An Experimental Approach to Mathematical Modeling in Biology

Glenn Ledder*
University of Nebraska-Lincoln, Department of Mathematics,
203 Avery Hall, Lincoln, NE 68588-0130
gledder@math.unl.edu.

October 2, 2007

keywords: mathematical modeling, population dynamics, biology, Leslie models, matrix models, population projection

Abstract

The simplest age-structured population models update a population vector via multiplication by a matrix. These linear models offer an opportunity to introduce mathematical modeling to students of limited mathematical sophistication and background. We begin with a detailed discussion of mathematical modeling, particularly in a biological context. We then describe Bugbox-population, a virtual insect laboratory that allows students to make observations and collect quantitative data easily, thereby learning mathematical modeling in the context of its use in scientific research. Creating a mathematical model for boxbugs involves the same intellectual work as creating a mathematical model for real insects, but without the difficulties involved in collecting real biological data. The analysis of the Bugbox-population data leads to the development of the eigenvalue problem for population projection matrices.

INTRODUCTION

To work together, mathematicians and biologists do not need to be experts on each other's subject, but they do need to have a good understanding of how mathematics and science interact. One interaction is in the use of mathematical techniques, in the form of statistics, to analyze biological data. This is fine as far as it goes.

Another interaction between mathematics and biology is in the use of mathematical models. Where statistics can only describe data, mathematical models can yield quantitative predictions and qualitative scientific insight. Unfortunately, few mathematicians and even fewer biologists have any training in mathematical modeling. Mathematics books generally contain what are called "applications" of mathematics, perhaps including some applications

---

*supported by NSF grant DUE 0531920

in biology. “Applications,” as presented in most mathematics texts, are an inadequate introduction to mathematical modeling because of several standard features:

1. The author has done the hard work of deciding what assumptions to use in the mathematical description.
2. Quantities that are dependent on circumstances, such as rates of radioactive decay, are assigned numerical values rather than being given as symbols representing a range of possible values.
3. There is no mention, much less discussion, of the need to test the predictions of the mathematical work before accepting the results as valid.
4. The questions are about specific details that can be answered by routine computations rather than general features requiring thoughtful analysis.

Because mathematical models are mathematical objects, I believe that it is primarily the responsibility of mathematicians to teach mathematical modeling. In this paper, I present an exposition of models and modeling along with a detailed example of how I teach mathematical modeling in its broader context of theoretical science. This task is facilitated by a software applet called Bugbox-population that stands in for a real scientific experiment.

MATHEMATICAL MODELS

Any discussion of mathematical modeling must begin with a definition of the term “mathematical model.”

mathematical model: *a self-contained set of formulas and/or equations based on an approximate quantitative description of real phenomena and created in the hope that the behavior it predicts will resemble the real behavior on which it is based.*

In this definition, I am thinking primarily of deterministic models; however, stochastic models fit the same definition if by “solution” we mean the probabilistic properties of the model rather than the results of a single instance.

As an example of a deterministic mathematical model, consider Newton’s model for projectile motion, which can be found in any calculus-based general physics book and most differential equations books. We may write the model as

$$\frac{d^2y}{dt^2} = -\frac{GM}{(R+y)^2}, \qquad y(0) = y_0 \geq 0, \quad \frac{dy}{dt}(0) = v_0, \tag{1}$$

where $y(t)$ is the height at time t of a projectile above a gravitational body of radius R and mass M, given that the height and velocity are initially y_0 and v_0, with G representing Newton’s universal gravitational constant. This model is self-contained because it has enough information to prescribe a unique solution in terms of M, R, y_0, and v_0. It is preferable to refrain from specifying values for these quantities until it becomes necessary to do so; even if we have the earth in mind as the specific gravitational body, we may still want to apply the model to Mars or the moon, which would require different values for M and R. The model is approximate because Newton's theory of gravity has been superseded by Einstein's theories of special and general relativity. It is also approximate in circumstances where other forces, such as air resistance, are present. We might create this model in the hope that it will be useful for replicating real data, a hope that will be realized if we restrict its use to situations where the approximations inherent in it are realistic. This model does very well at predicting the initial velocity needed for a rocket to escape the moon's gravity; it does very poorly at predicting the time it takes a feather to fall to the ground from an eagle's nest. It is not valid at all for $v_0 < 0$ if also $y_0 = 0$.

Mathematical models inhabit the ideal world of mathematics, not the messy world of reality. This is a strength, in that claims made about them can sometimes be proven. It is also a weakness, in that the results obtained from them may have questionable scientific value. The best models are simple enough for mathematical analysis, complex enough to capture the qualitative behavior of reality in a large range of cases, and accurate enough to make good predictions in a broad setting.

Narrow and broad views

Consider a mathematical model for the amount $y(t)$ of a radioactive isotope, given by

$$\frac{dy}{dt} = -ky, \qquad y(0) = y_0, \qquad k, y_0 > 0, \tag{2}$$

where y_0 is the initial amount of the isotope and k is a constant that represents the relative decay rate. In a typical question from a differential equations text, we might be given values of k and y_0 and asked to determine the amount of the isotope at some future time. To answer questions of this type, we solve the equations of the model to obtain the formula

$$y = y_0 e^{-kt} \tag{3}$$

and insert the appropriate numerical values. A similar question would be to give values for y_0 and also $y(t_1)$ for some time $t_1 > 0$ and ask for the value of k. This question is less direct, but it can still be answered using the solution formula. Both of these questions are examples of what I call the "narrow" view of mathematical modeling. In this view, the quantities y and t are variables and k and y_0 are constants.

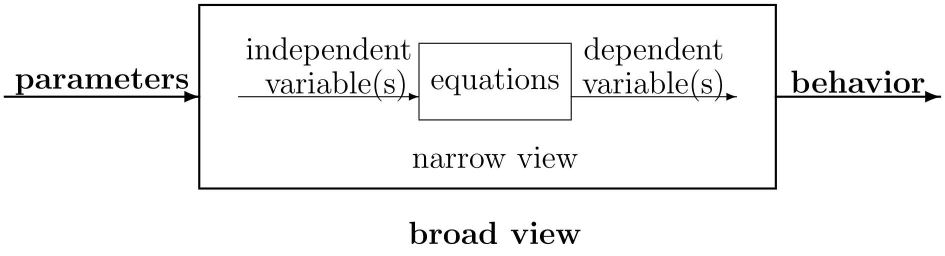

Figure 1: Narrow and broad views of mathematical models

Mathematical models can be seen in a broad view as well as the narrow view (see Figure 1). As an example, suppose we want to know how long it takes for a quantity of a radioactive substance to be reduced to a fraction ρ of the original amount. The question generalizes the concept of the half-life, for which $\rho = \frac{1}{2}$. We need the solution formula for this question too, but we use it to obtain an answer that is a formula rather than a number and applies to the general case rather than a specific case. Let t_ρ be the desired time. We set $y = \rho y_0$ and $t = t_\rho$, obtaining the result

$$t_\rho = \frac{\ln \rho^{-1}}{k}. \tag{4}$$

In this result, the quantities k and ρ play the role of independent variable and t_ρ serves as dependent variable. The original model variables y and t are absent from this broad view. Equation 4 is not particularly profound, but the problem it solves is certainly deeper than the routine questions of the narrow view. More sophisticated models have more profound questions that can be asked in this broad view. For example, we can use the projectile model (1) to determine how the velocity required to escape from a planet's gravitational field depends on the mass and radius of that planet.

The broad view of mathematical modeling requires an understanding of parameters, which we are now ready to define.

parameter: *a quantity in a mathematical model that is intermediate between a constant and a variable: it takes on a single value in the narrow view, in which the model is seen as prescribing the relationship between dependent and independent variables; it serves as an independent variable in the broad view, in which one seeks to characterize the model in terms of parameter values.*

Mechanistic and empirical models

Prior to the advent of computers, data had to be plotted by hand. Graph paper was made using linear scales and logarithmic scales, with the latter making it possible to graph the logarithm of a variable by plotting the variable values directly rather than having to calculate logarithms. In my science classes, we plotted data for variables x and y on regular, semi-log (one linear axis and one logarithmic axis) and log-log paper, keeping whichever graph appeared closest to a straight line. If this was the log-log graph, then we assumed $\log y$ and $\log x$ were related by a linear function, say

$$\log y = b + p \log x.$$

This led to the power function model

$$y = Ax^p, \tag{5}$$

where $A = 10^b$ and p are parameters.

Power functions remain popular as models. There are usually no scientific principles pointing to a power function model. Most often, they are chosen simply because they fit the data better than the other standard models. Models selected in this manner are called *empirical.*

empirical model: *a mathematical model based on reverse engineering of data.*

The radioactive decay model (2) can be chosen on empirical grounds, but it can also be derived from scientific principles. As a thought experiment, imagine that we have several different boxes filled with different quantities of the same substance. We let some time elapse and examine the boxes, whereupon we observe that the same fraction of objects seems to have disappeared from each box. Assuming that the relative rate of decay of a quantity $y(t)$ is constant, we obtain the mathematical statement

$$\frac{1}{y}\frac{dy}{dt} = -k, \tag{6}$$

where k is some positive parameter. A simple rewrite yields Equation 2. Models obtained in this manner are called *mechanistic.*

mechanistic model: *a mathematical model derived from a set of assumptions based on scientific principles.*

Empirical models can be useful for obtaining quantitative predictions, but they can yield no qualitative insight. Mechanistic models offer the potential for qualitative insight as well as quantitative prediction because of their intimate connection to scientific theory. Models for projectile motion are a good case in point. Galileo discovered that the height $y(t)$ of an object falling from initial height y_0 under the influence of gravity is

$$y = y_0 - kt^2, \tag{7}$$

where k is a fixed constant (approximately 4.9 m/sec^2 in modern units). This is an empirical model, as contrasted with the mechanistic model (1) of Newton. Using calculus[1], we can derive Galileo's model from Newton's and identify the k in Equation 7 as $GM/(2R^2)$. Galileo's model only works for falls from heights y_0 that are significantly less than R, a restriction that does not affect Newton's model.

Mechanistic models are more likely than empirical models to be useful beyond the narrow setting for which they were created. There is still a danger that a mechanistic model will be applied in a setting for which it is not appropriate. It is natural, for example, to apply the radioactive decay model for other decay processes, such as the change in temperature of a body that is hotter than its surroundings or the change in population size for a population of animals with no food. These applications may work out well, but we should be cautious. Only by comparing the predictions of a model with observation or experiment can we determine whether the model is successful in the specific context.

In applying mathematical models in novel settings, it is helpful to try to isolate the "approximate" from the "quantitative description" in the definition of a mathematical model. We can do this by writing a verbal description, or *conceptual model*, that corresponds exactly to the mathematical model. All of the approximation is then in the relationship between the real physical setting and the conceptual model used to approximate it. When the model behavior turns out to differ significantly from the real behavior on which it was supposedly based, the conceptual model is defective. We can often obtain a useful mathematical model by working out the defect in the equivalent conceptual model, a process that requires scientific thinking rather than mathematical thinking.

---

[1] Simply apply the approximation $R + y \approx R$ and integrate.

HOW MODELS ARE USED

In general, mathematical models can be used for quantitative and qualitative purposes. We consider these uses in turn.

Models and simulation

Scientific experiments can be costly, slow, and difficult. Experiments in biology are particularly problematic. Mathematical models offer great promise as a way to do experiments cheaply, quickly, and easily. An excellent example of this is the ATLSS model of the Everglades [2, 5], created by The Institute for Environmental Modeling (TIEM) of the University of Tennessee-Knoxville. Suppose we want to know what effect a new construction project would have on the Florida panther population in the Everglades. We cannot run a real experiment, but we can instead run the experiment with the ATLSS model. ATLSS is not guaranteed to give results that are absolutely quantitatively correct; however, it can be used to estimate the effects of a policy decision.

Simulation can be a valuable tool for engineering and policy making. However, there is a potential difficulty: The model used in the simulation could be wrong in the given context. A classic example in biology is the claim sometimes seen in mathematics books that when humans disturb a predator-prey system by hunting predators, the predator population does not decrease at all. This ridiculous result comes from analysis of the Lotka-Volterra predator-prey model (see Equation 8 below). The Lotka-Volterra model was very successful in the setting for which it was derived [4]. However, the model is not valid for all predator-prey settings. We'll return to the Lotka-Volterra model in the next section.

Simulations are used to make quantitative predictions. The quantitative accuracy is limited by uncertainty in the parameter values and the sensitivity of the model behavior to the parameter values. This is not usually a problem in physics, where many parameters are known to a high degree of precision. In biology, parameters can be very difficult to measure. As a simple example, suppose we are trying to predict the effect of a release of smallpox virus into a human population. The model must include parameters such as the fraction of smallpox victims that die. The only information available about these parameters is in historical records, which show a considerable variation. Death rates, for example, depend on the particular strain of the virus and can vary significantly for different groups of people.

Models and science

Science is the systematic search for simple explanations of our complicated world. It is a blend of the empirical, in the form of experiments and observations, and the rational, in the form of theories and mathematical models. Figure 2 indicates the role of mathematical modeling in science.

Science begins with coincidental or deliberate observations, but observations alone do not constitute science. Science requires an attempt to explain the observations through some conceptual framework, which is often in the form of a mathematical model. Those models that explain the observations are tentatively considered to be valid. Predictions made by the models are then tested with experiments. As long as the predictions of a model are verified

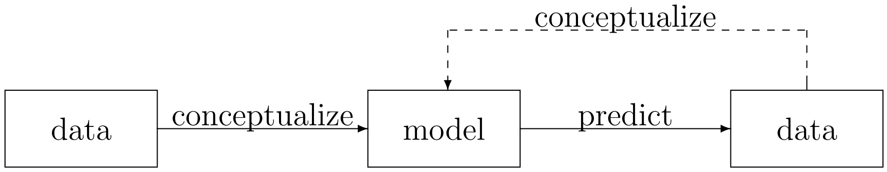

Figure 2: A schematic diagram indicating the role of mathematical modeling in science

by experiment, the situation is stable. This may last for a long time, and in the naïvety of the 19th century it was expected to last forever. The great revolutions in physics in the early 20th century changed the nature of science; it became clear that scientific theories are "useful" rather than "true." It now appears likely that all scientific theories have their limitations, and when these are found we must search for a new conceptualization.

Some areas of theoretical biology have progressed to the stage of mechanistic models. These models lack the predictive power of Newton's model of motion under gravity, but they are still useful.

MATHEMATICAL MODELING

Armed with an understanding of what mathematical models are and how they are used, we turn now to the process of mathematical modeling. I take the rather snobbish point of view that modeling with empirical models should be called "statistical modeling" and that the term "mathematical modeling" should be restricted to work done with mechanistic models. Thus restricted, mathematical modeling is a complete process consisting of conceptualization, characterization, testing, parameterization (usually), and prediction, as depicted in Figure 3. The dashed arrows are included when the purpose of the model is quantitative prediction.

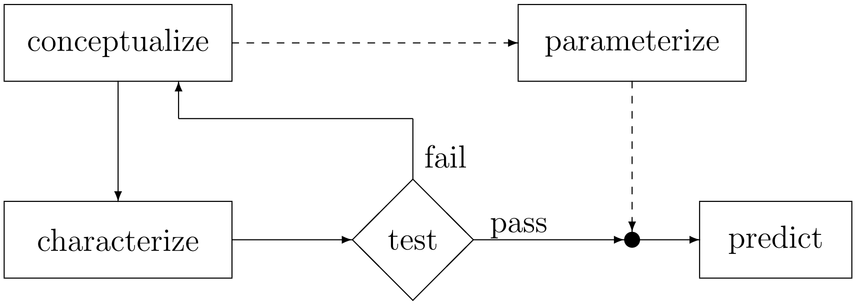

Figure 3: A schematic diagram indicating the logical sequence of mathematical modeling

Conceptualization involves creating a caricature of reality that can be expressed in verbal and mathematical form. By "caricature," I am invoking the image of a political cartoon, in which an oversimplified and exaggerated sketch is easily recognizable as some public figure.

It is common to go directly from the physical setting to the model; however, it is important to understand that every mechanistic model corresponds to some narrative caricature. Making the narrative version explicit makes it easier to interpret the model and to improve it if it is not successful. As an example, consider the standard SIR infectious disease model, which can be found in many differential equations books (or see [3] for a more complete treatment). The derivation assumes that contacts among people occur randomly, which corresponds to a conceptual model of people as molecules in a gas, moving endlessly and haphazardly, each indistinguishable from the others. This obviously flawed assumption does not necessarily render the model useless, but it should at least make us stop to think about the importance of social networks in any given setting. We should not use the SIR model to study an outbreak of a rare disease, because the human interactions of the individual who introduces the disease are clearly important in that setting.

I use the term "characterization," rather than "analysis," to describe the process of using mathematics to determine the behavior of the model over a range of parameter values. For most mathematicians, "analysis" carries a connotation of mathematical rigor, with emphasis on proving theorems. From a modeling point of view, the likelihood of bad conceptualization is much greater than the likelihood of bad characterization. One should focus on improving the conceptualization rather than verifying the characterization with proofs.

Parameterization is the process of choosing parameter values to fit data. In some theoretical investigations, there is no data with which to determine parameter values. In these cases, the question is whether the model can produce the observed behavior for some range of parameter values, and this is decided in the characterization.

Prediction can involve qualitative and/or quantitative features. Qualitative features are such general observations as "The predators die if their natural death rate is too high," with "too high" defined by some inequality comparing the natural death rate to some combination of the other parameters.

The modeling procedure includes a feedback loop starting with the testing step. The test question is

Is the model behavior qualitatively compatible with known observations?

If so, then use of the model is tentatively warranted. If not, then the conceptualization needs to be redone. This reality check is a crucial component of good mathematical modeling. We have to remember that models are caricatures created in the hope that their behavior resembles the real behavior on which they are based. If the model behavior does not adequately mirror the real behavior, then the model is wrong.

As an example, suppose we want to make a model to study the effect on a predator-prey community if we increase the predator death rate through human activity, such as hunting of predators. We might explore this issue by starting with the well-known Lotka-Volterra model, which we can write as

$$\frac{dx}{dt} = rx - bxy, \qquad \frac{dy}{dt} = cbxy - my, \tag{8}$$

where x and y are the prey and predator biomasses, respectively[2], and b, c, m, and r are

---

[2] We can use the variables to represent the number of individuals, but the models are generally more correct if we think of the variables as biomass.

positive parameters. The narrative version of the model is not generally stated, but we can discover it by examining the model in detail. First, note that there are three processes that contribute to the population change. The term rx represents the growth of the prey, while the term my represents the mortality of predators, with the parameter m taking larger values if people kill predators. The term bxy represents the rate at which prey biomass is lost through the interaction between species, and the extra factor $c < 1$ in the predator equation indicates the efficiency with which the predator converts prey biomass into its own biomass. We need not concern ourselves with the narrative implied by the algebraic forms bxy and my, although these are also subject to criticism. Our primary interest is in what the model says about the prey population growth, which is that there are no environmental restrictions on the relative growth rate r. The predation process is the only restriction on the growth of prey biomass because it is the only negative term in the prey equation.

We now turn to the characterization of the model. There are equilibria at $(0, 0)$ and $(m/cb, r/b)$. This alone is sufficient to point out a critical flaw in the model. Any reasonable predator-prey model should allow for the possibility of long-term prey survival coupled with predator extinction. Since predator absence reduces the model to one dimension, we require a model that has a stable equilibrium solution with $y = 0$ and $x > 0$, and there is no equilibrium solution with these properties. The model as it currently stands fails to pass this reality check; hence, we return to the conceptualization step.

The key to finding the flaw in the Lotka-Volterra model is to examine the reduced model corresponding to no predators. It is

$$\frac{dx}{dt} = rx. \tag{9}$$

The Lotka-Volterra model predicts unbounded prey growth in the absence of predators. Thus, if we try to eliminate predation, the prey population goes through the roof. That in turn makes the predator population viable because an extremely large value of x counters the large value of m. We can fix the error by stipulating a conceptual model that places a restriction on the prey population when predators are absent. One such model is given by

$$\frac{dx}{dt} = rx\left(1 - \frac{x}{K}\right) - bxy, \qquad \frac{dy}{dt} = cbxy - my. \tag{10}$$

In the absence of predators, the prey equation becomes

$$\frac{dx}{dt} = rx\left(1 - \frac{x}{K}\right). \tag{11}$$

The parameter K represents the largest prey biomass that the environment can support, because the prey population decreases if $x > K$.

We can now characterize this new model. We find equilibria

$$(0, 0), \qquad (K, 0), \qquad \left(\frac{m}{cb}, \frac{r}{b}\left[1 - \frac{m}{cbK}\right]\right), \tag{12}$$

with the last of these relevant only if $m < cbK$. Further analysis shows that this mutual survival equilibrium is asymptotically stable whenever it exists, and the predator extinction equilibrium $(K, 0)$ is asymptotically stable whenever $m > cbK$. In biological terms, the model makes the following prediction:

- Hunting of predators causes the predator population to decrease and the prey population to increase, up to the point where $m = cbK$; if m is made larger than that, the prey population will reach the environmental capacity while the predator population disappears.

This prediction is qualitatively compatible with observation. We can test it further by measuring the populations and parameter values for some specific predator-prey community.

BUGBOX-POPULATION: A TOOL FOR TEACHING MATHEMATICAL MODELING

It is not terribly difficult to teach students *about* mathematical modeling; that is exactly what I've been doing up to this point. Teaching students to *do* mathematical modeling is another matter entirely. Mathematical modeling is a skill that needs to be learned by experience. You can teach a person to play the piano, but you have to have a piano for them to use while they are learning. Similarly, you can teach mathematical modeling, but you have to have a way to provide students with the experience of modeling. This works best if the modeling is done in the larger context of science. Modeling is more meaningful to students when they create models based on their own observations and then discover the flaws in their models by doing their own verification. Since modeling is initially difficult, it helps if the setting to be modeled can be made progressively more challenging, starting at a very simple level. Finally, it helps if the mathematics is simple so that the students can focus on the modeling process without getting too stuck on the characterization step. This difficult set of requirements is met by a software tool called Bugbox-population. I wrote this tool using Python [10], a public domain programming language, with the auxiliary packages Pygame [9] and Pgu [8]; it can now be freely downloaded from my web page as a windows executable [6]. A similar tool, Bugbox-predator [7], is a virtual laboratory experiment to provide motivation for the derivation of the Holling type II functional response.

A real species in a virtual world

Bugbox-population was originally created for use in a course called *Research Skills for Theoretical Ecology*, which is a component of ***RUTE***[3], an interdisciplinary research program at the University of Nebraska-Lincoln. Students in the course have limited college experience or are new high school graduates. The course assumes no mathematics background beyond precalculus and no specific biology background. The students have been successful in high school or college, come with recommendations from teachers, and are highly motivated. The course focuses on the general subject of biological pest control. We study the population dynamics of aphids and also predator-prey interactions between ladybird beetles and aphids. The lecture and laboratory run concurrently. This poses a challenge, since we want the mathematics to be driven by the laboratory work. The solution to this difficulty is to use virtual laboratory work to motivate the mathematics.

---

[3] Research for Undergraduates in Theoretical Ecology

Aphids have four juvenile stages and an adult stage, all of which look the same except for minor size differences. During the summer, they live in colonies of females and give birth to live young. It takes eight to ten days for aphids to progress from newborn to adult. This makes for population growth that is quite rapid by biological standards, but it still takes four weeks to collect enough data for mathematical modeling. Once an aphid settles on a leaf or stem, it is likely to spend its entire life without any further movement; however, it is still difficult to count aphids because of their small size and high density and their tendency to cluster on the curved stem of a plant.

Bugbox-population displays a virtual world inhabited by a single species of insects called boxbugs. Like aphids, boxbugs are all female and give birth to live young. Their environment is two-dimensional, and they do not move for any reason. A day in bugbox time can take as much or as little clock time as the observer wants. Like ladybird beetles, boxbugs progress from larva to pupa to adult, with each stage distinctive in appearance. Boxbugs come in four varieties, with each successive variety having a slightly more complicated life history than the previous one. This combination of characteristics makes boxbugs the perfect species for population dynamics studies.

Modeling boxbug populations

Students are given a limited number of facts about boxbug biology:

- Boxbugs progress from larva to pupa to adult.
- Boxbugs do not move.
- Larva are produced by adult boxbugs and inhabit a location adjacent to that which is/was inhabited by the mother.

Beyond these elementary facts, students must use a model to describe their observations. Bugbox-population makes this convenient. When Bugbox-population is loaded, the student chooses one of the four species by marking a radio button. Another button begins the simulation by showing the initial state of the bugbox. Thereafter, the student clicks a button to advance the bugbox one time unit as many times as is desired. The Bugbox-population display (Figure 4) shows two consecutive states of the bugbox: the current state is on the right and the previous state is on the left. At each update, the software copies the state of the bugbox from the right picture to the left, calculates the new state, and displays the new state on the right. The student must study the changes to determine the life history details for the given species.

Figure 4 shows the first update in an experiment using species four. Because of the three basic principles of boxbug biology, we can track the development of individuals from the time 0 picture on the left to the time 1 picture on the right. Of the four larvae, two became pupae, one remained a larva, and one died. Both pupae became adults. One of the adults survived, while the other two died. The three adults had a total of nine offspring. This particular trial was chosen for Figure 4 because it includes all possible transitions.

Figure 4: The Bugbox-population display

Species one has a much simpler life history. In each time step, all larvae become pupae, all pupae become adults, and all adults produce new larvae and die. We can therefore write down equations to exactly determine the number of pupae and adults at the next time step, given the current populations:

$$P_{t+1} = L_t, \qquad A_{t+1} = P_t, \tag{13}$$

where L_t, P_t, and A_t are the larvae, pupae, and adult populations at time t. In this highly simplified case, these equations are not "models" because they are exactly correct. The only aspect of species one that is not automatic is that the number of offspring is not the same for each adult. This is not an easy point for students to grasp. They need to advance through enough time steps to realize that they can foretell the number of pupae and adults at each time step from the numbers of larvae and pupae at the previous time step; however, they can not accurately predict the population several steps into the future because the unpredictable results of reproduction carry forward in time. We can't exactly calculate the number of pupae at time 2 if we can't first calculate the number of larvae at time 1 from the known populations at time 0.

Unlike the development from larva to pupa and pupa to adult, the pattern of offspring production is stochastic, so there is no deterministic formula that can calculate the number of larvae at time $t + 1$ from the number of adults at time t. Nevertheless, we can attempt to construct a deterministic model, just as we can construct a deterministic model for the stochastic process of radioactive decay.

So far, boxbug population dynamics for species 1 has been described *exactly*. At this point, we need to *model* the mechanism of offspring production. In the absence of actual knowledge about boxbug fertility, a reasonable first guess is that there is some average number of offspring per adult. With this conceptual model, we have the mathematical statement

$$L_{t+1} = fA_t, \tag{14}$$

where f is a parameter that represents the average number of offspring per adult per time step. This model cannot be strictly true, but it might still yield good predictions in some circumstances.

Equations 13 and 14 are the result of the conceptualization step for species 1. At this point, we could recast the model in matrix form and do an eigenvalue analysis. My students do not have the background for this yet, so we postpone the characterization step of Figure 3. Instead, I have the students parameterize the model by collecting enough data to estimate the fecundity f. Then I have the students write a Matlab simulation[4] using the species 1 model with their value of f. They can use this simulation to predict boxbug populations into the future. They can then run several experiments to test their predictions. I generally ask them to run their simulation for just 8 steps because I don't want to go so far that environmental limits significantly affect the population size.

Before trying to teach more modeling skills, I prefer to give the students more practice with conceptualization of linear models. I have them work their way up from species 1 through the intermediate species to species 4. Each successive species adds another level of complexity to the model. The model for species 4,

$$\begin{aligned} L_{t+1} &= sL_t + fA_t, & (15)\\ P_{t+1} &= pL_t, & (16)\\ A_{t+1} &= P_t + aA_t, & (17) \end{aligned}$$

is as complex as the model they will eventually use for aphid population dynamics, except that it has only three stages rather than five or six[5]. From this point, we can work on learning how to characterize a matrix model. If desired, we can also work on improving the conceptualization with a nonlinear model. We discuss these issues separately.

Characterization of the linear boxbug model

Most of my students have done matrix multiplication in a high school algebra course, but this is the extent of their knowledge of linear algebra. This does not prevent us from a full characterization of the model of Equations 15–17, but it does constrain the methods we can employ. I use what I like to call "directed discovery," in which I ask the students to use a Matlab simulation to study their model in just the right way so that they discover the dominant eigenvalue and its eigenvector. Specifically, I have them plot

1. the relative population increases X_t/X_{t-1}, for each of the three variables L, P, and A, and
2. the ratios L_t/A_t and P_t/A_t

for species 4. A typical plot is illustrated in Figure 5. Figure 5a shows that the ratios X_t/X_{t-1} all approach the same numerical value, which I define to be the "stable growth rate" λ. Figure 5b shows the ratios L_t/A_t and P_t/A_t, each of which also approach specific numerical values, which I denote u and v respectively. Similar graphs can be plotted with different initial conditions, leading to the conclusion that the behavior of the solution, regardless of initial

---

[4] I almost never have even one student with computer programming experience. It takes one hour of class time for us to write this very simple Matlab program, one line at a time with frequent stops for explanation and testing. I don't expect the students to be able to write their own programs, but they do need to understand the program well enough to make the modifications needed for their aphid model.

[5] Returning to Figure 4, note the necessity of examining the mechanisms rather than just counting the numbers. Some larvae in the current population are new, but one of them is a"holdover" from the previous population. Similarly, some of the adults in the current population were pupae in the previous population, but one was already an adult. Based on the minimal information we have, we could estimate $f = 3.00$, $s = 0.25$, $p = 0.50$, and $a = 0.33$. These estimates will change if we collect more data.

conditions, approaches a steady state in which the population increases by a factor of λ each time step and is distributed among the stages in the ratio $u : v : 1$.

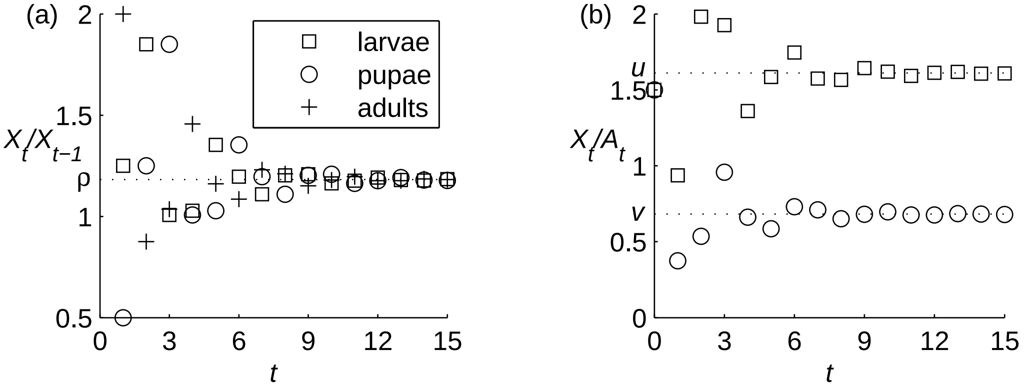

Figure 5: (a) The growth rates X_t/X_{t-1} for the larvae, pupae, and adults, boxbug species 4; (b) The ratios X/A for the larvae and pupae, boxbug species 4

Our task now is to find and solve a mathematical problem for the unknowns λ, u, and v. One way to do this is to consider what happens if we choose the initial populations to be $L_0 = u$, $P_0 = v$, and $A_0 = 1$. With this judicious choice, we have the steady-state solution right from the beginning. Thus, we know that each population grows by a factor λ:

$$L_1 = \lambda L_0 = \lambda u, \qquad P_1 = \lambda P_0 = \lambda v, \qquad A_1 = \lambda A_0 = \lambda. \tag{18}$$

We can also calculate the populations at time 1 by applying the model (15–17) with $t = 0$:

$$L_1 = sL_0 + fA_0 = su + f, \qquad P_1 = pL_0 = pu, \qquad A_1 = P_0 + aA_0 = v + a. \tag{19}$$

Combining equations (18) and (19) yields a system of three equations with unknowns u, v, and λ:

$$su + f = \lambda u, \qquad pu = \lambda v, \qquad v + a = \lambda, \tag{20}$$

or

$$(s - \lambda)u + f = 0, \qquad pu - \lambda v = 0, \qquad v + (a - \lambda) = 0. \tag{21}$$

These are, of course, the scalar equivalents of the equation $(\mathbf{M} - \lambda \mathbf{I})\mathbf{x} = \mathbf{0}$, where $\mathbf{x} = [u\ v\ 1]^T$ and $\mathbf{M}$ is the matrix consisting of the coefficients of equations 15–17.

Students with limited modeling experience have a difficult time keeping in mind that some of the symbols are unknowns and others represent fixed values. It may help to work the problem first with numerical values for the parameters. The equations are linear in u and v, so the simplest plan is to solve for these quantities in terms of λ, thereby obtaining the characteristic equation. We have $v = \lambda - a$ from the third equation and then $pu = \lambda(\lambda - a)$ from the second, whence the first equation becomes

$$pf = (\lambda - s)pu = \lambda(\lambda - a)(\lambda - s), \tag{22}$$

which is one form of the characteristic equation for the matrix $\mathbf{M}$. The function on the right side of this equation is monotone increasing for $\lambda \geq K \equiv \max(s, a)$, which proves that there is a unique solution $\lambda_1 > K$, and this is the value that corresponds to the asymptote in Figure 5a. Note that our form of the characteristic equation is more useful[6] than the standard form because it allows us to prove that there is a unique solution greater than K. Once λ is known, we obtain solutions for u and v, and these correspond to the asymptotes in Figure 5b. In summary, we have determined the stable growth rate and population ratios, which are the dominant eigenvalue and eigenvector respectively, but without the machinery of linear algebra. At this point, I show the students how the same problem is solved using the notation and machinery of linear algebra.

The world of Bugbox-population

Mathematicians reading this discussion may be interested in knowing the actual rules that govern the Bugbox-population world. There actually are parameters p, s, and a that represent the probability of a larva becoming a pupa, a larva surviving without becoming a pupa, and an adult surviving, respectively. The rule governing births is more complicated. The simulation identifies potential birth locations, which are empty spaces adjacent to spaces occupied by an adult in the previous step. Each space is filled with a larva with some probability b or else left empty. The birth rate is thus density-dependent, and the population self-limiting, because the number of potential birth spaces decreases as the population grows.

Building a nonlinear model

Depending on the level of sophistication of the students and the level of mathematical modeling one wants to teach, it is possible to continue working with boxbug models in an attempt to make scientific progress. (We don't do this in my course, partly because the students do not have enough background in analysis and partly because the environmental capacity for real aphid populations is so high that it does not affect population size until the plants have accumulated considerable damage.) The first step is to discover the serious flaw in the boxbug model: the lack of density dependence. If the students run their Matlab simulation for 20 time steps instead of 8, the population will continue to grow exponentially. Of course the real population must level off. The model is qualitatively incorrect, so it now fails the test in Figure 2. At this point, the students will have discovered the need for an improved conceptualization.

As previously noted, it is not possible to make a model that exactly corresponds to the scientific principles of the Bugbox-population world. (How much harder it is for the real world, in which we do not have a human creator who knows exactly the governing principles!) However, students can build density-dependence into their model, at the cost of losing linearity, by replacing the constant f with a function of the total population $N = L + P + A$. Models such as this have been studied in considerable detail. There is a nice discussion in the book by Linda Allen [1], based on a more thorough analysis by Silva

---

[6] This is one of many examples in which doing algebra carefully by hand yields results superior to the unsimplified results obtained with computer algebra systems.

and Hallam [11]. One reasonable idea is to assume a carrying capacity K, changing the L equation to

$$L_{t+1} = sL_t + f\left(1 - \frac{L_t + P_t + A_t}{K}\right) A_t, \tag{23}$$

where f is now the maximum average number of offspring per adult per time step. The resulting nonlinear model can be characterized by (1) finding the equilibrium states and (2) linearizing the model about each equilibrium state to determine stability. The students could estimate K experimentally, use a Matlab simulation to predict the time-dependent populations, and run further experiments to test these predictions. The ultimate result would be a mathematical model that, while not exactly capturing the "true" principles governing Bugbox-population, would describe the population dynamics of boxbugs with useful accuracy.

Conclusions

Even in a highly-simplified human-created virtual world, mathematical models are mathematical objects that are more or less successful at predicting actual scientific results. The connection between models and reality in the real world is more tenuous. This does not mean that models have no value, but it does determine how we must think about mathematical models. The most important task of mathematical modeling is the reality check. Without at least a qualitative confirmation of accuracy, mathematical models have no scientific value. The teaching of mathematical modeling must include a serious discussion of the place of models in science and their connections to the real world. In particular:

> We should always be skeptical about theoretical results that conflict with observation. We may have discovered a new scientific result, but it is far more likely that we are using an inappropriate model.

Acknowledgements

The author wishes to thank Sebastian Schreiber for the suggestion of obtaining the eigenvalue problem directly from the discrete population model, Clint Herron for help with programming issues, Brian J. Winkel for his critical reading of the manuscript, and Rebecca Ledder for her editorial assistance.

REFERENCES

1. Allen, L.J.S. 2007. *An Introduction to Mathematical Biology.* Upper Saddle River, NJ: Pearson Prentice Hall.

2. ATLSS: Across-Trophic-Level System Simulation: An Approach to Analysis of South Florida Ecosystems. Biological Resources Division, United States Geological Survey Technical Report, 1997.

3. Brauer, F. and C. Castillo-Chavez. 2001. *Mathematical Models in Population Biology and Epidemiology.* New York: Springer.

4. Braun, M. 1983. *Differential Equations and Their Applications*, 3rd ed. New York: Springer.

5. DeAngelis, D. L., L. J. Gross, W. F. Wolff, D. M. Fleming, M. P. Nott and E. J. Comiskey. 2000. Individual-based models on the landscape: applications to the Everglades, in J. Sanderson and L. D. Harris (eds.), *Landscape Ecology: A Top-Down Approach*. Boca Raton, FL: Lewis Publishers.

6. Ledder, G. 2007. Bugbox-population, at http://www.math.unl.edu/~gledder/bugbox/.

7. Ledder, G. 2007. Bugbox-predator, at http://www.math.unl.edu/~gledder/bugbox/.

8. Pgu, module for Pygame, at http://www.imitationpickles.org/pgu/.

9. Pygame, module for Python, at http://www.pygame.org.

10. Python, programming language, at http://www.python.org/.

11. Silva, J.A.L. and T.G. Hallam. 1992. Compensation and stability in nonlinear matrix models. *Math. Biosci.* 110: 67–101.

Biographical Sketch

Glenn Ledder is an Associate Professor of Mathematics at the University of Nebraska-Lincoln, where he joined the faculty in 1989. He received his B.S. in Ceramic Engineering from Iowa State University in 1977 and his Ph.D. in Applied Mathematics from Rensselaer Polytechnic Institute in 1990. He is the author of 18 research papers, 4 pedagogical papers, and a differential equations textbook, and has received 3 National Science Foundation grants for undergraduate education. He has extensive experience in mentoring undergraduate research and is co-director of the Research for Undergraduates in Theoretical Ecology (***RUTE***) program at the University of Nebraska-Lincoln.