CHAPTER 1

Mathematical Modeling in Biology

Chapter Goals:

- To develop an appreciation for the importance of mathematics in biology

Chapter Concepts:

- Variables
- Dynamics
- Parameters
- Principle of parsimony

1.1 Introduction

Mathematics permeates biology. Unfortunately, this is far from obvious to most students of biology. While many biology courses cover results and insights from mathematical models, they rarely describe how these results were obtained. Typically, it is only when biologists start reading research articles that they come to appreciate just how common mathematical modeling is in biology. For many students, this realization comes long after they have chosen the majority of their courses, making it difficult to build the mathematical background needed to appreciate and feel comfortable with the mathematics that they encounter. This book is a guide to help any student develop this appreciation and comfort. To motivate learning more mathematics, we devote this first chapter to emphasizing just how common mathematical models are in biology and to highlighting some of the important ways in which mathematics has shaped our understanding of biology.

Let's begin with some numbers. According to BIOSIS, 886,101 articles published in biological journals contain the keyword "math" (including math, mathematical, mathematics, etc.) as of April 2006. Some of these articles are in specialized journals in mathematical biology, such as the *Bulletin of Mathematical Biology*, the *Journal of Mathematical Biology*, *Mathematical Biosciences*, and *Theoretical Population Biology*. Many others, however, are published in the most prestigious journals in science, including *Nature* and *Science*. Such a coarse survey, however, misses a large fraction of articles describing theoretical models without using "math" as a keyword.

We performed a more in-depth survey of all of the articles published in one year within some popular ecology and evolution journals (Table 1.1). Given that virtually every statistical analysis is based on an underlying mathematical model, nearly all articles relied on mathematics to some extent. With a stricter definition that excludes papers whose only use of mathematics is through statistical analyses, 35% of *Evolution* and *Ecology* articles and nearly 60% of *American Naturalist* articles reported predictions or results obtained using mathematical models. The extent of mathematical analysis varied greatly, but mathematical equations appeared in almost all of these articles. Furthermore, many of the articles used computer simulations to describe changes that occur over time in the populations under study. Such simulations can be incredibly helpful, allowing the reader to "see" what the equations predict and allowing authors to obtain results from even the most complicated models.

TABLE 1.1
Use of mathematical models in full-length journal articles

Journal (in 2001)	Number of articles	General use of models[a]	Specific use of models[b]	Equations presented[c]
American Naturalist	105	96%	59%	58%
Ecology	274	100%	35%	38%
Evolution	231	100%	35%	33%

[a]General use: Used a mathematical model in the broadest sense, including statistical or phylogenetic analyses with a mathematical basis (e.g., ANOVA, regression, etc.).
[b]Specific use: Used a mathematical model to obtain results (excluding cases that involve only statistical or phylogenetic analyses); the model may or may not be derived in the paper.
[c]Equations presented: Excluding standard statistical equations.

An important motivation for learning mathematical biology is that mathematical equations typically "say" more than the surrounding text. Given the space constraints of many journals, authors often leave out intermediate steps or fail to state every assumption that they have made. Being able to read and interpret mathematical equations is therefore extremely important, both to verify the conclusions of an author and to evaluate the limitations of unstated assumptions.

To describe all of the biological insights that have come from mathematical models would be an impossible task. Therefore, we focus the rest of this chapter on the insights obtained from mathematical models in one tiny, but critically important, area of biology: the ecology and epidemiology of the human immunodeficiency virus (HIV). As we shall see, mathematical models have allowed biologists to understand otherwise hidden aspects of HIV, they have produced testable predictions about how HIV replicates and spreads, and they have generated forecasts that improve the efficacy of prevention and health care programs.

1.2 HIV

On June 5, 1981, the Morbidity and Mortality Weekly Report of the Centers for Disease Control reported the deaths of five males in Los Angeles, all of whom had died from pneumocystis, a form of pneumonia that rarely causes death in individuals with healthy immune systems. Since this first report, acquired immunodeficiency syndrome (AIDS), as the disease has come to be known, has reached epidemic proportions, having caused more than 20 million deaths worldwide (Joint United Nations Programme on HIV/AIDS 2004b). AIDS results from the deterioration of the immune system, which then fails to ward off various cancers (e.g., Karposi's sarcoma) and infectious agents (e.g., the protozoa that cause pneumocystis, the viruses that cause retinitis, and the bacteria that cause tuberculosis). The collapse of the immune system is caused by infection with the human immunodeficiency virus (Figure 1.1). HIV is transmitted

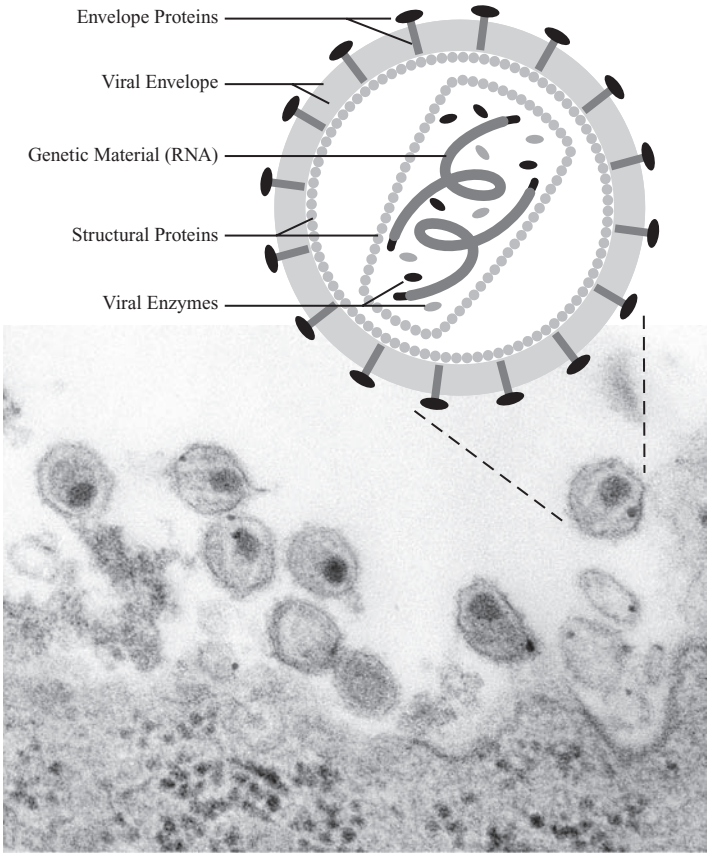

Figure 1.1: The human immunodeficiency virus. Electron micrograph shows HIV co-cultivated with human lymphocytes (courtesy of CDC; A. Harrison, P. Feorino, and E. L. Palmer).

from infected to susceptible individuals by the exchange of bodily fluids, primarily through sexual intercourse without condoms, sharing of unsterilized needles, or transfusion with infected blood supplies (although routine testing for HIV in donated blood has reduced the risk of infection through blood transfusion from 1 in 2500 to 1 in 250,000 [Revelle 1995]).

Once inside the body, HIV particles infect white blood cells by attaching to the CD4 protein embedded in the cell membranes of helper T cells, macrophages, and dendritic cells. The genome of the virus, which is made up of RNA, then enters these cells and is reverse transcribed into DNA, which is subsequently incorporated into the genome of the host. (The fact that normal transcription from DNA to RNA is reversed is why HIV is called a retrovirus.) The virus may then remain latent within the genome of the host cell or become activated, in which case it is transcribed to produce both the proteins necessary to replicate and daughter RNA particles (Figure 1.2). When actively replicating, HIV can produce hundreds of daughter viruses per day per host cell (Dimitrov et al. 1993), often killing the host cell in the process. These virus particles (or virions) then go on to infect other CD4-bearing cells, repeating the process. Eventually, without treatment, the population of CD4+ helper T cells declines dramatically from about 1000 cells per cubic millimeter of blood to about 200 cells, signaling the onset of AIDS (Figure 1.3).

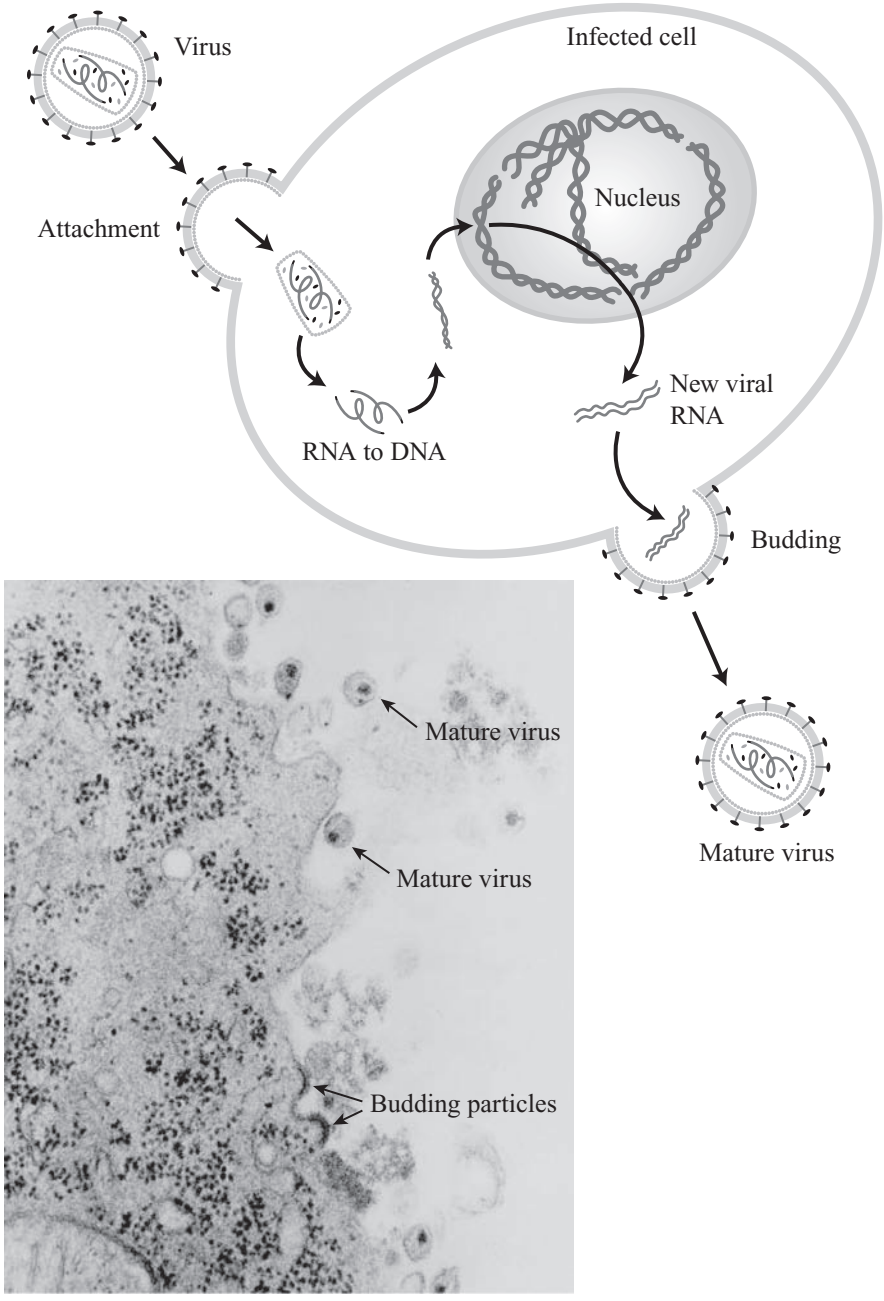

Figure 1.2: The life cycle of HIV within a host cell. Electron micrograph shows budding and mature HIV (courtesy of CDC; A. Harrison, E. L. Palmer, and P. Feorino).

Normally, CD4+ helper T cells function in the cellular immune response by binding to fragments of viruses and other foreign proteins presented on the surface of other immune cells. This binding activates the helper T cells to release chemicals (cytokines), which stimulate both killer T cells to attack the infected cells and B cells to manufacture antibodies against the foreign particles. What makes HIV particularly harmful to the immune system is that the

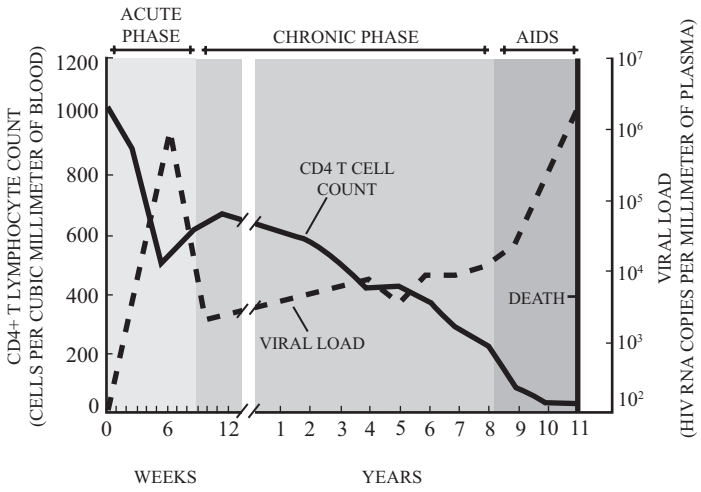

Figure 1.3: The time course of HIV infection within an individual. Viral loads and CD4+ T cell counts are plotted over time since infection. Based on data in Fauci et al. (1996).

virus preferentially attacks activated helper T cells; by destroying such cells, HIV can eliminate the very cells that recognize and fight other infections.

Early on in the epidemic, the median period between infection with HIV-1 (the strain most common in North America) and the onset of AIDS was about ten years (Bacchetti and Moss 1989). The median survival time following the onset of an AIDS-associated condition (e.g., Karposi's sarcoma or pneumocystis) was just under one year (Bacchetti et al. 1988). Survival statistics have improved dramatically with the development of effective antiretroviral therapies, such as protease inhibitors, which first became available in 1995, and with the advent of combination drug therapy, which uses multiple drugs to target different steps in the replication cycle of HIV. In San Francisco, the median survival after diagnosis with an AIDS-related opportunistic infection rose from 17 months between 1990 and 1994 to 59 months between 1995 and 1998 (San Francisco Department of Public Health 2000). Unfortunately, modern drug therapies are extremely expensive (typically over US$10,000 per patient per year) and cannot be afforded by the majority of individuals infected with HIV worldwide. Until effective therapy or vaccines become freely available, HIV will continue to take a devastating toll (Figure 1.4; Joint United Nations Programme on HIV/AIDS 2004a).

1.3 Models of HIV/AIDS

Mathematical modeling has been a very important tool in HIV/AIDS research. Every aspect of the natural history, treatment, and prevention of HIV has been the subject of mathematical models, from the thermodynamic characteristics of HIV (e.g., Hansson and Aqvist 1995; Kroeger Smith et al. 1995; Markgren et al. 2001) to its replication rate both within and among individuals (e.g., Funk et al. 2001; Jacquez et al. 1994; Koopman et al. 1997; Levin et al. 1996; Lloyd 2001; Phillips 1996). In the following sections, we describe four of these models in more detail. These models were chosen because of their implications

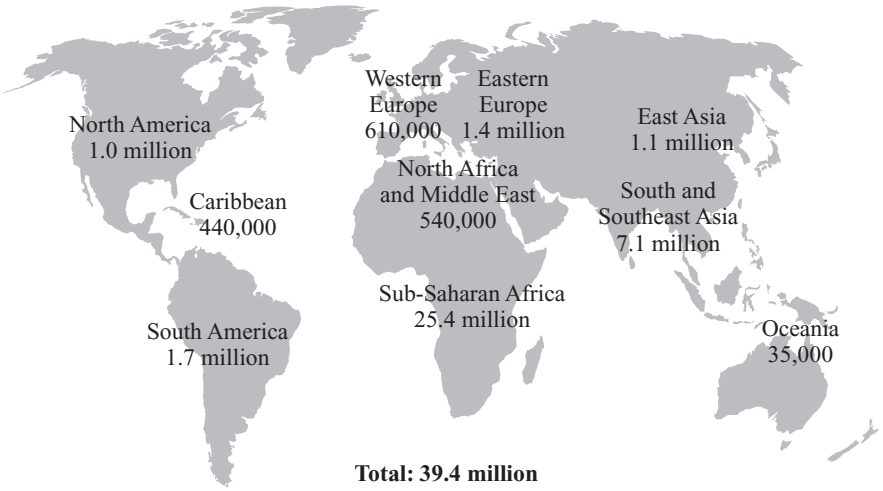

Figure 1.4: Number of individuals living with HIV. The number of adults and children estimated to be living with HIV is shown (Joint United Nations Programme on HIV/AIDS, 2004a).

for our understanding of HIV, but they also illustrate the sorts of techniques that are described in the rest of this book.

1.3.1 Dynamics of HIV after Initial Infection

After an individual is infected by HIV, the number of virions within the bloodstream skyrockets and then plummets again (Figure 1.3). This period of primary HIV infection is known as the acute phase; it lasts approximately 100 days and often leads to the onset of flu-like symptoms (Perrin and Yerly 1997; Schacker et al. 1996). The rapid rise in virus particles reflects the infection of CD4+ cells and the replication of HIV within actively infected host cells. But what causes the decline in virus particles? The most obvious answer is that the immune system acts to recognize and suppress the viral infection (Koup et al. 1994). Phillips (1996), however, suggested an alternative explanation: the number of virions might decline because most of the susceptible CD4+ cells have already been infected and thus there are fewer host cells to infect. Phillips developed a model to assess whether this alternative explanation could mimic the observed rise and fall of virions in the blood stream over the right time frame. In his model, there are four *variables* (i.e., four quantities that change over time): R, L, E, and V. R represents the number of activated but uninfected CD4+ cells, L represents the number of latently infected cells, E represents the number of actively infected cells, and V represents the number of virions in the blood stream. The *dynamics* of each variable (i.e., how the variable changes over time) depend on the values of the remaining variables. For example, the number of viruses changes over time in a manner that depends on the number of cells infected with actively replicating HIV. In the next chapter, we describe the steps involved in building models such as this one (see Chapter 2, Box 2.4).

A *variable* of a model is a quantity that changes over time.

The *dynamics* of a system is the pattern of changes that occur over time.

Phillips' model contains several *parameters*, which are quantities that are constant over time (see Chapter 2, Box 2.4). In particular, the death rate of

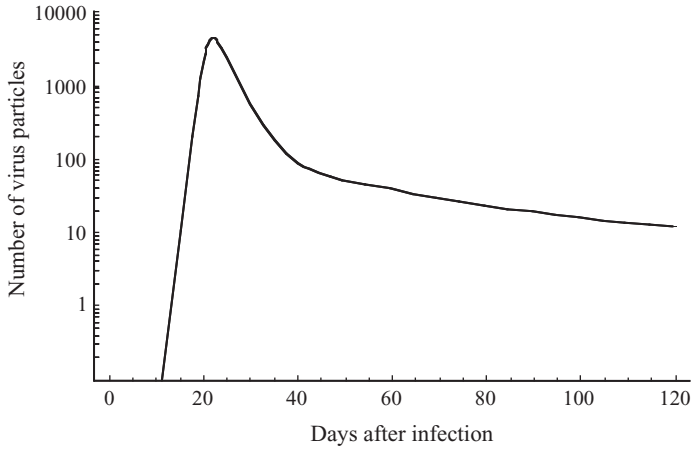

Figure 1.5: Number of virus particles in the blood stream. Based on the model and parameter values of Phillips (1996), the number of virions per mm^3 blood (V) is shown as a function of the number of days since primary infection (y-axis is plotted on a log-scale). Around a month after infection, the number of virus particles declines by about 100-fold even without any specific response by the immune system. See Box 2.4 in Chapter 2 for more details.

actively infected cells (δ) and the death rate of viruses (σ) are parameters in the model and are not allowed to change over time. (δ and σ are the lower-case Greek letters "delta" and "sigma." Greek letters are often used in models, especially for terms that remain constant ("parameters"). See Table 2.1 for a complete list of Greek letters.) Thus, Phillips built into his model the crucial assumption that the body does not get better at eliminating infected cells or virus particles over time, under the null hypothesis that the immune system does not mount a defense against HIV during the acute phase. To model the progression of HIV within the body, Phillips then needed values for each of the parameters in the model. Unfortunately, few data existed at the time for many of them. To proceed, Phillips chose plausible values for each parameter and numerically ran the model (a technique that we will describe in Chapter 4). The numerical solution for the number of virus particles, V, predicted from Phillips' model is plotted in Figure 1.5 (compare to Figure 1.3). Phillips then showed that similar patterns are observed under a variety of different parameter values. In particular, he observed that the number of virus particles typically rose and then fell by several orders of magnitude over a period of a few days to weeks. (An order of magnitude refers to a factor of ten. The number 100 is two orders of magnitude larger than one.)

A *parameter* of a model is a quantity that remains constant over time.

Phillips thus came to the counterintuitive conclusion that "the reduction in virus concentration during acute infection may not reflect the ability of the HIV-specific immune response to control the virus replication" (p. 497, Phillips 1996). The wording of this conclusion is critical and insightful. Phillips did not use his model to prove that the immune system plays no role in viral dynamics during primary infection. In fact, his model cannot say one way or the other whether there is a relevant HIV-specific immune response during this time period. What Phillips *can* say is that an immune response is not necessary to explain the observed data. This result illustrates an important principle in modeling: the *principle of parsimony*. The principle of parsimony states that one should prefer models containing as few variables and parameters as possible to describe the essential attributes of a system. Paraphrasing Albert Einstein, a model should be as simple as possible, but no simpler. In Phillips' case, he

According to the *principle of parsimony*, a simple explanation (or model) should be preferred over a complex explanation if both are equally compatible with the data.

could have added more variables describing an immune response during acute infection, but his results showed that adding such complexity was unnecessary. A simpler hypothesis can explain the rise and fall of HIV in the bloodstream: as infection proceeds, a decline in susceptible host cells reduces the rate at which virus is produced. Without having a good reason to invoke a more complex model, the principle of parsimony encourages us to stick with simple hypotheses.

Phillips' model accomplished a number of important things. First, it changed our view of what was possible. Without such a model, it would seem unlikely that a dramatic viral peak and decline could be caused by the dynamics of a CD4+ cell population without an immune response. Second, it produced testable predictions. One prediction noted by Phillips is that the viral peak and decline should be observed even in individuals that do not mount an immune response (i.e., do not produce anti-HIV antibodies) over this time period. Indeed, this prediction has been confirmed in several patients (Koup et al. 1994; Phillips 1996). Employing a more quantitative test, Stafford et al. (2000) fitted a version of Phillips' model to data on the viral load in ten patients from several time points during primary HIV infection; they found a good fit to the data within the first 100 days following infection. Third, Phillips' model generated a useful null hypothesis: viral dynamics do not reflect an immune response. This null hypothesis might be wrong, but at least it can be tested.

Phillips acknowledged that this null hypothesis can be rejected as a description of the longer-term dynamics of HIV. His model predicts that the viral load should reach an equilibrium (as described in Chapter 8), but observations indicate that the viral load slowly increases over the long term as the immune system weakens (the chronic phase in Figure 1.3). Furthermore, Schmitz et al. (1999) directly tested Phillips' hypothesis by examining the role of the immune system in rhesus monkeys infected with the simian immunodeficiency virus (SIV), the equivalent of HIV in monkeys. By injecting a particular antibody, Schmitz et al. were able to eliminate most CD8+ lymphocytes, which are the killer T cells thought to prevent the replication of HIV and SIV. Compared to control monkeys, the experimentally treated monkeys showed a much more shallow decline in virus load following the peak. This proves that, at least in monkeys, an immune response does play some role in the viral dynamics observed during primary infection. Nevertheless, the peak viral load was observed at similar levels in antibody-treated and untreated monkeys. Thus, an immune response was not responsible for stalling viral growth during the acute phase, which is best explained, instead, by a decline in the number of uninfected CD4+ cells (the targets of HIV and SIV).

1.3.2 Replication Rate of HIV

After the initial acute phase of infection, HIV circulates within the body at low levels until the onset of AIDS (Figure 1.3). These low levels suggest that virus particles might be produced at a low rate per day. This suggestion was, however, shown to be false using mathematical models in conjunction with

experimental data (Ho et al. 1995; Nowak et al. 1995; Wei et al. 1995). According to the mathematical models, low numbers of virus particles can result from a low rate of viral production, P, or from a high rate of clearing virus from the body, c (Ho et al. 1995). Determining which of these possibilities is correct is not possible using only the observed number of virus particles in untreated patients. These landmark papers pointed out, however, that you can tease apart these possibilities using mathematical models that predict viral dynamics following the application of antiretroviral drugs (Ho et al. 1995; Nowak et al. 1995; Wei et al. 1995). For example, Ho et al. (1995) treated HIV-infected patients with ABT-538, an antiviral drug that effectively prevents HIV replication (at least in the short term). Thus, the experimental treatment reduced viral production P to zero, causing the viral load to plummet within the bloodstream. The rate at which the viruses decreased in frequency was consistent with a simple mathematical equation that we will encounter in Chapter 6 (equation (6.10b)). Fitting the mathematical model to the data allowed the authors to obtain an important and surprising result: virus particles were rapidly cleared from the body, with the half-life of HIV in plasma being only a couple of days. The authors thus inferred that the production rate of viruses must normally be enormous, on the order of a billion new viruses produced per day, in order to maintain HIV in the face of high clearance rates. Later work, using more precise experimental data and more detailed modeling, demonstrated that the turnover of HIV is even more rapid, with the half-life of HIV being less than a day and with over 10 billion viruses produced per day. This is a remarkable insight, as it was once thought that relatively little was happening during the chronic phase of HIV infection (Perelson et al. 1996).

These papers had an enormous impact on our understanding of HIV. One of the most important conclusions to follow from this work was that we must expect genetic diversity to be rapidly generated in HIV as a result of the high rate of viral production. If resistance to an antiviral drug requires a particular mutation, it is virtually guaranteed that this mutation will arise rapidly. Only combination drug therapies, requiring multiple mutations for resistance, have a long-term chance of success given the enormous evolutionary potential of HIV.

Skip...

1.3.3 The Effects of Antiretroviral Therapy on the Spread of HIV

The specter of AIDS has softened following the development of effective antiretroviral therapies (ART), involving various drug combinations that have allowed people to live longer with HIV. Public health officials are concerned, however, that this respite will be short lived for two reasons: (a) people may be more inclined to engage in risky behavior knowing that ART exists and (b) HIV might evolve resistance to these drugs, causing the drugs to become ineffective. With these possibilities in mind, Blower et al. (2000) constructed a mathematical model to predict how drug therapy might affect the number of new cases of HIV and the number of deaths due to AIDS. Their model was tailored to data from the San Francisco gay community, where approximately 30% of men were infected with HIV (HIV+) and approximately 50% of these were taking combination ART.

study, the estimates provided by Williams et al. (2001) help to assess the risk of HIV infection as a function of age and can be used to better target programs designed to halt the spread of HIV in South Africa.

1.4 Concluding Message

Through reading this book, we hope that you will come to appreciate just how useful mathematical modeling can be in biology. Models can help to guide a scientist's intuition about how various processes interact; they can point out logical flaws in an argument; they can identify testable hypotheses, generate key predictions, and suggest appropriate experiments; and they can reshape fields by providing new ways of thinking about a problem.

But mathematical models also have their limitations. The results of an analysis are only as interesting as the biological questions motivating a model. And even if a scientist has identified an interesting question, it may turn out that a model addressing the question is hopelessly complicated and can be solved only by making a series of assumptions, some of which are dubious. Finally, models, by themselves, can only tell us what is possible. Models can tell us, for example, how HIV levels within the body or HIV incidence within a population might change over time. But without data, collected in the field or the lab, mathematical models can never tell us what has happened or what is happening. Thus, it would be foolish to promote mathematical biology above other areas in biology. Equally, it would be foolish to avoid mathematics altogether. Science will progress faster and further by a marriage of mathematical and empirical biology. This marriage will be even more successful if more biologists can use math, when needed, to advance their own research goals. It is toward this end that we devote this book.

Further Reading

For general information about the immune system and the evolution and ecology of infectious diseases, see

- Frank, S. A. 2002. *Immunology and Evolution of Infectious Diseases*. Princeton University Press, Princeton, N.J.

Further information concerning the life cycle, health impact, and societal implications of HIV is available through the links on the book website (http://press.princeton.edu/titles/8458.html) and at www.zoology.ubc.ca/biomath. Also, see

- Moore, R. D. and Bartlett, J. G. (1998) Improving HIV therapy. *Scientific American* 279: 84–93.

References

Bacchetti, P., and A. R. Moss. 1989. Incubation period of AIDS in San Francisco. *Nature* 338:251–253.

CHAPTER 2

How to Construct a Model

Chapter Goals:

- To describe the steps involved in developing a model
- To derive equations that describe the dynamics of a biological phenomenon

Chapter Concepts:

- Discrete-time model
- Continuous-time model
- Recursion equations
- Differential equations
- Life-cycle diagrams
- Flow diagrams
- Mass action

2.1 Introduction

If you have seen mathematical models but never constructed one, it may seem like an overwhelming task. Where do you start? What is the goal? How do you know whether the model makes sense? This chapter outlines the typical process of modeling and gives helpful hints and suggestions to break down the overwhelming task into manageable bits. The most important piece of advice is to *start*. Start thinking about problems that puzzle you. Grab a piece of paper and start drawing a flow diagram illustrating the various processes at work. The biggest hurdle preventing most biologists from modeling is the paralysis one feels in the face of mathematics; most of the technical problems that pop up along the way can be surmounted or sidestepped (at the very least by simulation). You will certainly make mistakes (we all do), but there are telltale signs of mistakes, and they can be corrected. Over time, you will learn more tools and techniques that will allow you to avoid pitfalls and to get further with the problems that interest you. Your intuition will develop to help you "see" when something is wrong with your model and to help you interpret your results.

Models can describe any biological phenomenon. In the core of this book, we focus on *dynamical* models, which describe how a system changes over time. Dynamical models are very common in biology as they provide insight into how various forces act to change a cell, an organism, a population, or an assemblage of species. Within dynamical models, two broad classes are distinguished: deterministic and stochastic. "Deterministic" is shorthand for the assumption that the future is entirely predicted (determined) by the model. "Stochastic" is shorthand for the assumption that random (stochastic) events affect the biological system, in which case a model can only predict the probability of various outcomes in the future. In the remainder of this chapter, as well as in Chapters 3–12, we focus on deterministic models. The steps for constructing stochastic models are similar, but we postpone further consideration of stochastic models until Chapters 13–15.

Box 2.1 describes, in seven steps, how to construct a dynamical model. This is like describing how to ride a bike in a series of steps; obviously we can only give an idea about how the process works. Mastering the steps requires practice, and the remainder of this chapter contains a series of seven sections, each corresponding to one of the seven steps in Box 2.1.

Box 2.1: Seven Steps to Modeling a Biological Problem

Step 1: Formulate the question
What do you want to know?
Describe the model in the form of a question.
Boil the question down!
Start with the simplest, biologically reasonable description of the problem.

Step 2: Determine the basic ingredients
Define the variables in the model.
Describe any constraints on the variables.
Describe any interactions between variables.
Decide whether you will treat time as discrete or continuous.
Choose a time scale (i.e., decide what a time step equals in discrete time and specify whether rates will be measured per second, minute, day, year, generation, etc.).
Define the parameters in the model.
Describe any constraints on the parameters.

Step 3: Qualitatively describe the biological system
Draw a life-cycle diagram (see Figure 2.2) for discrete-time models involving multiple events per time unit.
Draw a flow diagram to describe changes to the variables over time.
For models with many possible events, construct a table listing the outcome of every event.

Step 4: Quantitatively describe the biological system
Using the diagrams and tables as a guide, write down the equations.
Perform checks. Are the constraints on the variables still met as time passes? Make sure that the units of the right-hand side equal those on the left-hand side.
Think about whether results from the model can address the question.

Step 5: Analyze the equations
Start by using the equations to simulate and graph the changes to the system over time.
Choose and perform appropriate analyses.
Make sure that the analyses can address the problem.

Step 6: Checks and balances
Check the results against data or any known special cases.
Determine how general the results are.
Consider alternatives to the simplest model.
Extend or simplify the model, as appropriate, and repeat steps 2–5.

Step 7: Relate the results back to the question
Do the results answer the biological question?
Are the results counterintuitive? Why?
Interpret the results verbally, and describe conceptually any new insights into the biological process.
Describe potential experiments.

2.2 Formulate the Question

The first step, coming up with a question, can be more difficult than it sounds. In most biology classes, students are told what the questions are and what answers have been found. Rarely are students asked to formulate scientific questions for themselves. This is very unfortunate because, in any scientific enterprise (modeling or otherwise), the process begins with a question. One hint is to keep an eye out for things that do not make sense or that seem to conflict—there very well might be an interesting and nonintuitive resolution. For now, start simple and don't worry about how profound your question is. Look around you, find a living object, and think up one question about how it might change over time. We did this and came up with the following three questions, which we will use in this chapter to illustrate model construction. (i) How does the number of branches of a tree change over time? (ii) How does a cat change the number of mice in a yard? (iii) How does the number of people with the flu change over the flu season?

The above three questions are "toy" examples that will make it easier to show the steps of modeling. Nevertheless, these simple examples also embody many of the key elements that come together in various combinations when constructing more complicated and realistic models. As we will see, the tree branching model is a special case of a model describing population growth. The mouse model incorporates an important component of immigration that is commonly used in ecology. For example, Blower et al. (2000) used a similar model of immigration to describe individuals moving into the gay male community of San Francisco. Finally, the flu model highlights some important concepts related to interactions among variables. For example, the way that we will model flu transmission is fundamentally similar to the way that Phillips (1996) modeled the infection of cells by HIV. Thus, these toy models provide an excellent background for tackling more complex models.

2.3 Determine the Basic Ingredients

Once you have a question in mind, proceed to Step 2 in Box 2.1. First, think about what entities might change over time; these entities are the *variables* in your model. The number of variables will depend on the question of interest. In our toy examples, we might choose to follow (i) the number of branches on a tree, (ii) the number of mice in a yard, and (iii) the number of people with the flu and the number without the flu. In choosing variables to track, we must always simplify reality. For example, in keeping track of the number of branches, we lose information about their size and age. As a general principle, start simple, adding more variables only when the model fails to address the question.

Next, we assign a letter to represent each variable—it is easier to write "x" than "the number of branches on a tree." The letters n, p, x, and y are commonly used to represent variables, but the choice is arbitrary. A good idea is to choose letters that help you remember what the variable represents, e.g., "n" for number or "p" for proportion. If a model contains multiple variables

that are similar in nature, placing subscripts on the variables can help to emphasize their similarity, e.g., n_1 and n_2 for the numbers of two different species. For our models, we will use (i) $n(t)$ for the number of branches on a tree, (ii) $n(t)$ for the number of mice in a yard, and (iii) $n(t)$ for the number of people with the flu and $s(t)$ for the number of *s*usceptible people.

To remind ourselves that a variable, say n, varies over time, we can write it as $n(t)$ where t represents time and there is no space between the n and the (t). The parentheses tell us that our variable is a function of something else (time), and we read $n(t)$ as "n at time t." This notation helps to avoid math errors. For example, without this notation, we might forget that n takes on different values at different times and mistakenly treat it as a constant. Be aware, however, that not all authors use the same notation; they might write n_t instead or might simply state that n is a variable and not write it explicitly as a function of time. The important thing is to be consistent and to remember that, if we write a variable as $n(t)$, we mean "n at time t" not "n times t."

Another way to avoid math errors is to keep a list (at least a mental list) of any constraints that must remain true about the variables. For example, the number of branches on a tree should never become negative. The number of people with the flu and the number without the flu should never be negative and should sum up to the total population size. If a variable describes a frequency, a probability, or a fraction of a whole (e.g., the fraction of the total population with the flu), it should always lie between zero and one ($0 \leq p(t) \leq 1$). Ensuring that your equations and results obey the list of constraints is a good way to check that no errors have crept in.

Once you have a preliminary list of variables, the next step is to choose a type of dynamical model to describe changes in these variables. There are two main types of dynamical models, *discrete time* and *continuous time*, depending on whether time is represented in discrete steps or along a continuous axis. Discrete-time models describe how the variables change from one time unit (e.g., day, year, or generation) to the next. Continuous-time models track the variables over any period of time. Both discrete-time and continuous-time models are idealizations of reality, and they make somewhat different assumptions.

A *discrete-time model* tracks changes to variables in discrete time steps.

A *continuous-time model* allows variables to change at any point in time (i.e., time is treated as continuous).

Discrete-time models assume that changes cannot compound within a time unit. For example, in a discrete-time model for the number of branches on a tree, branches that arise during a time unit cannot give rise to new branches within the same time unit. As long as the time unit is short enough (e.g., a day), this assumption is often reasonable. If the time unit were long (e.g., a year), however, then some new branches might very well branch again within the year. These branching events would not be counted in a discrete-time model if the new branches were not present at the beginning of the year.

Continuous-time models assume that variables can change at any point in time, with increments or decrements occurring even within tiny intervals of time. As a consequence, it is possible for a change to occur in one small interval of time followed by the same type of change in the next small interval of time. But this may not be biologically realistic. For example, a continuous-time model might allow a newly formed branch to immediately produce its own new branch. In reality, the new branch must undergo enough cell divisions to

produce a new bud, which takes time. If the rate of branching is small, then this won't be much of a problem because the average time between branching events will be large. But if the rate of branching is high, then a continuous-time model will generate incorrect predictions unless it takes into account the time lag between the formation of a branch and the formation of buds on this new branch.

Because discrete- and continuous-time models treat the timing of events in different ways, they display different temporal dynamics. In discrete-time models the variables "jump" from one value to another from one time unit to the next, and the size of these jumps can be small or large depending upon the parameters of the model. In continuous-time models, on the other hand, the variables change smoothly over time. This means that, as a variable goes from one value to another, it passes through all intervening values along the way (Figure 2.1).

In either case, we must also choose a time scale over which changes to the variables are measured. We use a "day" as the basic unit of time for the toy models considered in this chapter. Specifically, we assume that each time step in discrete-time models reflects the passage of 24 hours and that all processes in continuous-time models occur at a rate measured per day.

Just as time can be modeled discretely or continuously, so too can the variables themselves. For example, the number of branches on a tree, the number of mice in a yard, and the number of people with the flu are all discrete, integer-valued quantities (i.e., they are integers such as 0,1,2, . . . , etc.). On the other hand, an organism's metabolic rate or an organism's weight can take on any of a continuum of possible values. Regardless of the true nature of the variables, the majority of models in ecology and evolution treat variables as being continuous, an approach that we follow throughout most of the book (except in Chapters 13–15, which incorporate random events and explicitly track the numbers of each type). There are three main justifications for treating variables as continuous. First, for many questions, the variables of interest take on large enough values that treating them as continuous will introduce very little error in the results (e.g., the number of HIV particles in the blood). Second, a reinterpretation of the variable (e.g., as the total biomass of mice rather than the number of mice) can sometimes justify the use of a continuous variable. Third,

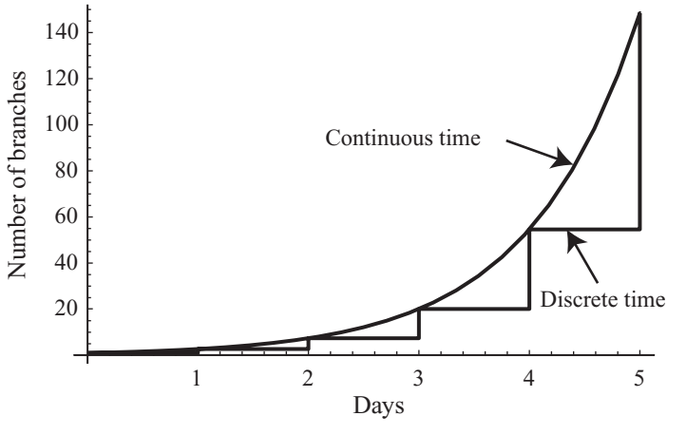

Figure 2.1: Tree branching. A plot of the number of branches on a tree over time using a discrete-time model and a continuous-time model.

it is typically easier mathematically to treat variables as being continuous rather than discrete. Remember, all models are abstractions of biological reality, and treating variables as continuous is often a reasonable abstraction.

A *recursion equation* describes the value of a variable in the next time step.

In discrete-time models, we track changes to a variable using a *recursion* equation, which describes the value of a variable (say, n) in the next time unit as a function of the variable in the current time unit:

$$n(t + 1) = \text{“some function of } n(t)\text{.”} \tag{2.1a}$$

Such equations are called recursions, because one can apply them recursively to find out how the variable changes across a number of time units (from t to $t + 1$, then from $t + 1$ to $t + 2$, etc.). An equivalent way to track a variable is to use a *difference* equation. A difference equation specifies how much a variable changes from one time unit to the next, and it is just the difference between the recursion equation for $n(t + 1)$ and the current value of the variable $n(t)$:

$$\Delta n = n(t + 1) - n(t) = \text{“some function of } n(t)\text{,”} \tag{2.1b}$$

where the capital Greek letter Δ (“Delta,” see Table 2.1) denotes “change,” and we read Δn as “the change in the variable n.” Recursion equations are more commonly used to describe the value of a variable in discrete-time models, but we will occasionally use difference equations when we want to understand how much a variable changes across a time step.

In continuous-time models, equations specify the rate of change of the variables over time:

$$\frac{\mathrm{d}(n(t))}{\mathrm{d}t} = \text{“some function of } n(t)\text{.”} \tag{2.1c}$$

A *differential equation* describes the rate at which a variable changes over time.

Such equations are called *differential* equations. Differential equations are distinct from the more familiar derivatives taught in introductory calculus courses (see Box 2.2). You can think of a differential equation as a description of the ebb and flow in a variable over time. To get a better feel for a differential equation, imagine plotting the value of the variable $n(t)$ as a function of time (see Chapter 4). The slope of the curve would be $\mathrm{d}(n(t))/\mathrm{d}t$ because the derivative of a function at a point gives the slope of the function at that point. If the variable is increasing over time, the slope and thus $\mathrm{d}(n(t))/\mathrm{d}t$ are positive. If the variable is decreasing over time, the slope and thus $\mathrm{d}(n(t))/\mathrm{d}t$ are negative. When the magnitude of $\mathrm{d}(n(t))/\mathrm{d}t$ is small, the variable changes slowly over time, whereas when the magnitude of $\mathrm{d}(n(t))/\mathrm{d}t$ is large, the variable changes rapidly. As we will see, this mental picture is the reverse of how we typically construct models. We usually start by describing how various biological forces change the value of the variable (i.e., contribute to $\mathrm{d}(n(t))/\mathrm{d}t$), and we then try to infer the value of the variable itself (i.e., $n(t)$).

Which type of model should be used? Sometimes, there is a natural choice. If you want to model the number of annual plants on an island, a discrete-time model using a year as the time unit is appropriate because the life cycle of annual

TABLE 2.1
Greek letters. Here, we list the Greek letters commonly encountered in biological models (with alternative characters in parentheses).

Lower case	Upper case	Name
α	A	alpha
β	B	beta
χ	X	chi
δ	Δ	delta
ε	E	epsilon
$\phi(\varphi)$	Φ	phi
γ	Γ	gamma
η	H	eta
ι	I	iota
κ	K	kappa
λ	Λ	lambda
μ	M	mu
ν	N	nu
o	O	omicron
$\pi(\varpi)$	Π	pi
$\theta(\vartheta)$	Θ	theta
ρ	P	rho
$\sigma(\varsigma)$	Σ	sigma
τ	T	tau
υ	Υ	upsilon
ω	Ω	omega
ξ	Ξ	xi
ψ	Ψ	psi
ζ	Z	zeta

plants is itself discrete; that is, the seeds produced during one year will not germinate until the following year. By contrast, if you want to model your blood sugar levels after a meal, a continuous-time model would be more natural because there are no clear demarcations in time. Conceptually, it is sometimes easier to think in terms of discrete-time models where changes describe what happens over an interval of time rather than continuous-time models where changes are described by instantaneous rates. Mathematically, however, continuous-time models can be easier to analyze because one can utilize the various rules of calculus summarized in Appendix 2 (see Chapter 6). As we discuss later (Box 2.6

Box 2.2: Derivatives and Differential Equations

Calculus is the mathematical study of rates of change. The most important concepts and rules of calculus are summarized in Appendix 2, including formulas for differentiating and integrating a variety of functions. For example, the derivative of the polynomial $y = ax^2 + bx + c$ with respect to x is $dy/dx = 2ax + b$. Here, the rate of change of the *dependent* variable y is a function only of the *independent* variable x. In many biological problems, however, the rate of change of the dependent variable is a function of the dependent variable itself, e.g., $dy/dx = \alpha y + \beta$. Notice that the variable on the right-hand side is y not x. An equation relating the derivative of a variable to a function of the variable itself is called a *differential equation*. Equations (2.8)–(2.10) are differential equations. For example, in equation (2.8), the derivative of the dependent variable describing the number of tree branches, $n(t)$, with respect to the independent variable (time t) is a function of $n(t)$, not t. Differential equations naturally arise in continuous-time biological models because we often expect the rate of change of a variable to be a function of its current value. For example, large trees can have more new branches, a cat can eat more mice if there are more mice available, and more people can catch the flu if there are more susceptible people within the population.

A derivative or differential equation describes how a variable changes. But what we usually want to know is the *value* of the dependent variable (e.g., $n(t)$) as a function of the independent variable (e.g., t). In a typical calculus course, we are taught how to solve for y by taking the antiderivative or integral of both sides. In other words, we could solve the equation $dy/dx = 2ax + b$ for $y(x)$ by integrating both sides with respect to x to obtain its *solution*, $y = ax^2 + bx + c$ (see Appendix 2), which gives us the value of y for any value of x. A common error that students make when they first encounter differential equations is to integrate the left-hand side of an equation like $dn(t)/dt = bn(t)$ with respect to t but the right-hand side with respect to $n(t)$. This would give $n(t) = bn(t)^2/2$. To see that this is incorrect, take the derivative of both sides with respect to t (see Appendix 2). This would give $dn(t)/dt = bn(t)dn(t)/dt$, which incorrectly has $dn(t)/dt$ on the right-hand side. The error in this procedure crept in when we took the antiderivative of the left-hand side with respect to t, but the antiderivative of the right-hand side with respect to a different variable, $n(t)$. To solve for $n(t)$ we would have to take the antiderivative of *both sides* with respect to t, i.e.,

$$\int \frac{dn(t)}{dt}\, dt = \int bn(t)\, dt. \tag{2.2.1}$$

The left-hand integral is $n(t)$, as before, but we cannot evaluate the right-hand integral because doing so requires $n(t)$, which is what we are trying to find. In Chapter 6, we will see how to obtain solutions to certain types of differential equations, like the ones presented in this chapter. For now, it is enough to recognize the distinction between derivatives and differential equations and to remember that care must be taken when integrating differential equations.

Before leaving the subject, it is worth mentioning that the term "differential equation" encompasses several types of equations, all of which arise in biology. Differential equations can be written as functions of more than one dependent variable. For example, in our flu model, the

(continued)

Box 2.2 ***(continued)***

differential equation (2.10a) for the number of people with the flu, $dn(t)/dt$, will depend on both the number of people with the flu, $n(t)$, and the number of susceptible individuals in the population, $s(t)$. Differential equations can also be written as functions of *both* the dependent variable $n(t)$ and the independent variable t. Such differential equations arise whenever we expect a variable to change as a function both of its current value and of time. For example, in a seasonal environment, the budding rate of a tree should depend on the time of year as well as on the number of branches on a tree. We can model this by treating b as some function of time, $b(t)$, rather than a constant. In addition, differential equations might depend on the past state of a variable as well as (or instead of) its current state. For example, in the tree branching example, the production of new branches at time t might depend on the total number of branches τ days ago, or $n(t - \tau)$, as these branches are now large enough to branch again. Revising equation (2.8) gives $dn(t)/dt = bn(t - \tau)$. Such equations, known as "delay differential equations," arise naturally when describing biological processes involving time lags.

All of the above examples have only one independent variable (time). These fall into the category known as "ordinary differential equations" (ODE). Many biological problems involve more than one independent variable (e.g., space as well as time), and such differential equations are known as "partial differential equations" (PDE).

and Chapter 4), discrete-time and continuous-time models can sometimes exhibit similar behavior over time, and it is possible to predict when they should behave similarly. Thus, in many cases, one is free to choose between the two.

The next step is to describe the *parameters* of the model; these are the various quantities that influence the dynamics of the model, but that remain fixed over time as the variables change. As with variables, each parameter is given its own symbol, which you are free to choose. Commonly used symbols for parameters are italicized roman letters (e.g., a, b, c, d, m, and r) and lower-case greek letters (e.g., α, β, Table 2.1).

A chief difference between discrete-time and continuous-time models is that parameters representing events per unit time are described as the *number of events* (or fraction of the population undergoing the event) per time step in discrete-time models but as the instantaneous *rate of events* per unit time in continuous time. In contrast, parameters that do not represent events per unit time (e.g., the probability that an event is one type or another) retain the same definition in the two types of models. We will discuss the difference in parameter units between discrete- and continuous-time models at greater length in Box 2.6, once we have described how their dynamical equations are derived.

Potential parameters for our discrete-time models include (i) the number of new branches that bud off each old branch per day, b; (ii) the fraction of mice in the yard eaten by the cat per day, d, and the number of mice born per mouse per day, b; (iii) the fraction of healthy people that are exposed to a flu carrier per day, c, and the probability of transmission of the flu between a healthy person and a flu carrier upon exposure, a. The analogous parameters in a continuous-time

model would be (i) the rate of budding for each old branch, b; (ii) the rate of consumption of mice, d, and the rate of births per mouse, b; and (iii) the rate of contact between a flu carrier and a susceptible person, c, and the probability of transmission of the flu between a carrier and a healthy person per contact, a. These parameters represent events per unit time and so have slightly different definitions for the discrete-time and continuous-time models except a, which always represents the probability of contracting the flu per contact.

As with variables, one should also keep track of any constraints imposed on each parameter. For example, can a parameter be negative? Does a parameter represent a fraction, proportion, or probability, in which case it must fall between zero and one? These constraints might well depend on the type of model. For example, the parameter d in the cat-mouse model is restricted to lie between zero and one in discrete-time models (because it represents the *fraction* of mice eaten by the cat), whereas the analogous parameter d in the continuous-time model can have any positive value (because it represents the *rate* of consumption of mice per unit time). This is another common difference in the parameters between discrete- and continuous-time models (described more fully in Box 2.6).

In addition to the absolute constraints on each parameter, it is worth keeping track of the range of parameter values that are biologically reasonable. For example, it is reasonable to assume that the number of new branches that bud off each old branch per day is small for most trees ($b << 1$). Similarly, the number of mice born per mouse per day (b) will be much less than one ($b << 1$). We write $b << 1$ to imply that b is much smaller than one. How much smaller depends on the context, but typically this statement implies that b is 0.1 or less. Having a list of constraints and reasonable ranges for parameters can help in two important ways. First, reasonable parameter values must be chosen to carry out realistic simulations and to plot relevant graphs. Second, results from a model often depend on the values of the parameters, e.g., whether a parameter is positive or negative, large or small, so that making accurate predictions from a model depends on choosing appropriate parameter values.

Before proceeding to the next step, it is a good idea to construct a table of all the variables and parameters in your model, as well as any constraints on these terms. You can later revisit this table to ensure that it includes the variables and parameters needed to capture the essence of the biological process and to address the question of interest. It is very common that the first version of a model includes too many variables and parameters, causing the model to be unnecessarily complex, or too few variables and parameters, causing a model to behave in unintended ways (e.g., populations grow to infinite size, or nobody ever recovers from the flu). If a model displays unintended behavior, then think about whether the biological system being modeled includes other processes that should also be incorporated into the model (e.g., competition, recovery).

2.4 Qualitatively Describe the Biological System

Before writing equations down, it is a very good idea to organize your model conceptually with the aid of a diagram or table. Diagrams and tables make it easier to see whether the necessary variables and parameters are included and