Opinion TRENDS in Ecology and Evolution Vol.19 No.10 October 2004

Simulation as experiment: a philosophical reassessment for biological modeling

Steven L. Peck

Department of Integrative Biology, Brigham Young University, Provo, UT 84602, USA

Some scientific modelers suggest that complex simulation models that mimic biological processes should have a limited place in ecological and evolutionary studies. However, complex simulation models can have a role that is different from that of simpler models that are designed to be fit to data. Simulation can be viewed as another kind of experimental system and should be analyzed as such. Here, I argue that current discussions in the philosophy of science and in the physical sciences fields about the use of simulation as an experimental system have important implications for biology, especially complex sciences such as evolution and ecology. Simulation models can be used to mimic complex systems, but unlike nature, can be manipulated in ways that would be impossible, too costly or unethical to do in natural systems. Simulation can add to theory development and testing, can offer hypotheses about the way the world works and can give guidance as to which data are most important to gather experimentally.

Simulation modeling has become vital for studying a range of complex systems, from sociology to the hard physical sciences, such as physics, astronomy and meteorology. Simulation modeling is enabled by recent advances in computer technology and has been applied in several different ways (Box 1). Here, I focus on complex computer simulations that are designed to mimic biological processes by creating a 'computer-world' to represent those processes. I argue that this kind of simulation modeling is the most effective way to study complex biological systems when simple models cannot capture the necessary complexity, and experiments are impossible owing to logistic, ethical or budget constraints.

Simulation has attracted the attention of philosophers and practitioners of science [1–3]. However, ironically, its utility is debated [3–9] and some ecologists and evolutionary biologists view it with suspicion and even contempt. For example, Oreskes *et al.* complain that numerical models are often misinterpreted and that attempts to verify, validate and confirm them are problematic [10]. Gavrilets condemns simulation modeling as being too specific, difficult to parameterize, difficult to draw

Box 1. What is simulation?

Winsberg identifies three views of simulation modeling: (i) it is used to solve analytic equations computationally; (ii) it is a new science; and (iii) it is used to capture and mimic real-world systems that, unlike real-world systems, can then be experimented upon [33]. These views are not mutually exclusive, but each gives insight into how simulations can be used.

Solving analytical equations and exploring statistical properties

The first view considers simulation to be a tool for using computational methods to solve complex sets of analytic equations; for example, when a system of differential equations is solved with numerical approximations using a computer. This approach is also used frequently within the field of statistics, in which Monte Carlo simulations are used to explore distributional properties of statistical models and formulations; for example, this type of simulation is used in phylogenetic tree reconstruction [38]. The distinguishing characteristic of this view of simulation is that it is used to explore analytic mathematical models or formal statistical models.

A new science

The second view looks at simulation as a new type of science that stands somewhere between experimental methods and purely theoretical analytic models. This view is typified by Wolfram, who calls cellular automata (a type of simulation modeling) a new kind of science [2]. In this view, simulation is worthy of study in its own right, where new methods must be developed and new ways of looking at models must be considered.

Capture and mimic real-world systems

In the third view, simulations are an attempt to mimic a real-world empirical system. In this view, simulation is seen as the creation of a possible world that is constructed *in silico* [39] using computer programs to represent the processes under consideration. These models seek to represent formally relevant aspects of the real system under investigation (e.g. the flow of energy through trophic levels, the effect of spatial substructuring on population genetics, etc.). This type of modeling has become the *sine qua non* for understanding complex systems and has been used successfully in developmental biology [3], astrophysics [40], physics [41], geomorphology [42], meteorology [43] and a host of other disciplines, including evolution and ecology. Agent-based models, in which individuals interact dynamically with each other as structural elements in the model world, exemplify this view of simulation modeling [30,44]. These types of simulation should be viewed as another kind of experimental system. The same sorts of experiment and manipulation that might be done in a real system are done using a computer, with the difference being that these experimental systems can be manipulated with ease in a way that real systems cannot.

Corresponding author: Steven L. Peck (steven_peck@byu.edu).
Available online 13 August 2004

www.sciencedirect.com 0169-5347/$ - see front matter doi:10.1016/j.tree.2004.07.019

generalizations from, open to wide interpretation and difficult to reproduce [11]. Barryman finds them too expensive to build, too complex to understand, and filled with redundancies [4]. Yet, in spite of these concerns, simulation can play an important role in furthering our understanding of complex sciences such as ecology and evolutionary biology.

What are the legitimate uses for simulations models? Philosophers and practitioners of science are recognizing that simulation models are a new kind of tool that defies the categories, uses and restrictions found in the historical uses of mathematical models [3,12–14]. Entire disciplines are being structured around simulation modeling (e.g. artificial life, in which digital 'creatures' are allowed to compete, reproduce and evolve in an computer 'environ-environment' [15,16]). A shortcoming in the views expressed by those concerned about using complex simulation models is that they often bundle three activities (data collection, statistical analysis and mathematical modeling) and only allow each of these activities to support the others (i.e. models are only valuable in light of data and their subsequent analysis). Are such concerns legitimate? To view simulation and its uses more aptly, we must first understand the philosophy behind current uses of modeling and place simulation models in that context. I approach this by looking at how models are used to represent reality, and then by discussing two programs of model building (models used to describe and those used to explain) I then explore the place of simulation in this context as a type of experimental system.

Models: representing reality

Models can be used to represent some aspect of the world, some aspect of our theories about the world, or both simultaneously. Hence, the representative power of a model lies in its ability to teach us something about the thing it represents. Models then mediate between the real world and our theories and suppositions about that world [14]. By understanding and manipulating the model, researchers hope to gain some sense of the way in which the world works. Models are a formalization of theory (i.e. models are theories) and I use the two terms interchangeably throughout this article [17,18].

Two broad programs in biological model building are often confused and conflated, in part because they are conflated. Ginzburg and Jensen characterize the difference between these programs as models that describe versus those that explain [17]. Much of the confusion about the role of simulation results from these two programs not being clearly recognized.

In the first program, statistical models are used to make predictions or find functional relationships among variables. In these models, data are used to fit the parameters used in the model [19]. Usually with these models, the fit is achieved through maximum likelihood, least squares, time-series analyses, or other methods of parameter fitting. In these types of model, it is not the causal story that is being modeled, but rather statistical relationships that are thought to hold among the individual components of the biological system. For example, in Hubble's unified neutral theory of biodiversity and biogeography, the predictive ability of the model is very high even though there is no clear causal explanation for why the statistical models proposed should hold [20,21].

In the second program, models are used to offer an explanation of biological phenomena [17,22]. In these types of model, the terms used mean something specific about the biology of the system. The model offers a theory about how the components of the system (be they biotic or abiotic) work causally together to produce a given outcome. These models can also be manipulated to make predictions by using biological data to fit model parameter values via the same techniques previously suggested: maximum likelihood, least squares, and so on [23]. Parsimony techniques can then be used to refine these predictions to find the best fit to the data [17,24] (Examples of these kinds of model illustrating how they are used within these programs are given in Box 2). In addition to prediction, these models can be used to test how well a particular theory, expressed as a model, fits the data, which gives an indication of how well the theory might describe underlying biological processes. Therefore, it is possible to test specific theoretical ideas about how the world works.

Models can be used as tools to gain a deeper understanding about the biological processes that are being explored. Cooper provides three ways in which models can be considered as tools [18]. First, he points out that models are always simplifications of real-world systems. Second, he notes that another use of theories is that, rather than seeing a particular model as providing corroborating evidence for a particular causal story, the model constrains the set of possibilities. In this view, modeling limits the search space of competing hypotheses being used to explain a particular data set. Third, Cooper highlights the fact that theories can be used as a framework to structure empirical investigations. Theory can guide which data to gather and will help inform researchers about which experiments to conduct. Examples of this can be seen in metapopulation dynamic theory [25], island biogeography [26] and Fisher's fundamental theorem in genetics [27]. In these examples, theory is used to determine which variables are likely to provide the most insight into the question being asked.

Briefly, models have legitimate uses other than fitting parameter values to the models from data. Perhaps one of the best-known examples is the Lotka–Volterra model in ecology. Although it has been demonstrated that few, if any, data sets fit it [28], it has provided a rich framework for further theoretical and empirical work about predator–prey interactions.

Within the framework of this second program, where models are seen as theoretical tools for understanding the world, where does simulation fit?

Model complexity: what is at stake?

Model building plays itself out with at least two competing goals: (i) fidelity to actual biological structure; and (ii) the need to simplify the system to represent it as a model, that is, the level of abstraction [18]. The complexity of ecological and evolutionary systems is profound. In

Box 2. Examples of model types

There are two modeling programs in ecological and evolutionary modeling. The first uses statistical models to fit data gathered in field studies and designed experiments to understand the statistical relationships found among several variables. In the second program, model variables are designed to be a causal explanation of system behavior. Models within this second program can be used to confront data in a similar manner to those from the first program [17,23]; however, complex simulations of this type must be viewed as another type of experimental system to exploit the gains that come from these computer models.

Statistical predictive models

To predict the emergence of pink bollworm *Pectinophora gossypiella*, Carrière *et al.* used a least squares regression (Eqn I, [45]):

$$Y = b + a_1 x + a_2 x^2 + a_3 x^2 + \varepsilon \qquad \text{(Eqn I)}$$

where x is a combined measure of time and temperature and Y is the cumulative emergence. The variables a_1, a_2 and a_3 are the parameters to be fit and ε is the error term. When fit to the data, the model returned an $R^2 = 0.94$, indicating a good fit. These types of model can be more sophisticated and can include time dimensions that use methods from time-series analysis. There are no causal mechanisms that explain the possible biological relationships in this model.

Phenomenological models: two extremes

Alstad and Andow developed a deterministic model of the development of insect resistance to transgenic crops with p proportion of resistant genes and w the proportion of susceptible genes (Eqn II):

$$\frac{dp}{dt} = \left[\frac{(1-r)pX + \dfrac{srpX}{s+G} + \dfrac{srwY}{s+G}}{\dfrac{dX}{dt}} \right] \qquad \text{(Eqn II)}$$

$$\frac{dw}{dt} = \left[\frac{(1-r)wY + \dfrac{GrwY}{s+G} + \dfrac{GrpX}{s+G}}{\dfrac{dY}{dt}} \right]$$

where r is the proportion of the population that moves from the site in which it was born; and s is a preference factor; G is a factor that gives the proportional size of toxic to non-toxic regions. Similar equations for the X and Y (the adult densities per unity area) were included in the original paper [46]. This model follows the resistant and susceptible alleles through time, and provides a biological explanation of how resistance develops. The model is tractable and data could be used to fit model behavior. Because the model is a differential equation, well defined techniques can be used to study the behavior of this model.

Complex simulation model

Storer *et al.* built a spatially explicit model written in the computer language C++ exploring the evolution of resistance in corn earworm *Helicoverpa zea* to transgenic corn. The model had 15 parameters, including population dynamics, genetics and several management variables. This model requires an experimental approach to explore the details and implications of this high dimensional model [47,48].

addition, there might be no general laws to be found in such systems [18,22] and local conditions might dominate [29–31]. To explore this complexity using theoretical modeling tools, choices must be made about what to include in the model. Deciding how much biological detail to add, and how much abstraction to accept, involves tradeoffs. For example, analytical models, such as a set of differential equations, include tools that have been developed over the past two centuries that make these models much more tractable, interpretable and understandable. However, the complexity that these models can handle is comparatively limited. Simple models are unlikely to capture the complexity that is inherent in real-world systems, where, as Cartwright suggests, local influences dominate [32]. This is especially true in biological systems, where ecological, genetic and evolutionary history can be tangled in a complex web of spatial and temporal interaction, causation and stochasticity. However, simulation models can handle almost unlimited complexity, but are hard to understand and explore, in some cases almost reaching the complexity of nature itself. Are there ways to handle this level of model complexity? Why is it important that we do so?

Simulation as experiment

The world is complex and we need all the tools that we can muster to understand it. Researchers across many disciplines are realizing that simulation models are a new kind of experimental system [1,3,13,33]. Simulations are becoming increasingly important in the physical sciences, including physics, astronomy, geology and meteorology. The use of simulation in these disciplines falls somewhere between traditional modeling formulations and experimental systems [1] (All of these sciences are more explanatory than predictive, a feature of many of the questions asked in evolution and ecology).

Simulation models are properly explored using the same experimental and statistical techniques that are used to explore real-world systems. As Winsberg notes, 'If simulationists want to learn about the general qualitative features of a class of systems, then they must apply all the usual tools of experimental science for analysing data: visualization, statistics, data mining, etc. If they want to discover functional dependencies, then they must also run a barrage of trials, looking at the results across a wide range of parameters. It is without a doubt this aspect of simulation that carries the most obvious methodological characteristics of experimental work' [33].

Simulations are experimental systems. Their complexity can make them closer cousins in complexity to nature itself than to simple analytic models, but with a powerful advantage over the real world: the modeler has complete control of the system. Thus, the advantage that simulation gives to scientific exploration is that the model system is strongly manipulable. If one has a good representation of a system, then one has created a world over which one has complete control. It enables experimentation that would be impossible, too costly, too time consuming, or unethical to do with a real system. Simulation models have proven particularly important in understanding spatial and stochastic dynamics in evolutionary and ecological systems, where simple models cannot capture the complexity of these high dimensional systems [34,35]. The insight that simulations are another kind of experimental system is an important restructuring of simulation modeling philosophy. Similar to simple mathematical models, simulations are theoretical constructs created to understand real systems. However, to interpret them requires that they are treated more like experimental systems, unlike simple analytic models, for which the functional relationships among model components can be read off the equations themselves [33].

Simulation modelers have been instinctively looking at simulation as experimental systems ever since the 'dawn' of computers during the 1950s, analyzing their data using techniques borrowed from statistical experimental design. However, many of their efforts have gone into making the square peg of simulation fit into the round hole of analytic modeling with limited success; hence, the criticisms of these kinds of model when simulation is compared with analytical modeling. Adopting the stance of simulation as experiment, currently being championed by philosophers and practitioners of science in the physical sciences, will help clarify the role that simulations can play in advancing ecology and evolutionary biology.

Building useful simulations

When a researcher builds a simulation model, they have created a world in which they have access to all of the laws and components of that world, and the relationships among those components. Not only do researchers have access to these things, but they can also manipulate them. To the extent that researchers can match their simulated world to the real world, they should be able to read things off the simulated world that will tell them something about the real world. But, how is that done? As Winsberg notes, 'Making the simulation work, and making it produce results that the simulationist is willing to sanction as reliable, is a skill that has been developed in a lengthy period of trial, error and comparison with both theory and known results from physical experiments. In sum, by the semiautonomy of a simulation model, one refers to the fact it starts from theory but one modifies it with extensive approximations, idealizations, falsifications, auxiliary information, and the blood, sweat and tears of much trial and error' [33].

Building a proper simulation model usually requires three steps: (i) certifying that the model is an accurate representation of the biological phenomenon under consideration; (ii) handling the high dimensionality of the parameter space; and (iii) conducting uncertainty and sensitivity analyses to understand how the parameters influence model behavior [36,37]. Surprisingly, there is nothing here about fitting the model to data. Sometimes, this can be done, but often it cannot. This is not to say that data are not used for this type of complex simulation modeling. Data are included from previous studies to construct the model itself. The kinds of experiment done with the simulation model give insight into future data-gathering efforts, test hypotheses that would be impossible to test otherwise and inform researchers about the implications of theoretical insights contained in the causal story that the model represents. Simulation is another experimental system with which to explore theories about how the real world works, using an artificial world that researchers can control. The simulation can point to areas for which more data are needed. It offers hypotheses for testing with simpler models and points to management options that might be reached in no other way. In short, it shows what the world would look like, if it really did work the way in which we think it does.

I am not advocating that we add things into the model that are unnecessary and that the model be a complex one just because it can be simulated. When an analytic model can be used, it should be used. They are simpler, clearer and usually can be fit to data to make their interpretation much easier. However, some processes are inescapably complex. This is the domain of the complex simulation model.

Recognizing that simulation models are another kind of experimental systems helps place their use in a new context. They are an important addition to the methods of science and their proper use can help researchers investigate the world in new ways by providing a method with which to explore ideas in ecology and evolution that would be impossible otherwise.

Acknowledgements

I thank Eric Winsberg, Mark Belk, Kent Hatch, Duane Jeffry, Sarah Jenkins, Dennis Shiozawa, Mark Sisterson, Jack Sites and Loreen Woolstenhulme for their insightful help and comments. Comments from three anonymous reviewers were also valuable.

References

1 Rohrlich, F. (1990) Computer simulation in the physical sciences. *Philos. Sci. Assoc.* 2, 507–518
2 Wolfram, S. (2002) *A New Kind of Science*, Wolfram Media
3 Keller, E.F. (2002) *Making Sense of Life: Explaining Biological Development with Models, Metaphors, and Machines*, Harvard University Press
4 Berryman, A.A. (1991) Population theory: an essential ingredient in pest prediction, management, and policy-making. *Am. Entomol.* 37, 138–142
5 Hess, G.R. (1996) To analyse, or to simulate, is that the question? *Am. Entomol.* 42, 14–16
6 Gilbert, N. (1999) Modeling sociality: the view from Europe. In *Dynamics in Human and Primate Societies* (Kohler, T.A. and Gumerman, G.J., eds), pp. 355–371, Oxford University Press
7 Wilson, W. (2000) *Simulating Ecological and Evolutionary Systems in C*, Cambridge University Press
8 Wilson, W.G. (1998) Resolving discrepancies between deterministic population models and individual-based simulations. *Am. Nat.* 151, 116–134
9 Taylor, B.L. *et al*. (2000) Evaluating dispersal estimates using mtDNA data: comparing analytical and simulation approaches. *Conserv. Biol.* 14, 1287–1297
10 Oreskes, N. *et al*. (1994) Verification, validation, and conformation of numerical models in the earth sciences. *Science* 263, 641–646
11 Gavrilets, S. (2003) Models of speciation: what have we learned in 40 years? *Evolution* 57, 2197–2215
12 Winsberg, E. (1999) Sanctioning models: the epistemology of simulation. *Sci. Cont.* 12, 275–292
13 Winsberg, E. (2001) Simulations, models, and theories: complex physical systems and their representations. *Philos. Sci.* 68, S442–S454
14 Morrison, M. and Morgan, M.S. (1999) Models as mediating instruments. In *Models as Mediators* (Morgan, M.S. and Morrison, M., eds), pp. 10–37, Cambridge University Press
15 Wilke, C.O. and Adami, C. (2002) The biology of digital organisms. *Trends Ecol. Evol.* 17, 528–532
16 Heudin, J. (1999) Virtual worlds. In *Virtual Worlds: Synthetic Universes, Digital Life, and Complexity* (Heudin, J.C., ed.), pp. 1–28, Perseus Books
17 Ginzburg, L.R. and Jensen, C.X.J. (2004) Rules of thumb for judging ecological theories. *Trends Ecol. Evol.* 91, 121–126
18 Cooper, G.J. (2003) *The Science of the Struggle for Existence: On the Foundations of Ecology*, Cambridge University Press
19 Thompson, J.R. (2000) *Simulation: a Modeler's Approach*, John Wiley & Sons
20 Hubbell, S.P. (2001) *The Unified Neutral Theory of Biodiversity and Biogeography*, Princeton University Press
21 Volkov, I. *et al*. (2003) Neutral theory and relative species abundance in ecology. *Nature* 424, 1035–1037

22 Shrader-Frechette, K.S. and McCoy, E.D. (1993) *Method in Ecology: Strategies for Conservation*, Cambridge University Press
23 Hilborn, R. and Marc, M. (1997) *The Ecological Detective: Confronting Models with Data*, Princeton University Press
24 Johnson, J.B. and Omland, K.S. (2004) Model selection in ecology and evolution. *Trends Ecol. Evol.* 19, 101–108
25 Hanski, I. and Ovaskainen, O. (2000) The metapopulation capacity of a fragmented landscape. *Nature* 404, 755–758
26 MacArthur, R. and Wilson, E.O. (1967) *The Theory of Island Biogeography*, Princeton University Press
27 Edwards, A.W.F. (2002) The fundamental theorem of natural selection. *Theor. Popul. Biol.* 61, 335–337
28 Hall, C.A.S. (1988) An assessment of several of the historically most influential theoretical models used in ecology and the data provided in their support. *Ecol. Model.* 43, 5–31
29 Durrett, R. and Levin, S. (1994) The importance of being discrete (and spatial). *Theor. Popul. Biol.* 46, 363–394
30 Halloran, M. *et al*. (2002) Containing bioterrorist smallpox. *Science* 298, 1428–1432
31 Cartwright, N. (1983) *How the Laws of Physics Lie*, Oxford University Press
32 Cartwright, N. (1999) *The Dappled World: A Study of the Boundaries of Science*, Cambridge University Press
33 Winsberg, E. (2003) Simulated experiments: methodology for a virtual world. *Philos. Sci.* 70, 105–125
34 Banks, J. (1998) Principles of simulation. In *Handbook of Simulation: Principles, Methodology, Advancing Application, and Practice* (Banks, J., ed.), pp. 3–30, John Wiley & Sons
35 Fix, A.G. (1999) *Migration and Colonization in Human Microevolution*, Cambridge University Press
36 Swartzman, G.L. and Kaluzny, S.P. (1987) *Ecological Simulation Primer*, MacMillan
37 Kleijnen, J. (1998) Experimental design for sensitivity analysis, optimization, and validation of simulation models. In *Handbook of Simulation: Principles, Methodology, Advancing Application, and Practice* (Banks, J., ed.), pp. 173–223, John Wiley & Sons
38 Lange, K. (1997) *Mathematical and Statistical Methods for Genetic Analysis*, Springer
39 McCulloch, A.D. and Huber, G. (2002) Integrative biological modeling. In *'In Silico' Simulation of Biological Processes* (Bock, G. and Goode, J.A., eds), pp. 4–19, John Wiley & Sons
40 Meza, A. *et al*. (2003) Simulations of galaxy formation in a ΛCDM universe. III. The dissipative formation of an elliptical galaxy. *Astrophys. J.* 590, 619–635
41 Zheng, C. *et al*. (2002) Quantum simulation of ferrocytochrome *c*. *Nature* 334, 726–728
42 Moresi, L. and Lenardic, A. (1997) Three-dimensional numerical simulations of crystal deformation and subcontinental mantle convection. *Earth. Planet. Sci. Lett.* 150, 233–243
43 Gillett, N.P. and Thompson, D.W.J. (2003) Simulation of recent Southern Hemisphere climate change. *Science* 302, 273–275
44 Duke-Sylvester, S.M. and Gross, L.J. (2002) Integrating spatial data into an agent-based modeling system: Ideas and lessons from the development of the across-trophic-level system simulation. In *Santa Fe Institute Studies in the Sciences of Complexity. Integrating Geographic Information Systems and Agent-Based Modeling Techniques for Simulating Social and Ecological Processes* (Gimblett, H.R., ed.), pp. 125–136, Oxford University Press
45 Carrière, Y. *et al*. (2001) Predicting spring moth emergence in the pink bollworm (Lepidoptera: Gelechiidae): implications for managing resistance to transgenic cotton. *J. Econ. Entomol.* 94, 1012–1021
46 Alstad, D.N. and Andow, D.A. (1995) Managing the evolution of insect resistance to transgenic plants. *Science* 268, 1894–1896
47 Storer, N.P. *et al*. (2003) Spatial processes in the evolution of resistance in *Helicoverpa zea* (Lepidoptera: Noctuidae) to Bt transgenic corn and cotton in a mixed agroecosystem: a biology-rich stochastic simulation model. *J. Econ. Entomol.* 96, 156–172
48 Storer, N.P. *et al*. (2003) Sensitivity analysis of a spatially-explicit stochastic simulation model of the evolution of resistance in *Helicoverpa zea* (Lepidoptera: Noctuidae) to Bt transgenic corn and cotton. *J. Econ. Entomol.* 96, 173–187