

Министерство образования Российской Федерации
Петрозаводский государственный университет

А. В. Коросов

**Имитационное моделирование
в среде MS Excel
(на примерах из экологии)**

Петрозаводск
2002

УДК 28.08:22.172
К686

Рецензенты:

*член.-корр. РАН, профессор, доктор биологических наук Э. В. Ивантер;
доцент, кандидат физико-математических наук В. Б. Ефлов;
доцент, кандидат физико-математических наук С. Р. Богданов*

*Печатается по решению редакционно-издательского совета
Петрозаводского государственного университета*

Коросов А. В.

К686 Имитационное моделирование в среде MS Excel (на примерах из экологии): Монография / ПетрГУ. Петрозаводск, 2002. 212 с. ISBN 5-8021-0168-7

В книге изложены основы общедоступного метода построения *арифметических* имитационных моделей в среде Excel. Метод позволяет конструировать описательные и динамические модели любой сложности и оценивать их параметры.

Рассмотрены идеология имитационного моделирования (основы системного подхода), процедура построения моделей (табличное программирование), приемы и примеры моделирования реальных биоэкологических объектов, в том числе аппроксимация функций, декомпозиция сложных криволинейных зависимостей на более простые, декомпозиция сложных распределений на серию нормальных, интеграция серии простых моделей в общую более сложную модель, описание динамики многокомпонентных систем с помощью латентных переменных и др.

Книга адресована широкому кругу читателей, использующих компьютер для расчетов, но в первую очередь научным сотрудникам (не только биологических специальностей) и студентам.

УДК 28.08:22.172

ISBN 5-8021-0168-7

© А. В. Коросов, 2002

© Петрозаводский государственный
университет, 2002

Издание осуществлено при финансовой поддержке Института "Открытое общество. Фонд содействия" (Фонд Сороса - Россия)

Содержание

Предисловие	3
Введение	4
Постановка проблемы	8
1. Идеология моделирования: системный подход	11
Принцип системности.....	11
Принцип целесообразности.....	12
Принцип структурно-функциональной организации.....	13
Принцип иерархичности.....	15
Принцип эмерджентности.....	16
Принцип интеграции (целостности).....	16
Принцип развития.....	17
Системный подход как система.....	18
2. Процедура моделирования	19
Модель.....	19
Этапы моделирования.....	21
Пример с популяцией гадюки.....	21
Виды моделей.....	23
Построение блок-схемы.....	26
Регрессионные и имитационные модели.....	26
Аналитические и имитационные модели.....	30
Переменные и параметры.....	31
Математическое описание модели.....	34
Главное правило моделирования.....	35
Приемы составления формул.....	36
Табличное программирование.....	40
Компоненты имитационной модели.....	41
Имитационная система.....	43
Адекватность и значимость.....	48
Статистические ошибки параметров.....	50
3. Приемы моделирования	53
Фреймы имитационных систем.....	53
Аппроксимация кривой.....	57
Пропуски в данных.....	58
Усреднение и параметризация.....	61
Индивидуализация переменных.....	65
Декомпозиция кривой.....	67
Анализ распределения.....	70
Скрытые переменные.....	75
Экстремумы.....	80
Сети связей.....	83
Прогноз как гипотеза.....	88
Структура понятий имитационного моделирования.....	91
4. Теория оптимизации	92
Модель с одним параметром.....	92
Модель с несколькими параметрами.....	94
Параметры макроса «Поиск решения».....	96
Приемы работы в MS Excel.....	99
Литература	101

"...Невозможно написать учебник по построению моделей".
Р. Л. Акоф, М. Сасиени¹

Предисловие

Лет 15 тому назад автор сформулировал для себя направление поиска приемов моделирования в виде небольшого эссе "Модели – биологу". Уже тогда было понятно – чтобы войти в широкую биологическую практику, моделирование должно быть максимально простым по использованию (в программировании свои бесконечные проблемы), но абсолютно жестким по логическому оформлению (биологические высказывания должны выверяться числом). Работа в этом направлении (Ивантер, Коросов, 1992; Коросов, 1995; Коросов, 1996) позволила в конце концов наметить контуры простой процедуры построения сложных количественных описаний природных объектов. Когда идеи оформились в метод (Коросов, 1999), он был успешно апробирован при подготовке докторской диссертации по зоолого-экологической тематике (Коросов, 2000). Его следует назвать *табличное программирование арифметических имитационных моделей* в среде Excel. Об этом методе и пойдет речь дальше.

Трудно точно определить жанр предлагаемой книги. Это и монография, поскольку содержит главы, посвященные методологии общедоступного моделирования, описание новых моделей, оригинальные результаты исследований. Это и учебное пособие, предназначенное для пользователей ЭВМ, желающих освоить новый количественный метод (автор читает учебный курс "Компьютерные методы в экологии"). Это и вариант научно-популярной литературы, вводящей читателя в круг проблем количественной биологии. Книга может быть интересна как научным сотрудникам (биоэкологам), так и студентам (разных специальностей), а также школьникам.

Автор будет признателен всем читателям, которые пришлют свои комментарии и положительные отзывы по адресу: 185640, Республика Карелия, г. Петрозаводск, пр. Ленина, 33. Петрозаводский госуниверситет. Кафедра зоологии и экологии.

E-mail: korosov@karelia.ru

"Честная неудача не позорна; позорен страх перед неудачей".
Г. Форд²

Введение

Эта книга написана для читателя, использующего компьютер. Действия по образцу очень быстро привьют навык применения предлагаемого метода имитационного моделирования. Сразу же усвоить общий порядок операций поможет решение небольшой проблемы, например такой: вычислить номер гадюки, которая укусит автора в пятый раз. Идеология метода формировалась в то время, когда автор (по специальности зооэколог) изучал островную популяцию обыкновенной гадюки. Методика мечения требовала отлова множества этих животных, которые иногда кусались. С приобретением опыта возрастала и осторожность, поэтому дистанция между укусами (и временная, и по номерам особей) постоянно увеличивалась. Накопившаяся статистика позволяет заглянуть в будущее и заранее узнать, с каким же следующим номером надо будет обращаться особенно осторожно? Для этого построим простую модель зависимости двух показателей и попробуем сделать предсказание.

Включаем компьютер и запускаем программу MS Excel.

Сначала введем данные на лист Excel (рис. 1).

¹ Акоф Р. Л., Сасиени М. Основы исследования операций. М.: Мир, 1971, С. 84

² Г. Форд. Моя жизнь, мои достижения. М., 1989. С. 24.

Рис. 1. Номера особей гадюки, кусавших автора

	А	В	С	
1				
2				
3	Укус №	N особи		
4	1	14		
5	2	99		
6	3	1128		
7	4	1724		
8				
9				

Следующий этап – математическое описание модели. Если построить диаграмму, то видно, что искомая зависимость, скорее всего, криволинейна, и может быть описана, например, степенной функцией: $Y = a \cdot X^b$.

Применив наши обозначения, получим предполагаемую модель:

$$N' = a \cdot Y^b,$$

где

N' – расчетный номер особи,

Y – номер укуса по порядку.

Условные приблизительные значения параметров модели (коэффициенты a и b) необходимо задать сразу же, введем их в ячейки листа, $C1 = 1$, $C2 = 1$ (рис. 2).

Теперь можно создавать модель на листе Excel – вводить модельные

Рис. 2. Ввод первичных значений параметров

	А	В	С	
1		a=	1	
2		b=	1	
3	Укус №	N особи		
4	1	14		
5	2	99		
6	3	1128		
7	4	1724		
8	5			
9				

формулы. Правила табличного программирования требуют, чтобы значения модели для каждого объекта выборки (или одного временного шага модели) вычислялись явным образом. Введем в ячейки $C3:C9$ формулы модели, начиная с $C4$:

$$C4 = \$C\$1 * A4^{\wedge} \$C\$2,$$

где

$\$C\1 - абсолютный адрес ячейки со значением параметра a ,

$\$C\2 - абсолютный адрес ячейки со значением параметра b ,

$A4$ - относительный адрес ячейки со значением величины Y .

$*$ $^{\wedge}$ - знаки арифметических операций, умножения и возведения в степень, соответствующие конструкции модели (рис. 3).

После нажатия клавиши **Enter** ячейка будет содержать рассчитанное (заведомо неточное) значение номера гадюки при текущих уровнях параметров a и b .

Рис. 3. Ввод первой модельной

	А	В	С	Д	Е
		a=	1		
		b=	1		
	Укус №	N особи	N'		
4	1	14	= \$C\$1 * A4^ \$C\$2		
5	2	99			

Теперь нужно рассчитать остальные модельные значения, используя процедуру «автозаполнение». Для этого выделяем ячейку $C4$, наводим курсор мыши на ее правый нижний угол: он из белого креста превращается в черный крестик (рис. 4.).

Рис. 4. Начало операции автозаполнения: курсор поменял формулу

	А	В	С
1		a=	1
2		b=	1
3	Укус №	N особи	N'
4	1	14	1
5	2	99	

Нажав левую кнопку мыши, тащим курсор до нижнего угла ячейки C7, отпускаем. Все ячейки заполнились стереотипными формулами, которые рассчитали модельные значения (рис. 5.).

Рис. 5. Завершение операции

5.

	А	В	С
1		a=	1
2		b=	1
3	Укус №	N особи	N'
4	1	14	1
5	2	99	2
6	3	1128	3
7	4	1724	4
8			

С помощью двойного щелчка левой кнопкой мыши (еще лучше нажать функциональную клавишу «F2») формула в любой ячейке становится доступной для редактирования. Таким образом можно убедиться, что в каждой из ячеек содержится формула, имеющая правильные абсолютные и относительные ссылки на ячейки со значениями параметров (a и b) и ведущей переменной (Y) (рис. 6).

Рис. 6. Содержимое ячеек – формулы

6.

	А	В	С	Д
1		a=	1	
2		b=	1	
3	Укус №	N особи	N'	
4	1	14	1	
5	2	99	2	
6	3	1128	=C\$1*A6^C\$2	
7	4	1724	4	

Формула каждой ячейки рассчитывает значение номера кусающейся особи для «своего» номера укуса (из колонки А).

Итак, модельные значения (N') подсчитаны, но они явно плохо согласуются с реальными значениями (N). Например, вторичный укус сделан 99-й отловленной гадюкой, а модель дала $N' = 2$. Понятно, что принятые нами значения параметров плохо характеризуют соотношение между изучаемыми признаками. Необходимо как-то их улучшить. Для этого в первую очередь рассчитаем суммарное отличие модели от реальности. Эту роль может выполнить квадрат разности между каждой парой значений модель – реальность:

$$\phi = (\text{Мод.} - \text{Реал.})^2 = (N_i' - N_i)^2.$$

На листе Excel эта формула примет такой вид, например, для строки 4:

$$D4 = (C4 - B4)^2 = (1 - 14)^2 = -13^2 = 169.$$

Используя процедуру «автозаполнение», нетрудно рассчитать такие же значения для всех пар переменных N и N' (для проверки стоит щелкнуть в D5 и нажать «F2»; рис. 7).

Обозначим полученную графу, например, через «ф». В ячейку D8 введем формулу подсчета суммы всех квадратов различий между реальной и модельной переменными (рис. 8):

$$\Phi = \sum \phi \text{ или } D8 = \text{СУММ}(D4:D7) = 4233603.$$

Рис. 7. Расчет
отличия

	A	B	C	D	E
1		a=	1		
2		b=	1		
3	Укус №	N особи	N'	φ	
4	1	14	1	169	
5	2	99	2	$= (C5-B5)^2$	
6	3	1128	3	1265625	
7	4	1724	4	2958400	

Значение 4233603, вычисленное в ячейке D8, характеризует обобщенное отличие расчетных модельных значений признака от исходных данных. Столь большая величина определяется, видимо, тем, что произвольно назначенные коэффициенты для модельных уравнений плохо соответствуют специфике реальных зависимостей.

Рис. 8.
Расчет
обобщенной
функции
отличий (Φ)

D8 = =СУММ(D4:D7)					
	A	B	C	D	E
1		a=	1		
2		b=	1		
3	Укус №	N особи	N'	φ	
4	1	14	1	169	
5	2	99	2	9409	
6	3	1128	3	1265625	
7	4	1724	4	2958400	
8			Φ=	4233603	

Можно предположить, что при определенных значениях коэффициентов модель будет точнее описывать реальность и функция отличий (значение в ячейке D8) снизится. Это соображение позволяет начать поиск лучших значений параметров модели, изменяя их и отслеживая снижение функции различий Φ.

Например, можно было бы многократно вводить в ячейки C1 и C2 различные значения параметров, уменьшающих Φ. Так, при $a = 100$ уровень функции отличий становится немного ниже предыдущего (2456157 против 4233603) (рис. 9).

Рис. 9.
Ручная
подгонка
значений
параметров

	A	B	C	D	E
1		a=	100		
2		b=	1		
3	Укус №	N особи	N'	φ	
4	1	14	100	7396	
5	2	99	200	10201	
6	3	1128	300	685584	
7	4	1724	400	1752976	
8	5		500		
9			Φ=	2456157	

К счастью, в среде пакета Excel ручная подгонка не нужна, поскольку там имеется встроенная программа (макрос), выполняющая процедуру поиска лучших параметров, процедуру оптимизации. Вызовем ее командой «Поиск решения» из меню «Сервис» (рис. 10).

Рис. 10. Настройка окна макроса «Поиск

	A	B	C	D	E	F	G	H
1		a=	1					
2		b=	1					
3	Укус №	N особи	N'	Ф				
4	1	14	1	169				
5	2	99	2	9409				
6	3	1128	3	1265625				
7	4	1724	4	2958400				
8			Ф=	4233603				
9	Поиск решения							
10	Установить целевую ячейку: \$D\$8							
11	Равной: ☐ максимальному значению ☒ значению: 0							
12	☐ минимальному значению							
13	Изменяя ячейки:							
14	\$C\$1:\$C\$2							
15	Предположить							
16								

Если в меню «Сервис» команды «Поиск решения» обнаружить не удалось, то, скорее всего, эта процедура просто не подключена, либо не установлена. Для подключения макроса «Поиск решения» нужно вызвать окно «Надстройки» меню «Сервис», где поставить галочку напротив заголовка «Поиск решения». После этого соответствующая команда появится в меню «Сервис». Если заголовка «Поиск решения» в окне «Надстройки» нет, нужно воспользоваться установочным диском MS Office, выбрать «Добавить/Удалить / Microsoft Excel / Надстройки», поставить галочку в окне «Поиск решения» и после установки подключить процедуру.

Окно макроса, которое появляется по команде «Поиск решения», нужно правильно заполнить (например, с помощью мыши) (рис. 10):

- Установить целевую ячейку: \$D\$8
- Равной значению: 0
- Изменяя ячейки: \$C\$1:\$C\$2 (содержащие значения параметров).

После щелчка по кнопке «Выполнить» появится окно итоговых сообщений (рис. 11), предупреждающее, что поиск не может найти решения. Это естественно, ведь задача для модуля состояла в том, чтобы свести функцию отливий к нулю, а макрос смог ее уменьшить лишь до уровня $\Phi = 117686$ (ячейка D8).

С "точки зрения" макроса (подробнее см. раздел **Параметры макроса...**) найденное решение неудовлетворительно, т.е. полученные значения коэффициентов неточны. Но с точки зрения эколога такое решение может быть вполне приемлемым: значения в столбцах N и N' стали довольно близкими, ну а статистическая значимость их отливий может (и должна) быть исследована статистическими методами. Поэтому есть все основания сохранить найденное решение и рассмотреть полученное уравнение:

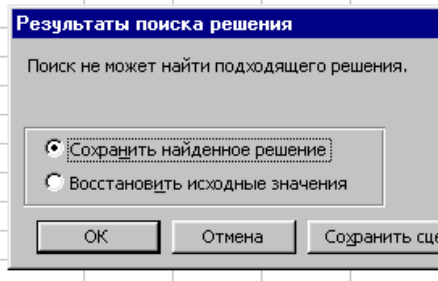
$$N' = 63.2 \cdot y^{2.41}.$$

Это и есть искомая модель. Она, конечно, крайне примитивна и вполне могла бы быть рассчитана другим способом – с помощью регрессионного анализа из блока «Анализ данных» меню «Сервис».

Тем не менее полученный результат принципиально важен для понимания существа моделирования. Теперь тот пользователь, что выполнил предложенный пример, уже освоил процедуру построения имитационной модели.

Рис. 11.
Окно
выполнения
макроса «Поиск

	A	B	C	D	E
1		a=	63.2425		
2		b=	2.41289		
3	Укус №	Н особи	N'	ф	
4		1	14	63.2425	2424.82
5		2	99	336.792	56545.1
6		3	1128	895.883	53878.4
7		4	1724	1793.56	4838.07
8			Ф=	117686	
9					
10					
11					
12					
13					
14					
15					
16					
17					
18					



Рассмотренный метод моделирования имеет такие же возможности, что и регрессионный анализ, когда дело касается простых взаимоотношений переменных (см. раздел **Регрессионные и имитационные модели**). Однако он становится незаменимым, когда экологу (биологу) приходится изучать динамические процессы.

Кроме того, гибкие средства электронной таблицы Excel содержат в себе возможности дорабатывать имитационную систему и создавать структуры для оценки статистических характеристик параметров (см. раздел **Имитационная система**).

Вводный тезис всей книги формулируется так: среда Excel позволяет создавать модели любых процессов и оценивать значения их параметров. Дальнейшее изложение посвящено его иллюстрации.

Что же касается прогноза, то предоставим читателю самостоятельно определить номер роковой змеи, подсказав направление поиска. Нужно добавить строку 8, ввести $A8 = 5$ и ...

Постановка проблемы

Количественное описание наблюдаемых явлений давно стало нормой экологического исследования. По большей части, правда, оно ограничивается каким-нибудь статистическим анализом (теперь все чаще – многомерным).

Примеры же построения динамических моделей экологических явлений пока достаточно редки. На наш взгляд, этому мешает широко распространенное (но ложное) мнение, что "полноценными" могут быть лишь аналитические модели, дающие *прогноз*, а также и другое (справедливое) мнение о чрезмерной сложности процедуры построения и решения системы дифференциальных уравнений. Они в зародыше губят мысль о моделировании. В этом печальном умозаключении нужно изменить обе посылки.

Прежде чем прогнозировать, экологические явления нужно *понять, объяснить*. Поэтому мы, во-первых, предлагаем отбросить взгляд на модель лишь как на средство прогноза и увидеть в ней способ изучения характеристик биосистем, в том числе и тех, что скрыты от непосредственного наблюдения. Во-вторых, для объяснения явлений нужно использовать иные, эффективные, средства, но доступные большинству исследователей, именно – имитационное моделирование.

Создавать имитационные модели намного легче, чем аналитические, поскольку они представляют собой компьютерные программы, которые могут строиться на базе простейших (линейных) алгебраических уравнений. Более того, системы алгебраических уравнений не приходится решать, т. к. значения параметров моделей подбираются с помощью специальных компьютерных алгоритмов (процедур

оптимизации). С этих позиций имитационные модели выглядят как числовые, арифметические, модели. Имитационные модели вполне позволяют "решать конкретные задачи арифметическими методами" (Смит, 1970, с. 67).

Итак, вместо составления и решения дифференциальных уравнений предлагается составлять *программы* и затем *настраивать* модельные параметры.

До последнего времени эта замена не многим упрощала ситуацию. Ведь даже при составлении простых программ, не говоря уже о программировании процедуры оптимизации, требуются глубокие знания какого-либо языка (Basic, Pascal, C++). Специализированные языки моделирования, такие как СЛАМ (Прицкер, 1987), ЭКО-САПФИР (Иванищев и др., 1989) и другие (Страшкраба, Гнаука, 1989; Угольницкий, 1999; Рыжиков, 2000), в основном труднодоступны и также требуют особой подготовки.

Ситуация в корне изменилась с распространением пакета Microsoft Excel, поскольку заявленные проблемы (программирование и настройка) в его среде оказываются чисто техническими и решаются очень просто (Коросов, 1999). На электронном листе Excel программирование стало *табличным*, а параметры вычисляются с помощью *встроенного* макроса "Поиск решения". Месяцы работы по составлению программ заменяются часами!

Смена целей построения моделей (объяснение) и серьезное упрощение этой процедуры (в среде Excel) должны развеять атмосферу необычайной сложности процесса моделирования, созданную вокруг него математиками, и помочь распространению этого метода.

Однако что означает "объяснить" экологическое явление, "понять" его механизм? С формальной точки зрения это значит найти удачную конструкцию модельных формул и количественно оценить значения их коэффициентов, параметров. Таким образом, в соответствии с предложенной инверсией целей моделирования во главу угла теперь ставится *оценка параметров процесса*.

Вместо того, чтобы замахиваться на невозможное, на проблему "построить модель изучаемого явления" (невозможно в одну модель втиснуть все знания специалиста об объекте), теперь требуется всего лишь "найти некоторые параметры данного процесса". Новая формулировка целей моделирования – "оценить параметры" – хорошая психологическая установка, она отвлекает настороженное подсознание от мыслей о возможных сложностях предстоящей работы, направляет его на осуществление конкретных действий.

В таком виде ожидания от предстоящего моделирования не должны быть более негативными, чем при построении привычных всем уравнений регрессии. Но что важно! Имитационная модель эффективнее регрессионной. Создание компьютерной имитации явления есть способ исследования, в результате которого вырабатывается представление о механизме изучаемого процесса, тогда как регрессионный анализ – это способ поверхностного описания фактов. Полностью справляясь с задачами регрессионного анализа, имитационная модель может дать новое знание об объекте моделирования и, кроме того, оптимизирует, *организует* ход исследования.

Моделирование – это не прием эффективного завершения научной работы. Моделирование и есть процесс (системного) исследования, который использует жесткую *количественную* логику. Мысленные представления о сложных экологических объектах не могут быть строгими во всех своих частях, они зачастую некорректны и, чтобы стать более адекватными, должны обрести каркас численного представления. Мы уверены ("по определению"), что можно построить машинную имитацию любого явления, о котором имеется хоть какое-нибудь мнение (числовые факты). Модель как количественно выраженное научное мнение обеспечивает рост объективности знаний о природе. Появившись в начале исследования в виде неточного, плохо структурированного, некорректного, неадекватного описания, модель, тем не менее, сразу начинает играть важную роль, помогая обнаруживать логические ошибки словесного описания, избавляться от мешающих стереотипов,

традиционных форм мышления, выявлять "белые пятна", указывая тем самым на актуальные направления дальнейших исследований.

Кстати, выявлять пробелы в знаниях можно и с помощью прогноза. Правда, в рамках имитационного моделирования его цели будут совсем иными. Если прогноз примет не утвердительную, а вопросительную форму ("Правда ли, что ...?"), он будет играть роль способа количественной постановки новой проблемы.

В остальном никакой экстраполяции в имитационных моделях не предполагается. Важно четко понимать, с какой точки зрения эколог-исследователь рассматривает свои материалы. Чисто технически параметры модели настраиваются по определенному набору данных, полученному некоторое время назад. Значит, исследуемые события уже произошли, реализовав динамику биосистемы. Взгляд специалиста, изучающего эти данные, всегда направлен в прошлое, он ретроспективен.

По этой причине динамические имитационные модели представляют собой ретроспекции и носят *спокойный характер интерполяции*. Расчетные параметры модели абсолютно обоснованы, поскольку относятся не ко всему, а только к наблюдавшемуся процессу. Правда, по этой же причине они оказываются ограниченными.

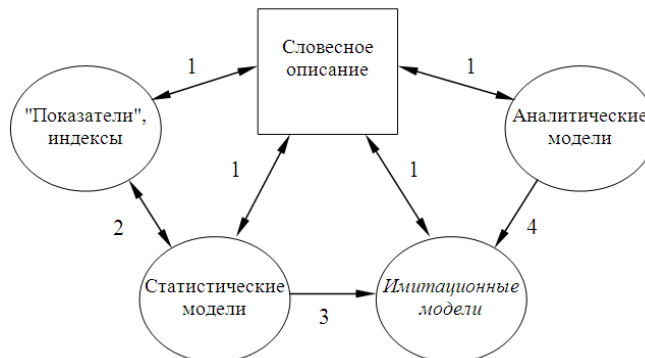
Несмотря на все позитивные моменты, имитационные модели не могут претендовать на уникальность – это лишь один из способов описания действительности, которым пользуется эмпирическое научное исследование. Подходя к классификации моделей, важно делать различие между видами исследований.

Задача *теоретического* исследования – определение (выявление и формулирование) законов, обуславливающих *возможность существования* природных объектов. Задача *эмпирического* исследования состоит в формировании знаний о *способе существования* конкретного природного объекта.

Строгая теория должна, по всей вероятности, выражаться на языке абстрактных аналитических высказываний, в первую очередь в форме дифференциальных уравнений. (Часто забывают, что это суждение имеет форму импликации, и обратное не верно: формулировка высказываний на языке дифференциальных уравнений вовсе не означает, что это теория, особенно биологическая). Что же до конкретного объекта, то его способ (процесс) существования можно успешно отобразить с помощью конкретных (портретных) количественных описаний.

К таким описаниям относятся всякого рода "показатели" и "индексы" (порожденные практикой предметных исследований), статистические описания (базирующиеся на теории вероятностей и математической статистике) и имитационные модели (рис. 12).

Рис. 12. Соотношение способов описания экологических объектов



Опыт работы автора с разными методами подсказывает, что для целей исследования *процессов* лучше всего подходит последний способ. Два достоинства делают его незаменимым. Во-первых, в рамках имитационной модели можно реализовать все мыслимые способы количественного

отображения явлений (включая статистические оценки) и решить все связанные с этим технические проблемы. Результатом выступает множество значений частных и общих параметров изучаемой системы. Во-вторых, имитационные модели позволяют воссоздать явление не по частям, но в целом, реконструировать причинно-следственные связи, обнаружить преемственность происходивших событий, т. е. представить информацию об изучаемой системе в форме знания, а не набора фактов. Эти два противоположных по направленности приема (дифференциация механизма явления и интеграция исследованных частей в целое) демонстрируют системный характер метода имитационного моделирования.

Модели любого вида базируются на словесном описании и после создания участвуют в становлении теории, равно как и в постановке вопроса о новой попытке моделирования (1). Многие выборочные "показатели" получают описание их статистической природы (2). Статистический аппарат используется для оценки адекватности имитационных моделей (3). Известные теоретические наработки, в том числе в виде дифференциальных уравнений, автоматов и пр., служат основой для построения формул имитационных моделей (4).

Здесь уместно заметить, что без системного подхода нет моделирования, точнее, моделирование и есть последовательное применение системных принципов. По этой причине особое место в книге отведено системной организации исследования. Системный подход в данном контексте выступает как общенаучная методология, но не область математики или кибернетики; в свою очередь, системные принципы представляют собой приемы эффективного мышления, в том числе и рефлексивного.

Для начала эту книгу можно прочесть "по диагонали", выбирая лишь некоторые главы. Важнейшие из них – это Введение, Этапы моделирования, Имитационная система, Фреймы имитационных систем, Структура понятий.

Иллюстрации, в основном это таблицы Excel, представляют собой готовые примеры для упражнений. Поскольку числа в таблицах прямо экспортированы из электронного листа, они содержат большое число цифр, что необходимо для контроля над ходом вычислений во время занятий. Перекрестные ссылки на разделы даны полужирным шрифтом Курьер.

Автор признателен всем сотрудникам научных учреждений Карелии, любезно предоставившим свои данные для использования в качестве примеров, в том числе А. П. Кутенкову, Н. М. Калинкиной, И. В. Литвиновой, Н. В. Медведеву, С. Н. Фомичеву, И. В. Пименовой, О. В. Сергеевой, Т. А. Прохоровой и др.

Важную роль в окончательной смысловой "настройке" текста книги сыграли рецензенты Э. В. Ивантер, Н. С. Ростова, В. Б. Ефлов и С. Р. Богданов, которых автор искренне благодарит за помощь.

1. Идеология моделирования: системный подход

Системный подход – это методика правильного мышления. Он опирается на проверенные веками приемы диалектической логики. Его функцией выступает выявление и обострение (а не погружение в подсознание) конфликтов между потребностью новых знаний и консервативной устойчивостью старых представлений. Системный подход предлагает конкретные приемы (системные принципы) оптимального хода исследовательской мысли.

Обычно говорят о четырех направлениях развития системных идей – это теория систем (философский аспект), системный подход (общенаучная методология исследования), системный анализ (методика решения проблем, в технике – ТРИЗ), системология (исследование сложных математических систем) (Кноринг, Деч, 1989; Хомяков, 1996).

В нашем представлении системный подход – это практика применения семи взаимосвязанных и взаимозависимых принципов мышления. Каждый из них

работает на скорейшее достижение конечной цели любого исследования – построение модели явления; сначала она выступает в словесной форме, затем в виде блок-схемы и, наконец, как количественное описание (Перегудов, Тарасенко, 1989).

Принцип системности

Любой объект природы (косный, живой, идеальный) можно рассматривать как систему, как совокупность взаимосвязанных частей. Принципы системного мышления приложимы к любой системе.

В рамках системного подхода понятие “система” используется как логическая категория, т. е. как способ мышления объектов природы. С этим термином возникает серьезная путаница, поскольку частные науки также используют термин “система” (экосистема, кровеносная система, солнечная система, неравновесная система и т. д.). Однако в этих случаях под системами понимаются объекты природы, но не способ их исследования. Поэтому логический термин “система” есть просто омоним “системы” частной науки. В системном подходе подразумевается относительная (целесообразная) обособленность объекта исследования, возможность применить к нему системные принципы мышления, например возможность (хотя бы умозрительно) разделить объект на части. Отсюда вытекает первая рекомендация принципа системности – *четко назвать объект исследования*.

Принцип системности выполняет описательную и объяснительную функции. Идея мыслить объект исследования “целым от природы”, “черным ящиком” работает не во всех областях биологии: в разное время один и тот же биологический объект при воздействии на него одного и того же уровня фактора может реагировать по-разному, а при воздействии разных факторов – одинаково. При системном описании специфика реакций отыскивается, в том числе и в особенностях организации биологического объекта, поэтому для объяснения наблюдаемых реакций требуется анализ, познание способа функционирования системы (Югай, 1985; Касаткин, 1988; Флейшман, 1982).

Если структурный подход требует разделения изучаемого объекта на части, то функциональный подход существенно дополняет его. В экологии объектом исследования служит, прежде всего, процесс, явление. Частями такой системы выступают частные функции, приводящие к общему результату. Соотношение между строением и функционированием системы рассматривается принципом структурно-функциональной организации. Следовательно, вторая рекомендация принципа системности – *взять объектом исследования процесс*.

Противоречие системного принципа ставит вопрос о реальности систем в условиях, когда любой объект можно мыслить системой, когда *несистем* нет, т. е. нет необходимого фона для определения систем, альтернативы определяемому. Противоречие разрешается в практике познания: систем вообще нет, есть объекты природы. Когда же один из них становится “системой” в нашем представлении, остальные мыслятся не системами (т. к. вообще не мыслятся), что и дает необходимый фон для определения объекта.

Для практики моделирования принцип системности открывает самое широкое поле деятельности, поскольку в явной форме показывает абсурдность поиска “естественных систем”, “целого от природы”, акцентирует внимание на функциональном критерии исследования, нацеливает на объяснение функции, процесса, явления.

Принцип целесообразности

Границы, содержание и поведение системы определяются целью. Система – это тот специально ограниченный набор объектов (элементов) реального мира, с помощью которого достигается определенная цель. Система – это то, что выполняет функцию.

Система определяется выполняемой функцией. Внешне эта функция выглядит как цель, то, “для чего” существует система (Аристотель, 1937).

Функция предписывает системе определенное строение и динамику (организацию). Система состоит только из тех элементов, которые участвуют в выполняемой функции. Говоря точнее, в состав системы (как самой общей модели объекта) включаются только те части, которые, на взгляд исследователя, и осуществляют системную функцию (Оптнер, 1969). Границы системы (совокупности ее частей) определяются тем, какая функция исследуется. Например, организм традиционно отделяется от среды как гомеостатическая система. Однако, исследования по “экологии человека” рассматривают другую систему: человек + воздух + вода + пища + загрязняющие компоненты. Социология рассматривает систему “группа людей”. Физиология рассматривает “функцию органа”. В каждом случае при определении границ системы доминирует цель исследования. В этом ключе важно не забывать, что некоторые цели исследователя могут быть неосознанными.

Убедиться в этом позволит небольшой опыт. Попробуйте сейчас составить небольшой список частей системы “экзамен”. Составили? Скорее всего, в нем оказались следующие объекты: студент, преподаватель, зачетка, знания, шпаргалка и что-нибудь еще в этом роде. Такой состав системы “экзамен” каждый из читателей видел с точки зрения студента, когда основной функцией экзамена видится “хороша оценка за знания”. Однако с точки зрения лаборанта, готовившего аудиторию для экзамена, его функция иная – “похвала преподавателя за подготовленную аудиторию”. Поэтому система “экзамен” для него будет состоять из чистой доски, стопки бумаги для записей, свежего воздуха, мебельной композиции и пр. Так и в экологии оказывается, что определение систем во многом определяется опытом и направлением исследований специалиста. В этой связи принцип целесообразности дает первую рекомендацию: *выявлять все цели, лежащие в основе дефиниции системы и установления ее границ.*

Детализируя рассмотрение процесса познания, принцип целесообразности обнаруживает три рода системных объектов – реальные объекты (процесс существования объектов), модели (отражение реальных объектов в сознании) и познание (процесс исследования реальных явлений и построение их моделей). Они отделимы друг от друга и характеризуются своими специфическими целями.

К **природным объектам** обычное понятие цели применимо лишь в том случае, если это бихевиоральная система (Разумовский, 1993), т. е. система, имеющая центр организации поведения (мозг животного). Для реальных объектов природы, не имеющих мозговых центров (биоценоз), применим термин *объективной цели* (Перегудов, Тарасенко, 1989). Факт функционирования (существования) биосистемы означает, что она достигает своей цели, т. е. приспособлена к среде; элиминация означает, что цель не достигнута. Природный объект состоит из “природных частей” (т. е. не тождественен любому нашему подразделению), сложно иерархически организован (иерархии, построенные нами с разных точек зрения, в реальности пересекаются). Критерием достижения цели служит “приспособительный результат”, способность пройти естественный отбор.

Цель **модели** (эмерджентная функция) состоит в том, чтобы в основных чертах функционировать *подобно* своему реальному прототипу. Если модель справляется с этой задачей хотя бы некоторое время, то она входит в систему знаний как ценное обобщение, иначе – отбраковывается. В состав модели входят логические конструкции, соотнесенные с природным объектом, например, уравнения, вычисляющие те или иные переменные на базе некоторых параметров. Модель имеет несколько уровней иерархии, начиная от среды моделирования (имитационная система) и заканчивая конкретными значениями рассчитанных переменных. Критерием качества модели служит ее адекватность реальности (практика).

Цель **моделирования** – это воплощение характеристик реального объекта в количественные характеристики идеального объекта, перенос особенных черт строения и функционирования с прообраза на дубликат, модель. Результатом моделирования должно стать понимание законов или способа функционирования реальной системы. На практике это позволяет решить проблемы (изменив режим функционирования объекта), в науке – ответить на поставленные вопросы и сформулировать новые задачи исследования. Составные компоненты процесса моделирования – это этапы построения модели, начиная с блок-схемы и заканчивая проверкой работы модели (см. ниже). Критерием моделирования выступает успешное решение проблем с помощью моделей, полноценное объяснение или предсказание изучаемого явления. Здесь цель есть предположение о результатах исследования; цель – антипод проблемы (Пойа, 1991).

Принцип целесообразности предписывает познанию рефлекссию, постоянную самооценку познавательных действий, выявление скрытых целей, критический анализ предпринятых шагов, реконструкцию побудительных причин к тому или иному действию. Самая распространенная ошибка в этом отношении – смешение *целей* и (традиционных) *средств* решения задач. Конечной целью эмпирического экологического исследования всегда выступает выяснение способа функционирования того или иного экологического объекта, а средством – использование (применение) той или иной методики, прибора, метода анализа. Отсюда вытекает вторая рекомендация принципа целесообразности: *сознательно и детально определить иерархию целей и средств их достижения*.

Противоречие принципа целесообразности состоит в том, что исследователь природных объектов всегда видит (понимает) только построенные им самим модели; хотя считает, что видит природные объекты. Однако практика постоянно подбрасывает примеры контринтуитивного поведения природных объектов, что и указывает на принципиальную неполноту любого знания.

Для практики моделирования принцип целесообразности имеет даже большее значение, чем принцип системности, поскольку вынуждает рефлексировать в процессе исследования, видеть познание в единстве трех компонентов – объекта, субъекта и метода, помогает выявлять скрытые цели, делает исследование (моделирование) прозрачным.

Принцип структурно-функциональной организации

Мыслить систему – значит мыслить элементы, из которых она состоит, и мыслить связи между элементами, объединяющие их в систему. Конечная цель системного подхода – структурированное представление системы в виде модели.

Понятие элемента как физически неделимой части целого со временем стало отражать условную неделимость как объекта, в делении которого пока нет необходимости. Элемент – это категория прекращения деления целого. Понятие “связь”, “отношение” говорит о наблюдении синхронности сосуществования двух частей. В понятие “функция” включено представление о “смысле” или “цели” отношения и о некоторой длительности его осуществления, функция есть целенаправленный процесс осуществления отношений (Аристотель, 1937; Кювье, 1937; Оптнер, 1969; Анохин, 1979).

Понятие “отношение” обнажает источник свойств вещи. Явление вещи может происходить только как ее *взаимодействие* с другими вещами (доступное для наблюдения). Его и фиксируют как “свойство” объекта исследования. При этом из поля зрения как бы выпадает внешний акцептор этих свойств. Свойство определяется как “одноместное отношение” (Сарабьянов, 1990; Перегудов, Тарасенко, 1989). Множество свойств (реализуемых отношений), которыми может быть охарактеризован объект исследования, позволяет говорить о множестве языков, с помощью которых можно выполнить его описание, что одновременно служит основанием для декомпозиции этой системы. Например, такие характеристики особи (элемента), как пол, возраст, зрелость, генотип,

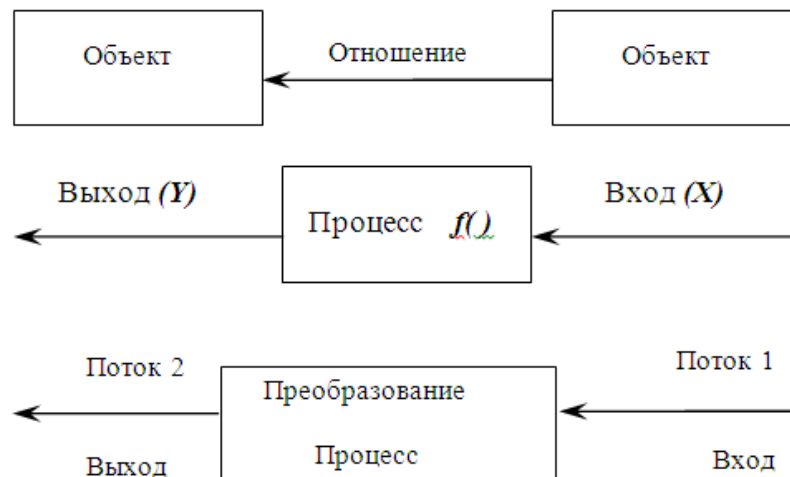
позволяют описать популяцию (систему) на языке состава – полового (2 группы), возрастного (х групп), половозрелого (3 группы), генотипического (3 группы по одному биаллельному гену). Имеются два источника происхождения языков описания: это совокупность свойств объекта (совокупность отношений объекта с окружением) и набор изучаемых уровней иерархии. Первая рекомендация рассматриваемого принципа – установление *списка языков* для описания системы.

Среди множества типов отношений (классификации, включения, порождения) для моделирования первостепенное значение имеет отношение *преобразования*. В модели статический (материальный) объект фиксируется как набор количественных характеристик, численно заданных свойств. Отношение между статическими объектами предстает как функция (динамический объект), например, как “влияние”, “изменение”. В числовом формате (в простейшем случае) это отношение представлено двумя рядами чисел (X, Y) – либо характеристиками двух объектов в разные моменты времени (тогда говорят о динамике), либо значениями двух признаков у набора вариантов (тогда говорят о выборке) (см. раздел **Приемы составления формул**). С этой точки зрения объект предстает как поток чисел, а отношение между объектами – как преобразование одного потока (X) в другой (Y).

Иначе, количественное описание отношения предстает как функция преобразования ресурса – в продукт, преобразование одних данных (множество X) – в другие (множество Y) (Ракитов, 1977).

Здесь структурная блок-схема *строения отношений* между объектами (объект – связь – объект) превращается в динамическую блок-схему *осуществления отношения* (входной элемент – процесс преобразования – выходной элемент). ВХОД преобразуется в ПРОЦЕССЕ функционирования в ВЫХОД (Винер, 1983). В таком контексте отношение предстает математической функцией и может быть выражено в виде уравнения преобразования потоков (рис. 1.1).

Рис. 1.1. Способы отражения отношения преобразования



Для графического представления отношений строят *блок-схему*. Блок-схема есть способ “визуального мышления”, который позволяет оперировать со “зримым явлением сущности”. Системный подход, предлагая жесткие правила, превращает эту операцию в исследовательскую процедуру. На наш взгляд, в основе понятийной структурно-функциональной блок-схемы должны лежать всего лишь два графических элемента:

- *стрелки* представляют “потоки”, количественные характеристики элементов системы, признаки; они всегда имеют единицы измерения; за стрелками стоят морфологические элементы системы, это – переменные;
- *прямоугольники* представляют отношения между элементами, или функции преобразования одних потоков – в другие; функции не имеют единиц размерности, они лишь производят измеряемый поток.

Эти правила только кажутся элементарными, в практике моделирования они позволяют построить жесткую основу для модели, выявляя неполноту наших представлений (Ладенко, 1981). Вторая рекомендация принципа – выражать отношения системных элементов сначала в форме понятийной, но *жестко структурированной блок-схемы*, затем в виде модельных уравнений. Система описывается, как минимум, на двух уровнях иерархии.

Структурно-функциональный подход ставит вопрос о способах декомпозиции целого. Рассматривая живой объект либо как данность, либо как процесс существования, его анализ можно провести двумя разными способами. В первом случае система может быть разделена на статические (структурные, топографические, морфологические) части, или связанное функциями множество частей. Во втором случае на части (этапы) делится процесс существования объекта, образуя совокупности динамических (функциональных, физиологических) элементов, или связанное потоками множество функций. Однако для реальных объектов статические компоненты (агенты исполнения динамики системы) неотделимы от динамических компонентов (способа связи статических компонентов). Характеризуя разные стороны системы, ни термин “структура”, ни термин “функция” оказываются не в состоянии целиком охватить существо объекта. Объединяющим понятием становится “организация”. Временной аспект понятия “организация” предполагает поочередное рассмотрение то органа (элемента), то его функции (связи с другим органом) в порядке их реализации (Паавер, 1976; Мороз, 1987; Саламатов, 1991). Организационная блок-схема представляет собой временную развертку структурной блок-схемы. Этот прием структурирования системы очень важен для построения имитационных моделей (см. раздел **Виды моделей**).

Структурирование биосистем осложняется множеством нестрогих корреляций между функциями их частей. Ключевыми понятиями выступают *мультифункциональность* (любой орган выполняет множество функций) и *полиструктурность* (любая функция выполняется множеством органов) (Северцов, 1980; Славин, 1989; Пучковский, 1998). Любые две функции выполняются двумя *пересекающимися* множествами органов, т. е. наблюдается *наложение* функций друг на друга (*транссессия*). Для адекватного (контринтуитивного) представления действительности (помимо описания прямых функций органов) на более глубоком уровне иерархии приходится строить модели *субстратного (органного) взаимодействия* между функциями (Коросов, 1999). Не только уровни потоков (переменных) связаны друг с другом, но также и параметры функции преобразования разных потоков.

В соответствии с третьей рекомендацией принципа следует акцентировать внимание на возможной зависимости параметров разных функций друг от друга. Система описывается уже на трех уровнях иерархии – как множество элементов (2 верхних уровня) и как множество функций в пределах элемента (2 нижних уровня) (подробнее см. **Принцип эмерджентности**).

Противоречие структурно-функционального принципа ставит вопрос о границах систем в условиях всеобщей взаимосвязи. В рамках конкретного исследования граница изучаемой системы определяется целью этого исследования.

Принцип иерархичности

Каждый элемент – система. Элемент системы участвует в ее образовании при взаимодействии с соседними элементами; одновременно он сам состоит из взаимодействующих частей и поэтому выступает в роли системы нижележащего уровня.

Понятие иерархии в первую очередь отражает простую и очевидную возможность членения любого объекта на все более мелкие детали (Рассел, 1957; Ленин, 1969; Гегель, 1974). В то же время иерархические ряды одних и тех же объектов (например, экосистем, биоты в целом), построенные разными авторами, сильно различаются. Это свидетельствует об искусственности

иерархических подразделений, которые определяются целями построения иерархий, а также критериями, подсказанными опытом и интуицией авторов. В рамках поставленных целей одни и те же объекты природы могут попадать на разные уровни иерархии: клетка как *часть* взрослого организма и как *весь* организм зародыша (Энгельс, 1987). С позиции рассматриваемого принципа логическое противоречие всегда выступает в виде границы между объектом как элементом (одно) и объектом как системой (многое) (Лосев, 1994).

Иерархии зиждутся на отношении включения (а не на отношении соподчинения): компоненты нижнего уровня иерархии суть части вышележащего уровня. Критерием выделения элементов одного уровня служит рассмотренный выше структурно-функциональный принцип: множество элементов относятся к одному уровню иерархии, если они непосредственно взаимодействуют и совместно выполняют общую функцию. Назвать уровень иерархии означает назвать множество элементов, множество отношений между ними, т. е. систему. Понятно, что иерархическую структуру можно выявить как у статических объектов (системы – подсистемы – ... – элементы – ...), так и у динамических объектов (процессы – ... – этапы – подэтапы – ...).

Для моделирования принцип иерархии представляет собой еще один способ выбрать языки описания системы. Так, динамику экосистем эффективно описывают на языке трансформации энергии (физический уровень иерархии, самый низкий), вещества (химический), численности (популяционный), биомассы (видовой), трофической структуры (ценотический). Основная рекомендация принципа иерархичности состоит в осознанном (для достижения поставленной цели) выборе некоего уровня иерархии в качестве базового языка модели; это – центральный уровень модели.

Противоречие принципа иерархичности состоит в том, что решение задачи описания данной системы возможно лишь при решении задачи описания этой системы как элемента более общей системы, но решение последней задачи возможно лишь после решения первой. Конкретное исследование всегда опирается на неполные данные и дает не исчерпывающую картину явления, но лишь более приближенную к реальности.

Принцип эмерджентности

Система обладает особым свойством, которое отсутствует у слагающих ее элементов.

В силу универсальности системных принципов эмерджентные (эмергентные) качества присущи любой системе, в том числе и неживой, их число бесконечно. Системный подход рассматривает эмерджентное свойство как не *абсолютно* новое, но как *существенно* новое свойство. Новое свойство “целого” есть лишь свойство некоего элемента этой системы, *усиленное* участием всех остальных элементов системы и “продленное” за границы системы.

Практика моделирования эмерджентных свойств демонстрирует два подхода – холизм и редукционизм.

Холизм рассматривает (простую) систему как неделимую данность, как “черный ящик”, имеющий неизвестную внутреннюю структуру (Одум, 1975; Саати, Кернс, 1991). Свойства системы при взаимодействии с внешними системами рассматриваются как *реакции* системы на *действие* фактора. Такие эмерджентные качества выражаются аналитическими или регрессионными моделями. Модельное описание охватывает *два уровня иерархии* – собственно систему и среду, в которой она находится, т. е. надсистему (состоящую из других систем). При этом свойства системы видятся как “зависимые”, обусловленные внешними влияниями.

Редукционизм явно декомпозирует систему, описывает элементы и их связи; эмерджентное качество определяется как результат взаимодействия элементов (Лаплас, цит. по: Клайн, 1988; Блауберг, Юдин, 1972; Флейшман, 1982). Для объяснения сложных систем строят имитационные модели. Редукционное описание базируется обычно на нескольких языках. Многие элементы поэтому

оказываются мультифункциональными и играют несколько ролей в динамике системы, т. е. характеризуются несколькими переменными. В силу общности субстрата (один элемент) эти переменные коррелируют друг с другом. При редукционистском подходе для описания также используются *два уровня иерархии* – система и ее элементы. При этом свойства системы видятся как “независимые”, “свободные”, обусловленные внутренней имманентной сущностью системы.

Системный подход к исследованию реальности в двух методах (редукционизме и холизме) видит две стороны единого описания противоречивой системы. Биологический объект (явление) есть одновременно и “одно” (одна *система*), взаимодействующее со средой как целое, и “многое” (система *элементов*), сложенное из частей, осуществляющих ее функции и поведение (Ильенков, 1984; Лосев, 1993). Принцип эмерджентности рекомендует для моделирования рассматривать систему в обеих ипостасях параллельно, т. е. изучать объект и строить модель на трех уровнях иерархии: на уровне субэлементов, системы и надсистемы.

Противоречие принципа эмерджентности состоит в постоянном конфликте “свободных” свойств системы (как множества элементов) и “вынужденных” свойств системы (как единичного элемента надсистемы более высокого уровня иерархии).

Принцип интеграции (целостности)

Свойства части системы (элемента) определяются свойствами соседних системных элементов, навязываются части, делают ее зависимой от целого.

Беря часть, следует видеть тот контекст, в котором она существует, функционирует, используется. Целое появляется при “сотрудничестве” (Богданов, 1990), “взаимосодействии” (Анохин, 1979) частей. В пределах системы элементы приобретают особенные дополнительные (эмерджентные) качества. В отличие от эмерджентности, направленной “наружу” системы, новые качества элементов ориентированы “вовнутрь” целого (Акоф, Сасиени, 1971; Гегель, 1974; Ракитов, 1977). Мысль о целостности должна явно разделять эти эмерджентности, т. е. должна вводить понятие *границы* системы (Оптнер, 1969; Блауберг, Юдин, 1973).

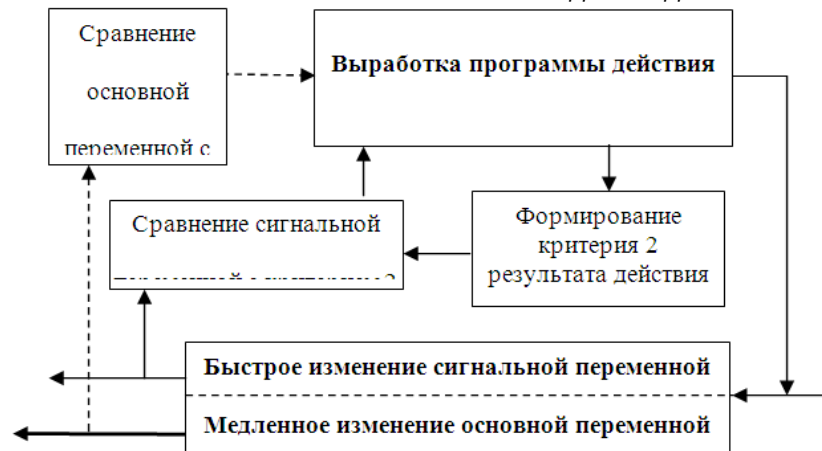
Наиболее характерным признаком целостности динамических (в первую очередь живых) систем выступает сохранение относительного постоянства внутренней среды, или *гомеостаз* (Волькенштейн, 1981). Механизм гомеостаза обычно описывается с помощью понятий корреляции (Шмальгаузен, 1982), обратной афферентации (Анохин, 1979), отрицательной обратной связи (Винер, 1983). Обратная связь – это контроль за конечным результатом какого-либо процесса. Изменения на выходе передаются (влияют) на вход в динамическую (поведенческую) систему. Отрицательная обратная связь *ограничивает* систему от резких колебаний ее переменных и обеспечивает ей достаточно независимое существование (функционирование) (Саати, Кернс, 1991).

В биосистеме прямые и обратные связи осуществляются одновременно. В имитационной модели биосистемы, где время задается дискретно, как множество последовательных состояний системы, реализация зависимых процессов осуществляется последовательно и выглядит как чередование этапов системного процесса (в моменты времени $t_0, t_2 \dots$) и следующих за ними этапов работы управляющей обратной связи (в моменты времени $t_1, t_3 \dots$), т. е. предстает как организация системы.

Полная базовая блок-схема обратной связи сложного поведенческого акта имеет две петли (Анохин, 1979) (рис. 1.2). Первый контур воспринимает отклонение от нормы *основного* параметра внутренней среды организма и вызывает построение реакции (поведенческого акта) для его компенсации; ключевым значением здесь служит “встроенный” первый критерий нормы. Второй контур (обратная афферентация) воспринимает отклонение от нормы *сигнального* параметра, который несет информацию о результатах

компенсаторных реакций организма, и либо тормозит, либо не тормозит повторную реализацию поведенческого акта; второй критерий “нормы” формируется непосредственно в процессе поведенческого акта.

Рис. 1.2. Блок-схемы основных видов моделей



Между основным и сигнальным параметрами имеется тесная, эволюционно выверенная корреляционная связь: поступление сигнала по второй петле обратной связи означает возможную скорую нормализацию внутренней среды (как следствие предпринятых действий). Сигнальный параметр имеет высокую скорость изменения, основной параметр много более инерционен.

Представленная динамическая система поведения не содержит специфических компонентов, универсальна и приложима к широкому спектру управляемых систем. Она указывает на ключевые параметры сложной поведенческой динамики биосистемы и на основе их оценки позволяет строить модели сложной реактивности организма, включая поведение.

Противоречие принципа целостности: мысль о системе неполноценна без мысли об ее элементах, но последняя отвергает первую; мысль об элементах уже не есть мысль о целом.

Принцип развития

Любая система представляет собой результат исторического хода событий. С системных позиций развитие есть, во-первых, дифференциация системы (усложнение организации: рост числа элементов и связей) и, во-вторых, усиление ее интеграции (структурирование: усиление взаимозависимости частей системы, ведущее к усилению эмерджентных качеств).

В наибольшей степени принцип развития (историзма) используется для отражения хода процесса построения моделей, а также в близкой области – при техническом конструировании (ТРИЗ) (Альтшуллер, 1979).

Идейным источником принципа историзма выступает диалектическая логика – метод мышления противоречий, логика всеобщего развития. Г. В. Ф. Гегель выделяет три движения мысли: логику абстрактную (или рассудочную), диалектическую (или отрицательно-разумную) и спекулятивную (или положительно-разумную). Мысли и плоды деятельности аккумулируются, выверяются и в конце концов формируют знание, правильно отражающее мир. При этом правильность мышления понимается как соответствие знаний человека – окружающей действительности. В триаде Гегеля отслежен лишь закон логического становления мысли, но не апробация ее правильности. Такое дополнение в виде четвертого звена сделано и в философии (Лосев, 1995) и осуществляется в жизни. Известны следующие четыре этапа познания (Энгельс, 1952; Гегель, 1974; Бэкон, 1977; Ленин, 1982; Кант, 1994; Лосев, 1995).

Положение (тезис): объект определяется в первом приближении как данность, как контуры будущего знания (результат – первая дефиниция).

Анализ (антитезис, отрицание): объект исследуется во всех своих проявлениях – частях, связях, этапах, функциях (результат – накопленная информация об исследуемом объекте).

Синтез (отрицание отрицания): все разрозненные знания об объекте сводятся в единую (динамическую) систему, организацию, свойства объекта поняты как явления его содержания (результат – модель объекта).

Апробация (факт, отбор, практика): сопоставление знаний с реальностью, проверка идей и моделей в практике их использования (результат – решение проблемы, формулировка более полной дефиниции).

Понятие, определенное диалектически, есть не просто абстрактное умополагание или эмпирический факт, но их синтез, правильность которого выверена практикой. Последовательная детализация, переосмысление, “опосредование” объекта раскрывает его сущность с новых сторон, вскрывают все большее его содержание. Знание не просто растет, но развивается, структурируется, становится более функциональным, объяснительным и прогностичным.

Противоречие принципа развития констатирует, что завершение исследования есть начало нового его этапа, т. е. вовсе не завершение.

Системный подход как система

Рассматривая любой объект (процесс) как систему, системный подход обязан применять свои методы рефлексивно, постоянно уяснять собственную системную природу (Саати, Кернс, 1991). Свою эмерджентную функцию он видит в решении проблем или получении ответа на научный вопрос. Помимо структуры (7 взаимосвязанных принципов) он предписывает себе динамику. Исторический опыт говорит о *четырёх* специфических этапах процесса познания. На наш взгляд, это и есть этапы системного подхода: *определение системы, декомпозиция, агрегация, реализация*. Рассмотренные ранее системные принципы мышления выступают сквозными идеями всего процесса. На каждом этапе вырабатывается специфический промежуточный результат: *дефиниция, информация, блок-схема (модель), решение проблемы*. Эти результаты выступают в роли *связей* между элементами (этапами) исследования, это информационные *потoki* (табл. 1.1).

Успешная апробация полученной модели переводит весь процесс в начальное положение: данная модель представляет собой новое, более развитое, но бесконечно неполное определение системы. Отбор (реализация) – суть *обратная связь* процесса системного подхода, этот этап содержит критерий успешности проведенного синтеза.

В отличие от нашей схемы число этапов по другим источникам широко варьирует (до 20). На наш взгляд, противоречие можно разрешить с системных позиций, если каждый этап рассматривать как отдельную систему (процесс), состоящий из тех же *четырёх* последовательных этапов. На 2-м уровне иерархии весь системный процесс состоит уже из 16 этапов. Каждый из них может быть, в свою очередь, разделен на 4 и т. д.; моделирование требует углубления до 5 уровня иерархии процесса познания. Все исследование и каждый его этап представляют собой итеративный процесс, который заставляет возвращаться на предыдущие ступени, если достигнутые результаты не удовлетворяют заранее выдвинутому критерию, или переопределяются сами критерии.

Таблица 1.1. Дробные этапы системного подхода (2 уровня иерархии)

Определение системы	Определение проблемы, объекта, целей
Определение	Назвать эмерджентное качество системы, которое нужно усилить
Декомпозиция	Составить перечень языков для описания системы. Определить окружение системы и участников исследования

Определение системы	Определение проблемы, объекта, целей
Агрегация	Назвать качества, которые объект моделирования, в известное время, проявляет по отношению к известным соседям. Определить цели как моделирование эмерджентных качеств. Определить проблему как улучшение эмерджентного качества
Отбор	Критерий: понятность объекта и полнота информации Итерация: смешение целей и средств, мало данных
Декомпозиция системы	Установление состава, структуры, организации
Определение	Планировать вид таблицы декомпозиции
Декомпозиция	Составить перечень оснований декомпозиции объекта (языки, иерархия, структура, функция, этапы, исходы, воздействия) Планировать дополнительные исследования
Агрегация	Провести научные исследования объекта. Декомпонировать систему до известной глубины иерархии по известным основаниям на известных языках. Составить иерархические таблицы декомпозиции (и блок-схему)
Отбор	Критерий: соответствие уровня детности целям исследования Итерация: слабое структурирование, провисание связей, пробелы блок-схемы, избыток или недостаток языков, мало информации
Агрегация системы	Описание динамики, построение модели
Определение	Планировать методы статистического анализа и вид моделей для описания систем
Декомпозиция	Составить перечень агрегатов (статистики, структуры, иерархии, организации, классификации, типы моделей)
Агрегация	Составить блок-схему модели Моделирование и модельное решение проблемы
Отбор	Критерий: получение модельного решения Итерация: отсутствие решения, неадекватность, высокая стоимость
Отбор системы	Определение направления желательного развития
Определение	Разработка критерия удовлетворительной работы
Декомпозиция	Оценка работы модели в реальных условиях
Агрегация	Выработка решения о пересмотре модели
Отбор	Критерий: проблема решена Итерация: проблема не решена

2. Процедура моделирования

Модель – это мысль о системе (Коган, 1968; Форрестер, 1971). Системный подход конкретизирует неточные мысленные и словесные построения и переводит их в форму структурированных описаний с помощью изобразительных, аналоговых и символических (математических) моделей (Акоф, Сасиени, 1971). Процесс познания при моделировании трансформируется: в поле зрения исследователя попадает не только реальность, но и ее “слепок” – модель (Либа, Мауринь, 1970). Сложность моделирования в том и состоит, что модель выступает одновременно и как результат исследования (форма обобщения известной информации о реальной системе), и как объект этого же исследования (существующая в ЭВМ искусственная система). Неполнота построенной модели инициирует продолжение исследования.

Модель

Вообще говоря, любая мысль об окружающем мире есть его модель; в системном подходе моделью служит представление объекта в виде системы. “Модель – это мысленное представление системы” (Кноринг, Деч, 1989, с. 25).

“Модель определяется как представление системы” (Оптнер, 1969, с. 128). “Понять феномен – значит построить его мысленную модель” (Левченко, 1993, с. 302). “Любое теоретическое представление, сложившееся в результате наблюдений или экспериментов, является в известном смысле качественной моделью изучаемых процессов” (Коган, 1968, с. 221). “Всякая гипотеза является обобщенной моделью некоторого явления” (Шмальгаузен, 1968, с. 218).

Осознанные мысленные модели воплощаются в словесное описание. “Словесное описание есть модель, наше мысленное представление ... тоже модель” (Форрестер, 1971, с. 46). Основная проблема словесных моделей – это их неточность. “Интуитивная емкость ... <словесных> описаний, являющаяся в известном смысле их достоинством, не может служить достаточным основанием для того, чтобы судить о полноте и точности отображения с их помощью соответствующих объектов” (Ладенко, 1981, с. 197).

Как следует из предыдущего раздела, логическое понятие “система” также есть модель реального объекта, точнее, способ представления объекта в виде модели. При дальнейшем структурировании образ реального предмета может воплотиться в какой-либо материал и стать моделью иного рода. “Обычно используют модели трех типов: изобразительные (модели геометрического подобия), аналоговые и символические (математические)” (Акоф, Сасиени, 1971, с. 80). Объект природы предстает перед сознанием в трех ипостасях: как независимая от исследователя данность, как система, как модель.

В соответствии с нашими целями заострим внимание на математических (символьных, или знаковых) представлениях объекта исследования. “Словесная модель и математическая модель... обе являются абстрактным описанием реальных систем. Математическая модель более упорядочена... более “точна”... <но> не обязательно более правильна, чем словесная...” (Форрестер, 1971, с. 47). (Еще раз заметим в скобках, что реальны лишь объекты, называемые в частных науках системами, “система” системного подхода суть логическая абстракция, см. раздел **Принцип системности.**)

Поставив своей целью более точное описание изучаемого объекта, словесные описания переводятся в математические формы. “На этапе *моделирования* качественные представления переходят в количественные” (Пэнтл, 1979, с. 34).

При этом определяются строгие математические соотношения между исследуемыми характеристиками системы. “Математическая модель представляет собой систему математических соотношений, формул, функций, уравнений... описывающих те или иные стороны изучаемого объекта, явления, процесса” (Антомонов, 1977, с. 10). “Математической моделью системы называют математические соотношения (уравнения, неравенства и пр.) или программы, описывающие некоторые характеристики этой системы” (Розенберг, 1984, с. 7).

В выборе средств для такого описания почти не существует ограничений, хотя большинство моделей традиционно пишутся на языке дифференциальных уравнений. “В настоящее время практически все методы классической и современной математики используются при анализе биологических и медицинских систем и решении прикладных задач” (Методы..., 1980; с. 160).

Среди математических моделей наряду с аналитическими, стохастическими, матричными, многомерными, оптимизационными, эволюционными и другими (Джефферс, 1981, с. 38; Ляпунов, 1968, с. 67; Розенберг, 1984, с. 25) выделяют особый тип – имитационные модели, связанные с тесным использованием ЭВМ и наиболее доступные для биологов и экологов. “Под имитационной моделью понимают программу, которая в процессе ее реализации на ЭВМ позволяет имитировать поведение реальной системы в разных условиях” (Методы..., 1980, с. 74). “Имитационная модель – логико-математическое описание системы, которое может быть исследовано в ходе проверочных экспериментов на цифровой ЭВМ и ... может считаться лабораторной версией системы (Прицкер, 1987, с. 14).

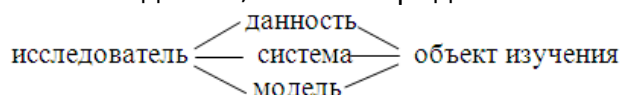
Имитационная модель представляет собой очень гибкий инструмент описания действительности. В ее потенциях стать сложной и детализированной или, напротив, простой и обобщающей. Назначение (широта) модели определяется целью исследования; только явная формулировка желаемого результата позволяет справиться с постоянной коллизией простоты и полноты модельного описания. “Исследователь является творцом модели” (Флейшман, 1982, с. 55).

В зависимости от целей модель может охватить как самые общие явления в целом, так и все многочисленные его частности. Чем более грубо рассмотрены свойства модели, чем меньше она содержит деталей (переменных и параметров), тем более выразительно (и эстетично) выглядит модель. “Степень понимания явления обратно пропорциональна числу переменных, фигурирующих в его описании” (Акоф, Сасиени, 1971, с. 82). “Чем более широкое значение призвана иметь модель, тем более она должна быть свободна от частных подробностей” (Рыжков, 1968, с. 198).

Если же преследовать целью аккумуляцию в модели всего знания о реальном объекте, тогда число компонентов должно возрасти. “...При попытке воссоздать “портретный” образ экосистемы стараются использовать всю имеющуюся информацию... В имитационной модели обычно стремятся учесть как можно больше деталей” (Федоров, 1983, с. 19).

Имитационные модели служат полноправными (но не всегда полноценными) представителями реальных прототипов. “Модель... – вспомогательный объект, который в процессе исследования способен замещать объект-оригинал так, что его изучение дает нам новую информацию об объекте-оригинале” (Лиёпа, Мауринь, 1970, с. 6).

В поле зрения исследователя оказывается не только реальность, но еще и более или менее точный ее слепок – модель. Сам процесс исследования при этом трансформируется: часть усилий идет на изучение природного явления, часть – на создание ее модельного отображения. “Процесс познания характеризуется схемой: исследователь – объект изучения. При моделировании та же схема имеет вид: исследователь – модель – объект изучения (оригинал)... Необходимую информацию об объекте получают посредством исследования модели” (Лиёпа, 1982, с. 4). Фактически схема еще сложнее, поскольку прежде, чем стать математической моделью, объект предстает как система:



На наш взгляд, методологическая сложность моделирования состоит именно в том, что модель выступает одновременно и как результат исследования (форма обобщения известной информации о реальной системе), и как объект этого же исследования (самостоятельно существующая в ЭВМ искусственная система). Процесс познания складывается из чередования этапов: конструирования модели (на базе натурных данных и системного образа объекта) и выработки гипотез для следующего исследовательского цикла (на основании изучения поведения модели).

Увлечение первым или вторым этапом этого процесса может завести исследование в тупик. “Модель есть некоторое абстрактное отражение моделируемого объекта, формальное упрощение реального мира, подменяющее изучение реального объекта изучением более искусственной системы... Поэтому модель всегда беднее той реальности, которую описывает” (Федоров, 1983, с. 17).

Модель лишь подобна, но не идентична реальной системе. В этом отношении она, конечно, беднее реальности. Но такова общая проблема познания: известное всегда меньше неизвестного, любая информация – лишь часть истины. Вместе с тем эта ограниченность в определенном смысле оказывается не бедой, но достоинством модели, путем к пониманию окружающего многообразия в его принципиальных отношениях. Это поиск законов,

управляющих миром. “Моделирование в биологии и медицине всегда имеет дело с представлением сложной системы посредством более простой и более доступной для исследования” (Купий, 1989, с. 90).

Этапы моделирования

Моделирование как система имеет своей функцией (целью) синтез всех известных элементов в неизвестную модель, это способ агрегации структурно (в блок-схеме) и количественно (в массиве чисел) заданных элементов и последующего конструирования количественного прототипа реальной системы. Моделирование еще не стало обычной процедурой биологического исследования, и в литературе обнаруживается большое разнообразие алгоритмов этого процесса (Антомонов, 1977; Розенберг, 1984; Прицкер, 1987; Шеннон, 1970). На наш взгляд, проводники идей моделирования не уделяют должного внимания его системной упорядоченности. Мы постарались применить здесь идеи об организации исследовательской деятельности, которая проходит четыре этапа: Определение, Декомпозицию, Агрегацию, Проверку (см. раздел **Идеология моделирования**); так получаем схему этапов моделирования (табл. 2.1). Пять шагов агрегации (A1-A5) мы выделяем потому, что каждый из них преследует отдельную явную цель, т. е. все они выполняют равнозначные функции. В этом смысле они находятся на одном уровне иерархии: выход одного этапа становится входом для другого. Можно полагать, что пятишаговый процесс построения модели есть устоявшийся фрейм.

Пример с популяцией гадюки

Рассмотрим задачу построения имитационной модели для определения демографических параметров островной популяции обыкновенной гадюки по реальным данным повторного отлова меченых животных. Общая цель всего исследования состоит в оценке абсолютной численности островной популяции.

Таблица 2.1. Этапы моделирования (три уровня иерархии)

Этап	Подэтап	Содержание этапа (подэтапа)
О		Цель: Найти модельное решение проблемы
Д		Определить типы и виды моделей, которые будут использоваться (статистические, имитационные), описать структуру данных (экземпляры, показатели, переменные, параметры), выборку, выбрать алгоритмы оптимизации (идентификации параметров) компонентов имитационной системы
А		Описать динамику, построить модель, провести моделирование, найти решение, найти ответ на конкретный научный вопрос
A1		Построение блок-схемы
	О	Цель: Структурировать словесную модель
	Д	Составление таблиц декомпозиции
	А	Построение идеограммы блок-схемы по правилам
	П	Критерий: все компоненты объекта включены в блок-схему
A2		Математическое описание модели
	О	Цель: Создать аналитическое описание блок-схемы
	Д	Математическое представление всех связей (преобразований) переменных, выявление обратных связей, нелинейности
	А	Формирование системы уравнений со счетчиком времени
	П	Критерий: все компоненты блок-схемы описаны уравнениями
A3		Программирование модели
	О	Цель: Подготовить программу, реализующую алгоритм
	Д	Выбор языка и средств программирования, организация дискретного хода счета (датчика времени), описание переменных, параметров, программирование и отладка блоков

Этап	Подэтап	Содержание этапа (подэтапа)
	А	Синтез программы в целом, отладка программы
	П	Критерий: программа работает правильно
А4	Настройка модели	
	О	Цель: Подготовить модель к экспериментам
	Д	Подбор данных, организация имитационной системы
	А	Идентификация (настройка параметров), верификация модели с помощью рандомизации
	П	Критерий: модель адекватна исходным данным, описаны существенные причинно-следственные связи
А5	Эксперименты с моделью	
	О	Цель: спланировать и провести моделирование
	Д	Стратегическое и тактическое планирование, эксперименты при разных режимах, условиях, заданиях
	А	Статистическое описание и оценка результатов
	П	Критерий: обнаружено оптимальное решение или получен ответ на поставленный вопрос
П	Доказана статистическая значимость ответа Критерий: решение найдено, ответ получен	

На этапе **Определение системы** важно обрисовать объект исследования в самых общих чертах. Таким объектом выступает динамика островной популяции, т. е. процесс поддержания численности примерно на одном уровне: убыль от гибели животных компенсируется пополнением за счет размножения. Реальным объектом исследования выступает популяция гадюки на о. Кижы, что в Онежском озере (Карелия). Для нее характерно примерное постоянство общего уровня численности (Коросов и др., 1999б). Популяция имеет две количественно заданные компоненты. Одна часть популяции – это меченые животные, другая – немеченые. Поскольку метки не влияли на жизнеспособность животных, то судьба и меченых и немеченых животных оказывалась одинаковой – они постепенно погибали. На этом определение системы для последующего моделирования заканчивается.

Начиная следующий этап, **Декомпозицию**, во-первых, нужно описать структуру эмпирических данных. Мечение гадюк на о. Кижы проводилось в течение нескольких лет. В повторных пробах определяли встречаемость меченых особей. Для наглядности мы воспользуемся данными только по одному году мечения. Так, в 1994 г. отловили $n_0 = 158$ экз., всех поместили и выпустили $M_0 = 158$ особей. В последующие годы (1995–1998 гг.) животных с метками обнаруживали в пробах разного объема ($n_1 = 365$, $n_2 = 273$, $n_3 = 214$, $n_4 = 238$ экз.), но число их уменьшалось: $m_1 = 18$ экз. в 1995 г., затем $m_2 = 10$, $m_3 = 10$ и $m_4 = 9$ экз. в 1998 г. Продолжение декомпозиции рассмотрено в следующем разделе.

Виды моделей

Существует великое множество классификаций видов моделей, так же как и классификаций видов систем, – все зависит от принятого основания. Чаще всего эту роль играет либо *процедура* моделирования, либо *функция*, которую выполняет модель.

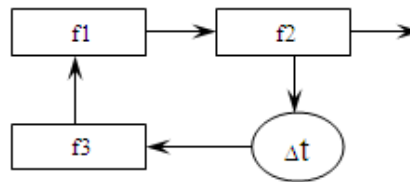
В первом случае выделяют четыре вида моделей: *аналитические*, *самоорганизующиеся*, *имитационные* (портретные, дискретные) и *эмпирико-статистические* (Розенберг, 1984, с. 25–26). Во втором случае различают модели *аналитические* (конструктивные, модели поведения, детерминистические) и *стохастические* (дискретивные, описательные) (Страшкраба, Гнаук, 1989, с. 26; Акоф, Сасиени, 1971, с. 81; Джефферс, 1981, с. 33–35). Каждый из классов, в свою очередь, может дробиться на подклассы. Так, среди дискретных

имитационных моделей с позиции способа выполнения имитации выделяют **событийный** подход (вычисление характеристик каждого нового состояния модели синхронизируется внутренним календарем событий) и подход **сканирования** (переход модели в новое состояние задается условиями начала и конца событий); синтетический процессно-ориентированный подход предполагает совмещение последовательной обработки с расчетами по условию достаточности (Прицкер, 1987, с. 321).

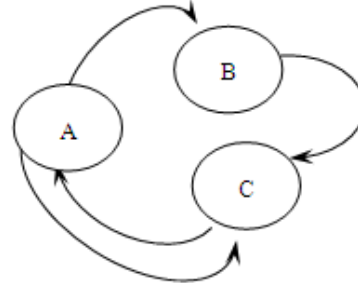
Основываясь на описаниях видов моделей, можно предложить еще одну классификацию с точки зрения предопределения структуры будущей блок-схемы заранее выбранным видом модели. С этой позиции отчетливо различаются модели **аналитические** (непрерывные), **автоматные** (сканирование), **стохастические** и **имитационные** (дискретные, портретные) (рис. 2.1).

Рис. 2.1. Блок-схемы основных видов моделей

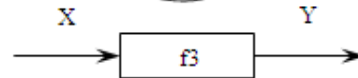
а. Аналитическая модель



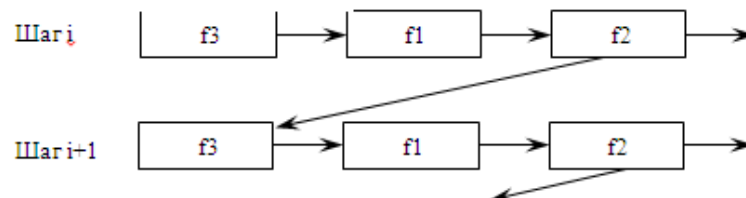
б. Автоматная модель



в. Регрессионная модель



г. Имитационная модель



Аналитические модели (системы дифференциальных уравнений) служат для выражения законов. Параметрам аналитических моделей стремятся придать биологический смысл. Они включают время в качестве одной из переменных, что позволяет в принципе находить общее решение, т. е. оценивать состояние системы (значение ее переменных) в любой конкретный момент времени, без реконструкции ее полной динамики. Для этого, правда, требуется предварительно решить систему уравнений (найти корни – параметры).

Этот метод, разработанный в основном под влиянием физики и успешно в ней применяемый, оказался неприемлемым в случае со многими (в том числе живыми и экономическими) системами (Коган, 1968, с. 213; Гиг, 1981, с. 57) по причине их чрезвычайной сложности (Форрестер, 1971, с. 21) и дискретного характера осуществления жизненных процессов (Жаков, Меншуткин, 1982, с. 56). Решение сложных систем дифференциальных уравнений можно получить только на ЭВМ, приводя их к дискретному виду и осуществляя процедуры последовательных приближений. В этой связи дифференциальные уравнения зачастую стали играть лишь роль удобного способа символьной записи скоростей процессов, протекающих в био- и экосистемах (Федоров, Гильманов, 1980). На следующем шаге эти уравнения перерастают (деградируют?) в дискретную алгебраическую форму имитационных моделей.

Блок-схема аналитической модели представляет собой компактную идеограмму, которая отображает связи компонентов системы друг с другом (рис. 2.1, а). Дифференциальное уравнение, характеризующее динамику какого-либо процесса, может послужить основой для построения имитационной модели.

Автоматные модели (дискретная смена состояния) отличаются, во-первых, дискретным ходом времени, во-вторых, дискретными изменениями значений переменных. Жизнь автомата представляет собой последовательность тактов (временных шагов), за один такт происходит одно изменение состояния автомата (Гильдерман, 1974, с. 403). Весь возможный диапазон изменений переменных определен заранее и разбит на конечное (очень ограниченное) число возможных значений (состояний автомата). Никаких расчетов значений переменных не проводится, переход в новое состояние определяется по матрице перехода в зависимости от текущего состояния модели и от характера внешнего воздействия. Содержание матрицы перехода выступает в роли неизменных параметров модели. Блок-схема автоматной модели (рис. 2.1, б) предстает в виде графа, вершины которого соответствуют возможным состояниям модели, а ребра – направлению переходов, заданных матрицей смежности вершин (Касаткин, 1975, с. 8; Оре, 1980, с. 30; Розенберг, 1984, с. 141). Разработанная теория автоматов позволяет создавать различные аналитические модели, например поведения (Поспелов, 1970). В то же время дискретная структура автоматных моделей может быть использована для составления имитационных моделей.

Стохастические модели (статистические показатели, коэффициенты, уравнения регрессии, главные компоненты, многомерная статистика, анализ временных рядов) используются для описания отдельных компонентов системы и их частных взаимозависимостей, отношений, зависимостей, связей, корреляций, синхронности динамики. В отличие от аналитических моделей регрессионные уравнения выступают в роли частных эмпирических описаний, параметры которых редко имеют биологический смысл, хотя наиболее обоснованы со статистической точки зрения. В отличие от имитационных моделей регрессия не предназначена для описания структуры и динамики многокомпонентных нелинейных систем, но может эффективно описывать отдельные функции (линейные и криволинейные уравнения). Время в явном виде, как правило, вообще не фигурирует в этих моделях. Обычно оно выступает просто источником варьирования изучаемых переменных, когда за период сбора проб объекты изменяют свое состояние (Коросов, 1996, с. 11–17). Главным достоинством стохастических моделей является аппарат оценки статистической значимости параметров моделей и их адекватности исходным данным (Ивантер, Коросов, 1992). В силу особенностей своей конструкции стохастические модели не рассчитаны на описание контуров обратной связи, поэтому их блок-схемы отличаются простотой, они отслеживают прямые зависимости одних переменных от других (рис. 2.1, с). Законы поведения случайных величин могут быть использованы для построения имитационных моделей.

Имитационные модели (программы на языке ЭВМ) сначала получили широкое распространение как средство для поиска численных решений дифференциальных уравнений, а затем как способ создания портретного образа исследуемой системы (Федоров, 1983, с. 18). Цель имитации состоит в том, чтобы изучить взаимодействие между частями системы и ее возможную реакцию на внешние возмущения: “по совокупности элементарных актов, протекающих в определенных условиях, выяснить в целом картину, к которой должны привести эти элементарные акты” (Ляпунов, 1968, с. 103). В силу этого имитационные модели представляют собой наиболее гибкий метод моделирования систем любой сложности, линейных и нелинейных, с обратной связью и сетями управления. Все характеристики, признаки и свойства системы получают численное представление. Среди них те показатели, которые изменяются в процессе “жизни” модели, названы *переменными*. Значения

переменных вычисляются по формулам и могут изменяться вполне плавно. Другие величины, значения которых определяют режимы изменения переменных, остаются, как правило, неизменными; это *параметры* (константы). Термин “дискретные модели” относится к способу задания времени, которое выведено за рамки системы уравнений; имитация его хода организуется программно с помощью специального датчика (Яковлев, 1975, с. 21). Блок-схема дискретной модели помимо представления частей системы должна явно отображать временную (организационную) последовательность их связей друг с другом, т. е. должна быть развернутой по вектору времени (рис. 2.1d). Поскольку все взаимодействия между компонентами модели выражены простым алгебраическим языком (в нашем случае – это формулы Excel), это алгебраические (и даже арифметические) имитационные модели.

Главная особенность имитационных моделей состоит в том, что они “живут” в ЭВМ, что они в определенном смысле самодостаточны. Исходные данные необходимы им далеко не всегда и то лишь как стартовые или опорные значения. Если реальные величины переменных могут быть определены непосредственно в природе, как правило, только для некоторых моментов времени, разделенных более или менее длительной паузой, то модель может вычислить их и для любого промежуточного момента времени. Когда значения переменных невозможно определить в природе в принципе, модельный путь состоит в том, чтобы их вычислять, используя “атмосферу взаимозависимости” реальных переменных.

Для построения имитационных моделей мы использовали три способа математического описания – алгебраический, автоматный, стохастический.

Алгебраические (арифметические) модели в форме алгебраических уравнений выражают предполагаемые соотношения между изучаемыми переменными (признаками) и используются для обобщенного и частного описания динамики наблюдаемых процессов. В таких моделях применяются в основном простейшие арифметические действия (+, -, •, /), аналоги которых встречаются в живой природе (объединение, миграции, размножение, гибель). Модельные формулы могут вовсе не содержать логарифмов, степенных выражений, тригонометрических функций (математические приемы обобщенного описания), предоставляя биологу возможность для содержательной интерпретации рассчитанных параметров. Однако нет препятствий для ввода этих и других операций. Задача подобных моделей состоит в описании всего исследуемого процесса, а также его механизма. Принципиально важно, что записанные буквами уравнения не приходится решать по алгебраическим правилам, их корни вычисляются с помощью внешней процедуры настройки модели (оптимизация). Получается, что модельные уравнения выражают исключительно последовательность арифметических действий и в этом отношении являются арифметическими!

Стохастические модели исследуют сложное поведение случайных величин и для расчетов используют формулы законов распределения. Кроме того, для них требуется специфическая организация расположения частей модели на листе Excel. Объектами настройки в таких моделях выступают параметры распределений – средние, дисперсии, объемы выборок.

Автоматные модели, отражающие дискретные события и поведение, должны содержать логические функции, в первую очередь функцию листа Excel =ЕСЛИ(). Эта функция определяет смену состояний моделируемой системы в соответствии с изменившимися внешними условиями среды. При этом динамика состояния окружающей среды может быть описана алгебраическими моделями. Цель автоматного моделирования состоит в определении критических (штатных) уровней переменных, на стабилизацию которых направлено поведение системы.

Пример с популяцией гадюки

На этапе **Декомпозиция** выбираем для нашего случая (см. раздел **Этапы моделирования**) алгебраическую модель.

В популяционной экологии для оценки репродуктивных параметров популяции по реальным данным повторного отлова меченых животных обычно используются детерминистические модели (Коли, 1979). Простейший метод расчета численности (метод Петерсена) основан на “химической” пропорции, учитывающей “растворение” меченых и выпущенных особей среди остальных членов популяции: доля меченых особей в частной пробе (m/n) равна доле меченых животных во всей популяции (M/N):

$$m/n = M/N,$$

где m – число меченых особей в пробе, n – объем пробы, M – число меченых особей в популяции, N – численность всей популяции.

Отсюда вытекает известная формула для расчета общей численности популяции: $N = M \cdot n / m$,

а также формула для расчета числа меченых особей в отдельной пробе: $m = M \cdot n / N$.

В отношении островной популяции гадюк этот метод частично пригоден (отсутствуют массовые миграции особей с острова), частично не пригоден (в промежутках между отловами животные гибнут). Для того чтобы метод подходил к конкретному случаю, следует модифицировать базовую формулу, введя, во-первых, дополнительную переменную, ежегодную смертность (N_d), во-вторых, развернуть модель по вектору времени. Конструирование модели показано далее.

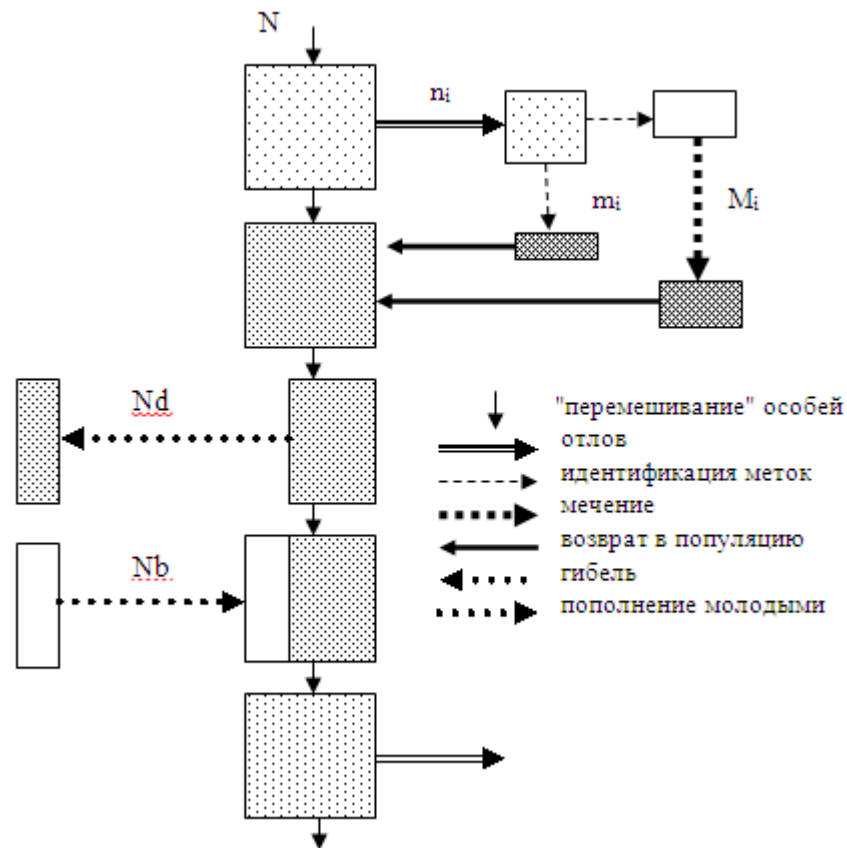
Построение блок-схемы

На этом этапе моделирования можно эффективно воспользоваться простыми правилами, описанными в разделе **Системный подход**. Схема может быть структурно-функциональной (блоками обозначены потоки, а стрелками – функции) или функционально-структурной (блоками обозначены функции, стрелками – потоки).

Пример с популяцией гадюки

Процесс мечения и повторного отлова гадюк представлен в виде структурно-функциональной блок-схемы (рис. 2.2). Она соответствует модели Петерсена, дополненной блоками гибели и пополнения животных. На каждом этапе мечения (в каждый год) из популяции численностью N особей отлавливается выборка из n_i особей. Если это не первый отлов, то среди них может оказаться m_i меченых животных. Остальных M_i немеченых метят, затем всех отпускают. После “перемешивания” концентрация меченых особей в популяции увеличивается. В последующий период до очередного мечения часть животных в популяции погибает (N_d), часть переходит из младших возрастных групп в старшие и начинает встречаться на маршрутах (N_b). Таким образом концентрация меченых особей в популяции уменьшается.

Рис. 2.2. Блок-схема судьбы меченых животных на протяжении одного цикла отловов



Регрессионные и имитационные модели

Для обоих видов моделей основой выступают конкретные характеристики реальной системы. Однако способ вычисления модельных параметров и область применения моделей существенно отличаются в пользу имитационных моделей. Они могут полностью выполнить функции регрессионного анализа; для этого строятся *описательные* модели (когда значения переменных определяются текущим состоянием системы). Кроме того, имитационные модели могут отображать преемственное развитие процесса, т. е. они могут быть *динамическими* (когда значения переменных определяются предыдущим состоянием системы). Эти отличия связаны с конструктивными особенностями моделей разных видов.

Конструкция и настройка

Имитационная модель самодостаточна. Во-первых, она физически существует в электронной среде, она реализуется отдельными формулами для каждого шага, для каждой варианты. Во-вторых, ее параметры не вычисляются, но назначаются – либо самим исследователем, либо *внешней* процедурой оптимизации. Реальные данные используются лишь для придания смысла уже существующим значениям модельной выборки или модельной динамики. По этой причине для настройки модели достаточно всего двух исходных значений – стартового и контрольного.

Регрессионная модель не самостоятельна. Вычисление коэффициентов регрессии обеспечивает виртуальное существование модели (уравнения регрессии): это потенциальный способ описания связи переменных, который может быть и не представлен реально в виде теоретических значений переменных. При этом регрессионная модель целиком и полностью определяется имеющимися наблюдениями, значениями исходных переменных.

Описание и имитации

По определению, статистическая варианта, или значение случайной величины, состоит из двух частей: из доли, привнесенной влиянием на изучаемый объект систематических (регулярных, доминирующих,

регистрируемых, изучаемых) факторов, и доли, добавленной действием случайных (неизвестных) факторов:

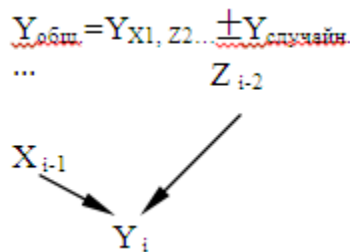
$$Y_{\text{общ.}} = Y_{\text{систем.}} \pm Y_{\text{случайн.}}$$

В центре регрессионного анализа стоит варианта, несущая пару значений, X_i и Y_i , привязанных к текущему i -му моменту наблюдений. Роль систематического фактора отводится независимой переменной X :

$$X_i \rightarrow Y_i \text{ и формула варианты имеет вид: } Y_{\text{общ.}} = Y_X \pm Y_{\text{случайн.}}$$

Это выражение означает, что величина признака X отразилась на величине признака Y , а для каждой варианты – что значение X_i уже полностью сказалось на значении Y_i . История становления значения Y_i не рассматривается, техника расчетов не принимает во внимание никакие предыдущие значения X_{i-1} или Y_{i-1} , участвующие в формировании текущего значения Y_i .

Применяемая в имитационном моделировании процедура настройки модельных параметров снимает практически любые ограничения не темпоральную удаленность влияний. В результате меняется и строение варианты, и строение модели, которая становится динамической. Отдельное значение варианты любой переменной рассматривается как результат влияния предыдущих значений других переменных, причем как известных, так и неизвестных:

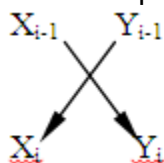


Тем самым каждое значение переменной оказывается обусловленным не только текущим состоянием системы, но всей предшествующей историей ее существования. При этом включение в модель эффектов влияния предшествующих значений переменных происходит без изменения исходной структуры данных. Произвольно изменяя продолжительность задержки влияния одних переменных на другие, можно обнаружить источники этих задержек, в частности, выйти на недостающие звенья в причинно-следственной цепи событий.

Аналогичный анализ можно провести и с помощью регрессионных моделей, смещая временные ряды переменных на один или несколько временных шагов для каждого варианта пересчета модели. Однако для этого каждый раз требуется перестраивать исходную матрицу, что особенно сложно выполнить для многомерных моделей. Пример динамической модели с задержками рассмотрен в разделе **Сети связей**.

Моделирование обратных связей

Существенные преимущества имитационной модели перед регрессионной состоят в возможности учета *синхронного взаимовлияния* переменных, когда предыдущее значение переменной X_{i-1} определяет текущее значение переменной Y_i , и одновременно с этим предыдущее значение переменной Y_{i-1} определяет текущее значение переменной X_i . Речь идет о количественном описании обратной связи.



Здесь регрессионная модель серьезно проигрывает. Чисто технически два уравнения системы ($X_i = f(Y_{i-1})$ и $Y_i = f(X_{i-1})$) много проще получить, настроить и изучить путем имитации, безо всякой перестройки исходной матрицы данных,

тогда как для расчета соответствующих двух уравнений регрессии требуется сформировать два разных массива:

$X_2, \dots, X_i, \dots, X_n$

$Y_1, \dots, Y_{i-1}, \dots, Y_{n-1}$ — для расчета зависимости $X_i = f(Y_{i-1})$ и

$Y_2, \dots, Y_i, \dots, Y_n$

$X_1, \dots, X_{i-1}, \dots, X_{n-1}$ — для расчета зависимости $Y_i = f(X_{i-1})$.

Если переменных больше, чем две, задача учета взаимовлияния требует гигантской вычислительной работы (Славин, 1989, с. 169–190), становясь практически неразрешимой; сеть обратных связей нельзя изучить иным способом, нежели их имитацией. Коэффициент корреляции не может быть выходом из этой сложной ситуации, поскольку он сообщает лишь “силу” зависимости переменных, но не дает формулу преобразования потоков.

Описание данных с пропусками

Сильной стороной имитационной системы выступает регулируемая степень зависимости модели от исходных данных. Имитационная модель “живет своей жизнью”, лишь косвенным путем черпая в исходных данных информацию об уровнях параметров — через сравнение блока расчетных значений переменных и блока аналогичных реальных значений. При этом натурные данные могут содержать пропуски, могут быть даже фрагментарны и отстоять друг от друга на значительные промежутки времени. Все равно, используя технику минимизации функции невязки, можно оценить величины параметров модели.

Более того, построив “самостоятельно живущую” модель, появляется возможность организовать проверку совпадения модели с реальностью раздельно на разных участках рядов наблюдений. Тем самым можно последовательно приближаться к наилучшему решению (оптимальным значениям параметров), постепенно добавляя все новые “точки соприкосновения” модели с реальностью при расчете функции отличий.

Регрессионная же модель требует, чтобы все пары вариант содержали числа. Фрагментарность в данных недопустима. Если в исходных материалах встречаются пропуски, то применяется специальная техника для их заполнения. Например, вместо отсутствующих реальных значений переменных вставляются их средние значения или величины, рассчитанные на базе компонентного анализа (Енюков, 1986, с. 59). В любом случае доля пропусков должна быть невелика, иначе такая модель станет характеристикой способа заполнения пробелов в наблюдениях. Особенно ощутимо для регрессионных моделей отсутствие данных во временных рядах, поскольку при этом утрачивается преемственность явлений и анализ лишается смысла. Необходимый пример показан в разделе **Пропуски в данных**.

Имитация систем без явных данных

С помощью имитационного моделирования можно не откладывать “на потом” анализ остатков, как это практикуется в регрессионном исследовании, но одновременно с настройкой параметров по известным (явным) переменным пытаться изучить и неизвестные факторы, влияющие на систему, введя их в модель как **скрытые** переменные (см. разделы **Описание переменных**). Рассчитанные в процессе моделирования значения этих переменных могут рассматриваться в качестве оценок неких факторов (или их совокупности), идентифицировать которые можно, если сопоставить эти оценки с дополнительными наборами характеристик околосистемной среды. Примеры на эту тему рассмотрены во многих последующих разделах (см. **Скрытые переменные**).

Упрощение модельной структуры

Одно из развитых направлений биометрических исследований с помощью регрессионного анализа — это поиск сложных эмпирических уравнений, как можно более точно описывающих взаимосвязь переменных. В вариационной статистике разработаны достаточно простые методы преобразования исходных

данных, позволяющие в конечном итоге получать модель любого вида. С точки зрения биоэколога, такого рода обобщенные описания служат лишь первым шагом в изучении причин явлений. Дальнейшие шаги должны быть направлены на выяснение причин сложного вида зависимости переменных. Исходные уравнения при этом должны быть декомпозированы на составные компоненты, на более простые уравнения, описывающие отношения между элементами системы. Так удастся установить тонкие механизмы изучаемого процесса (содержащего обратные связи), и, кроме того, во многом упростить вид описывающих его частных моделей. Становится возможным перейти от одного сложного уравнения кривой к нескольким более простым или разделить общее сложное распределение на серию универсальных (например, нормальных), имеющих прозрачную биологическую интерпретацию. Сделать это, оказывается, проще всего с помощью имитационной модели. Эти приемы показаны в разделах **Декомпозиция кривой**, **Анализ распределения**.

Точность описания криволинейных зависимостей

Построение имитационных моделей позволяет упростить процедуру оценки значений параметров криволинейных зависимостей и даже *уточнить* эти значения. Общеизвестно, что коэффициенты регрессии вычисляют с помощью метода наименьших квадратов, который требует *линейной* зависимости между переменными (Браунли, 1977). Если исходные признаки фактически связаны криволинейно, то для получения уравнения криволинейной регрессии эти данные предварительно преобразуют (логарифмируют, извлекают корень, находят обратные значения и т. д.) (Лакин, 1973; Зайцев, 1990 и др.). Выбор вида преобразования зависит от предполагаемой формы связи переменных и в случае сложной зависимости представляет отдельную проблему (Тьюки, 1981, с. 105–106). После этого этапа рассчитываются параметры *линейной* модели для *преобразованных* данных. Наконец, в завершение всего выполняют обратное преобразование, но уже *параметров* модели. Уравнение при этом принимает вид криволинейной зависимости.

Но вот выясняется, что первоначальное преобразование данных не полностью компенсируется последующим обратным преобразованием параметров, при этом накапливается систематическая ошибка (Гильманов, 1987, с. 88). В частности, при расчете параметров степенной функции исходные данные логарифмируются, после чего отыскивается наименьшая сумма квадратов отклонений логарифмированных данных от линии регрессии. Но квадраты логарифмов значений при обратном преобразовании отличаются от квадратов исходных значений, поэтому и местоположение построенной таким способом линии регрессии отличается от динамики групповых средних. В отличие от описанной процедуры настройка параметров имитационной модели не требует предварительного преобразования, она рекомендуется для получения более точных оценок параметров криволинейных моделей любого вида (Zar, 1968; цит. по: Гильманов, 1987, с. 88).

В качестве примера рассмотрим модель из **Введения**. Настройка имитационной модели по исходным данным дает степенное уравнение $N_{м.} = 63.2 \cdot Y^{2.41}$ (табл. 2.2, блок параметров C1:C2). Следующего, пятого, укуса приходится ждать от гадюки под номером 3073. Регрессионная модель, построенная по тем же данным, имеет другие коэффициенты $N_{р.} = 12.3 \cdot Y^{3.68}$ и дает другой прогноз – роковой станет 4555 особь. В поисках источника различий построим блок расчета функции невязки с этими значениями параметров (табл. 2.2, блок E1:F9). Хорошо видно, что регрессионная модель хуже описывает исходные данные, поскольку для нее функция невязки оказалась выше, чем для имитационной модели ($268702 > 117686$). Ясен и источник – третья и четвертая точки данных (квадраты разности, значения ϕ , для первой больше, чем для второй: $185004 > 53879$ и $79333 > 4839.2$).

Что этот результат не является изъяном процедуры настройки параметров, доказывает настройка параметров по предварительно прологарифмированным

данным (т. е. в соответствии с процедурой построения регрессионной степенной модели) (табл. 2.3).

Таблица 2.2. Результаты настройки имитационной и регрессионной моделей

	A	B	C	D	E	F
1	63.2				12.3	
2	2.41				3.68	
3	У_к ус	N экз.	N м.	φ	N п.	φ
4	1	14	63.2	2424.7	12.3	2.9346
5	2	99	337	56544	157	3361.9
6	3	1128	896	53879	697	18600
7	4	1724	179	4839.2	200	79333
8	5		307		455	
9			3		5	
			Φ =	11768	Φ =	26870
				6		2

Таблица 2.3. Имитационная модель по логарифмам данных из табл. 2.2

	A	B	C	D
1	12.2		1.08944	
2	9		4	
3	3.67540			
4	4			
5	Lg(Y)	Lg(N)	Lg(N п.)	φ
6	0	1.14	1.08944	0.00
7		6	4	3
8	0.3	1.99	2.19585	0.04
9	0	6	1	
10	0.4	3.05	2.84305	0.04
11	8	2	7	4
12	0.6	3.23	3.30225	0.00
13	0	7	8	4
14	0.6		3.65844	
15	9		1	
16			Φ =	0.09
17				1

Надо заметить, что логарифмирование изменяет и форму связи, степенное уравнение становится линейным: вместо $Y = aX^b$ получаем $\lg Y = \lg a + \lg X \cdot b$ или в формате Excel $C5 = C\$1 + A5 * C\2 (табл. 2.3, ячейка C5). Настроенные параметры модели в точности соответствуют регрессионным:

$$\lg(N \text{ п.}) = 1.089444 + \lg(Y) \cdot 3.675404 \text{ или}$$

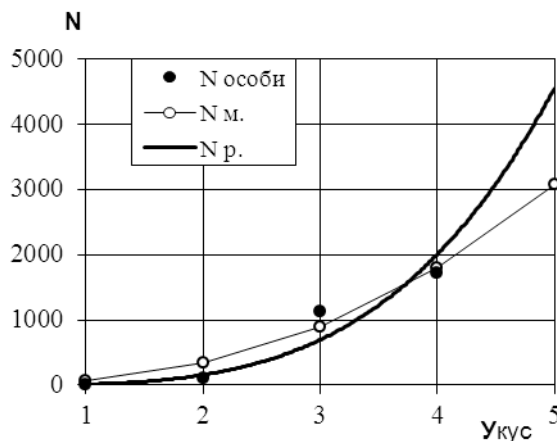
$$N \text{ п.} = 12.29 \cdot Y^{3.675404}, \text{ поскольку } 10^{1.089444} = 12.29 \text{ (ячейка B1).}$$

Сопоставляя результаты “прямого” и “логарифмического” путей определения параметров модели, становится ясным, что во втором случае логарифмы играют роль *весовых коэффициентов*, преувеличивающих роль

небольших величин и преуменьшающих влияние больших чисел на результаты моделирования.

Здесь возникает закономерный вопрос, существуют ли какие-либо теоретические соображения о правомерности такого формального способа взвешивания данных? Автору такие положения неизвестны. Более того, с содержательной (интуитивной?) точки зрения ход модельной кривой выглядит гораздо более убедительным, чем ход линии регрессии (рис. 2.3).

Рис. 2.3. Номера кусающихся особей: данные и модели



Аналитические и имитационные модели

Помимо перечисленных внешних отличий аналитических и имитационных моделей (формула и программа, время как переменная и как счетчик, решение и настройка) уместно будет коснуться еще двух вопросов. Важное сходство этих типов моделей состоит в том, что эти модели могут быть динамическими, с их помощью можно одновременно изучать и скорости протекания наблюдаемых процессов и их результат. Формула имитационной модели в отдельной ячейке электронного листа может отражать приращение некой функции за фиксированный отрезок времени (модельный шаг); по существу это числовой дискретный аналог производной. Пошаговое суммирование всех частных эффектов (аналог интегрирования) дает динамику обобщающей функции (аналог первообразной). И хотя имитационное моделирование родилось как способ численного решения сложных дифференциальных уравнений, сейчас оно переросло в самостоятельный метод, сохранив важнейшие функции дифференциальных уравнений – описание динамических систем.

Еще одно сходство заключается в форме записи. И там и тут мы находим уравнения, т. е. реализацию отношения эквивалентности, или эквиваленции (Свинцов, 1987) между правой и левой частями модельных формул. Однако это сходство лишь внешнее.

Основная цель аналитического моделирования состоит в поиске общего решения системы уравнений, независимо от величины модельных параметров. Отношение эквиваленции играет в этом случае структурообразующую роль, поддерживая жесткие связи между переменными и параметрами в чреде аналитических преобразований формул в процессе поиска корней. Эквиваленция выступает необходимым условием сохранения постоянства заложенных в модель теоретических оснований. Это “внутреннее” средство поддержания сохранности теоретических конструкций. Аналитическое моделирование есть метод теоретического исследования.

При имитационном моделировании, напротив, конструкция модели, по определению, должна быть не постоянной, но податлива для произвольных изменений. По этой причине знак равенства в формулах обслуживает другое отношение эквиваленции, а именно: между реальными эмпирическими данными

и модельными параметрами (посредством модельных переменных). Цель имитационного моделирования состоит в поиске частного решения системы модельных уравнений, т. е. в подборе таких значений параметров, которые обеспечивают объяснение существа наблюдаемого конкретного явления. Эквиваленция здесь играет роль критерия, устанавливающего соответствие между конструкцией модели и организацией реального объекта. В этом качестве отношение эквиваленции используется как “внешнее” средство для объяснения частных явлений. Имитационное моделирование есть метод эмпирического исследования.

По мнению автора, этот тезис очень важен для биологов. Он показывает, что имитационное моделирование не есть приложение высшей математики. По способу записи модели – это алгебра, по способу расчетов – арифметика. Четырех арифметических операций во многих случаях оказывается достаточно для воплощения предположений биологов о характере функционирования изучаемого объекта. Реализацию отмеченного отношения эквивалентности принимает на себя имитационная система. Такая ситуация снимает с биологов ответственность за известную математическую неграмотность и позволяет с легким сердцем пользоваться эффективным количественным инструментом познания, арифметическим моделированием.

Конечно, автор не призывает “кухарку управлять государством”, но лишь, как водится, – *учиться “управлять”*. К сожалению, такая позиция, по опыту автора, не находит отклика в сердцах математиков. Видимо, коллизии между “математиками от биологии” и “биологами от математики” достаточно обычны. Иначе чем объяснить такую настоятельную рекомендацию автора одной из современных монографий по имитационному моделированию: “Совершенно недопустима демонстрация превосходства и тем более высокомерия, свойственная отдельным математикам...” (Угольницкий, 1999, с. 57)?

Конечно, знание математики не может помешать построению моделей. Но в любом случае лучше на первых порах ошибаться в попытках построить модель, чем в угоду предвзятым специалистам вовсе отказываться от этого метода исследования биосистем.

Переменные и параметры

Перед составлением системы модельных уравнений следует тщательно описать входящие в их состав численные показатели. Все количественные характеристики модели делятся на переменные величины и константы (параметры).

Переменные характеризуют среду, окружающую изучаемую систему, изменчивое состояние самой системы в целом и ее элементов в частности. В терминах блок-схемы переменные – это потоки, количественное проявление способа существования статических компонентов системы (см. раздел **Идеология моделирования**). Отдельный элемент может быть описан несколькими переменными. Поэтому говорят о разных языках описания системы: каждый язык относится к одному виду потоков данного уровня иерархии. Потоки преобразуются в процессе функционирования системы: либо изменяют свою величину, либо трансформируются в другие потоки.

Параметры количественно выражают режимы (скорость, интенсивность) преобразования потоков (изменения значений переменных). В отличие от переменных величин параметры обычно задаются неизменными, во всех модельных расчетах они остаются независимыми от состояния системы (другое дело, когда они изменяются, “оптимизируются”, в процессе настройки модели).

Установление состава изучаемых признаков – это первый шаг в их описании (Методы..., 1980, с. 58). Чем больше привлекается переменных, тем ближе к оригиналу будет протекать “жизнь” модели, напротив, чем меньше переменных, тем проще оказывается объяснить причинно-следственные связи между элементами системы, и тем легче будет настроить модель. Противоречие “простота – полнота” модели (Перегудов, Тарасенко, 1989, с. 281) заставляет на

разных этапах стремиться к оптимальной конструкции. Во-первых, на стадии составления блок-схемы соображения о существовании биологического процесса позволяют отобрать переменные, исходя из иерархической композиции объекта (Безель, 1987, с. 23). Во-вторых, в процессе настройки модели приходится улучшать модель путем добавления или исключения переменных с контролем ее адекватности системе-оригиналу (Дженнрич, 1986, с. 78). Формальным методом для этого служит оценка статистической значимости параметров, которая позволяет достаточно обоснованно изменять их состав. В арсенал эффективных приемов исправления модели входят агрегация переменных (Акоф, Сасиени, 1971, с. 105), декомпозиция исходных переменных на две или более составляющих, конструирование скрытых переменных (Страшкраба, Гнауик, 1989, с. 53). При организации матрицы исходных данных важно следить за тем, чтобы число строк (объектов наблюдения, шагов динамики) было минимум в два раза больше числа столбцов (признаков, контролируемых явных переменных).

Природа переменных

На втором шаге выясняется происхождение и тип изучаемых величин (Акоф, Сасиени, 1971, с. 104). Измеренные в разных шкалах (номинальной, порядковой, интервальной) признаки могут быть дискретными (класс, ранг, балл) или непрерывными (шкалы отношений и абсолютные шкалы). В целях упрощения описания модели часть непрерывных переменных зачастую переводят в дискретную форму. Так, преобразование плавной периодической функции в форму дискретной (прямоугольной) функции во многом упрощает анализ динамики процесса, почти не снижая его точности (Мерсер, 1964, с. 148). Гладкую кривую нормального распределения непрерывного признака переводят для упрощения расчетов в дискретную форму, основанную на интервальной шкале (Ивантер, Коросов, 1992, с. 21-22). В других случаях возникает задача преобразования переменных. Многие химические, биохимические и физиологические показатели имеют логнормальное распределение (Израэль и др., 1985, с. 10; Безель, 1987, с. 26). Логарифмирование этих значений дает нормальное распределение, что более удобно для статистического анализа.

Определение состава выборок

Третий шаг – это формирование массивов исходной информации (совокупности вариантов, множества объектов, серия шагов). Эти данные количественно выражают состояние изучаемой системы в разные моменты времени, в разных точках пространства, при разных внешних воздействиях. Состав выборки, следовательно, отражает те направления изменчивости, с учетом которых должны строиться и модель взаимоотношений между переменными, и интерпретация результатов моделирования. В массивах биологических данных отчетливо выделяются несколько общих направлений изменчивости, вызывающих корреляцию между признаками (Мина, 1975, с. 177; Мина, Клевезаль, 1976, с. 10-12; Коросов, 1996, с. 11-17):

- множество значений характеризует одну особь (объект) в разные моменты ее жизни, в разных состояниях (индивидуальная, онтогенетическая или временная, аллометрическая динамика),
- множество значений характеризует индивидуальные особенности особей одного возраста (статуса) (внутригрупповая, аллометрическая изменчивость),
- множество значений характеризует особей разного возраста (межгрупповая или “групповая онтогенетическая” изменчивость, динамика),
- множество значений характеризует отличия особей из разных популяций (межпопуляционная, географическая или пространственная изменчивость),
- множество значений характеризует отличия особей из серии выборок, испытывающих действие разных доз внешнего фактора (факториальная изменчивость).

Исследование природы переменных создает основу для последующих этапов моделирования в виде двух таблиц: в одной представлена характеристика всех переменных, в другой – исходная матрица (таблица) реальных значений

переменных, полученных в природе или эксперименте. Структура будущей модели во многом предопределяется структурой матрицы эмпирических данных.

Типы переменных

Не все модельные переменные играют одинаковую роль в построении модели. Обычно указывают на три функционально различных группы (Акоф, Сасиени, 1971, с. 81; Суходольский, 1972, с. 373; Славин, 1989, с. 43-44; Страшкраба, Гнау, 1989, с. 53-54; Боровиков, Боровиков, 1997, с. 530):

- *независимые* (управляющие, движущие, вынуждающие, экзогенные);
- *зависимые* (управляемые, состояния, местоположения, эндогенные);
- *скрытые* переменные (неизвестные, внутренние, фиктивные, латентные).

В этой классификации нет четкой грани между исходными переменными (наборами значений, полученных в природе) и модельными переменными (наборами значений, вычисляемых в процессе моделирования). К тому же более удобно, на наш взгляд, представление типов переменных в форме дихотомического дендрита (Угольницкий, 1999, с. 18). Модифицированная типология переменных в рамках имитационной модели выглядит следующим образом (рис. 2.4).



Рис. 2.4. Классификация переменных в имитационных моделях

Наблюдения поставляют ряды чисел – **реальные** (*внешние*, исходные, эмпирические, наблюдаемые, неуправляемые) переменные. По отношению к объекту исследования некоторые переменные, характеризующие среду, будут *независимыми* (например, суточная динамика температуры), другие, характеризующие состояние объекта, фактически, причинно оказываются *зависимыми* от первых (например, уровень метаболизма животных). Значения внешних переменных в пределах имитационной системы никак не изменяются, они характеризуют состояние среды, влияющей на объект исследования, и состояние этого объекта. Эти значения получены эмпирическим путем и известны еще до начала моделирования (хотя обычно модель требует проведения дополнительных исследований). Из них формируется матрица эмпирических данных.

В рамках имитационной системы независимые реальные переменные служат основой для расчета зависимых *модельных* переменных. Зависимые реальные переменные служат основой для настройки параметров модели, а также для оценки близости модели к прототипу.

Уравнениями имитационной модели вычисляются другие ряды чисел – **модельные** (*внутренние*, расчетные, управляемые) переменные; при этом используются значения реальных независимых переменных, а также значения других модельных переменных. Если каждой из них в действительности соответствует реальная зависимая переменная, количественные наблюдения, то эти модельные переменные следует назвать *явными*. Они используются для настройки параметров модели и для проверки ее адекватности.

Помимо названных, имитационные модели содержат *скрытые* переменные (*скрытые модельные* переменные). Такое название появилось потому, что эти характеристики не регистрируются в природе и в силу этого в рамках модели скрыты от прямого контроля со стороны реальных данных. Для них не существует аналогов в рядах натурных наблюдений из-за невозможности, сложности или высокой стоимости сбора данных (хотя предполагается, что

такой характеристикой природный объект обладает). Смысл введения этих гипотетических переменных состоит в попытке оценить значения тех экологических факторов или внутренних функций биосистем, которые трудно изучить непосредственно. Эти переменные играют важнейшую роль в моделировании, поскольку выступают заместителями неизвестной реальности, служат способом анализа природных явлений, дают возможность глубже проникнуть в структуру исследуемой системы. Латентные переменные всегда оказываются вычисляемыми, т. е. формально зависимыми. Однако чаще всего интересен как раз тот случай, когда оцениваются значения переменной, внешней по отношению к изучаемому объекту, определяются уровни неизвестного фактора среды, т. е. значения, по существу, независимой переменной.

Наряду с важной гносеологической функцией, использование скрытых переменных таит большую опасность заблуждения. Дело в том, что ввод большого числа бесконтрольных компонентов модели приводит к ее полной (псевдо-) адекватности исходным данным, поскольку при неправильной организации модели внутренние переменные могут забирать на себя всю необъясненную изменчивость (см. раздел **Скрытые переменные**).

Описание параметров

Блок-схема, предопределяющая структуру модельных переменных, диктует и список модельных параметров. Определиться с составом параметров необходимо в первом приближении еще перед построением модельных формул, выполняя главное правило моделирования (см. следующий раздел). Модельные параметры сосредотачивают на себе весь смысл построения модели, это квинтэссенция теоретических представлений о динамике реальной системы, главная характеристика способа ее существования. Поэтому формирование блока параметров – самый ответственный этап в моделировании. Число параметров обычно меняется в процессе создания формул и при доводке модели до уровня приемлемой адекватности, однако общие ориентиры следует наметить заранее.

Кроме теоретических соображений о структуре реальной системы росту сложности модели препятствует ряд формальных ограничений. Вообще, чем больше уравнений и параметров представляют систему, тем более точной, “портретной”, оказывается модель (Розенберг, 1984, с. 30, 95; Перегудов, Тарасенко, 1989, с. 281–282). Однако общее пожелание формулируется так: “поменьше параметров”. Во-первых, не может быть эффективной настройка модели, содержащей более 10 параметров; в этом случае увеличивается количество разных решений одной и той же системы уравнений, а методов формального выбора лучшего решения не существует (Страшкраба, Гнаук, 1989, с. 56). Кроме того, для обеспечения достаточной точности имитации (ошибка прогноза менее 100%) в среднем каждое уравнение должно содержать менее 4-х параметров (Федоров, 1983, с. 21). Эти ограничения не означают, что модель должна быть “маленькой”, т. е. должна описывать ограниченное число переменных. Если каждый из модельных коэффициентов контролирует аналогичные процессы, происходящие в разных точках пространства с разными объектами или в разные моменты времени, то число переменных может достигать десятков и сотен. В частности, всего 4 параметров было достаточно для имитации динамики загрязнения в 16- камерной модели крупного водоема и лишь ограниченное число точек отбора проб в этом случае помешало увеличить число камер до нескольких десятков, причем без увеличения числа параметров (см. раздел **Скрытые переменные**).

Если система-оригинал имеет сложную структуру, то проблему включения многих параметров обходят, используя блочный принцип составления модели. Всю систему разбивают на относительно независимые блоки, состоящие из небольшого числа уравнений, отдельно настраивают параметры, а затем связывают в общую модель, часто через интерфейсные (внешние) переменные

(Левченко, 1993, с. 311; Иванищев и др., 1989, с. 53–55). Пример такого синтеза показан в разделе **Сети связей**.

Еще один эффективный подход к описанию сложных систем состоит в том, чтобы последовательно строить преемственный ряд все более сложных моделей до момента удовлетворительного описания реальной системы (Галицкий, 1981, с. 104). При первоначально кажущейся большой сложности объект удастся смоделировать, затратив меньшие усилия. Этот способ характерен для большинства представленных примеров.

Вообще говоря, чрезмерное количество параметров свидетельствует, скорее всего, о плохо продуманной структуре модели, слабо структурированной блок-схеме и смешении элементов разного уровня иерархии.

Помимо сказанного следует учитывать ограничение, связанное с неполнотой исследования, с обычным для экологии невысоким качеством натурных наблюдений (речь идет о “невоспроизводимости” наблюдений, а не о небрежности экологов). С одной стороны, значения параметров и строение модели подгоняются под выборочные данные, которые, понятно, есть лишь тень исследуемого объекта, но не он сам (Прицкер, 1987, с. 18), тогда как модель должна копировать не просто выборку, но объект исследования. С другой стороны, выборочные данные могут быть настолько фрагментарны, неполны, “загрязнены” стохастическим шумом, что настройка системы уравнений становится неразрешимой задачей (Страшкраба, Гнау, 1989, с. 56). Тем не менее лучше построить неказистую модель, в попытке активизировать дополнительные размышления о своем объекте исследования, нежели пренебречь этой возможностью под гнетом тяжелого впечатления от неточных описаний: “атмосфера взаимосвязей” изученных показателей не менее информативна, чем конкретные числа!

Математическое описание модели

К этому этапу создания модели уже построена блок-схема преобразования потоков, описаны сами потоки (переменные), намечен состав параметров. Остается определить, каким образом эти потоки преобразуются, как система переходит из одного состояния в другое. Конструирование уравнений есть процесс установления соотношений между переменными и параметрами. Отдельные уравнения (математические выражения) модели соответствуют блокам блок-схемы и описывают процессы преобразования одних потоков в другие (обмен, передача, функциональная связь, темп) (Форрестер, 1971, с. 42; Меншуткин, 1993, с. 8).

Составление блок-схемы и воплощение ее в систему уравнений есть по существу “навязывание” представлений исследователя о происходящих процессах – структуре модели (Розенберг, 1984, с. 95). Цель моделирования в том и состоит, чтобы соотнести наши предположения, выраженные в форме модели, с поведением объекта исследования. Логика оформляется количественно и сопоставляется с реальностью. Моделирование как раз и “является эффективным средством исследования структуры”, направленным на то, чтобы стали “понятны причины связи между управляемыми и неуправляемыми переменными” (Акоф, Сасиени, 1971, с. 114, 83).

Имитационная модель может описывать как выборку объектов, так и временную динамику системы. В последнем случае важно определиться с порядком вычислений. Для этого организуется дискретный (чаще всего равномерный) ход времени, когда реально непрерывное течение времени разбивается в модели на множество отдельных моментов (периодов, промежутков). Реальные данные могут и не соответствовать каждому из этих шагов модели, важно другое, чтобы динамика реальной системы в целом отображалась в электронном аналоге. В течение одного промежутка времени (характерное время процесса) происходят модельные *события* – одновременное изменение всех переменных при условии постоянства внешних воздействий (Ляпунов, 1968, с. 70; Мастецкий, 1997, с. 10). Шаг за шагом система переходит

из одного состояния в другое, текущее определяется предыдущим. Это значит, что в расчетах нового значения каждой переменной на i -м временном шаге используются те значения определяющих их переменных, которые они принимали на предыдущем, $i-1$ -м временном шаге:

$$Y_{1i} = a \cdot X_{1i-1} + \dots;$$

$$Z_{1i} = b \cdot Y_{1i-1} + \dots; \dots$$

Все вычисления, проводимые на данном временном шаге, оказываются независимыми друг от друга, а полученные значения – постоянными в течение этого периода. Такая пошаговая организация вычислений снимает проблему заикливания (циклических ссылок между ячейками Excel). Техника составления формул и организации временной развертки показана в следующих разделах.

Главное правило моделирования

Как и при аналитическом моделировании, главная проблема создания имитаций состоит в конструировании модельных формул. Однако на этом их сходство и заканчивается. Цель аналитической модели – выразить закон явления. Цель имитационной модели – рассчитать параметры механизма конкретного явления. Поэтому требования к формулам имитационных моделей весьма специфичны – они обязательно ориентированы на фактический материал. Имитационная модель должна не просто математически описать объект, но описать его таким образом, чтобы вычислялись те же самые переменные, что наблюдались. На листе Excel нужно воспроизвести такую же структуру массива зависимых модельных переменных (Y_{mj}), что имеет массив зависимых реальных переменных (Y_j). Это необходимо для последующей настройки модели. Понятно, что этот расчет должен базироваться на основе тех параметров ($a_1, a_2, \dots$), которые (на наш взгляд) и определяют содержание (механизмы) наблюдаемых явлений. Иначе говоря, модельная формула должна, во-первых, выражать суть механизма явления, во-вторых, рассчитывать аналог матрицы исходных данных.

В этом и заключается главное правило составления модельных формул:

**выразить известные переменные (Y_j)
через неизвестные параметры (a_j).**

Как способов, так и технических приемов конструирования моделей может быть очень много. И все же личный опыт автора показывает, что широкий спектр экологических явлений можно количественно описать, опираясь лишь на три раздела математики: математическую статистику, алгебру, теорию графов. При этом большинство описаний использует именно алгебру, основы которой заложены в старших классах школы и понятны любому биологу. (В скобках надо заметить, что речь здесь идет не о примитивности описаний, а о создании основ для роста интереса к моделированию, к его использованию хотя бы в простых формах).

Пример с популяцией гадюки

В состав данных по повторному отлову меченых гадюк входят известные переменные m_i, M_0, n_i . При этом переменные M_0, n_i являются независимыми, а m_i – это реальная зависимая переменная. Модельные уравнения должны вычислять именно эту переменную (m'_i) для того, чтобы иметь возможность сравнивать с реальностью (m_i). В качестве неизменных параметров должны выступать значения начальной численности популяции N_0 и уровень ежегодной смертности N_d ; их предстоит подобрать в процессе настройки. Для упрощения задачи примем ежегодную численность и смертность в островной популяции гадюки неизменными, $N = \text{const}$, $N_d = \text{const}$. Теперь осталось построить модельные формулы.

Приемы составления формул

Формулы имитационной модели вычисляют значения модельных переменных, для которых имеются (явные переменные) или отсутствуют (скрытые переменные) аналоги в массиве исходных данных (среди зависимых

переменных). Существует два принципиально разных пути модельного отображения реальности (два типа имитационных моделей): описание процесса в целом (описательная модель) и характеристика скорости процесса (динамическая модель).

Описательная модель выражается одним уравнением, которое охватывает процесс целиком и позволяет непосредственно вычислить значения переменных при любых значениях независимой переменной (в любой момент времени). По своим результатам такое моделирование подобно регрессионному анализу. В зависимости от целей и наличных данных можно выделить три конструкции обобщенной модели (табл. 2.4).

Таблица 2.4. Способы обобщенного описания зависимостей

Индекс	Тип зависимости	Тип данных	Общий вид модели
O1	Зависимость переменной от времени	Временной ряд	$Y = f(A, T)$
O2	Зависимость одной переменной от другой	Выборка сопряженных значений	$Y = f(A, X)$
O3	Зависимость процесса изменения одной переменной от других	Несколько временных рядов сопряженных значений	$Y = f(A, X_1, X_2, \dots, T)$

Обобщенное описание зависимости процесса от времени (O1) можно выразить формулой:

$$Y = f(A, T) \text{ (например, } y_i = a \cdot i + b),$$

она рассматривает два ряда исходных значений: ряд временных шагов (T):

$$i = 1, 2, \dots, T,$$

а также временной ряд значений переменной, изменяющейся с течением времени:

$$Y = \{y_1, y_2, \dots, y_i, \dots, y_T\}.$$

В этом случае речь идет о *ряде значений*, упорядоченных *во времени*. Для задания формы соотношения $f()$ можно воспользоваться самыми различными конструкциями, в числе которых наиболее простые – это линейное, степенное, экспоненциальное уравнения зависимости, среди более сложных – параболическое, логистическое (кривая Ферхюльста) (табл. 2.5).

Пример такого рода (зависимость размеров тела от возраста) можно найти в разделе **Адекватность и значимость**. Понятно, что на базе основных типов зависимости можно сконструировать формулы любой сложности с любым числом коэффициентов.

Таблица 2.5. Уравнения зависимости процесса от времени

Тип зависимости	Уравнение $f()$
Линейная	$y_i = a \cdot i + b$
Степенная (аллометрическая)	$y_i = b \cdot i^a$
Экспонента	$y_i = b \cdot e^{ai}$
Парабола (i-й степени)	$y_i = a_0 + a_1 \cdot i + a_2 \cdot i^2 + \dots + a_j \cdot i^j$
Логистическая кривая	$y_i = C + A / (1 + e^{a \cdot i + b})$
Примечание: a , b – коэффициенты пропорциональности, j – номер переменной	

Обобщенное описание зависимости одной переменной от другой (O2) можно выразить формулой:

$$Y = f(A, X) \text{ (например, } y_i = a \cdot x + b),$$

она рассматривает два ряда исходных значений – выборку зависимой переменной (Y):

$$Y = \{y_1, y_2, \dots, y_n\},$$

и выборку независимой, заданной переменной (X):

$$X = \{x_1, x_2, \dots, x_n\},$$

$$i = 1, 2, \dots, n.$$

В этом случае речь идет об изучении *выборки вариантов* (объемом n), несущих по два признака. Для задания формы соотношения $f()$ между переменными можно воспользоваться теми же уравнениями, включая полиномиальную (табл. 2.6). Простой пример такого рода модели приведен во **Введении**.

Описание зависимости процесса *изменения одной переменной от нескольких* (ОЗ) есть обобщение первых двух случаев. Такой пример можно найти в разделе **Сети связей**.

Все рассмотренные уравнения характеризуют изучаемое явление в целом, не вникая в нюансы его динамики. Обычно для описания экологических явлений приходится пользоваться моделями сложного вида (см. раздел **Аппроксимация кривой**). Предлагаемые приемы помогут быстро и легко найти *точные* параметры общих уравнений любой сложности.

Таблица 2.6. Уравнения зависимости двух переменных

Тип зависимости	Уравнение $f()$
Линейная	$y_i = a \cdot x + b$
Степенная (аллометрическая)	$y_i = b \cdot x^a$
Экспонента	$y_i = b \cdot e^{ax}$
Парабола (j-й степени)	$y_j = a_0 + a_1 \cdot x + a_2 \cdot x^2 + \dots + a_j \cdot x^j$
полином	$y_{ji} = a_0 + a_1 \cdot x_{1i} + a_2 \cdot x_{2i} + \dots + a_{ji} \cdot x_{ji}$
Логистическая кривая	$y_i = C + A / (1 + e^{a \cdot x + b})$
Примечание: a, b – коэффициенты пропорциональности, j – номер переменной, n – объем выборки, $i=1, 2, \dots, n$	

Преимущество имитационного моделирования перед регрессионным анализом раскрывается при разделении сложных зависимостей на слагаемые. Вообще говоря, любая криволинейная связь двух переменных является отражением *процесса формирования* их значений. Поскольку непосредственные явления мира биоты (притяжение – отталкивание; добавление – утрата) реализуются всегда линейно, криволинейность связи появляется при объединении нескольких линейных процессов, а сложная криволинейность – при их дальнейшей интеграции. Имитация позволяет найти эти простые составляющие, раскрывая тем самым источники их формирования (см. раздел **Декомпозиция кривой**).

Динамическая модель служит для описания скорости и результата процесса, она характеризует причинный механизм его протекания, поскольку дает возможность сконцентрировать внимание на исследовании каждого момента осуществления процесса. В отличие от единственного уравнения описательных моделей динамические имитационные модели представлены как минимум двумя уравнениями. Одно из них для каждого шага позволяет рассчитать только сиюминутный результат процесса. Второе служит для аккумуляции частных результатов, полученных на всех предыдущих шагах. На листе Excel полная модель занимает блок ячеек. Получается, что для расчета значений в заданный конкретный момент времени (на данном шаге, в отдельной ячейке Excel) требуется воссоздать ход всего процесса с самого начала. В зависимости от целей исследования и наличных данных можно говорить о четырех основных конструкциях динамических моделей (табл. 2.7).

Таблица 2.7. Способы описания скоростей изменения переменных

Индекс	Тип зависимости	Тип данных	Общий вид модели
--------	-----------------	------------	------------------

C1	Зависимость текущего значения переменной от предыдущего значения этой же переменной	Один автономный процесс	$Y_i = f(A, Y_{i-1}, T)$
C2	Зависимость текущего значения переменной от предыдущего значения другой переменной	Один зависимый процесс	$Y_i = f(A, X_{i-1}, Y_{i-1}, T)$
C3	Зависимость текущих значений серии переменных от предыдущих значений соседних переменных	Серия зависимых процессов, обратная связь	$Y_i = f(A, X_{i-1}, T)$ $X_i = f(B, Y_{i-1}, T)$
C4	Зависимость текущих значений переменных и параметров от предыдущих значений этих же или соседних переменных	Серия зависимых процессов (с двумя контурами обратной связи)	$Y_i = f(A_i, Y_{i-1}, T)$ $A_i = f(C, Y_{i-1}, T)$

В случае автономной, *самообусловленной*, модели текущее значение модельной переменной определяется ее предыдущим значением (текущее состояние моделируемой системы определяется лишь предыдущим ее состоянием) (табл. 2.7, C1). Модель комплектуется из двух формул. Одна из них вычисляет частный результат процесса, достигнутый в каждый отдельный момент времени (скорость процесса):

$$dy_i = a \cdot y_{i-1} + b;$$

другая формула ответственна за интеграцию всех частных результатов в один общий:

$$y_i = y_{i-1} + dy_i,$$

где

y_i – текущее значение переменной (характеристика состояния),
 y_{i-1} – значение переменной на предыдущем ($i-1$ -ом) временном шаге (характеристика предыдущего состояния),
 dy_i – прирост переменной за i -й временной шаг,
 a, b – коэффициенты пропорциональности,
 $i = 1, 2, \dots, T$ – индекс временных шагов модели.

Член dy_i аналогичен производной, вычисленной для каждого момента времени (скорость изменения x на данном шаге i), а член y_i подобен первообразной (результата изменения переменной к данному моменту).

С точки зрения статистической значимости модели одни и те же исходные данные могут быть одинаково хорошо описаны и формулами *обобщенного описания*, и формулами *скорости процесса*, но во втором случае модельные параметры имеют гораздо более прозрачную биологическую интерпретацию и могут эффективно сравниваться друг с другом. Пример можно найти в разделе **Пропуски в данных**, где рост рассматривался как автономный процесс.

Иную ситуацию описывает модель *обусловленности средой*, когда текущее значение модельной переменной определяется, во-первых, предыдущим значением (предыдущим состоянием моделируемой системы), во-вторых, значением внешней переменной (табл. 2.7, C2). Таким образом может быть описано влияние среды на внутреннее состояние системы или воздействие одной переменной на другую переменную системы. Число формул увеличивается:

$$d_1 y_i = a \cdot y_{i-1} + b;$$

$$d_2 y_i = c \cdot x_{i-1} + d;$$

$$dy_i = d_1 y_i + d_2 y_i;$$

$$y_i = y_{i-1} + dy_i,$$

где d_1y_i , d_2y_i – приросты переменной за счет внутренних потенций системы и в результате внешнего влияния, a , b , c , d , e , f – коэффициенты пропорциональности.

Такого рода модели очень сложно выразить в форме обобщенных аналитических выражений, тем более с помощью регрессионного анализа. Однако в рамках имитационной системы они имеют очень простое оформление. Этот тип модели был использован для имитации автономного роста численности популяции гадюки, на который накладывается влияние среды в виде гибели особей по разным причинам (см. сквозной пример).

Важно отметить, что эта конструкция позволяет очень просто моделировать динамику системы с *запаздыванием* ее отклика на внешние воздействия путем простого изменения индекса (ссылки) y переменной. Так, для учета влияния пред-предыдущего значения средовой переменной в предыдущую модель вносятся следующие изменения или дополнения:

$$d_3y_i = e \cdot x_{i-2} + f;$$

$$dy_i = d_1y_i + d_2y_i + d_3y_i.$$

Пример модели такого типа представлен в разделе **Сети связей**, где изучается влияние отдаленных (позапрошлогодных) событий на численность групп взрослых особей.

Модель *взаимообусловленности, обратной связи* (табл. 2.7, С3), рассматривает ситуацию, когда текущее значение одной модельной переменной определяется предыдущим значением другой переменной и одновременно текущее значение второй переменной определяется предыдущим значением первой:

$$dx_i = a \cdot y_{i-1} + b;$$

$$dy_i = c \cdot x_{i-1} + d;$$

$$y_i = y_{i-1} + dy_i,$$

$$x_i = x_{i-1} + dx_i,$$

где x_i , y_i – две модельные переменные, a , b , c , d – коэффициенты пропорциональности.

Здесь речь идет о системе с обратной связью, для которой общего аналитического решения получить нельзя, но можно исследовать ее динамику численными методами – с помощью моделирования (см. раздел **Усреднение или параметризация?**).

Во всех рассмотренных случаях параметры представленных моделей задавались неизменными на всех шагах процесса. Так в модели воплощается упрощенное и заведомо неверное предположение о неизменности скорости жизненных процессов биологической системы-оригинала при любых ее состояниях. Однако совершенно очевидно, что реальные прототипы параметров моделей не обладают таким постоянством и во многом определяются текущим состоянием системы. В связи с этим в модель иногда имеет смысл вводить *переменные параметры* – параметры, изменяющиеся либо автономно, либо зависящие от некоторых переменных моделируемой системы (табл. 2.7, С4). С точки зрения строения имитационной системы можно говорить о создании новой скрытой переменной, которая влияет на некоторую явную переменную. Модель с *автономными переменными параметрами* будет иметь следующий вид:

$$da_i = c \cdot a_{i-1} + d;$$

$$dy_i = a_i \cdot y_{i-1} + b;$$

$$y_i = y_{i-1} + dy_i,$$

где a_i , a_{i-1} – текущее и предыдущее значения параметра, da_i – прирост параметра на одном временном шаге.

Как выясняется, такие модели много более точно описывают динамику системы, учитывая, например, возрастные изменения скорости накопления тяжелых металлов (Медведев, Коросов, в печати), или возрастное *изменение скорости* потребления пищи (см. раздел **Пропуски в данных**).

В рассмотренных выше модельных конструкциях скорость выражена *линейной* функцией. Для имитации динамики сложной многокомпонентной

системы рекомендуется использовать именно наиболее простые виды соотношений (Страшкраба, Гнауик, 1989, с. 30; Боровиков, Боровиков, 1997, с. 530). В то же время при необходимости можно взять любую другую форму зависимости, например, из числа приведенных в табл. 2.6.

На базе простейших типов моделей, которые были рассмотрены выше, можно построить модели, сочетающие различные их комбинации и усложнения. Возможно, одну из самых сложных конструкций будет иметь модель, учитывающая *два контура* обратной связи (сочетание C2, C3 и C4). Первый контур – это тот способ, с помощью которого текущие значения одних переменных контролируют будущие значения других переменных. Второй контур – это способ изменения величины параметров, зависящих от достигнутых уровней внутренних и внешних переменных. Первая петля обратной связи через посредство параметров регулирует уровни переменных, а вторая петля регулирует параметры регуляции переменных. Такую сложную структуру имеют биологические явления запаздывания, насыщения, скачки, пороги реакции, мертвые зоны, ограничения, петли гистерезиса (Страшкраба, Гнауик, 1989, с. 30; Гиг, 1987, с. 613; Безель, 1987, с. 20; Николис, Пригожин, 1979, с. 71; Волькенштейн, 1981, с. 464). При этом граница между параметрами и переменными размывается и архитектура модели приближается к виду *сети обратных связей* (Новосельцев, 1989, с. 67). Автор избегал построения моделей такого типа ввиду их полной беззащитности перед лицом любой критики.

Пример с популяцией гадюки

Для моделирования динамики отлова меченых животных использовалась динамическая модель второго типа (табл. 2.7, C2).

В соответствии с блок-схемой модель предполагает, что в каждый момент отлова (*i*-й год) численность равна N , а между двумя этапами отлова популяция теряет N_d особей; в тот же момент столько же особей добавляется из числа молодых (N_b). Эту идею можно выразить алгебраической формулой с условиями:

$$N_i = N_{i-1} - N_d + N_b, \quad N_i = N_{i-1} = N, \quad N_d = N_b \quad (i=1991, 1992, \dots, 1998).$$

Чтобы перейти к наблюдаемым в природе величинам, нужно их выразить через популяционные параметры. Предположим, что... (нет, нельзя писать на математическом сленге!). Как было сказано выше, в пробах каждого (*i*-го) года подсчитывалось число особей, помеченных в предыдущем году; это число равно m_i . Его и надо выразить формулой. Нам известно и другое значение – число особей, помеченных и выпущенных в предыдущем *i*-1-м году; оно равно M_{i-1} . Далее найдем, сколько меченых особей выжило к текущему *i*-му году (M_i). Так, если число погибших составляет N_d , то их доля во всей популяции – N_d/N . Когда метки достаточно хорошо перемешаны в популяции, то доля меток, потерянных за счет гибели, будет равна доле погибших особей (N_d/N), и общие потери меток между текущим и предыдущим (*i*-1-м и *i*-м) годами равно $(N_d/N) \cdot M_{i-1}$. Понятно, что число сохранившихся меток в этот (*i*-й) год составит:

$$M_i = M_{i-1} - (N_d / N) \cdot M_{i-1}.$$

Используя пропорцию Петерсена, выразим число меченых животных, которые должны быть отловлены из популяции в *i*-м году: $m_i = M_i \cdot n_i / N$.

Итак, состояние популяции гадюки в каждый год выражается тремя формулами:

$$N_i = N_{i-1} - N_d + N_b;$$

$$M_i = M_{i-1} - (N_d / N) \cdot M_{i-1};$$

$$m_i = M_i \cdot n_i / N,$$

где $i = 1, 2, 3, 4$ (судьба однажды помеченных особей отслеживалась в течение четырех лет).

В этой модели переменные m , M_0 , n известны, а параметры N_0 и N_d ($N_b=N_d$) предстоит оценить.

Табличное программирование

После того как построена математическая основа модели, можно приступать к программированию, переносу математических выражений в среду ЭВМ. В

рамках развиваемого подхода программирование означает создание имитационной системы на листе Excel. В отличие от программирования последовательного (Basic), структурного (Fortran, Pascal, C, TurboBasic, QuickBasic), блокового (СЛАМ, Sepath в Statistica, "I think") и объектного (Visual Basic, Delphi и др.) организация имитационной системы на листе Excel есть программирование *табличное*. Оно имеет целый ряд отличительных черт.

1. Программировать в электронной таблице *очень просто*, это доступно практически любому пользователю, в общих чертах знакомому с пакетом Excel. Ввод формул в ячейки листа Excel и есть программирование. Для этого достаточно только:

знать правила ввода формул (проще всего создавать формулы с помощью мыши и затем, при необходимости, редактировать),

знать правила формирования ссылок на ячейки (абсолютные ссылки, которые не изменяются при автозаполнениях, снабжены значками валюты \$),

уметь среди множества функций листа Excel выбрать нужную (они вызываются кнопкой f), обычно требуется не больше 10 функций (СУММ(), СТАНДОТКЛОН(), СЛЧИС(), СЧЁТ(), ...).

2. Программируя на электронном листе, не нужно заботиться об организации ввода-вывода и интерфейсе, поскольку Excel великолепно реализует эти функции самым "дружелюбным" образом. Не только данные, но и вся имитационная система с моделью и ее графическим отображением сохраняются в одном файле. Понятно, что в случае необходимости объем данных можно изменить, увеличив длину ряда. При этом важно следить за тем, чтобы диапазоны ссылок в формулах (суммирование и др.) включали бы и новые данные.

3. На листе Excel не нужно конструировать циклы, необходимые в других пакетах для организации хода времени или для перебора вариантов выборки. Табличная форма позволяет "развернуть" временную последовательность в пространстве электронного листа (например, сверху вниз). При этом ячейки одной строки содержат значения переменных, характеризующих состояние исследуемого объекта в данный определенный момент времени, а набор строк отражает последовательную смену состояния объекта во времени. Большое число строк листа Excel (16384 в Excel 7 и 65563 в Excel 97) позволяет отобразить сколь угодно продолжительную динамику (или сколь угодно большую выборку). Модель уже не есть единственная формула, встроенная в тело многократно повторяющегося цикла, как в алгоритмических языках программирования. Модель на листе Excel предстает как множество стереотипных формул, рассчитывающих значения всех переменных для каждого момента времени. Как уже было показано, ввод множества формул не представляет проблемы, благодаря предусмотренной в Excel процедуре "автозаполнение". При этом остается только следить за правильностью ссылок на параметры (они должны быть абсолютными). Некоторые *полезные приемы работы в Excel* даны в одноименном разделе четвертой главы.

4. Программа на листе Excel оказывается абсолютно прозрачной, поскольку все данные и результаты счета моментально визуализируются в численной форме и их можно охватить буквально одним взором. Сложные формулы можно разделять на ряд простых и на листе Excel отражать все промежуточные результаты счета (промежуточные переменные). Быстро ориентироваться в структуре перекрестных ссылок позволяют команды панели "Зависимости", связывающие стрелками ячейки вызываемые и вызывающие ячейки.

Компоненты имитационной модели

Численное моделирование имеет дело с тремя видами значений: исходными данными, рассчитанными значениями переменных и со значениями параметров. На листе Excel массивы с этими значениями занимают обособленные области.

Исходные реальные данные, выборки или ряды чисел, получают при непосредственном натурном наблюдении или в опытах. В рамках процедуры

моделирования они остаются неизменными (понятно, что при необходимости можно дополнить или уменьшить наборы значений) и играют двоякую роль. Часть из них (независимые переменные среды, X) служат основой для расчета модельных переменных; чаще всего это характеристики природных факторов (ход времени, фотопериод, температуры, обилие корма, доза токсиканта, объемы сброса поллютантов и др.). Другая часть данных (зависимые переменные объекта, Y) представляет собой количественную характеристику состояния, реакций или поведения объекта исследований, которая была получена в тех или иных условиях, при действии зарегистрированных факторов среды. В биологическом смысле первая группа значений не зависит от второй; напротив, переменные объектов зависят от переменных окружения. На лист Excel данные вводят с клавиатуры или из файла в обычном режиме работы с электронной таблицей.

Модельные расчетные данные воспроизводят теоретически мыслимое состояние объекта, которое определяется предыдущим состоянием, уровнем наблюдаемых факторов среды и характеризуется ключевыми параметрами изучаемого процесса. В ординарном случае при расчете модельных значений (Y_{Mi}) для каждого временного шага (i) используются параметры (A), характеристика предыдущего состояния (Y_{Mi-1}) и текущие уровни факторов среды (X_i):

$$Y_{Mi} = f(A, Y_{Mi-1}, X_i, i),$$

где $f()$ – принятая форма соотношения параметров и переменных среды, вид модели,

$$i = 1, 2, \dots T \text{ или } i = 1, 2, \dots n.$$

Расчеты характеристик системы по модельным формулам для каждого временного шага (для каждого состояния) позволяют сформировать массив модельных явных переменных (Y_M), который должен в точности повторять структуру массива реальных зависимых переменных (Y), что необходимо для последующей настройки модельных параметров. Формулы для расчета модельных переменных вводят в ячейки листа Excel вручную (см. раздел **Полезные приемы**).

Параметры модели (A) составляют третью группу значений. Все параметры можно представить как множество:

$$A = \{a_1, a_2, \dots, a_j, \dots, a_m\},$$

где j – номер параметра, m – общее число параметров,

и расположить в отдельном блоке. Понятно, что число параметров определяется структурой принятых модельных формул.

Занимая на листе Excel обособленное положение, они играют самую значительную роль в моделировании. Параметры призваны характеризовать самое существо, механизм осуществления наблюдаемых явлений. Параметры должны иметь биологический (физический) смысл. Для некоторых задач необходимо, чтобы параметры, рассчитанные для разных массивов данных, можно было сравнить. Значит, они иногда должны сопровождаться своими статистическими ошибками.

Первичные значения параметров вводят с клавиатуры. Часто достаточно задать единицы, $a = 1$. В других случаях следует вводить биологически правдоподобные величины. В дальнейшем точные значения параметров рассчитывает специальная процедура “Поиск решения” (см. раздел **Имитационная система**).

Пример с популяцией гадюки

В разделе **Приемы составления формул** были построены уравнения, описывающие снижение числа меченых животных в популяции и в повторных отловах. Цель настоящего этапа – преобразовать алгебраические уравнения в формат табличной модели, развернутой во времени.

Блок реальных данных занимает на листе Excel диапазон ячеек A1:C6 (табл. 2.8). Блок параметров в нашем примере должен содержать значение общей численности (N), число ежегодно гибнущих особей (N_d) и объем пополнения (N_b ,

$N_b=N_d$) популяции. Эти величины необходимо задать сразу же в первом приближении. На электронном листе Excel они помещены снизу: $F8 = 5000$, $F9 = 500$, $F10 = F9$, $D2 = F8$.

Таблица 2.8. Имитационная модель снижения числа меченых особей (M') в островной популяции гадюки в предположении постоянства численности (N), объемов отхода (N_d) и пополнения (N_b); приведены значения до настройки параметров

	A	B	C	D	E	F	G
1	Год	n	m	N'	d'	M'	m'
2	1994	158		5000		158	
3	1995	365	18	5000	0.1	142	10
4	1996	273	10	5000	0.1	128	7
5	1997	214	10	5000	0.1	115	5
6	1998	238	9	5000	0.1	104	5
7							
8					$N=$	5000	
9					$N_d=$	500	
10					$N_b=$	500	
11					d %=	10	

В соответствии с построенной ранее моделью численность популяции ежегодно поддерживается балансом процессов гибели и пополнения:

$$N_i = N_{i-1} - N_d + N_b.$$

Эта динамика в формате Excel примет вид:

$$D3 = D2 - \$F\$9 + \$F\$10,$$

$$D4 = D3 - \$F\$9 + \$F\$10,$$

$$D5 = D4 - \$F\$9 + \$F\$10,$$

$$D6 = D5 - \$F\$9 + \$F\$10 \quad (\text{табл. 2.8, столбец D, или графа } N').$$

Ежегодная смертность равна:

$$d_i = N_d/N, \text{ или в формате Excel:}$$

$$E3 = \$F\$9/D3, \dots (\text{графа } d').$$

Число выживших меченых особей будет равно:

$$M'_i = M'_{i-1} - d_{i-1} \cdot M'_{i-1}, \text{ или}$$

$$F3 = F2 - F2 \cdot E2, \dots (\text{графа } M').$$

В соответствии с моделью Петерсена число меченых особей в пробе составит:

$$m'_i = M'_i \cdot n_i / N, \text{ или}$$

$$G3 = F3 \cdot B3/D3, \dots, G6 = F6 \cdot B6/D6 (\text{графа } m').$$

После ввода всех формул в таблице Excel отображаются результаты расчетов; в данном случае численность сохраняется неизменной $N_i = 5000$ экз. (графа N').

Как видно из таблицы, модельное число повторно отловленных гадюк уменьшается (m'), но не совсем так, как реальные значения (m). Это говорит о том, что произвольно взятые величины N_0 и N_d требуют уточнения; модель ждет настройки. Она проведена в следующем разделе.

Имитационная система

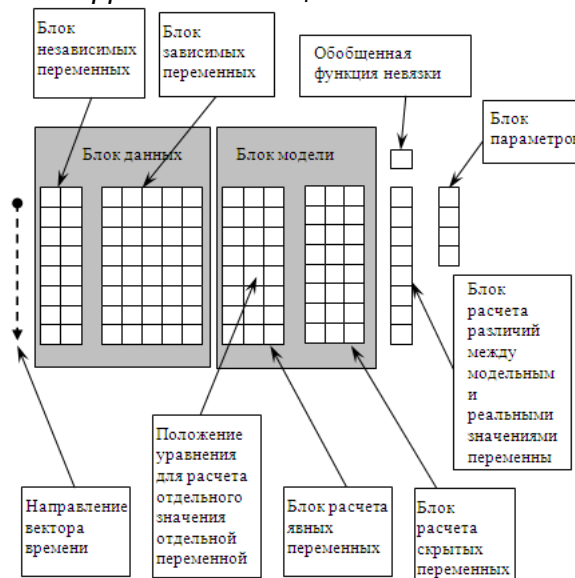
Имитационной системой называют информационную среду, в арсенал которой должны входить средства ввода и сохранения данных, программирования, отладки, настройки модели и презентации результатов (Моисеев, 1981, с. 213).

Расположившиеся на листе Excel три блока имитационной модели (реальные данные, расчетные данные, значения параметров) еще не позволяют выполнить основную задачу имитационного моделирования – определить оптимально точные значения параметров, характеризующих исследуемый процесс.

В идеале величина параметров должны быть такой, чтобы значения модельных переменных (Y') в точности совпали со значениями, наблюдавшимися в природе (опыте) (Y). Если значения реальных и модельных переменных различаются (как в нашем случае, табл. 2.8), это значит, что либо текущие параметры модели не вполне точно характеризуют механизм наблюдаемого процесса, либо принятый вид модели, модельные формулы, не соответствуют специфике явления.

В первом случае нужно приложить все усилия, чтобы уменьшить наблюдаемое отличие путем изменения значений параметров. Следуя этой логике, к трем блокам имитационной модели следует добавить, как минимум, еще три: 4) блок расчета отличий реальных и модельных переменных (расчет функции отличия или функции невязки), 5) блок процедуры последовательного изменения параметров с целью снижения “функции отличий” до нуля, 6) блок оценки статистической значимости уравнения. Если к этим блокам добавить еще и 7) диаграмму, изображающую реальность и модель, то совокупность из семи этих частей сформирует полноценную “имитационную систему” – среду для построения, настройки и исследования имитационной модели реального процесса (рис. 2.5).

Рис. 2.5. Базовый фрейм имитационной системы в среде Excel



Блок функции отличий (невязки) образован множеством формул, рассчитывающих отличие каждой пары комплиментарных реальных и модельных значений вида (Пэнгл, 1979):

$$\phi = B \cdot (Y' - Y)^2,$$

где B – “вес” данной переменной в общем массиве переменных.

Если переменных больше, чем одна, формула принимает вид (для одного временного шага):

$$\phi = \sum B_j \cdot (Y'_j - Y_j)^2,$$

где j – индекс переменной, всего m переменных ($j = 1, 2, \dots, m$),

Σ – суммирование по всем m переменным.

На листе Excel массив из этих формул образует блок различий между всеми парами значений модельных и реальных переменных. Сумма отличий по всем шагам образует искомую функцию отличий:

$$\Phi = \Sigma \Sigma B_j \cdot (Y'_{ji} - Y_{ji})^2, ,$$

где i – индекс шага модели, всего n шагов ($i = 1, 2, \dots, n$),

Σ – суммирование по всем n шагам и по всем m переменным.

Эта сумма занимает отдельную ячейку и используется для настройки модели. В простейшем случае, когда “вес” единственной переменной принимают за единицу, имеем:

$$\Phi = \Sigma (Y'_i - Y_i)^2.$$

Блок настройки модели служит для поиска параметров модели. Механизм настройки состоит в исследовании качества работы модели с разными наборами значений параметров и выборе лучшего из них. Исходно вычисляют функцию невязки (Φ_1) при некоторых заданных базовых значениях параметров. Затем изменяют их на некоторую небольшую величину и рассчитывают новое значение функции невязки (Φ_2). Если $\Phi_2 < \Phi_1$, значит, новые значения параметров лучше прежних, их принимают за очередные базовые и процедура повторяется. Существуют эффективные алгоритмы быстрого поиска лучших значений параметров. Процедура заканчивается, когда функция отличий снизится до минимально возможного уровня, в идеале – до нуля: при $A \rightarrow \text{оптимум}$, $\Phi \rightarrow 0$.

Можно полагать, что в этом случае полученные значения параметров будут оптимально характеризовать механизм наблюдаемого явления. Часто говорят не “настройка параметров модели”, а “настройка модели”, подразумевая, что модель приобретает некоторые свойства реального природного объекта.

Поставленная задача – уменьшение значений функции отличий до нуля при изменении неких параметров – есть математическая задача минимизации (частный случай задачи оптимизации), которая решается с помощью численных методов. Теоретическая подоплека процедуры оптимизации детально изложена в четвертой главе **Теория оптимизации**.

На первых порах экологу нет необходимости проникать во все нюансы метода, первая главная задача состоит в освоении процедуры построения моделей и их настройки в среде Excel. Однако ответственное отношение к результатам моделирования требует понимания существа процесса, происходящего во время настройки.

С практической точки зрения достаточно того, что пакет Excel располагает встроенной программой оптимизации – это “Решатель” (“Solver”), который вызывается командой “Поиск решения” (меню “Сервис”). Помимо действий по заполнению открывшегося окна этого макроса, описанных во **Введении**, может потребоваться перенастройка параметров его работы (см. раздел **Параметры макроса**) или задание условий выполнения настройки (см. раздел **Процедура моделирования, Прогноз как гипотеза**).

Блок статистической оценки служит для решения вопроса: соответствует ли модель реальным данным? Распространены четыре способа:

- 1) расчет коэффициента корреляции между модельными и опытными переменными (Акоф, Сасиени, 1971, с. 457),
- 2) проверка работы модели на независимо полученных данных (Розенберг, 1984),
- 3) оценка значимости модельных параметров с использованием статистической ошибки, полученной путем рандомизации (Безель, 1987),
- 4) оценка остаточной дисперсии и дисперсионный анализ (Ивантер, Коросов, 1992).

Рассмотрим последний достаточно простой способ. Как правило, многие значения изучаемой реальной переменной y не совпадают с модельными значениями той же переменной Y и на диаграмме располагаются по соседству с модельной линией графика. При этом отклонение отдельной варианты y_{ij} от общей средней Y будет состоять из двух слагаемых: отклонения варианты от

своего модельного прогноза и отклонения прогноза от средней величины признака для всей выборки: $(y_{ij} - Y) = (y_{ij} - Y_i) + (Y_i - Y)$.

Обобщая эту модель для всех вариантов, приходят к схеме дисперсионного анализа (Браунли, 1977, с. 309-320): изменчивость признака y складывается из варьирования, учтенного моделью, и из варьирования по случайным причинам, т. е. остаточного (табл. 2.9).

На листе Excel некоторые из этих величин уже рассчитаны. Значение функции отклонения численно равно остаточной сумме квадратов: $\Phi = C_{\text{остат.}} = \sum (y_i - Y_i)^2$;

остаточная дисперсия определяется из отношения: $s^2_{\text{остат.}} = \Phi / (n-1)$.

Общую сумму квадратов ($C_{\text{общ.}} = \sum (y_i - Y)^2$) можно вычислить, используя встроенную в Excel функцию стандартного отклонения (=СТАНДОТКЛОН(блок ячеек)), возведя полученное значение в квадрат и умножив на число степеней свободы. Модельная сумма квадратов ($C_{\text{мод.}} = \sum (Y_i - Y)^2$) есть разность между общей и остаточной:

$$C_{\text{мод.}} = C_{\text{общ.}} - C_{\text{остат.}} ;$$

это же значение равно модельной дисперсии, поскольку ее число степеней свободы $df_{\text{мод.}} = 1$:

$$s^2_{\text{мод.}} = C_{\text{мод.}} / df_{\text{мод.}} = C_{\text{мод.}} / 1 = C_{\text{мод.}}$$

После предварительных расчетов с помощью критерия Фишера можно проверить нулевую гипотезу $H_0: a=0$, параметр модели значимо от нуля не отличается (Браунли, 1977, с. 314), или предсказания модели в целом неадекватно описывают исходные данные. Конструкция критерия исследует вопрос: превышает ли варьирование, учтенное моделью, случайное (остаточное) варьирование? Критерий Фишера вычисляется как отношение модельной и остаточной дисперсии:

$$F_{(0.05; 1, n-2)} = s^2_{\text{мод.}} / s^2_{\text{остат.}}$$

Если значение критерия окажется выше табличного, значит, дисперсия реального признака y приближается по величине к дисперсии модельного признака Y , т. е. существенно превышает (случайные) отличия между ними. Значение критерия ниже табличного свидетельствует о существенных отличиях между реальными и модельными данными, о плохом согласовании модели с реальностью, о неадекватности модели.

Таблица 2.9. Дисперсионный анализ для проверки адекватности модели исходным данным

Составляющие дисперсии	Суммы квадратов C	Формулы расчета суммы квадратов	Степени свободы df	Дисперсии
Наклон модельной линии	$C_{\text{мод.}} = \sum (Y_i - Y)^2$	C_{xy}^2 / C_x^2	1	$s^2_{\text{мод.}} = C_{\text{мод.}} / df_{\text{мод.}}$
Отклонения вариант от точек модели	$C_{\text{ост.}} = \sum (y_i - Y_i)^2$	$C_{\text{общ.}} - C_{\text{мод.}}$	$n-2$	$s^2_{\text{остат.}} = C_{\text{ост.}} / df_{\text{остат.}}$
Общая дисперсия (всего)	$C_{\text{общ.}} = \sum (y_i - Y)^2$	$\sum y_i^2 - (\sum y_i)^2 / n$	$n-1$	$s^2_{\text{общ.}} = C_{\text{общ.}} / df_{\text{общ.}}$

Имитационные модели позволяют рассчитать значения не одной переменной (Y), но нескольких ($Y_1', Y_2' \dots$), каждая из которых может иметь свою статистическую природу. Комплексным показателем может служить *показатель точности имитации* (Безель, 1987, с. 28):

$$S = \sqrt{100 \sum [(a^3 - a^M) / a^M]^2 / (m-1)},$$

где a^3, a^M – эмпирические и модельные средние арифметические для m переменных.

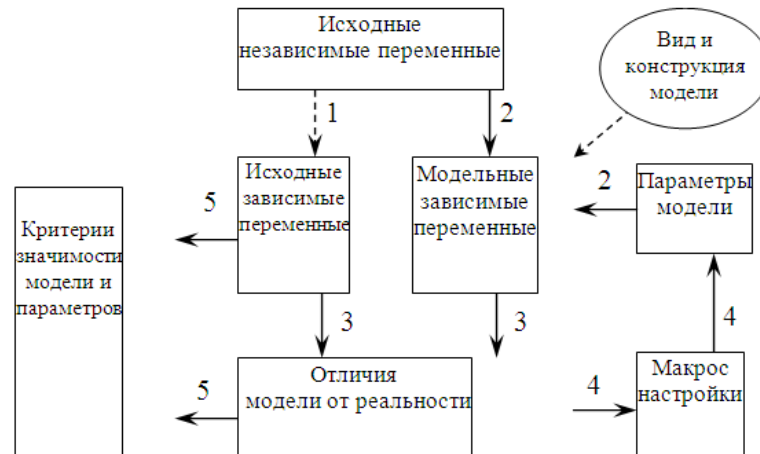
Для обобщенной оценки адекватности имитационной модели используют метрики сходства корреляционных (или ковариационных) матриц, рассчитанных

по исходным данным и модельным прогонам. Значение таких методов состоит в возможности статистического доказательства адекватности модели с помощью критерия хи-квадрат.

Блок графического представления результатов есть ни что иное, как диаграмма; возможности для визуализации данных в среде Excel необычайно широки. Отображение связи между зависимыми и независимыми переменными уже на самых первых шагах моделирования поможет определить вид модели, выбрать подходящие формулы для ее построения. Диаграмма помогает найти уклонения модельных приближений от действительности.

Отношения между компонентами имитационной системы формируют функциональное единство, ориентированное на достижение общей цели – оценку параметров модели (рис. 2.6, табл. 2.10).

Рис. 2.6. Блок-схема имитационной системы



В осуществлении отдельных функций, обозначенных стрелками, одновременно участвуют по несколько элементов. С тем чтобы не загромождать картину, на схеме не отражены блоки графического представления и рандомизации. Имитационная система призвана обслуживать любые изменения конструкций модели, которые в случае необходимости могут быть внесены исследователем. Базовые конструкции имитационных систем, а также возможные пути их декомпозиции и интеграции представлены в разделе **Фреймы имитационных систем**.

Таблица 2.10. Частные функции имитационной системы

№ стрелки	Выполняемая функция
1	Наблюдаемое природное явление: независимые реальные переменные влияют на зависимые
2	Расчет значений модельных переменных с использованием значений независимых реальных переменных и значений параметров
3	Расчет функции отклонения между реальными и модельными переменными
4	Настройка параметров модели с помощью макроса "Поиск решения" (оптимизация)
5	Оценка адекватности модели (дисперсионный анализ: расчет значений общей и остаточной дисперсии)

Блоки имитационной системы в среде Excel располагаются на одном листе. В дальнейшем мы в основном будем придерживаться указанного порядка размещения компонентов имитационной системы (рис. 2.5). Специальные задачи (рандомизация, графическое представление результатов и пр.) могут увеличивать число компонентов.

Пример с популяцией гадюки

Предварительно рассчитанная динамика встречи меченых гадюк в пробах плохо соответствует реальным значениям. Рассчитаем степень отличия модели от наблюдений, используя формулу:

$$\phi = (m - m')^2, \text{ или}$$

$$[I3] = (C3 - G3)^2, \dots (\text{табл. 2.11, графа } \phi),$$

и суммировав все найденные отличия:

$$[I8] = \text{СУММ}(I3:I6).$$

В таблице 2.11 величина обобщенного отличия (функция невязки) составила $\Phi = 109$. Понятно, что если бы модель абсолютно точно описывала реальность, то это значение было бы равно нулю.

Для настройки модели следует вызвать макрос “Поиск решения” и заполнить его окно (рис. 2.7). В качестве целевой выступает ячейка I8 со значением функции невязки, которое после настройки должно стать равным нулю; для этого нужно изменять значения в ячейках F8:F9. Результаты поиска решения представлены в таблице 2.12.

После настройки модельная динамика снижения числа меченых животных стала почти такой же, что наблюдали в поле, при численности островной популяции обыкновенной гадюки, равной $N = 3086$ экз., и при уровне ежегодной смертности $N_d = 228$ экз. ($d = 7.4\%$). Однако функция невязки так и не обнулилась и составила $\Phi = 6$.

Таблица 2.11. Имитационная система с моделью снижения числа меченых гадюк до настройки параметров

	A	B	C	D	E	F	G	H	I
1	Год	n	m	N'	d'	M'	m'		φ
2	1994	158		5000		158			
3	1995	365	18	5000	0.1	142	10		58
4	1996	273	10	5000	0.1	128	7		9.1
5	1997	214	10	5000	0.1	115	5		26
6	1998	238	9	5000	0.1	104	5		17
7									
8		S=	4.2		N=	5000		Φ=	109
9		df=	3		Nd=	500		D _{ост.} =	36
10		SQ=	53		Nb=	500		F=	-1.6
11		D _{мод.} =	-57		d	10			
					%=				

Для оценки адекватности модели проведем дисперсионный анализ. Остаточная дисперсия определяется из отношения: $D_{\text{ост.}} = \Phi / (n - 1)$, или $[I9] = I8 / C9 = 2$.

Общую сумму квадратов рассчитываем, используя квадрат функции стандартного отклонения и число степеней свободы: $[C8] = \text{СТАНДОТКЛОН}(C3:C6) = 4.2$

$$[C9] = \text{СЧЁТ}(C3:C6) - 1 = 3; [C10] = C9 * C8^2 = 53.$$

Рис. 2.7. Заполнение окна макроса “Поиск решения”

Поиск решения

Установить целевую ячейку:

Равной: ☐ максимальному значению ☒ значению: ☐ минимальному значению

Изменяя ячейки:

Ограничения:

Таблица 2.12. Имитационная система с моделью снижения числа меченых гадюк после настройки параметров

	A	B	C	D	E	F	G	H	I
1	Год	n	m	N'	d'	M'	m'		φ
2	1994	158		3086		158			
3	1995	365	18	3086	0.074	146	17		0
4	1996	273	10	3086	0.074	135	12		4
5	1997	214	10	3086	0.074	125	9		2
6	1998	238	9	3086	0.074	116	9		0
7									
8		S=	4.2		N=	3086		Φ=	6
9		df=	3		Nd=	228		D _{ост.} =	2
10		SQ=	53		Nb=	228		F=	23
11		D _{мод.} =	47		d%=	7.4			

Модельная сумма квадратов есть разность между общей и остаточной ($[C11]=C10-I8$) = 47; это же значение равно модельной дисперсии ($D_{\text{мод.}}$), поскольку число степеней свободы модели равно $df_{\text{мод.}}=1$. Величина критерия Фишера составит:

$$F = D_{\text{мод.}}/D_{\text{ост.}}, \text{ или } [H10] = C11/I9 = 23;$$

оно превышает табличное значение $F(0.05,1,3) = 6.6$. Модель в целом адекватна наблюдаемым данным. Этот вывод позволяет говорить о том, что численность взрослых особей островной популяции гадюки действительно приближается к 3000 экз. Для полученного значения численности можно рассчитать и статистическую ошибку, но только в результате дополнительных экспериментов с моделью.

Адекватность и значимость

Адекватность полученной модели – необходимое условие для осмысленной интерпретации ее параметров. В общем случае неадекватность модели связана с тремя причинами:

- состав модели существенно менее полон, чем реальная система, т. е. уравнения модели всегда включают в себя лишь небольшую часть из числа существенных переменных;

- связи между изучаемыми переменными описаны неточно, с использованием не той функции, или сделана попытка описания несуществующих связей;
- описание носит детерминистический характер, тогда как реальность всегда стохастична, т. е. любые данные содержат стохастический шум неизвестных, неучтенных и малозначительных факторов, но модель “старается” сгладить это варьирование.

Попытаться достичь большей адекватности (на одном и том же наборе данных) можно двумя различными способами: изменяя модельные уравнения и внедряя в них латентные (скрытые) переменные.

Внедрение латентных переменных – специфическая проблема имитационного моделирования, и поэтому ниже она будет обсуждаться подробнее (см. раздел **Скрытые переменные**). Здесь же, на первый случай, важно отметить главные функции скрытых переменных. Одна из них состоит в разбиении сложных зависимостей на несколько простых, что делает модель более прозрачной. Кроме того, скрытые переменные призваны играть роль неизвестных факторов (не доступных для наблюдения характеристик объекта исследования), забирать на себя “лишнюю” изменчивость явных переменных, контролируемых со стороны исходных данных. Это придает модели большую “гибкость”, она лучше согласуется с реальными данными. В то же время злоупотребление этим приемом, введение в модель избыточного числа скрытых переменных отрицательно сказывается на жесткости модельной конструкции и ведет к тому, что ее параметры утрачивают содержательную интерпретацию.

Изменение конструкции модели должно строиться на базе углубленного изучения явления. В частности, известно, что ростовые процессы в общем описывать лучше всего степенной функцией, фенологические – параболой, токсикологические – логистической кривой. Большая адекватность модели достигается также с увеличением порядка полинома.

При построении имитационных моделей интуитивно кажется, что чем лучше расчеты соответствуют эмпирическим данным, тем лучше модельные параметры характеризуют механизм исследуемых процессов. Однако при неограниченном вводе новых членов модели такое заключение оказывается неправомерным. Рассмотрим два конкурирующих описания зависимости размеров тела самцов гадюки (L_t) от числа прожитых лет (данные автора за 1998 г.) (табл. 2.13, рис. 2.8). Простая степенная функция (колонка E):

$$L_t = 25.6 \cdot x^{0.29}$$

обеспечивает адекватность модели на уровне $p < 0.001$ при $F = 73$. Полином 5-й степени (колонка C) дает линию, проходящую практически через все эмпирические точки:

$$L_t = 16.5 + 3.59 \cdot x + 6.54 \cdot x^2 - 2.6 \cdot x^3 + 0.35 \cdot x^4 - 0.0157 \cdot x^5,$$

а критерий Фишера достигает уровня $F = 864$, $p < 0.000001$.

Несмотря на рост адекватности, описание данных становится только хуже. Во-первых, интерпретация многочисленных коэф-фициентов полинома стала практически невозможна. Во-вторых, явно случайное рассеяние эмпирических значений вокруг среднего уровня новая модель восприняла как закономерные отличия, строго связанные с возрастом животного (x), подменяя тем самым знание о явлении в целом (его дает гладкая степенная кривая) детальным знанием особенностей конкретной выборки. Более адекватная модель оказывается менее интересной для биолога.

Таблица 2.13. Описание возрастного изменения длины тела гадюки (L_t) полиномиальной (L_{tm1}) и степенной (L_{tm2}) моделями

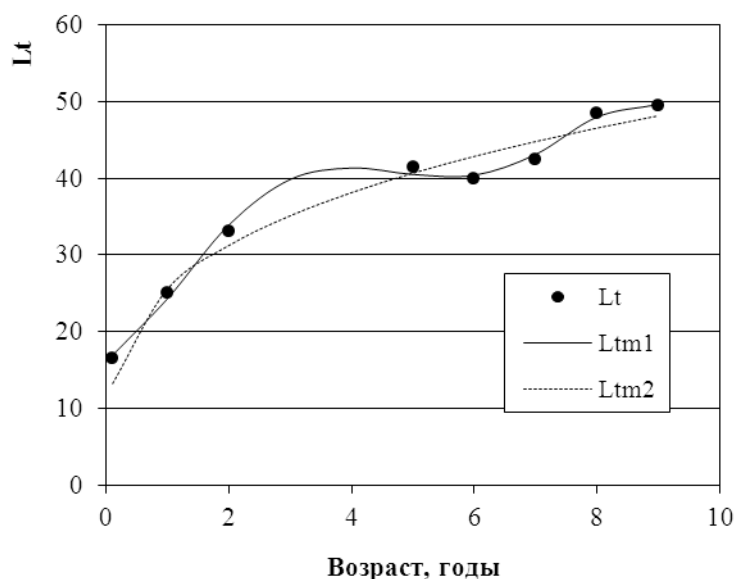
	A	B	C	D	E	F
1		$a_0 =$	16.5	$a_0 =$	25.6	
2		$a_1 =$	3.58	$a_1 =$	0.29	
3		$a_2 =$	6.54			

4		a3= -2.63				
5		a4= 0.35				
6		3				
6		a5= -0.02				
7	годы	Lt	Ltm1	φ	Ltm2	φ
8	0.1	16.5	16.9	0.2	13.2	10.7
			3			
9	1	25	24.3	0.4	25.6	0.4
			4			
10	2	33	33.9	0.9	31.3	3.0
			3			
11	3		39.8		35.1	
			4			
12	4		41.3		38.1	
			3			
13	5	41.4	40.4	0.8	40.7	0.5
			9			
14	6	40	40.3	0.1	42.8	8.1
			8			
15	7	42.5	43.1	0.4	44.8	5.2
			1			
16	8	48.5	47.9	0.3	46.5	3.8
			8			
17	9	49.5	49.6	0.0	48.1	1.9
			1			
18						
19			Φ= 3.1	Φ= 33.7		
20			F= 863	F= 73.0		

В отличие от регрессионного анализа в контексте имитационного моделирования терминам “адекватность” и “значимость” следовало бы придать разный смысл. Между уравнением регрессии и модельной динамикой есть глубокие различия. Расчеты регрессионных коэффициентов основаны на стройной и жесткой статистической теории, предполагающей, что все отклонения вариант от линии регрессии есть отклонения по случайным причинам, т. е. их варьирование соответствует нормальному закону (Браунли, 1977). Отсюда выводятся формулы расчета коэффициентов регрессии, их статистических ошибок и критерий значимости (оценка близости выборочных параметров к генеральным). В регрессионном анализе вывод о значимости параметров есть одновременно и вывод об адекватности модели.

Иное дело имитация, где ход модельной кривой целиком и полностью определяется замыслами автора модели, а законы распределения переменных не определены. В центр внимания становится, помимо выборки исходных данных, еще и подобная ей выборка модельных значений, полученная в результате расчета по формулам. Генеральная совокупность как-то уходит из поля зрения во время настройки параметров модели. По желанию исследователя модель может охватить все точки исходных значений или, напротив, пройти плавной линией между ними.

Рис. 2.8. Полиномиальная (L_{tm1}) и степенная (L_{tm2}) модели описания роста гадюки



В каждом случае доля “случайной” изменчивости вариант будет разной, а “значимость” параметров будет определяться *принятым* видом модели, т. е. будет субъективна.

Абсурдность ситуации можно снять, если отличать понятия “адекватность” (степень соответствия динамики модели – исходным данным) и “значимость” (степень соответствия структуры модели – реальности), и, соответственно, применять разные методы вычисления ошибок и критериев.

Для оценки адекватности модели исходным данным достаточно использовать коэффициент корреляции или дисперсионный анализ с критерием Фишера (см. предыдущий раздел). Для оценки же значимости необходима множественная настройка параметров модели на разных наборах исходных данных (см. следующий раздел). Только так можно определить, как конструкция модели воспринимает разнородные данные, т. е. в какой мере механизм динамики модели подобен механизму реального явления. В конце концов, оценка адекватности – не панацея от заблуждений, но индикатор степени незнания.

Статистические ошибки параметров

Оценка значимости параметров имитационных моделей, имеющих зачастую нелинейное поведение (с эффектами запаздывания, накопления, пороговых реакций, с изменением величины параметров в зависимости от состояния системы в целом), оказывается нетривиальной задачей. В литературе описаны три метода ее решения.

1. Оценка стандартных отклонений параметров методом наименьших квадратов по ковариационной матрице системы уравнений, полученной в процессе ее решения, например, методом последовательных приближений Гаусса-Ньютона (Попов, 1976, с. 50). Автору известна только одна программа, вычисляющая значимость параметров, – STATISTICA (Боровиков, Боровиков, 1997, с. 560).

2. Второй метод определения ошибок параметров состоит в многократном повторении процедуры настройки, каждый раз с новым массивом аналогичных *реальных* данных. Например, при определении параметров переноса тяжелых металлов между органами в организме животного в качестве отдельных переменных выступают уровни концентраций металла в различных органах, поразному накапливающих загрязнитель. В этом случае один массив эмпирической информации образует выборка разновозрастных особей. Исследование нескольких (n) групп подопытных животных дает необходимый набор

аналогичных массивов, по которым можно определить серию значений каждого модельного параметра ($a_1, a_2, \dots, a_n$) и непосредственно вычислить оценку их случайного варьирования (Безель, 1987, с. 22), т.е. статистическую ошибку параметров: $m_a = S_a / n^{0.5}$.

3. В тех случаях, когда многократное проведение экспериментов или наблюдений невозможно, серию из n аналогичных массивов исходных данных подготавливают искусственно, на базе одной исходной матрицы данных. Для этого, используя соображения о характере варьирования переменных, задают случайные отклонения от эмпирических значений и получают новые массивы (условно) исходных данных, поправленные на эти отклонения, т. е. искусственно организуют случайную изменчивость. Затем эти массивы используют, как описано выше, для расчета серии значений параметров, а затем и величины их ошибки (m_a) (Розенберг, 1984, с. 101). Последний метод наиболее доступен, хотя и достаточно сложен для применения, рассмотрим его подробнее на примере.

Рандомизация

Задача рандомизации состоит в том, чтобы изменить исходный ряд значений случайным образом. Это значит, что каждый новый столбец значений исходных данных должен быть похож на новую выборку из одной и той же генеральной совокупности. Поскольку в большинстве случаев признаки имеют нормальное распределение, рандомизация означает имитацию варьирования, многократное “искажение” каждого значения переменной с помощью случайной величины, распределенной нормально.

С технической точки зрения можно выделить три момента:

- организацию датчика случайной нормальной величины,
- определение параметров распределения (среднюю и дисперсию),
- расчет новых значений исходных данных.

Простейший датчик случайной величины, имеющей нормальное распределение, основан на датчике случайной величины, равномерно распределенной на интервале от 0 до 1. Проще всего распределение можно получить с помощью функции Excel =СЛЧИС(), которая реализует конгруэнтный метод (Прицкер, 1987, с. 47):

$$z_{i+1} = (a z_i + b) \pmod{c},$$

где z – случайное число в диапазоне 0–1,

a, b, c – любые константы,

$\pmod{c}$ – возвращение остатка от деления $(az + b) / c$.

При желании эту формулу, или ее упрощенный аналог,

$$z_{i+1} = az_i \pmod{c} \text{ (Акоф, Сасиени, 1971, с. 126),}$$

можно использовать для создания собственной случайной величины. Хорошие результаты дают значения $a=125, c=8192$ (Нивергельт и др., 1977, с. 210); остаток от деления возвращает функция Excel =ОСТАТ(делимое, делитель).

Датчик нормального распределения можно получить на этой основе с использованием центральной предельной теоремы. Величина t аппроксимирует нормальное распределение:

$$t = (1/n) (\sum z_i - 0.5) (12n)^{0.5},$$

где z_i – значения случайных величин, распределенных равномерно на интервале $[0...1]$, $i=1, 2, \dots, n$;

t – значение случайной величины, распределенной нормально со средней $M=0$ и стандартным отклонением $S=1$ (Нивергельт и др., 1977, с. 219). Для приблизительных расчетов те же авторы рекомендуют упрощенную формулу суммирования двенадцати значений z ($n=12$):

$$t = \sum z_i - 6, i=1, 2, 3, \dots, 12.$$

Рассчитать новые рандомизированные исходные значения x можно по формуле:

$$x = M \pm t \cdot S,$$

поскольку значения t представляют собой нормированные отклонения значений исходных переменных от своих средних, $t = (x - M) / S$.

Значения M и S назначаются по результатам предварительных исследований статистической природы изучаемых признаков. На первый случай в качестве M можно взять реальное исходное значение “варьируемой” переменной, а в качестве S – значение, равное 15% от величины M , поскольку основное большинство биологических признаков имеет коэффициент вариации ($CV = S/M$) не более 20 % (Яблоков, 1968).

Пример с популяцией гадюки

Рассмотренная выше модель динамики гибели меченых гадюк (раздел **Имитационная система**, табл. 2.12) позволила рассчитать некие средние уровни численности и смертности животных в популяции. Однако ясно, что реальный уровень численности отличается от расчетного и затушеван случайной изменчивостью,

$$N_{\text{мод.}} = N_{\text{реал.}} + N_{\text{сл.}}$$

В свою очередь, случайные флуктуации оценок численности вызваны случайной изменчивостью числа меченых животных в пробах. Среди причин, определяющих изменение вероятности попадания в очередную пробу ранее меченого животного, важно назвать две – относительную неравномерность исследования территории острова и изменение статуса особей (повзрослевшую особь встретить и отловить много легче, чем молодую).

На основании этой информации можно приблизительно рассчитать величину случайного уменьшения или увеличения вероятности отлова отдельной меченой змеи. Это отклонение служит основанием для превращения реального числа отловленных животных (m) в измененное, новое число повторно отловленных животных ($m_{\text{н.}}$).

По нашим данным, в разные годы число меченых особей в пробе колебалось от 18 до 9 экз. (табл. 2.12). Грубый расчет (для $CV=15\%$) позволяет получить возможные значения дисперсии:

$$S = 18 \cdot 0.15 = 2.7 \text{ и } S = 9 \cdot 0.15 = 1.35 \text{ особи.}$$

Специальный анализ (Коросов, 2000, с. 189) показал близкие значения:

$$S = 4 \text{ (от 0 – до 10) экз.}$$

Теперь можно вычислять новые, теоретически случайные, объемы повторно отловленных особей в каждой пробе по формуле:

$$m_{\text{н.}} = m \pm t \cdot S,$$

где m – реальное число меток в пробе,

S – стандартное отклонение, $S = 4$ экз.

$\pm t$ – значение случайной величины, имеющей нормальное распределение.

Сначала на листе Excel была подготовлена таблица для расчета флуктуирующих значений меченых особей (табл. 2.14). Значения равномерной случайной величины (z_i) рассчитывались с помощью функции =СЛЧИС(), значения нормальной случайной величины (t) вычислялись по формуле $t = \Sigma z_i - 6$. Новые (случайно измененные) значения числа меченых животных в повторных пробах определялись по формуле: $m_{\text{н.}} = m \pm t \cdot 4$, а эмпирические значения взяты из исходной табл. 2.8 (с. 84).

Таблица 2.14. Расчет одного набора (m_{10}) случайно измененных значений числа меченых животных в пробах

Год	m	m1	t	z1	z2	z3	z4	z5	z6	z7	z8	z9	z1	z1	z1
		0											0	1	2
1995	18	22	1	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.1	0.8	0
	8			9	8	8	7	2	9	5	4	8			
1996	10	5.3	-	0	0.	0.	0.	0.	0.	0	0.	0.	0.1	0.8	0
	0		1.2		2	8	8	3	5		8	6			
1997	10	14	0.9	0.	0.	0.	0.	0.	1	0.	0	0.	1	0.8	0
	0			8	5	4	4	2		9		9			
1998	9	8.5	-	0.	0.	0.	0.	0.	0.	0.	0	0.	0.9	0.5	0.2
			0.1	6	8	3	6	3	9	1		5			

Варьирование около фактического значения числа повторно отловленных животных предполагает, что эти значения достаточно близки к генеральному среднему уровню меченых особей в популяции; задача сводится только к созданию серии новых рандомизированных рядов повторных отловов. Так были получены 10 новых массивов данных ($m_1 - m_{10}$), которые затем были использованы для новой настройки параметров модели (N, N_d) (табл. 2.15).

Таблица 2.15. Изменчивость параметров модели, настроенных по варьирующим данным

Год	m_1	m_2	m_3	m_4	m_5	m_6	m_7	m_8	m_9	m_{10}	M	S	m
1995	20	19	17	15	15	22	9	18	15	22	17	4	
1996	12	10	4	13	11	11	7	12	10	5	9	2.9	
1997	7	16	23	13	14	13	16	10	10	14	14	4.3	
1998	9	10	8	2	10	12	9	11	7	8	9	2.9	
N	2448	3171	3420	2912	3575	2667	4555	3330	3643	2541	3226	632	200
N_d	386	0	0	430	0	123	0	52	173	345	151	174	55

Выборка из десяти значений параметров (N, N_d) позволяет рассчитать их статистическую ошибку (в качестве средних принимаем результаты расчета по реальным значениям меченых). По этим данным, численность гадюки на о. Кизи составляет $N \pm m_N = 3086 \pm 200$ экз., а смертность $N_d \pm m_{N_d} = 228 \pm 55$ экз.

Полученные оценки ошибок, конечно, не точны вследствие небольшого объема повторных расчетов ($n=10$). Это заметно по тому, что базовая и средняя оценки численности не совпали ($N=3086$ экз. против $N=3226$ экз.), то же и для числа меченых животных в пробах для отдельных лет (реальный ряд: 18, 10, 10, 9 экз. против расчетного: 17, 9, 14, 9 экз.). Возможно, увеличение выборки смогло бы нивелировать различия и уточнить оценки ошибок.

Недостатком рассмотренного метода следует назвать ничем не обоснованное предположение о том, что обнаруженные в повторных отловах объемы меченых животных (m_i) хорошо отражают концентрацию меченых животных в популяции и могут фигурировать в качестве генеральных средних при рандомизации. Снять это возражение можно, лишь предложив способ получения более репрезентативных характеристик повторных отловов. В нашем исследовании такие величины удалось получить, вычисляя средние многолетние доли меток в повторных пробах (Коросов, 2000, стр. 191, 207).

3. Приемы моделирования

Строение имитационной системы определяется, во-первых, ее предназначением (необходимо оценить параметры имитационной модели), во-вторых, структурой доступных исходных данных (независимых и зависимых внешних переменных) и, в-третьих, тем, какова должна быть функция модели (аппроксимация, интерполяция, декомпозиция, интеграция). Ниже различные фреймы (конструкции, каркасы, варианты) имитационных систем рассмотрены с позиций системного подхода. Это значит, что создание каждого нового фрейма, соответствующего новой задаче и содержащего новые модельные компоненты (дифференциация, декомпозиция), идет с параллельной разработкой способа объединения этих компонентов в целостную систему (интеграция, агрегация), способную выполнить свою главную функцию – корректно оценить параметры модели. Конкретно речь идет об оптимальной перестройке структуры

имитационной системы, пополнении ее различными вспомогательными блоками или операциями, среди которых наиболее значимую роль играют организация непрерывной динамики модели и весовые коэффициенты.

Фреймы имитационных систем

Основные варианты конструкций имитационной системы представлены на иллюстрациях, где прямоугольниками отображены блоки ячеек листа Excel, а стрелками – отношения между этими блоками (ссылки), которые означают использование значений из одних ячеек для расчета в формулах, находящихся в других ячейках. Хотя на иллюстрации блоки связаны одной стрелкой, фактически ссылки индивидуальны, все ячейки одного блока связаны с соответствующими ячейками (обычно той же строки) другого блока. В наиболее разветвленных схемах очевидные стрелки не изображены.

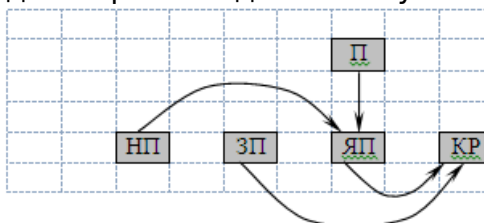
1. Базовый фрейм имитационной системы

Простейшая схема содержит 5 блоков объемом по одной ячейке. Значение внешней независимой переменной (НП) используется в формуле расчета значения модельной явной переменной (ЯП) с участием значения параметра (П): $ЯП = П * НП$.

Значение внешней зависимой переменной (ЗП) и расчетное значение явной (ЯП) переменной используются для оценки отличий модели от реальности в форме квадрата разности между ними (КР):

$КР = (ЗП - ЯП)^2$ (по существу, это функция невязки).

С помощью этой конструкции имитационной системы можно, например, оценить соотношение особей разного пола в популяции гадюк на изучаемом нами о. Кизи: из 2300 животных самок было 1250 экз. Иными словами, $НП = 2300$, $ЗП = 1250$. Приняв для начала $П = 1$, получим $ЯП = П * НП = 2300$, $КР = (ЗП - ЯП)^2 = 1102500$. Проведя настройку с помощью макроса "Поиск решения" (для $КР \rightarrow 0$), получим $П = 0.5345$, т.е. доля самок составляет около 53%. Конечно, эту величину можно было найти как частное от деления $ЗП/НП$, однако динамические модели простым делением уже не настроить.



2.1. Увеличение числа значений независимой переменной

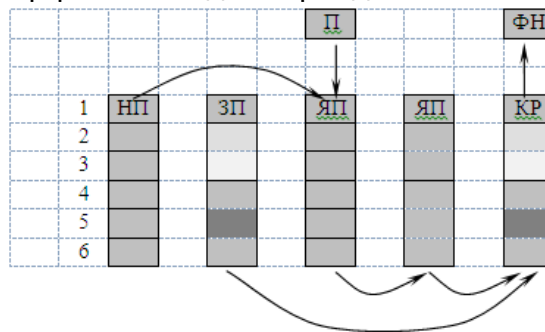
Ставя задачей поиск закономерностей, эмпирические данные группируют в массивы; блоки реальных независимых и зависимых переменных представлены уже не единичным значением, а сериями. Можно говорить о четырех основных способах тиражирования (декомпозиции) значений независимой переменной (нетрудно представить их комбинации):

- временные ряды (серия шагов по временной, возрастной шкале),
- градация фактора (серия доз, уровней факторов, расстояния и пр.),
- статус, габитус (множество объектов разного качества),
- выборка (наборы однотипных, но индивидуально различимых вариантов).

Чисто технически расширение имитационной системы состоит во вводе дополнительных строк, каждой из которых отвечает новый временной шаг (доза, варианта). Интеграция усложнившейся системы идет по двум направлениям. Это, во-первых, обобщение всех квадратов разности в одно значение функции невязки (ФН), во-вторых, организация блока критерия адекватности (КА) модели исходным данным. Вместе с ними имитационная система состоит уже из 7 блоков, что соответствует детальной схеме, рассмотренной выше в разделе **Имитационная система**, и примеру из **Введения**.

– модели, аккумулирующей результаты первого блока в соответствии со структурой исходных данных.

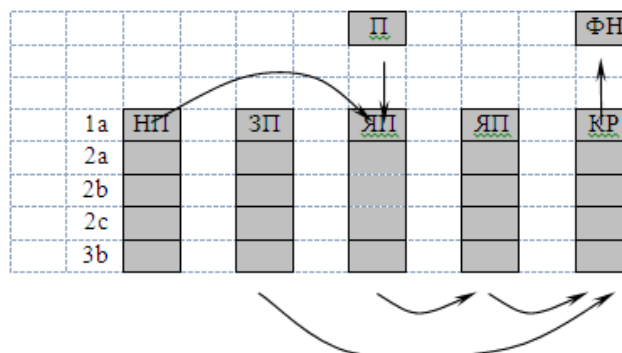
Такой случай характерен при изучении возрастного состава популяций, когда известна численность животных не каждого возраста отдельно, а лишь для разновозрастных групп, объединяющих особей разного возраста – молодых, взрослых, старых; пример расчетов дан в разделе **Сети связей**.



2.5 Дифференциация шагов

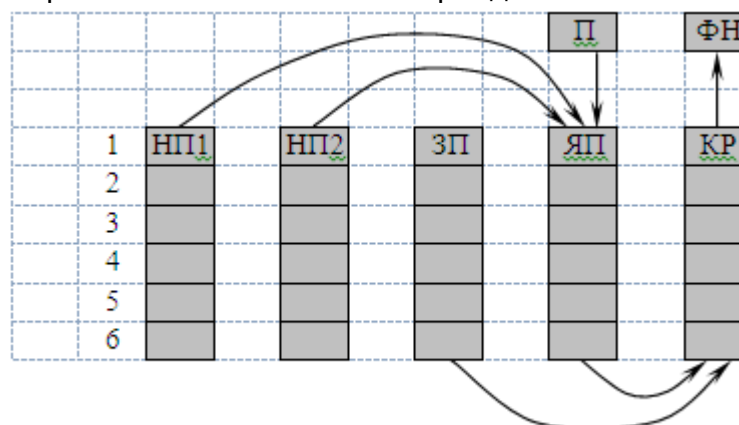
Часто встречается ситуация, когда отдельные временные шаги (градации фактора) представлены выборками вариант (причем разного объема). Здесь становится важным не только отразить общую динамику системы, но и не утратить информацию по индивидуальным отличиям (см. раздел **Индивидуализация модельных переменных**). Модельный блок вновь включает две части:

- общую модель, реализующую "плавный", последовательный ход времени (шаг 1, 2, 3...),
- модель, имитирующую особенности вариант в пределах каждого шага (1a, 1b... 2a, 2b, 2c...).



3. Увеличение числа независимых переменных

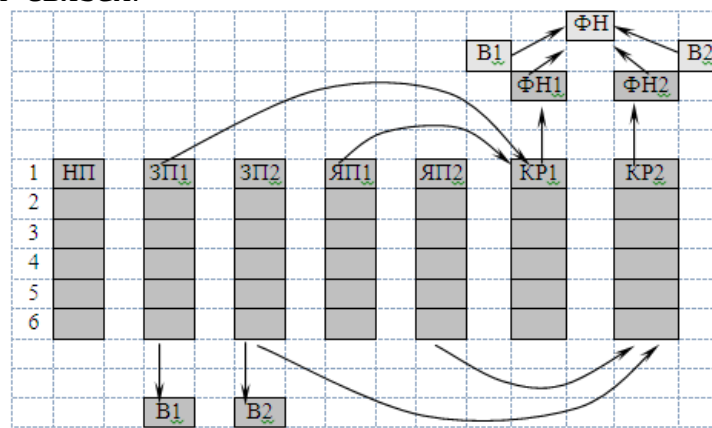
В природе, как правило, состояние объектов определяется несколькими факторами внешней среды. Модель, изучающая влияние нескольких независимых переменных на одну зависимую, должна содержать уже несколько наборов параметров (П1 и П2), на основе которых вычисляются модельные переменные; пример на этот счет показан в разделе **Сети связей**.



4. Увеличение числа зависимых переменных

Обычно у изучаемых объектов регистрируют несколько признаков. В этом случае блок внешних данных включает несколько зависимых переменных (ЗП1, ЗП2). Понятно, что и модельный блок должен содержать соответствующее число явных переменных (ЯП1, ЯП2) со своими наборами параметров (на рис. 3.1 они не указаны, также опущены связи с независимыми переменными).

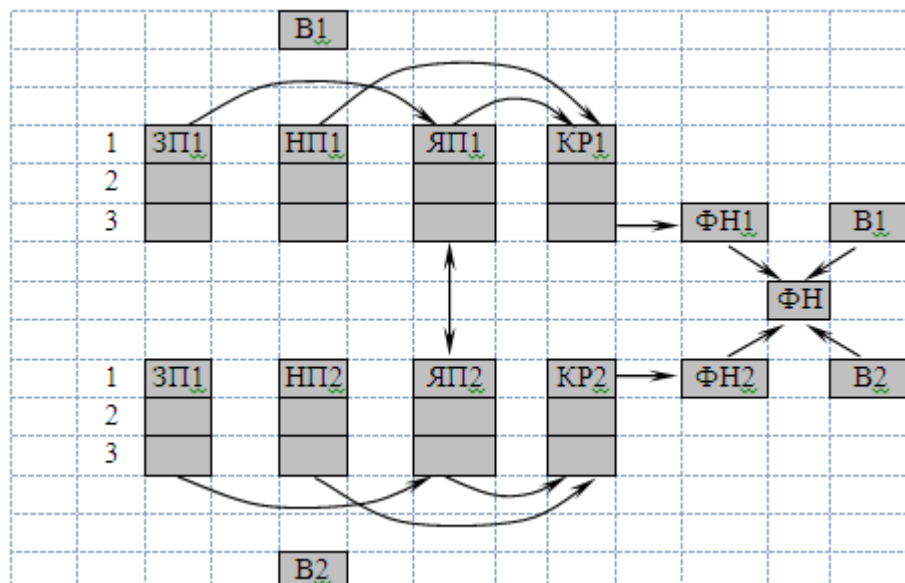
По существу, речь идет о создании двух соподчиненных моделей. Их интеграция в рамках имитационной системы должна идти путем формирования общей функции невязки. Каждая комплиментарная пара зависимых и явных переменных (ЗП1 и ЯП1, ЗП2 и ЯП2) служит для расчета своей, частной, целевой функции (ФН1 и ФН2). Когда единицы измерения или порядок величин разных переменных отличаются, тогда нельзя вычислять общую целевую функцию путем простого суммирования этих значений, необходимо вводить весовые коэффициенты для отдельных переменных (В1, В2). В простейшем случае в качестве весов можно принять обратную величину от средней арифметической данной зависимой переменной. Такой вариант рассмотрен в разделах **Усреднение..., Сети связей**.



5. Синтез нескольких моделей

Этот случай формально похож на предыдущий. Когда один и тот же объект исследования описан разными частными моделями, может появиться задача сведения их в единую общую модель (подробнее см. раздел **Сети связей**). С позиций перестройки имитационной системы это означает:

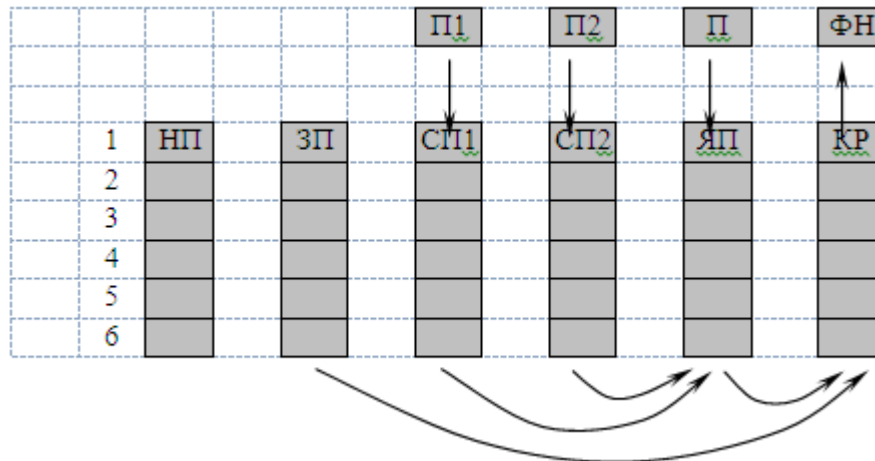
- установление взаимосвязи между моделями (создание перекрестных ссылок в модельных формулах),
- взвешенное обобщение частных модельных функций невязки в одно общее значение.



6.1. Увеличение числа модельных переменных: наведение

Самая интригующая сторона имитационного моделирования – это изучение *ненаблюдаемых* переменных. Модельный блок расширяется за счет организации скрытых переменных. Эти скрытые переменные (СП1, СП2) со своими параметрами (П1, П2) участвуют в формировании явной переменной (ЯП), которая вместе с независимой переменной используется для расчета функции невязки.

В простейшем случае скрытые переменные выступают в роли неизвестного фактора среды (имитация независимой переменной), влияние которого сказывается на динамике системы. Пример приведен в разделе **Скрытые переменные**.



6.2. Увеличение числа модельных переменных: декомпозиция

Второе применение латентных переменных состоит в декомпозиции изучаемого объекта на слагаемые. Скрытые переменные (СП1, СП2) в сумме формируют явную переменную, которая и сопоставляется с независимой переменной. Наиболее интересны случаи декомпозиции сложной функции (раздел **Декомпозиция кривой**), сложного распределения (**Анализ распределения**), динамики многокомпонентной модели (**Усреднение или параметризация**), а также в особых случаях создания переменного параметра (**Пропуски в данных**).

Все рассмотренные фреймы имитационных систем применялись автором для решения разнообразных проблем, возникающих в практике биолого-экологического исследования.

Аппроксимация кривой

В среде пакета Excel есть возможность строить уравнения простых зависимостей (линейной, степенной, экспоненциальной) с помощью встроенной функции “Добавить линию тренда”. Доступ к этой функции открывается, если выбрать стрелкой предварительно построенную диаграмму типа “Линия” или “Зависимость x-y” и вызвать контекстное меню (правой кнопкой мыши). В случаях, когда зависимость между изучаемыми признаками имеет более сложную форму, приходится пользоваться специальными статистическими пакетами типа StatGraphics.

Предлагаемая нами система моделирования позволяет строить уравнения зависимости переменных самого разнообразного характера и снабжать их оценкой адекватности исходным данным.

В качестве примера рассмотрим описание характера реакции представителей пресноводного зоопланктона Карелии *Simoccephalus serrulatus* на калий – основной компонент сточных вод Костомукшского горно-обогатительного комбината (данные любезно предоставила Н. М. Калинкина). Изучали две группы рачков – взятых непосредственно из лабораторной культуры и выросших в среде с небольшой добавкой токсиканта (20 мг/л). Учитывали долю (D_{реал.}, %) погибших животных в средах с разной концентрацией ионов калия [K]. Результаты представлены на рис. 3.2. Они хорошо могут быть описаны

S-образной кривой, характеризующей реакции типа “доза-эффект” большого класса биосистем (Прозоровский, 1962; Безель и др., 1994).

Для количественной характеристики этой зависимости используется логистическое уравнение (Лакин, 1973) вида:

$$Y = C + A / (1 + e^{ax+b}),$$

где C – нижний предел изменения зависимой переменной Y ,
 A – верхний предел.

В нашем случае $C=0$ (нет гибели в низких концентрациях), $A=100$ (полная гибель в высоких концентрациях). В табл. 3.1 представлена имитационная система, использующая формулу логистической кривой для построения модели вымирания по данным опыта:

$$D_{\text{мод.}} = A / (1 + e^{a[K]+b}).$$

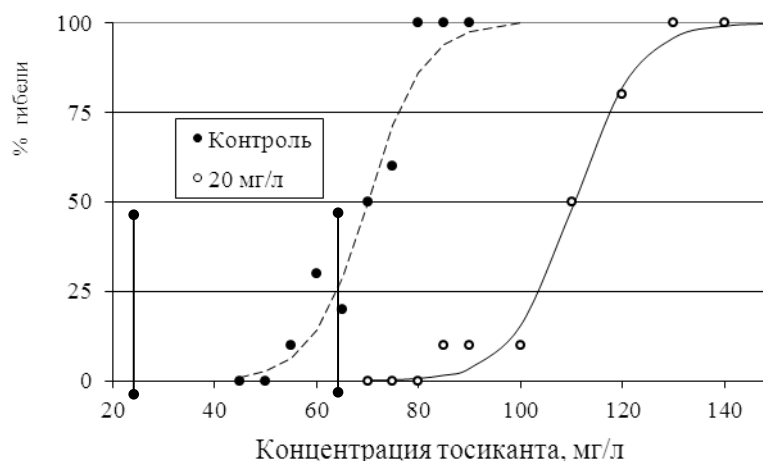
В формате Excel уравнение принимает следующий вид (для ячейки C6):

$$[C6] = C\$3 / (1 + \text{EXP}(C\$1 * A6 + C\$2)).$$

При настройке модели в качестве изменяемых параметров были взяты только a и b , третий, $A=100$, оставался неизменным. Модель значима ($p < 0.001$).

Расчеты позволили построить гладкие кривые логистической зависимости “доза-эффект”, по которым можно графически определить важный токсикометрический параметр – среднюю смертельную концентрацию, LC_{50} , опустив нормаль на ось абсцисс (C) из точки кривой $D=50\%$. Как видно по этим данным (рис. 3.2), акклиматизированные рачки переносят гораздо более высокие концентрации калия ($LC_{50}=112$), чем контрольные ($LC_{50}=68$ мг/л).

Рис. 3.2. Усиление резистентности группы рачков после акклиматизации в токсиканте (20 мг/л)



Пропуски в данных

Реальные ряды наблюдений зачастую содержат пропуски, пробелы и разрывы, когда в очередной момент времени определенное значение какой-либо переменной осталось неизмеренным, неизвестным.

При статистическом анализе многомерного массива данных исследователь всегда оказывается перед дилеммой: либо выбросить ряд (или столбец) с пробелом, либо заполнить его каким-либо значением. В практике многомерного анализа используется заполнение значением средней арифметической, рассчитанном прямо или по регрессии, а также с помощью метода главных компонент (Енюков, 1986, с. 56-61). После заполнения всех пропусков массив данных может быть обработан одним из методов. Когда результаты статистических расчетов начинают характеризовать специфику метода заполнения пробелов, их ликвидация становится неоправданной.

Таблица 3.1. Фрагмент листа Excel с результатами расчета параметров логистической зависимости доли погибших рачков от концентраций токсиканта.

	A	B	C	D	E
1		a =	-		
2		b =	0.181		
3		A =	12.684	Φ	722
4	Кон ц.	Dреал.	Dмод.	=	
5	45	0	1		1
6	50	0	3		7
7	55	10	6		14
8	60	30	14		253
9	65	20	29		79
10	70	50	50		0
11	75	60	71		128

1	80	100	86	195
2				
1	85	100	94	38
3				
1	90	100	97	7
4				
1				
5				
1	Варьирование		C	df S2
6				
1	Общее		1712	10
7			2	
1	Остаточное		722	8 90
8				
1	Модельное		1640	1 1640
9			0	0
2	F =		182	
0				

Корень вопроса состоит в том, что статистические процедуры требуют неразрывных массивов исходных данных. Статистические модели не могут быть рассчитаны при наличии разрывов.

Другое дело имитационная модель, которая самодостаточно существует как массив модельных значений, рассчитанный на каждый момент времени. Наблюдаемые данные служат лишь основой настройки модельных параметров. Для этой процедуры достаточно, чтобы наблюдения проводились не во все, а хотя бы в некоторые моменты времени, они-то и сыграют роль опорных точек. В принципе настроить модель можно и по единственной опорной точке, единственному реальному значению. В этом случае, правда, возникает проблема проверки статистической адекватности модели, поскольку по единственному значению нельзя провести дисперсионный анализ. С технической позиции неучет пробелов в исходных данных означает лишь, что в этой точке разность между модельными и реальными данными (ф) не рассчитывается. Роль функции невязки играет сумма таких разностей в остальных, "полноценных", точках.

В качестве примера рассмотрим модель роста кольчатой нерпы (данные любезно предоставлены О. В. Сергеевой). Одна особь нерпы выращивалась в бассейне института Северных морей (г. Мурманск). Условия освещенности, температуры и прочие соответствовали природным. Кроме этого, нерпу тренировали с возрастающей интенсивностью. Во время этого опыта регистрировали массу рыбы (пищи), скормленной за день (Мп), и один раз в неделю – массу самой нерпы, жившей в неволе (Мн). Цель моделирования состояла в том, чтобы описать динамику весового роста в связи с условиями кормления, а также дать рекомендации по выкармливанию нерп.

Общеизвестно, что из всего объема съеденной пищи лишь часть усваивается и идет на поддержание жизнедеятельности, активность и на весовой рост. Модель весового роста содержит коэффициент ассимилированной пищи, обеспечившей прирост массы тела (а):

$$dM_n = a \cdot M_{pi},$$

$$M_{ni}' = M_{ni-1}' + dM_n,$$

где i – день наблюдения,

M_{pi} – масса скормленной рыбы,

dM_n – ежедневный прирост массы особи, зависящий от объема съеденной пищи.

Предварительный анализ данных показал, что с течением времени доля пищи, идущей на рост (а), постепенно возрастает независимо от характера кормления. Это заставило усложнить модель и отразить в ней изменения

коэффициента ассимиляции. По сути, это новая скрытая переменная, лучше сказать, *переменный параметр*. Уравнение принимает вид:

$$a_i = a_{i-1} + b \cdot a_{i-1},$$

где b – коэффициент ежедневного увеличения коэффициента ассимиляции пищи. Полная модель теперь выглядит так:

$$a_i = a_{i-1} + b \cdot a_{i-1},$$

$$dM_n = a \cdot M_{pi},$$

$$M_{n_i}' = M_{n_{i-1}'} + dM_n,$$

$i=1, 2, \dots, 85$ (продолжительность опыта, дни).

Теперь модель имеет 4 переменные величины:

M_p – масса рыбы, кг, съеденной за день (независимая переменная),

M_n – еженедельные значения массы тела, кг, (зависимая переменная),

M_{n_i}' – расчетные значения массы тела (явная модельная переменная),

a_i – коэффициент использования пищи на рост массы тела (скрытая переменная),

а также два параметра:

a_1 – начальное значение коэффициента ассимиляции,

b – коэффициент (скорость) изменения коэффициента ассимиляции.

В соответствии с идеологией *неучета* пробелов разность между известной массой нерпы и массой, рассчитанной по модели, составит:

$$\phi = (M_n - M_{n_i}')^2,$$

вычисляем один раз в неделю 16.06.86, 23.06.86 ... 1.09.86, 8.09.86, а не для каждого дня жизни модели (взвешивание проводилось один раз в неделю); всего получаем не 85 значений (дней), а 13 значений (недель). Их сумма дает целевую функцию, которую минимизируем с помощью макроса "Поиск решения", изменяя два параметра: a_1 и b .

По результатам настройки можно видеть, что при исходном значении коэффициента усвоения $a_1 = 0.045$ и ежедневном увеличении в $b = 0.006$ раз (на 0.6 %) к осени коэффициент усвоения пищи увеличивается до $a_{85} = 0.075$, а в среднем за сезон составит $\bar{a} = 0.059$. Нерпа при этом увеличивает свою массу с 8.9 кг в начале лета до 17.4 кг в конце (табл. 3.3, блок "Весовой рост нерпы"). Модель показывает плавное увеличение массы тела нерпы; она хорошо согласуется с эмпирическими данными ($p < 0.001$, расчеты статистики F не показаны).

Таблица 3.3. Модель роста нерпы при разных режимах кормления

Дата	День опыта	Весовой рост					Масса пищи		
		M_p	M_n	a_i	M_{n_i}'	ϕ	M_{o_i}'	M_d	M_{n_i}''
16.06.86	1	2	8.9	0.045	8.9	0	0.6	7.9	7.9
17.06.86	2	1.4		0.045	9		0.4		7.9
18.06.86	3	1.7		0.046	9		0.5		7.9
19.06.86	4	1.5		0.046	9.1		0.4		7.9
20.06.86	5	2		0.046	9.2		0.6		8.0
21.06.86	6	1.5		0.046	9.3		0.4		8.0
22.06.86	7	0.4		0.047	9.3		0.1		8.0
23.06.86	8	1.1	9.6	0.047	9.3	0.07	0.3		8.0

24.06.86	9	1.6	0.047	9.4		0.5	8.1
...	...	...	...	...	...	...	...
30.08.86	76	2.8	0.071	15.8		0.8	9.9
31.08.86	77	2.8	0.071	16		0.8	10
01.09.86	78	2.8	0.072	16.2	0	0.8	10
02.09.86	79	2.4	0.072	16.4		0.7	10.1
03.09.86	80	2.4	0.073	16.6		0.7	10.1
04.09.86	81	2.4	0.073	16.7		0.7	10.2
05.09.86	82	2.4	0.074	16.9		0.7	10.2
06.09.86	83	2.4	0.074	17.1		0.7	10.3
07.09.86	84	2.4	0.074	17.3		0.7	10.3
08.09.86	85	2.4	0.075	17.4	0.11	0.7	10.4
		$a_1 =$	0.045			$c = 0.29$	
		$b =$	0.006				
				$\Phi =$	1.36	$\Phi =$	0.0015
		$\bar{a} =$	0.059				

Как уже говорилось, настроить динамическую модель в принципе можно и по единственному наблюдаемому значению. Продолжим рассмотрение примера с выкармливанием нерпы. В неволе к концу сезона она достигла массы 17 кг, тогда как средний вес диких животных в это же время составляет 10.4 кг. Свидетели говорят о сильном ожирении животного из неволи, вызванного, во всей видимости, перекармливанием. Новый вопрос формулируется так: каков должен быть объем пищи, чтобы масса нерпы в неволе стала равна нормальной массе дикой нерпы?

Рассмотренная выше модель связывает массу нерпы с массой скормленной рыбы. Уменьшить привес животного можно уменьшая ежедневную порцию пищи, например, в c раз. Дополним модель новым уравнением:

$$a_i = a_{i-1} + b \cdot a_{i-1},$$

$$Mo_i' = c \cdot M_{Pi},$$

$$dM_n = a \cdot Mo_i',$$

$$M_{n_i}'' = M_{n_{i-1}}'' + dM_n,$$

где

Mo_i – оптимальная масса суточного потребления нерпы,

c – коэффициент приведения реального объема пищи к оптимальному.

В отличие от прежней модели здесь вычисляется, во-первых, новый, оптимальный, объем пищи (Mo_i') и, во-вторых, масса нерпы, которую выкармливают по новому рациону M_{n_i}'' (табл. 3.3, блок "Масса пищи"). Изменено и стартовое значение, 6.9 кг (для 16.06.86), наблюдавшееся в природе.

Самое главное состоит в том, что в качестве функции невязки выступает единственное отличие модельной массы нерпы (8 сентября, 85 день) от массы дикой нерпы (M_d), измеренной в это же время:

$$\Phi = \phi = (M_d - M_{N85})^2.$$

Настройка новой модели (минимизация Φ по параметру c) дает величину $c = 0.29$. Для достижения оптимальной ("дикий") массы нерпе надо скормить лишь 29 % от того объема пищи, что был принят, или в среднем по 0.5 кг рыбы в сутки вместо 1.7 кг, что давали на самом деле. В этом случае динамика массы тела животного в неволе должна быть, видимо, примерно такой же, что и в природе.

Усреднение и параметризация

Одна из форм интуитивной критики имитационного моделирования состоит в противопоставлении ему результатов общей статистической обработки тех же исходных материалов, на которой строится модель. Если обобщенные описания можно получить простыми общепринятыми методами статистики, зачем создавать более сложные построения? Например, многие параметры функционирования популяций можно получить усреднением многолетних наблюдений.

При такой постановке вопроса из внимания упускается тот факт, что модель рассматривает временной ряд в его преемственном изменении, а при усреднении те же данные выступают в роли выборки независимых вариантов; информация о причинно-следственных связях между значениями соседних дат и рядами параллельных процессов по большей части утрачивается. В отличие от этого модельные параметры характеризуют не видимую данность, но ее подоплеку, структуру взаимоотношений переменных. Положим целью дальнейшего изложения попытку построить модель, которая реализовала бы усредненное описание динамики экологической системы.

В качестве объекта моделирования возьмем популяцию темной полевки, детально изученную в Карелии (Ивантер, 1975; Ивантер Э. В., Ивантер Т. В., 1986). Моделирование проводилось для оценки числа полевок, изымаемых из конкретной островной популяции хищником, обыкновенной гадюкой (Коросов, 2000). Мы хотели, чтобы модель, оставаясь очень простой, все же давала достаточно точные оценки демографических параметров популяции.

Наша модель имитирует годовой цикл воспроизводства популяции. Темная полевка, как и многие другие мелкие млекопитающие, – вид-эфемер. Перезимовавшие особи весной созревают и приступают к размножению, новорожденные быстро растут и наряду с зимовавшими принимают участие в размножении. К осени зимовавшие успевают дать по 3–5 выводков, а сеголетки – по 1–3 выводка. Особи, рожденные в конце лета и осенью, не растут, не созревают и остаются зимовать до следующей весны (Ивантер, 1975, с. 172–189).

Шаг модели составляет 1 месяц, что соответствует представлению данных в литературе. Это, конечно, огрубление, но не очень сильное. Взрослые самки могут повторно рожать уже через 22 дня, а у прибылых с момента их зачатия до родов следующего поколения проходит минимум 40 дней (20 дней идет на развитие до полового созревания плюс 20 дней беременности); потом срок приближается к показателям зимовавших. Однако приведенные числа характеризуют лишь максимальную скорость плодоношения (один раз в 20–22 дня). В среднем репродуктивный цикл особи приближается как раз к одному месяцу. Это значит, что для расчета числа новорожденных, появившихся за один шаг модели (30 дней), можно использовать среднюю плодовитость самок и их количество.

Структурируя рассмотренное выше описание в единицах численности, для каждого шага модели получаем следующее (табл. 3.4, настройка модели уже проведена). Зимовавшие особи (численностью N_{ad}) входят в состав половозрелых (N_a) и постепенно вымирают (выживаемость S_a). Новорожденные (N_n) выживают лишь частично (S_n), оставшиеся (N_{ns}) пополняют группу

сеголетков (N_{ji}). Кроме них в эту группу входят все молодые особи, выжившие на предыдущем шаге (N_{ji-1}). Какая-то доля особей из этой смешанной ювенальной группы выживает (N_{js} , S_j). Среди прибылых есть половозрелые (N_{ja_i}), их число рассчитывается как доля (A_j) от выживших прибылых (N_{js}). Вместе с зимовавшими взрослые прибылые особи образуют группу взрослых (N_a). Все половозрелые животные размножаются (плодовитость E). Зимовавшие и прибылые в сумме и образуют популяцию (численностью N'). В модели рассчитаны и относительные объемы групп: доля в популяции зимовавших особей (Pa_d') и доля в популяции половозрелых особей (зимовавших и сеголетков совместно) (Pa').

Таким образом, в число модельных переменных входят оценки численности 8 групп (N_{ad} , N_n , N_{ns} , N_j , N_{js} , N_{ja} , N_a , N') и относительные объемы двух групп (Pa_d' , Pa'). Среди них только три явных переменных – N' , Pa_d' , Pa' , которые сравниваются со своими аналогами в блоке эмпирических данных. Все остальные переменные (N_{ad} , N_n , N_{ns} , N_j , N_{js} , N_{ja} , N_a) – скрытые. Такая дробность представления модели введена для большей прозрачности расчетов. Базовые уравнения просты и очевидны; для примера приведены формулы из блока F7:O7 (табл. 3.5). Общая архитектура модели отображена на рис. 3.2.

Таблица 3.4. Имитационная модель динамики численности и состава темной полевки

Таблица 3.4. Имитационная система динамической модели
воспроизводства популяции темной полевки

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1					E	Sa	Sn	Sj											
2					2.85	0.74	0.45	0.84							Модельные				
3		Усредненные данные													данные				
4		N	Pad	Pa	Aj	Nad	Nn	Nns	Nj	Njs	Nja	Na	N'	Pad'	Pa'	φN	φPad	φPa	φB
5	май	0.05	1	1	0	0.05	0	0	0	0	0	0.05	0.05	1	1	0	0	0	
6	июнь	0.09	0.33	0.66	0.46	0.04	0.14	0.07	0.07	0.06	0.03	0.06	0.10	0.36	0.61	0.00	0.00	0.00	
7	июль	0.25	0.18	0.73	0.68	0.03	0.18	0.08	0.14	0.12	0.08	0.11	0.16	0.17	0.64	0.01	0.00	0.01	
8	август	0.3	0.08	0.55	0.48	0.02	0.30	0.14	0.25	0.21	0.10	0.12	0.27	0.07	0.45	0.00	0.00	0.01	
9	сент.	0.35	0.01	0.12	0.35	0.02	0.35	0.16	0.37	0.31	0.11	0.13	0.39	0.04	0.33	0.00	0.00	0.04	
10	B=	4.81	3.13	1.64											Φ=	0.01	0.002	0.062	
11															Φ*B=	0.046	0.005	0.101	0.152

Таблица 3.5. Уравнения динамической модели для полевков

Описание	Алгебраическое уравнение	Формула Excel
Выживает часть зимовавших	$N_{ad_i} = N_{ad_{i-1}} \cdot Sa$	F7=F6*F\$2
Взрослые размножаются	$N_{n_i} = N_{a_{i-1}} \cdot E$	G7=L6*E\$2
Выживает часть новорожденных	$N_{ns_i} = N_{n_i} \cdot Sn$	H7=G7*G\$2
Новорожденные пополняют группу прибылых	$N_{j_i} = N_{ns_i} + N_{js_{i-1}}$	I7=H7+J6
Выживает часть прибылых	$N_{js_i} = N_{j_i} \cdot Sj$	J7=I7*H\$2
Среди прибылых есть взрослые	$N_{ja_i} = N_{js_i} \cdot Aj$	K7=J7*E7
Расчет числа всех взрослых	$N_{a_i} = N_{ad_i} + N_{ja_i}$	L7=F7+K7
Расчет общей численности	$N_i = N_{ad_i} + N_{j_i}$	M7=F7+I7
Расчет доли зимовавших	$Pa_{d_i} = N_{ad_i} / N_i$	N7 = F7/M7
Расчет доли взрослых	$Pa_i = N_{a_i} / N_i$	O7 = L7/M7

В число параметров модели входят следующие величины:

– выживаемость для 3 групп (Sa , Sn , Sj), их значения настраивались с помощью макроса "Поиск решения";

- доля созревающих особей для каждого месяца (A_j), значения получены пересчетом таблицы 5 первоисточника (Ивантер Э. В., Ивантер Т. В., 1986, стр. 72),

- плодовитость (E), которая в среднем по всем группам составляет 5.69 экз./самку; поскольку в модели расчеты проводились не "на самку", а "на особь", в расчетах использовали значение $E = 5.69/2 = 2.85$.

Для настройки модели в качестве исходных данных взяты материалы из очерка по биологии темной полевки; это – усредненная за много лет динамика трех популяционных характеристик: общей численности, N (наблюдения за 36 лет; Ивантер, 1975, с. 189, рис. 57), доли зимовавших зверьков, Pa_d , доли взрослых особей всех возрастов, Pa (наблюдения за 26 лет; Ивантер Э. В., Ивантер Т. В., 1986, с. 74, табл. 9,6). На листе Excel эти значения помещены в блок B4:E9 (табл. 3.4). Комплиментарные им значения явных модельных переменных вычисляются в блоке M4:O9.

В качестве функции невязки выступали суммы квадратов разности исходных и модельных переменных (расчеты в блоке P4:R10). Как и в случае с динамикой популяций лягушек (см. раздел *Сети связей*), настройка данной модели идет по нескольким переменным одновременно. Поскольку всего рассчитываются три отдельных значения функции невязки, необходимо задать степень участия каждого из них в общей целевой функции, т. е. следует рассчитать весовые коэффициенты частных целевых функций. Весами служат обратные значения средних арифметических для каждого из реперных параметров (расчет приведен в блоке B10:D10). Общая функция невязки есть сумма произведений частных функций невязки на свои весовые коэффициенты (ячейка S11).

Настройка модели дает однозначное решение при разных стартовых значениях параметров (от 0.01 до 1, хотя иногда требуется двукратный запуск макроса): коэффициенты выживаемости составляют для взрослых $Sa = 0.74$, для новорожденных $Sn = 0.46$, для сеголетков в целом $Sj = 0.84$.

При этих параметрах динамика модельных переменных в общих чертах соответствует эмпирическим усредненным данным (табл. 3.4, рис. 3.3, 3.4). Модель статистически адекватна. Тем не менее цена этой адекватности невелика, поскольку полного совпадения статистических и модельных описаний нет! С одной стороны, статистические демографические характеристики получены для длинных рядов данных, они в высшей степени устойчивы и не подвержены случайному варьированию, добавление данных по новому году учетов практически не смещает средних оценок. С другой стороны, модельные переменные связаны жесткими зависимостями и взаимно абсолютно детерминированы. Это значит, что любые отклонения модельных и усредненных рядов зависят не от статистических ошибок, а от метода описания популяционной динамики. Как видно на иллюстрациях, отличия модели от реальности касаются в основном уровня численности (N) и доли половозрелых особей (Pa); снижение доли зимовавших (Pa_d) оба способа описывают практически идентично.

При рассмотрении усредненной динамики численности полевок бросается в глаза "горб": в июле натурные оценки (N) существенно превышают модельные (N'). То же можно сказать и для доли взрослых особей: в июле и августе их в среднем больше (Pa), чем выходит по модели (Pa').

Рис. 3.3. Динамика численности темной полевки
(N , экз./ 100 ловушко- суток)

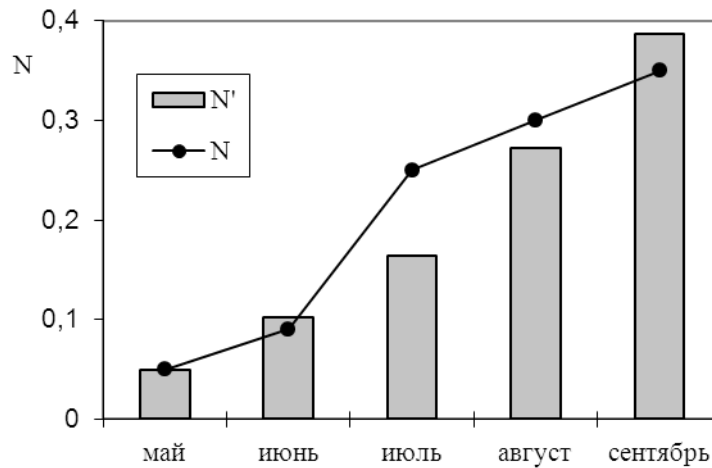
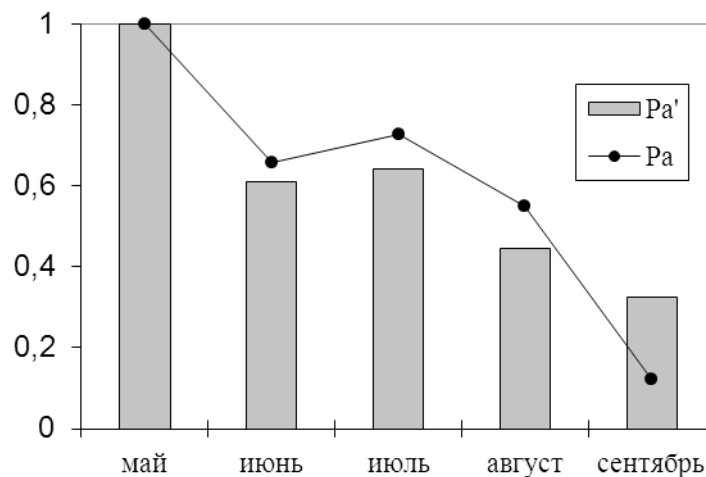


Рис. 3.4. Доля половозрелых прибылых



Причина этих различий кроется, на наш взгляд, в том, что для получения показателей N и P_a усреднялись разнородные данные, характеризующие существенно разный ход динамики параметров воспроизводства популяции, численности и половозрелой части популяции в разные годы ее жизни.

В годы пика численности (7 из 36 лет наблюдений) размножение начиналось раньше и максимум численности наблюдался в июле – августе. В годы с низким обилием зверьков (29 из 36) максимум достигался лишь к октябрю (рост численности был близок к экспоненте). При взвешенном усреднении столь разных динамических кривых должен образоваться "горб", но не так ярко выраженный, как при пике, а сглаженный (N). В то же время общий уровень численности оказывается много более низким, чем на пике, но все же много выше, чем во время депрессии.

Сходная картина складывается и при усреднении данных по динамике доли половозрелых особей (переменная P_a , рис. 3.3). В годы пика численности доля половозрелых особей велика весной и в начале лета, а во второй его половине и осенью резко снижается (за счет вымирания зимовавших и прохолодания сеголетков). В годы с низкой численностью, напротив, доля взрослых приближается к 100% даже осенью, когда размножение еще не закончилось. Усреднение сглаживает отличия между месяцами и к отчетливому "горбу" в середине лета добавляет лишь очень слабо выраженный "хвост" в сентябре.

Как было определено сначала, перед моделью стояла задача связать все усредненные описания популяции темной полевки в единую динамическую систему. Оказалось, что этого сделать нельзя. С одной стороны, модели навязан довольно большой объем размножающихся животных (через A_j и P_a), а с другой

– довольно низкие (и неравномерные!) темпы нарастания численности (через N). Модельная популяция со *средней многолетней* динамикой возрастного состава не может поддерживать *среднюю многолетнюю* динамику численности.

Не помогает и критика модели как достаточно грубого построения, мол, если бы она учитывала характеристики всех генераций прибылых, картина бы прояснилась. Это не так, поскольку все равно мы получаем одну и ту же ситуацию: какие бы из генераций ни доминировали, но при данном (среднем) уровне численности мы имеем данный (средний) объем половозрелых животных, которые не могут воспроизвести данный (средний) уровень численности, а могут только – *иной* (у нас – более низкий летом и более высокий осенью).

Моделирование выявляет внутреннее противоречие операции усреднения: наиболее репрезентативное описание динамики популяции оказывается не столь уж и адекватным, популяция не живет "в среднем".

Не менее важен и другой вывод. Как выяснилось, усредненные данные содержат много дополнительной информации, которую моделирование позволяет извлечь. Наиболее интересны оценки ежемесячной смертности животных разного возраста. Интуитивное понимание того, что выживаемость ниже всего у новорожденных, самая высокая у взрослых сеголетков, а зимовавшие в этом смысле занимают промежуточное положение, находит количественное подтверждение: $S_a = 45\%$, $S_j = 84\%$, $S_a = 74\%$ (табл. 3.4; блок F1:H2). Без использования трудоемкой методики мечения с повторным отловом удалось осветить самые интимные стороны демографической ситуации в популяции темной полевки. Развитие представленной модели видится в поиске аналогичных показателей для лет с разным уровнем численности, темпами воспроизводства и половозрастной структурой.

Индивидуализация переменных

Как было показано выше, для настройки своих параметров имитационная модель должна быть подобна реальности, точнее говоря, модель должна вычислять столько же модельных значений переменных, сколько их наблюдалось в природе, массив расчетных данных должен иметь идентичный состав и структуру с массивом исходных данных. Сделать это достаточно просто, если эмпирические материалы были получены для всего процесса, когда каждому его шагу соответствует одна единица информации.

Обычно же наблюдаемые данные отличаются не только неоднородностью и пробелами, но и неравномерностью описания разных стадий процесса. В частности, при изучении возрастных изменений какого-либо признака одни возраста будут представлены множеством индивидов, тогда как другие – единичными особями, либо они могут вообще отсутствовать в выборке.

В такой ситуации требуется проведение довольно трудоемкой процедуры подготовки модельного массива данных для сравнения с массивом исходной информации. Модель оказывается представленной несколькими блоками. Первый блок реализует модельную динамику всего процесса в целом, второй блок адаптирует плавную картину процесса – к индивидуальным особенностям реально наблюдаемых представителей этого процесса.

Рассмотрим пример "индивидуализации" общей модели по данным накопления одного из тяжелых металлов во внутренних органах лосей Карелии (табл. 3.6) (Медведев, 1995).

У добытых животных определяли возраст, весовые признаки, в число которых входят масса печени и масса почек. Пробы тканей этих органов анализировали на предмет концентрации кадмия. Значения концентраций были пересчитаны в единицы общего содержания металла (мг) во всем органе. Цель моделирования состояла в восстановлении картины возрастного накопления поллютанта в органах, а также в выяснении аккумуляционной способности каждого из них.

Поскольку возраст лосей оценивался с точностью 1 год, временной шаг модели также приняли за 1 год. В простейшем варианте считали, что текущее

содержание металла в органе определяется, в конце концов, его "аккумуляционной способностью" (балансом процесса поступления и экскреции): каждый год количество элемента увеличивается на некоторую величину dW , ту долю от общего поступления металла в организм с пищей и по другим каналам ($W_{п}$), которую орган в состоянии аккумулировать. Эту его способность характеризует "коэффициент аккумуляции" a :

$$dW_1 = a \cdot W_{п1},$$

$$W_1 = 0 + dW_1$$

...

$$dW_i = a \cdot W_{пi},$$

$$W_i = W_{i-1} + dW_i,$$

где W_i – содержание металла в органе,

$W_{пi}$ – объем годового потребления,

i – возраст особи.

В таблице 9.1 (настройка параметров уже проведена) этот процесс отражен в блоке "Модель накопления Cd по годам". В качестве объема "годового потребления" кадмия ($W_{п}$) взята гипотетическая величина 100 (для первого года жизни), которая в дальнейшем увеличивается пропорционально объему потребляемой пищи (до 211 на 10-м году жизни). Из этого количества некоторая часть аккумулируется в печени и почках, некоторая – в других органах. Коэффициенты аккумуляции $a_{геп}$ и $a_{рен}$ показывают относительную скорость накопления кадмия разными органами.

Представленная выше модель рассматривает процесс накопление кадмия с шагом один год и ничего не сообщает о конкретных особях, с которыми и надо сравнивать модельные расчеты. Для создания блока модельных данных (сопоставимого с блоком исходных данных) можно пойти двумя путями.

Простой путь состоит в том, чтобы для каждой особи известного возраста взять рассчитанное по модели содержание металла. Например, в соответствии с моделью в печени двухлетней особи должно содержаться 1.1 мг Cd. Можно, вообще говоря, приписать это значение всем двухлетним особям (№№ 1, 24, 61). Выполнив такую процедуру, мы можем получить блок модельных данных, по структуре идентичный блоку реальных данных. Однако в этом случае утратится информация об индивидуальных особенностях лосей одного возраста, в нашем случае – о разной массе органов и связанным с этим разным содержанием кадмия в них.

Правильнее все же не игнорировать индивидуальные особенности и учесть разный вес органов у разных особей. Так, вес печени для особи № 1 определен как $M_{геп} = 3.0$ кг, для № 24 – 4.2 кг, № 61 – 4.6 кг. Понятно, что и содержание загрязнителя в них будет различаться. Учесть индивидуальные массы органов, значит, что индивидуальное содержание кадмия в органе в той же мере отличается от среднего, что и индивидуальный вес органа – от среднего веса органа для текущего возраста. Такая постановка вопроса предполагает предварительный расчет средней массы органа M_o' (для этого строится обобщенная модель изменения массы органов с возрастом), затем поиск пропорции отличий индивидуальной массы от средней:

$$p = M_o / M_o',$$

где M_o' – расчетная средняя масса органа для данного возраста,

M_o – реальная масса органа данной особи,

и, наконец, расчет индивидуального содержания кадмия в органах данной особи:

$$W' = p \cdot W_i,$$

где i – возраст особи.

Таблица 3.6. Модель накопления Cd (мг) в печени (геп) и почках (рен) самцов лося в течение жизни (сокращенные данные Н. В. Медведева)

Особь		Содержание Cd (исходные данные, W)		Реальная масса органа (Mo)			Поправка для массы органа особи (p)		Расчет W' (Cd в особях)		Отличия $\phi =$ $= (W' - W)^2$	
№	лет	her	ren		her	ren	her	ren	her	ren	her	ren
1	2	2.4			3.0		0.8		0.9		2.3	
2	2	1.5	0.3		4.2	0.6	1.1	1.0	1.2	0.8	0.1	0.3
4												
6	2	2.4	0.1		4.6	0.7	1.2	1.2	1.4	0.9	1.1	0.6
1												
4	3	2.5			5.0		1.1		2.1		0.2	
5												
5	3	2.7	1.8		3.9	0.6	0.9	0.9	1.6	1.1	1.2	0.4
6												
5	3		2.8			0.8		1.2		1.6		1.6
8												
6	4	1.6	1.7		5.1	0.8	1.1	1.1	2.8	2.0	1.5	0.1
5												
1	5	2.7			5.5		1.1		3.8		1	
4												
6	6	2.4	2.0		5.1	0.8	1.0	1.0	4.2	3.1	3.2	1.2
6												
7	7	13.2			5.8		1.1		5.6		58	
2												
5	7	8.0			6.1		1.1		5.9		4.5	
1												
0	8	2.3	5.3		4.6	0.7	0.8	0.9	5.1	3.8	7.6	2.5
1												
3	8	7.5	7.8		5.5	0.8	1.0	1.0	6.0	4.4	2.3	11
5												
1	10	2.8	2.9		5.8	0.9	1.0	1.0	7.9	5.9	26	9.4
Результаты				Модель роста мас-сы органа (p, Mo')			Модель накопления Cd по годам (W _i ')				Потребление Cd в год (W _{pi})	
		her	ren	лет	her	ren		лет	her	ren	лет	у.е.
$\Phi = \sum \phi_i =$		102	35	1	2.9	0.5		1	0.5	0.3	1	100
				2	3.8	0.6		2	1.1	0.8	2	133
				3	4.4	0.7		3	1.8	1.3	3	153
	a =	0.0076	0.0056	4	4.8	0.7		4	2.6	1.8	4	167
				5	5.1	0.7		5	3.5	2.4	5	177
				6	5.3	0.8		6	4.4	3	6	186
				7	5.5	0.8		7	5.3	3.6	7	194
				8	5.7	0.8		8	6.2	4.3	8	200
				9	5.9	0.8		9	7.2	5	9	206
				10	6	0.8		10	8.2	5.7	10	211

Как показали расчеты, в двухлетнем возрасте масса печени в среднем составляет 3.8 кг, что отличается (p) от массы печени у особей № 1, 24, 61 соответственно в 0.8, 1.1, 1.2 раз. Модельное содержание кадмия в органах двухлеток составляет (с учетом средней $W_2 = 1.1$ мг) для № 1 $W' = p \cdot W_2 = 0.8 \cdot 1.1 = 0.9$, для № 24 – 1.2, для № 61 – 1.4 мг/ орган.

Индивидуализация модельных данных заставила дополнить имитационную систему тремя блоками – блоком выборки значений реальной массы органов (Mo), блоком расчета возрастной динамики массы органов (Mo'), блоком поправок для перехода от средних значений к индивидуальным (p). Сама же

модель предстает в двух ипостасях - как многолетняя динамика накопления кадмия в органах (W_i') и как адаптированная модельная выборка разновозрастных особей, аналогичная по структуре с выборкой реальных особей (W'). Понятно, что построение этих блоков предшествует настройке параметров модели; мы же для простоты рассматривали процесс индивидуализации переменных по значениям уже настроенной модели.

В результате настройки получено обобщенное описание процесса накопления кадмия в печени и почках, а также - их относительная аккумуляционная способность: $a_{\text{hep}} = 0.0076$ и $a_{\text{ren}} = 0.0056$. Судя по коэффициентам, почки накапливают кадмий менее интенсивно, чем печень, что характерно и для других млекопитающих (Безель, 1987). Существенное отличие приведенных расчетов от представленных в литературе состоит в том, что коэффициенты накопления были рассчитаны нами не по лабораторным наблюдениям, а по выборке разновозрастных животных, взятых из природы. Иными средствами, без применения моделирования, такие результаты не могут быть получены.

Декомпозиция кривой

Одна из функций моделирования - упрощение. Это не только инструмент преобразования нагромождения фактов в емкую модельную формулу или в структурированное количественное описание динамики экологических явлений. Пожалуй, более интересна возможность расчленять эту динамику на компоненты, представлять сложный процесс как объединение более простых взаимодействующих процессов. Любая сложная кривая может быть не только описана в виде уравнения, но и, что более важно, с легкостью разложена на составные более простые кривые, параметры уравнений которых приобретают биологический смысл.

Одной из таких часто встречающихся кривых является параболическая зависимость. Многие явления в жизни природы (вспышки численности, прирост размеров всего организма или его частей и т. п.) подчинены сезонной цикличности, когда постепенное увеличение уровня изучаемого признака сменяется плавным его падением. При этом за восходящую ветку кривой, как правило, ответственны одни процессы (например, размножение), а за нисходящую - другие (смертность). Результирующая параболическая кривая оказывается следствием синтеза двух противоположных тенденций.

Рассмотрим пример модельного анализа одного сложного процесса - изменения плодовитости рачка дафнии (*Daphnia magna* Straus) во все возрастающих концентрациях лигнина, основного компонента сточных вод целлюлозно-бумажных комбинатов (Калинкина, 1993) (рис. 3.5, табл. 3.7). Опыт состоял в следующем. Дафний одного возраста (7 дней) сажали в растворы с разными концентрациями лигнина. Ежедневно в течение месяца собирали живую молодежь. Число погибших молодых особей регистрировать было невозможно, но в высоких концентрациях на дне сосудов наблюдались скопления останков (карапаксов) погибшей молодежи. К концу опытов даже невооруженным глазом было видно, что дафнии в контроле существенно мельче дафний из опытных сред.

Эти простые наблюдения позволяют определить два основных фактора, ответственных за изменение плодовитости. Во-первых, лигнин явно стимулировал рост животных и крупные особи стали давать больше молодежи; для *Cladocera* это явление широко известно (Иванова, 1985). Во-вторых, в высоких концентрациях возросла смертность новорожденных, что и снизило общую плодовитость; этот эффект также обычен для интоксигированных дафний (Брагинский и др., 1987).

Параболическая кривая плодовитости, следовательно, образована наложением двух простых кривых - монотонно возрастающей (рост потенциальной плодовитости с ростом размеров тела) и монотонно убывающей

(снижение плодовитости с ростом концентрации лигнина). Определение параметров этих уравнений и составляет задачу моделирования.

Рис. 3.5. Плодовитость дафний (E, экз./самку/сутки) в растворе лигнина (Lig)

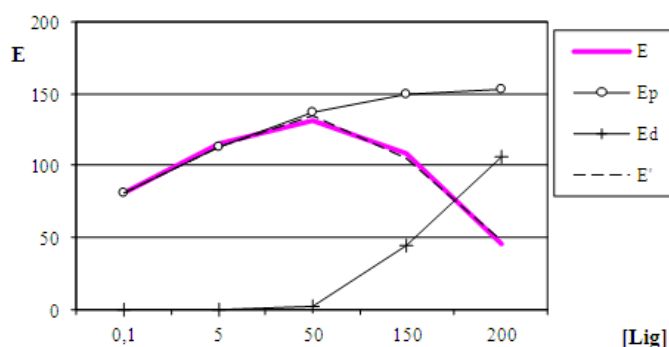


Таблица 3.7. Имитационная система модели изменения плодовитости дафний (E') в серии возрастающих концентраций лигнина [Lig] до настройки параметров

[Lig]	E	Ep	Ed	E'	φ		
0.1	81	0.1	0.1	0	6561	a	1
						=	
5	11	5	5	0	1322	b	1
	5				5	=	
50	13	50	50	0	1716	c	1
	1				1	=	
15	10	15	15	0	1188	d	1
0	9	0	0		1	=	
20	45	20	20	0	2025		
0		0	0				
					Φ	5085	F
					=	3	=

Решить эту задачу можно даже в отсутствие данных по размерам тела, напрямую связав рост плодовитости рачков с увеличением концентрации лигнина (мг/л). Тогда модель должна быть представлена тремя уравнениями (табл. 3.7):

– связь потенциальной плодовитости (экз./самку/30 дней) и размеров тела, зависящих от концентрации лигнина:

$$E_p = a \cdot [Lig]^b,$$

– зависимость смертности от концентрации лигнина:

$$E_d = c \cdot [Lig]^d$$

– результирующая, наблюдаемая плодовитость:

$$E' = E_p - E_d,$$

где a, b, c, d – модельные параметры.

Здесь две модельные переменные, потенциальная плодовитость (E_p) и смертность (E_d), оказываются скрытыми, они используются для расчета одной явной переменной – результирующей плодовитости (E'). При этом обе скрытые переменные определяются одной общей для них независимой переменной [Lig]. Для того чтобы модельные уравнения правильно отображали тенденции изменения соответствующих переменных, значения их параметров необходимо задать, хотя бы приблизительно, еще перед настройкой. Так, рост плодовитости может начаться только со значения 81 экз. (плодовитость в контроле); примем начальную величину первого параметра на уровне $a = 80$. Поскольку рост плодовитости идет не очень интенсивно, примем степенной параметр равным $b = 0.1$ (с этими параметрами первые значения потенциальной и реальной плодовитости почти совпадут). Смертность, в отличие от плодовитости, напротив, возрастает очень резко (примем $d = 2$), начавшись с очень низкого уровня (например, $c = 0.1$). Подставив примерные величины параметров в

имитационную систему, получим несколько абсурдные результаты (отрицательные значения смертности) и очень низкую оценку адекватности (табл. 3.8).

Тем не менее даже на этой базе удастся (с помощью макроса оптимизации) получить хорошее решение ($p < 0.00001$) (табл. 3.9, рис. 3.5). Дальнейшее варьирование стартовых условий позволяет найти наиболее устойчивые значения параметров (глобальный экстремум целевой функции) и отвергнуть менее эффективные решения (локальные экстремумы) (см. раздел **Экстремумы**).

Таблица 3.8. Имитационная система модели изменения плодовитости дафний (E') в серии возрастающих концентраций лигнина [Lig]. Параметры заданы в первом приближении

[Lig]	E	Ep	Eded	E'	φ		
0.1	81	63.5462	0.00	63.5452	304.667	a	80.00
		6	1	6		=	
5	11	93.9695	2.5	91.4695	553.683	b	0.10
	5	2		2		=	
50	13	118.300	250	-	69010.9	c	0.10
	1	6		131.699		=	
150	10	132.038	2250	-	4959359	d	2.00
	9	1		2117.96		=	
200	45	135.891	4000	-	1528112		
		7		3864.11	7		
				Φ =	5077589	F	0.6078
						=	4

Таблица 3.9. Имитационная система моделирования изменения плодовитости дафний (E') в серии возрастающих концентраций лигнина [Lig] после настройки параметров.

[Lig]	E	Ep	Eded	E'	φ		
0.1	81	81.5276	1.01E-08	81.5	0.27840	a	98.72343
		5			9	=	
5	11	112.853	0.00145	112.	4.61336	b	0.083115
	5	6	3	9	6	=	
50	13	136.656	1.57726	135.	16.6410	c	0.000009
	1	6	4	1	9	=	1
150	10	149.722	44.2833	105.	12.6807	d	3.035573
	9	4	9	4	5	=	
200	45	153.345	106.047	47.3	5.27963		
		5	8		2		
				Φ =	39.4932	F	113.4543
					5	=	

В заключение следует остановиться на двух моментах: обсудить понятие "жесткость модели" и дать биологическую интерпретацию модельных параметров.

Представленная модель имеет одну независимую переменную, это концентрация лигнина ([Lig]), и четыре зависимых:

- плодовитость (E), реальная зависимая переменная;
- результирующая плодовитость (E'), явная модельная переменная;
- потенциальная плодовитость (E_p), скрытая переменная;
- смертность молоди (E_d), скрытая переменная.

Наибольший интерес имеют, конечно, последние две компоненты, поскольку они демонстрируют явление, которое почти невозможно наблюдать в действительности. Здесь возникает принципиальный вопрос, насколько "правильно" модельные уравнения вычисляют эти переменные? Насколько можно доверять расчетам таких скрытых переменных, соответствуют ли они действительности?

На наш взгляд, верить таким расчетам можно, если:

- сохраняется уверенность в плавности наблюдаемых реальных зависимостей (без скачков и разрывов),
- выбранные функции соответствуют природе изучаемых явлений,
- принята достаточно "жесткая" конструкция модели.

В нашем случае плодовитость дафний в серии возрастающих концентраций токсиканта изменялась вполне плавно, аллометрический характер этих процессов показан в литературе (Иванова, 1985), а модель имеет абсолютно жесткий фрейм (каркас) (рис. 3.6.).

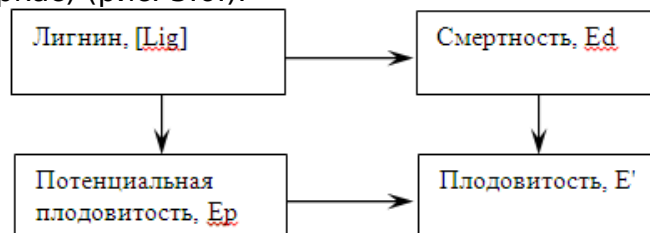


Рис. 3.6. Блок-схема модели плодовитости дафний

Каждая из латентных переменных (E_p , E_d) опирается на две заданные опытом переменные: одна из них служит аргументом в уравнениях ($[Lig]$), другая – основой для настройки параметров (E). В рассмотренном выше смысле модель не имеет ни одной степени свободы. "Упругость" аллометрической функции оставляет лишь по одной возможной траектории для каждой скрытой переменной. Поэтому есть все основания полагать, что найденные зависимости соответствуют реальности.

Проверить это предположение можно на модели с привлечением материалов по замерам дафний, выполненных в этом же опыте (табл. 3.10). Средняя длина тела (L_t , мм) вводится в качестве еще одной реальной зависимой переменной, она используется для расчета потенциальной плодовитости:

$$E_p = a \cdot L_t^b,$$

$$E_d = c \cdot [Lig]^d,$$

$$E' = E_p - E_d.$$

Параметры a , b теперь будут характеризовать реальную зависимость плодовитости от размеров тела.

Таблица 3.10. Имитационная система моделирования изменения плодовитости дафний (E) в серии возрастающих концентраций лигнина $[Lig]$ после настройки параметров

$[Lig]$	E	L_t	E_p	E_d	E'	ϕ		
0.1	81	3.7	83.6963	2.01E-08	83.6963	7.27008	a	0.33513
			1		1	5	$=$	4
5	11	4.5	116.297	0.00209	116.295	1.67704	b	4.21943
			1	3		4	$=$	
50	13	4.1	129.067	1.8818	127.185	14.5475	c	0.00001
			7		9	5	$=$	8
150	10	4.9	157.795	48.2890	109.506	0.25653	d	2.95370
			5	4	5	4	$=$	4
200	45	4.3	157.795	112.948	44.8469	0.02341		
			5	5	8	5		

		$\Phi =$	23.7746	F	45.5781
			3	=	5

В результате настройки новой модели были получены значения параметров, весьма близкие к старой модели ($c = 0.0000091$ и 0.0000180 , $d = 3.035573$ и 2.9537045), а прогнозы оказались почти идентичными: коэффициент корреляции между значениями потенциальной плодовитости в обоих вариантах модели составил $r = 0.982$. Иными словами, точное описание связи плодовитости с размерами тела в возрастающих концентрациях лигнина первая модель смогла дать, обойдясь без значений длины тела.

Достаточная жесткость модели сообщает латентным переменным статус реально наблюдаемых признаков!

Некоторые параметры моделей имеют ясный биологический смысл. Значение 99 практически равно плодовитости дафний в единичной концентрации лигнина $[Lig] = 1 \text{ мг/л}$ ($E_p = 99 \cdot [Lig]^{0.1}$) (табл. 3.9). Степень 4 близка к литературным оценкам и является отголоском кубической зависимости между объемом особи и ее линейными размерами ($E_p = 0.3 \cdot L_t^4$). Показатель 3 в уравнении зависимости плодовитости от лигнина ($E_d = 0.000018 \cdot [Lig]^3$), характеризует, по всей вероятности, видовую специфику реакции дафний на данное вещество (табл. 3.10).

Вычисление скрытых переменных оправдано и при менее жесткой структуре моделей, и в отсутствие проверочной информации. В этом случае они приобретают смысл количественных гипотез, выдвигаемых для первичного описания исследуемых явлений.

Анализ распределения

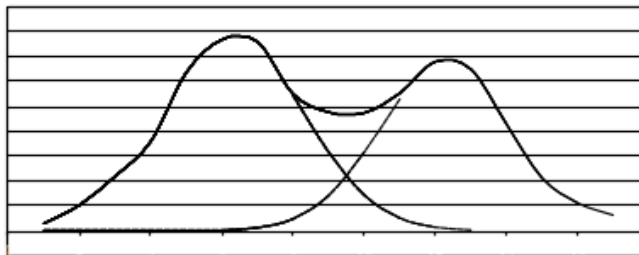
Метод имитационного моделирования позволяет декомпозицировать частотные распределения. В практике биоэкологических исследований достаточно часто встречается ситуация, когда частотное распределение какого-нибудь признака имеет не симметричную колоколообразную форму, соответствующую нормальному (или биномиальному) распределению, но может быть бимодальным (две вершины) или мультимодальным. Другая ситуация – отчетливая асимметрия и длинный “хвост” вариант в области больших значений (по типу логнормального распределения).

В таких случаях есть все основания считать, что данное распределение сформировано из двух (или нескольких) распределений, подчиненных нормальному закону. Первая посылка для этого умозаключения такова: нормальное распределение, несмотря на сложность внутренней структуры, отражает наиболее банальное, самое распространенное поведение непрерывных случайных величин (для дискретных признаков обычно биномиальное). Причина заключается в том, что каждое значение почти любой биологической случайной величины сформировано под действием большого множества неизвестных причин, иначе говоря, это значение представляет собой обобщение, сумму множества частных случайных величин. В силу математических законов оказывается, что такие суммативные, интегрированные случайные величины должны обладать нормальным распределением (см. центральную предельную теорему любого учебника по математической статистике). Когда же к действию случайных факторов добавляется действие систематических, то варианты образуют более или менее обособленные группы. Если влияние таких сильных воздействий не контролируется при наблюдении и разнокачественные варианты попадают в общую совокупность, их частотное распределение будет либо асимметричным (два фактора, один действует реже), либо двухвершинным (два фактора, действуют одинаково часто), либо мультимодальным (множество действенных неслучайных факторов).

Один из наглядных примеров – это когда исследуются объекты разного качества (например, особи разного возраста), но разграничение которых методически невозможно. В результате две или несколько разнокачественных

групп сливаются в одну выборку и образуют одно распределение с неразлично широкой зоной трансгрессии (рис. 3.7). Таким данным нельзя дать точные статистические характеристики (средние, дисперсии) ни для одной из групп.

Рис. 3.7. Бимодальное распределение



Здесь возникает важный исследовательский вопрос, нельзя ли для апостериорной дискриминации объектов использовать форму их совместного частотного распределения? Такая постановка проблемы означает, что модель должна реконструировать несколько ненаблюдаемых переменных (ряды частот), которые бы в сумме образовали эмпирический частотный ряд.

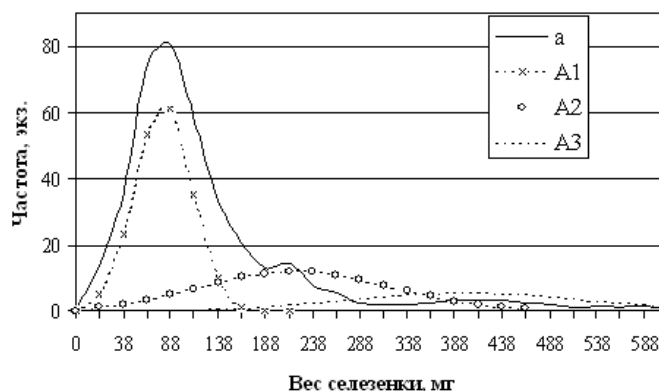
Рассмотрим пример изучения распределения веса селезенки у ювенальных прибылых красно-серых полевков в летний сезон (сборы автора в Южном Прибайкалье за 1984–1987 гг.) (табл. 3.11, рис. 3.8).

Таблица 3.11. Имитационная системы модели декомпозиции мультимодального распределения; до настройки параметров.

	A	B	C	D	E	F	G
1		M	80	22	40	$\Phi =$	176
2		S	30	10	13		1
3		n	190	12	70	38	384
4		C	37	45	33	0	
5		V					
6		C	158.	30	13		
7		a	A1	A2	A3	A	ϕ
8	0	0	0	0	0	0	0
9	13	1	5	1	0	6	50.7
10	38	3	23	2	0	26	108
11	63	7	53	3	0	57	310
12	88	8	61	5	0	66	205
13	11	5	35	7	0	42	285
14	13	3	10	9	1	19	206
15	16	2	1	10	1	13	67.6
16	18	1	0	11	1	13	0.05
17	21	1	0	12	2	14	0.19
18	23	8	0	12	2	14	41.3

1	26	5	0	11	3	14	76.2
2	3						
3	28	3	0	10	4	13	112
4	8						
5	31	2	0	8	4	12	98.1
6	3						
7	33	2	0	6	5	11	74
8	8						
9	36	2	0	4	5	9	49.5
10	3						
11	38	3	0	3	5	8	25.2
12	8						
13	41	3	0	2	5	7	16.3
14	3						
15	43	3	0	1	5	6	9.02
16	8						
17	46	2	0	1	5	5	9.32
18	3						
19	48	2	0	0	4	5	6.84
20	8						
21	51	1	0	0	4	4	6.23
22	3						
23	53	2	0	0	3	3	2.09
24	8						
25	56	1	0	0	2	2	1.96
26	3						
27	58	2	0	0	2	2	0.13
28	8						
29	61	1	0	0	1	1	0.48
30	3						
31	63	1	0	0	1	1	0.01
32	8						
33	66	2	0	0	1	1	0.71
34	3						

Рис. 3.8. Исходное распределение (а) полевок по весу селезенки и его три возможных слагаемых (A1-A3)



Селезенка как орган кроветворения и депо крови с возрастом изменяется вполне закономерно: у ювенильных животных она имеет малые размеры и небольшой вес, при созревании увеличивается и достигает максимальных размеров у беременных самок и взрослых самцов (до 1200 мг); это относится

как к зимовавшим, так и к прибылым особям (Ивантер и др., 1984). Вместе с тем, характерной особенностью селезенки является ее исключительная индивидуальная изменчивость (до $CV = 170\%$), которая связывается с действием разнообразных экологических факторов. В чем причина такой широкой изменчивости?

Ответ на этот вопрос дает анализ полигона частотного распределения веса селезенки ($n = 384$ экз.). Отчетливо различаются три пика: в области минимальных значений с модой $M_0 = 80$ мг, в области средних значений $M_0 = 220$ мг и у особенно крупных органов $M_0 = 400$ мг. Знание биологии этой группы позволяет утверждать, что первый пик соответствует неполовозрелым животным, второй – созревающим, но не имеющим пока диагностически значимого развития половых органов, третья группа – это стрессированные животные, скорее всего, в связи с какой-либо патологией. Логично предположить, что для каждой из групп распределение веса селезенки должно отвечать нормальному закону, поскольку большинство других, менее лабильных морфофизиологических признаков, имеет именно нормальное распределение.

На этой теоретической основе возникает задача реконструкции каждого из трех распределений как в форме полигона частот, так и в виде основных параметров – средней, стандартного отклонения, объема выборки (M, S, n).

Используем метод имитационного моделирования для расчета таких частных распределений (в терминах имитационной системы – трех скрытых переменных), чтобы их частоты в сумме давали исходное распределение.

Процедура расчета частот распределения, подчиненного нормальному закону (A_j), хорошо известна (Ивантер, Коросов, 1992):

$$A = C \cdot p_i,$$

$$p_i = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{(X_i - M)^2}{2S^2}},$$

$$C = dX \cdot n / S,$$

где

A – расчетная частота нормального распределения,

M – выборочная средняя арифметическая,

S – стандартное отклонение,

n – объем выборки,

dX – ширина интервала по оси OX ,

p_i – ординаты нормальной кривой,

X_i – значение на оси OX (центр i -го интервала),

i – номер интервала, $i=1, 2, \dots 34$.

Для реализации модели сначала в таблице Excel разместим исходные данные (использованы строки 8:34, табл. 3.11): в столбец A – значения веса селезенки, W (ширина интервала принята $dX=25$ мг, см. комментарий в конце раздела), в столбец B – частоты встречаемости вариант (особей) в этих интервалах (a_i). Сразу же рассчитаем общий объем выборки (N):

$$G3 = \text{СУММ}(B7:B34).$$

Далее введем примерные значения искомых параметров выборок (M, S, n). В качестве средних примем значения моды: для первой группы $M_1=80$, для второй $M_2=220$, для третьей $M_3=400$ (мг). Заносим на лист Excel:

$$C1=50; D1=220; E1=400.$$

Для примерной оценки стандартных отклонений нужно иметь в виду важное свойство нормального распределения: в границах $M \pm 1.96 \cdot S$ находятся 95% вариант совокупности (Колемаев и др., 1991). Иными словами, размах изменчивости вариант в выборке содержит примерно 4 стандартных отклонения (Ашмарин и др., 1975), минимальные значения выборки отстоят от средней примерно на $2S$, то же – и максимальные значения. Это позволяет для

предварительного расчета стандартного отклонения использовать следующие формулы:

$$S \approx \text{Lim}/4 \approx (\max - \min)/4 \approx (\max - M)/2 \approx (M - \min)/2.$$

Находим для нашего случая $S1 \approx (80 - 10)/2 \approx 30$, $S2 \approx 100$, $S3 \approx 130$ (мг). Заносим значения на лист Excel: C2=30; D2=100; E2=130.

На основе этих данных можно рассчитать иллюстративный коэффициент вариации: $CV = 100 \cdot S/M$. Его значения даны в блоке ячеек C4:E4. Объемы выборок можно предварительно назначить, ориентируясь на форму полигона частот (рис. 3.8). Соотношение объемов групп в общей выборке, видимо, недалеко от 50:30:20%; при общем объеме $N=384$, примерные объемы групп составят $n1=190$, $n2=120$, $n3=70$ (экз.). Заносим их на лист Excel: C3=190; D3=120; E3=70.

Кроме этого, для каждого ряда необходимо рассчитать константу C , адаптирующую объем выборки к стандартному отклонению и ширине интервала, выбранного для построения распределения. В нашем случае ширина составила $dX=25$ мг. С использованием уже внесенных значений формула расчета константы примет вид: $C5=25 \cdot C3/C2$. "Автозаполним" ею также ячейки D5, E5.

Теперь можно вводить рассмотренную выше модельную формулу для расчета частот (A) в каждой группе особей отдельно. В формате Excel она примет вид:

$$C8=(C\$5)*(1/\text{КОРЕНЬ}(2*\text{ПИ}()))*\text{EXP}(-0.5*($A8-C\$1)^2/C\$2^2)).$$

В этой записи заслуживает внимания форма ссылок на ячейки, где находятся значения параметров. Ссылка $C\$1$ указывает на величину средней арифметической для первой выборки, содержащейся в первой строке, которая и зафиксирована в ссылке, т. е. стала абсолютной ($\$1$). В то же время ссылка на индекс столбца осталась относительной (C), потому что впереди будет операция автозаполнения, позволяющая распространить формулу на колонки D и E, которые содержат параметры и частоты двух других частных выборок.

Рассмотренную формулу путем автозаполнения следует ввести в ячейки C8:C34, затем в C8:E34. Блоки ячеек каждого из столбцов C, D, E будут отображать распределения, которые строго соответствуют нормальному закону и базируются на заданных выше параметрах.

Объединение этих расчетных частот частных распределений (скрытых переменных) должно соответствовать исходному распределению эмпирических частот. Рассчитаем их суммы в каждом интервале. Для ячейки F8 сумма равна: $F8=\text{СУММ}(C8:E8)$. Выполним автозаполнение этой формулой ячеек F8:F34. В заключение найдем общий объем объединенного распределения $F3=\text{СУММ}(F8:F34)$.

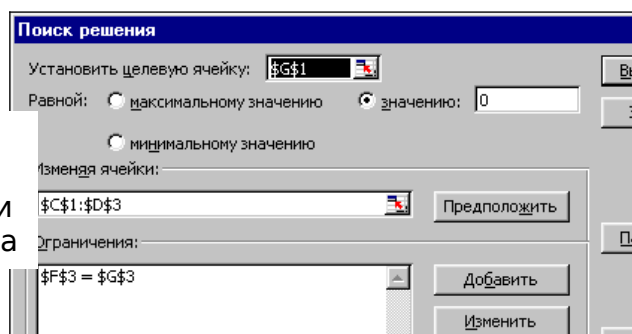
Достроим имитационную систему блоком оценки сходства реального и общего модельного рядов частот: $G8=(B8-F8)^2$, автозаполним блок G8:G34. Значение функции отклонения (невязки) рассчитаем в ячейке G1:

$$G1=\text{СУММ}(G8:G34).$$

Для построения диаграммы, отображающей результаты моделирования, зададим имена переменным: $B6=a$, $C6=A1$, $D6=A2$, $E6=A3$, $F6=A$, $G6=f$. Выделим с помощью мыши и клавиши **Ctrl** области A6:B34; <Ctrl>F6:F34 и построим линейную диаграмму (рис. 3.8). Как видно, расчетные частоты (A) плохо соответствуют исходному распределению (a).

Далее, в окне макроса "Поиск решения" (рис. 3.9) укажем, что целевую ячейку G1 следует обнулить $G1 \rightarrow 0$, изменяя значения в ячейках $\$C\$1:\$E\3 . (Для отработки разных вариантов настройки в качестве изменяемых ячеек, возможно, имеет смысл разделять объемы выборок, средние и стандартные отклонения.) Кроме этого необходимо задать условие $\$F\$3 = \$G\3 , поскольку объемы исходной и расчетной совокупности должны, естественно, совпадать. Результаты настройки параметров представлены в табл. 3.12 и на рис. 3.10.

Рис.
3.9.
Заполнени
я окна



Редкий случай сравнения частот позволяет воспользоваться критерием Пирсона хи-квадрат (Ивантер, Корсов, 1992):

$$\chi^2 = \sum ((F8 - B8)^2 / B8).$$

Критерий Фишера также вполне применим; оба показателя дают высокий уровень значимости полученного описания частот ($p < 0.0001$).

Перерасчет коснулся всех статистических параметров. Общая особенность состоит в том, что уровень изменчивости веса селезенки в каждой из групп существенно уменьшился и приблизился к значениям, характерным для большинства морфофизиологических показателей, $CV = 20-40\%$. Интересен и характер перераспределения частот между группами. Вторая группа (созревающие особи) локализовалась неожиданно широко, а центр оказался левее, чем интуитивно ожидалось. Она вобрала в себя не только правостороннюю асимметрию распределения первой группы, но и ее избыточный эксцесс. Третья группа, практически полностью изолированная от первых двух, вобрала в себя большинство весовых значений крупных селезенки, подчеркнув симметричность частоты встреч патологий.

Таблица 3.12. Имитационная система модели декомпозиции мультимодального распределения после настройки параметров

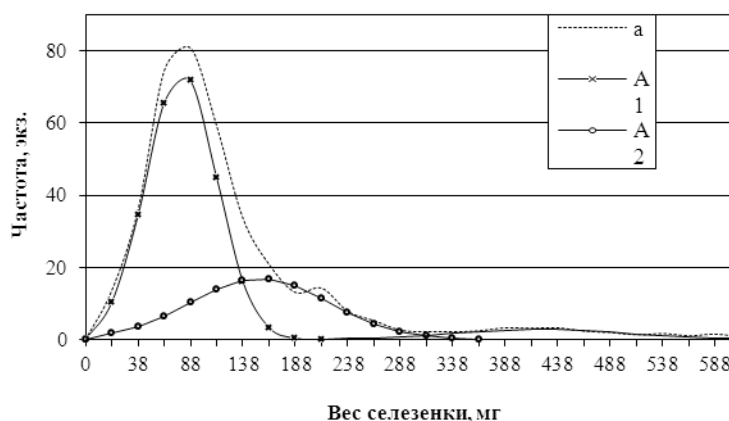
	A	B	C	D	E	F	G
1		M	79.0	154	42	Φ=	32
2		S	33.5	65.	84		
3		n	249	112	23	38	384
4		CV	42	43	20		
5		C	185.	42.	6.9		
6		a	A1	A2	A3	A	φ
7	0	0	0	0	0	0	0
8	13	14	10	2	0	12	2.6
9	38	36	34	4	0	38	3.8
10	63	75	66	6	0	72	6.3
11	88	81	72	10	0	82	1.1
12	11	59	45	14	0	59	0.1
13	13	34	16	16	0	33	1.2

14	16	21	3	17	0	20	0.4
15	3	13	0	15	0	15	3
16	8	14	0	11	0	12	5.2
17	21	3	0	8	0	8	4
18	23	8	0	8	0	8	7.0
19	8	5	0	4	0	5	2
20	26	3	0	2	1	3	0
21	28	3	0	2	1	3	0.1
22	8	3	0	2	1	3	0.1
23	31	2	0	1	1	2	0
24	33	2	0	0	2	2	0.0
25	8	2	0	0	2	2	0.0
26	36	3	0	0	3	3	0.5
27	38	3	0	0	3	3	2
28	8	3	0	0	3	3	0.2
29	41	3	0	0	3	3	0.3
30	43	8	0	0	3	3	3

26	46	2	0	0	2	2	0.0
	3						1
27	48	2	0	0	2	2	0
	8						
28	51	1	0	0	2	2	0.0
	3						3
29	53	2	0	0	1	1	0.3
	8						9
30	56	1	0	0	1	1	0.1

	3						6
31	58	2	0	0	0	0	1.3
	8						1
32	61	1	0	0	0	0	0.2
	3						6
33	63	1	0	0	0	0	0.6
	8						5
34	66	2	0	0	0	0	2.2
	3						5

Рис. 3.10. Три слагаемых (A1-A3) общего распределения
(а) животных по весу селезенки



В заключение следует сказать несколько слов о процедуре построения частотного распределения в среде Excel. Она выполняется с помощью макроса "Гистограмма", вызываемого командой меню "Сервис" \ "Анализ данных". Одна из важных подготовительных операций перед запуском макроса – это определение "интервала карманов", т. е. разбиения всего диапазона изменчивости признака на серию интервалов одинаковой ширины (dX), и ввод на лист Excel их граничных значений. В нашем примере при ширине интервала dX = 25 исходно были взяты следующие значения "карманов": 0, 25, 50, 75, и т. д. до 675 мг. Макрос "Гистограмма" подсчитывает число вариантов, попавших в "карман", и в результирующей таблице ставит эти частоты в соответствие с *левой* границей интервала.

Если гистограмму (полигон частот) строить по таким данным, распределение окажется смещенным влево на величину, равную половине ширины интервала. Однако наиболее представительной вариант для каждого интервала выступает его *центр*, и именно с ним нужно соотносить подсчитанные частоты. На простой диаграмме такой нюанс, скорее всего, останется незамеченным, но при имитационном моделировании эта скрытая особенность макроса приведет к смещению оценок средних арифметических.

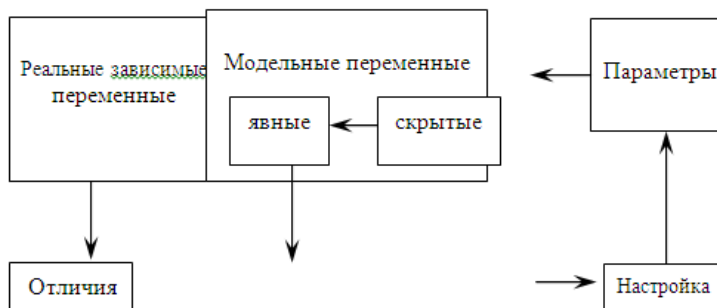
Исходя из этих соображений, полученный вначале ряд *левых* (меньших) значений начала интервалов для нашего примера был преобразован в ряд значений *центров* интервалов (первое значение 0 внесено для наглядности иллюстраций и в расчетах участия не принимает): 0, 12.5, 37.5, 62.5, 87.5, 112.5 и т. д. до 662.5 мг.

Именно эти числа были использованы для расчета теоретических частот (A) каждого из частных распределений и для построения соответствующих диаграмм.

Скрытые переменные

Скрытые переменные входят в группу модельных (внутренних) переменных (рис. 3.11), играют важную роль в построении имитационных моделей и требуют к себе особого внимания. Их общая функция состоит в детализации модельного описания, в направленном поиске неизвестных причин динамики изучаемого объекта.

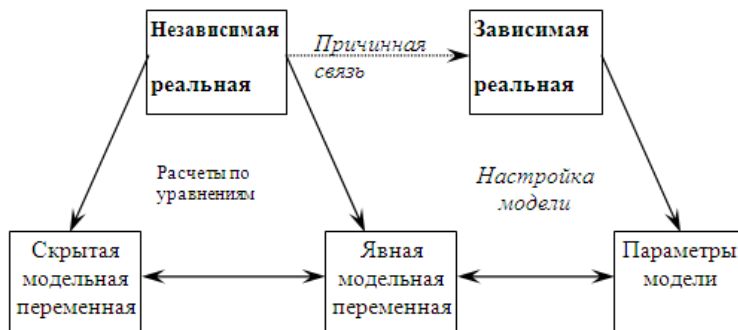
Рис. 3.11. Место скрытых переменных в имитационной системе



В рамках имитационной системы между всеми модельными переменными (независимыми, явными, скрытыми) устанавливаются вполне определенные отношения. Скрытые переменные играют роль смягчающего буфера, который позволяет вписать в жесткий каркас эмпирических данных упругие модельные построения.

В целом получается, что имитационная модель контролируется реальными данными с двух сторон – со стороны опорных *реальных независимых* переменных (основа для расчетов модельных переменных) и со стороны *реальных зависимых* переменных (основа для настройки параметров модели) (рис. 3.12).

Рис. 3.12. Отношения между переменными в имитационной системе



Латентные переменные всегда оказываются вычисляемыми, т. е. формально зависимыми. Однако чаще всего интересен как раз тот случай, когда оцениваются значения переменной, внешней по отношению к изучаемому объекту, определяются уровни неизвестного фактора среды, иными словами, когда оцениваются значения по существу независимой переменной.

Несмотря на отсутствие непосредственного контроля со стороны реальных данных, в рамках имитационной системы можно организовать косвенную зависимость латентных переменных от реальных – через структуру модели. Существуют два пути опосредованного контроля: 1) через общность модельных параметров, и 2) с помощью "упругих" конструкций модели, т. е. функций, заданных лаконичными уравнениями определенного типа.

Второй путь рассмотрен в разделе **Декомпозиция кривой** (разделение противоположных эффектов, аккумулированных в одной реальной зависимой переменной), а также в разделе **Анализ распределения** (разделение мультимодального распределения на серию унимодальных нормальных).

Первый же способ показан ниже на примере камерной модели распространения загрязнителя в водной среде – распространения сточных вод Кондопожского ЦБК в Кондопожской губе Онежского озера (Литвинова, Коросов, 1998).

Кондопожский ЦБК издавна загрязняет воду Онежского озера и в первую очередь – органическими веществами. Поэтому он всегда привлекал пристальное внимание экологов, ведущих мониторинг состояния озера. При этом одним из основных вопросов был и остается характер распространения загрязняющих веществ в Кондопожской губе и интенсивность их поступления в открытое Онего. В попытках решить этот вопрос использовались как простые расчеты, так и моделирование гидрологической ситуации. Первый метод не может с достаточной полнотой учесть сложность процессов переноса веществ. Второй метод требует детального предварительного гидрологического исследования, что представляет отдельную проблему. Известным компромиссом может выступить имитационное моделирование, соединяющее простоту средств количественного описания с возможностью исследовать динамику процесса загрязнения.

В нашей модели принципиальным оказывается именно игнорирование всех реальных процессов перераспределения загрязнителя на акватории залива. В центр внимания попадают не частные процессы, но генеральный результат распространения стоков – более или менее устойчивое сохранение градиента концентраций загрязнителя в направлении от вершины залива к его устью. Речь идет не столько о моделировании процесса выноса загрязнителя, сколько об имитации его результата.

Для моделирования процесса распространения стоков ЦБК выбрана камерная модель (Безель, 1987). Моделируемая система предстает как множество камер, содержащих то или иное количество вещества (переменные), которое по каналам обмена может переходить из камеры в камеру. Доля передаваемого вещества выступает в качестве *констант переноса* (параметры модели).

Мы располагали данными по объемам промышленных (стоки КЦБК) и естественных (сток р. Суна) поступлений БО (бихроматная окисляемость, мг/л; эта величина принята нами в качестве оценки содержания органических соединений в воде), а также данными замеров БО для каждой из камер в 1985–1989 гг. По направлению к выходу из губы концентрация БО в воде падает (табл.3.13). В модели использовали общее количество (содержание) загрязнителя в каждой из камер (Y), рассчитанное как произведение объема камеры на концентрацию загрязнителя.

Таблица 3.13. Различие между концентрациями БО в камерах и средней величиной БО (19.31 мг/л)

Глубина	Зоны поверхности			
	зона 1	зона 2	зона 3	зона 4
уровень 1	11.5	2.9	-0.99	-0.68
уровень 2		0.35	-1.46	-0.92
уровень 3		0.67		-2.62
уровень 4				-3.62

Весь объем воды Кондопожской губы мы условно разделили на 10 камер (4 зоны и 4 слоя) объемом 0.12–1.0 км³ (табл. 3.14). Неравенство объемов камер не играет никакой роли, поскольку частные механизмы перераспределения веществ между камерами (например, скорость течения, испарение, осадки), в нашей модели не находят никакого отражения, важен лишь градиент концентраций.

Таблица 3.14. Направления переноса загрязнителя между камерами
(→ поступление, → горизонтальный, ↓ вертикальный, ↓ осаднение на дно)

Глубины		Зоны акватории губы			
		зона 1	зона 2	зона 3	зона 4
КЦБК		→ з1↓ →	з2↓ →	з3↓ →	з4↓ →
уровень 1	у 1	з1у1↓ →	з2у1↓ →	з3у1↓ →	з4у1↓ →
уровень 2	у 2		з2у2↓ →	з3у2↓ →	з4у2↓ →
уровень 3	у 3		з2у3↓		з4у3↓ →
уровень 4	у 4				з4у4↓ →
уровень 5					

Загрязнитель попадает в первую камеру (зона 1 уровень 1 или з1у1, вершина губы). Поскольку в среднем градиент концентраций из года в год остается одним и тем же, то весь годовой сброс распределяется в течение года между камерами (частично разлагается и выпадает на дно). Модель имитирует природную композицию камер, организуя “перенос” веществ по следующим направлениям:

разложение: з1у↓ → деструкция

по горизонтали: з1у1 → з2у1, з2у1 → з3у1, з3у1 → з4у1, з4у1 → в озеро;

по вертикали: з2у1 → з2у2, з2у2 → з2у3, з3у1 → з3у2, з4у1 → з4у2, з4у2 → з4у3; з4у3 → з4у4;

на дно: з1у1 → на дно, з2у3 → на дно, з3у2 → на дно, з4у5 → на дно.

Содержание БО в каждой камере складывается из того, какое количество БО в ней было, сколько поступило, ушло, разложилось.

Например, содержимое камеры з3у2 составит:

"стало = было + пришло слева + пришло сверху - - ушло вправо - разложилось - осело на дно",

$з3у2' = з3у2 + а1 * з2у2 + а2 * з3у1 - а1 * з3у2 - а3 * з3у2 - а4 * з3у2$,

где а1 - константа горизонтального переноса,

а2 - параметр вертикального переноса,

а3 - параметр разложения,

а4 - параметр оседания на дно.

То же выражение в формате Excel (табл. 3.15) примет вид:

$C9 = B9 + H8 + I6 - H9 - J9 - K9$, точнее

$C9 = B9 + D8 * \$F\$15 + D7 * \$F\$17 - D9 * \$F\$15 - D9 * \$F\$16 - D9 * \$F\18 .

Для наглядности величина каждого потока была рассчитана отдельно (табл. 3.15). Поскольку ни один из 27 промежуточных потоков загрязнителя между всеми камерами не измерялся, можно говорить о 27 скрытых переменных. Лишь одна переменная (модельное содержание БО, **Y'**) оказывается явной, которая сравнивается с аналогичной реальной переменной (**Y**).

Чтобы не выписывать все довольно очевидные расчетные формулы, остановимся на ключевых моментах.

С модельной точки зрения весь процесс длится один год. Поэтому в модели находят отражение лишь две стадии: начальное (реальное, среднестатистическое) распределение объемов БО по камерам (табл. 3.15, столбец Объемы БО Y; B4:B13) и конечное (модельное, через год)

распределение (табл. 3.15, столбец Объемы БО Y' ; C4:C13). В то же время модель отражает результат распространения стоков по всем камерам.

Самое главное здесь – однонаправленность этого процесса. Загрязнитель идет слева направо и сверху вниз. Так, реальный сток объемом 103 тыс. т (ячейка B3) поступает в камеру 31у1 (E4). Здесь вещество добавляется к тем 8 тыс. т., что уже были в этой камере (111 тыс. т, D4). От этого объема 37 % (41, H4) в течение года уходит дальше – в камеру 32у1 (41, E5), а также оседает на дно (59, K4). Поскольку камера 31у1 поверхностная, ничего сверху в нее не добавилось ($F4=0$), также ничего не уходит в нижние слои ($I4=0$), но лишь оседает на дно. Мы разделили результаты вертикального переноса и осаждения потому, что в их основе лежат явно разные процессы (проигнорированные в модели), например, возможность весеннего и осеннего перемешивания водной толщи.

Настройку параметров, констант разложения (a_4) и переноса (a_1 , a_2 , a_3) органических веществ, проводили методом минимизации различий между исходными и модельными значениями содержания БО (минимизация Φ , L15) с помощью макроса "Поиск решения" (табл. 3.15, табл. 3.16).

Таблица 3.15. Имитационная система модели распространения стоков (тыс. т/год) КЦБК по камерам водной толщи Кондопожской губы (после настройки параметров)

	A	B	C	D	E	F	G	H	I	J	K	L
1	Объемы БО			Пришло			Ушло					ф
2	Y	Y'		Σ	$\rightarrow$	$\downarrow$	Σ	$\rightarrow$	$\downarrow$	$\times$	на дно	
3	Сто 10	к 3										
4	31у 8	11		11	10		100	41	0	59		11
5	32у 28	29		68	41		40	25	15	0		1
6	33у 17	17		42	25		24	15	9	0		1
7	34у 18	14		34	15		20	12	7	0		17
8	32у 17	13		32	15		18	12	7	0		15
9	33у 14	3		34	12	9	31	13	0	18		10
10	34у 18	16		38	13	7	22	14	8	0		5
11	32у 3	1		9	7		9	3	0	5		3
12	34у 4	5		12	8		7	4	3	0		1
13	34у 2	1		2	0		1		0	1		1
14												
15	F= 26.6			a1 0.3 (течение)					Φ= 15			
16	3			= 7					9			
17				a2 0 (разложен								
18				= ие)								
19				a3 0.2 (оседание)								
20				= 1								

1		a4 0.5 (осаждени	
8		= 3 e)	

Таблица 3.16. Константы переноса загрязнителя между камерами

Перенос	a1 0.3
горизонтальный	= 7
Перенос	a2 0.2
вертикальный	= 1
Разложение в толще	a3 0
	=
Оседание на дно	a4 0.5
	= 3

Расчеты показали, что поддержание текущего распределения БО в воде Кондопожской губы обеспечивается в первую очередь интенсивным осаждением веществ на дно ($a_4 = 0.53$) из придонных камер, в которые довольно быстро ($a_2 = 0.21$) вещества поступают из верхних слоев; в открытое озеро загрязнитель выносится с более низкой скоростью ($a_1 = 0.37$), и при этом почти не разлагается в водной толще ($a_3 = 0$). Модель в целом адекватна реальности ($p < 0.001$).

Несмотря на множество камер (всего 10) и связывающих их потоков (всего 41), модель использует лишь 4 параметра. Каждый параметр участвует в расчете потоков для нескольких камер. Благодаря этому расчетные коэффициенты становятся независимыми как от объема камер, так и от концентрации в них БО и играют роль относительных констант переноса. Три первых коэффициента (a_1 , a_2 , a_3), *общих* для многих уравнений, формируют достаточно жесткий каркас для всей модели. Все рассчитанные с их помощью переменные (объемы потоков между камерами) оказываются жестко связанными и сориентированными на обеспечение сходства между реальными и модельными объемами БО в камерах. И только один параметр (a_4 , коэффициент осаждения) вычисляет неподконтрольную переменную, объем донных накоплений. Тем не менее некоторый контроль за поведением этой переменной есть, поскольку параметр a_4 также общий для уравнений расчета объемов осадков из всех 4 придонных камер.

Попытки ограничиться тремя первыми параметрами не давали адекватного модельного описания ($p > 0.05$). Эти расчеты показывали, что существуют еще какие-то пути распространения загрязнителя, кроме переноса. Предположение о его оседании на дно и ввод в уравнения дополнительного параметра a_1 обеспечили высокое сходство модельных и реальных значений БО ($p < 0.001$).

На наш взгляд, очень важно глубже посмотреть на эту операцию. Вопрос не в том, чтобы любыми средствами "подогнать" расчетные значения к реальным данным (см. подробнее раздел **Адекватность модели**). Речь идет о том, как сделать это с минимальным ослаблением жесткости модели, к тому же на базе ясной теории и хотя бы косвенных фактов. Только при таких условиях ввод неуправляемой латентной переменной оправдан.

В случае с Кондопожском ЦБК накопление донных отложений есть безусловный факт (Васильева, 1990). В этом отношении ввод соответствующего параметра и переменной был уместен. Более того, рассмотренная модель "заставляет" исследователей оценить объемы донных накоплений для формирования цельной картины протекающих процессов. Как ни странно, объемы донных отложений в Кондопожской губе изучались недостаточно полно (Поляков, 1973), тем не менее эти данные подтверждают основные выводы, сделанные по результатам моделирования (максимальные накопления наблюдаются недалеко от места сброса стоков).

Для того чтобы пройти путь создания скрытых параметров до конца, в заключение рассмотрим нашу попытку ввести еще 4 скрытых переменных для

каждой придонной камеры (z_{1y1} , z_{2y3} , z_{3y2} , z_{4y4}), эти переменные управляются своими константами осаждения: a_{41} , a_{42} , a_{43} , a_{44} . В этом случае для поддержания постоянства концентраций во время настройки все коэффициенты почти обнулились, кроме констант выпадения на дно в первых трех зонах ($a_{41}=0.9$, $a_{42}=1$, $a_{43}=0.2$). Модель симитировала моментальное осаждение загрязнителей в непосредственной близости от места выпуска, идеально описывая модельные данные. Однако высокая степень адекватности ($p \ll 0.001$) не компенсирует встроенный дефект конструкции, ее излишнюю гибкость, вследствие чего параметры модели утратили относительность, а значит, и возможность осмысленной интерпретации. Платой за большую адекватность модели становится утрата биологического смысла.

Предложенное описание потоков миграции поллютантов пока не может претендовать на существенную новизну в научном плане, однако оно показывает сильные стороны метода количественного описания распространения загрязнений с помощью имитационного моделирования, возможность абстрагироваться от частных характеристик гидрологических процессов при восстановлении ситуации в водоеме. Имитационное моделирование может обеспечить не только детализацию, но и генерализацию исследуемого явления.

Экстремумы

В задачах оптимизации обычна ситуация, когда обнаруживается несколько экстремумов (для нас важны минимумы) подконтрольной функции – один глобальный и множество локальных (Иванищев и др., 1989). Иначе говоря, существует несколько наборов значений параметров, при которых целевая функция невязки прекращает уменьшаться. Получить эти наборы можно, если использовать разные стартовые значения параметров при каждой очередной настройке модели. В качестве примера рассмотрим задачу изучения плодовитости дафний в разных растворах токсиканта (см. раздел **Декомпозиция кривой**, табл. 3.9). Если в качестве начальных значений параметров взять единицы, результаты настройки будут существенно отличаться от тех, что были получены (табл. 3.17).

Таблица 3.17. Имитационная система изменения плодовитости дафний (E) в серии возрастающих концентраций лигнина [Lig]. После настройки модели при исходных значениях параметров, равных единице

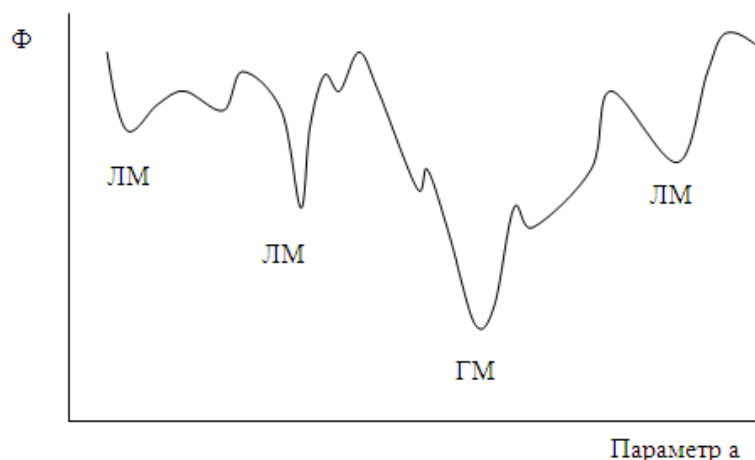
[Lig]	E	Eabs	Ed	Emod	ϕ		
0.1	81	9	0	9	5143	a=	24.5
5	115	48	0	48	4466	b=	0.42
50	131	127	1	126	30	c=	0.000001
						d=	3
150	109	202	69	133	567		3.55
200	45	228	191	36	73		
					$\Phi =$	2570	F= 1.2

Несмотря на то, что эта настройка не дает адекватной модели ($F=1.2$, $p>0.1$), функция невязки дальше не уменьшается. Достигнут локальный минимум, который явно больше по величине, чем было получено при других стартовых условиях. Начав с отдаленной области "ландшафта целевой функции", макрос оптимизации "застрял" в ближайшем понижении этого рельефа.

Возникает вопрос, существуют ли процедуры, обеспечивающие поиск и выбор лучших решений? Существует ли метод обнаружения глобального минимума? Исходя из общих формальных соображений, можно ожидать, что

для глобального минимума характерна и более широкая "воронка" на ландшафте минимизируемой функции (рис. 3.13).

Рис. 3.13. График условной целевой функции (Φ) с глобальным (ГМ) и локальными (ЛМ) минимумами



Видимо, это соображение и порождает предложения вначале отыскивать окрестности глобального минимума с помощью быстрых, простых и грубых методов оптимизации, например метода случайного поиска или метода φ -преобразования, а затем искать и собственно глобальный минимум, но уже с помощью точных методов таких, например, как метод Ньютона, метод градиента (Иванищев и др., 1989).

Второй подход, который называют "методом биологического правдоподобия" (Прицкер, 1987), состоит в том, что в качестве стартовых величин параметров берут значения, имеющие биологический смысл (в то время как модельные формулы имеют прозрачную биологическую интерпретацию). По существу, речь идет о *содержательном* способе поиска окрестностей глобального минимума целевой функции. Последующая настройка с помощью "сильных" методов дает "лучшие" значения параметров. На наш взгляд, для небольших исследовательских моделей такой путь наиболее приемлем. Он не только вынуждает глубоко прорабатывать тему, заранее планируя результат (возможные значения параметров), но и обеспечивает осмысленный подход к построению моделей, поскольку заставляет вести поиск неформальных доводов по использованию тех или иных модельных формул, которые должны иметь "биологический смысл".

Пример с дафниями как раз из этой области: знание биологии вида и общие биологические соображения позволили предложить стартовые значения коэффициентов, которые оказались лишь немного исправлены в процессе настройки. Эти значения мало похожи на те, что получены при свободном и "бессмысленном" поиске (ср. табл. 3.9 и табл. 3.17), особенно в отношении потенциальной плодовитости ($a=97.8$ и $b=0.84$ против $a=24.5$ и $b=0.42$).

Теоретические соображения позволяют обнаружить лишь область глобального минимума функции невязки, но не самый минимум. Более того, в пределах этой области может быть обнаружена серия "конкурирующих" минимумов. Зачастую оказывается, что если брать разные стартовые значения модельных параметров, то их результирующие оценки могут существенно варьировать.

Причины различий результатов настройки становятся ясными из анализа функции невязки. Например, для модели смертности гадюк (табл. 2.12, см. раздел **Имитационная система**) общее значение Φ рассчитывается как сумма частных отличий (ϕ_i) модельных переменных от реальных (графа Φ). Сравнение отдельных значений ϕ_i в этой графе показывает, что они отличаются. Так, наименьшее значение относится к 1996 году ($\phi=9.1$), а

наибольшие – к 1995 г. ($\phi=58$). Поскольку задача настройки состоит в обнулении целевой функции Φ , программа оптимизации будет стремиться свести к нулю “в первую очередь” наибольшее значение – для 1995 г. Понятно, что оцененный таким образом приоритет тех или иных точек (объектов) оказывается формальным, иначе говоря, случайным. Очевидно, что для правильной настройки нужно специально (неформально) определить приоритеты разных объектов, а затем учесть их при расчете целевой функции.

Этот третий подход к определению наилучшего решения (к оценке вклада отдельных объектов в общее значение целевой функции) есть определение весовых коэффициентов B_i в исходной формуле целевой функции (см. раздел **Имитационная система**):

$$\phi_i = B_i \cdot (Y_{im} - Y_i)^2,$$

$$\Phi = \sum \phi_i.$$

Здесь i – индекс объекта (точки, строки в модели, шаги модели...), для которого известно реальное значение переменной Y , а также рассчитывается модельное значение Y_m .

В случае с гадюкой формула принимает вид:

$$\phi_i = B_i \cdot (m'_i - m_i)^2,$$

где

i – это отдельный год,

m – число отловленных меченых особей,

m' – их расчетное число для того же года.

Как определить коэффициент B ? Если он должен выражать степень “важности”, значимости отдельного значения переменной на фоне прочих, то первое, что приходит в голову, это оценка репрезентативности каждой их точек и пропорциональное назначение весов для них.

В случае с отловами гадюк отдельные значения m имеют случайную компоненту, связанную с разными объемами выловов как в год учета числа меченых в популяции (n_i), так и в предыдущий год мечения (n_{i-1}). Это значит, что оценить вес отдельных значений m можно через произведение $B = (n_i) \cdot (n_{i-1})$. Формула целевой функции примет вид:

$$\phi_i = (n_i) \cdot (n_{i-1}) \cdot (m'_i - m_i)^2.$$

В этом случае чем больше был объем выборки, по которой определяли значения повторных встреч m_i , тем больший вклад оно привнесет в общее значение целевой функции Φ . Можно видеть, что с точки зрения репрезентативности данных оценки за 1996 г. ($B\% = 0.37$) почти в два раза “весомее”, чем оценки за 1995 г. ($B\% = 0.22$) (табл. 3.18, графа В, %).

Понятно, что процедура оптимизации, стремящаяся минимизировать общее значение Φ , будет изменять параметры таким образом, чтобы преимущественно уменьшались самые большие значения ϕ_i , модель будет настраиваться по наиболее репрезентативным точкам. Если приоритеты расставлены верно, будет получено единственное решение.

Таблица 3.18 Имитационная модель снижения числа меченых гадюк (M') в островной популяции в предположении постоянства численности (N), объемов отхода (N_d) и пополнения (N_b) (до настройки параметров)

	A	B	C	D	E	F	G	H	I	J	K
1	Год	n	m	N'	dm	M	m	Φ	$n_i \cdot n_{i-1}$	B, %	$B \cdot \Phi$
2	1994	158		5000		158					
3	1995	365	185	5000	0.12	1420	10	58	57670	0.22	12.6
4	1996	273	103	5000	0.128	127	9.1	99645	0.37	3.39	

5	199	21	10	500	0.1	11	5	26	58422	0.2	5.63	
	7	4		0		5				2		
6	199	23	9	500	0.1	10	5	17	50932	0.1	3.16	
	8	8		0		4				9		
7												
8			N =	500				Σ	10	26666	1	24.7
				0					9	9		
9			Nd =	500								
10			Nb =	500								
0												

В нашем случае введение весовых коэффициентов обеспечило однозначную настройку с параметрами $N = 3145$, $N_d = 242$ (табл. 3.19). Это результат, хотя и незначительно, но отличается от настройки без взвешивания, когда были получены значения $N = 3086$, $N_d = 228$ (табл. 2.12). Вместе с тем этот пример показал работоспособность принципа взвешивания: отличие модели от реальности в исходном варианте составило для 1996 г. 3.9 (табл. 2.12), а при взвешивании – 2.8 (табл. 3.19). Взвешивание изменило приоритеты оптимизации.

Аналогичная задача возникает при одновременной настройке нескольких моделей, связанных общими переменными. В этом случае каждая из моделей вычисляет разные переменные, имеющие отличие в единицах измерения и по абсолютным значениям (различия уровней переменных могут достигать нескольких порядков). Поскольку одновременная настройка нескольких моделей означает определение суммарной целевой функции по всем переменным, то понятно, что при этом отличия “крупномасштабных” переменных также будут большими и могут “поглотить” отличия “мелкомасштабных” переменных, взять на себя формальный приоритет. Поддерживая паритет неформальной значимости показателей, необходимо нивелировать этот эффект. Результат будет достигнут, если умножить функцию невязки для отдельной модели на обратную величину от средней арифметической данной переменной. Этот случай расчетов весовых коэффициентов рассмотрен в разделе **Сети связей**.

Таблица 3.19. Имитационная модель снижения числа меченых гадюк (M') в островной популяции гадюки в предположении постоянства численности (N), объемов отхода (N_d) и пополнения (N_b) (после настройки параметров по взвешенным переменным)

	A	B	C	D	E	F	G	H	I	J	K
1	Год	n	m	N'	dm'	M	m'	Φ	ni • ni-1	B, %	B • Φ
2	1994	158		3145		158					
3	1995	365	18	3145	0.077	146	17	1.2	57670	0.22	0.25
4	1996	273	10	3145	0.077	135	12	2.8	99645	0.37	1.06
5	1997	214	10	3145	0.077	124	8.5	2.4	58422	0.22	0.52
6	1998	238	9	3145	0.077	115	8.7	0.1	50932	0.19	0.02
7											
8			N =	3145			Σ	6.5	266669	1	1.85

9		Nd	242
		=	
1		Nb	242
0		=	

В конце концов результаты моделирования связаны не с точностью полученной модели, а с углублением понимания изучаемого вопроса. Но это определяется уже не функцией невязки, а связанностью аргументов.

Сети связей

До сих пор в нашем поле зрения в основном находились простые модели, формирующие на выходе одну переменную, $Y = f(A, X, T)$. Чаше возникает задача описания сложной системы, в которой отдельные переменные взаимно определяют друг друга. В природной среде экологическая система не может быть замкнутой, поэтому помимо внутренних системных переменных модель должна учитывать влияние внешних (независимых) переменных. В общем плане сложная многокомпонентная модель призвана имитировать динамику нескольких переменных (Y), зависящих друг от друга и от факторов внешней среды (X):

$$Y = f(A, Y, X, T),$$

где

$Y = \{Y_1, \dots, Y_m\}$ – множество зависимых переменных (всего m);

$X = \{X_1, \dots, X_p\}$ – множество независимых переменных (всего p);

$A = \{a_1, \dots, a_k\}$ – множество параметров (всего k);

$i = \{1, 2, \dots, T\}$ – счетчик времени.

Общая модель такой системы оказывается сложенной из нескольких более простых моделей (всего q), описывающих динамику каждой отдельной переменной:

$$Y_1 = f(A_{k1}, Y_{m1}, X_{p1}, T),$$

$$Y_2 = f(A_{k2}, Y_{m2}, X_{p2}, T) \dots$$

$$Y_q = f(A_{kq}, Y_{mq}, X_{pq}, T).$$

Каждая частная модель может иметь форму уравнения множественной регрессии:

$$Y_i = a_0 + a_1 \cdot Y_{m,i-1} + a_2 \cdot X_{p,i} + \dots$$

В соответствии с блочным принципом конструирования сложных моделей каждая частная модель должна быть сначала настроена самостоятельно (поскольку опирается на собственные реальные данные), а затем отрегулирована в контексте общей модели (поскольку имеет общие переменные с другими моделями). Речь идет об увязывании нескольких независимых моделей, когда выход одной модели играет роль входа для другой модели. Каскад моделей, перенимающих друг у друга результаты внутренних расчетов, позволяет провести совместную настройку всех параметров и скорректировать ранее полученные значения. По существу, происходит синтез набора модельных блоков в единую супермодель, которая может существенно изменить представление о динамике системы, сложившееся при раздельной настройке частных моделей. Такой синтез будет эффективным, если предварительно получить достаточно полное представление о поведении ее частей, т. е. иметь набор примерных значений всех параметров. Таким образом, настройка сложных моделей должна идти в два этапа: сначала настраиваются частные модели, затем – общая модель в целом.

В качестве примера рассмотрим модель локальной популяции травяной лягушки в заповеднике “Кивач”. Детальное описание динамики численности отдельных внутривидовых групп, их взаимной обусловленности и зависимости от факторов среды отражено в литературе (Kutenkov, Mosiyash, 2000). Для описания изменения численности каждой группы использовано 5

уравнений множественной регрессии, включающих только значимые коэффициенты:

– число кладок (NS) пропорционально встречаемости взрослых лягушек год назад:

$$NS_i = a_1 \cdot Nad_{i-1};$$

– число не погибших кладок (NL) пропорционально числу отложенных кладок и зависит от соотношения между испаряемостью и обилием осадков (ER):

$$NL_i = b_1 \cdot NS_i + b_2 \cdot ER_i;$$

– встречаемость сеголетков (N0+) пропорциональна числу не погибших кладок и средней температуре ее развития в текущий летний период (TS):

$$N0_{i+1} = c_0 + c_1 \cdot NL_i + c_2 \cdot TS_i;$$

– встречаемость годовиков (N1+) определяется числом сеголетков в предыдущий год и, кроме того, зависит от температуры в период их нагула (в июле предыдущего года, TJ) и глубины снежного покрова во время осенних морозов предыдущего года (HS):

$$N1_{i+1} = d_0 + d_1 \cdot N0_{i-1} + d_2 \cdot TJ_{i-1} + d_3 \cdot HS_{i-1};$$

– встречаемость взрослых особей пропорциональна встречаемости годовиков один и два года назад, а также зависит от температуры текущей зимовки (TW):

$$Nad_i = e_0 + e_1 \cdot N1_{i-2} + e_2 \cdot N1_{i-1} + e_3 \cdot TW_i,$$

где a, b, c, d, e – коэффициенты регрессии в разных моделях.

Соответствующие коэффициенты регрессии (параметры регрессионных моделей) приведены в табл. 3.21, A, например, вторая и четвертая модели с параметрами регрессионных моделей имеют вид:

$$NL_i = b_1 \cdot NS_i + b_2 \cdot ER_i = 0.521 \cdot NS_i + 4.419 \cdot ER_i.$$

$$N1_{i+1} = d_0 + d_1 \cdot N0_{i-1} + d_2 \cdot TJ_{i-1} + d_3 \cdot HS_{i-1} = 28.415 + 0.058 \cdot N0_{i-1} + 0.001 \cdot TJ_{i-1} + 1.522 \cdot HS_{i-1}.$$

Те же уравнения со значениями параметров после обобщенной настройки стали такими:

$$NL_i = 0.497 \cdot NS_i + 6.668 \cdot ER_i.$$

$$N1_{i+1} = 18.652 + 0.0214 \cdot N0_{i-1} - 0.002 \cdot TJ_{i-1} + 0.8329 \cdot HS_{i-1}.$$

Все модели адекватны, расчетные переменные коррелируют с исходными на уровне около 0.9, общий коэффициент корреляции между всеми исходными и расчетными данными составляет 0.89 (данные были предварительно нормированы для нивелировки отличий по размерности признаков, см. Коросов, 1996: с. 44).

Цель последующей работы состоит в том, чтобы все представленные модели "замкнуть" друг на друга, когда выход любой частной модели на каждом временном шаге становится входом для другой модели. Поскольку в природной популяции любая возрастная группа через год формирует следующую возрастную группу, в модели организуется каскад ссылок на значения переменных, разнесенных на один временной шаг ($Y1_i$ влияет на $Y2_{i+1}$, $Y2_{i+1}$ влияет на $Y3_{i+2}$, а $Y3_{i+2}$ влияет на $Y1_{i+3}$). В этом смысле можно говорить о построении динамической модели с сетью прямых и обратных связей.

На листе Excel этот каскад зависимостей предстает в форме множества ссылок с предыдущих шагов модели на следующие. Для примера рассмотрим представление всех частных моделей в формате Excel на 5-м временном шаге (блок M10:Q10) (табл. 3.22):

– число кладок: $NS_i = a_1 \cdot Nad_{i-1}$ или M10=M\$2*Q9;

– число не погибших кладок: $NL_i = b_1 \cdot NS_i + b_2 \cdot ER_i$ или N10=N\$2*M10+N\$3*G10;

– встречаемость сеголетков: $N0_{i+1} = c_0 + c_1 \cdot NL_i + c_2 \cdot TS_i$ или O10=O\$1+N9*O\$2+O\$3*N10;

– встречаемость годовиков: $N1_{i+1} = d_0 + d_1 \cdot N0_{i-1} + d_2 \cdot TJ_{i-1} + d_3 \cdot HS_{i-1}$ или

$$P_{10} = P\$1 + P\$2 \cdot O_9 + P\$3 \cdot I_9 + P\$4 \cdot K_9;$$

– встречаемость взрослых особей: $Nad_i = e_0 + e_1 \cdot N_{i-2} + e_2 \cdot N_{i-1} + e_3 \cdot TW_i$ или

$$Q_{10} = Q\$1 + Q\$2 \cdot P_8 + Q\$3 \cdot P_9 + Q\$4 \cdot J_{10}.$$

Таблица 3.21. Оценка адекватности моделей, настроенных разными способами (коэффициенты корреляции)

Параметр ов ↓ номера индексы →	Переменные в частных моделях					Обща я моде ль
	NS	NL	N0+	N1+	Nad	
	a	b	C	d	e	
А. Параметры частных регрессионных моделей						
0			-11.992	-97.1	28.415	
1	184.7	0.521	0.00013	3.059	0.058	
2		4.419	0.99981	6.013	0.001	
3				0.614	1.522	
r =	0.90	0.87	0.64	0.96	0.97	0.89
В. Совместная настройка частных моделей (стартовые значения – коэффициенты регрессии)						
0			-12.379	-31	16.101	
1	229.7	0.488	0.00083	9.77	0.0293	
2		6.569	0.98363	0.48	0.0172	
3				0.93	0.6939	
r =	0.98	0.91	0.66	0.96	0.89	0.91
С. Совместная настройка частных моделей (стартовые значения – единицы)						
0			-12.793	-1.24	18.652	
1	218.3	0.497	0.00089	12.32	0.0214	
2		6.668	1.00749	-1.88	-0.002	
3				1.009	0.8329	
r =	0.97	0.90	0.65	0.96	0.91	0.91

В качестве начальных значений в модели приняты некоторые исходные данные. Так, для первых трех лет (1981–1983 гг.) учет встречаемости взрослых лягушек не проводился, поэтому в модели использованы реальные значения числа кладок (выделены курсивом). Для всех последующих лет число кладок рассчитывалось. Так же дело обстоит и с числом взрослых особей: нужны два года учета годовиков, чтобы рассчитать число взрослых. Поэтому значения для взрослых особей за 1983–1984 гг. взяты из исходной таблицы (выделены курсивом). Все остальные модельные величины рассчитаны.

В обобщающей таблице 3.22 приведены результаты моделирования после настройки модельных параметров. В соответствии с алгоритмом настройка общей модели должна идти с использованием единственного значения целевой функции. Получить его можно, суммировав значения целевых функций всех частных моделей. В табл. 3.22 эти частные значения

содержатся в блоке ячеек S2:W2, например, функция невязки для модели числа не погибших кладок равна $T2 = \text{СУММ}(T6:T15)$.

Понятно, что каждая частная функция невязки рассчитана как сумма квадратов отклонений модельных значений от исходных, например, $T10 = (N10 - C10)^2$.

Настройка многокомпонентной модели сопряжена с проблемой разной размерности моделируемых переменных. При суммировании небольшие значения функции невязки затеряются среди больших значений. В нашем примере доля целевых функций для моделей встречаемости особей разного возраста (0+, 1+, ad) составляет 0.01% (11+141+25) от суммарного значения целевой функции ($\Phi = 1748085$), в котором доминируют функции для числа кладок. В этом случае модель будет настраиваться, в основном ориентируясь на крупномасштабные переменные, в ущерб прочим.

Для решения этой проблемы достаточно ввести весовые коэффициенты перед каждой частной функцией невязки:

$$\Phi = \sum B_i \cdot \phi_i,$$

$$\phi_i = \sum (Y'_{ij} - Y_{ij})^2.$$

По сравнению с порядком отличия переменных (например, штуки встреченных особей и сотни подсчитанных кладок) значения целевых функций отличаются в квадратичной степени. Поэтому весовые коэффициенты должны быть обратно пропорциональны квадратам значений отдельных зависимых переменных, относящихся к разным модельным блокам. Кроме того, они не должны изменяться во время настройки, чтобы не исказить ее ход.

Таблица 3.22. Имитационная система модели динамики численности и состава популяции травяной лягушки

	A	B	C	D	E	F		H	I	J	K	L	M	N	O	P	Q	S	T	U	V	W	X
1												a0=			-12	-31	16.1	ϕ_1	ϕ_2	ϕ_3	ϕ_4	ϕ_5	Φ
2												a1=	230	0.49	0	9.8	0.03	717013	1030895	11	141	25	1748085
3		Число	Успеш-									a2=		6.57	0.98	0.5	0.02	0.2565	1.62048	1.4	0.5	0.4	4.24964
4		кладок	ность	Возраст особей								a3=				0.9	0.69	0.98	0.91	0.66	0.96	0.89	0.95
5	Год	Sp	Suc	0+	1+	>3	Mul	Ts	Tj	Tw	Hs		Sp	Suc	0+	1+	>3	Sp	Suc	0+	1+	>3	
6	1981	2806	1749				53.8						2806	1724					637				
7	1982	1631	189	3			14.4	14.3	16		0		1631	891	3				492960	0			
8	1983	1723	364	4	14	6	-33	15.2	17	-13	32		1723	627	3	7	6		68975	0.5	47		
9	1984	1350	6	1	45	13	-71	14.5	16	-11	5		1378	205	2	39	13	801	39754	2	33		
10	1985	3177	2530	5	7	14	105	15.2	14	-10	5		2986	2149	3	5	10	36353	145293	5.1	4.7	16	
11	1986	2061	749	2	7	9	-15	14	16	-15	0		2292	1022	3	7	7	53483	74294	1.4	0.1	3	
12	1987	1320	1157	2	5	4	41.5	12.8	14	-17	20		1670	1088	1	7	5	122735	4713	0.9	6.2	0.8	
13	1988	711	632	5	2	4	-47	16.4	19	-15	21		1130	242	5	5	6	175630	151727	0.1	7.4	3.9	
14	1989	865	30	2	39	7	-102	15.3	16	-15	11		1371	0	3	43	6	255742	900	0.8	16	1.3	
15	1990	1076	570	1	10	8	-48	14.3	15	-14	20		1345	343	2	15	8	72268	51641	0.5	26	0.1	
16																							
17	B=	4.E-07	2.E-06	0.13	0.004	0.02																	

* Параметры среды:

Mul – осадки минус испарение (см); Ts – T°C летом; Tj – T°C в июле; Tw – T°C января; Hs – глубина снега в ноябре (см)

Для этой роли лучше всего подходит обратная величина квадрата среднего значения каждой исходной зависимой переменной. Например, для модели числа кладок:

$$B17 = 1/\text{CPЗНАЧ}(B6:B15)^2 \text{ (табл. 3.22).}$$

Умножив частные целевые функции на соответствующие веса, получаем вполне сравнимые значения (ряд S3:W3), которые теперь можно объединить в общее значение целевой функции ($X3 = \text{СУММ}(S3:W3) = 4.24964$) и использовать при настройке всех модельных блоков одновременно.

Начинать настройку можно с разных стартовых позиций, задав в качестве исходных значений параметров рассчитанные ранее коэффициенты регрессионных моделей, или присвоить им любые другие значения, например единицы. Для контроля результатов счета удобно пользоваться коэффициентами корреляции между исходными и расчетными рядами

значений для каждой переменной, а также для всех (нормированных) переменных вместе.

Вначале рассмотрим результаты настройки параметров общей модели в целом, взяв полученные коэффициенты регрессии в качестве стартовых. Новые модели несколько отличаются от регрессионных (табл. 3.21, В). Часть модельных переменных стала лучше описывать исходные данные, соответствующие коэффициенты корреляции повысились на величину 0.01–0.08; одна из переменных, численность взрослых особей, стала хуже соотноситься с исходной ($r = 0.89$ против $r = 0.97$). Немного увеличилась и общая коррелированность всех параметров ($r = 0.91$ против $r = 0.89$). Эти отличия, конечно, не значимы, но важен сам факт модификации моделей. Сходные результаты (табл. 3.21, С) дает настройка с единицами в качестве стартовых значений параметров (если макрос запускается несколько раз, см. раздел Параметры макроса “Поиск решения”). Это говорит о закономерности происшедших изменений.

Почему же обобщенная модель отличается от серии уравнений регрессии? Как это вообще возможно – улучшение какой-либо частной модели по сравнению с регрессионным уравнением? По всей видимости, дело во *взаимодействии* частных моделей. В регрессионном анализе используется единственный ряд реальных данных, которые однозначно детерминируют значения коэффициентов регрессионной модели. Иное дело настройка модельных параметров; в этом повторяющемся процессе рассчитываются все новые и новые значения модельных переменных. На каждом шаге итерации любая частная модель использует в качестве аргументов значения, рассчитанные другой моделью, отличные от значений предыдущих итераций. Эти новые расчетные значения, в свою очередь, передаются по цепочке дальше, следующим моделям. Таким образом все модели влияют на все модели, а все параметры настраиваются по всему массиву переменных. Это очень важный момент структурирования многокомпонентной модели. Оказывается, что если изменение параметров одной частной модели может без особого ущерба для своей модели улучшить результаты другой частной модели, то эти параметры изменятся. Некоторые параметры могут “подыграть” чужой, а не “своей” модели, за счет чего и происходит улучшение адекватности отдельных частных моделей и всех моделей в целом. В нашем случае улучшение коснулось всех моделей, кроме модели изменения численности взрослых особей.

Почему же происходит ухудшение некоторых частных моделей? Казалось бы, при этом не может идти запланированная минимизация, а должна увеличиваться общая функция невязки! Такая ситуация напоминает результаты компонентного анализа, когда факторные нагрузки первой главной компоненты распределяются между признаками неравномерно: большая часть признаков получает высокие нагрузки, меньшая часть – низкие. Плеяда коррелированных признаков, большая по объему, условно говоря, “перекачивает в себя” изменчивость из плеяд меньшего размера, обеспечивая своим членам высокие уровни нагрузок. Видимо, аналогичная ситуация складывается и с распределением участия параметров отдельной модели в настройке соседних частных моделей. Если изменение параметров отдельной модели улучшает адекватность сразу нескольких других моделей, то такое изменение будет происходить даже с некоторой утратой адекватности самой этой модели (до тех пор, пока выигрыш превышает проигрыш, пока повышение частной функции невязки компенсируется снижением общей функции невязки). По существу, группа моделей начинает пользоваться параметрами единичной модели. Именно таким “донором” оказалась модель встречаемости взрослых лягушек (Nad), ее адекватность данным снизилась с $r = 0.97$ до $r = 0.89$ (табл. 3.21, А, В).

Можно показать на конкретных цифрах общий механизм такого перераспределения ролей. Например, изменение параметров модели Nad (табл. 3.21) позволило уменьшить модельную численность для 1985 г. (с 14 до 10 экз.), что привело к снижению расчетного числа кладок (NS) в 1986 г. (с 2500 до 2292 кладок) и приближению его к уровню реальных значений, т. е. отличие по кладкам в данный год снизилось с 192 503 до 53 483 (ячейка S11, табл. 3.22), а общая функция невязки уменьшилась с 935 194 до 717 013 (ячейка S2).

При рассмотрении результатов счета, изменивших исходные значения параметров (табл. 3.21, B), напрашивается мысль, что изменения могли бы быть более существенными, если в качестве стартовых значений принять бессмысленные величины, например единицы. Такой расчет был проведен (табл. 3.21, C). Несмотря на ожидания, в целом настройка дала очень похожие результаты, правда, запускать макрос поиска решения пришлось несколько раз (см. подробнее раздел Параметры макроса). Основные изменения коснулись четвертой модели встречаемости годовиков (N1+). В новой версии модели увеличилась роль встречаемости сеголетков в предыдущем году (12.32 против 9.77), в результате свободный член практически потерял свое значение (-1.23 против -31). Модель годовиков оказалась переориентированной на внутривидовые факторы, внешние и необъясненные влияния сократили свое участие.

Как же следует относиться к такого рода перестройкам частных моделей "за счет внутренних ресурсов"? В чем состоит смысл модификаций частных моделей в рамках общей? На наш взгляд, здесь все дело в коллизии между детерминистическим характером моделей и стохастическими по природе реальными переменными. Полученные в природе оценки численности разновозрастных групп являются выборочными и не точно соответствуют истинным значениям популяционных характеристик.

В детерминистической же модели все исходные значения переменных выступают как абсолютно репрезентативные характеристики популяции. Одни переменные участвуют в прогнозе других как истинные популяционные показатели. В рамках общей модели все переменные оказываются взаимосвязанными в самой жесткой форме: численность каждой группы в данное время по цепям связей детерминирует свои значения и численность других групп на много шагов вперед.

Когда дело доходит до настройки детерминистических моделей по стохастическим данным, открывается широкое поле для свободного выбора приоритетов. При настройке каждой частной модели по отдельности (процедура множественного регрессионного анализа) полученное уравнение характеризует основной тренд изменения переменной, изменение численности изучаемой группы. При этом варьирование исходных данных ограниченного объема в рамках частной модели может приобрести статус закономерных отклонений, и модель получится неточной. В результате при объединении частных (и неточных) моделей в одну общую частные тренды в изменении численностей разных групп могут не согласовываться друг с другом!

Изменение параметров при дополнительной настройке общей модели в целом есть не что иное, как процедура согласования частных трендов друг с другом. По существу, это избавление от избыточной изменчивости, *сглаживание эмпирических рядов данных*, в котором участвует весь объем полученной информации. Значения модельных переменных оказываются более "правильными", "репрезентативными", более близкими к общепопуляционным характеристикам, нежели исходные реальные данные. Это позволяет использовать модель в более обоснованной оценке влияния абиотических факторов на жизнь данной конкретной популяции, в частности травяной лягушки.

Какие возможности открывает построенная модель?

Во-первых, можно детализировать описание возрастного состава популяции, если в структуру модели ввести скрытые переменные, отражающие численность разновозрастных животных, объединенных в модели в рамках одной группы "взрослые". Необходимую дополнительную информацию можно получить, например, изучив возрастной состав даже на единичной выборке взрослых особей.

Это же явление можно исследовать иным способом, поскольку модель позволяет очень просто воссоздавать эффекты запаздывания реакций на действие факторов. В модели встречаемость взрослых в данном (i -м) году определялась численностью годовиков год ($i-1$) и два ($i-2$) года назад. При добавлении еще пяти членов может быть организована задержка до 8 лет.

Наконец, современная форма представления модели в среде Excel позволяет оперативно исследовать влияние новых факторов среды, вводя в формулы дополнительные слагаемые. В отличие от регрессионного анализа перед расчетом значений параметров не требуется производить никаких предварительных преобразований исходных данных, поскольку типы частных моделей при необходимости с легкостью могут быть перестроены. Например, в уравнение можно ввести квадратичный член, логарифм и т. д. Значимость параметров новых членов можно определить по изменению общей адекватности модели. Таким путем можно достаточно быстро отобрать факторы, определяющие динамику исследуемого признака. Повышение уровня адекватности общей модели по сравнению с суммарной адекватностью частных моделей важно использовать как критерий репрезентативности исходных данных.

Прогноз как гипотеза

Имитационная модель биосистемы – это не конечный результат исследования, но промежуточный этап процесса исследования. Когда некоторый эмпирический материал получает количественное оформление в виде имитационной модели, открывается несколько путей для продолжения исследований на этой базе. В первую очередь – это экстраполяция на область неизвестных значений зависимых переменных. На первый взгляд, упоминание об экстраполяции как исследовательской процедуре противоречит духу этой книги, пропагандирующей интерполяционные модели. Однако в определенном смысле экстраполяцию можно рассматривать как способ построения более общих интерполяционных моделей, как очередной шаг моделирования. И вот почему.

Допустим, что исследовалась биосистема на ограниченном (во времени или пространстве) наборе данных и была построена имитационная модель, хорошо их описывающая. По мере накопления новых материалов появлялась возможность проверить работу модели на независимых массивах. Если при этом модельные предсказания уровня зависимых переменных становились все хуже, значит (как подсказывает опыт), либо параметры модели не улавливают всего размаха варьирования переменных, либо конструкция модели не учитывает неких факторов (независимых переменных), реально участвующих в динамике изучаемой системы. Такое же положение неопределенности складывается, когда на одних и тех же эмпирических данных настройка модели до уровня адекватности дает разные значения параметров. Если при этом сделать серию прогнозов (рассчитать неизвестные значения модельных переменных, выходящих за известный диапазон изменчивости реальных данных), используя *разные* модельные параметры, то появится перспективная задача выяснить, какой из прогнозов ближе к реальности. В этих случаях модель укажет на изъян, вынудит сконцентрировать внимание на тех моментах анализа природного объекта, которые исходно были плохо

структурированы, т. е. будет явно диктовать направление дальнейшей работы.

Построение отдельной модели есть лишь единичный шаг в итеративном процессе исследования. Тем не менее в публикациях обычно указывается лишь некий окончательный модельный результат длительного пути создания моделей, что создает ложное впечатление о смысле моделирования. Примеры анализа неадекватности моделей в литературе достаточно редки (Безель, 1987, с. 100-101; Павлов, 1984, 1986). В этой связи было бы интересно отследить как раз динамику исследования, возникновение неадекватных прогнозов и пути их нейтрализации.

Рассмотрим пример прогноза абсолютной численности популяции травяной лягушки на о. Кизи (Коросов, 2000). На протяжении двух лет (1997–1998 гг.) подсчитывалось число кладок икры в репродуктивных водоемах острова. Оценки совпали $N=3467$ и $N=3649$, что в случае равного представительства полов позволяет оценить абсолютную численность взрослых особей в 7000 экз.

Для расчета численности *всей* популяции необходимо каким-то образом реконструировать недостающие значения по молодым лягушкам. Предположив постоянство численности островной популяции (на это указывают как наши оценки, так и благоприятный для развития личинок гидрологический режим приозерных болот), можно реконструировать стационарное возрастное распределение и на его основе рассчитать общую абсолютную численность (табл. 3.20), опираясь на оценки демографических параметров (Северцов и др., 1998; Кутенков и др., 1990).

Величина группы годовиков рассчитывается как число икринок, из которых развились успешно перезимовавшие особи, а число икринок – как суммарная плодовитость всех самок:

$$N_0 = N_{\text{fadi}} \cdot e, [B5] = [B1] * [B2],$$

$$N_1 = N_0 \cdot s_0, [B6] = [B1] * [C1],$$

где N_{fadi} – число половозрелых самок, e – плодовитость самки.

Всего на острове учтено 3500 самок, что при средней плодовитости $e = 1800$ икринок дает $N_0 = 6300000$. Выживаемость личинок летом и сеголетков зимой сильно варьирует – от долей до нескольких процентов, для первого случая берем округленное модальное значение $s_0 = 1\%$ ($[C1] = 0.01$). В пределах возрастного распределения текущая численность каждой возрастной группы равна численности младшей возрастной группы год назад минус число погибших особей, т.е. равна численности выживших особей:

$$N_{i+1} = N_i \cdot s_i, [B7] = [B6] * [C2], \dots [B11] = [B10] * [C3] \dots$$

где N_i – численность i -го возрастного класса ($i = 1, 2, \dots, 9$), s_i – выживаемость.

Из популяционной экологии известно, что обычно молодые и взрослые особи имеют разную выживаемость. Это позволяет ввести в модель и оценивать лишь два параметра – выживаемость младших (1-5 лет) и старших (6-8 лет) особей:

$$s_{1-5} = s_1 = s_2 = s_3 = s_4 = s_5, \text{ячейка } [C2],$$

$$s_{6-8} = s_6 = s_7 = s_8, \text{ячейка } [C3].$$

Используя приведенные уравнения, уже можно сконструировать модель (столбцы A:B), но еще нельзя приступить к настройке параметров модели (C2:C3). Для этого необходимо рассчитать некие модельные значения переменных, которые следует сравнивать с реальными значениями.

В нашем распоряжении имеется единственное реальное значение переменной – число всех половозрелых особей, $N_{\text{ad}} = 7000$ экз. Необходимо поэтому рассчитать соответствующее модельное значение. Из литературы известно, что в среднем продолжительность жизни травяной лягушки составляет 8 лет; с четырех лет все особи становятся половозрелыми ($[D3]=1$), а среди трехлеток доля половозрелых составляет 20% ($[D2]=0.2$), остальные особи – ювенальные (Северцов и др., 1998; Кутенков, 1996). В

модели число взрослых особей рассчитывается как часть ото всех особей (столбец C8:C13):

$$Nad_i = N_i * p_i,$$

где Nad_i – число половозрелых особей и p_i – их доля.

Сумма этих значений дает искомое число взрослых особей: [C15]=СУММ (C8:C13).

Таблица 3.20. Реконструкция возрастной структуры популяции лягушки при разных уровнях смертности личинок и сеголетков

для $s_0=0.01$ %					для $s_0=0.001$ %				
	A	B	C	D		A	B	C	D
1		3500	0.01		1		3500	0.00	
2		1800	0.37	0.	2		1800	0.71	0.
3			0.37	1	3			0.71	1
4	ле	Все	Взросл		4	ле	Все	Взросл	
	т	особи	ые			т	особи	ые	
5	0	630000			5	0	630000		
		0					0		
6	1	63000			6	1	6300		
7	2	23574			7	2	4472		
8	3	8821	176		8	3	3174	635	
			4						
9	4	3301	330		9	4	2253	2253	
			1						
10	5	1235	123		10	5	1599	1599	
			5						
11	6	462	462		11	6	1135	1135	
12	7	173	173		12	7	806	806	
13	8	65	65		13	8	572	572	
14					14				
15		100631	700		15		20311	7000	
			0						

Теперь осталось только вызвать макрос настройки параметров и указать соответствующие ссылки: установить целевую ячейку [C15] равной значению 7000, изменяя ячейки [\$C\$2:\$C\$3]. При этом на возможные значения параметров следует ввести явные ограничения. Так, они не должны превышать 1 (не может выжить более 100 % особей) или быть меньше 0.2 (обычно выживаемость выше 20 %). Кроме этого, выживаемость амфибий в старших возрастных группах обычно ниже, чем в младших (Коли, 1979, с. 165), $C3 \leq C2$. С помощью мыши вводим эти ограничения в окно настройки, используя кнопку “Добавить” (рис. 3.14).

Рис. 3.14. Определение условий настройки модели

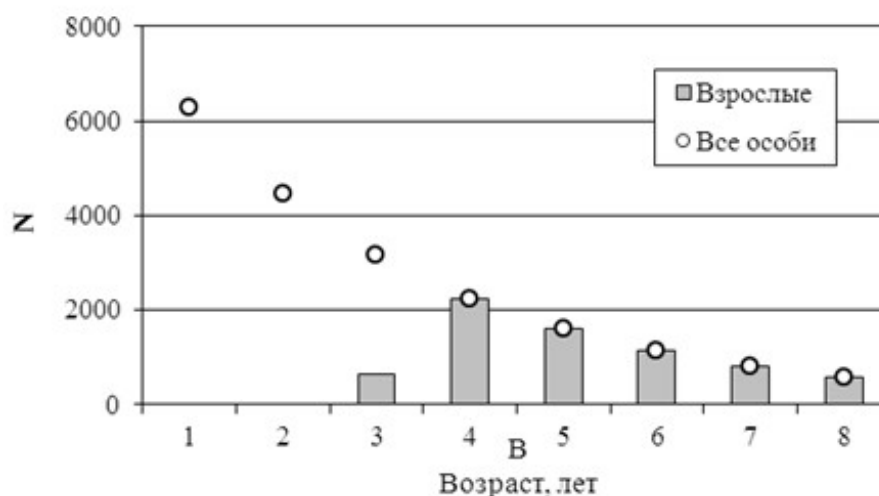
В результате настройки получили единственное решение: выживаемость в возрасте 2-5 лет равна $s_{1-5} = 37\%$, то же и в возрасте старше 5 лет - $s_{6-8} = 37\%$. Общая расчетная численность популяции травяной лягушки составила около 100 000 экз.

Представленная модель оказалась очень чувствительной к параметру "выживаемость личинок и сеголетков", s_0 , [C1]. Взяв другое значение, на уровне одной десятой процента $s_0 = 0.001\%$, в результате новой настройки получаем среднюю выживаемость для других групп $s_i = 0.71\%$, а общую численность на уровне 20 000 экз. (табл. 3.20, рис. 3.15).

Как видно из таблицы 3.20, результаты двух попыток моделирования довольно сильно различаются. Какая из этих двух моделей (данной структуры) ближе к действительности, поможет решить лишь независимая оценка одного из параметров популяции, например, выживаемости взрослых особей, или абсолютной численности одного из младших возрастных классов. Получить такие данные можно только после дополнительных исследований с использованием трудоемкого метода мечения с повторным отловом.

Моделирование показало, что для реконструкции даже усредненного возрастного состава и численности популяции лягушки первичных данных не достаточно, нужны дополнительные факты или теоретические посылки. Этот вывод очень важен, поскольку конкретизирует направления поиска. Имитационное моделирование указывает пути дальнейшего исследования.

Рис. 3.15. Реконструкция возрастного состава популяции лягушек для $s_0 = 0.001\%$



Структура понятий имитационного моделирования

Идеологическая сложность и новизна предлагаемой технологии имитационного моделирования заставляет дать общий обзор основных понятий, затронутых в книге. Стремясь к максимально емкой форме, мы составили блок-схему основных понятий, используя, как и во всех других случаях, системный подход (рис. 3.16, табл. 3.23, 3.24).

В соответствии со *структурно-функциональным* принципом схема должна содержать части (у нас – понятия), связанные друг с другом. Всего было выделено 11 базовых понятий и 12 связей между ними. Горизонтальные стрелки отображают отношения преобразования (переход информации из одной формы в другую).

Рис. 3.16. Структура основных понятий моделирования в среде Excel



Таблица 3.23. Определения понятий

Понятие	Содержание понятия
Реальность, природа	Мир, существующий вне сознания
Теория, знания	Научная форма мышления о природе
Объект исследования	Исследуемая часть природы
Исследуемая проблема	Понимание неполноты знаний в данной области
Словесное описание	Характеристика способа существования объекта
Блок-схема	Идеограмма логики мышления об объекте
Реальные данные	Любая информация об объекте (описание состояния объекта, среды, условий наблюдений)
Модельные уравнения	Представление объекта в виде формул
Переменные исходные	Множество чисел, полученных при наблюдении

Переменные модели	Множество чисел, рассчитанных по модели
Оценка адекватности	Оценка возможности использования модели
Функция невязки	Общая оценка отличия модели от реальности
Макрос настройки	Средство поиска минимума функции невязки

Принцип *иерархичности* заставляет искать соподчиненность понятий. В нашей системе с очевидностью выделяются 5 уровней иерархии процесса эмпирического исследования, самый верхний из которых есть системный подход как методология эмпирического исследования, а самый глубокий, внутренний, частный – процедура настройки параметров модели. Вертикальные стрелки блок-схемы отображают отношения включения.

Целесообразность построения в том и состоит, чтобы одного взгляда на схему было достаточно для полного охвата проблематики моделирования.

В свою очередь, *целостность* понятийной структуры обеспечена жесткой преемственностью между понятиями смежных уровней иерархии. Кстати, проведение этого принципа обеспечило очень быстрое самоструктурирование блок-схемы, но и потребовало переопределить некоторые отношения, сложившиеся в блок-схеме после использования предыдущих принципов.

Таблица 3.24. Определения отношений (показаны стрелками)

Ин-декс	Содержание отношений
1.1	Выбор доступного объекта исследования
1.2	Определение актуальности исследования
1.3	Определение цели и методов исследования
2.1	Упорядочение известной информации
2.2	Целенаправленное исследование, сбор недостающих сведений
2.3	Выбор вида модели
3.1	Организация базы данных, формирование 2 блоков переменных
3.2	Работа главного правила: выразить переменные через параметры
3.3	Программирование (структурирование) модели на листе Excel
4.1	Статистический вывод об адекватности модели
4.2	Выбор фрейма имитационной системы, завершение ее интеграции
4.3	Настройка модели с помощью макроса "Поиск решения"

Эмерджентные качества блок-схемы мы видим в дидактическом эффекте: ее лаконизм должен компенсировать необходимую детальность предыдущего описания: из набора понятий должна сложиться система.

4. Теория оптимизации

Термин «оптимизация», «улучшение ситуации», в математике обозначает процесс поиска наилучших с некоторой точки зрения параметров системы, в моделировании – поиск таких значений параметров модели, которые обеспечивают наибольшее сходство модельных и реальных переменных, минимизируют функцию невязки (целевую функцию оптимизации). Обозначим параметр модели (характеристику системы) через a , функцию невязки – через $F(a)$. Общая задача оптимизации состоит в том, чтобы найти такие a , при которых

$$F(a) = 0 \quad (\text{Иванищев и др., 1989, с. 68}).$$

Для решения задач такого рода используются численные методы последовательного приближения к правильному решению, методы итераций. В их числе метод Ньютона, градиентные методы.

Модель с одним параметром

Возьмем для начала простейшую модель с одним параметром a :

$$Y' = f(a, X, Y).$$

Пусть модель строится для имитации изменения ряда реальных значений переменной Y , зависящей от внешней переменной X . Отличие расчетных значений от модельных характеризует функция невязки, или целевая функция оптимизации, $F(a) = \sum(Y' - Y)^2$. Общая задача оптимизации состоит в том, чтобы найти такое значение параметра модели a , при котором отличие модельных предсказаний от реальных данных было бы минимальным, т. е. целевая функция отличий $F(a)$ обнулилась:

$a \rightarrow \text{оптимум},$

$$F(a) \rightarrow 0.$$

В этом плане оптимизация есть последовательный (повторяющийся, итеративный) поиск все *новых* значений параметра a_{i+1} , которые были бы лучше, чем *предыдущие* значения a_i , лучшие в том смысле, что при подстановке их в модельное уравнение функция отличий приблизилась бы к нулю:

если $0 < F(a_{i+1}) < F(a_i)$, то a_{i+1} лучше, чем a_i .

По этой причине задача настройки параметров модели есть задача *минимизации*.

При расчете нового значения параметра a_{i+1} обычно используют предыдущее значение a_i , измененное на некоторую величину $V(a_i)$ (уравнение итерации) (Васильев, 1988, с. 261):

$$a_{i+1} = a_i + k \cdot V(a_i), \quad [1]$$

где i – номер шага оптимизации (шаг итерации),

k – коэффициент пропорциональности со своим знаком.

Это вычисление требуется для «улучшения» значения параметра. Но это значит, что прибавка $V(a_i)$ еще до расчета нового значения a_{i+1} должна каким-то образом улавливать, что это улучшение произойдет. Контроль за будущим значением параметра осуществляется двояким образом: *по величине и по знаку* прибавки $V(a_i)$.

Во-первых, чем больше функция невязки, т. е. чем «хуже» прежнее значение параметра, тем больше должно быть приращение, чтобы за один шаг максимально улучшить модель. Иными словами, приращение должно быть пропорционально функции невязки:

$$V(a_i) \sim F(a_i).$$

Новая формула итерации примет вид:

$$a_i = a_{i+1} + k \cdot F(a_i). \quad [2]$$

Второй момент вычисления приращения $V(a_i)$ состоит в том, чтобы определить направление оптимального изменения параметра (уменьшать его или увеличивать), т. е. распознать знак перед приращением. Для этого требуется уловить динамику процедуры оптимизации, динамику изменения функции $F(a)$. Задача состоит в том, чтобы определить, как изменится целевая функция, если параметр изменится на величину Δa_i , какой знак будет иметь отношение приращения целевой функции

$$\Delta F(a_i) = F(a_i) - F(a_i + \Delta a_i),$$

к приращению параметра на величину Δa , т. е. какой знак будет иметь величина

$$F'(a_i) = \Delta F(a_i) / \Delta a = [F(a_i) - F(a_i + \Delta a)] / \Delta a ?$$

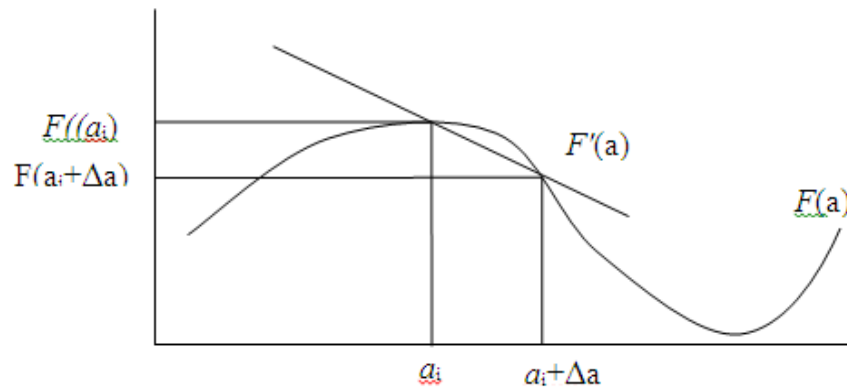
Это уравнение вычисляет дискретный аналог производной, характеризующей как раз скорость и направление преобразования некой функции (в нашем случае – целевой функции) (рис. 4.1), в нем фигурируют три члена:

Δa – условное приращение параметра; это малая величина, она задается в условиях программы оптимизации (см. свойства макроса «Поиск решения»),

$F(a_i)$ – значение целевой функции, вычисленное при прежнем значении параметра,

$F(a_i + \Delta a)$ – значение целевой функции, вычисленное при измененном значении параметра.

Рис. 4.1. Целевая функция и дискретный аналог ее производной



Если положительное приращение Δa уменьшает целевую функцию $F(a_i) < F(a_i + \Delta a)$ и сохраняет положительный знак перед приращением $\Delta F(a_i)$, значит, знак перед $F'(a_i)$ остается положительным. Если же функция невязки увеличилась, то знак перед $F'(a_i)$ поменяется на противоположный.

Рассмотренные идеи автоматического определения и величины, и знака приращения значения параметра лежат в основе метода Ньютона:

$$V(a_i) = F(a_i) / F'(a_i)$$

или

$$a_{i+1} = a_i + F(a_i) / F'(a_i) \quad (\text{Турчак, 1987; с. 159}). \quad [3]$$

С технической стороны вся процедура оптимизации выглядит следующим образом. Сначала для произвольно взятого значения модельного параметра (a_0) вычисляется исходное значение целевой функции $F(a_0)$ (как было показано выше, исходные значения параметров моделей задаются с учетом знания о существовании моделируемого явления). После этого отыскивается первое условное приращение параметра Δa , исходя из установленных свойств программы оптимизации, затем вычисляется «рабочее» новое значение параметра $a_0 + \Delta a$ и «рабочее» новое значение целевой функции $F(a_0 + \Delta a)$. Так получают характеристики двух шагов итерации, после чего появляется возможность для расчета величины и знака приращения параметра $V(a_0)$, а также и самого нового значения параметра a_1 по формуле [3]. Используя рассчитанное обновленное значение параметра, вычисляется новое значение целевой функции $F(a_1)$, после чего процедура повторяется по тем пор, пока решение не станет «оптимальным», т. е. пока целевая функция не станет равна нулю или не перестанет уменьшаться.

Здесь уместно сравнить величины приращений Δa и $V(a_i)$, каждая из которых в свое время прибавляется к текущему значению параметра a_i . Отличие их состоит в том, что сумма $(a_0 + \Delta a)$ служит только для определения «оптимального» направления дальнейшего изменения параметра, но не участвует в получении его окончательного нового значения. Сумма $(a_i + V(a_i))$, напротив, используется только для получения очередного, «более оптимального» значения, параметра a_{i+1} .

Серия вычислений производит последовательность значений параметра: $a_1, a_2, \dots, a_{i-1}, a_i, a_{i+1}, \dots, a_n$, которая может *сходиться*, т. е., в конце концов, приблизиться к некоторому стабильному значению a_i . В наилучшем случае функция невязки превращается в нуль. Отсюда следует основной смысл уравнений итерации [1, 2, 3]: когда целевая функция $F(a_i)$ становится равной нулю, найденный параметр a_i считают решением уравнения (Заварыкин и др., 1990, с. 13–17).

Обычно между моделью и реальностью остается определенное несоответствие, целевая функция $F(a_i)$ отличается от нуля, значения модельных переменных лишь в общих чертах демонстрируют сходство своей динамики с реальной системой. Более того, с каждой итерацией величина ее изменений становится все менее существенной. Процесс схождения «до нуля» может затянуться до бесконечности при отсутствии существенных улучшений модельного описания. По этой причине процесс оптимизации важно вовремя остановить, а именно в том момент, когда достигнутая величина очередного приращения станет, на наш взгляд, несущественной, меньше некоторой изначально заданной малой величины (ε), т. е. когда выполнится условие:

$$F(a_i) - F(a_{i+1}) < \varepsilon.$$

Значение этой относительной погрешности выбирают порядка $\varepsilon = 0.000001$; его можно изменить в параметрах макроса «Поиск решения».

Имитационные модели, как правило, содержат больше одного параметра. В этом случае используются многомерные методы оптимизации, в числе которых методы градиентные, случайного поиска, квазиньютоновский метод.

Модель с несколькими параметрами

Общая структура модели с несколькими параметрами задается формулой:

$$Y' = f(A, X, Y),$$

где $Y' = \{y_1, y_2 \dots\}$ – вычисляемые переменные,

$X = \{x_1, x_2 \dots\}$ – реальные внешние переменные,

$A = \{a_1, a_2, \dots a_j, \dots a_m\}$ – множество параметров модели,

$j = 1, 2, \dots m$ – индекс модельного параметра.

Отличие расчетных значений от модельных характеризует функция невязки, или целевая функция оптимизации, зависящая от множества параметров $F(A) = \sum (Y' - Y)^2$.

Задача оптимизации остается той же, что и в однопараметрической модели, – найти такие значения параметров, чтобы целевая функция отличий $F(A)$ обнулилась или перестала существенно изменяться, $F(A_i) - F(A_{i+1}) < \varepsilon$.

Также сохраняется прежняя итеративная формула, вычисляющая приблизительные значения каждого (j -го) из модельных параметров на каждом (i -м) шаге итерации:

$$a_{ji+1} = a_{ji} + k \cdot V_j(a_{ji}), \quad [4]$$

где k – некоторая константа,

$i = 1, 2, \dots n$ – номер шага итеративного процесса,

$j = 1, 2, \dots m$ – индекс параметра,

$V_j(a_{ji})$ – необходимое приращение параметра, преобразованная целевая функция, относительная скорость изменения функции невязки при изменении j -го параметра на i -м временном шаге.

В то же время процедура расчета приращения параметра $V_j(a_{ji})$ усложняется, поскольку на каждом i -м шаге оптимизации приходится, во-первых, учитывать влияние многих параметров (A) на одно общее значение целевой функции $F_i(A)$, и, во-вторых, использовать приращение общей целевой функции для вычисления приращений $V_j(a_{ji})$ каждого параметра.

Один из методов настройки многопараметрической модели – *градиентный метод*. Он использует понятие «градиента», направления максимальной скорости изменения функции.

Значение приращения каждого j -го параметра на каждом i -м шаге оптимизации вычисляется по формуле

$$V_j(a_{ji}) = F'_j(a_{ji}) / -\text{grad}F, \quad [5]$$

где $F'_j(a_{ji})$ – скорость изменения функции невязки при изменении отдельного (j -го) параметра на величину Δa , когда остальные параметры остаются неизменными;

$\text{grad}F$ – градиент функции невязки, скорость изменения функции невязки при совместном изменении всех параметров на величину своего шага Δa .

Прежде чем начать более детальное рассмотрение этой формулы, важно отметить общее сходство с аналогичным выражением [3] для однопараметрической модели. Числитель формулы характеризует направление (знак) и степень (величину) изменения модели при изменении *одного* j-го параметра на величину Δa . Знаменатель формулы характеризует направление (знак) и степень (величину) изменения модели при изменении *всех* параметров на величину Δa .

В целом формула вычисляет *относительный вклад* отдельного параметра в общее «улучшение» модели (контролируемое целевой функцией невязки $F(A)$), она позволяет усилить эффективность изменения наиболее «перспективных» параметров (в смысле минимизации целевой функции). Чем более сильное уменьшение функции невязки вызывает изменение параметра на величину Δa , тем большей получится величина необходимой прибавки $V_j(a_{ji})$ к этому параметру.

Смысл членов уравнения состоит в следующем.

Числитель $F'_j(a_{ji})$ как скорость изменения функции невязки в математическом плане есть производная от этой функции: $F'_j(a_{ji}) = dF_j/da$, которая в дискретном виде приближенных компьютерных расчетов определяется по формуле: $F'_j(a_{ji}) = dF_j/da \approx \Delta F_j/\Delta a =$

$$= [F(a_{1i}, a_{2i}, \dots a_{ji} + \Delta a, \dots a_{mi}) - F(a_{1i}, a_{2i}, \dots a_{ji}, \dots a_{mi})]/\Delta a,$$

где $a_{ji} + \Delta a$ – новое значение j-го параметра на i-м шаге итерации,

Δa – прирост значения параметра за один шаг (Турчак, 1989, с. 186; Шуп, 1990, с. 169).

Знаменатель, градиент функции невязки, $\text{grad}F$, представляет собой обобщение изменений целевой функции при изменении всех параметров. В геометрическом смысле градиент есть вектор, величина и направления которого равны максимальной скорости *увеличения* многомерной функции (скалярного поля) (Кудрявцев, Демидович, 1989, с. 397). Если функция задана одним параметром, то градиент в определенной точке a_i по знаку и величине равен производной функции в этой точке. Если функция задана по m параметрам, то градиент вычисляется как средняя квадратичная по всем m параметрам:

$$\text{grad}F = \sqrt{(dF_1/da_1)^2 + (dF_2/da_2)^2 + \dots + (dF_j/da_j)^2 + \dots + (dF_m/da_m)^2}$$

(Баранов, Рябчук, 1988, с. 121). В дискретной форме выражение принимает вид:

$$\text{grad}F = \sqrt{(\Delta F_1/\Delta a)^2 + (\Delta F_2/\Delta a)^2 + \dots + (\Delta F_j/\Delta a)^2 + \dots + (\Delta F_m/\Delta a)^2}.$$

Градиент показывает максимальную скорость *увеличения* функции. Поскольку задача оптимизации модели состоит в минимизации функции невязки, для расчета направления следующего итеративного шага значение градиента берут с *обратным знаком*, в этом случае он указывает направление максимальной скорости *уменьшения* функции невязки.

Порядок итерации остается таким же, но для настройки многопараметрической модели в пределах одного шага итерации приходится делать много промежуточных вычислений. На каждом шаге вычисляются m новых значений $a_j + \Delta a$, m новых значений $F'_j(a_{ji})$, общее значение градиента, m новых значений необходимой прибавки $V_j(a_{ji})$ и, наконец, m новых «улучшенных» значений параметров a_{j+1} . Процесс итеративных расчетов завершается, если выполнилось условие:

$$F(A_i) - F(A_{i+1}) < \varepsilon.$$

Одна из распространенных модификаций градиентного метода названа *методом наискорейшего спуска*. Имеется в виду спуск к минимуму функции невязки по склону некоего ландшафта возможных значений модельных параметров. В этом случае новое направление градиента на каждом шаге не рассчитывается. Вместо этого до тех пор, пока функция невязки уменьшается, при определении новых значений параметров a_{i+1} используют одно и то же приращение $k \cdot V_j(a_{ji})$, вычисленное в начале процедуры спуска. При этом изменение целевой функции специально контролируют. Как только она при очередном шаге перестает убывать, проводят переопределение направления

дальнейшего спуска (перерасчет приращений $V_j(a_{ji})$) и вновь устремляются к минимуму $F(A)$. Скорость схождения рядов $a_1, a_2, \dots, a_n$ зависит от величины коэффициента k , который может изменяться программно, существенно ускоряя процесс; для этого разработаны эффективные процедуры (Шуп, 1990).

Параметры макроса «Поиск решения»

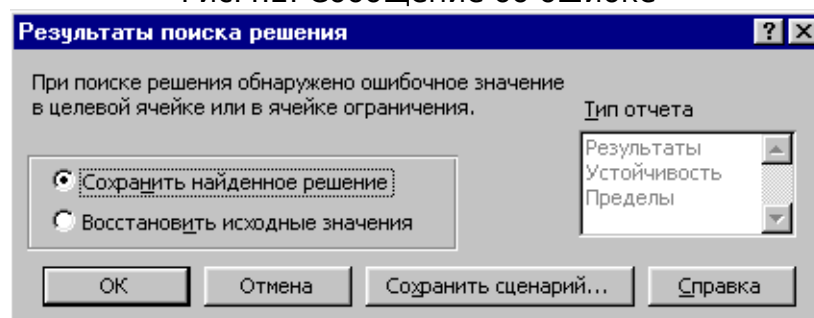
Справочная информация о порядке пользования и смене настроек макроса «Поиск решения» представлена в нескольких вкладках «Справочной системы» пакета MS Excel (знак вопроса на главной панели меню). В «Содержании» – это пункт «Выборка и анализ данных/ Нелинейная множественная оптимизация», в «Предметном указателе» – рубрика «поиск решений». На вкладке «Поиск» новую информацию можно найти по ключевым словам «поиск решен...», «solver». Основные же подсказки дает кнопка «Справка» на панелях окна макроса.

Вызов процедуры оптимизации осуществляется командой «Поиск решения...» меню «Сервис». В появившемся окне необходимо сделать некоторые изменения, чтобы макрос стал работать так, как нужно. В основном справочная информация достаточно хорошо освещает все этапы этого процесса. Некоторые варианты заполнения главного окна макроса «Поиск решения» уже были рассмотрены в предыдущих разделах. Здесь важно сделать несколько дополнений по контролю за ходом и результатами процедуры настройки.

«Выполнить» – кнопка запуска макроса. Казалось бы, объяснять ее назначение излишне, после нажатия начинается оптимизация модели. Хитрость заключается в том, что иногда для настройки требуется неоднократно запускать эту программу! При этом каждая новая попытка улучшает модель, иногда возникает два устойчивых набора параметров, один из которых все же лучше другого. Причина этого «явления» состоит в том, что при каждом новом запуске заново переопределяется величина приращения параметра (Δa), служащая для расчета новых «поисковых» значений целевой функции. Значение Δa рассчитывается автоматически и определяется стартовой «атмосферой» модели. Изменение исходных значений параметров (после каждой новой настройки) влечет за собой изменение величины приращения и самого характера оптимизации. По этой причине результаты поочередных актов настройки могут отличаться. Правда отличия сохраняются всего лишь на протяжении 3–5 итераций. Впрочем, так происходит далеко не всегда, а только в случае с моделями, имеющими «мягкий» каркас, несколько скрытых переменных или плохо структурированные данные.

«Ограничения» – диалог ввода, удаления и изменения граничных условий изменения переменных и параметров. В рамках имитационной системы это свойство макроса используется просто как выходной фильтр. Фактически же программа позволяет решать задачи математического программирования. Условия могут накладываться как на отдельные значения, так и на блоки ячеек, как на значения параметров, так и на модельные значения. Использовать ограничения следует в двух случаях. Во-первых, это теоретические соображения о биологически правдоподобном диапазоне возможных значений параметров и переменных. Например, многие природные характеристики не могут быть отрицательными (численность, размеры, масса, концентрация), хотя модельные уравнения могут вычислять такие значения. Во-вторых, для исправления вычислительных ошибок, если во время настройки модельный блок заполнится значками диеза (##, за ними прячутся сообщения об ошибках: #ДЕЛ/0!, #ЧИСЛО!) и появится окно «Результаты поиска решения» с текстом, предупреждающим об обнаружении ошибки (рис. 4.2).

Рис.4.2. Сообщение об ошибке

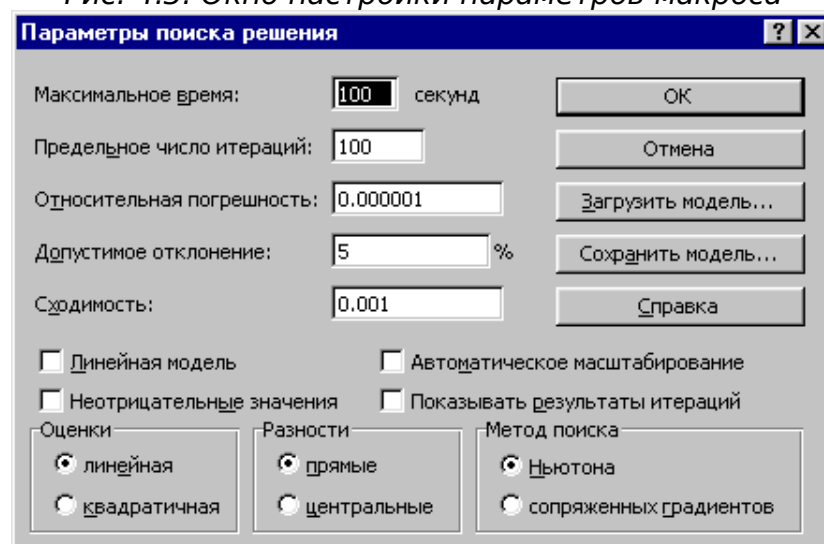


Источники ошибок в арифметических моделях почти всегда одинаковы: деление на ноль, попытка вычислить логарифм или извлечь корень из отрицательного числа. Выход из ситуации один – отменить результаты настройки и задать условия, ограничивающие область изменения значений, которые вызвали ошибку. Для этого нужно нажать кнопку «Добавить» и, выделив мышью ячейки (или ячейку), задать им граничные значения (например, «больше или равно нулю», ≥ 0).

«Параметр» – вызов окна (рис. 4.3) для изменения характера работы макроса. Рассмотрим важнейших из них.

«Предельное число итераций» – ограничение длительности настройки. Одна итерация состоит в двукратном пересчете функции невязки (с базовыми и измененными значениями параметров), в вычислении направления и величины изменения базовых параметров, в расчете новых значений базовых параметров. Каждая итерация улучшает модель. Хорошо структурированные модели сходятся менее чем за 10 шагов, плохо структурированные – пытаются преодолеть барьер в 100 итераций. Если появляется окно «Текущее состояние поиска решения» с вопросом «Продолжить?», значит, скорее всего, модель считает «что-то не то» и требует критического пересмотра.

Рис. 4.3. Окно настройки параметров макроса



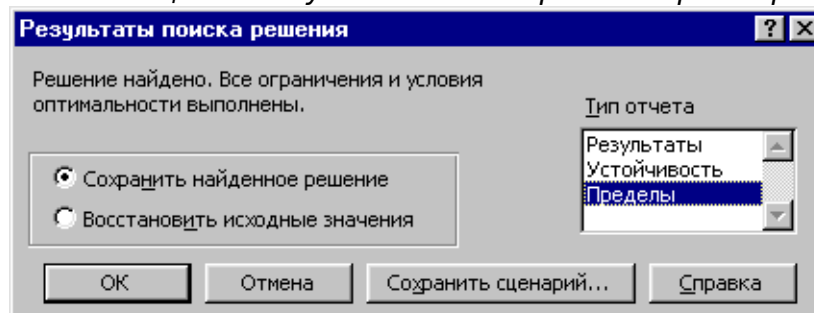
«Сходимость» – предельная минимальная величина разности между значениями целевой функции, полученными в смежных итерациях, ε . По ее достижении (когда $F(A_i) - F(A_{i+1}) < \varepsilon$) процедура настройки модели прекращается. Чем больше величина ε , тем быстрее закончится настройка, тем грубее будет модельное описание. Этот параметр относительный, он контролирует скорость изменения целевой функции. Для рассмотренных биоэкологических моделей достаточна уже величина $\varepsilon = 0.1$, однако по умолчанию назначается величина $\varepsilon = 0.001$, которая дает немного более

точные результаты, почти не сказываясь на времени выполнения процедуры. Этот параметр можно никогда не изменять. Для большей обоснованности прекращения настройки берут усредненное значение по пяти последним итерациям.

«Относительная погрешность» (или точность) – предел округления величин $F(A_i)$ и $F(A_{i+1})$ при расчетах разности между ними. Этот параметр, как и предыдущий, также существенно определяет ход процесса настройки, но это абсолютный показатель, он устанавливает не предельную скорость, но предельное отклонение целевой функции от заданного уровня.

«Допустимое отклонение» – различие между значением, к которому стремится целевая функция (оно устанавливается с самого начала), и ее значением, достигнутым в процессе оптимизации. Этот параметр влияет не на ход настройки, а на возможность отобразить ее результаты. Когда достигнутое и запланированное значения целевой функции отличаются друг от друга не более чем на величину «допустимого отклонения», в окне «Результаты поиска решения» появляется текст «Решение найдено, все условия и ограничения оптимально выполнены», а также предоставляется возможность создать отчет по его результатам, выбрав соответствующий пункт в окошке «Тип отчета», например «Пределы» (рис. 4.4).

Рис. 4.4. Сообщение об успешной настройке параметров модели



Когда же отклонение больше допустимого, текст в окне «Результаты поиска решения» меняется: «Поиск не может найти подходящего решения» и доступа к выбору отчета нет. Поскольку при имитационном моделировании обычно «запланировано» значение целевой функции, равное 0 (которое, понятное дело, *никогда* не достигается), то и настройщик модели всегда видит только такое окно.

Вообще говоря, для настройки имитационной модели эти сообщения безразличны, поскольку имитационная система имеет блок оценки адекватности, дающий содержательный ответ на вопрос о близости модели к реальности. Однако если пользователь хочет в *любом случае* получать отчеты по результатам настройки, нужно сделать следующее. Сначала настроить модель с помощью макроса «Поиск решения» с автоматически заданными установками. Затем вновь запустить макрос «Поиск решения» и «установить целевую ячейку, равную значению» только что полученного значения функции невязки с листа Excel; нажать кнопку «Выполнить». Сразу же откроется окно «Результаты поиска решения» первого типа.

«Линейная модель» – способ быстрой и грубой предварительной оценки параметров модели. На наш взгляд, гораздо продуктивнее «метод биологического правдоподобия», первичное назначение параметров, исходя из знания предмета и теоретических соображений.

«Метод Ньютона» – процедура вычисления направлений поиска оптимума, рассмотренная выше. Характеризуется большим объемом вычислительной работы, но более высокой скоростью поиска решения. Приемлем для большинства задач, но может пропустить глобальный экстремум, имеющий узкое основание на условном ландшафте целевой функции.

«Метод сопряженных градиентов» – процедура вычисления направлений поиска оптимума, рассмотренная выше. От метода Ньютона отличается меньшим объемом вычислительной работы на каждом шаге, но относительно невысокой скоростью поиска решения. Приемлем для большинства задач, поскольку более детально (значит, и медленнее) обследует структуру ландшафта целевой функции. Для рассмотренных моделей оба метода давали одинаковые результаты.

«Оценка», «Производные» – варианты методики оптимизации, мало влияющие на результаты настройки рассмотренных моделей.

Приемы работы в MS Excel

В «море» литературы, посвященной Excel, все приемы освещены с исчерпывающей полнотой. При этом многие возможности программы остаются невостребованными, их просто не удастся найти в бесконечной череде советов. В частности, автор еще не встречал пользователя Excel, который бы знал о существовании встроенного макроса оптимизации. Практика построения множества имитационных систем позволяет предложить несколько советов по оптимальной организации среды для ее создания.

Быстрая команда «автозаполнения».

Во введении показана процедура «автозаполнения» с протяжкой мыши. Еще быстрее можно выполнить эту операцию, дважды кликнув левой кнопкой на черном крестике. Формула или значения из текущей ячейки автоматически заполнят столько нижележащих ячеек, сколько заполненных ячеек содержится в соседнем столбце (приоритет у левого столбца). Стоит поэкспериментировать с автозаполнением, если выделены несколько ячеек, например с числами 1 и 2.

Разделители.

После установки Windows в качестве разделителя целой и дробной части числа устанавливается запятая. Это очень неудобно по многим причинам. Во-первых, после экспорта данных из среды Excel в Word приходится исправлять все запятые на точки, в соответствии с новыми требованиями к формату чисел в публикациях. Во-вторых, числа с запятыми не воспринимаются другими программами, если их экспортировать для специальных сложных расчетов. Таким «капризным» пакетом является, например, Statgraphics for Windows. Не менее несуразно использование по умолчанию точки с запятой в качестве разделителя в тех же экспортных списках. Заменить запятую на точку, а точку с запятой на запятую можно в окне «Языки и стандарты» вкладки «Числа». Для этого нажмите кнопку «Пуск» панели Windows, выберите «Настройка» \ «Панель управления» \ «Языки и стандарты».

Специальные символы и другие полезные кнопки.

Большая часть функций, с помощью которых создаются формулы в ячейках Excel, имеет кириллическое написание (и кто только это придумал!). Поэтому удобнее работать с клавиатурой в русском регистре. Тем не менее главный знак формулы, организующий абсолютную ссылку, – это знак валюты (\$), для него приходится часто переключаться в латинский регистр и обратно. Еще один часто используемый знак – возведение в степень (^) – также доступен только в латинице. Лишние переключения требуют времени, внимания и увеличивают число ошибок. Упростить ситуацию можно, если вывести указанные символы как кнопки на панель инструментов, например, сразу же после знака вопроса на главной панели. В этом случае для ввода символа в нужное место формулы (с позиции курсора) достаточно щелкнуть соответствующей кнопкой.

Поместить новые кнопки на главную панель можно следующим образом. Установите курсор мыши на любой панели, правой кнопкой откройте контекстное меню, выберите «Настройка», вкладка «Команды». В Категории

«Вставка» найдите значки «Знак возведения в степень» и «Знак доллара». Мышью перетащите значки на главную панель, закройте окно настройки. Как минимум, еще одна очень полезная кнопка должна быть представлена на панели инструментов. Это кнопка «Тип диаграммы». Открыв то же окно настройки, выберите категорию команд «Диаграмма». Почти в самом конце списка команд найдите кнопку (не строку меню) «Тип диаграммы» и поместите ее на панель, например, рядом с кнопкой «Мастер диаграмм».

Кнопка «Тип диаграммы» позволяет не только быстро изменять тип диаграммы, но и моментально строить диаграмму нужного типа. Достаточно отметить область данных, после чего, нажав кнопку, выбрать нужный тип диаграммы и отпустить кнопку мыши. Полезно также создать собственную небольшую панель, например, «Связи», поместив на нее несколько кнопок с панели «Зависимости».

Тип шрифта.

Казалось бы, не все ли равно, какой шрифт будет в окне Excel. Однако при выполнении любой работы стоит заглядывать в будущее и видеть результат научной деятельности в виде отчета или публикации. Становится ясным, что данные из среды Excel придется переносить в документ отчета, составленный, скорее всего, в среде Word. К документам же всегда предъявляются жесткие требования по формату шрифтов, общепринят Times New Roman Cyr 12. Именно такой шрифт имеет смысл установить в Excel, предвидя операции копирования данных из электронной таблицы Excel в текстовые таблицы Word. Эта установка делается в окне «Сервис»\ «Параметры» на вкладке «Общие». Там же уместно установить по 3 листа в новой книге и свой каталог для записи файлов. После изменения типа шрифта для улучшения обзора **масштаб листа следует уменьшить**. Например, при разрешении дисплея 600*800 оптимален масштаб листа Excel 77%.

Вставка и форматирование таблицы Excel в среде Word.

Выделяете нужный блок ячеек в таблице Excel, копируете в буфер обмена, переключаетесь в Word, устанавливаете курсор в нужном месте, вставляете блок из буфера. Проблемы возникают с расположением и форматом таблицы и чисел в ней: вдруг оказывается, то таблица расположена не по центру, имеет разномастные столбцы, слишком длинную дробную часть чисел, «прилипших» к правому краю ячеек. Для придания таблице достойного вида часть операций необходимо выполнить перед экспортом в среде Excel, часть – в Word. Быстрый способ отформатировать ширину столбца по размеру содержимого ячеек состоит в следующем. Установите курсор мыши между буквенными названиями столбцов, из белого креста он превратится в черный со стрелками по бокам. Теперь двойной клик изменит ширину столбца точно по ширине наибольшей записи, содержащейся в какой-либо ячейке левого от курсора столбца.

Для форматирования чисел, выделив нужный блок ячеек в таблице Excel, откройте окно «Формат»\ «Формат ячейки». На вкладке «Число» выберите из «Числовых форматов» «Числовой» и установите нужное число десятичных знаков. Чтобы не загромождать таблицу, устанавливайте минимальное число знаков после запятой. Если исходные данные имеют хх значащих цифр, значение средней арифметической и стандартного отклонения могут содержать на одну значащую цифру больше (ххх), значение ошибки средней – на две значащих цифры (хххх). Например, если промеры имеют порядок 3.8, 5.6, 8.0, средняя может быть 5.32, стандартное отклонение – 1.23, ошибка – 0.412. Обозначения всех граф и настройку их ширины удобнее выполнять в среде Excel, чем в Word.

Таблица из Excel вставляется в Word выровненной по левому краю. Для центрирования таблицы на листе Word сначала следует ее правильно выделить. Для этого подведите курсор мыши к верхнему краю первого столбца таблицы, где он преобразуется в черную вертикальную стрелку.

Нажав левую кнопку мыши, ведите ее направо, выделяя все столбцы, и далее – за правую границу таблицы; отпустите кнопку мыши. Помимо почерневших столбцов справа от таблицы появится колонка добавочных черных квадратиков. В этом режиме щелкните кнопку выравнивания «по центру» панели «Форматирование». Таблица переместится в центр листа. Теперь нажмите Shift и один раз стрелку влево. Дополнительные черные квадратики исчезнут. Вновь щелкните кнопку выравнивания «по центру» панели «Форматирование». Текст в каждой ячейке отформатируется по центру.

Литература

- Акоф Р. Л., Сасиени М.** Основы исследования операций. М.: Мир, 1971. 534 с.
- Альтшуллер Г. С.** Творчество как точная наука. М.: Сов. радио, 1979.
- Анохин П. К.** Системные механизмы высшей нервной деятельности. М.: Наука, 1979. 453 с.
- Антомонов Ю. Г.** Моделирование биологических систем: Справочник. Киев: Наукова думка, 1977. 260 с.
- Аристотель.** О частях животных. М., 1937. 220 с.
- Ашмарин И. П., Васильев Н. Н., Амбросов В. А.** Быстрые методы статистической обработки и планирования экспериментов. Л.: Изд-во ЛГУ, 1975. 78 с.
- Баранов А. В., Рябчук Э. В.** Численные методы в инженерных задачах. Волгоград, 1988. 128 с.
- Безель В. С.** Популяционная экотоксикология млекопитающих. М.: Наука, 1987. 130 с.
- Безель В. С., Большаков В. Н., Воробейчик Е. Л.** Популяционная экотоксикология. М.: Наука, 1994. 80 с.
- Блауберг И. В., Юдин Б. Г.** Понятие целостности и его роль в научном познании. М.: Знание, 1972. 48 с.
- Блауберг И. В., Юдин Б. Г.** Становление и сущность системного подхода. М.: Наука, 1973. 270 с.
- Богданов А. А.** Системная организация материи // На переломе: Философские дискуссии 20-х годов. М.: Политиздат, 1990. С. 422-435.
- Боровиков В. П., Боровиков И. Д.** STATISTICA – статистический анализ и обработка данных в среде WINDOWS. М.: Филинь, 1997. 608 с.
- Брагинский Л. П., Величко И. М., Щербань Э. П.** Пресноводный планктон в токсической среде. Киев: Наукова думка, 1987. 180 с.
- Браунли К. А.** Статистическая теория и методология в науке и технике. М.: Наука, 1977. 468 с.
- Бэкон Ф.** Сочинения в двух томах. Т.1: О достоинстве и приумножении наук. М.: Мысль, 1977. 567 с.
- Васильев Ф. П.** Численные методы решения экстремальных задач. М.: Наука, 1988. 549 с.
- Васильева Е. П.** Донные отложения // Экосистема Онежского озера и тенденции ее изменения. Л.: Наука, 1990. С. 147-174.
- Винер Н.** Кибернетика. М.: Наука, 1983. 342 с.
- Волькенштейн М. В.** Биофизика. М.: Наука, 1981. 576 с.
- Вольтерра В.** Математическая теория борьбы за существование. М.: Наука, 1976. 286 с.
- Галицкий В. В.** О моделировании продукционного процесса в растительном сообществе // Моделирование биогеоценотических процессов. М.: Наука, 1981. С. 104-118.
- Гегель Г. В. Ф.** Энциклопедия философских наук. Т.1: Логика науки. М.: Мысль, 1974. 452 с.
- Гиг Дж., ван.** Прикладная общая теория систем. М.: Мир, 1981. 733 с.
- Гильдерман Ю. И.** Лекции по высшей математике для биологов. Новосибирск: Наука, 1974. 411 с.
- Гильманов Т. Г.** Введение в количественную трофологию и экологическую биоэнергетику позвоночных в наземных экосистемах. Ч. 1: Основные модели. Пойкилотермные животные. М.: Изд-во МГУ, 1987. 180 с.
- Гиляров А. М.** Популяционная экология. М.: Изд-во МГУ, 1990. 191 с.
- Гиляров А. М.** Экология, обретающая статус науки // Природа. 1998. № 2. С. 89-99.
- Дженрич Р. И.** Пошаговая регрессия // Статистические методы для ЭВМ. М.: Наука, 1986. С. 77-112.
- Джефферс Дж.** Введение в системный анализ: применение в экологии. М.: Мир, 1981. 256 с.
- Енюков И. С.** Методы, алгоритмы, программы многомерного статистического анализа: Пакет ППСА. М.: Финансы и статистика, 1986. 232 с.
- Жаков Л. А., Меншуткин В. В.** Практические занятия по ихтиологии. Ярославль, 1982. 112 с.
- Завадский К. М.** Основные формы организации природы // Философские проблемы современной биологии. М.; Л.: Наука, 1966. С. 29-46.
- Заварыкин В. М., Житомирский В. Г., Лапчик М. П.** Численные методы. М.: Просвещение, 1990. 176 с.
- Зайцев Г. Н.** Математика в экспериментальной ботанике. М.: Наука, 1990. 296 с.
- Иванищев В. В., Михайлов В. В., Тубольцева В. В.** Инженерная экология. Л.: Наука, 1989. 144 с.
- Иванова М.Б.** Продукция планктонных ракообразных в пресных водах. Л.: Наука, 1985. 222 с.
- Ивантер Э. В., Ивантер Т. В., Туманов И. Л.** Адаптивные особенности мелких млекопитающих. М.: Наука, 1984. 318 с.
- Ивантер Э. В., Ивантер Т. В.** Экологическая структура и динамика населения мелких млекопитающих Приладожья // Фауна и экология птиц и млекопитающих Северо-Запада СССР. Петрозаводск, 1983. С. 72-99.
- Ивантер Э. В.** Земноводные и пресмыкающиеся. Петрозаводск, 1975. 96 с.
- Ивантер Э. В.** Популяционная экология мелких млекопитающих таежного Северо-Запада СССР. Л.: Наука, 1975. 246 с.
- Ивантер Э. В., Ивантер Т. В.** К экологии темной полевки (*Microtus agrestis* L.) // Экология наземных позвоночных Северо-Запада СССР. Петрозаводск, 1986. С. 64-91.
- Ивантер Э. В., Коросов А. В.** Основы биометрии. Петрозаводск, 1992. 168 с.

- Израэль Ю. А., Анохин Ю. А., Остромогильский А. А., Пословин А. Л., Белова Н. И.** Некоторые результаты осуществления мониторинга природной среды в регионе озера Байкал // Совершенствование регионального мониторинга состояния озера Байкал. Л.: Гидрометеиздат, 1985. С. 4-22.
- Ильенков Э. В.** Диалектическая логика. М.: Политиздат, 1984. с.320.
- Калинкина Н. М.** Эколого-токсикологическая оценка опасности сульфатного лигнина для гидробионтов: Дис. ... канд. биол. наук. Петрозаводск, 1993. 172 с.
- Кант И.** Критика чистого разума. М.: Мысль, 1994. 591 с.
- Касаткин А. М.** Эвристические методы в исследовании сложных систем // Математическое моделирование жизненных процессов. М.: Мысль, 1988. С. 198-210.
- Касаткин В. Н.** Семь задач по кибернетике. Киев: Вища школа, 1975. 97 с.
- Клайн М.** Математика. Поиск истины. М.: Мир, 1988. 295 с.
- Кноринг Л. Д., Деч В. Н.** Геологу о математике. Л.: Недра, 1989. 208 с.
- Коган А. Б.** О моделировании процессов нервной деятельности // Математическое моделирование жизненных процессов. М.: Мысль, 1968. С. 211-221.
- Колемаев В. А., Староверов О. В., Турундаевский В. Б.** Теория вероятностей и математическая статистика. М.: Высшая школа, 1991. 400 с.
- Коли Г.** Анализ популяций животных. М.: Мир, 1979. 362 с.
- Коросов А. В.** Алгоритмы биометрии. Петрозаводск, 1995. 32 с.
- Коросов А. В.** Имитация экологических объектов в среде Excel // Биологические основы изучения, освоения и охраны животного и растительного мира, почвенного покрова Восточной Финляндии: Международная конф. и выездная научн. сессия Отделения общей биологии РАН. Петрозаводск, 6-10 сентября 1999 г. Петрозаводск, 1999а. С. 83-84.
- Коросов А. В., Хилков Т. Н., Фомичев С. Н.** Кижские острова – «газья мекка» // Острова кижского архипелага. Биогеографическая характеристика. Тр. КНЦ РАН, сер. "Биогеография Карелии". Вып. 1. Петрозаводск, 1999б. С. 91-95.
- Коросов А. В.** Плейстная организация признаков как отражение аллометрических зависимостей // Актуальные вопросы биологии и рациональное использование ресурсов Карелии. Петрозаводск, 1989. С. 42-44.
- Коросов А. В.** Развитие системного подхода к изучению островных популяций животных (на примере обыкновенной гадюки, *Vipera berus* L.): Дисс. ... докт. биол. наук. Петрозаводск, 2000. 333 с.
- Коросов А. В.** Экологические приложения компонентного анализа. Петрозаводск, 1996. 152 с.
- Кудрявцев В. А., Демидович Г. П.** Краткий курс высшей математики. М.: Наука, 1975. 384 с.
- Купий В. Т.** Моделирование в биологии и медицине. Философский анализ. Л.: Изд-во ЛГУ, 1989. 188 с.
- Кутенков А. П.** Земноводные и пресмыкающиеся // Фауна и флора заповедников СССР. Фауна заповедника "Кивач". М., 1988. С. 9-11.
- Кутенков А. П., Панарин А. Е., Шкляревич Ф. Н.** Экология бесхвостых амфибий Карелии и Кольского полуострова // Наземные позвоночные животные в заповедниках Севера Европейской части РСФСР. М., 1990. С. 54-70.
- Кутенков А. П., Панарин А. Е., Шкляревич Ф. Н.** Экология размножения бесхвостых амфибий Карелии и Кольского полуострова // Наземные позвоночные животные в заповедниках Севера Европейской части РСФСР. М.: ЦНИЛ Главохоты РСФСР, 1990. С. 54-70.
- Кювье Ж.** Рассуждения о переворотах на поверхности земного шара. М.; Л., 1937. 368 с.
- Ладенко И. С.** Имитационные системы (методология исследований и проектирования). Новосибирск: Наука, 1981. 300 с.
- Лакин Г. Ф.** Биометрия. М.: Высшая школа, 1973. 344 с.
- Левченко В. Д.** Модели в теории эволюции. СПб.: Наука, 1993. 384 с.
- Ленин В. И.** Избранные произведения. Еще раз о профсоюзах. Т. 3. М.: Политиздат, 1982. 498 с.
- Ленин В. И.** Материализм и эмпириокритицизм. М.: Политиздат, 1969. 392 с.
- Лиела И. Я.** К математическому моделированию в экологии // Моделирование и прогнозирование в биоэкологии. Рига, 1982. С. 3-41.
- Лиела И., Мауринь А.** Об основных этапах моделирования биологических объектов // Моделирование в ботанике. Рига, 1970. С. 4-14.
- Литвинова И. А., Коросов А. В.** Имитационное моделирование распространения сточных вод КЦБК в Кондопожской губе Онежского озера // Антропогенное воздействие на природу Севера и его экологические последствия. Апатиты, 1998. С. 116-118.
- Лосев А. Ф.** Диалектика художественной формы // Форма – Стиль – Выражение. М.: Мысль, 1995. С. 5-296.
- Лосев А. Ф.** Самое само // Миф – Число – Сущность. М.: Мысль, 1994. С. 299-526.
- Лосев А. Ф.** Философия имени // Бытие – имя – космос. М.: Мысль, 1993. С. 613-880.
- Ляпунов А. А.** О математическом подходе к изучению жизненных явлений // Математическое моделирование жизненных процессов. М.: Мысль, 1968. С. 65-107.
- Мастецкий Л. М.** Математические модели в экологии. Тверь, 1997. 40 с.
- Медведев Н. В.** Птицы и млекопитающие Карелии как индикаторы антропогенных загрязнений: Дис. ... канд. биол. наук. Петрозаводск, 1995. 151 с.
- Медведев Н. В., Коросов А. В.** Оценка накопления кадмия в популяции лосей Карелии с помощью имитационного моделирования // Современные проблемы биоиндикации и биомониторинга. Сыктывкар, 2001. С. 127-128.
- Меншуткин В. В.** Имитационное моделирование водных экологических систем. СПб.: Наука, 1993. 160 с.
- Мерсер Ф.** Аналитические методы исследования периодических процессов, замаскированных случайными флуктуациями // Биологические часы. М.: Мир, 1964. С. 126-152.
- Методы математической биологии.** В 8 т. Т. 1: Общие методы анализ биологических систем. Киев: Вища школа, 1980. 239 с.
- Мина М. В.** Аллометрический рост // Количественные аспекты роста организмов. М.: Наука, 1975. С. 176-180.
- Мина М. В., Клевезаль Г. А.** Рост животных. Анализ на уровне организма. М.: Наука, 1976. 291 с.
- Моисеев Н. Н.** Математические задачи системного анализа. М.: Наука, 1981. 487 с.
- Мороз А. И.** Курс теории систем. М.: Высшая школа, 1987. 304 с.

- Николис Г., Пригожин И.** Самоорганизация в неравновесных системах. М.: Мир, 1979. 512 с.
- Новосельцев В. Н.** Организм в мире техники. Кибернетический аспект. М.: Наука, 1989. 240 с.
- Одум Ю.** Основы экологии. М.: Мир, 1975. 742 с.
- Одум Ю.** Экология. Т. 2. М.: Мир, 1986. 367 с.
- Оптнер С. Л.** Системный анализ для решения деловых и промышленных проблем. М., Советское радио, 1969. 216 с.
- Оре О.** Теория графов. М.: Наука, 1980. 336 с.
- Паавер К.** Вопросы синтетического подхода в биоморфологии. Таллин: Валгус, 1976. 256 с.
- Павлов Б. К.** Методология прогнозирования численности наземных животных // Прогнозирование экологических процессов. Новосибирск: Наука, 1986. С. 185-190.
- Павлов Б. К.** Причины ошибок долговременных прогнозов состояния популяций животных при использовании аналого-инерционных экстраполяций // Геологические и экологические прогнозы. Новосибирск: Наука, 1984. С. 177-180.
- Перегудов Ф. И., Тарасенко Ф. П.** Введение в системный анализ. М.: Высшая школа, 1989. 367 с.
- Пианка Э.** Эволюционная экология. М.: Мир, 1981. 400 с.
- Пойа Д.** Как решать задачи. Львов: Квантор, 1991. 216 с.
- Поляков Ю. К.** Донные отложения // Кондопожская губа Онежского озера в связи с ее загрязнением протокками ЦБК. Петрозаводск, 1973. Ч. 2. С. 166-184.
- Попов Б. В.** Определение параметров многомерных моделей // Комплексообразование и метаболизм радиоактивных изотопов. Свердловск, 1976. С. 48-52.
- Поспелов Д. А.** Вероятностные автоматы. М.: Энергия, 1970. 87 с.
- Прицкер А.** Введение в имитационное моделирование и язык СЛАМ. М.: Мир, 1987. 644 с.
- Прозоровский В. Б.** Использование метода наименьших квадратов для пробит-анализа кривых летальности // Фармакология и токсикология. 1962. № 1. С. 115-119.
- Пучковский С. В.** Избыточность жизни. Ижевск, 1998. 376 с.
- Пэнгл Р.** Методы системного анализа окружающей среды. М.: Мир, 1979. 214 с.
- Разумовский О. С.** Бихевиоральные системы. Новосибирск: Наука, 1993. 240 с.
- Ракитов А. И.** Философские проблемы науки. М.: Мысль, 1977. 270 с.
- Рассел Б.** Человеческое познание. М., 1957. 555 с.
- Розенберг Г. С.** Модели в фитоценологии. М.: Наука, 1984. 265 с.
- Рыжков В. Л.** О простой биологической модели // Математическое моделирование жизненных процессов. М.: Мысль, 1968. С. 198-210.
- Рыжиков Ю. И.** Решение научно-технических задач на персональном компьютере. СПб.: Питерпрес, 2000. 272 с.
- Саати Т., Кернс К.** Аналитическое планирование. Организация систем. М.: Радио и связь, 1991. 224 с.
- Саламатов Ю. П.** Система развития законов творчества // Шанс на приключение. Петрозаводск, 1991. с.5-174.
- Сарабьянов В. Н.** Качество с точки зрения практического подхода // На переломе. Философские дискуссии 20-х годов: Философия и мировоззрение. М.: Политиздат, 1990. С. 484-491.
- Сарабьянов В. Н.** Назревший вопрос // На переломе. Философские дискуссии 20-х годов: Философия и мировоззрение. М.: Политиздат, 1990. С. 259-276.
- Свинцов В. И.** Логика. М.: Высшая школа, 1987. 287 с.
- Северцов А. С.** Функциональная дифференциация организма в ходе филогенеза // Уровни организации биологических систем. М.: Наука, 1980. С. 41-48.
- Северцов А. С., Ляпков С. М., Сурова Г. С.** Соотношение экологических ниш травяной (*Rana temporaria*) и остромордой (*Rana arvalis*) лягушек (*Anura*, *Amphibia*) // Журн. общ. биол. 1998. 59. № 3. С. 279-301.
- Славин М. Б.** Методы системного анализа в медицинских исследованиях. М.: Медицина, 1989. 304 с.
- Смит Дж.** Математические идеи в биологии. М.: Мири, 1970. 179 с.
- Страшкраба М., Гнаука А.** Пресноводные экосистемы. Математическое моделирование. М.: Мир, 1989. 376 с.
- Суходольский Г. В.** Основы математической статистики для психологов. Л., 1972. 430 с.
- Турчак Л. И.** Основы численных методов. Л.: Наука, 1987. 320 с.
- Тьюки Дж.** Анализ результатов наблюдений. М.: Мир, 1981. 694 с.
- Угольницкий Г. А.** Управление эколого-экономическими системами. М.: Вузовская книга, 1999. 132 с.
- Федоров В. Д.** К стратегии экологического прогнозирования // Человек и биосфера. М., 1983. С. 4-30.
- Федоров В. Д., Гильманов Т. Г.** Экология. М.: Изд-во МГУ, 1980. 464 с.
- Флейшман Б. С.** Основы системологии. М.: Радио и связь, 1982. 368 с.
- Форрестер Дж.** Основы кибернетики предприятия (индустриальная динамика). М.: Мир, 1971. 340 с.
- Хомяков М. П.** Общая характеристика системного анализа и системного подхода на основе опыта исследований в системной экологии // Проблемы окружающей среды и природных ресурсов. М., 1996. №19. С. 1-10.
- Шеннон Р.** Имитационное моделирование систем – искусство и наука. М.: Мир, 1970. 418 с.
- Шмальгаузен И. И.** Организм как целое в индивидуальном и историческом развитии // Избранные труды. М.: Наука, 1982. С. 12-228.
- Шмальгаузен И. И.** Кибернетические вопросы биологии. Новосибирск: Наука, 1968. 224 с.
- Шуп Т. Е.** Прикладные численные методы в физике и технике. М.: Высшая школа, 1990. 254 с.
- Энгельс Ф.** Анти-Дюринг. М.: Политиздат, 1952. 376 с.
- Энгельс Ф.** Диалектика природы. М.: Политиздат, 1987. 349 с.
- Югай Г. А.** Общая теория жизни (диалектика формирования). М.: Мысль, 1985. 256 с.
- Яблоков А. В.** Изменчивость млекопитающих. М.: Наука, 1966. 364 с.
- Яковлев Е. И.** Машинная имитация. М.: Наука, 1975. 158 с.
- Kutenov A. P., Mosiyash S. E.** On the dynamics of population of common frog (*Rana temporaria*) on the North-West of Russia // Russian Journal of Herpetology. 2000. Vol. 7. № 2. P.123-134.