Qualitative, rule-based modeling

In an African savanna, there are distinct wet and dry seasons each year. In some years, the dry season is relatively short, in others it is a month or two longer. In some years, rainfall during the wet season is low, in other years it is abundant.

Imagine that a study of calf mortality among browsing antelope living in the acacia woodlands indicates that rainfall influences calf mortality in several complex ways related to browse availability and the fat reserves of the mothers. How can one build a model to explore how the findings relate to the population dynamics of the antelope?

The standard population models in the literature are of two forms: logistic models (in terms of a differential or difference equation for the total population) and age-structured Leslie-matrix models (Starfield and Bleloch 1986). An age-structured model might seem promising, because calf survival is one of its parameters, but it would also require data on the survival of the other age classes. If this data is not available for the antelope study, must you abandon the idea of building a model? This article describes a modeling approach that builds on what you know and copes with a lack of information. It is called qualitative, rule-based modeling (Widman et al. 1989). The antelope example used to illustrate this approach is loosely based on observations of kudus by Norman Owen-Smith (described in Starfield and Bleloch 1986).

A simple antelope-population model

The population is driven by rainfall patterns. We begin by looking for the simplest description of those patterns. The dry season, for example, can be characterized as "long" or "short," so we employ those two descriptions instead of the more conventional number of days or months since the last rains. We introduce a qualitative variable, *dryseason*, which represents one of two states. State 1 indicates a short dry season, whereas state 2 indicates a long dry season.

We can similarly characterize the rainfall during the wet season. Instead of a numerical variable, such as inches of rainfall, we introduce the qualitative variable *rainfall*: state 1 is low, state 2 is average, and state 3 is high. On the basis of field experience, we can denote what we mean by these three states.

If we were to specify the variables *rainfall* and *dryseason* for each year; our model should then be able to compute the state of the antelope population. We might let *antelope* represent one of four population states: 1 is low, 2 is medium low, 3 is medium high, and 4 is high.

A conventional model would consist of equations enabling one to predict the antelope population from the rainfall and dry season input data. In a qualitative model, rules replace equations. The rules can be based on research findings. Suppose that the study of calf mortality gave the following results:

- Calves are dropped at the beginning of the wet season. Rainfall in the previous wet season is an important factor because it determines the fat reserves in calf-bearing females at the beginning of the dry season.
- If the dry season preceding birth is long, browse becomes scarce and calf-bearing females are stressed, particularly if their fat reserves are already low.
- Browse available to the calves and their mothers during the wet season of their birth is also important for survival during their first year. Rainfall is strongly correlated with available browse. It follows that rainfall divided by antelope numbers or density is a good measure of browse available to an individual.

We use this information to modify an internal variable (let us call it *recruit*) that balances the factors favoring increases or decreases in the antelope population and accumulates the net effects from year to year. To demonstrate how this process works, we develop a set of rules.

The third result of the study indicates that calf survival will be low if this year's rainfall is low and the antelope population is high. This finding suggests:

Rule 1: If *rainfall* = 1 and *antelope* = 4, then subtract 1 from *recruit*.

The other observations on calf mortality suggest:

Rule 2: If *previous rainfall* = 1 and *dryseason* = 2, then subtract 1 from *recruit*.

Rule 3: If *previous rainfall* = 2 and *antelope* = 4 and *dryseason* = 2, then subtract 1 from *recruit*.

The following rules relate to conditions of high recruitment:

Rule 4: If *rainfall* = 3 and *dryseason* = 1, then add 1 to *recruit*.

Rule 5: If *rainfall* = 2 and *previous rainfall* > 1 and *dryseason* = 1, then add 1 to *recruit*.

We assume that one event reducing recruitment is not sufficient to lower the antelope population. Therefore,

by Anthony M. Starfield

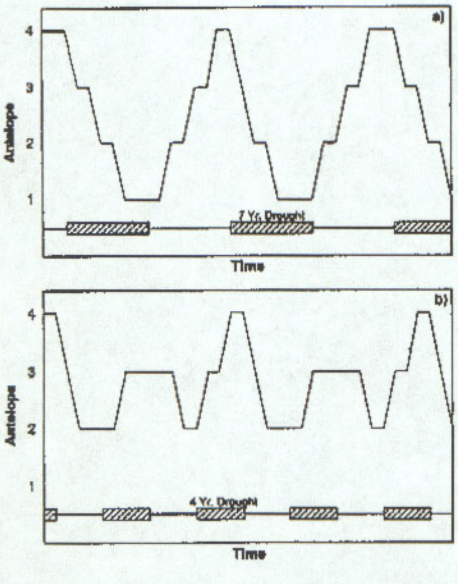

Figure 1. Output from the model, showing how the antelope population responds to cycles of wet and dry years. a. A 14-year cycle. b. An 8-year cycle. Antelope population is represented by four levels: 1 = low, 2 = medium-low, 3 = medium-high, 4 = high.

we update the state of the antelope population with the final rules:

Rule 6: If *recruit* > 1, then add 1 to *antelope* and subtract 2 from *recruit*.

Rule 7: If *recruit* < 1, then subtract 1 from *antelope* and add 2 to *recruit*.

In this way, we reset *recruit* whenever it has become big or small enough to affect the state of the antelope population. Application of the last two rules would, of course, have to be constrained, because *antelope* can never drop below 1 or rise above 4.

Model results. The seven rules are easy to program in almost any computer language. As in a conventional model, we manipulate the inputs (*rainfall* and *dryseason*) and observe the output (*antelope*).

For example, we could explore how the antelope respond to cyclic weather conditions. Suppose seven good years (*rainfall* = 3 and *dryseason* = 1) alternate with seven bad years (*rainfall* = 1 and *dryseason* = 2). Figure 1a shows the predicted state of the antelope population during a period of 35 years. Figure 1b shows what our model predicts if we reduce the 14-year cycle to an 8-year cycle (i.e., if we alternate four good years with four bad years).

Are these results plausible? How do they relate to the dynamics of an actual antelope population? If there are inconsistencies, can we deal with them by making small adjustments to our model, or is the basic structure flawed? We can ask such questions as we explore what the model can tell us and how that relates to reality.

Adding a predator to the system

How easy is it to add another trophic level to this kind of model? Let us populate our savanna with lions. The lions live mainly off large herds of migratory animals that invade the region during the wet season. In years of low rainfall, the migrant herds stay in the region for only a short time. However, the lions do not migrate with the herds and during the dry season are dependent on antelope and other local residents. If the lion population is high and the rainfall low, the lions will reduce the antelope population.

Let us introduce a new variable *lion* with three states: 1 is low, 2 is medium, and 3 is high. We can then insert a new rule:

Rule 5a: If *rainfall* = 1 and *lion* = 3, then subtract 1 from *recruit*.

Occasional predation by lions will not noticeably affect the antelope population. However, the converse may not be true: availability of antelope may be important for the survival of lion cubs, particularly during lean conditions. If we introduce another internal variable *lion-recruit*, we might postulate the following set of rules:

Rule 8: If *rain* = 3 or (*rain* = 2 and *antelope* = 4), then add 1 to *lion-recruit*.

Rule 9: If (*rain* = 1 and *antelope* < 4) or (*rain* = 2 and *antelope* = 1), then subtract 1 from *lion-recruit*.

Rule 10: If *lion-recruit* > 2, then add 1 to *lion* and subtract 3 from *lion-recruit*.

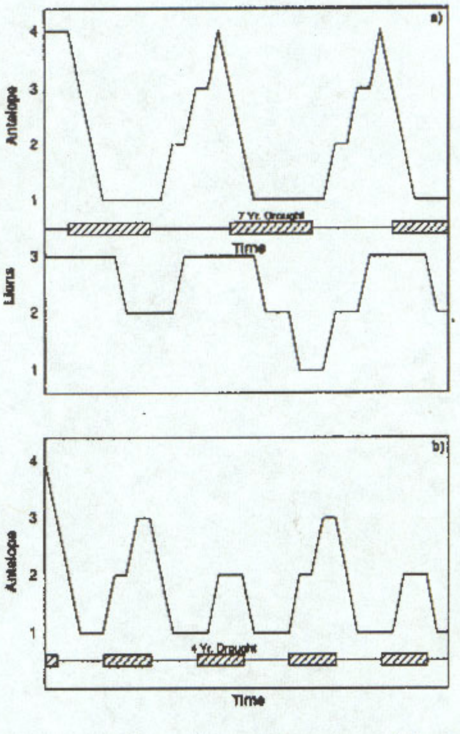

Figure 2. How the antelope population responds to weather cycles in the presence of a predator. a. A 14-year cycle. b. An 8-year cycle. Changes in the predator population are also shown in a. In b, the lion population was not plotted because it remained constant at the maximal level.

Rule 11: If *lion-recruit* < 2, then subtract 1 from *lion* and add 3 to *lion-recruit*.

Notice that rules 10 and 11 suggest that the lions respond more slowly to changing conditions than do the antelope. Figure 2a shows how both the predator and antelope respond to 14-year cycles of good and bad years. Comparing Figures 1a and 2a, we see that the lions have a noticeable but not major impact on the antelope. We also notice that there are large fluctuations in the lion population.

In Figure 2b, we repeat the experiment with 8-year cycles. Here we only show the state of the antelope population. The lion population remained constant and high, because the shorter weather cycles enable the lions to maintain their population throughout the drought. Comparing Figures 2b and 1a, we see that under these conditions the lions have a major impact on the antelope. Our qualitative model is revealing qualitative differences in the behavior of the rainfall-prey-predator system.

• **Stochastic input data.** An alternative way of exercising our model is with stochastic input data. For example, we could specify the probability that rainfall will be low, medium, or high, as well as the probability that the dry season will be short or long. The computer will then generate stochastic weather patterns. Running the model many times, we could use the results to compute the probability of finding the antelope in a given state under that particular rainfall regime. Figure 3 summarizes results of this kind. It is one way of showing, relatively simply, how a variety of different conditions might affect the antelope population.

Rule-based modeling

Rule-based modeling is an outgrowth of developments in artificial intelligence and expert systems, an area that is now being applied to ecology (Rykiel 1989). Starfield and Bleloch (1986) first showed how rules might be used to modify conventional, quantitative models and suggested how qualitative dynamic models could be built. Their ideas were subsequently implemented in a model that predicted how the biota in a large, shallow estuarine lake might respond to changing conditions of salinity and water level (Starfield et al. 1989a).

The example of an estuarine lake has two things in common with the rainfall-antelope-lion model developed here. First, both systems are driven by natural events. Interactions between the components of the system may at times be important, but they are dominated by external driving forces such as weather and water quality. Qualitative models lend themselves to this type of problem.

Second, in both cases there were insufficient data to construct more detailed models. In fact, in the case of the estuarine lake there were almost no formal data, but the rules elicited much informal information on the behavior of the system. Qualitative models have an important role to play in allowing construction of dynamic models that draw on what is known without introducing a host of unknown (perhaps unknowable) parameters.

Rule-based models, therefore, fill a niche. Within that niche, they are much easier to explain and understand than most conventional models. The workings of the model are more accessible, and the assumptions are more explicit. Connections between input and output are easier to trace and explain.

Rule-based models thus promote discussion. They can therefore draw on information from different sources; for example, in our hypothetical model an antelope researcher can easily incorporate information on predation.

The ability of these models to synthesize information, without getting caught up in intricate details, suggests additional uses. By facilitating discussion between ecologists and engineers, a qualitative model might provide an effective vehicle for introducing environmental concerns into the wider context of management and design (Starfield et al. 1989b).

It is now generally recognized that large system models, such as those generated by the Biome projects of the 1960s and 1970s (e.g., Innis 1978) tend to collapse under their own weight. It would be interesting to build much simpler ecosystem models, at a level of resolution consistent with a qualitative approach.

Modeling in the classroom

The rainfall-antelope-lion example is a model that might be introduced to a class of undergraduate or graduate

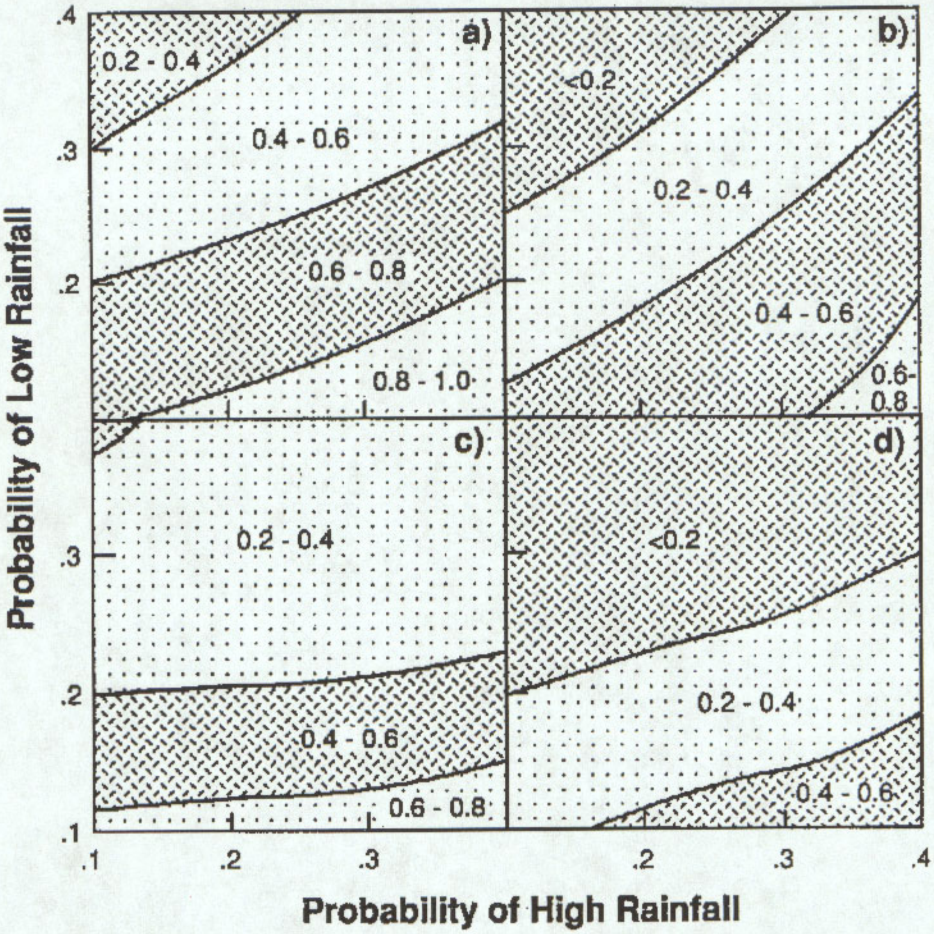

Figure 3. The probability of finding the antelope population in its highest state under different weather scenarios. The graphs may be used to predict for a given area of the world, where the probability of high rainfall and the probability of low rainfall are known, how well an antelope population is likely to do. The graphs differ in a third weather condition—probability of a long dry season—and in whether predators are present. **a.** No predators, 0.3 probability of a long dry season. **b.** No predators, 0.6 probability of a long dry season. **c.** With predators, 0.3 probability of a long dry season. **d.** With predators, 0.6 probability of a long dry season. The graphs show that the presence of predators generally reduces the likelihood of a high antelope population, as does a more frequent, long dry season.

ecology students. It can be used to create an active learning environment (Starfield et al. 1990). It has pedagogic advantages over conventional, numerical, or mathematical models.

This example introduces modeling without mathematics. Often students miss the point of modeling, because they are preoccupied with the mathematics. It is also easy to implement on a computer, so the student is not distracted by computational details. All effort is directed at the modeling process. Ease of implementation can be enhanced even further by developing rule-based modeling shells[1] or compilers that enable students to define their variables and states, build rules, and then test and run the model in a supportive, user-friendly environment.

Rule-based modeling encourages parsimony and a thoughtful choice of variables and resolution. At each stage of the modeling process, explicit decisions have to be made. Why, for example, did we need to introduce two rainfall variables (*rain* and *dryseason*)? Why did we specify *rain* in terms of three states but *dryseason* in terms of only two? Why specify *antelope* in terms of four states rather than three or five or six?

The introduction of each additional variable or state extracts a high price in terms of the complexity of the rule set. The student is forced to walk the fine line between a model that is so simple that it is trivial and one that is so complex that it is useless. The student can learn to find that line adaptively, introducing or deleting additional states or variables as needed.

In making these decisions, the student learns perhaps the most important lesson in modeling: to match the model to its purpose and to the quality of the available information. Modeling is perceived as a purposeful tool rather than a mathematical exercise or a precise attempt to mimic reality.

The simplicity of both the input and output variables helps the student learn the discipline of planning computer experiments and presenting the results in a cogent manner.

Rule sets are easy to modify. Students learn how to ask "what if?" questions and to explore the sensitivity of results to the structure and assumptions of the model.

The device of internal variables (such as *recruit* and *lion-recruit*) can be used to assign to each organism or process its own time scale. Qualitative models are useful for showing how events that proceed at different rates can interact. How, for example, would the results of Figure 1b differ if the predator responded as fast as its prey?

Finally, the language of rule-based models fits the language and thought processes of ecologists. Most students of ecology learn to make hypotheses and build verbal models. Rule-based modeling breathes life into verbal models, allowing the student to see if his or her ideas really do hang together and if the consequences are as expected.

By using a qualitative, rule-based model, the students learn the importance of modeling as a means of communication. Despite its simplicity, they can incorporate good biology, and they do so in an explicit fashion. It is easy to sustain a biological argument about the rules.

The qualitative states, which at first might appear to be purely arbitrary, quickly take on a life of their own. The modeler develops a picture in his or her mind of the difference between a low or medium-low population. The processes of model building and interpretation go hand-in-hand. The student learns to relate the model to the real world in an abstract way and to look for qualitative rather than numerical differences. It follows that rule-based modeling encourages not only good modeling, but also good ecology.

A disadvantage of rule-based modeling is that the student does not learn concepts such as equilibrium, stability, catastrophe, and optimization, which come so neatly out of mathematical modeling. Rule-based modeling is not an alternative to other types of modeling but rather another item in the modeler's tool kit. It is a good tool for the beginner.

There is no need for rule-based models to compete with conventional models. Examples should be chosen (such as succession or systems that are driven by external events) that are appropriate to the technique. When the student has mastered the elements of modeling, it is easy to present a problem that is clumsy in a rule-based form and so motivate the student to move on to other forms of modeling.

Acknowledgments

This work was supported by a grant-in-aid from the Graduate School of the University of Minnesota. The author would like to thank Rick Taylor and John Field for fruitful discussions and Mark Quadling for his design and implementation of a modeling shell.

References cited

Innis, G. L., ed. 1978. *Grassland Simulation Model.* Ecological Studies, Analysis and Synthesis 26. Springer-Verlag, New York.

Rykiel, E. J. Jr, ed. 1989. Artificial intelligence and expert systems in ecology and natural resource management. *Ecol. Modell.* 46: 1–2.

Starfield, A. M., and A. L. Bleloch. 1986. *Building Models for Conservation and Wildlife Management.* Macmillan, New York.

Starfield, A. M., B. P. Farm and R. H. Taylor. 1989a. A rule-based ecological model for the management of an estuarine lake. *Ecol. Modell.* 46: 107–119.

Starfield, A. M., K. A. Smith, and A. L. Bleloch. 1990. *How to Model It: Problem-solving for the Computer Age.* McGraw-Hill, New York.

Starfield, A. M., R. H. Taylor, and L. S. Shen. 1989b. Rule-based models at the interface between engineering and ecology. In *Proceedings of the Sixth Conference on Computing in Civil Engineering.* American Society of Civil Engineers, Washington, DC.

Widman, L. E., K. A. Loparo, and N. R. Nielsen, eds. 1989. *Artificial Intelligence, Simulation and Modeling.* John Wiley & Sons, New York.

Anthony M. Starfield is a professor in the Department of Ecology, Evolution and Behavior at the University of Minnesota, Minneapolis, MN 55455.

[1]A Beta-test version of a rule-based modeling shell (called Rule and developed by M. Quadling, A. M. Starfield, and J. G. Field) is available for DOS (IBM-compatible) computers by writing to J. G. Field, Zoology Department, University of Cape Town, 7700 Rondebosch, South Africa. Enclose a check or bank draft for $25.00 to cover the cost of the disk, packaging, and postage and specify whether you require a 5.25" or 3.5" diskette.

NEW PRODUCTS

Heated stage. The Kitazato Micro-Warm Plate® model MP-10DM microscope stage heater is designed for observing temperature-sensitive specimens. The transparent glass plate is heated by a warming element inside the glass. Temperatures are controlled to 0.1° C of the desired surface temperature. The MP-10DM glass plate is 84×106×1 mm and can be