NATURAL RESOURCE MODELING
Volume 14, Number 3, Fall 2001

RESEARCH NOTES

GETTING "RESULTS": THE PATTERN-ORIENTED APPROACH TO ANALYZING NATURAL SYSTEMS WITH INDIVIDUAL-BASED MODELS

1. Introduction. Individual-based models (IBM's) are increasingly used to simulate natural resources and their response to management and exploitation. Many IBM's represent animal populations, but others (mainly unpublished) simulate how fishing regulations affect fish populations and harvest (including models with fishing boats simulated as individuals), interactions among recreational boaters seeking anchoring sites, interactions among hikers and motorized tourists in wilderness situations, and even the evolution of activist organizations in response to competition for members.

Many of us who conduct research with IBM's have had the experience of demonstrating a model to colleagues and immediately being asked, "But how do you get 'results'?—it's just a stochastic simulator". This response to IBM's is surprisingly common, especially from modelers who use conventional differential equation-based approaches. This question may seem presumptuous but it deserves attention. As the review of Grimm [1999] points out, many IBM's have been built but relatively little has been learned about ecology or natural resource management from these models. We do indeed need ways to develop "results"—general principles and conclusions—from IBM's. This paper presents an approach to using IBM's in a hypothesis-testing framework to draw general conclusions about the characteristics of the individuals and the system-level properties that emerge from individual characteristics.

While the rest of this paper addresses IBM's, it is worth remembering that this "how do you get results" question is just as easily applied to differential equation models. Differential equation models require simplifying assumptions so the equations can be defined or solved. There is a tendency to assume that, because these models represent aggregations of individuals, the model results are applicable to individuals in general. The opposite is often true—a hard look at all the assumptions in such models and the conditions under which these assumptions really apply may indicate that the differential equation model's results are rarely if ever really applicable. My colleagues and I (Railsback et al. [1999]) examined an example of this problem, the "result" of a mathematical derivation (Gilliam and Fraser [1987]) showing the fitness of animals to be maximized when they select habitat to minimize the ratio of mortality risk over energy intake. Upon examination, this "result" was found to be based on a number of assumptions that are rarely met in nature. (One of these assumptions is that no habitat fails to provide positive growth rates; in reality, animals are continuously faced with the challenge of finding adequate food). We should not automatically assume that IBM's are less "general" than are more aggregated models.

One of the most important reasons for using IBM's is to overcome the limitations of conventional modeling approaches, especially their dependence on restrictive and unrealistic assumptions (Huston et al. [1988]). However, to do science with these models we need ways to test hypotheses and reach general conclusions.

2. The pattern-oriented approach.

2.A. Microanalysis of complex systems. Getting "results" from IBM's is the natural resource modeler's version of the fundamental problem in understanding all complex systems: determining the relations between the characteristics of individuals and system-level responses that emerge from these characteristics (Levin [1999, p. 114]). The only thorough, explicit discussion of this issue with which I am familiar is by Auyang [1998]. Auyang rejects both the reductionist and holist approaches to science—examining only the parts or only a whole system is inadequate. Instead, Auyang recommends an approach called *synthetic microanalysis* because it integrates synthesis at the system level with analysis of the individuals that compose the system.

Synthetic microanalysis begins with synthesis of a broad conceptual framework of the system being analyzed. The boundaries of the system to be studied are delineated along with the external driving variables by which the rest of the world affects the otherwise-isolated study system. System-level measures of state are defined; these can be aggregate measures of the state of the system's individuals (e.g., total number or biomass of organisms) or statistical distributions of individual states (e.g., distributions of age and size). Synthesis also includes defining the individuals and their characteristics, using information gathered from smaller-scale analyses like laboratory studies of individual organisms. For the natural resource modeler, this synthesis phase constitutes designing an IBM for the system of interest (often, an animal or plant population and its exploiters); it is highly system-specific and I do not address it further in this paper (but see Railsback [in prep.]).

In the microanalysis phase, *system*-level concepts are posed and the *individual* behaviors and mechanisms ("traits") that explain the system-level concepts are determined. This phase includes determining how, in IBM's, to model individual organisms and their environment so that realistic population-level response patterns emerge from individual traits. Microanalysis provides the mechanistic understanding of the links between individual traits and system responses that allows prediction of how system dynamics respond to factors acting at the individual level, the "results" we are after.

2.B. Pattern-oriented microanalysis. My colleagues and I (e.g., Railsback and Harvey [in prep.]) developed an approach to microanalysis that allows us to test and possibly reject models of individual traits, and therefore infer what characteristics of individuals can produce realistic emergent responses at the system level. The idea for this approach came from the "pattern-oriented" approach to ecological modeling recommended by Grimm et al. [1996]. A similar approach was proposed by Clark and Mangel [2000] for testing models of decision making by individual animals.

Grimm et al. [1996] suggested that ecological models should be designed to address specific patterns observed in nature: modelers should identify important observed patterns and attempt to understand and represent the mechanisms that cause the patterns. This is a way to focus a modeling study, providing guidance on such issues as how aggregated the model should be, what processes to include, and what scales are relevant. For individual-based natural resource modeling, the key benefit of pattern-oriented modeling is that it assures that the model produces predictions suitable for testing: models can be tested by whether they reproduce the patterns they were designed to address.

Pattern-oriented analysis can be applied to natural resource IBM's by using it as a way of identifying and testing hypotheses about how individuals behave. The following steps can be used; these constitute the microanalysis phase defined by Auyang [1998]. An example is provided below.

(1) Define a set of "testing patterns", observed patterns of system-level (or individual) responses to known stimuli that the IBM is designed to explain and reproduce. These patterns may be identified before the model is designed as a way to guide model development. If the model has already been designed, the testing patterns should be explicitly defined before model analysis begins so the modeler cannot be accused of picking only those testing patterns that confirm the model's formulation.

Specific mechanisms in a model can be analyzed and tested by picking test patterns that are all driven by the selected mechanism (as in the following example concerning habitat selection mechanisms in a trout model). This allows individual traits of the individuals in a model to be analyzed separately.

Useful patterns are often easily extracted from existing literature on the system being studied. Picking test patterns driven by a wide range of stimuli allows microanalysis to be more comprehensive and conclusive, but the model must include the mechanisms driving the patterns. Our experience (Railsback and Harvey [in prep.]) shows that even the simplest test patterns can be sufficient to draw interesting conclusions. In fact, the analysis is more likely to be convincing if the test patterns are simple, obvious, and noncontroversial.

(2) Build a model that includes the mechanisms and individual traits believed to drive the testing patterns. The IBM must be designed so that the tested responses emerge from simpler individual behaviors. Modelers must be sure that the test patterns are *emergent responses* of the model and are *not* hardwired in.

(3) Pose alternative formulations for individual traits as hypotheses that will be tested. The hypothesized traits will often be rules for decision making, but can also be alternative formulations for such processes as feeding or mortality risks. It can be interesting to pose the assumptions used in conventional aggregated models as hypotheses for

individual traits.

(4) With the IBM, simulate the conditions under which each test pattern has been observed to occur. Repeat the simulations with each alternative formulation for individual traits. Reject hypothesized formulations that do not cause the test patterns to emerge from the model.

Completion of these steps allows the modeler to at least reject model rules that fail to produce emergence of the test patterns, and, perhaps, identify rules that succeed in producing a wide range of realistic emergent system-level behaviors. These results are likely to allow meaningful and general inferences to be drawn.

3. Example: habitat selection by stream trout. We applied the pattern-oriented approach to test model rules for how fish select habitat among patches varying in survival probability and food intake (Railsback and Harvey [in prep.]). This experiment is summarized here as an example of pattern-oriented model analysis. (Habitat selection is a key component of our model that predicts the effects of dam operations on stream fisheries).

The primary individual behavior simulated in our trout model is movement to select new habitat, in reaction to changes in river flow, temperature, and trout density. These factors affect the spatially explicit food availability (which is affected by hydraulics and competition for food with other trout) and mortality risks. Habitat selection is simulated at a daily time step and spatial scale of a stream reach (200 m length) with patch sizes of several m^2. Habitat selection is simulated by letting each fish move to the habitat patch providing the highest fitness, where fitness is defined as the predicted probability of surviving and growing to reproductive size over a future time horizon. The predicted probability of surviving over the time horizon is a function of food intake as well as predation risks—if a fish does not obtain sufficient food, it risks starving to death within the time horizon (Railsback et al. [1999]).

Our pattern-oriented analysis was designed to evaluate the trout model's ability to simulate habitat selection realistically. The analysis used six patterns of habitat shifts observed in real trout. From the fisheries literature, we identified these patterns:

- Where food availability is highly variable over space, trout exhibit "hierarchical" feeding behavior. The dominant (largest) fish gets the best feeding site, and if the dominant fish is removed the next-dominant fish moves into the best site and other fish move up in the hierarchy.
- Trout respond to flood flows simply by moving to the stream margins where velocities are low. Juvenile trout use higher velocities on average when in competition with another, larger, trout species.
- Juvenile trout use faster and shallower habitat on average in the presence of predatory fish.
- Trout use higher velocities on average when temperatures are higher (metabolic rates increase with temperature, so more food is needed to avoid starvation; food intake is higher at higher velocities).
- When food availability is decreased, trout shift to habitat with higher food intake and higher mortality risks.

We simulated the conditions under which these response patterns have been observed. We posed three alternative rules as hypotheses for how trout select habitat. In addition to our approach (maximizing predicted survival and growth to reproductive size over an upcoming time horizon), we tested (1) maximizing instantaneous growth rate and (2) maximizing instantaneous survival probability.

Simulations with habitat selection to maximize instantaneous growth reproduced three of the six habitat selection patterns, maximizing survival reproduced two patterns, and our "maximize predicted survival and growth" approach reproduced all six patterns. Two patterns (shifts in habitat with temperature and food availability) were not reproduced by the rules that consider only instantaneous growth and risk, but were explained by our rule that simulates prediction of how future starvation risk depends on current energy reserves and food intake. Many models of habitat selection by animals have been based on instantaneous growth or risk, but our experiment provides strong evidence that these conventional models cannot explain some observed behaviors. Our pattern-oriented microanalysis provided evidence that realistic models of animal behavior need to simulate how animals consider the future consequences of decisions. (This result is no surprise to those familiar with complex systems research—prediction-based decision making is a key trait of models in which complex, lifelike behavior emerges from adaptive agents; Railsback [in prep.])

This experiment illustrated that we can use IBM's and pattern-oriented analysis to reject some hypothesized model mechanisms and to identify individual traits that produce realistic emergent population-level responses.

4. An unattractive alternative: testing against response magnitudes. Although model validation is often assumed to require the comparison of model output to observed population data, such comparisons are generally not useful for developing general results about traits of individuals and their links to population responses.

A key advantage of the pattern-oriented approach is that it allows testing of a model against *patterns* of observed system responses instead of against the *magnitude* of observed responses. In other words, it allows us to test a model without having to go through the expensive, tedious, and sometimes impossible steps necessary to attempt simulation of the exact conditions under which observed data were collected, and allows us to avoid the statistical pitfalls of comparing model results to observations. The following are reasons why testing IBM's by direct comparison to the magnitude of observed system responses is an unattractive alternative to pattern-oriented analysis.

First, it is often very difficult to simulate a system precisely enough to expect response magnitudes to be testable, but it can be relatively simple and easy to simulate the conditions under which a well-understood response pattern should emerge. Testing a model against response magnitudes requires (1) providing input accurately representing the magnitude of all the forces driving the response and (2) correct estimation of the parameters linking the driving forces to the individual response. Such quantitative information is difficult to obtain with sufficient completeness and accuracy. Estimating parameters typically involves calibration, and data used for calibration are no longer useful for testing against model results. As opposed to testing the magnitude of responses, the pattern-oriented approach allows us to test an IBM even when input data and parameter values are relatively limited and uncertain. The focus remains on getting the mechanisms right instead of on unnecessary detail and calibration. Pattern-oriented model analysis can proceed before expensive data collection and input processing is required.

Second, identifying a comprehensive range of independent tests is much more likely with the pattern-oriented approach than with comparisons to observed response magnitudes. A wide range of system or individual response patterns are often readily obtained from the literature, along with sufficient understanding of the stimuli causing the pattern, to allow the pattern to be reproduced in an IBM. On the other hand, the difficulty and expense of testing an IBM's ability to reproduce the magnitude of system responses make it unlikely that a diverse and comprehensive set of such tests could be performed.

Third, and perhaps most importantly, it is difficult to make inferences about specific parts of an IBM's formulation by comparing model predictions to observed population data. If the model fails to reproduce observed population data, it may not be clear whether this failure resulted from problems with one or more of the many parts of the model's formulation, or from poor parameter values. On the other hand, if the model does reproduce observed responses, then nothing is learned about how well its formulation represents individual traits and emergent responses: one certainly cannot assume that all parts of a complex model are correct if it reproduces a small number of observed data sets. Calibration is also likely to mask the success or failure of individual model components: in models as complex as IBM's, calibration can lead to apparent model success at reproducing population observations whether or not there are flaws in any one key component. Pattern-oriented analysis can thoroughly test one model mechanism at a time; without such focused tests there is likely to be insufficient information developed for concluding whether any one part of the model does or does not work well.

Finally, comparing model results to observed responses can lead the unwary into several statistical pitfalls. Conventional hypothesis-testing statistics must be used cautiously in testing IBM's. One problem is that statistical comparisons are inappropriate between data sets where uncertainty results from different sources. Results from replicate IBM simulations (generated with input unchanged except for the random number sequence) have variability due only to the model's stochastic components, which is not comparable to the variability typically observed in natural systems, which also results from external driving forces and initial conditions. Statistical comparisons (e.g., t-tests) are useful for comparisons among results of different IBM simulations, but rarely valid for comparing IBM results to observed data. Another problem is that IBM's can produce output with arbitrary and high sample sizes that exaggerate statistical significance. This problem is important when comparing the distribution of some individual characteristic (e.g., size or selected habitat) between model individuals and real organisms. We can easily observe each of a model's thousands of individuals, so sample sizes can be very large. Virtually any difference becomes statistically significant with such sample sizes, even differences that are clearly not biologically meaningful. The pattern-oriented approach avoids these problems by eliminating the need to statistically quantify differences between observed and simulated populations.

5. Conclusions. A key criticism of individual-based modeling is that IBM's have rarely led to improved understanding of basic ecological and resource management issues. However, general conclusions and results can be obtained from individual-based simulations when the simulations are designed to test alternative model assumptions against observed patterns of system-level response. Comparing alternative models of individual traits by their ability to cause emergence of realistic system-level behavior can be a powerful way to infer the mechanisms by which individuals respond to each other and their environment. Testing models against patterns of response instead of against the magnitude of observed responses allows model mechanisms to be tested comprehensively with a reasonable level of effort and expense. Pattern-oriented analysis of IBM's can lead to understanding of how system dynamics emerge from the traits of individual agents, one of the most important problems in ecology and management.

Acknowledgments. This work was supported by EPRI, the Electric Power Research Institute, under agreement WO6953-02 with Lang, Railsback & Associates, and by the USDA Forest Service, Redwood Sciences Laboratory, under research joint venture agreement no. PSW-99-007-RJVA with Lang, Railsback & Associates. Collaborators in developing the trout model and testing scenarios were Walt Duffy (California Cooperative Fisheries Research Unit), Roland Lamberson (Department of Mathematics, Humboldt State University) and especially Bret Harvey (Redwood Sciences Laboratory). Software support was provided by Steve Jackson, Jackson Scientific Computing, McKinleyville, California and by Glen Ropella and Chris Langton, The Swarm Corporation, Santa Fe, New Mexico.

This paper is largely the product of the author's interactions with the complex adaptive systems' modeling community associated with the *Swarm* agent-based simulation package and the Santa Fe Institute. An earlier version of this paper appeared in the proceedings of SwarmFest 2000, the *Swarm* users conference, Logan, Utah, March 11–13th. We thank the Swarm Development Group (Santa Fe, New Mexico; www.swarm.org) for providing the tools and climate that facilitated these ideas.

REFERENCES

S.Y. Auyang [1998], *Foundations of Complex-System Theories in Economics, Evolutionary Biology, and Statistical Physics*, Cambridge Univ. Press, New York.

C.W. Clark and M. Mangel [2000], *Dynamic State Variable Models in Ecology*, Oxford Univ. Press, New York.

J.F. Gilliam and D.F. Fraser [1987], *Habitat Selection under Predation Hazard: Test of a Model with Foraging Minnows*, Ecology **68**, 1856–1862.

V. Grimm [1999], *Ten Years of Individual-Based Modelling in Ecology: What Have We Learned and What Could We Learn in the Future?*, Ecological Modelling **115**, 129–148.

V. Grimm, K. Frank, F. Jeltsch, R. Brandl, J. Uchmanski and C. Wissel [1996], *Pattern-Oriented Modelling in Population Ecology*, The Science of the Total Environment **183**, 151–166.

J.H. Holland [1995], *Hidden Order: How Adaptation Builds Complexity*, Perseus Books, Reading, Massachusetts.

M. Huston, D. DeAngelis and W. Post [1988], *New Computer Models Unify Ecological Theory*, BioScience **38**, 682–691.

S.A. Levin [1999], *Fragile Dominion: Complexity and the Commons*, Helix Books, Reading, Massachusetts.

S.F. Railsback [in prep.], *Concepts from Complex Adaptive Systems as a Framework for Individual-Based Modelling*, Ecological Modelling, submitted.

S.F. Railsback and B.C. Harvey [in prep.], *Comparison of Salmonid Habitat Selection Objectives in an Individual-Based Model*, Ecology, submitted.

S.F. Railsback, R.H. Lamberson, B.C. Harvey and W.E. Duffy [1999], *Movement Rules for Spatially Explicit Individual-Based Models of Stream Fish*, Ecological Modelling **123**, 73–89.

Steven F. Railsback
Lang, Railsback & Associates
250 California Avenue
Arcata, CA 95521, USA
E-mail: LRA@Northcoast.com