

A Terminal Post-Calculus-I Mathematics Course for Biology Students[†]

Glenn Ledder¹, Department of Mathematics, University of Nebraska-Lincoln

Name of Institution: University of Nebraska-Lincoln	
Size	about 24,000 students
Institution Type	large state university with PhD program
Student Demographic	recent high school graduates with high potential and interests in math and/or biology
Department Structure	Mathematics and Biology are individual departments in the College of Arts and Sciences

Abstract

Most innovative mathematics curricula for biology students assume the form of a two-semester calculus sequence that diverges quickly from the standard calculus sequence. However, there are important groups of students for whom this plan is not ideal. The University of Nebraska-Lincoln has developed a model in which the first course can be any introductory calculus course and the second course is a terminal calculus-based mathematics course that focuses on modeling, probability, and dynamical systems. In this paper, we consider the pedagogical issues raised by such a course and provide details on the topics taught in the course and the methods used to teach them.

Course Structure

- Weeks per term: 15
- Classes per week/type/length: five 1-hour lecture/lab/recitation periods each week
- Labs per week/length (if any): 1-hour laboratory periods as needed
- Average class size: 5-10 students in one section
- Enrollment requirements: One calculus course
- Faculty/dept per class, TAs: Taught by one mathematics instructor, with a TA for grading.
- Next course: This course gives students the necessary background to take more advanced interdisciplinary courses in mathematical biology. Both the Department of Mathematics and the School of Biological Sciences have such a course.

Introduction

The need for a new mathematics curriculum for biology students has been well documented (National Research Council, 2003). Most of the response to this need has come in the form of two-semester curricula that distinguish biology students from other science students at the beginning of their academic careers. We have taken a different approach at UNL because

1. In the American educational system, many students enter college uncertain of their major and career plans. It is desirable to have a flexible curriculum that allows them to defer their choice of program until after their first semester. This suggests a two-semester curriculum that is sufficiently similar to the standard engineering calculus to allow students to transfer from one to the other after the first semester.
2. In addition to an undergraduate major in biology, UNL has a major in fisheries and wildlife biology. Students in this program are required to take any one calculus course, and most choose the 3-credit

[†]supported by NSF grant DUE 0536508

¹gledder@math.unl.edu

business calculus course over the 5-credit engineering calculus or biocalculus courses. After encountering quantitative problems in their major courses, some of these students are interested in developing more mathematics background, but lack the prerequisites for a biocalculus II course. This suggests a terminal course in useful mathematics for which any calculus I course can serve as a prerequisite.

3. UNL has faculty in the life sciences who are interested in mathematical biology. These faculty members have PhD students who need to learn mathematical tools for their research. Such students do not have the background for courses in differential equations, linear algebra, or mathematical biology, and they lack the time to take an entire calculus-for-biology sequence. This suggests a terminal course in useful mathematics beyond calculus—a course that allows them to succeed in spite of having no recent work in calculus.

These considerations led to the development of a course called Mathematical Methods for Biology and Medicine (MMBM), which was taught at UNL five times from 2005 through 2010. The course gradually evolved in an effort to better meet the needs of its several student populations:

- Undergraduate biochemistry majors who have recently taken a 5-credit calculus I course;
- Undergraduate biology majors who have recently taken a 5-credit calculus I course;
- Undergraduate fisheries and natural resources majors who have recently taken a 3-credit business calculus course;
- Graduate students from various life science departments, for whom calculus is but a distant memory.

The needs of these groups were determined by consultation with the undergraduate curriculum committees of the client departments and the mathematically-oriented ecology and natural resources faculty. The standard biochemistry curriculum requires two 5-credit calculus courses but lacks statistics or differential equations. The standard biology curriculum requires a 5-credit calculus I course and either the second calculus course or a 3-credit non-calculus-based statistics course; most biology students take the latter because of the smaller number of credits, and nearly all students find it to be of little value.

Course Overview

We decided to offer Mathematical Methods for Biology and Medicine as a 5-credit course to accommodate the biochemists' desire for an alternative to calculus II worth the same number of credits. This provided ample time for important topics but made the course less attractive to undergraduate biology majors. Because the course met the needs of the biochemistry students better than calculus II, our hope was that it would eventually become the default second course for biochemistry majors, assuring us of a stable enrollment. The class met for a 50-minute period five days each week in a classroom with computers. Having me in the classroom every day gave me the flexibility of being able to mix lecture and computer laboratory activities according to pedagogical needs. For continuing viability, we would have eventually needed to have a teaching assistant take the course one day each week for computer laboratory activities and review—this move is necessary for institutionalization of the course because faculty loads are never more than seven credits in a semester.

Table 1 shows the outline of topics for the course. The first four items are lecture units, which will be discussed in more detail later. The student presentations are informal, with each student giving a talk about ten minutes long using either a slide presentation or the board. The students are asked to find a journal article or book chapter in which mathematics is used for research in a life science area. I provide guidance if they need help finding a source. They are instructed to focus on the biological setting, a very general description of the mathematics used, and the biological significance of the results. The mathematics must go beyond mere statistical analysis of data, but I do not expect them to understand the details. The point of this assignment is to give the class a sample of the many uses of mathematics in biology. I allow a couple of minutes for friendly questions, and I can almost always add a little additional information to the presentations using my general knowledge of mathematical models in biology.

Table 1. General Outline of Mathematical Methods for Biology and Medicine.

Mathematical Modeling	2-3 weeks
Review of Calculus	1 week
Probability	4-5 weeks
Dynamical Systems	5 weeks
Student Presentations	1 week

The lecture portion of the course used my lecture notes, as I could not find a suitable textbook.² The main reasons for this are that there is no published treatment of mathematical modeling as I envision it and that all the published materials on dynamical systems require more background knowledge than I can expect from my students. In the following discussion of the lecture materials, I emphasize my strategies for teaching these advanced subjects to students with minimal background.

Mathematical Modeling

Mathematical modeling involves constructing mathematical models from a conceptual model (a simplified narrative description of the biological principles believed to be at work), determining parameter values from data, characterizing and testing models, and using them to make predictions about scientific experiments. These skills are not generally taught anywhere in the mathematics curriculum. I introduce them in a precalculus setting so that I can start the course with an area that is familiar to everyone and the modeling skills can be reinforced in the calculus review portion of the course. The topics in the modeling unit are listed in Table 2; each is discussed in some detail below.

Table 2. Mathematical Modeling Topics

Functions with Parameters
Concepts of Modeling
Fitting Models to Data
Empirical and Statistical Modeling
Mechanistic Modeling

1. Functions with Parameters

Parameters are an essential feature of mathematical models, as they provide ways to incorporate numerical data that may be uncertain or may differ among cases. Previously, students have learned to interpret symbols as either dependent or independent variables, and they find it difficult to understand parameters and to manipulate functions that contain them. Lack of experience with parameters is one reason why calculus students have trouble understanding the definition of the derivative, which is based on a family of secant slopes with the horizontal distance between the points as the parameter. Understanding parameters is useful, but not central, in calculus; most students complete calculus in spite of not understanding them. We make this possible by avoiding problems with parameters on exams. In contrast, parameters are so integral to mathematical modeling that students are unable to learn the subject at all without a minimal facility.

In my introduction to parameters, I focus on the ways they affect the graphs and analytical properties of functions. Figure 1 illustrates the function $f(x) = x^3 - 2x^2 + bx$. The parameter b changes the shape of the graph, qualitatively as well as quantitatively. When we reexamine this function in the calculus review unit, we can

²My book *Mathematics for the Life Sciences: Calculus, Modeling, Probability, and Dynamical Systems* will be published by Springer-Verlag in 2013.

determine the bifurcation point; that is, the value of the parameter where the function changes from monotone to not monotone.

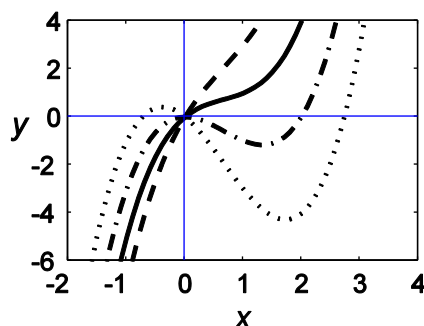


Figure 1. $f(x) = x^3 - 2x^2 + bx$ with $b=-2$ (dotted), $b=0$ (dash-dot), $b=2$ (solid), and $b=4$ (dashed)

2. Concepts of Modeling

In their previous experience with applications of mathematics, students have been misled into thinking that the scientific value of a model is equivalent to its mathematical value. It is important to disabuse students of this notion. At best, we may claim that a mathematical model is *valid* or *useful*; we should not be so bold as to say that it is *correct* or *true*. Results obtained from mathematical models are often mathematically correct but counterintuitive. Such results are usually wrong. This is not a mathematical issue. Mathematics can determine the properties of models, but it cannot determine their validity. Validation of models requires comparison with data.

I want my students to understand that mathematical models can be treated in two ways. The general case of a model can be analyzed mathematically using the ideas learned in an initial study of parameters. Specific instances of a model can be worked out by numerical simulations, which is useful if a valid set of parameter values is known. It is also important for students to appreciate that the same model can take different symbolic forms by using different parameters. As an example, the mathematical formulations $f(t) = f_0 e^{-kt}$ and $f(t) = B e^{-t \ln 2/T}$ represent the same model, despite the different symbols for the initial value parameter and the use of the decay constant k in one and the half-life constant T in the other. This is far from obvious to students and needs emphasis. Similarly, the relationship between these dimensional models and the corresponding dimensionless model $F(\tau) = e^{-\tau}$ is difficult for students to understand.

3. Fitting Models to Data

The parameter values used in the examples for the previous sections were chosen for mathematical convenience. In practice, students need to know how to choose parameter values to fit a set of data. I focus on three simple models,

$$Y = mX, \quad y = b + mx, \quad z = Ae^{-kt},$$

which I do using linear least squares. I've never seen the 1-parameter model $Y = mX$ treated as a separate case, but I find it useful when the graph of the function is expected on scientific grounds to pass through the origin. The mathematical benefit from treating this special case first is that we can define a one-variable function,

$$F(m) = \left(\sum X^2 \right) m^2 - 2 \left(\sum XY \right) m + \left(\sum Y^2 \right)$$

to represent the residual sum of squares for any slope m . Its minimum is the vertex of the parabola, which can be found algebraically if necessary (for students who have not had any calculus). Once the students have learned to fit the one-parameter straight line, the two-parameter straight line follows easily from the assumption that the best fit

line must go through the mean (\bar{x}, \bar{y}) of the data (this can be proven using multi-variable calculus, but I ask the students to accept it on intuitive grounds). The original xy data can be converted to XY data by

$$X = x - \bar{x}, Y = y - \bar{y}.$$

The slope m can then be found as before, and the intercept b follows by replacing X and Y in the model $Y = mX$ by the same formulas.

4. Empirical and Statistical Modeling

In biology, it is common to choose mathematical models by reverse engineering from data. As a simple example, we can try to choose between a power function model and an exponential function model by plotting the data as $\ln y$ vs $\ln x$ and as $\ln y$ vs x ; a power function model is used if the former plot suggests a straight line, while an exponential function model is used when the latter plot is more suggestive of a straight line. In scientific practice, choosing between competing models by looking at graphs has been supplanted by the use of the Akaike Information Criterion (AIC), which assesses the relative statistical support for competing models, given a particular data set, by balancing residual error against complexity (see McQuarrie and Tsai (1998)). It is curious that this useful tool has not become part of the standard elementary statistics syllabus since its discovery in 1974. Perhaps this is because the derivation of the formula is quite difficult.³ Nevertheless, its application is simple, and anyone doing mathematical modeling in biology should be able to use it. The simplest example of its benefit is that it shows the one-parameter model $y = mx$ to be superior to the two-parameter model $y = b + mx$ in just those cases for which scientific principles suggest the simpler model. For example, if x is the number of prey animals available to a predator and y is the number of prey animals consumed in a unit time interval, it is a logical necessity that $y(0) = 0$. A minimum error fit of real data to the model $y = b + mx$ yields nonzero b ; the more complicated model has smaller error, but is qualitatively wrong. With reasonably good data, AIC is almost certain to select the simpler model.

It is particularly important for students to understand why complexity in a formula is a drawback. It is not merely because simple formulas are easier to work with. It is because complex formulas often have complicated behavior that does not match the simple behavior expected by theory. Figure 2 shows a data set generated by adding a small amount of random noise to the function $Y = 0.3X$, along with two attempts to fit the data to a model. The dashed line is the model $Y = mX$, while the curve is a quartic polynomial; both models were parameterized using the odd-numbered points, with the result shown in Figure 2a. The quartic polynomial fits the data exactly and would therefore be preferred if accuracy were the only criterion. However, the inclusion of the even-numbered points in Figure 2b shows that the quartic polynomial is oversensitive to variations in the data, which more than compensates for the advantage of perfect accuracy of fit to a small data set. Of course we could fit a quartic polynomial to the full data set, but the fit would again be problematic if we added more data points.

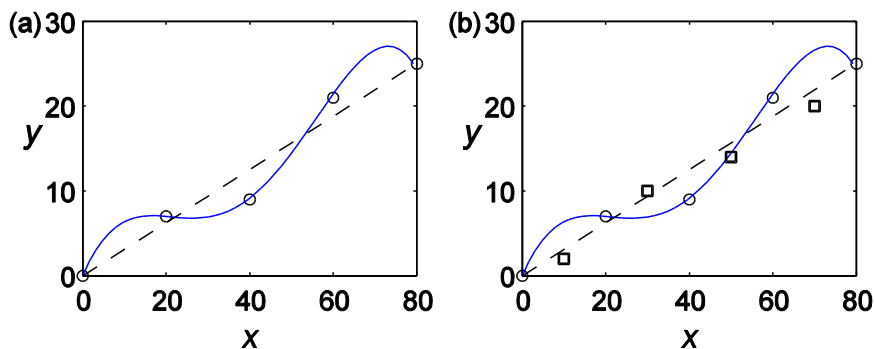


Figure 2. A dataset fitted by linear and quartic models; the parameters were determined using only the odd-numbered points, as shown in (a), while (b) shows the same parameter choices with all of the data.

³While I agree that we should not teach unmotivated (black box) methods, I think most mathematicians go too far with their insistence that everything must be rigorously proved. I like to characterize mathematicians as people who believe that you shouldn't drive a car unless you have built it yourself.

5. Mechanistic Modeling

Here we discuss the derivation of models from assumptions based on biological principles, a topic that has been discussed elsewhere (Ledder, 2008). I also indicate a relatively simple method for parameterizing models of the form $y = mf(x; p)$, where m and p are parameters. The method utilizes the special structure of a model in which one parameter enters nonlinearly while the other appears only as a scaling factor:

1. Let $t_i(p) = f(x_i; p)$ for each data point (x_i, y_i) and a range of p .
2. Define a function $G(p)$ by

$$G(p) = \min_{m \in \mathbb{R}} \sum [y_i - mt_i(p)]^2.$$

3. The best fit for p is the value that minimizes the function G .
4. The best fit for m is the value that minimizes the sum when p is the best fit value.

The method works because the nonlinear model reduces to the linear model $y = mt$ for any particular p , allowing us to determine the residual sum of squares as a function of p . The function $G(p)$ can be computed numerically and plotted to obtain an estimate for the best fit value of p . Of course we could use nonlinear optimization software as a black box, but it is much more instructive to use a method that students can understand.

Calculus

The course requires one calculus course as a prerequisite, but some of the students have not seen calculus in several years. This poses a difficulty, as some of them could use a semester-long refresher course while others need at most a day or two of review. I address this problem by deemphasizing use of techniques of calculus in the subsequent units on probability and dynamical systems and using functions with parameters whenever possible. I expect my students to be able to use the basic differentiation rules for simple functions, the fundamental theorem for definite integrals that require only elementary antiderivative formulas and substitution, and Simpson's rule for definite integrals that cannot be calculated by the fundamental theorem. The primary focus of the review is on the concepts of calculus, because it is the concepts that generate the applications. Thinking of the derivative as the slope of a graph leads naturally to its use for optimization. Similarly, thinking of the definite integral as accumulation in time, space, or structure leads to more obvious applications, such as obtaining distance from velocity or total population from spatial population density, and to more subtle applications, such as obtaining the total population from an age-distribution function. From here, one can derive Euler's equation for the growth rate of an age-structured population fairly simply. If $l(x)$ and $m(x)$ are the probability of survival to age x and the rate of offspring production per parent of age x respectively, then the population of parents of age x is $B(t - x)l(x)dx$ and so the total rate of offspring production is

$$B(t) = \int_0^\infty B(t - x)l(x)m(x)dx.$$

However, if the population grows at rate r , then the birth rate must also grow at rate r ; hence, we also have $B(t) = B(0)e^{rt}$. Substituting this formula into the integral equation yields Euler's equation

$$1 = \int_0^\infty e^{-rx}l(x)m(x)dx,$$

which students can explore by using specific functions for survival and fecundity, obtaining an estimate for r by trial and error (Comar, 2008).

Probability

Table 3 lists the topics in the unit on probability. The topics and their order are based on two principles:

1. Students learn best when probability is approached experimentally rather than axiomatically.
2. Biologists are most interested in having their students learn how to work with probability distributions.

Table 3. Probability Topics

Characterizing Data
Concepts of Probability
Discrete Distributions
Continuous Distributions
Distribution of Sample Means
Conditional Probability
Estimating Parameters with Likelihood

My first principle is motivated by the pedagogical work of Lock and Lock (2008). The second is the result of conversation with biologists at my institution. A consequence of these principles is that I don't spend much time on the usual introductory problems that use a uniform discrete distribution with combinatorics. I do work with probabilities of various nucleotide combinations in DNA, but I avoid problems with card games or colored marbles in an urn. Another consequence is that I do probability distributions before conditional probability. It is logical to progress from studying sample means, which are the results of multiple experiments, all similar, to studying probability in multiple experiments, each different.

I think of a probability distribution as model of an infinite population, so it is natural to begin with elementary descriptive statistics of finite samples and move directly to probability distributions. Our study of distributions emphasizes the definitions of the fundamental distributions, using the probability distribution function for discrete distributions and the cumulative distribution function for continuous distributions and using them to calculate probabilities. I spend a lot of time on distribution of sample means because they are central to understanding what biologists actually measure. Each measurement made by a biologist corresponds to drawing a number from a probability distribution. However, the variances of distributions in biology are high. Biologists thus take multiple measurements, so measurements in biology are means of samples from a distribution. Biologists need to understand that their measurement constitutes a single number drawn from a distribution of sample means rather than a number drawn from a distribution of individual experimental values. If an experiment is conducted flawlessly, there is still a 4% chance that the true mean of the population lies more than two standard deviations (of the distribution of means, not the distribution of individuals) from the mean calculated from the experimental data. This probability of anomalous results cannot be decreased; the only way to be confident in a biological measurement is to average together so many individual measurements that an error of two standard deviations in the mean is acceptable.

My coverage of the final two topics depends on the makeup of the class. Conditional probability is particularly important in the understanding of medical test design. Rare conditions are hard to test for because a small probability of false positives can nevertheless yield more false positives than true ones. The last topic is of special value to practicing ecologists, so I present it only if there are ecology graduate students in the class.

Dynamical Systems

Discrete and continuous dynamical systems are usually treated separately. I find it beneficial to treat them together in the context of one-dimensional systems before separating them for multi-dimensional systems. The outline of topics appears in Table 4. My presentation follows several pedagogical principles:

1. Students should have experience with dynamical systems in scalar form before attempting a more mathematical treatment with vector notation.
2. The primary goal of mathematical analysis is to determine the long-term behavior of a system in different regions of the parameter space. (In particular, I have my students obtain analytical solutions only for the trivial exponential growth equations.)

Table 4. Dynamical Systems Topics

Single Variable Dynamics
Discrete Models
Continuous Models
Cobweb Plots
The Phase Line
Stability Analysis
Discrete Dynamical Systems
Discrete Linear Models
A Matrix Algebra Primer
Eigenvalues and Eigenvectors
Theoretical Results
Continuous Dynamical Systems
Continuous Models
The Phase Plane
Stability for Linear Systems
Stability for Nonlinear Systems

Elementary genetics serves as a nice introductory example for discrete one-variable models (Comar, 2008). Suppose p is the gene pool frequency of the malaria susceptibility trait **A** and $q = 1 - p$ is the frequency of the sickle cell trait **a**. Suppose these traits are randomly distributed⁴ among individuals in a population. Then the genotypes **AA**, **Aa**, and **aa** have frequencies p^2 , $2pq$, and q^2 , respectively. If we assign a value of 1 to the fitness of the heterozygote **Aa**, then the fitness of the homozygote **AA** is some number $m \in (0,1]$ whose value depends on the risk of malaria. The homozygote **aa** results in sickle cell anemia, with fitness 0. Given these assumptions, the relative frequencies of parents are in the ratio $(1 - m)p^2 : 2pq$. Hence, the next generation has traits **A** and **a** with relative frequencies given by the ratio $[2(1 - m)p^2 + 2pq] : 2pq$, from which we can obtain the difference equation

$$q_{t+1} = \frac{q_t}{(1 - m)(1 - q_t) + 2q_t}$$

for the sickle cell trait. Students can experiment with this model, with questions such as

1. If the risk of malaria is extremely high ($m \rightarrow 0$), what happens to the gene pool?
2. Suppose a population with a given initial q moves to a non-malarial environment. How does the frequency of the sickle cell trait change over 100 generations?

My favorite continuous model is the resource utilization model

$$\frac{dx}{dt} = f(x) - Cg(x),$$

where x is the level of some resource, f is a function representing its natural change, g is a function representing the decrease in level caused by an average consumer, and C is the fixed number of consumers. We can use the prey equation from any predator-prey model, with the level of predators taken as a parameter rather than a dynamic variable. We use Matlab's built-in numerical differential equation solver to run simulations with models of this form. Students quickly discover the concepts of stable and unstable equilibria, which are then explored mathematically in the subsequent sections on the phase line and stability computation.

I have my students develop a stage-structured population model as an example of a discrete system. Numerical experiments (see Ledder et al (2012)) lead naturally to the discovery of the dominant eigenvalue and its eigenvector as the long-term stable growth rate and stage-distribution ratio. I present the formal theory of eigenvalues only after this initial exploration.

⁴ This discussion is facilitated by my opting to discuss probability before studying dynamical systems.

There are a number of good models to use as introductory examples for continuous dynamical systems. Pharmacokinetics provides a good example of a linear system, and Michaelis-Menten kinetics provides a relatively simple example of a nonlinear system. Depending on the student populations, I will also use the Perelson and Nelson (1999) HIV model or a predator-prey model with logistic growth in the prey and a Holling Type 2 functional response (Brauer and Castillo-Chavez, 2001). These examples help students to learn nullcline analysis and linearized stability analysis.

Postscript

Unfortunately, the 5-credit course described here never caught on with biochemistry freshmen, in spite of its acceptance by the biochemistry faculty. The difficulty was that students are more likely to take advice from their peers than from their professors. Establishment of a course depends on satisfying one of two conditions: either the course must be required, or the curriculum must be such that the course is clearly the path of least resistance for a large number of students. Biochemistry students could not be persuaded to select an unfamiliar course to replace the familiar Calculus II. I am now working with the School of Biological Sciences at UNL to repackage the course material in a way that makes the course part of the path of least resistance for biology majors. This will require the course to be offered as an alternative to Calculus I, which means a 5-credit course in calculus and dynamical systems. The irony is that this change will take me back to tracking students into biomath courses in the first semester, precisely what my project was intended to avoid. The clear lesson, stated in terms of the parts of this volume, is that no *model* can survive unless it is part of a successful *process*.

Conclusions

Mathematical Modeling for Biology and Medicine never had enough students to produce formal data regarding its success. However, I collected feedback from the students each time I taught the course, and I kept track of student success with the various topics. Students have consistently given positive evaluations of the course, but my analysis of the outcomes of the course is mixed. The dynamical systems portion of the course has been most successful. While topics such as nullcline analysis and linearized stability analysis are normally taught late in a differential equations course, they do not require background other than facility with algebra and graphing, an understanding of the derivative as a rate of change, and some ability to compute derivatives. The probability portion of the course has improved in the last two runs, perhaps owing to the topic list and pedagogical methods presented here, which were somewhat different from prior implementations of the course. The mathematical modeling portion of the course is the one that seems most difficult to the students. They generally do well with fitting models to data, but they have trouble working with parameters, in spite of the attention I give to the topic. In particular, my classes have not done well in understanding dimensionless form. However, the idea presented here of beginning the course by studying functions with parameters is new and as yet untested. I am optimistic that the mathematical modeling portion of the course can be as successful as the other portions, given enough class time and a pedagogy that subdivides the topic into a hierarchy of skills.

Ultimately, the success of any new course depends on attaining two goals:

1. It must develop a steady enrollment so that it can be taught as a regular (not externally-funded) course offering.
2. It must become a department course rather than an individual instructor course; that is, the materials and pedagogy need to be sufficiently well documented that other faculty can teach it. I am hopeful that future publication of my materials in book form will make it possible for courses such as mine to be successful.

References

- Brauer, F., and C. Castillo-Chavez, 2001: *Mathematical Models in Population Biology and Epidemiology*, Springer-Verlag.
- Comar, T.D., 2008: The integration of biology into calculus courses. *PRIMUS* **18**, 49--70.
- Ledder, G., 2008: An experimental approach to mathematical modeling in biology. *PRIMUS* **18**, 119--138.

Ledder, G., B. Tenhumberg, and G.T. Adams, 2012: An interdisciplinary research course in mathematical biology for young undergraduates. This volume, xx--yy.

Lock, R.H. and P.F. Lock, 2008: Introducing statistical inference to biology students through bootstrapping and randomization. *PRIMUS* **18**, 39--48.

McQuarrie, A.D.R. and C.-L. Tsai, 1998: *Regression and Time Series Model Selection*, World Scientific.

National Research Council, 2003: *Bio 2010: Transforming Undergraduate Education for Future Research Biologists*, National Academies Press.

Perelson, A.S. and P.W. Nelson, 1999: Mathematical analysis of HIV-1 dynamics *in vivo*. *SIAM Rev.* **41**, 3--44.