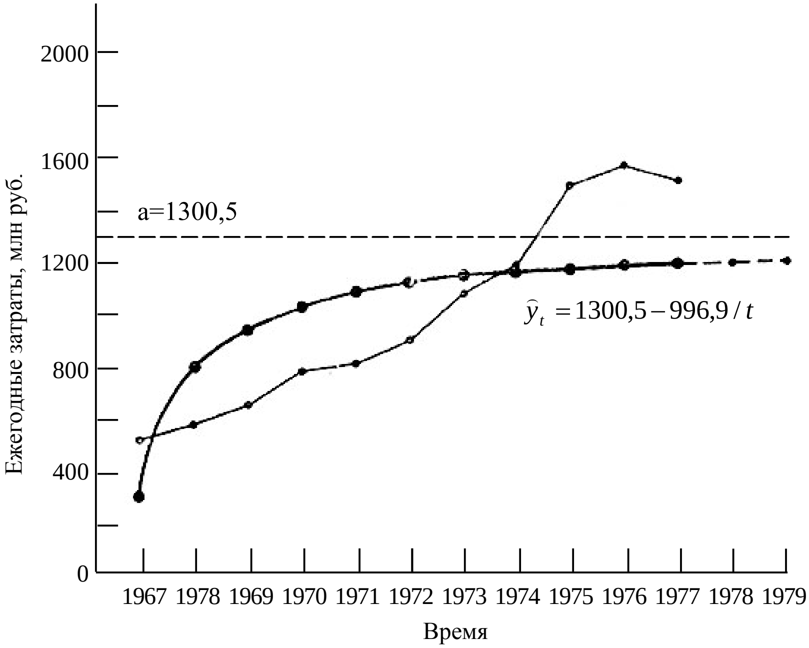


Рис. 5.13. Гиперболическая кривая I типа и прогноз

Гиперболическая кривая II типа

Этот тип гиперболы задают уравнением

$$\hat{y}_t = 1/(a + bt). \tag{5.37}$$

Данная гипербола сводится к линейному уравнению с помощью обратного преобразования зависимой переменной $\hat{y}_t$ (для гиперболы I типа такое преобразование касалось независимой переменной). Таким образом, если обозначить $Y_t=1/y_t$, то уравнение (5.37) будет выглядеть следующим образом:

$$\hat{Y}_t = a + bt. \tag{5.38}$$

На рис. 5.14 показан пример гиперболической кривой II типа. Для гиперболы такого типа при $b>0$ значение $\hat{y}_t$ стремится к нулю при неограниченном увеличении времени t; при $b<0$ значение $\hat{y}_t$ стремится к бесконечности при стремлении t к a/b. Последняя ситуация на практике маловероятна.

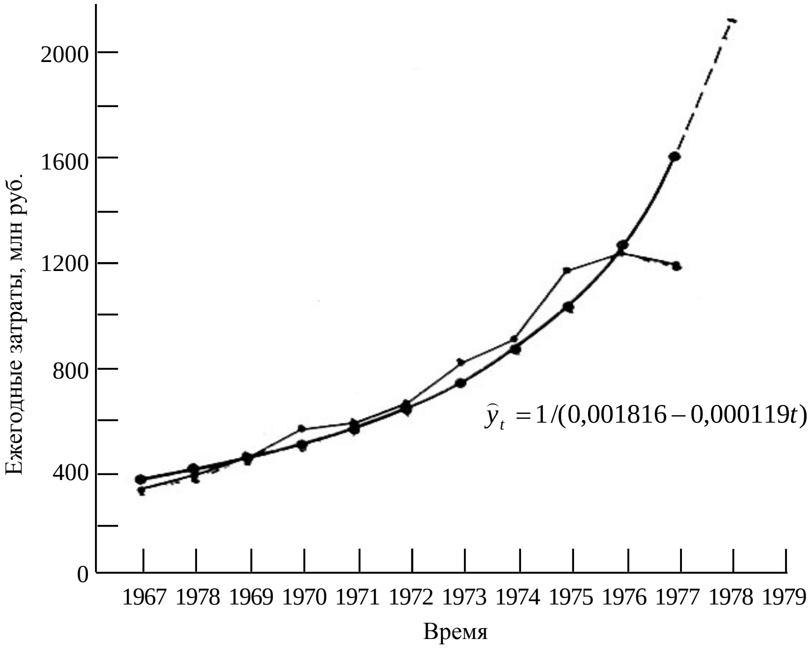


Рис. 5.14. Гиперболическая кривая II типа и прогноз

Гиперболическая кривая III типа (простая рациональная зависимость)

Такой тип гиперболической зависимости задается уравнением:

$$\hat{y}_t = t/(a+bt). \tag{5.39}$$

Этот несколько более сложный тип гиперболической зависимости (иногда называемый простой рациональной зависимостью) сводится к линейному уравнению переходом к обратным величинам зависимой ($Y_t=1/y_t$) и независимой ($T=1/t$) переменных. После подобных преобразований уравнение (5.39) перепишется в линейном виде:

$$\hat{Y}_t = a'T + b'. \tag{5.40}$$

При переходе от исходной зависимости к линейной смысл коэффициентов меняется: коэффициент *b'* определяет начальный уровень при *T*=0, а коэффициент *a'* – коэффициент наклона. На рис. 5.15 показана гипербола III типа. Для этого типа гиперболы, независимо от коэффициента *b*, при *t*=0 $\hat{y}_t=0$. Для положительных значений *b* значение $\hat{y}_t$ возрастает и асимптотически стремится к величине *1/b* при неограниченном увеличении *t*. При отрицательном *b* эта кривая, как и гипербола второго типа, становится неустойчивой при *t=a/b*.

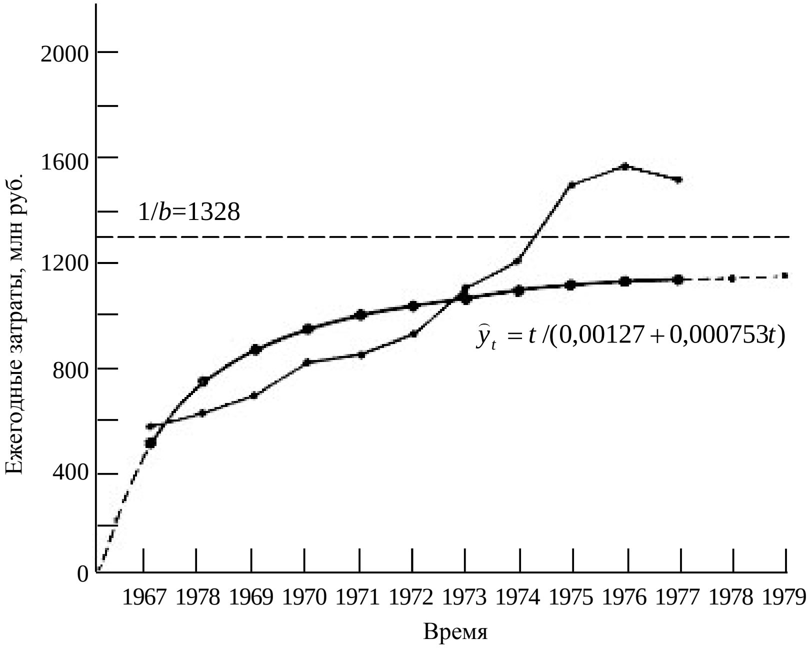


Рис. 5.15. Гиперболическая кривая III типа и прогноз

Логарифмическая кривая

Логарифмическая кривая задается уравнением

$$\hat{y}_t = a + b\ln t\,. \tag{5.41}$$

Заменяя в этом уравнении ln*t* на *T*, придем к линейной зависимости:

$$\hat{y}_t = a + bT\,. \tag{5.42}$$

Пример кривой с наилучшим качеством подгонки показан на рис. 5.16.

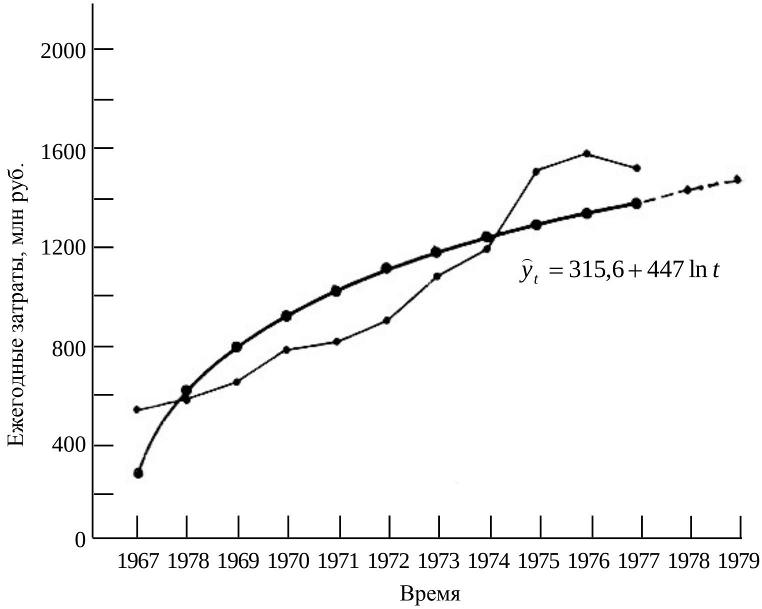


Рис. 5.16. Логарифмическая кривая и прогноз

S-образная кривая

S-образная кривая определяется выражением

$$\hat{y}_t = e^{a+b/t} . \qquad (5.43)$$

От обеих частей уравнения возьмем натуральный логарифм, тогда уравнение примет вид

$$\ln \hat{y}_t = a + b/t . \qquad (5.44)$$

После логарифмического преобразования зависимой переменной ($\hat{Y}_t = \ln \hat{y}_t$) и обратного преобразования независимой переменной (*T=1/t*) уравнение (5.44) сведется к линейному:

$$\hat{Y}_t = a + bT . \qquad (5.45)$$

Пример подобранной кривой и прогноза приведены на рис. 5.17.

На самом деле эта кривая имеет форму *S* только при отрицательном значении *b* и при условии, что его абсолютное значение больше *a*. Для приведенной на рисунке кривой *a*=7,169, *b*=-1,08, поэтому она не имеет формы S. Для иллюстрации возможного поведения кривых на рис. 5.17 показана S-образная кривая при *a*=7,69, *b*=-10. Если кривая (5.43) действительно имеет форму *S*, она используется для описания полного цикла развития динамических процессов. Полный цикл такого процесса начинается с медленного роста, затем следует фаза бурного развития и, наконец, развитие заканчивается периодом насыщения (т.е. асимптотического приближения к некоторому уровню, в данном случае к величине e_a). Подобная смена фаз характерна для многих процессов, протекающих в экосистемах. Для S-образной кривой точкой перегиба, т.е. точкой, в которой рост коэффициента наклона касательной сменяется падением, будет точка $t=-b/2$.

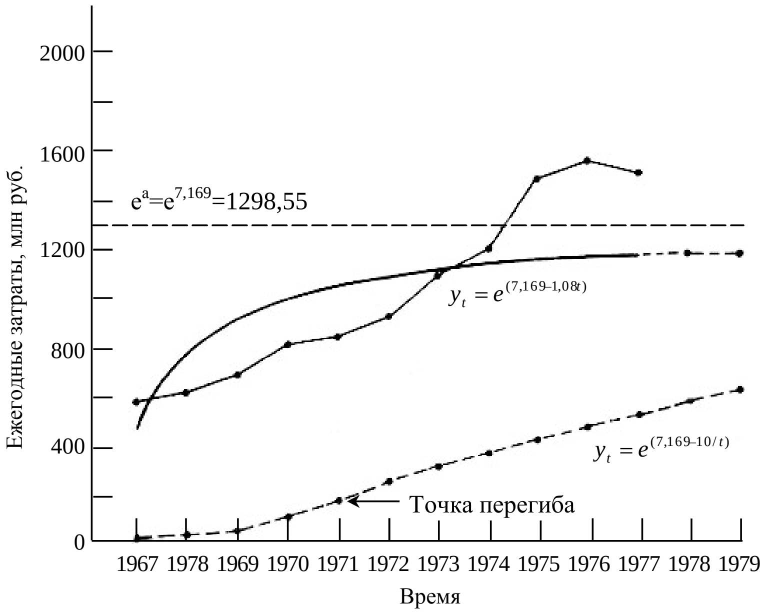


Рис. 5.17. S-образная кривая и прогноз

Обратнологарифмическая кривая

Обратнологарифмическая кривая определяется уравнением

$$\hat{y}_t = 1/(a + b\ln t). \tag{5.46}$$

Произведем обратное преобразование зависимой переменной y_t $(\hat{Y}_t = 1/y_t)$, а независимой – логарифмическое (T=ln t). Тогда уравнение (5.46) примет вид

$$\hat{Y}_t = a + bT\ . \tag{5.47}$$

Обратнологарифмическая кривая встречается в практике прогнозирования очень редко. Она рассматривается здесь лишь с точки зрения полноты возможных комбинаций двух преобразований: семидвойного логарифмирования и/или обратного преобразования, сводящих нелинейную зависимость к линейной. На рис. 5.18 приведен пример подогнанной кривой и построенного прогноза.

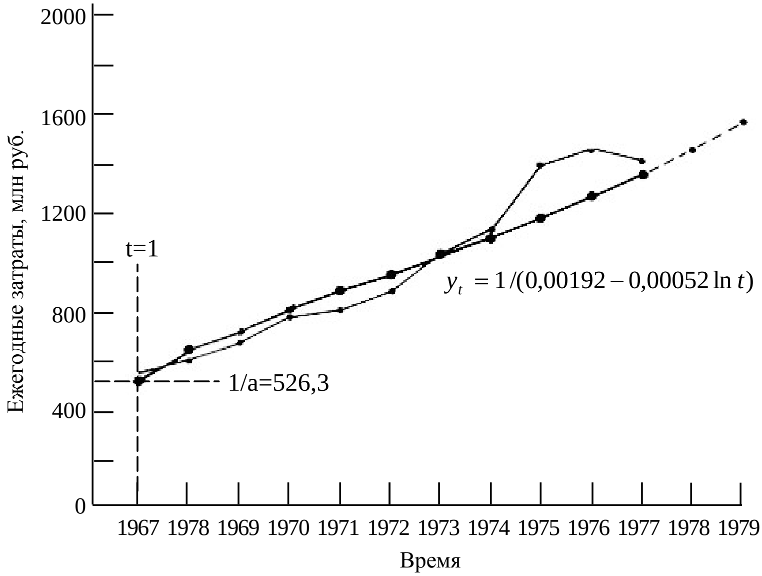


Рис. 5.18. Обратнологарифмическая кривая и прогноз

5.2.2. Выравнивание по кривым, сводящимся к модифицированной экспоненте

Все кривые, кроме S-образной, описывают ситуации, когда коэффициент наклона касательной либо возрастает, либо убывает. Однако иногда встречаются данные, которые необходимо описывать кривыми, имеющими точку перегиба, т.е. точку, где рост наклона касательной сменяется падением или, наоборот, падение сменяется ростом. При этом динамика явления такова:

- вначале рост довольно медленный, затем он убыстряется;
- промежуточный период роста сменяется третьим периодом;
- третий период – уменьшение роста и приближение к уровню насыщения.

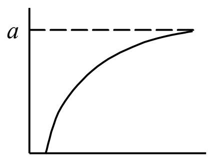
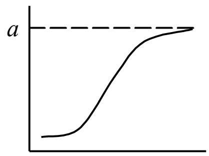
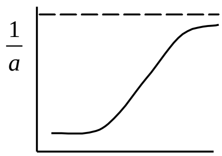
Подобная динамика часто возникает при описании экологических процессов, в частности при описании полного цикла. Широко распространенными кривыми, обладающими точкой перегиба, являются логистическая кривая и кривая Гомперца. Они точнее всего описывают процессы полного цикла.

Логистическая кривая и кривая Гомперца могут быть получены из другой кривой, известной как модифицированная экспонента, тем же способом, каким были получены кривые из обычной линейной регрессии.

Кривые, построенные по модифицированной экспоненте, задаются тремя параметрами (вместо двух параметров при линейной зависимости). Вследствие этого эти кривые будут давать лучшие результаты подгонки, однако они требуют и больших вычислительных затрат. В табл. 5.6 даны уравнения трех кривых, соответствующие преобразования и их стандартный вид.

Таблица 5.6

Кривые, сводящиеся к модифицированной экспоненте некоторым преобразованием

Название кривой	Уравнение	Преобразование	Вид кривой
Модифицированная экспонента	$\hat{y}_t = a + bc^t$	–	
Кривая Гомперца	$\hat{y}_t = ab^{c^t}$	$y_t = \lg y_t$	
Логистическая	$\hat{y}_t = 1/(a + bc^t)$	$y_t = 1/y_t$	

Модифицированная экспонента

Модифицированная экспонента сама по себе не имеет точки перегиба. Уравнение этой кривой следующее:

$$\hat{y}_t = a + b\tilde{n}^t \; . \qquad (5.48)$$

Для ее задания необходимо определить три параметра: a, b и c. Обычная процедура метода наименьших квадратов непосредственно к модифицированной экспоненте неприложима.

Подгонка модифицированной экспоненты

Параметры модифицированной экспоненты в порядке их вычисления определяются следующим образом:

$$\tilde{n} = \frac{(n-1)\sum_{t=0}^{n-1} y_t y_{t+1} - \sum_{t=1}^{n-1} y_t \sum_{t=1}^{n-1} y_{y+1}}{(n-1)\sum_{t=1}^{n-1} y_t^2 - (\sum_{t=1}^{n-1} y_y)^2}, \tag{5.49}$$

$$b = \frac{n\sum c^t y_t - \sum c^t \sum y_t}{n\sum c^{2t} - (\sum c^t)^2}, \tag{5.50}$$

$$a = \frac{\sum y_t - b\sum c^t}{n}. \tag{5.51}$$

Модифицированная экспонента, как правило, служит базовой кривой, на основе которой с помощью некоторых преобразований получаются используемые чаще всего кривая Гомперца и логистическая кривая. Однако это не означает, что по модифицированной экспоненте нельзя получить хорошей подгонки данных и прогноза. Так, если в начальный период времени величина прогнозируемого показателя довольно высока, и скорость увеличения его также значительна, а с течением временем она имеет тенденцию к стабилизации или насыщению, т.е. фаза медленного роста отсутствует, то наилучшими качествами подгонки будет обладать как раз модифицированная экспонента. На рис. 5.19 приведен пример кривой, построенной на основе модифицированной экспоненты.

Кривая Гомперца

Кривая Гомперца определяется уравнением

$$\hat{y}_t = ab^{c^t}. \tag{5.52}$$

Возьмем от обеих частей этого уравнения натуральный логарифм или логарифм по основанию 10, получим

$$\lg \hat{y}_t = \lg a + c^t \lg b, \tag{5.53}$$

что приводится к виду

$$\hat{Y}_t = a' + b'c^t, \tag{5.54}$$

где

$$\hat{Y}_t = \lg \hat{y}_t; \qquad a' = \lg a; \qquad b' = \lg b.$$

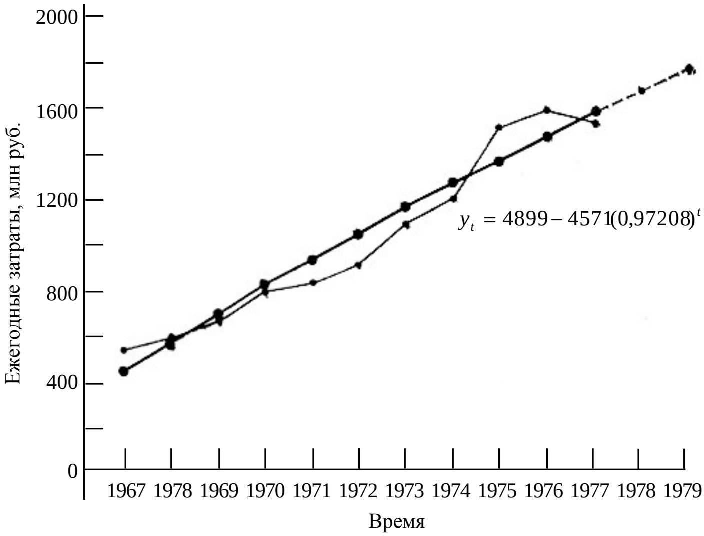


Рис. 5.19. Модифицированная экспонента и прогноз

Уравнение (5.54) имеет вид модифицированной экспоненты, поэтому параметры *a',b',c* могут быть получены по формулам (5.49) – (5.51). Выравненные значения и прогноз находятся из уравнения (5.52). Кривая Гомперца имеет S-образную форму и поэтому часто применяется для описания процессов полного цикла. Точкой перегиба для этой кривой будет

$$t_p = \frac{1}{\ln c}\ln(-\frac{1}{\ln b}) \qquad (5.55)$$

со значением функции $\hat{y}_{t_p} = a/e$, e – основание натурального логарифма. На рис. 5.20 приведен пример кривой Гомперца с отмеченной точкой перегиба. В этой точке прирост затрат достигает максимума и равен 120 единицам, далее этот прирост постепенно падает.

Логистическая кривая

Логистическая кривая задается уравнением

$$\hat{y}_t = 1/(a + b\tilde{n}^t)\,. \qquad (5.56)$$

Произведем обратное преобразование левой и правой частей этого уравнения:

$$\hat{Y}_t = a + bc^t\,, \qquad (5.57)$$

где $\hat{Y}_t = 1/\hat{y}_t$. Уравнение (5.57) имеет вид модифицированной экспоненты и ее параметры легко находятся из уравнений (5.49) – (5.51). Пример выравненных значений и прогноз на две точки вперед приведены на рис. 5.21. Логистическая кривая имеет S-образную форму с точкой перегиба, равной

$$t_p = \frac{1}{\ln c} \ln \frac{a}{b}. \tag{5.58}$$

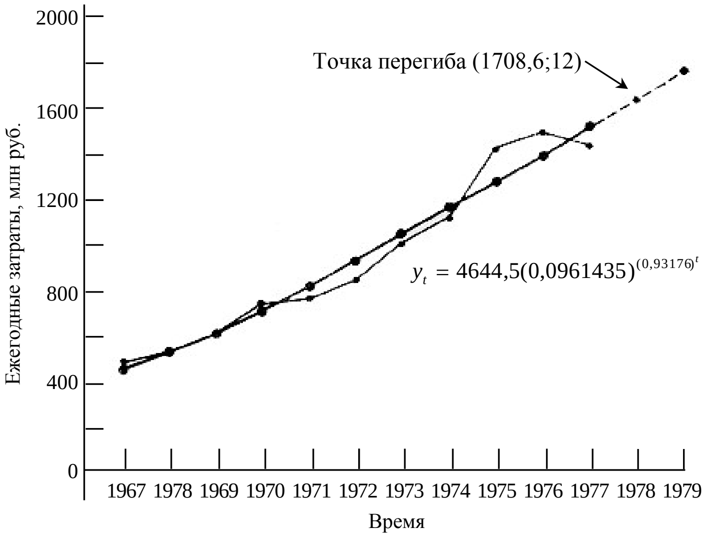


Рис. 5.20. Кривая Гомперца и прогноз

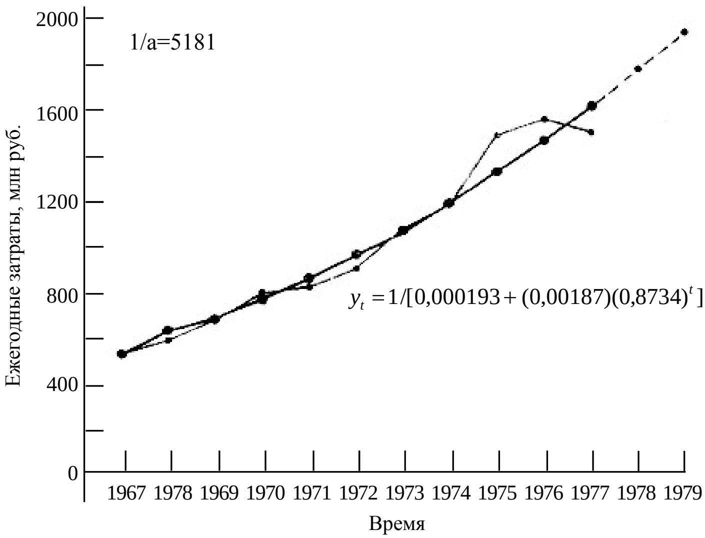


Рис. 5.21. Логистическая кривая и прогноз

Значение $\hat{y}$ в точке перегиба равно $\hat{y}_{t_p} = 1/2a$.

5.3. Элементы планирования эксперимента

5.3.1. Сущность метода

Проблемы выборочного метода и вопросы планирования эксперимента при экосистемных исследованиях имеют большое значение для минимизации затрат при проведении экспедиционных и натурных наблюдений и оптимизации исследования различных характеристик экологических систем. Под экспериментом принято понимать создание некоторого комплекса условий U, в результате которых могут наступать (или не наступать) события из некоторого заданного множества V. Если на результаты эксперимента могут влиять случайные события той или иной вероятностной природы, то этот эксперимент называется статистическим.

Понятие и цели планирования эксперимента. Последовательность действий экспериментатора можно условно разбить на три этапа:

- определение условий проведения эксперимента;
- реализация эксперимента, сопровождающаяся некоторыми измерениями характеристик объекта;
- обработка полученных данных и анализ результатов.

Условия эксперимента обычно зависят от одного или нескольких факторов, которые исследователь может выбирать по своему усмотрению. Такой эксперимент называется активным, а сами факторы – контролируемыми переменными. Каждый отдельный фактор может иметь качественную или количественную природу. Для качественных факторов не существует понятия «расстояния» и введение какой-либо метрики не имеет смысла. Количественные факторы принадлежат к некоторому метрическому пространству (Заславский, Полуэктов, 1988).

Обозначим область изменения контролируемых переменных через Σ и введем для элемента этой области обозначение x. В конечномерном пространстве Σ может ассоциироваться с частью евклидова пространства, $\Sigma \subset R^n$, а x – с некоторым n-мерным вектором, $x=(x_1,x_2,\ldots,x_n)$. 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 $x^{(1)},x^{(2)},\ldots,x^{(N)}$ и числом опытов N. При этом число опытов N может заранее не фиксироваться и *a'priori* быть неограниченным, а величины $x^{(i)}$ не всегда могут быть различны.

Рассмотрим частный случай эксперимента, так называемый «факторный» эксперимент. Пусть Σ – некоторая область евклидова пространства R^n. «Классическая» схема эксперимента строится следующим образом. Все факторы, кроме одного, фиксируются на определенных уровнях, а один из них, например x_i, изменяется, принимая значения $x_i^{(1)}, x_i^{(2)}, ..., x_i^{l_i}$, 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 $N=l_1,l_2,...,l_n$ точках пространства контролируемых переменных. Число и расположение точек в факторном пространстве и называется «планом факторного эксперимента».

В процессе реализации эксперимента (второй этап) производится измерение характеристик объекта, т.е. в каждой из N точек фиксируется некоторый результат $y^{(i)}$. Переменная y может быть скалярной, векторной величиной. В статистическом эксперименте значение $y^{(i)}$ является реализацией некоторой случайной величины или случайного процесса. Каковы источники этой случайности? В простейшем случае ошибки измерений могут рассматриваться как аддитивные случайные величины с заданными корреляционными свойствами. С другой стороны, сам объект может иметь стохастическую природу.

Пусть ω – элементарное событие из некоторого вероятностного пространства Ω. Тогда результаты эксперимента в точке $x^{(i)}$ факторного пространства можно представить как реализацию случайной функции:

$$y^{(i)} = A(x^{(i)}, \omega) \qquad (i = \overline{1, N}), \tag{5.59}$$

где A – некоторый оператор, действующий из ΣxΩ в Y.

Рассмотрим содержание третьего этапа. В стохастическом эксперименте анализ и обработка данных должны производиться 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 (5.59). 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. Поскольку в этих случаях используются методы регрессионного анализа, эти эксперименты получили название регрессионных.

Задача дисперсионного эксперимента – получить ответ на вопрос о влиянии контролируемых переменных на «выход» объекта y. Следует оценить значимость так называемых «главных эффектов» факторов и их взаимодействий. При этом каждый из факторов может принимать конечное число градаций, 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

$$M\{y^{(i)} \big| x^{(i)}\} = \eta(x^{(i)}, \theta), \tag{5.60}$$

где $y^{(i)}$ – результат измерения в точке $x^{(i)}$, $\eta(x,\theta)$ – функционал, вид которого известен, а $\theta=(\theta_1,\theta_2,\ldots,\theta_m)^*$ – вектор неизвестных параметров. Оператор $M\{\ldots\}$ 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 (5.59) в некоторой области измерения контролируемых переменных Σ. Наиболее развиты и хорошо обоснованы методы получения линейных оценок для тех случаев, когда неизвестные параметры входят в модель линейно.

Сформулируем понятие плана эксперимента. Планом эксперимента называется совокупность величин

$$(x^{(1)}, r_1), (x^{(2)}, r_2), \ldots, (x^{(l)}, r_l), \tag{5.61}$$

где r_i – так называемые повторности эксперимента. 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. 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.

Каковы же эти критерии? Постановка единичного (т.е. без повторностей) опыта в каждой точке множества Σ требует затрат определенных ресурсов, независимо от того, идет ли речь о натурных 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 эксперимента. Например, в эксперименте по проверке 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 (5.61) 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 исследователя часто интересует качество оценок параметров, мерой чего может явиться дисперсионная матрица оценок.

Для осуществления первых двух этапов прежде всего необходимо решить проблему организации выборочного наблюдения, которое и ляжет в основу эксперимента.

Необходимо составление таких планов исследований, которые давали бы возможность получить как можно больше информации из каждого эксперимента. Неправильная организации выборочных обследований, с точки зрения оснований приложения статистических приемов оценки из результатов, способна свести на нет всю работу, приводя к ложным выводам или не позволяя ответить на поставленные вопросы. Точно так же и план проведения эксперимента или натурных наблюдений должен быть правилен с точки зрения статистики. Казалось бы, на первый взгляд, что нет ничего общего между выборочным наблюдением и экспериментом. Выборочное наблюдение – это по преимуществу средство изучения популяции в широком смысле, когда исследователь стремится распространить его результаты на всю совокупность экземпляров данного вида или группы и т.п. Экспериментатор пытается ответить на вопрос об эффективности действия изучаемого в опыте фактора (или группы факторов) на исследуемое явление или процесс. Здесь общее заключается в том, что и теория выборочного метода и планирование эксперимента опираются на одни и те же законы – законы, управляющие случаем. Случайная ошибка выборки или эксперимента является здесь единственным мерилом, которое лежит в основе всех суждений исследователя. Поэтому и принципиальные логические черты, характеризующие оба этих процесса, остаются одними и теми же. Мы не будем пытаться охватить все многообразие форм организации выборочных наблюдений и планирования эксперимента, а рассмотрим только их основные элементы с некоторыми примерами, которые могут пригодиться при проведении экосистемных исследований и анализе данных (Дружинин, 1977).

5.3.2. Основные формы организации выборочного статистического наблюдения и планирования эксперимента

На первых стадиях развития организация выборочного обследования и планирования экспериментальной работы была менее сложной, чем в настоящее время. Первоначально эта теория касалась лишь простой случайной выборки, требовавшей соблюдения лишь одного условия: не нарушать законы, управляющие случаем. В организационном плане это требовало формирования выборочных совокупностей из единиц, отбираемых так, чтобы не отдавалось никакого предпочтения одним перед другими. В экспериментальной же работе это требование обусловило принцип рандомизации вариантов опыта, или размещения их в случайном порядке. Очень часто экспериментатор встречается с реальными затруднениями в осуществлении случайного отбора. Эти затруднения возникают в связи с тем, что исходная совокупность, из которой производится отбор, или условия эксперимента зачастую оказываются неоднородными. Разумеется, что отбор из такой неоднородной совокупности приводит к нарушению законов случая, которое выражается в «смещенной» оценке результатов отбора. Что касается планирования опыта, то его рандомизация возможна, вероятно, только в лабораторных экспериментах, когда вся обстановка опыта позволяет устранять различные посторонние «возмущающие» влияния. Это может относиться, например, к физическим или химическим экспериментам. Но в большинстве случаев эксперименты приходится проводить в неоднородной среде, исключающей их полную рандомизацию. Из этих затруднений, встречающихся на пути организации чисто случайной выборки, необходимо было найти какой-нибудь выход.

Этот выход для выборочного наблюдения заключается в приеме, который носит название стратификации (расслоения) и является одним из способов ограничения случайного отбора. Он состоит в отборе единиц из предварительно расчлененной на «страты» или группы исходной совокупности. Случайная ошибка выборки вычисляется на основе средней величины внутригрупповых дисперсий. Но если при определении величины ошибки учитывается только внутригрупповая вариация данных, то, следовательно, в образовании этой ошибки уже не участвует вариация межгрупповая, обусловленная различиями между группами. Общая дисперсия σ_x^2, если произведена группировка данных, может быть разложена на два компонента: дисперсию групповых средних – $\sigma_{\bar{x}_i}^2$ и среднюю величину внутригрупповых дисперсий – $\overline{\sigma_{x_i}^2}$. При стратификации при вычислении случайной ошибки используется лишь второй компонент, который характеризует вариацию данных внутри групп и поэтому может рассматриваться как показатель только случайной вариации. Чем удачнее будут отражены при установлении групп условия неоднородности, тем точнее будет определена и случайная ошибка отбора. Главная задача здесь заключается в правильном выборе признака для расчленения общей совокупности на группы. Стратифицированную выборку принято называть «типической» потому, что для выделения групп-страт должен быть выбран какой-то признак, обуславливающий выражение того или иного «типа» условий, в которых совершаются изучаемые явления, или «типа» самих этих явлений.

Изменение средних величин признака по стратам можно представлять как эмпирическую регрессию, характеризующую зависимость этих изменений от изменения признака стратификации. Степень тесноты этой зависимости в большей мере обусловлена числом групп-страт. Увеличение числа групп приводит, как правило, к ослаблению корреляционной связи. Если эту связь выразить величиной коэффициента детерминации (долей «объясненной» дисперсии в общей дисперсии) и допустить, например, что коэффициент детерминации $R^2 = 0{,}64$, то, применяя модель $R^2/H^2+(1\text{-}R^2)$ (Kish, 1967), где H^2 – квадрат числа страт, видим, что уже шесть страт сократят величину коэффициента детерминации до (0,018+0,36)=0,378, а 12 страт – до 0,364. Здесь вместе с ослаблением корреляционной связи будет уменьшаться и вариация средних по стратам. Таким образом, увеличивая число страт, исследователь может лишиться возможности устранения первого компонента общей вариации данных при определении величины случайной ошибки.

Тем не менее, в приеме стратификации содержится возможность сокращения случайной ошибки выборки при преодолении в то же время нежелательного влияния условий неоднородности, в которых проводится выборочное наблюдение. Важен лишь удачный выбор признака стратификации, позволяющий создать ряд однородных групп, характеризующихся сравнительно небольшой вариацией признака внутри себя. Если опять обратиться к терминам корреляции, то такое требование будет удовлетворяться при одновременном возрастании коэффициента детерминации и так называемого коэффициента внутриклассовой или внутригрупповой корреляции:

$$\bar{r} = \frac{\dfrac{n\sigma_m^2}{\sigma_0^2} - 1}{(n-1)}, \tag{5.62}$$

где n – число страт (групп); σ_m^2 – дисперсия средних по группам; σ_0^2 – общая дисперсия. Чем больше положительная величина этого показателя, тем теснее связь между единицами, входящими в состав групп, или, иными словами, тем однороднее состав этой группы.

Применительно к экспериментальной работе, где речь идет о размещении вариантов опыта, нужно уже говорить не об ограничении случайного отбора, а об ограничении рандомизации в условиях неоднородности. Способы ограничения рандомизации разнообразны. Простейшим примером таких способов является так называемый «парный метод», который очень распространен в современной экспериментальной биологии. Принцип, на который опирается этот метод, состоит в том, что вовсе не требуется, чтобы единообразие условий соблюдалось для всех повторений опыта, важно обеспечить это единообразие лишь для каждой пары сравниваемых его вариантов. Таким образом, исследователь получает ряд разностей, происхождение которых почти свободно от всяких внешних случайных влияний. Можно считать, что эти разности обусловлены ничем не «возмущаемым» действием испытываемых в опыте факторов. Но они при повторении опыта изменяются, подвергаясь влиянию тех неоднородностей, в условиях которых производятся повторения.

Широко применяемое планирование опытов в виде так называемых «латинских квадратов» может служить более сложным примером ограничения рандомизации в условиях неоднородностей, когда сравниваются уже не два, а несколько вариантов опыта. Принцип латинского квадрата состоит в размещении вариантов опыта так, что они встречаются только один раз в каждой строке и в каждой колонке этого квадрата. Например, при четырех вариантах опыта схема латинского квадрата может быть представлена так:

А	В	С	Д
В	С	Д	А
С	Д	А	В
Д	А	В	С

Такое расположение вариантов опыта дает возможность исключить из экспериментальной ошибки влияние неоднородностей уже в двух направлениях: по рядам и по столбцам. Если схема латинского квадрата применяется в опыте с кормлением рыбы, то способ разведения и условия содержания рыбы окажутся теми двумя источниками неоднородности, влияние которых должно быть исключено из экспериментальной ошибки.

Таковы основные из тех приемов, которые применяются при организации выборочного наблюдения или при планировании эксперимента для того, чтобы избавиться от ненужного влияния неоднородности условий проведения этих работ.

5.3.3. Принципы стратификации

Эффективность стратификации зависит от выбора признака стратификации. Стратифицированная выборка, как уже говорилось, носит еще и название «типической». Примером такой типической выборки служит совокупность промышленных предприятий города при их выборочном изучении как источников загрязнения атмосферы или акватории водного бассейна. Эту исходную совокупность при одинаковой специализации предприятий можно разделить на группы, находящиеся в разных частях города, при разной специализации, – на группы производственного назначения. Выбор признака стратификации не всегда прост, сложность заключается в том, что важно не только выбрать признак, который носил бы характер признака-фактора, но и наметить такое число групп, при котором отчетливо проявилось бы действие этого признака. Каждый отдельный случай требует особого подхода к решению этой задачи. При стратификации слишком большое число групп, уменьшая вариацию групповых средних, может свести на нет ее преимущества перед простой случайной выборкой.

При стратификации важно добиваться такого положения, чтобы средние по группам сколько возможно отличались друг от друга при однородности единиц внутри групп. Это означает, что чем больше положительная величина внутриклассовой корреляции $\bar{r}$, т.е. чем выше степень сходства единиц, входящих в состав групп, тем более выгодное создается положение: возрастает величина дисперсии групповых средних ($\sigma^2_{\bar{x}_i}$) и уменьшается средняя величина внутригрупповых дисперсий ($\overline{\sigma^2_{x_i}}$).

Важной задачей при осуществлении стратифицированной выборки является размещение всего числа отбираемых единиц внутри страт. Обычно, если численность групп-страт, на которые расчленяется исходная совокупность, неизвестна, эта задача на практике решается упрощенно путем отбора из каждой страты равного числа единиц. Но в тех случаях, когда состав исходной совокупности известен, требуется иной подход, заключающийся в том, что доля отбираемых в каждой группе единиц соответствовала бы удельному весу этой группы в исходной совокупности, т.е. $\frac{n_1}{n} = \frac{N_1}{N}$ (где n_1 – численность отбираемых в группе единиц; n – общая численность выборки; N_1 – численность единиц в группе; N – численность исходной совокупности). Такой отбор носит название пропорциональной стратифицированной выборки. Примером такой выборки служит табл. 5.7, данные которой характеризуют случайный отбор 500 ферм из общего их числа 2072, сгруппированных по размерам площади для определения числа голов скота в среднем на ферму (Дружинин, 1977).

Из табл. видно, что $\frac{n_i}{n} = \frac{N_i}{N}$, т.е. отношение числа отобранных в каждой группе ферм к общей численности выборки действительно соответствует весу группы в исходной совокупности. Оценке в этой выборке подлежит численность скота в среднем на ферму, которая равна $y = \frac{6094}{500}$ =12,19.

Таблица 5.7

Пропорциональная стратифицированная выборка ферм

Группы ферм по размерам площади	Число ферм в группе N_i	Доля группы в общей численности ферм $N_i/N=W_i$	Число ферм в случайной выборке n_i	Внутригрупповые дисперсии для численности скота s^2_i	Численность скота на фермах y_i
До 15	635	0,307	153	20,19	619
16–30	570	0,275	138	69,96	1423
31–50	475	0,229	115	63,13	1758
51–75	303	0,146	73	170,32	1691
76–100	89	0,043	21	184,9	603
N=2072		1,0	n=500		6094

Для вычисления дисперсии выборочных средних или квадрата ошибки стратифицированной выборки при случайном отборе единиц из групп можно воспользоваться следующей формулой:

$$m^2_{(\bar{y})} = \sum_{i=1}^{k} W_i^2 (1-\frac{n_i}{N_i}) \frac{s_i^2}{n_i} = \sum_{i=1}^{k} \left(\frac{1}{n_i} - \frac{1}{N_i} \right) W_i^2 s_i^2 , \qquad (5.63)$$

где k – число групп.

Все необходимые расчеты даны в табл. 5.8.

Таблица 5.8

Вычисление квадрата ошибки

W_i^2	$W_i^2 \quad s_i^2$	$\left(\frac{1}{n_i} - \frac{1}{N_i} \right)$	$\left(\frac{1}{n_i} - \frac{1}{N_i} \right) W_i^2 \quad s_i^2$
0,094	1,9	0,0049	0,009
0,076	5,32	0,0054	0,029
0,052	3,28	0,0066	0,022
0,021	3,58	0,0104	0,037
0,002	0,37	0,0364	0,013
			0,11

Можно видеть, что не только все $\frac{n_i}{n} = \frac{N_i}{N}$, но также, поскольку отбор был пропорциональным, все $\frac{n_i}{N_i} = \frac{n}{N}$. Таким образом, сомножитель

$(1-\dfrac{n_i}{N_i})$ в формуле (5.63) может быть заменен выражением $(1-\dfrac{n}{N})$ и вынесен за знак суммы.

Если представить n_i как nW_i, то формулу квадрата ошибки пропорционально стратифицированного отбора можно записать так:

$$m^2_{(\bar{y})} = \sum W_i^2 \frac{s_i^2}{n}\left(\frac{N-n}{N}\right) = \left(\frac{1}{n} - \frac{1}{N}\right)\sum W_i^2 s_i^2. \qquad (5.64)$$

Согласно этой формуле $m^2_{(\bar{y})} = \left(\dfrac{1}{500} - \dfrac{1}{2072}\right)\cdot 72{,}7115 \approx 0{,}11$.

Эта задача может быть решена и с помощью дисперсионного анализа. Суть метода заключается в сравнении общей дисперсии, дисперсии внутри групп и межгрупповой дисперсии. Если дисперсия, характеризующая вариацию данных внутри групп, на основе которой и вычисляется случайная ошибка выборки, будет меньше общей дисперсии, образующей ошибку простой случайной выборки, то стратификация окажется эффективной.

Организация пропорционального стратифицированного отбора не всегда возможна. Иногда статистическое распределение исходной совокупности оказывается сильно асимметричным, и группы с более высокими показателями становятся малочисленными до такой степени, что при пропорциональном отборе из них пришлось бы отбирать ничтожное число единиц, соответственно, такой отбор станет нерепрезентативным. В таких условиях лучше проводить так называемый оптимальный отбор единиц из групп, основанный на учете степени их вариации, т.е. отбирать больше единиц там, где больше их вариация, и меньше – при меньшей вариации. Практически это означает, что число единиц, которые надо отобрать из групп, должно определяться как $n_i = n\left(\dfrac{W_i s_i}{\sum W_i s_i}\right)$, т.е. как произведение численности всей выборки (n) и отношения среднеквадратического отклонения данной группы (s_i), взвешенного ее долей в исходной совокупности (W_i), к сумме квадратических отклонений всех групп, также взвешенных долями этих групп $(\sum W_i s_i)$.

Вычисление ошибки при оптимальном отборе может быть произведено по формуле (5.63) с учетом того, что $n_i = n\left(\dfrac{W_i s_i}{\sum W_i s_i}\right)$. Тогда формула (5.63) преобразуется следующим образом:

$$m^2_{(\bar{y})} = \frac{(\sum W_i\ s_i)^2}{n} - \frac{\sum W_i s_i^2}{N}. \qquad (5.65)$$

В этой формуле $\frac{(\sum W_i\ s_i)^2}{n} = \sum \frac{W_i s_i^2}{n_i}$.

Пропорциональное или оптимальное распределение отбираемых единиц не исключают и какого-либо другого порядка отбора единиц по группам. Так, например, если точные границы трудно определить, то прибегают иногда к так называемому «равному» распределению, т.е. из каждой группы отбирают одинаковое число единиц.

Преимущество стратифицированного отбора может быть легко показано и при сравнении формул квадратов ошибок. Такая сравнительная характеристика стратифицированной выборки с простой случайной выборкой открывает возможность для общей их оценки. Если для простоты в этих формулах опустить множитель $(\frac{N-n}{N})$, предполагая небольшую долю повторной выборки из конечной совокупности, то квадраты ошибок могут быть представлены так:

$$m_1^2 = \frac{s^2}{n}, \qquad m_2^2 = \frac{\sum N_i s_i^2}{nN}, \qquad m_3^3 = \frac{(\sum N_i s_i)^2}{nN^2},$$

где $m^2{}_1$ – при простом случайном отборе; $m^2{}_2$ – при пропорциональном размещении; $m^2{}_3$ – при оптимальном размещении. Общая сумма квадратов отклонений равна сумме квадратов отклонений внутри групп плюс сумма квадратов отклонений средних по группам от общей средней:

$$Ns^2 = \sum N_i s_i^2 + \sum N_i (\overline{y_i} - \overline{y})^2 .$$

Отсюда

$$m_1^2 = \frac{s^2}{n} = \frac{\sum N_i s_i^2}{nN} + \frac{\sum N_i (\overline{y_i} - \overline{y})^2}{nN} = m_2^2 + \frac{\sum N_i (\overline{y_i} - \overline{y})^2}{nN} .$$

Следовательно, квадрат ошибки при пропорциональном размещении меньше квадрата простой случайной выборки на величину того компонента последней, который обусловлен вариацией средних по группам.

Различие между квадратами ошибок при пропорциональном и оптимальном размещениях таково:

$$m_2^2 - m_3^2 = \frac{1}{nN}\left(\sum N_i s_i^2 - \frac{(\sum N_i s_i)^2}{N}\right) = \frac{1}{nN}\sum (s_1 - \bar{s})^2 ,$$

где $\bar{s} = \frac{\sum W_i s_i}{N}$ – средняя величина средних квадратических отклонений в группах. Таким образом,

$$m_1^2 = m_3^2 + \frac{\sum N_i(s_i - \bar{s})^2}{nN} + \frac{\sum N_i(\overline{y_i} - \bar{y})^2}{nN},$$

т.е. уменьшение квадрата ошибки при оптимальном размещении сравнительно с простой случайной выборкой происходит уже в результате исключения не только вариации средних по группам, но и вариации величин средних квадратических отклонений. Последнее обстоятельство обусловливается самим принципом оптимального размещения, заключающимся в отборе такого числа единиц из групп, которое оказывалось бы пропорциональным показателю их вариации, т.е. среднему квадратическому отклонению.

Преимущества стратифицированной выборки перед простой случайной очевидны, что не всегда бывает так при сравнении оптимального размещения с пропорциональным или при сравнении обоих этих размещений с каким-либо произвольным. Это происходит потому, что даже при существенном отклонении размещения от оптимального величина квадрата ошибки не возрастает настолько, чтобы можно было говорить о большом снижении точности выборки.

Таким образом, в стратифицированной выборке не всегда пропорциональное или оптимальное размещение отбираемых единиц приведет к такому повышению точности отбора, которое оправдает затраты труда на организацию этих видов отбора, более сложных по сравнению с каким-либо произвольным, а в особенности по сравнению с простой случайной выборкой.

5.3.4. Планирование эксперимента в условиях неоднородности

Как уже было сказано, одной из важнейших задач при планировании эксперимента является устранение влияния на его результаты неоднородности условий, в которых он проводится. При организации выборочного наблюдения также можно столкнуться с неоднородностью условий проведения этого наблюдения. Какие же средства применяются в экспериментальной работе для преодоления последствий этих условий?

Одним из таких средств является уже упоминавшийся «парный» метод. Идея этого метода заключается в том, что при планировании эксперимента вовсе не требуется, чтобы все его повторения осуществлялись в сходных условиях. Размещать их в различных условиях даже выгодно: вариация, обусловленная этими различиями, будет исключена и это уменьшит ошибку опыта. Важно обеспечить сходство условий для сравниваемых вариантов опыта лишь в пределах каждой их пары. Таким образом, экспериментатор получает ряд разностей, происхождение которых почти свободно от всяких посторонних влияний и которые вызваны лишь действием испытываемого в опыте фактора. Но при повторении опыта они могут образоваться всякий раз в отличных друг от друга условиях. Различия в этих условиях создаются уже не действием испытываемого фактора, а различными посторонними по отношению к этому действию влияниями.

Вариация этих разностей может быть основанием для определения величины случайной ошибки опыта, но только в том случае, когда осуществлена рандомизация размещения вариантов опыта в каждой их паре. В противном случае получится систематическое искажение этой ошибки. Например, в эксперименте с кроликами сравниваемые пары составлялись так, что в одну группу отбирались крупные кролики, в другую только мелкие. Разумеется, при этих условиях в окончательный результат была введена систематическая ошибка. Любое расположение вариантов опыта, при котором просматривается какая-либо система, недопустимо.

При организации опыта желательно, чтобы все число пар или повторений разделялось на две равные части, из которых одна содержала бы «лучший» член каждой пары – вариант А, а вторая – вариант Б. В каких именно парах той и другой части повторений «лучшие» члены будут содержать тот или иной вариант, должен решить жребий. Именно здесь организация опыта следует принципу рандомизации, ограничение которого состоит в том, что размещение пар или повторений уже не рандомизировано.

Парный метод находит аналогию в стратифицированном выборочном наблюдении. Оба этих метода направлены на то, чтобы получить надежную, несмещенную оценку результатов исследования, протекающего в условиях неоднородности. Парный метод рассчитан на сравнение лишь двух вариантов эксперимента или на испытание действия лишь одного фактора. При увеличении числа сравниваемых вариантов или при испытании не одного, а нескольких факторов, «пары» вариантов должны заменяться блоками. Блоки должны формироваться так, чтобы варианты опыта или способы обработки, составляющие блок, находились бы по возможности в однородных условиях. Вариация, обусловленная различиями между блоками, подлежит устранению точно так же, как вариация, обусловленная повторениями при парном методе. Составляя блоки так, чтобы входящие в них испытываемые варианты находились по возможности в сходных условиях, можно добиться, чтобы различия между средними значениями вариантов опыта были независимы от различий между блоками и порядок их расположения внутри блоков исключал возможность систематического получения лучших результатов для некоторых из них. В табл. 5.9 приведен пример нежелательного расположения опытных делянок в поле при изменении его плодородия вдоль блоков, при котором создаются систематически ухудшающиеся условия для вариантов Б, В и Г по сравнению с А.

Таблица 5.9

Нежелательное расположение вариантов опыта

Блок	Варианты			
1	А	Б	В	Г
2	А	Б	В	Г
3	А	Б	В	Г
4	А	Б	В	Г
5	А	Б	В	Г

Такое положение вещей чревато смещением оценки результатов опыта, и поэтому необходимо провести рандомизацию размещения вариантов внутри блоков. Такой план эксперимента называется «случайными блоками» и ничем, практически, не отличается от парного метода. Статистическая обработка данных опыта, запланированного как эксперимент в случайных блоках, остается такой же, как и при парном методе, с той только разницей, что в дисперсионной модели вариация, обусловленная различиями между блоками, соответствует вариации, обусловленной повторениями опыта при парном методе. В табл. 5.10 приведен пример опыта в случайных блоках.

Таблица 5.10

Опыт в случайных блоках

Блок	Варианты			
1	Г/22	Б/18	А/16	В/28
2	Б/22	В/26	А/20	Г/30
3	Б/20	Б/20	Г/28	А/22
4	В/20	В/20	Г/26	Б/26
5	Г/24	Г/24	Б/24	В/24

Порядок планирования эксперимента в блоках аналогичен принципу стратификации при проведении выборочного статистического наблюдения. Как при расчленении исходной совокупности на страты, так и при составлении блоков в эксперименте руководящей остается одна и та же идея: по возможности создать однородные условия внутри групп или блоков и сохранить различия между ними. Между планом проведения эксперимента в блоках и стратифицированной выборкой имеется и существенное различие. Оно состоит в том, что если при стратифицировании выборок вся вариация признака внутри групп рассматривается как случайная, то при проведении эксперимента вариация внутри блоков, вероятно, обусловлена не только влиянием случайных причин, но и действием самих испытываемых в опыте факторов или его вариантов.

Планирование опыта в случайных блоках предусматривает группировку вариантов опыта лишь по одному признаку – по блокам. Проведение опыта в этой форме вполне оправдывает себя, если значительная доля вариации обусловливается лишь одним источником неоднородности. Но в действительной обстановке эксперимента часто присутствует не один, а несколько таких источников. Например, в условиях полевого опыта плодородие почвы может изменяться не в одном, а в двух направлениях. Несколько источников неоднородности может быть создано самой организацией опыта.

Если имеются два источника неоднородности, то естественно, что опыт должен быть спланирован так, чтобы исключить из статистической оценки его результатов вариацию, обусловленную обоими источниками. Со статистической точки зрения это означает двойную группировку данных опыта. Этому требованию отвечает упомянутый ранее латинский квадрат (в строках и столбцах квадрата каждый вариант опыта должен встречаться лишь один раз). При планировании эксперимента стандартная форма латинского квадрата обычно подвергается рандомизации: как в столбцах, так и в строках квадрата размещению вариантов опыта придается случайная последовательность. Рандомизация размещения вариантов опыта в латинском квадрате необходима по тем же соображениям, по каким она производится при парном методе или в рандомизированных блоках: для устранения преимуществ одних вариантов перед другими, способных привести к систематическому искажению оценки результатов эксперимента.

Главным затруднением в организационном плане при проведении эксперимента методом латинского квадрата является то, что число повторений должно быть равно числу вариантов опыта. При увеличении числа повторений должны увеличиться как по строкам, так и по столбцам состоящие из различных вариантов опыта блоки. Но это может сместить оценку результатов опыта, поскольку увеличение блоков может привести к однородности условий внутри их.

Вариация по строкам и столбцам латинского квадрата может иметь двоякое значение: она может рассматриваться как вариация, источником которой являются условия неоднородности и как результат действия каких-либо изучаемых в опыте факторов. В последнем случае опыт в форме латинского квадрата приобретает черты факторного эксперимента. Примером факторного эксперимента высокого порядка, проводимого по принципу латинского квадрата, является греко-латинский квадрат или латинский квадрат второго порядка. Такой квадрат образуется путем комбинации двух ортогональных латинских квадратов, т.е. таких квадратов, в которых при наложении один на другой каждая пара букв или чисел, обозначающая варианты опыта, встречается только один раз.

A	B	C
B	C	A
C	A	B

1	2	3
3	1	2
2	3	1

A_1	B_2	C_3
B_3	C_1	A_2
C_2	A_3	B_1

Если в опыте, спланированном как греко-латинский квадрат, признаки группировки по строкам и столбцам не являются обозначением факторов, то в таком двухфакторном эксперименте вариация по строкам и по столбцам по-прежнему будет характеризовать влияние условий неоднородности.

5.3.5. Серийный отбор и его компромиссная роль в выборочном наблюдении

При проведении выборочного наблюдения необходимо соблюдать требования уменьшения затрат сил, времени и средств. Такие требования часто реализуются в форме серийного отбора. Например, при изучении производительности труда в качестве единицы отбора может выступить не отдельный работник, а производственная бригада. Такой серийный отбор может применяться при выборочном изучении размеров посевных площадей в личных хозяйствах: отбираются не одиночные хозяйства, а населенные пункты, в которых сплошь обследуются все хозяйства. Населенный пункт или производственная бригада может быть, следовательно, рассматриваться как серия элементарных единиц.

Серийный отбор имеет большие организационные преимущества перед отбором элементарных единиц, заключающиеся в легкости отбора и изучения отобранных единиц. Однако серийный отбор имеет и недостатки: точность его сравнительно с отбором элементарных единиц обычно снижается. Таким образом, исследователь оказывается перед дилеммой: сделать свою работу более легкой и дешевой, снижая ее точность, или стремиться к большей ее точности, хотя бы и удорожая работу. Задача, в конечном итоге, сводится к тому, чтобы подыскать какой-то оптимальный размер единицы отбора.

При серийном отборе изменение точности исследования по сравнению с простой случайной выборкой зависит от того, как поведет себя коэффициент внутриклассовой корреляции, выражающий степень однородности элементов, которые составляют серию как большую единицу отбора. Представим соотношение квадрата ошибки серийного отбора и квадрата ошибки простой случайной выборки элементов при той же общей численности их в сериях равенством

$$\frac{s_a^2}{a}=\frac{s^2}{n}[1+(b-1)\bar{r}] \text{ откуда } \frac{s_a^2/a}{s^2/n}=[1+(b-1)\bar{r}],$$

где a – число серий; b – число элементов в серии.

Это соотношение называют иногда мерой эффекта планирования. Оно может быть больше или меньше единицы в зависимости от знака перед коэффициентом внутриклассовой корреляции. Если этот коэффициент положителен, то эффект планирования превысит единицу. Это значит, что при серийном отборе дисперсия больше, чем при простой случайной выборке элементов, или, иными словами, этот отбор менее точен, чем случайная выборка. При крайнем положительном значении $\bar{r}$ =+1, когда все элементы в сериях одинаковы и источником вариации служат лишь различия между сериями, эффект планирования окажется равным $[1+(b-1)\bar{r}]=b$, т.е. числу элементов, составляющих серию. Это означает, что дисперсия серийных средних будет во столько раз больше дисперсии выборочных средних при простой случайной выборке элементарных единиц, сколько этих единиц входит в состав серий. Этот крайний случай, если бы он имел место в действительности, характеризует более невыгодное положение для серийного отбора, т.е. нужно отказаться от организации серийного отбора, т.к. при полном сходстве элементов, входящих в состав серий, достаточно отобрать по одному элементу из каждой серии.

Отрицательному крайнему значению $\bar{r}=[-\frac{1}{(b-1)}]$ соответствует $\frac{s_a^2/a}{s^2/n}=[1+(b-1)\bar{r}]=0$, т.е. нулевая дисперсия для серийных средних. Это равносильно отсутствию ошибки, что, конечно, невозможно представить на практике.

Отрицательные значения $\bar{r}$ вообще вряд ли могут встретиться при серийном отборе. Элементарные единицы, входящие в состав серий, обычно оказываются похожими друг на друга. Таким образом, можно ожидать, что отношение, характеризующее эффект планирования при серийном отборе в большинстве случаев будет превышать единицу. Оно будет равно единице в случае, когда распределение элементарных единиц среди серий будет полностью рандомизированным и можно считать, что $\bar{r}\approx 0$. Тогда точность серийной выборки не будет отличаться, очевидно, от точности простой случайной выборки элементарных единиц.

При организации серийного отбора может встать проблема, связанная с тем, что отбираемые серии часто бывают неравными. Возможны и такие случаи, когда запланированные равные серии на деле оказываются неравными вследствие неточностей, возникающих в самом процессе отбора.

При отборе с равной вероятностью неравных серий могут быть указаны три способа обработки, позволяющие избегать грубых смещений оценки результатов. Один из способов заключается в том, что надо оперировать не простой, а взвешенной средней из размеров серий. Например, если среднее число жителей на одно жилище в отдельных населенных пунктах представить как y, а число жилищ в этих пунктах как x, то формула вычисляемой средней приобретает вид средне взвешенной: $\bar{y} = \frac{\sum yx}{\sum x}$. Величина дисперсии в этом случае будет зависеть от того, насколько сильно разнятся величины y и x: чем больше они отличаются друг от друга, тем больше и величина дисперсии, а, следовательно, и величина ошибки. Она может быть больше, чем ошибка при невзвешенной средней. Однако в последнем случае, поскольку отбор неравных серий производится с равной вероятностью, оценка результатов оказывается смещенной, в то время как при взвешенной средней эта оценка не имеет систематической ошибки.

Рассмотренные методы отбора неравных серий позволяют устранить или, во всяком случае, уменьшить неблагоприятное влияние этого обстоятельства на оценку результатов исследования. Все они предусматривают отбор серий с равной вероятностью. Однако наиболее эффективный выход следует искать в организации отбора с неравными вероятностями, пропорциональными числу элементарных единиц в серии. Техника этого отбора достаточно проста. Надо представить данные исходной совокупности в виде кумулятивного ряда числа элементарных единиц в сериях. Пользуясь таблицами случайных чисел, можно отобрать намеченное число серий. Так как число элементарных единиц в сериях представлено нарастающим итогом, то неизбежно окажется, что более крупных серий будет отобрано больше, чем мелких. Затем полученные таким образом данные, обрабатываются обычным путем, причем средняя выборочная может быть вычислена как невзвешенная средняя. Это возможно сделать потому, что при отборе серий с вероятностями, пропорциональными числу элементарных единиц в них, происходит действительное взвешивание средней. Ошибка выборки может оказаться и большей, чем при отборе неравных серий с равной вероятностью, но она, в отличие от последней, не будет уже смещена.

По сравнению с простой случайной выборкой отбор неравных серий с пропорциональными им величинами вероятности все же может проиграть в точности результатов. Это зависит от степени корреляционной связи элементарных единиц внутри серий, величины вариации серийных средних и от среднего числа элементарных единиц в серии. На практике вместе с увеличением среднего числа элементарных единиц в серии влияние первых двух факторов уменьшается. Но увеличение среднего числа элементарных единиц в серии само по себе в конечном результате обусловливает существенное увеличение вариации выборочных данных. Практически отбор неравных серий с вероятностями, пропорциональными их размерам, обычно применяется на первой стадии многоступенчатого отбора, об организации которого будет сказано впоследствии.

Отбор серий может быть стратифицированным. Стратификация при серийном отборе даже желательна, к тому же провести ее для серий гораздо проще, чем для элементарных единиц. При стратификации серий с пропорциональным размещением может быть получена большая относительная выгода, чем при стратификации элементарных единиц. Это происходит потому, что отбираемые серии, представляющие известную группировку элементарных единиц, при стратификации как бы подвергаются вторичной группировке, и, таким образом, часть вариации, обусловленная различием средних по стратам и устраняемая при определении случайной ошибки, получает относительно больший вес.

5.3.6. Компромиссы в планировании эксперимента

Как уже подчеркивалось, эксперименты, проводящиеся обычно в условиях неоднородности, важно организовать так, чтобы эта неоднородность сказывалась лишь в различиях между повторениями или блоками, а внутри них создавались бы по возможности условия однородные. На это направлены не только все описанные в предыдущих разделах методы: парный метод, метод рандомизированных блоков, латинские и греко-латинские квадраты, но и многие другие планы.

Экспериментатор, проводящий опыт со многими его вариантами, может столкнуться с невозможностью осуществления этого принципа на деле. Нужно было найти какой-то выход из создавшегося положения. Решение было найдено в различных вариантах: в виде расщепляемых блоков, в форме планов неполных блоков или планов дробных, или планов дробных факторных экспериментов. Цель всех этих методов одна: организовать эксперимент так, чтобы при ограниченных размерах повторений или блоков получить возможность испытывать много вариантов опыта или способов обработки без смещения оценки его результатов.

Эти методы связаны с так называемым «смешением» тех или иных эффектов. Неполный состав блоков затрудняет некоторые сравнения, и экспериментатор вынужден отказываться от изучения некоторых эффектов. В этих условиях перед ним встает дилемма: расширить блоки, сделав эксперимент полноблочным (если это позволяют обстоятельства), но с риском создать условия неоднородности уже внутри блоков и, следовательно, идти заведомо на смещение оценки результатов опыта, либо сделать блоки неполными, но сузить информацию, «смешав» некоторые эффекты с межблоковой вариацией данных. Если экспериментатор выбирает второй путь, то он должен по крайней мере пожертвовать такой информацией, которая менее важна. Обычно это не главные эффекты, а эффекты каких-либо взаимодействий. Экспериментатор при планировании своего эксперимента идет здесь на определенный компромисс.

Стремление отыскать компромисс между необходимостью обеспечить однородность условий для каждого повторения опыта и желанием увеличить число испытываемых в опыте вариантов проявляется и при планировании эксперимента в форме неполных блоков. Идея такой организации эксперимента очень проста. Только часть изучаемых в опыте вариантов включается в каждое повторение или в каждый блок так, чтобы не сделать его слишком большим и не нарушить однородность условий внутри него и в то же время иметь возможность исключить влияние различий между блоками при сравнении вариантов опыта между собой. Именно потому, что в каждое повторение опыта или в каждый блок входит неполный набор вариантов опыта, такой план опыта и называется «неполноблочным». Неполноблочный план опыта может быть как сбалансированным, предусматривающим появление пар вариантов опыта одинаковое число раз, так и частично сбалансированным, в котором это условие нарушается. Если исследователь иногда обращается к плану неполноблочного частично сбалансированного опыта, то это обусловливается тем, что для сбалансирования опыта при большом числе вариантов опыта, размещаемых в блоке, требуется и большое число блоков, что не всегда возможно осуществить на практике.

Для пояснения можно привести пример с опытом, в котором использовались пять способов обработки на пяти партиях сырья, которые выступали в качестве блоков (табл. 5.11).

Таблица 5.11

Схема неполноблочного опыта

Партия сырья (блок)				
1	2	3	4	5
А *В* *С* *Д*	*А* *В* *С* *Е*	*А* *В* *Д* *Е*	*А* *С* *Д* *Е*	*В* *С* *Д* *Е*

Ни один из блоков не охватывает всех способов обработки. План опыта сбалансирован, т.к. каждый способ обработки испытывается одинаковое число раз и таким образом каждая пара вариантов может быть сравниваема также одинаковое число раз: *А* и *В* в 1, 2 и 3 блоках; *А* и *С* в 1, 2 и 4 блоках; *А* и *Д* в 1, 3 и 4 блоках и т.д. Внутри блоков принят случайный порядок расположения вариантов, но в общем этот план также принадлежит к числу планов с ограничением рандомизации. Хотя внутри блоков расположение вариантов опыта носит случайный характер, но сам принцип ограничения сравнений определенным числом раз является таким ограничением рандомизации.

Поскольку в несбалансированном неполноблочном эксперименте не все варианты или испытания появляются в одном и том же блоке, то это приводит к частичному смешиванию эффектов вариантов с эффектами блоков. Поэтому при статистической оценке результатов эксперимента как эффект вариантов, так и эффект блоков обычно подвергается корректированию. Корректирование проводится с применением метода наименьших квадратов, предусматривающего решение ряда нормальных уравнений.

К планам эксперимента, преследующим одни и те же цели, что и план неполных блоков, относится и план дробного факторного эксперимента. Сущность его заключается в том, что в повторениях опыта представляется не весь эксперимент, а только известная его часть. В этом обстоятельстве выражается та же идея компромисса.

5.3.7. Некоторые более сложные формы организации отбора

Имеются другие пути организации выборочного наблюдения, лучше отвечающие характеру изучаемого материала, и при использовании которых может даже улучшиться баланс между точностью наблюдения и затратами времени и труда. Одним из таких путей является многоступенчатый отбор, предполагающий подвыборку, которая заключается в отборе более мелких единиц из уже отобранных крупных. Многоступенчатая выборка может состоять из нескольких ступеней, но чаще всего встречается двухступенчатая ее форма. Широкое применение на практике этой формы выборки обусловлено ее большой гибкостью, позволяющей упрощать весь процесс организации наблюдения. Обычно объект и круг обследования определяются уже на первой ступени, облегчая таким образом отбор на следующих ступенях. Данные, получаемые относительно первичных более крупных единиц, могут содержать в себе информацию и об элементарных единицах.

В общем организация двухступенчатой выборки окажется более удобной, чем отбор такого же числа элементарных единиц. Конечно, легче установить список отбираемых единиц (если выборка носит характер систематического отбора) или намечать порядок их отбора при помощи таблицы случайных чисел (если выборка случайная) из сравнительно небольшого числа первичных крупных единиц, чем из большой исходной совокупности. Это немаловажное обстоятельство и побуждает часто исследователя прибегать к двухступенчатой выборке вместо прямого многочисленного отбора элементарных единиц.

Но эта легкость организации процесса отбора сопровождается, как правило, увеличением ошибки выборки, т.к. последняя состоит уже из двух слагаемых: к ошибке, получаемой при отборе крупных единиц на первой ступени, присоединяется ошибка, которой сопровождается отбор элементарных единиц на второй ступени выборки. Если, например, на первой ступени было отобрано n больших единиц, содержащих каждая m элементарных единиц, то при определении ошибки выборочной средней, вычисленной для признака y, которым обладают элементарные единицы и которая оценивается, необходимо определить как дисперсию, характеризующую вариацию этого признака между первичными большими единицами, так и среднюю величину дисперсий внутри этих больших единиц.

Ошибка двухступенчатой выборки должна быть больше ошибки простой случайной выборки. Ее величина зависит от соотношения между n и m, т.е. от числа отбираемых единиц на первой и второй ступенях. Увеличение числа и тех и других сокращает величину ошибки.

При одинаковом числе элементарных единиц, входящих в состав первичных единиц, двухступенчатая выборка оказывается более точной, чем серийная. Это происходит вследствие того, что при серийной выборке элементарные единицы, составляющие серию, оказываются, как правило, связанными положительной корреляцией. При этом иногда такая внутрисерийная корреляция очень тесна. При двухступенчатом отборе связь между элементарными единицами внутри больших первичных единиц уже нарушается или значительно ослабевает, т.к отбирается только часть их, причем обычно в случайном порядке. Чем меньшую долю составляют отобранные элементарные единицы в отобранных больших первичных единицах, тем больше оснований ожидать, что результаты выборки будут более точными. Это можно объяснить тем, что при сравнительно небольшом числе элементарных единиц, отбираемых в случайном порядке и поэтому разбросанных по всей большой единице, степень сходства между ними должна ослабевать.

Отличительным признаком многоступенчатой выборки является то, что на разных ступенях ее отбираются единицы разного типа или порядка: сначала, например, отбираются фермы, а затем на них – известное количество голов скота.

Если при многоступенчатой выборке речь идет о подвыборке, которая извлекается из полученной предварительно более многочисленной выборочной совокупности, то существует и такая форма организации выборочного наблюдения, которая предполагает расчленение общей совокупности отбираемых единиц на несколько независимых друг от друга выборок с применением одного и того же способа отбора. Это так называемые «взаимопроникающие выборки». Для такой формы организации выборочного процесса характерно то, что существует возможность уточнения ошибки путем сравнения результатов таких взаимопроникающих выборок. Если не одна, а несколько выборок приведут к сходным результатам, то это будет свидетельствовать, что истинные величины и соотношения репрезентированы достаточно точно. Но для того, чтобы такая возможность сравнения действительно существовала, необходима полная идентичность условий, в которых производятся эти выборки. Различия в характере объектов отбора или в среде, в которой эти объекты находятся, сделают несопоставимыми итоги работы. Нужно, чтобы выборочные совокупности, образованные как взаимопроникающие выборки, действительно представляли собой как бы извлечения из одной общей совокупности.

Для вычисления общей ошибки подходящим является использование дисперсионного анализа, в результате которого необходимо определить три суммы квадратов отклонений:

1) общую, составляющую отклонения данных всех взаимопроникающих выборок от их общей средней;

2) сумму квадратов отклонений средних значений признака в отдельных взаимопроникающих выборках также от их общей средней;

3) остаточную, которая может быть получена как разность между первыми двумя.

Вторая из этих сумм представит ту вариацию, которая обусловлена различиями между взаимопроникающими выборками. Эта вариация всегда существует, и она, конечно же, при анализе данных должна быть исключена, и тогда величина случайной ошибки определится уже на основе вариации данных внутри выборок, выраженной в таблице дисперсионного анализа как остаточная сумма квадратов отклонений. Сравнение дисперсий, вычисленных на основе двух последних сумм квадратов отклонений, может показать, при использовании критерия Фишера, можно ли игнорировать различия между результатами отдельных выборок. Например, если остаточная дисперсия значительно превосходит дисперсию выборочных средних, то нет оснований говорить о существенности различий в результатах, полученных в выборках и опасаться смещения оценок этих результатов.

5.3.8. Отбор, не основанный на законах случая

Как при организации выборочного наблюдения, так и при планировании эксперимента исследователь должен считаться с законами, управляющими случаем, иначе он не сможет правильно оценить результаты своей работы. Однако существует такая форма выборочного статистического наблюдения, когда исследователь полностью отказывается следовать тем принципам, на которых покоится вероятностная теория выборочного метода и эксперимента. Это – систематический отбор.

Выборка при систематическом отборе состоит из n единиц с порядковыми номерами $i, i+k, i+2k, \ldots, i+(n-1)k$, где k – интервал, через который отбираются эти единицы из исходной совокупности. Так как этот интервал остается постоянным для всей выборки, то систематический отбор обычно называется механическим. Широкая распространенность систематического отбора сравнительно со случайной выборкой обусловливается большей простотой и гибкостью его применения. Проще отобрать через определенный интервал единицы из общего списка, нежели подвергать жеребьевке члены исходной совокупности, нередко очень большой, или же обращаться к таблицам случайных чисел.

Систематический отбор действительно приводит к более точным результатам по сравнению со случайной выборкой в смысле приближения выборочной средней к средней исходной совокупности. Но преимущество в точности воспроизведения изучаемой характеристики в исходной совокупности, которое имеет систематический отбор сравнительно со случайной выборкой, может сохраняться лишь при том условии, что в организации этого отбора не содержится моментов, способствующих смещению как самой выборки, так и оценки ее результатов.

При систематической выборке, помимо проблемы организации, возникает и другой очень важный вопрос. Он связан с возможностью вычисления самой случайной ошибки выборки, величина которой и является для исследователя показателем точности отбора. На практике эту задачу решают довольно легко, допуская вычисление ошибки систематического отбора как ошибки простой случайной выборки, т.е. систематический отбор рассматривается как одна из форм стратифицированной выборки. Разделение совокупности, из которой производится отбор, на равные части по числу отбираемых единиц является операцией, подобной расчленению этой совокупности на группы единиц. Но при вычислении ошибки систематического отбора невозможно поступить так же, как и при стратифицированном отборе, т.е. определить величину этой ошибки на основе средней величины внутригрупповых дисперсий. Для этого потребуется отбирать по крайней мере две единицы из каждой группы-страты. Но сама организация систематического отбора исключает это условие, следовательно, вычисление ошибки систематического отбора как ошибки простой случайной выборки неправомерно: ошибка, вычисленная таким способом, обязательно приведет к смещению оценки результатов отбора. Она может оказаться больше или меньше действительной величины ошибки. Если состав совокупностей систематического отбора таков, что между ее членами обнаруживается положительная корреляционная связь (внутригрупповая или внутриклассовая), то ошибка, вычисленная по формуле случайного отбора, уменьшит действительную величину ошибки. В противном случае (коэффициент внутриклассовой корреляции отрицателен) ошибка, вычисленная по методу случайного отбора, будет преувеличена.

Может показаться, что преувеличение действительной величины ошибки является положительным фактом, поскольку лучше преувеличить возможные неприятности, чем преуменьшить их. Но это не так. В границах установленных доверительных интервалов, на основе которых исследователь будет распространять полученные результаты на всю исходную совокупность, могут быть допущены значительные просчеты, а следовательно искажены результаты статистического исследования.

Выходом из этого положения является рандомизация исходной совокупности. Если речь идет об ошибке систематического отбора из полностью рандомизированной бесконечно большой совокупности, в которой отсутствует какой-либо тренд и корреляция между ее членами, то в среднем для всех возможных выборок величина этой ошибки равна величине ошибки случайного отбора.

Иногда эту проблему можно решить следующим образом. Если в систематической совокупности наблюдается линейный тренд (объекты в совокупности располагаются в убывающем порядке значений признака), то репрезентативность средней величины совокупности обеспечивается отбором единиц из центра интервала. Возможно также прибегать и к корректированию концов ряда, которое заключается в том, что при вычислении выборочной средней первому и последнему членам придается вес согласно формуле $1 \pm \frac{n(2i-k-1)}{2(n-1)k}$, где n – численность выборки; I – порядковый номер объекта, отбираемого в интервале; k – интервал. Корректирование концов ряда может существенно приблизить возможные выборочные средние к средней исходной совокупности.

Существует еще способ приближения величины вычисляемой ошибки систематического отбора к ее действительной величине. Если в ряду изучаемых данных наблюдается линейная тенденция, то определенная польза может быть получена при применении «модели последовательных разностей». Суть этого метода заключается в том, что ошибка систематического отбора определяется на основе суммы квадратов последовательных разностей $(y_i - y_{i+k}),[y_{i+k} - y_{i+2k}]$ и т.д., т.е. суммы квадратов первых разностей. Квадрат ошибки или дисперсии выборочных средних вычисляются следующим образом:

$$m^2_{(\bar{y})} = \frac{N-n}{Nn} \frac{\sum [y_i - y_{i+k}]^2}{2(n-1)}.$$

Смысл предлагаемой модели состоит в том, что исчисление первых разностей означает исключение автокорреляции или линейного тренда, которым обусловливается возникновение этой автокорреляции. Таким образом, можно считать, что вычисление ошибки основывается только на случайном компоненте вариации данных. Результаты, получаемые по этой формуле, не свободны от смещения: все же остается известная связь между соседними разностями; кроме того, первый и последний члены ряда имеют сравнительно малый вес в определении случайного компонента.

В заключение можно сказать, что если систематический отбор при принятых мерах против смещения выборки в случаях наличия каких-либо трендов в исходной совокупности может быть с успехом применяем для уточнения величины выборочной средней, приближения ее к величине средней в исходной совокупности, то о точном определении случайной ошибки выборки не может идти речи, поскольку отбор не опирается на законы случая.

Контрольные вопросы

1. Различие между краткосрочным и среднесрочным прогнозированием. Методы прогнозирования.
2. Понятие стационарного ряда.
3. Преимущества экспоненциально взвешенного среднего перед скользящим средним.
4. Понятие тренда, характер тренда, типы трендов.
5. Меры точности прогноза.
6. В каких случаях возникает необходимость в низкочувствительном прогнозе?
7. Автокорреляция. Предел изменения коэффициента автокорреляции.
8. Какой характер будет иметь тренд, если его автокоррелограмма содержит большое число максимальных и минимальных значений коэффициентов автокорреляции?
9. Наиболее распространенные способы сведения кривой к линейному тренду.
10. Отличие гиперболических кривых I, II и III типа.
11. Влияет ли на случайную ошибку выборки при стратифицированном отборе межгрупповая дисперсия данных?
12. Основные принципы стратификации. Всегда ли возможно организовать пропорциональный стратифицированный отбор?
13. Преимущества и недостатки серийного отбора в выборочном наблюдении.
14. От чего зависит ошибка двухступенчатой выборки?
15. Способы уменьшения ошибки систематического отбора.

Глава 6
ЭЛЕМЕНТЫ МЕХАНИЗМА ЭКОЛОГО-ЭКОНОМИЧЕСКОГО ПРИРОДОПОЛЬЗОВАНИЯ С ПОЗИЦИЙ СИСТЕМНОГО АНАЛИЗА

6.1. Концепция устойчивого развития

Взаимодействие общества и природы синтезирует элементы экономической, социальной и природной систем. Широкое применение в настоящее время получил термин «рациональное природопользование», под которым обычно понимается плановая организация экономики, предусматривающая снятие конфликтов между развитием производительных сил и состоянием природной среды (Жариков, 1993).

В нормальном состоянии геосистема находится в динамическом равновесии, где сохраняется баланс вещества и энергии, что обеспечивается биологическим круговоротом, включающим создание органического вещества и его разрушение по принципу безотходных технологий. В природных системах эти процессы сбалансированы, а количество органического вещества, участвующего в круговороте, относительно постоянно. Влияние хозяйственной деятельности на сбалансированность природных систем происходит двояко: во-первых, уменьшается емкость биологического круговорота за счет изъятия биологической продукции и запасов органического вещества; во-вторых, за счет неуклонного роста отходов производства в круговорот поступают вещества, не имеющие аналогов в природе, что ведет к нарушению сложившихся трофических цепей и, в конечном итоге, к нарушению состояния равновесия. По мере роста хозяйственной деятельности проблема равновесия все более обостряется, а по некоторым параметрам уже достигла критических значений (примером может служить парниковый эффект или уменьшение озонового слоя). Ясно, что социально-экономическое развитие общества может осуществляться только в границах экологической емкости геосистем, выход за которые чреват экологическими катастрофами. К тому же беда заключается в том, что по отдельным видам загрязнения, не говоря уже о комплексных видовых воздействиях, границы устойчивости природной среды неизвестны.

Разрешение противоречий в системе общество-природа, достижение устойчивого развития общества – сложная комплексная задача, требующая применения междисциплинарного подхода, синтеза знаний разных наук и использования практического опыта.

Устойчивое развитие – это такое развитие, которое удовлетворяет потребности настоящего времени, но не ставит под угрозу способность будущих поколений удовлетворять свои собственные потребности. Оно содержит два ключевых понятия:

- понятие потребностей, в частности потребностей, необходимых для существования беднейших слоев населения, которые должны быть предметом первостепенного приоритета;
- понятие ограничений, обусловленных состоянием технологии и организацией общества, накладываемых на способность окружающей среды удовлетворять нынешние и будущие потребности.

Имеются и более краткие определения устойчивого развития, отражающие его важнейшие экономические аспекты. Среди таких определений можно выделить следующие:

- развитие, которое не возлагает дополнительные затраты на следующие поколения;
- развитие, которое обеспечивает постоянное простое и/или расширенное воспроизводство производственного потенциала на перспективу;
- развитие, при котором человечеству необходимо жить только на проценты с природного капитала, не затрагивая его самого.

Приведенные выше определения устойчивого развития можно рассматривать и сквозь призму экономических отношений поколений: внутри современного поколения (в частности, социальный аспект, проблема бедности) и между поколениями (эколого-экономический аспект). Таким образом, задачи экономического и социального развития должны быть определены с учетом его устойчивости, соответствия экологическому императиву во всех странах – развитых и развивающихся, странах с рыночной и другими видами экономики.

Можно выделить четыре критерия устойчивого развития на длительную перспективу, которые основываются на классификации природных ресурсов и динамике их воспроизводства:

- количество возобновимых природных ресурсов (земля, лес и пр.) или их возможность продуцировать биомассу должны, по крайней мере, не уменьшаться с течением времени, т.е. должен быть обеспечен режим простого воспроизводства. Например, для земельных ресурсов это означает сохранение площади наиболее ценных сельскохозяйственных угодий или в случае уменьшения их площади сохранение/увеличение уровня производства продукции земледелия, кормового потенциала земель для сельскохозяйственных животных и т.д.;
- максимально возможное замедление темпов исчерпания запасов невозобновимых природных ресурсов с перспективой в будущем их замены на другие нелимитированные виды ресурсов (например, частичная замена нефти, газа, угля на альтернативные источники энергии – солнечную, ветровую и пр.);
- возможность минимизации отходов на основе внедрения малоотходных, ресурсосберегающих технологий;

• загрязнение окружающей среды (как суммарное, так и по видам) в перспективе не должно превышать его современный уровень. Возможность минимизации загрязнения до социально и экономически приемлемого уровня («нулевого» загрязнения ожидать нереально).

Эти четыре критерия (их может быть и больше) должны быть учтены в процессе разработки концепции устойчивого развития. Их учет позволит сохранить окружающую среду для следующих поколений и не ухудшит экологические условия проживания. Среди экономических показателей эффективными критериями устойчивого развития является уменьшение трудоемкости экономики и структурный показатель, отражающий уменьшение удельного веса продукции и инвестиций отраслей природоэксплуатирующих секторов. Переход к устойчивому развитию предполагает ограничение потребностей в товарах и услугах, в отличие от техногенного развития с его максимизацией потребления.

В самом общем виде устойчивое развитие во времени с учетом основных параметров можно представить как

$$F_t(L,K,P,I) \le F_{t+1}(L,K,P,I)\,, \tag{6.1}$$

где $F_t(L,K,P,I)$ – функция устойчивого развития, L – трудовые ресурсы, K – искусственно созданный физический капитал, средства производства, P – природные ресурсы, I – институциональный фактор, $T \geq 0$.

В определенной степени функция устойчивого развития (6.1) является «расширением» производственной функции. Однако включенные новые параметры – природные ресурсы и институциональный фактор – принципиальны. Соотношение (6.1) показывает необходимость сохранения и увеличения во времени некоторого агрегатного производственного потенциала, определяемого в основном тремя видами капитала. Здесь природный капитал может уменьшаться до такой степени, пока это уменьшение может быть компенсировано увеличением применения искусственно созданных средств производства (заводы, технологии, дороги и пр.), повышением квалификации работников и т.д. Часто институциональный фактор не рассматривается, однако для устойчивого развития эта составляющая очень важна. Культурные традиции, религия, институты собственности и т.д. оказывают огромное влияние на выбор эколого-экономической политики. Например, в некоторых восточных странах вода считается даром Бога и поэтому нельзя устанавливать на нее цену и плату за ее использование, т.е. нельзя использовать те экономические инструменты, которые являются очевидными для рационального природопользования. Все это делает индивидуальным формирование устойчивого типа развития в каждой стране при сохранении его общих принципов.

Для более детального анализа устойчивого развития используются понятия слабой устойчивости и сильной устойчивости. Сторонники сильной устойчивости занимают жесткую, зачастую «антиэкономическую» позицию по многим вопросам экономического развития, таким как стабилизация или уменьшение масштабов экономического роста, приоритет прямого регулирования, жесткое ограничение потребления и т.д. Сторонники слабой устойчивости предпочитают модифицированный экономический рост с учетом экологического, «зеленого» измерения экономических показателей, широкое использование экономических инструментов (плата за загрязнение и пр.), изменение потребительского поведения и т.д. При всех различиях позиций обе они противостоят техногенной концепции развития, которая базируется на неограниченном развитии свободного рынка, ориентации на чисто экономический рост, эксплуатацию природных ресурсов, вере в бесконечные возможности научно-технического прогресса, максимизации потребления и пр.

6.2. Современная экологическая обстановка и принципы рационального природопользования

В процессе эволюции биосфера всегда обладала определенной степенью устойчивости. Но в связи с бурным развитием промышленности и сельского хозяйства в основных компонентах биосферы стали наблюдаться серьезные изменения.

Экологическая проблема не знает национальных границ и для своего решения требует объединения стран и народов. Биосфера как среда обитания является единой и неделимой и поэтому все развитые страны мира должны использовать свой научно-технический потенциал для защиты мировых естественных источников жизни.

В последние годы 20 столетия мировым сообществом принят ряд новых экологических программ.

Так, Генеральная Ассамблея ООН одобрила Всемирную хартию природы, которая возлагает на все государства планеты ответственность за сохранение природной среды нашей планеты и выработала программу по охране окружающей среды (ЮНЕП). Все национальные и международные программы базируются на одних и тех же общечеловеческих принципах природопользования. Природопользование – это теория и практика воздействия человечества на природную среду в процессе ее хозяйственного использования с учетом экологических последствий. Теория природопользования разрабатывает общие принципы человеческой деятельности, связанной с использованием природных ресурсов и определяемой потребностями хозяйствования. Это научная дисциплина стратегического значения, базирующаяся на следующих принципах:

- рационально размещать все отрасли производства с учетом наличия источников сырья, энергии и трудовых ресурсов, физико-географических факторов и т.д. Например, там, где природные условия наиболее благоприятны для выращивания хлопчатника, целесообразно сразу строить и текстильные предприятия и машиностроительные заводы, производящие оборудование для текстильной промышленности и машины для хлопководства. На территории, где разведаны медные руды, надо наряду с горнодобывающими предприятиями создать электротехническое машиностроение и другие отрасли, потребляющие в больших количествах медь;

- исключить вредное влияние нашей техногенной деятельности на природные ресурсы и окружающую среду. Для этого надо изучить условия региона, выполнить типизацию этих условий по устойчивости к техногенному воздействию и составить прогнозные технологические схемы по перспективам размещения производительных сил. Например, следует полностью исключить строительство гидроэлектростанций на равнинах. Создаваемые на реках водохранилища при этом повышают уровень грунтовых вод на прилегаемой территории. Происходит ее заболачивание; акваторией водохранилища затапливаются плодородные земли, разрушаются берега и проходящие вдоль берега дороги; подмываются растущие по берегам деревья;
- обеспечить воспроизводство используемых ресурсов. Рыболовство, охота должны производиться таким образом, чтобы рыбы и звери успели вновь размножиться и вырасти. На реках, перегороженных плотинами, надо создавать специальные устройства – рыбоходы для прохода рыб на нерест;
- использовать все природные ресурсы комплексно. Так, например, на Алтае впервые созданы комплексные леспромхозы, которые занимаются не только заготовкой деловой древесины, но и переработкой отходов в древесно-стружечные плиты, приготовлением из хвои хвойно-витаминной муки, заготовкой грибов, ягод и орехов, производят восстановление леса на лесосеках, т.е. ведут производство безотходным способом;
- создать здоровую среду обитания для людей и биосферы в целом путем предупреждения ее заражения и ликвидации естественных, существующих в ней вредных компонентов. Например, Колхидская низменность в свое время была рассадником малярии. Посадка там эвкалиптов привела к осушению, а выращивание цитрусовых культур превратило ее в цветущий уголок Земли, прекратилась малярия. Для проведения в жизнь мероприятий по рациональному природопользованию необходимо, чтобы население было знакомо с проблемами природопользования, с возможностями воспроизводства природных ресурсов;
- контролировать и учитывать качество и количество используемых и остающихся природных ресурсов. При строительстве различных промышленных объектов занимаются большие площади плодородных земель, портится и уничтожается растительность, почва, распугиваются

животные, загрязняется почва, воздух и вода. Сократить эти потери, предотвратить загрязнение воды и воздуха можно часто еще на стадии проектирования. Поэтому все руководители предприятий должны предусматривать возможный ущерб, наносимый природной среде.

Для соблюдения перечисленных принципов необходима полная перестройка общества, оно должно умерить свою власть над природой, создать новые международные политические и экономические механизмы, чтобы работать в мире ограниченных ресурсов.

Мировой опыт развития взаимоотношений между обществом и природой однозначно показывает, что основным противоречием в практике регулирования экологических проблем окружающей среды является то, что проблемы, порожденные экономическими причинами, в силу исторически сложившейся практики игнорирования экологических факторов в хозяйственной деятельности в первую очередь решаются административно-законодательными методами.

Одна из причин создавшегося положения – ориентация предприятий на получение максимальной прибыли и обусловленная ею объективная необходимость экономии на всех статьях затрат, в том числе и расходов на ООС. Загрязнение окружающей среды хотя и является производной от хозяйственной деятельности, но лежит вне сферы экономических интересов предприятий, ибо экономическая система, настроенная на бесплатную эксплуатацию природных благ, не заинтересована в пересмотре сложившихся отношений с окружающей средой, по крайней мере, до сих пор, пока они не станут тормозом для ее дальнейшего развития.

Существует ли возможность согласования экономических интересов общества с интересами природными, и если да, то заключается ли она только в соблюдении принципов рационального пользования?

В теоретическом плане решение задач согласования интересов основывается на принципе согласованного оптимума, известного как принцип Парето. Суть подхода состоит в следующем.

Пусть $G=f(x)$ – целевая функция интересов некоторого субъекта, где x – обозначает ситуацию, характеризующуюся параметрами $x=(x_1,x_2,\dots x_n)$, в качестве которых могут выступать различные характеристики действий субъекта. Субъект обладает свободой выбора параметров x, если $x \in X$ (X – некоторая область, в которой может быть реализовано любое значение x). Применительно к рассматриваемому случаю интересы экономической системы (например в лице предпринимателей и т.п.) могут быть выражены функцией $G_1=f_1(x_1,x_2)$, а природной системы (например в лице комитета по природным ресурсам) – $G_2=f_2(x_1,x_2)$, где x_1,x_2 – наборы параметров, управляемые соответственными субъектами.

В реальной жизни экономическая система реализовывала целевую функцию $G_1=f_1(x_1,x_2)$, а природная система – функцию $G_2=f_2(x_1,x_2)$, по принципу «каждому – свое». Учитывая, что министерства и ведомства являлись одновременно и потребителями природных благ, и отвечали за их сохранение, то, в основном, все сводилось только к максимизации функции $G_1=f_1(x_1,x_2)$ при бессознательном и сознательном игнорировании ситуации, определяемой аргументом x_2.

В соответствии со здравым смыслом предприятие, загрязняющее окружающую среду, должно платить за стоимость очистки сбросов (выбросов) до технически достигнутого уровня, за остаточный ущерб от загрязнения, а также за использование ассимиляционного потенциала экосистемы.

Другими словами, должен соблюдаться принцип эквивалентного обмена. Тогда задача согласования интересов экономической и природной систем сводится к аналитической игре с целевыми функциями

$$G_1=f_1(x_1,x_2) - \max, \tag{6.2}$$

$$G_2=f_2(x_1,x_2) - \min \tag{6.3}$$

и ограничением

$$g=f_1(x_1,x_2)=0. \tag{6.4}$$

Реализация принципа эквивалентного обмена может быть достигнута, если в качестве ограничения принять

$$C_1 \cdot x_1 - C_2 \cdot x_2 = 0,$$

где C_1 – вектор цен на блага, производимые экономической системой,
C_2 – вектор цен благ природной среды.

Решение этой задачи (6.2–6.3) может быть получено в виде линии согласованного оптимума, так называемого геометрического места точек оптимума, соответствующих вариациям ограничения (6.4). Единственное решение дает точка пересечения линии согласованного оптимума с линией ограничений.

Таким образом, теоретическая задача согласования интересов имеет единственное решение. Однако, как показывает практика, реализация рассмотренного подхода в реальном регулировании взаимоотношений общество-природа далека от совершенства. Одна из причин состоит в том, что природные блага (минерально-сырьевые, лесные, водные ресурсы, атмосферный воздух и др.) обладают потребительскими стоимостями для экономической системы, но не имеют цены. Следовательно, на старте развития общество имело и сейчас еще имеет бесплатный капитал в форме экологической устойчивости окружающей среды, причем самовоспроизводящийся, хотя и с определенными сбоями (леса, атмосферный воздух и др. уже требуют экологической защиты), который в принципе может быть подсчитан как стоимость воспроизводства природных ресурсов и ассимиляционной емкости.

6.3. Нормирование качества природной среды

В общем случае проблема соизмерения существующих темпов загрязнения окружающей среды с ее возможностями самоочищения должна решаться на основе оценки способности природных систем к динамическому накоплению и разрушению загрязняющих веществ с сохранением основных природных характеристик. Это приводит к необходимости решения задач экологического нормирования антропогенной нагрузки и определению экологического резерва (Жариков, 1993).

К настоящему времени в России разработаны и введены предельно-допустимые концентрации (ПДК) для нескольких сотен загрязняющих веществ, тем не менее экологическая ситуация у нас одна из худших в мире. И не из-за того, что ПДК занижены или завышены, а из-за бытовавшего упрощенного отношения к самой проблеме загрязнения и путям ее решения. Ведь совершенно ясно, что на выходе из источника загрязнения никакая логика ПДК не работает, и очевидно, что она может быть реализована только в границах распространения загрязняющих веществ, т.е. на уровне экосистемы. Локальный подход не позволил своевременно среагировать на начавшиеся глобальные процессы деградации окружающей среды за счет перехода количественных изменений в качественные. Осознание качественных изменений привело к пониманию необходимости построения экологической политики на уровне экосистем.

На современном этапе, с начала 80-х годов прошлого столетия, задача сохранения природной основы социально-экономического воспроизводства в границах экологической емкости геосистем решается путем разработки нормативов, регламентирующих выбросы в атмосферу и водную среду, так называемых предельно-допустимых выбросов (ПДВ) и предельно-допустимых сбросов (ПДС).

ПДВ – научно-технический норматив, рассчитываемый как предельно-эмиссионная величина, предусматривающая, что концентрация загрязняющих веществ в приземном слое воздуха от всех источников загрязнения не превышает нормативной концентрации этих веществ.

Под ПДС в водный объект понимается предельно-допустимая масса веществ в отводимых сточных водах при условии обеспечения качества воды в контрольном пункте. Согласно методическим основам расчета ПДС, их целесообразно устанавливать по бассейновому принципу с учетом ПДК веществ в местах водопользования, ассимилирующей способности водного объекта и оптимального распределения массы сбрасываемых веществ между водопользователями, сбрасывающими сточные воды. В случаях, когда концентрация загрязняющих веществ в контрольных точках превышает ПДК и невозможно снижение ПДВ (ПДС), устанавливаются временные согласованные выбросы, как бы узаконивающие права предприятий загрязнять окружающую среду. При сбросе загрязненных вод в прибрежные воды морей ПДС устанавливается дифференцированно на каждом выпуске сточных вод.

Несмотря на то, что данный подход рассматривается как экосистемный, тем не менее он еще далек от совершенства. Во-первых, в формулы расчетов ПДВ (ПДС) в основном заложены только условия рассеивания загрязняющих веществ. Во-вторых, оговорка, что нормы выбросов загрязняющих веществ устанавливаются с учетом ассимилирующей способности природной среды, относится скорее к желанию, чем к действительности, в силу ряда причин. Имеющееся решение на уровне контроля ПДК веществ в природной среде, хотя и в расширенном варианте за счет введения ПДВ (ПДС), крайне недостаточно, ибо происходит аккумуляция загрязняющих веществ в экосистемах, время выведения которых достаточно дифференцированно и может составлять десятки лет. Особенно это относится к полузамкнутым водным объектам, для которых, в отличие от водных объектов с проточными водами, недостаточно отработан механизм установления ПДС. Так, для акваторий морских заливов сброс сточных вод с концентрациями, удовлетворяющими лимитированным показателям (в частности ПДК), не гарантирует экологической безопасности из-за процессов накопления токсичных веществ в мелководной зоне, а также в животных и растениях. Большой градиент плотности воды на границе зон мелководье – мористая часть препятствует оттоку загрязняющих веществ из прибрежной зоны и способствует их аккумуляции в зоне геохимического барьера.

Утрата экосистемой естественной способности к самоочищению приводит к ее деградации, и не может быть и речи об ее использовании в качестве приемника антропогенного загрязнения. Следовательно, при разработке ПДВ (ПДС) недостаточно соблюдения норм ПДК, а необходимо исходить из самоочищающей способности экосистем воздушного и водного бассейнов.

Для решения этой проблемы необходимы экспериментальные и теоретические разработки, необходим системный подход. Наиболее существенные результаты в решении данной проблемы получены в морской экологии, где разработана концепция ассимиляционной емкости, которая позволяет подойти к решению задач нормирования антропогенной нагрузки на морские экосистемы и к прогнозу экологической ситуации (Израэль, 1983,1985).

По определению, A_J характеризует ассимиляционную емкость акватории – максимальную динамическую вместимость экосистемой такого количества j-го загрязняющего вещества, которое за единицу времени может быть накоплено, разрушено, трансформировано и выведено за счет процессов седиментации, диффузии или любого другого переноса за пределы экосистемы без нарушения ее нормального функционирования. Ее расчетная величина для j-го загрязняющего вещества определяется из соотношения

$$A_j = k_j \cdot c_j \cdot V / \tau_j, \qquad (6.5)$$

где V – объем морской воды в акватории, τ_j – время пребывания j-го загрязняющего вещества в морской воде, k_j – коэффициент запаса, характеризующий возможности акватории по j-му загрязняющему веществу, c_j – пороговая концентрация j-го загрязняющего вещества.

Коэффициент запаса характеризует возможности акватории по приему j-го загрязняющего вещества без превышения и нарушения ее ассимиляционной емкости, а его пороговая концентрация – предельную дозу для массовых видов морских организмов, превышение которой ведет к их гибели.

Состояние морской среды считается благополучным, если

$$F_j(t) \le A_j \text{ и } u_j \le k_j \cdot c_j, \qquad (6.6)$$

относительно благополучным, если

$$F_j(t) > A_j \text{ и } u_j \le k_j \cdot c_j, \qquad (6.7)$$

неблагополучным, если

$$F_j(t) > A_j \text{ и } u_j > k_j \cdot c_j, \qquad (6.8)$$

где $F_j(t)$ – общий сток j-го загрязняющего вещества в акваторию в t промежуток времени, u_j – концентрация j-го загрязняющего вещества в морской воде.

Задача определения ассимиляционной способности экосистем довольно емкая, требующая постановки продолжительных по времени опытов, анализа огромного эмпирического материала и т.п. Однако только ее решение открывает возможность изучения пределов использования экосистем в качестве приемников загрязнения.

При наличии количественных оценок способности экосистем к самоочищению задача поддержания качества природной среды решается на основе анализа баланса динамики уровней загрязнения окружающей среды. В морской экологии для этих целей используется уравнение

$$\Delta A_j = \sum A_{j1} - \sum A_{jd}, \qquad (6.9)$$

где A_{j1} – приток в акваторию j-го загрязняющего вещества по 1-му каналу поступления (с речным стоком, атмосферными осадками и т.д.), A_{jd} – отток из акватории j-го загрязняющего вещества по d-му каналу (за счет водообмена, биохимического разложения, перехода в атмосферу и т.п.).

Аналогичное уравнение может быть использовано и при получении оценок для воздушной среды, отличие только во времени осреднения при расчете ассимиляционной емкости, которое практически равно нулю.

6.4. Размещение производств с учетом экологических ограничений

Интегральный показатель экологической устойчивости экосистемы (показатель ассимиляционной емкости) – это своего рода индикатор. Уменьшение его базового значения указывает на ухудшение экологической ситуации в природной системе. С другой стороны, это основа прогнозирования экологической ситуации, без анализа которой нереально оптимальное размещение производительных сил с позиций комплексного системного подхода. Существующая практика размещения производительных сил, игнорируя учет экологического фактора, противоречит региональным интересам, что выражается в чрезмерной концентрации производства и населения в ограниченном числе городов и агломераций, которые на сегодняшний день сосредоточили в себе основной узел экологических, социальных, а через них и экономических проблем. Поэтому сегодня очень важно, чтобы вопросы сохранения окружающей среды рассматривались в качестве одного из важнейших факторов, обуславливающих необходимость перераспределения производительных сил между различными по уровню промышленного развития регионами (Жариков, 1993).

Зарубежный опыт структурной перестройки экономики показывает эффективность использования программных методов, в которых природоохранные мероприятия являются частью общего плана социально-экономического развития. Хорошо зарекомендовали себя и комплексные природоохранные программы, а также программы, направленные на решение конкретных экологических проблем. Наибольший эффект от использования нормативных методов в регулировании проблем охраны окружающей среды достигается при их реализации через национальные экологические программы, когда экологические проблемы становятся неотъемлемой частью жизни каждого общества.

Программы разных уровней должны отличаться по целям и методам их реализации. Например, эффективность региональных природоохранных программ обуславливается их конкретной направленностью, механизмом поэтапного снижения загрязнения, точным расчетом необходимых и в то же время допустимых капитальных вложений с позиций их реализуемости и устойчивого развития производства. Для этих целей рассчитывается уровень загрязнения при сложившихся и прогнозируемых темпах развития, который затем сравнивается с нормативным.

Разработка программы заканчивается оптимальным отбором мероприятий, обеспеченных материальными, финансовыми и трудовыми ресурсами, направленных на достижение запланированных целей.

Реализация программ неразрывно связана с территориальной организацией экономики, единой системой, включающей социально-экономические, экологические и прочие особенности. На территории (экосистеме) осуществляется производственная деятельность предприятий, являющихся одновременно источниками техногенного воздействия на компоненты окружающей среды в определенных пространственных зонах.

Исходя из концепции устойчивости для сохранения видового разнообразия флоры, фауны и в конечном итоге выживания общества необходимо соблюдение условий устойчивости для каждой из иерархически организованных природных систем, рассчитываемых на основе их ассимиляционных емкостей. Другими словами, должно быть соблюдено условие:

$$\sum_{i \in j} \text{ÏÄÂ}_{ij} \, (\text{ÏÄÑ}_{ij}) \le A_j \, , \qquad (6.10)$$

где $ПДВ_{ij}$ ($ПДС_{ij}$) – предельно допустимый выброс (сброс) j-го загрязняющего вещества i-м предприятием в экосистему.

По мере улучшения экологической ситуации в экосистемах значения ПДВ (ПДС) должны корректироваться в сторону уменьшения, с тем чтобы в перспективе полностью отказаться от эксплуатации естественной способности природных систем к самоочищению.

Выработка стратегии структурной перестройки экономики, в том числе и под влиянием экологического фактора, требует анализа большой совокупности различных вариантов, в том числе и взаимоисключающих друг друга. Процесс их отбора на основе системного анализа значительно упрощается при использовании экономико-математических моделей. Решение оптимизационной задачи на уровне отдельной экосистемы обеспечивает учет эффектов комплексности развития объектов, рационализации схем территориального и целевого использования ресурсов, структуры конечного потребления при соблюдении налаженной системы ограничений.

Не рассматривая здесь решения других сложных комплексных проблем, связанных с улучшением экологической ситуации, остановимся на возможности применения экономико-математической модели для установления ассимиляционных стандартов загрязняющих веществ предприятием исходя из потенциальной способности природной системы к самоочищению. Процедуру распределения между предприятиями объемов квот сбросов загрязняющих веществ рассмотрим применительно к морской акватории. Для решения поставленной задачи можно использовать линейную оптимизационную модель развития и размещения производства с добавлением экологического блока:

$$\sum_{i=1}^{m} \sum_{r=1}^{R_i} C_i^r \cdot Z_i^r + \sum_{j=1}^{n} \sum_{k=1}^{K} \sum_{i=1}^{m} \sum_{r=1}^{R_i} C_{jki}^r \cdot a_{ki}^r \cdot Z_i^r \to \min \, , \qquad (6.11)$$

при условиях:

$$\sum_{i=1}^{m}\sum_{r=1}^{R_i} a_{ki}^{r} \cdot Z_i^r \geq b_k \ \ (k=1,2,...K), \tag{6.12}$$

$$\sum_{r=1}^{R} Z_i^r \leq 1 \ \ (i=1,2,...m), \tag{6.13}$$

$$\sum_{k=1}^{K}\sum_{i=1}^{m}\sum_{r=1}^{R_i} \beta_{jki}^{r} \cdot a_{ki}^{r} \cdot Z_i^r \leq A_j \ \ (j=1,2,...n), \tag{6.14}$$

где a_{ki}^{r} – объем выпускаемой продукции k-го вида на i-м предприятии по r-й технологии, β_{jki}^{r} – удельный сброс в акваторию j-го загрязняющего вещества i-м предприятием при производстве k-го вида продукции по r-й технологии, b_k – ограничение на выпуск k-го вида продукции, A_j – ограничение на сброс в акваторию j-го загрязняющего вещества, C_i^r – затраты i-го предприятия на производство продукции по r-й технологии, C_{jki}^{r} – удельный ущерб от сброса в акваторию j-го загрязняющего вещества i-м предприятием при производстве k-го вида продукции по r-й технологии, Z_i^r – искомая интенсивность выпуска продукции i-м предприятием по r-й технологии.

Для нормального состояния акватории, оптимального с экологической и экономической точек зрения, режим антропогенной нагрузки на нее, выраженный этими условиями, должен подчиняться условиям среды.

Квоты сбросов загрязняющих веществ в акваторию для каждого предприятия определяются по формуле:

$$M_{ij} = \sum_{k=1}^{K}\sum_{r=1}^{R_i} \beta_{jki}^{r} \cdot a_{ki}^{r} \cdot Z_i^{*r}, \tag{6.15}$$

где M_{ij} – допустимый объем сброса в акваторию j-го загрязняющего вещества i-м предприятием, Z_i^{*r} – оптимальная интенсивность выпуска продукции i-м предприятием по r-й технологии из решения оптимизационной задачи (6.11–6.14).

Контрольные вопросы

1. Основные понятия концепции устойчивого развития.
2. Критерии устойчивого развития.
3. Принципы рационального природопользования.
4. Соотношения между ассимиляционной емкостью водной среды и коэффициентом запаса акватории.
5. В чем заключается способность экосистем к самоочищению? Условия устойчивости природных систем.

СПИСОК ЛИТЕРАТУРЫ*

Базыкин А.Д. Математическая биофизика взаимодействующих популяций. М.: Наука, 1985. 182 с.

Бокс Дж., Дженкинс Г. Анализ временных рядов. Прогноз и управление. М.: Мир, 1974. Вып. 1.

Боровиков В.П., Боровиков И.П. Statistica. Статистический анализ и обработка данных в среде Windows. М.: Информ. изд. дом «Филинъ», 1997. 608 с.

Дружинин Н.К. Выборочное наблюдение и эксперимент. М.: Статистика, 1977. 176 с.

Дулепов В.И., Лескова О.А., Майоров И.С. Системная экология: Учеб. пособие. Владивосток: Изд-во ВГУЭС, 2004. 252 с.

Дулепова Е.П., Дулепов В.И., Ефимкин А.Я. Сравнительный анализ питания кеты и горбуши в Беринговом море в летний период // Известия ТИНРО, 2005.

Дулов В.Г., Цибаров В.А. Математическое моделирование в современном естествознании: Учеб. пособие. СПб.: Изд-во СПб. ун-та, 2001. 244 с.

Жариков Е.П. Регулирование качества окружающей среды. Владивосток: Дальнаука, 1993. 120 с.

Заславский Б.Г., Полуэктов Р.А. Управление экологическими системами. М.: Наука, 1988. 294 с.

Израэль Ю.А., Цыбань А.В. Экология и проблемы комплексного глобального мониторинга Мирового океана // Комплексный глобальный мониторинг Мирового океана. Л., 1985. Т. 1. С. 19–49.

Израэль Ю.А., Цыбань А.В. Об ассимиляционной емкости Мирового океана // Доклады АН СССР, 1983. Т. 272. № 3. С. 702–705.

Лебедева Л.П. Моделирование структуры фитопланктонного сообщества // Экология. 1971. № 4. С. 5–11.

Меншуткин В.В. Математическое моделирование популяций и сообществ водных животных. Л.: Наука, 1971. 196 с.

Меншуткин В.В. Опыт прогнозирования динамики численности озерновской красной на основе кибернетической модели этого стада // Труды ВНИРО. 1969. № 67. С. 88–100.

Меншуткин В.В. Теоретические основы математического моделирования водных экологических систем // Общая биология. 1974. № 35. С. 34–42.

Страшкраба М., Гнаук А. Пресноводные экосистемы. Математическое моделирование. М.: Мир, 1989. 376 с.

Фрисман Е.Я. Математические модели динамики численности локальной однородной популяции: Метод. разработка к курсу «Дина-

* Рекомендуемая литература выделена шрифтом.

мика численности промысловых популяций». Владивосток: Изд-во Дальрыбвтуза, 1996. 59 с.

Хикс Ч. Основные принципы планирования эксперимента: Пер. с англ. М.: Мир, 1967. 406 с.

Box G.E.P., Jenkins C.M. Some statistical aspects of adaptive optimization and control. New York: Journal of the Royal Statistical Society (ser. B). № 24. 1962. 297 p.

Brown R.G. Smoothing, Forecasting and Prediction of Discrete Time Series. New Jersey, Prentice-Hall, 1962.

Brown R.G. Statistical Forecasting for Inventory Control. New York, McGraw-Hill,1959.

Chow W.M. Adaptive control of the exponential smoothing constant. Journal of Industrial Engineering. № 5. 1965. 314 p.

Harrison P.J. Short-term sales forecasting. Applied Statistics, ser. C. № 14. 1964. 102 p.

Holt C.C. Forecasting seasonals and by exponentially weighted moving averages. Naval Research Memorandum. № 52. 1957.

Kish Z. Survey sampling. New York, 1967. 643 p.

Kish Z. Syrvey sampling. N.Y., 1967. 643 p.

Lewins R. The strategy of model building in population biology. Amer. Sci. 54, 1966. P. 421–431.

Lewis C.D. The versatility of exponential smoother’s portfolio of skills/ Production and Inventory Management. № 19. 1978. 53 p.

Mesarovic M. Systems Theory and Biology. Springer, Berlin, 1968.

Mesarovic M., Maeko D., Takahars Y. Theory of Hierarchical Multilevel Systems. New York: Academic Press, 1970.

Muir A. Automatic sales forecasting. Computer Journal. № 1. 1958. 113 p.

Patten B.C. Negentropy flow in communities of plankton. Limnol. Oceanogr. № 6. 1961. P. 26–30.

Patten B.C. The biocoenotic process in an Estuarine Phytoplankton Community. ORNL Techn. Rept. 3946 UC 48, Oak Ringe, 1966. P. 1–97.

Riley G.A., Stommel H., Bumpus D.F. Quantitative ecology of the plankton of the western North Atlantic. Bull. Bingham Oceanogr. Coll. 12. 1949. P. 1–169.

Shone M.L. Viewpoints. Operational Research Quarterly. № 3. 1967. P. 318.

Straskraba M. Cybernetic categories of ecosystem dynamics. ISEM J.2, 1980. P. 81–96.

Thamara T. Exponential smoothing with Automatic Weight Assignment, paper presented at the TIMS/ORSA Joint National Meeting. San Francisco, 1968.

TriggD.W., Leach A.G. Exponential smoothing with adaptive response rate. Operational Research Quarterly. № 1. 1967.

Ward D.H. Comparison of different systems of exponentially weighted prediction. New York: Statistician. № 13. 1963. 173 p.

Winter P.R. Forecasting sales by exponentially weighted moving averages. New Jersey: Management Science. № 6. 1960. 324 p.

ТЕРМИНЫ

Автокорреляция – степень зависимости внутри самой переменной, т.е. «переменной на себя».

Агломеративные методы кластеризации – методы последовательного объединения наиболее близких объектов в один кластер.

Адаптация – способность системы обнаруживать целенаправленное приспосабливающееся поведение в сложных средах, а также сам процесс такого приспособления.

Аддитивный тренд – тренд, в котором фактические значения отклоняются от среднего в положительную или отрицательную сторону приблизительно на одинаковую величину.

Анализ чувствительности – определение реакции модели на изменение значений параметров.

Аналитическое моделирование – моделирование при известных входных переменных и уравнений состояния.

Базовый шаг модели – интервал модельного времени между двумя последовательными состояниями модели.

Банк моделей – совокупность моделей, решающих задачи определенной предметной области и реализуемых в некоторой операционной среде на основе единого математического, программного, технического и информационного обеспечения.

Биогеоценоз – сообщество живых организмов вместе со средой обитания, ограниченное некоторыми природными границами и имеющее одно или несколько устойчивых состояний.

Блок модели – программная единица, реализующая численный метод решения некоторой математической задачи, описывающей группу функционально однородных процессов, происходящих в исследуемой системе.

Вектор параметров модели – набор констант, который должен быть задан для осуществления прогона модели.

Вектор состояния блока – массив, определяющий совокупность выходных переменных данного блока в конкретный момент времени.

Видовая структура экосистем – число видов и их представленность количеством особей, биомассой или другими показателями.

Внутрифенотипическая компонента (ВФК) экологической ниши – величина, описывающая уровень изменчивости в использовании ресурса отдельными особями.

Движущие переменные – переменные, приводящие к изменениям переменных состояния системы и не зависящие от этих переменных состояния.

Дендрограмма – дерево объединения.

Дескриптивная модель – модель, позволяющая получить информацию о взаимосвязях между наиболее важными переменными экосистемы.

Динамика популяций – изменение численности, полового и возрастного состава популяций, определяемое внутрипопуляционными процессами и взаимодействиями популяций разных видов.

Динамика системы – изменение ее состояния во времени.

Динамика численности – изменение плотности видового населения организмов в определенном биотопе или территории.

Динамические методы – методы с учетом временно́й переменной.

Дискриминантный анализ – анализ для разделения или классификации объектов.

Дисперсионный анализ – способ качественного и количественного изучения влияния одной или нескольких переменных на результаты эксперимента.

Емкость среды – число особей, потребности которых могут быть удовлетворены ресурсами данного местообитания. Степень способности природного окружения обеспечивать нормальную жизнедеятельность определенному числу организмов и их сообществ без заметного нарушения самого окружения.

Закон периодичности цикла – колебания численности жертв и хищников являются периодическими, и их период зависит от коэффициентов размножения и смертности, а также от начальных условий для численности обоих видов.

Закон сохранения средних значений – средние значения численности жертв и хищников постоянны, каковы бы ни были начальные значения численностей особей обоих видов, если сохраняются постоянными коэффициенты размножения и смертности обоих видов, а также условия нападения и защиты.

Закон эволюции биосистем – изменение с течением времени численности биологических особей какого-либо вида с любыми допустимыми индивидуальными характеристиками обусловлено естественным отбором.

Имитационное моделирование – воспроизведение динамических свойств исследуемой системы с использованием численных методов и ЭВМ.

Канонический корреляционный анализ – анализ для нахождения максимальных связей между двумя группами переменных с совместным распределением.

Кластер – это группа объектов, обладающая свойством плотности, дисперсией, отделимостью от других кластеров, формой, размером.

Корреляционный анализ – анализ для установления степени связи между двумя или бо́льшим числом стохастических переменных, а также для определения степени стохастической зависимости, существующей между ними.

Коэффициент детерминации – доля общего разброса данных (относительно выборочного среднего), которую объясняет построенная регрессионная прямая.

Коэффициент корреляции – мера зависимости двух величин или мера взаимной согласованности в изменчивости двух или нескольких признаков, явлений.

Критерий согласия – статистический критерий значимости гипотезы о том, что выборочное значение частот согласуется с некоторой теоретической моделью.

Линейный тренд – закон изменения среднего, при котором среднее возрастает или убывает со временем по линейной зависимости.

Межфенотипическая компонента (МФК) экологической ниши – величина, описывающая уровень изменчивости в использовании ресурса среди особей всей видовой популяции.

Модель поведения – модель для описания поведения систем во время переходного состояния.

Мультипликативный тренд – тренд, в котором увеличение или уменьшение фактического значения составляет приблизительно одинаковый процент относительно среднего, определяемого характером тренда.

Параметризация – определение количественных значений параметров.

Переменная состояния – переменная, подчиняющаяся закону сохранения массы и энергии; ее различные изменения описываются с помощью дифференциальных уравнений.

Популяция – совокупность особей одного вида, объединенных общностью места и времени проживания, которая воспроизводит себя в течение большого числа поколений, длительно занимает определенное пространство.

Постоянные местоположения – переменные, которые характеризуют физические границы, в пределах которых происходят изучаемые процессы.

Предсказание – это субъективная оценка будущего.

Прогноз – это результат экстраполяции прошлого в будущее.

Прогнозирование экологическое – предсказание возможного поведения природных систем, определяемого естественными процессами и воздействием на них человечества.

Прямые методы оценки параметров – методы для оценивания параметров в течение одного шага исходя из массива данных измерений на входе и выходе сигнала из системы.

Разностные уравнения – уравнения, описывающие изменения переменных состояния, вызываемые незначительными различиями независимых переменных.

Регрессия – зависимость изменений одного признака от изменений другого или нескольких признаков.

Сезонный тренд – закон изменения среднего, изменяющегося циклически.

Сообщество – популяции растений, животных и микроорганизмов, взаимодействующих друг с другом в пределах данной среды и образующих тем самым особую живую систему со своим собственным составом, структурой, взаимоотношениями с окружающей средой, развитием и функциями.

Состояние модели – объединение векторов состояния всех блоков модели.

Статистические методы – методы, не учитывающие время в качестве переменной.

Стохастическое моделирование – моделирование при неизвестных параметрах системы, причем входные параметры и переменные состояния системы изменяются во времени.

Структура экосистемы – естественное функционально-морфологическое членение экосистемы на подсистемы и блоки, играющие в экосистеме роль структурных элементов.

Трофическая структура – организация сообщества, основанная на пищевых взаимоотношениях популяций.

Трофический уровень – положение в трофической цепи, определяемое числом этапов передачи энергии.

Устойчивое развитие – развитие, которое удовлетворяет потребности настоящего времени, но не ставит под угрозу способность будущих поколений удовлетворять свои собственные потребности.

Устойчивость экосистемы – внутренне присущая системе способность противостоять изменениям.

Факторный анализ – анализ для изучения соотношений между случайными переменными, обусловленных общими факторами, а также для вывода этих соотношений.

Ширина экологической ниши – общая сумма всего разнообразия ресурсов, используемых популяцией вида.

Экологическая эффективность или эффективность пищевой цепи – относительное количество энергии, передающейся от одного трофического уровня к другому.

Экосистемный анализ – исследование структуры и функциональных особенностей экологических систем с целью установления закономерностей на экосистемном уровне и возможности прогнозирования их развития, а также динамики изменения основных компонентов экосистем.

Эффективность ассимиляции – доля потребленной организмом энергии (выраженная в процентах).

Эффективность чистой продукции – доля потребленной организмом энергии (выраженная в процентах).

Эффективность эксплуатации – относительная доля (в процентах) потенциальной жертвы или кормовых растений, поглощаемых хищниками и растительноядными животными.

ОГЛАВЛЕНИЕ

Учебное издание

Дулепов Владимир Иванович
Лескова Ольга Алексеевна

ЭКОСИСТЕМНЫЙ АНАЛИЗ

Учебное пособие

Редактор С.Г. Масленникова
Корректор Л.З. Анипко
Компьютерная верстка С.Ю. Заворотной

Лицензия на издательскую деятельность ИД № 03816 от 22.01.2001

Подписано в печать 19.04.2006. Формат 60×84/16.
Бумага писчая. Печать офсетная. Усл. печ. л. 14,4.
Уч.-изд. л. 16,0. Тираж 270 экз. Заказ

Издательство Владивостокского государственного университета
экономики и сервиса
690600, Владивосток, ул. Гоголя, 41
Отпечатано в типографии ВГУЭС
690600, Владивосток, ул. Державина, 57