Preface

Mathematical biology is growing rapidly. Mathematics has long played a dominant role in our understanding of physics, chemistry, and other physical sciences. However, wholesale application of mathematical methods in the life sciences is relatively recent. Now questions about infectious diseases, heart attacks, cell signaling, cell movement, ecology, environmental changes, and genomics are being tackled and analyzed using mathematical and computational methods.

While the application of quantitative analysis in the life sciences has borne fruit in the research arena, only recently has it impacted undergraduate education. Until a few years ago, the number of undergraduate texts in mathematical biology could be counted on one hand. Now this has changed dramatically. Recent undergraduate texts range from simple introductions to biological numeracy (Burton [35, 36]), freshman calculus for students in the life sciences (Adler [1], Neuhauser [125]), modeling with differential equations (Taubes [155], Edelstein-Keshet [51], Britton [29]), computer algebra (Yeargers, Shonkwiler, and Herod [168]), and dynamical computer-based systems (Hannon and Ruth [78]), to name but a few.

Despite the plentitude of new books, mathematical biology is still rarely offered as an undergraduate course. This book is designed for undergraduate students. Our target audience are students in mathematics, biology, physics, or other quantitative sciences at the sophomore or junior level. Our aim is to introduce students to problem solving in the context of biology. The focus in our presentation is on integrating analytical and computational tools in the modeling of biological processes.

The book stems from pedagogic material developed by the authors for a 7–11 day workshop in mathematical biology, which has been taught since 1995 at the University of Tübingen (Germany) and since 2001 at the University of Alberta (Canada). Additional material has been added to make the book suitable for use in a full-term course in mathematical biology.

There are three parts to this book: (I) analytical modeling techniques, (II) computational modeling techniques, (III) problem solving.

Part I covers basic analytical modeling techniques. We discuss the formulation of models using difference equations, differential equations, probability theory, cellular automata, as well as model validation and parameter estimation. We emphasize the modeling process and qualitative analysis, rather than explicit solution techniques (which can be found in other textbooks). Classical models for disease, movement, and population dynamics are derived from first principles. Each section provides a number of biologically motivated exercises.

Part II introduces computational tools used in the modeling of biological problems. Students are guided through symbolic and numerical calculations with Maple (for readers who prefer an alternative software package, such as *Mathematica* or MATLAB, see "How to Use This Book" below). Many of the examples and exercises of this part relate directly to the models discussed in Part I. This part of our book has been designed such that students can work through the material independently and at their own pace. Readers without any programming background will pick up valuable computational skills. Readers who already have programming background will be able to skip some elementary exercises and focus attention on the biological applications.

Part III provides open-ended problems from epidemiology, ecology, and physiology. Each problem is formulated in a way that makes it accessible to students. In most cases, questions will guide the student through the modeling process. These problems can be used as the basis for extended investigation, for example, as a term project or as a team project. We conclude Part III with a detailed presentation of two projects (cell competition and the chemotactic paradox) based on solutions developed by teams of undergraduate students who participated in one of our workshops.

The field of mathematical biology is, admittedly, immense. This book does not attempt to achieve a comprehensive introduction to the field. Subjects are tempered by the test of being able to teach them effectively in a short period of time. Problems are biased towards the authors' interests, but are sufficiently wide-ranging to include something of interest for most students. Ultimately, we hope that this book offers the first step into a detailed modeling of problems in the life sciences.

How to Use This Book

We envision that this book can be used in a number of ways. Here we list some ideas about how a course could be designed based on the material of this book.

Full-Term Course: During a full-term course, material from Part I can be covered. Students should have access to computers to complete Part II and for the project work of Part III. Although students can work through the computer tutorial on their own, we recommend a two-hour computer lab during which an instructor is available to help the students get started. Projects from the open-ended problems from Part III may be assigned early in the course, with students submitting a written report, or presenting the project in class (or both) towards the end of the course.

10-Day Workshop: During the first half of the workshop, the focus should be on learning modeling with analytical and computational tools, based on Parts I and II of this book. Ideally a mixture of discrete-time equations, differential equations, and stochastic models should be covered. Specific topics would depend on the background of the instructor(s). We feel that Sections 2.2, 3.1–3.4, 4.3, and 5.1–5.6 should be included in any course. Lectures on these topics may be supplemented by homework. In our experience, students need about 15 hours to work through the computer tutorial of Part II.

During the second half of the workshop, students should work in teams of two (maybe three) on one of the open-ended problems from Part III. Under the guidance of an instructor, students develop a model, analyze and/or simulate the model, and prepare a presentation. We have found it important to stress that problems are open-ended, and have no "right

solution" *per se*. It is the process of model development that is most important, not necessarily the end product. In many cases, students will need to simplify their problem and build a hierarchy of models, each model incorporating additional realism from the original problem.

Substituting Maple with Other Software: Although we have based Part II of this book on Maple, we do not wish to give the impression that Maple is necessarily the ideal software to be used. In fact, we believe that it does not really matter which software package is used. Instructors or students proficient with other software, such as *Mathematica* or MATLAB, will readily be able to adapt the examples and exercises of Part II for the alternative software. A version of Part II in Mathematica is available at http://www.siam.org/books/mm12.

Working on Open-Ended Projects: Since the problems from Part III are open ended there is a danger of aiming too high. Some of the problems are currently being studied by experienced researchers, and it would be impossible to follow all the relevant literature. For a beginning modeler, we give the following guidelines.

From the project description, readers should be able to understand the biological problem at hand to a certain extent. Some reading of supplemental material might be useful. For most projects, a specific reference is given, and the Internet is always a good resource. It is not required to study the biological topic at length. Initial efforts in mathematical modeling require only the identification of basic mechanisms.

When the biological problem at hand is understood, students should determine first which of the model classes presented in Part I might be useful (discrete/continuous, deterministic/stochastic). With the help of an instructor, they then proceed to develop a mathematical model. After the students and the instructors agree upon a reasonable model, the students work on it, do the analysis, and write the software. Many projects are accompanied with data. In that case, data fitting will be an important element of the project. Last but not least, the model should be used to explain important aspects of the biological phenomenon and to make predictions for other experiments or observations.

Internet Resources: A webpage related to this book that contains solutions to most of the exercises and the computer tutorial of Part II in *Mathematica* can be found at http://www.siam.org/books/mm12.

Acknowledgments

Although five authors worked on this text, it would not have been written without support from various sources. We express our thanks to Andrew Beltaos, an undergraduate student who helped with editing, making figures, and solving exercises. We are grateful to the Universities of Tübingen and Alberta for providing the environment to run our workshops. We also are grateful to the Pacific Institute for the Mathematical Sciences (PIMS), the Mathematics of Information Technology and Complex System (MITACS), and the Department of Mathematical and Statistical Sciences at the University of Alberta for significant financial support for the Alberta workshops. Moreover, we express our thanks to our colleagues who have taught with us in our workshops; in alphabetical order: Gary de Young, Leah Edelstein-Keshet, K. P. Hadeler, Christina Kuttler, Michael Li, Frithjof Lutscher, Michael Mackey, Annette Räbiger, Rebecca Tyson, and Pauline van den Driessche. Thanks to Michael Baake, Fred Brauer, three anonymous reviewers, and many of our students for valuable comments

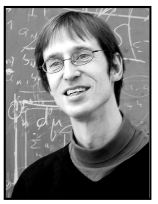
on earlier versions of this text. We would particularly like to thank all student participants of our workshops. Their enthusiasm and interest in mathematical modeling was our motivation to write this book. Patient help and guidance from the SIAM editorial staff was invaluable in the finishing stages of this work. We thank our families for their encouragement and support.

June 2006

Gerda de Vries is Associate Professor in the Department of Mathematical and Statistical Sciences at the University of Alberta, Canada.

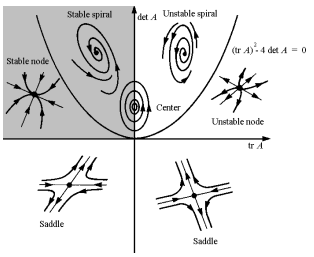
Thomas Hillen is Professor in the Department of Mathematical and Statistical Sciences at the University of Alberta, Canada.

Mark Lewis is Professor and Senior Canada Research Chair in Mathematical Biology in the Department of Mathematical and Statistical Sciences and Department of Biological Sciences at the University of Alberta, Canada.

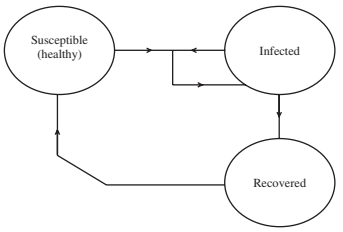


Johannes Müller is Professor of Mathematical Methods in Molecular Biology and Biochemistry in the Centre of Mathematical Sciences at the Technical University Munich, Germany.

Birgitt Schönfisch is a Scientific Employee in the Department of Medical Biometry at the University of Tübingen, Germany.



Part I
Theoretical Modeling Tools



Chapter 1
Introduction

1.1 The Modeling Process

Mathematical biology dwells at the interface of two fields: applied/computational mathematics and biology. Individually, these fields are growing quickly due to rapidly changing technology and newly emerging subdisciplines. Coupled together, the fields provide the basis for the emerging scientific discipline of mathematical biology, whose focus is interdisciplinary scientific problems in quantitative life sciences.

What can biology offer mathematics and computation? Biological models offer a seemingly endless supply of challenging and interesting nonlinear problems to solve. These nonlinear problems can provide a testing ground for applied mathematical and computational methods, and generate the impetus to develop new mathematical and computational methods and approaches.

What can mathematics and computation offer biology? Mathematics and computation can help solve a growing problem in biological research. Data collection, varying from gene sequencing to remote sensing via satellites, is now inundating biologists with complex patterns of observations. The ability to collect new data outstrips our ability to heuristically reason mechanisms of cause and effect in complex systems. It is the analysis of mathematical models that allows us to formalize the cause and effect process and tie it to the biological observations.

The mathematical model describes interactions between biological components. Analysis of the model, via computational and applied mathematical methods, allows us to deduce the consequences of the interactions. For example, voltage-dependent data on movement of electrically charged ions across a nerve membrane are inputs for models of electrophysiology. The output is a prediction of the dynamics of electrical activity in nerves. The behavior and survival of newly infected individuals are inputs to disease models. The output is a prediction of when and where the disease will outbreak, and how it can be controlled.

To become a successful modeler, modeling tools are required. The first part of this book gives an introduction to some of the more powerful modeling tools, such as discrete models, ordinary differential equations, partial differential equations, stochastic models,

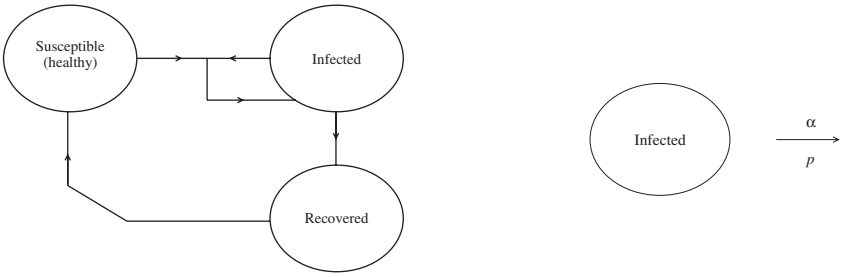


Figure 1.1. (a) *Arrow diagram for a simple epidemic model, showing the relationships between the classes of susceptible, infected, and recovered individuals.* (b) *Subgraph of the arrow diagram in* (a) *representing the recovery of infected individuals, with probability p or rate α.*

cellular automata, and parameter estimation techniques. The second and third parts of the book apply the modeling tools to biological problems.

1.2 Probabilities and Rates

We start with the derivation of a simple epidemic model for the spread of an infectious disease, such as influenza, through a population of healthy individuals. Assume that one infected individual is introduced into the population. In addition, assume that the infection is spread from individual to individual through contact, and that the infected recover after a certain period of time (two weeks for influenza). Recovered individuals are not available to catch the disease again.

Thus, after some time, the population consists of three types of individuals, namely, susceptible (healthy), infected, and recovered individuals. The relationships between these three classes are shown in Figure 1.1 (a). Note that in the diagram, recovered individuals can become susceptible again. In this case, we can think of recovered individuals being temporarily immune to the disease. Individuals return to the susceptible class when the immunity wears off.

In order to create a model for this situation, we need to quantify this diagram. To do that, we follow these three steps:

1. First, we identify the important quantities (the *dependent variables*) to keep track of. In our example, there are three classes of individuals. Let S be the number of susceptibles in the population, I the number infected, and R the number recovered.

2. Second, we identify the *independent variables*, such as time t, space x, or age a, and so on. For our example, we write $S(x,t)$, $I(x,t)$, and $R(x,t)$ if we wish to include time and space dependence, but not age dependence.

3. Finally, we quantify the transitions and/or interactions between these classes, as indicated by the arrows in Figure 1.1 (a). To do this, we use either *probabilities* or *rates*, as explained below.

To explain the use of probabilities versus rates, we consider a subgraph of Figure 1.1 (a), concerning only the recovery of infected individuals, shown in Figure 1.1 (b). In the discussion below, note that we ignore the generation of infected individuals through contact between infected and susceptible individuals (the full epidemic model will be treated in Section 3.3.3). In order to create a model representing this particular process, we apply the three steps outlined above:

1. The dependent variable is the number of infected individuals, I.
2. As time progresses, infected individuals recover. Thus, the independent variable is time, t.
3. If we assume that 2 out of every 100 infected individuals recover per day, then the probability of recovery in a single day is $p_1 = \frac{2}{100}$. The corresponding rate, α_1, is defined as the probability per unit of time, that is,

$$\alpha_1 = \frac{p_1}{\text{unit of time}} = \frac{2}{100} \cdot \frac{1}{\text{day}} = \frac{1}{50}\frac{1}{\text{day}}.$$

Similarly, the probability of recovery in two days is $p_2 = \frac{4}{100}$ (we use p_n to denote the probability of recovering in n days). The corresponding rate, α_2, is then

$$\alpha_2 = \frac{4}{100} \cdot \frac{1}{2\text{ days}} = \frac{1}{50}\frac{1}{\text{day}}.$$

For a time unit of $\frac{1}{2}$ of a day, we get $p_{\frac{1}{2}} = \frac{1}{100}$ and

$$\alpha_{\frac{1}{2}} = \frac{1}{100} \cdot 2\frac{1}{\text{day}} = \frac{1}{50}\frac{1}{\text{day}}.$$

We find that the rate α is independent of the time unit chosen, whereas the probability depends on the chosen time unit. Since the rate is independent of the chosen time unit, we can generalize. Let Δt denote a general unit of time, and let $p_{\Delta t}$ be the probability of recovering in Δt. Then the number of infectives after one unit of time is given as

$$I(t + \Delta t) = I(t) - p_{\Delta t} I(t).$$

With some rearrangements, we get

$$\begin{aligned} I(t + \Delta t) - I(t) &= -p_{\Delta t} I(t), \\ \frac{I(t + \Delta t) - I(t)}{\Delta t} &= -\frac{p_{\Delta t}}{\Delta t} I(t), \\ \frac{I(t + \Delta t) - I(t)}{\Delta t} &= -\alpha I(t), \end{aligned}$$

where now the rate $\alpha = \frac{p_{\Delta t}}{\Delta t}$ appears.

Since α is constant for all values of Δt, we can take the limit as $\Delta t \to 0$. On the left, we obtain the differential quotient, and we obtain the following equation governing the dynamics of $I(t)$:

$$\frac{d}{dt} I(t) = -\alpha I(t).$$

To summarize, for the simple subgraph shown in Figure 1.1 (b), we found two models, namely, a *discrete-time model* with probabilities,

$$I(t + \Delta t) = I(t) - p_{\Delta t} I(t), \tag{1.1}$$

and a *continuous-time model* with rates (a *differential equation*),

$$\frac{d}{dt} I(t) = -\alpha I(t). \tag{1.2}$$

Both models can be solved, analyzed, and simulated. For the discrete-time model, (1.1), we have to specify a time unit, say $\Delta t = \frac{1}{2}$ day. Then $p_{\Delta t} = p_{\frac{1}{2}} = \alpha \cdot \frac{1}{2}$ day. If we define $I_n := I(n \cdot \Delta t)$, then we obtain the simple *difference equation*

$$I_{n+1} = \left(1 - P_{\frac{1}{2}}\right) I_n,$$

which has the solution

$$I_n = \left(1 - P_{\frac{1}{2}}\right)^n I_0, \qquad n \geq 1,$$

where I_0 denoted the initial number of infected individuals. The differential equation, (1.2), is solved by an exponential, $I(t) = I(0)\, e^{-\alpha t}$. The latter solution indicates that the number of infected individuals decreases with time, as expected intuitively (recall that the generation of new infected individuals has been ignored).

In Figure 1.2, we compare the solutions of the discrete-time and continuous-time models over a time period of 15 days, starting with 100 infected individuals ($I(0) = I_0 =$

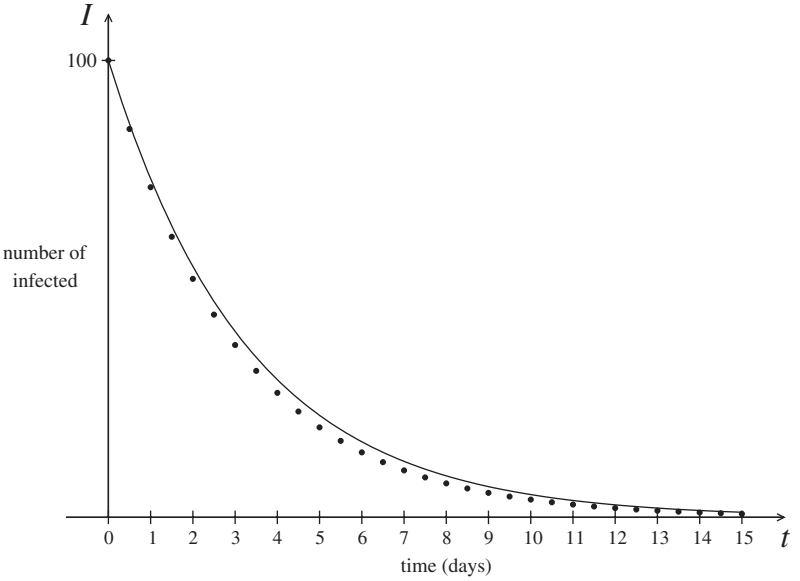


Figure 1.2. *Comparison of the solutions to the discrete-time model,* (1.1)*, and the continuous-time model,* (1.2)*, starting with* 100 *infected individuals (*$I(0) = I_0 = 100$*), and using a recovery rate of* $\alpha = 0.3$*.*

100) and using a recovery rate of $\alpha = 0.3$ per day and a time increment of $\Delta t = 1/2$ day. The full epidemic model corresponding to the arrow diagram shown in Figure 1.1 (a) will be discussed in detail in Section 3.3.3.

In Figure 1.2, the agreement of the discrete and the continuous models is quite convincing. However, this is not always the case. In Exercise 1.4.2, the reader is asked to vary Δt and to investigate if the agreement is still good.

1.3 Model Classes

In the previous section, we derived two models for the recovery of infected individuals, namely, a discrete-time or difference equation, (1.1), and a differential equation, (1.2). The difference between these models is that the time variable is discrete for the difference equation, whereas it is continuous for the differential equation. So far, both models appear suitable. The final choice of model depends on the scientific question asked, the purpose of the model, the available data, etc.

The independent or state variables also can be chosen to be either discrete or continuous. For example, a discrete state variable may represent the number of individuals in a population, whereas a continuous variable may represent a density or a concentration.

Both of the above models are called *deterministic*. That means that if you know the state of the system at a certain point in time t, you can *determine* all future states by solving the corresponding model. Sometimes, however, *stochastic* effects play a dominant role. For example, in a laboratory setting you can predict that a pair of healthy rabbits will produce offspring. Outside the laboratory, life is less predictable, and the same pair of rabbits may not reproduce. In general, stochastic variations are more important for small population sizes. A model for small populations and unpredictable environments should include the uncertainty via a stochastic formulation. Large populations in constant environments (such as an aggregate of cellular slime molds, which contains about 100,000 cells) usually are modeled by deterministic models.

The number of choices presented above generates many types of models. A discussion of all types of models is beyond the scope of this book. We have chosen to restrict the material in this book to the most common model classes, summarized in the following list:

Difference Equations: The state (or dependent variable) can be discrete or continuous but the time is always discrete. Discrete models are suitable for seasonal events. We treat deterministic difference equations in Chapter 2 and stochastic difference equations in Chapter 5.

Ordinary Differential Equations (ODEs): ODEs are used to describe population evolution over a continuous time period. Deterministic ODEs are one of the major modeling tools and are discussed in detail in Chapter 3. The theory of stochastic differential equations is quite involved and is not covered in this book.

Partial Differential Equations (PDEs): PDEs are used if two or more continuous independent variables are used, for example, time and space, or time and age. We discuss age-structured models and reaction-diffusion equations for spatial spread in Chapter 4. Historically, stochastic PDEs were used primarily in the context of statistical

physics. Only recently have such models been considered to describe population dynamics (see, e.g., [76]).

Stochastic Processes: Stochastic processes and *Markov chains* are completely stochastic model classes. They are particularly useful for small populations. We treat them in detail in Chapter 5.

Cellular Automata: Cellular automata and related models are fully discrete models. All independent variables (such as time and space) and all dependent variables (such as population sizes) are discrete. The analysis of cellular automata is mainly restricted to computer analysis and numerical simulation. We give an introduction in Chapter 6. Cellular automata can be either deterministic or stochastic, using a random number generator.

1.4 Exercises for Modeling

Exercise 1.4.1: Discrete-time versus continuous-time models. *Assume you have a culture of bacteria growing in a petri dish, and each cell divides into two identical copies of itself every* 10 *minutes.*

(a) *Choose a unit of time, and find the corresponding probability of cell division.*

(b) *Write down a discrete-time model which balances the amount of cells at time t and at time $t + \Delta t$.*

(c) *Define the growth rate, and derive the corresponding continuous-time model.*

(d) *Solve both the discrete-time and continuous-time models, and compare the solutions.*

(e) *When is a discrete-time model appropriate? When is a continuous-time model appropriate?*

Exercise 1.4.2: Comparison of discrete and continuous models. *Study the two models* (1.2), (1.1) *which lead to Figure* 1.2 *and vary the time increment Δt (e.g., try $\Delta t = \frac{1}{4}$ day, $\frac{1}{8}$ day,* 1 *day,* 2 *days,* 10 *days). What do you observe? Which choice of Δt gives the best, and which gives the worst agreement? Can you explain why?*

Exercise 1.4.3: Structured populations.

(a) *Give examples of spatially structured problems. What kind of effects cannot be understood without spatial structure?*

(b) *Give an example of an age-structured problem.*

(c) *Give an example of a size-structured problem.*