ANNUAL REVIEWS **Further**

Click here for quick links to Annual Reviews content online, including:

- Other articles in this volume
- Top cited articles
- Top downloaded articles
- Our comprehensive search

Agent-Based Models and Microsimulation

Daniel Heard,[1] Gelonia Dent,[2] Tracy Schifeling,[1] and David Banks[1]

[1]Department of Statistical Science, Duke University, Durham, North Carolina 27708-0251; emails: dph11@stat.duke.edu, tab43@stat.duke.edu, banks@stat.duke.edu

[2]Department of Mathematics, Medgar Evers College, City University of New York, Brooklyn, New York 11125; email: gdent@mec.cuny.edu

Annu. Rev. Stat. Appl. 2015. 2:259–72

First published online as a Review in Advance on November 6, 2014

The *Annual Review of Statistics and Its Application* is online at statistics.annualreviews.org

This article's doi: 10.1146/annurev-statistics-010814-020218

Keywords

emulators, individual-based models, inference, model reduction

Abstract

Agent-based models (ABMs) are computational models used to simulate the actions and interactions of agents within a system. Usually, each agent has a relatively simple set of rules for how he or she responds to his or her environment and to other agents. These models are used to gain insight into the emergent behavior of complex systems with many agents, in which the emergent behavior depends upon the micro-level behavior of the individuals. ABMs are widely used in many fields, and this article reviews some of those applications. However, as relatively little work has been done on statistical inference for such models, this article also points out some of those gaps and recent strategies to address them.

1. INTRODUCTION

Agent-based modeling differs from other types of statistical modeling because it describes only the behavior of the individuals, or agents, in a system, rather than global properties of the system. In Europe and in some disciplines, these models are called individual-based models. Agent-based models (ABMs) are not an appropriate methodology for all problems; e.g., this methodology would probably not be helpful in forecasting earthquakes. But when the behavior of interest is driven by an ensemble of small-scale (often local) actions, agent-based modeling can be a powerful tool.

To fix the concepts, consider a simple ABM that could be used to estimate the time required to evacuate a building. The environment is the floor plan of the building. The agents are "people" who are located at random within the building. Each agent has simple rules: (*a*) when the fire alarm goes off, go to the nearest exit; (*b*) if an exit is blocked, go to the nearest exit not already tried; and (*c*) do not pass through another agent. Each agent also has a parameter that indicates how quickly he or she walks. The analyst can then run the model many times, randomizing the initial locations of the agents each time, and create a histogram of the time required for evacuation. The analyst can also explore "what if?" scenarios, in which certain exits are blocked. In addition, agents need not be sentient; in weather forecasting, for example, an agent can be a cubic kilometer of atmosphere that exchanges pressure, temperature, and moisture with neighboring agents according to laws of atmospheric physics.

The inception of ABMs is rooted in cellular automata, developed by John von Neumann and Stanislaw Ulam at Los Alamos in the 1940s (Wolfram 1983). Cellular automata are ABMs in which the geography is a grid, and the grid cells interact according to specific rules. The most famous of these is Conway's Game of Life (**Figure 1**) (Gardner 1970).

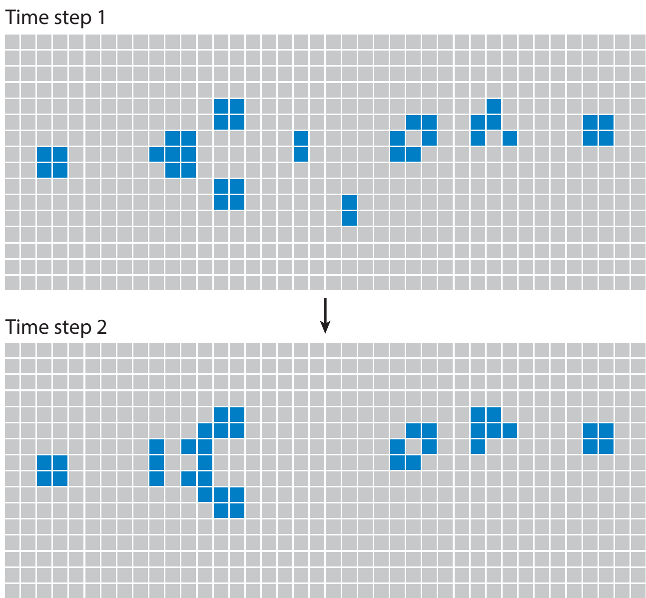

Figure 1

A single time point in the evolution of Conway's Game of Life simulation. The blue cells are "alive," and the gray cells are not. If a live cell has either fewer than two or more than four live neighbors among eight possible neighbors, it "dies" at the next time step; otherwise it continues to live. In addition, a gray cell with three live neighbors becomes a blue living cell at the next time point. The goal is to understand the configurations of live and dead cells that show persistent patterns.

ABMs grew in popularity in the 1990s because of the ease of implementation that came with improvements in available computer technology. Social sciences began using ABMs to explore social phenomena and the growth of societies over time; work by Epstein & Axtell (1996) provides a thorough discussion of this matter, example simulations of which can be seen at **https://www.youtube.com/watch?v=SAXWoRcT4NM** and **http://sugarscape.sourceforge.net/**.

In general, an ABM is built by specifying the properties and rules of the agents and their environment. Many of the decisions that a modeler should consider when building an ABM are outlined in the "Overview, Design Concepts, and Details" (ODD) protocol discussed in Grimm et al. (2006). The ODD protocol provides a comprehensive template for describing a general ABM and is intended to make model descriptions more structured, more complete, and more easily understood, especially in terms of reporting ABM work in the academic literature. This, in turn, enables reproducibility studies of ABMs and addresses other implementation concerns associated with ABMs, as detailed in an article by Lorek & Sonnenschein (1999). An extensive discussion of the ODD protocol can be found in Polhill et al. (2008), and a discussion of the growing use of this protocol is presented in Grimm et al. (2010).

The "Overview" portion of the ODD protocol contains "Purpose," "State Variables and Scales," and "Process Overview and Scheduling" subsections. The "Purpose" subsection explains what the model is intended to do and its general goals. The "State Variables and Scales" subsection outlines the structure of the model and identifies all of the entities within it, including the types of agents, the spatial structure, and other local and global variables. The variables that determine the state(s) of these entities at any given point during the simulation are also specified. The "Process Overview and Scheduling" subsection lists all processes that occur in the model and the order in which they occur. This subsection encompasses the hierarchy of agent behaviors, any effects these behaviors have on the environment, and associated updates to the states of entities.

The "Design Concepts" section of the ODD protocol describes the general concepts upon which the model is based. The primary purpose of this section is to provide an understanding of why certain design decisions were made. Such concepts include a summary of the emergent behavior, agent objectives, agent interactions, and a description of the perception(s) and thought processes of the agents, if these exist.

The "Details" portion of the protocol contains subsections on "Initialization," "Input," and "Submodels." The "Initialization" subsection identifies how the model is started and often provides references to support initial values of the variables. The "Input" subsection describes any other external inputs to the model (such as time series market data for financial models or temperature data for climate models). The "Submodels" subsection explains in detail the equations and algorithms used in the model and defines any parameters included in those equations and/or algorithms. Although the parameters in the model are often set to realistic values, the modeler may also need to calibrate some parameters from historical data to ensure that the output is similar. This need raises statistical questions related to goodness of fit and model calibration.

One way to validate an ABM is to compare output from the model with real data. More specifically, a modeler can choose several patterns observed in real data and investigate whether the ABM output shows similar patterns. This method of comparing patterns in the model and in real data is known as pattern-oriented modeling. If the ABM produces outputs that match well with multiple patterns that capture different aspects of the real data, then the modeler becomes more confident that the ABM is realistic.

ABMs are used in a wide range of fields, including finance and economics, ecology, and biology and epidemiology. This article reviews the ongoing development of ABMs and the breadth of their application.

2. ECONOMICS AND FINANCE

The flexibility of ABMs in incorporating a range of individual behaviors has led to the development of agent-based computational economics (ACE) and agent-based computational finance (ACF). The use of ABMs in the fields of economics and finance is growing. Predicting the behavior of individuals in settings in which resources are limited is important, and the value of ABMs in simulating and understanding the economy has been discussed in special issues of several prominent journals [*Computational Economics*, Volume 18(1), *Journal of Economic Dynamics and Control*, Volume 25(3,4), and *IEEE Transactions on Evolutionary Computing*, Volume 5(5)]. An online repository of ACE resources is maintained at **http://www2.econ.iastate.edu/tesfatsi/ace.htm**.

Because an economy is largely the result of interactions among autonomous individuals, ABMs are a natural tool for exploring the subject of economics. ABMs describe the microeconomic actions of adaptive agents, which give rise to emergent behavior in the form of macroeconomic structures. These structures, in turn, influence agent decisions. This influential cycle is an important component of ACE models. Such models are used in the context of financial institutions, retail, and internet auctions. Chen et al. (2009) explored the bridging of agent-based computational economic models and econometrics to enrich the content within these ABMs. The importance of examining ACE has led several journals to devote special issues to the topic.

As discussed in Tesfatsion (2002), which provides an overview of ABMs in economics and finance, research in these fields is divided into five main areas. We describe these areas and related topics in the following subsections.

2.1. Learning

We have already mentioned the adaptive nature of markets, financial institutions, and agents. To allow realistic incorporation of this adaptive behavior, an essential component of ACE research is the creation of algorithms to represent the learning process of agents. Gintis (2000) proposes local learning behavior in which agents coevolve within different neighborhoods and develop strategies based on individual utility functions. This behavior allows for the incorporation of game theoretic concepts that take account of long-run outcomes. It also enables examination of the degree to which an implementation is sensitive to the learning algorithms used. (For example, researchers can ask "What happens if agents learn to satisfice rather than maximize expected utility when margins are small?")

A common strategy to study learning in ACE is to use experimental data from human subjects as an aid in calibrating the learning of agents. Duffy (2001) performed parallel experiments with human subjects and computational agents to determine whether a generally accepted exchange medium emerges. In his experiment, each agent desires to consume some good other than the one he or she produces. Agents are randomly paired and decide whether to exchange goods on the basis of either desired consumption of alternate goods or intent to save their own goods (at some storage cost) for use in a future trade. The primary objective is to observe whether all agents converge on a specific good that acts as currency and is accepted in trades despite lacking consumption value.

2.2. Evolution of Behavioral Norms

An important capability in ABMs has been the fact that agents can adjust behavior in various social settings over time. Axelrod (1997) demonstrated that cooperation can develop among unrelated, self-interested agents through reciprocity. This work strongly influenced the way in which researchers consider individual strategies and the resulting social norms.

Extending earlier results, Epstein (2002) examined a more in-depth component of the evolution of norms, studying the extent to which individuals think about behaviors using an agent-based computational approach. His model proposed a set of agents, each of whom is assigned a binary norm and a sampling radius. In this model, each agent adjusts his or her norm based on observations of other agents within that radius, simply by matching his or her own norm to those of the majority of agents within his or her radius. By varying the initial distribution of norms, Epstein's research found an inverse relationship between the effort that an agent expends in polling about a behavior and the strength of the norms related to it.

2.3. Bottom-Up Modeling

A major limitation of traditional models of financial markets has been the inability to explain basic empirical market features, such as the heavy-tailed distribution of returns on assets and high trading volumes. Such issues have drawn ACE modelers to this area of research. Models of financial markets using ACE have provided possible explanations for some of the observed features of financial data (Hommes 2002, Lux & Marchesi 2000).

In the interest of obtaining models that better fit observed financial market data, LeBaron (2001) developed an agent-based computational stock market model calibrated with aggregate data. Under this approach, investors evaluate the performance of their stock trading rules on the basis of past performance and use memories of varying length. This model generated features consistent with those of real-world financial time series data.

Izumi & Ueda (2001) tried a similar approach for analysis of foreign exchange markets, which have proven more difficult to model. They use field data to construct behavioral rules that govern the learning and interactions of agents. The agents in their model compete to develop methods to predict future alternations in exchange rates. The study provided possible explanations for emergent behavior consistent with empirical features of foreign exchange markets.

2.4. Formation of Economic Networks

Modeling the strategic interaction of agents in a competitive economic setting is an important component of ACE. Transaction networks form among agents and evolve over time as market demands and agent strategies change. A particular type of transaction network of interest to economists because of its potential optimality properties is the small-world network (**Figure 2**) (Wilhite 2001). This is a connected network in which (*a*) each node is linked to a relatively densely connected set of neighbor nodes, and (*b*) shortcut connections between some nodes reduce the average minimum path length between nodes.

Tassier & Menczer (2001) examined job referral networks in United States labor markets. Their ACE labor market model features workers engaging in direct job search and in social network formation. Workers must acquire sufficient resources to meet a minimum survival requirement in order to remain viable. The model found that agents were inefficient in the amount of time spent on direct job searches compared with that spent on social network formation, corresponding to a lack of global efficiency in the labor market.

2.5. Modeling Organizations

Organizations within the economic community are an important aspect of ACE. Dawid et al. (2001) present an approach to modeling organizations that examines how the decisions of individual firms affect the optimal behavior rules of all firms within the market. In this approach, all of the firms

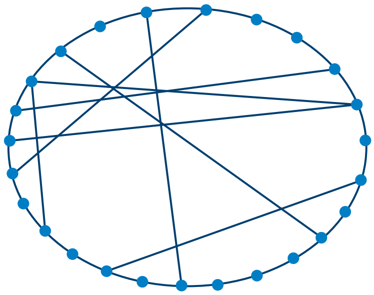

Figure 2

An example of a small-world network. Without the interior edges, it would be a large-world network, as the shortest path connecting two random nodes would have to traverse the circumference, and in most cases, that would be a long path. However, the interior edges mean that the shortest path between two random nodes can take a shortcut through the center, leading to a much smaller average number of steps.

compete in an industry, and, at each time step, each firm decides whether to produce an existing product or to introduce a new one. Introduction of new products is motivated by a diminishing demand for existing products after a stochastically determined period of time. Each firm has a rule to determine whether or not to "innovate" and produce a new product. The decision rules used by the firms coevolve over time based on expected profit. The authors explore how innovation rules should adapt to industry structure in order to optimize long-run profitability.

2.6. Summary

ACE and ACF are prominent means of leveraging the complex simulation power of ABMs. Implementation of adaptive agents allows realistic simulations of individual and organizational development and evolution. Additional work on quantifying uncertainty of these simulations will contribute to the utility of ABMs in economics and finance.

3. ECOLOGY

ABMs are widely used in ecology (where they are usually referred to as individual-based models). Their popularity reflects the fact that many ecological systems, such as trees in a forest or animals on a veld, are naturally described in terms of agents whose interactions depend on simple rules. In addition, ecological questions often pertain to ensemble behavior, such as succession in a forest or the extinction of a zebra herd. Previous analytical techniques, such as predator–prey equations, were generally too abstract to capture important and realistic behavior. In contrast, the rule sets in ABMs do capture such behavior, but as a result, they become needlessly elaborate. The penalty for that is typically longer computational time, not infidelity to the phenomenon of interest.

Ecologists are interested in issues such as land use planning, biodiversity, and the effects of climate change, which are impacted by resource management, water distribution, and population distributions. Thus, major questions in this discipline fit the traditional ABM structure in which there are both an environment that has specified properties and multiple types of agents that interact with the environment and with each other. To answer ecological questions, it is vital to first explore alternative scenarios for, say, water distribution and to then determine the effect(s) these scenarios have on species. The ability to examine alternative scenarios is often outside the reach of differential equation models and other numerical techniques.

Although an ABM must be tailored to a specific ecological application, there are generally common challenges, which can be addressed within the ODD framework. First, one must represent the ecological process. Second, one must describe the spatial environment. Third, one must couple or integrate the ABM with relevant input models describing, for example, annual temperature variation or rainfall. And fourth, one must use the outputs of the ABM to make statistical inferences on the predictions generated by it. These aspects are discussed below.

3.1. Representation of Processes

How one represents an ecological process and the spatial environment depends on the context of the problem. In applications of agent-based models with socioecological systems (ABM/SESs) (cf. Filatova et al. 2013), the main benefit of the ABM simulation technique is its ability to represent human behaviors that require strategic thinking, compromise, and regulation, among other processes. ABM/SESs are used to study methods for managing ecological resources, for example, by modeling payoff scenarios or the effect of network incentives on farmland versus nature preserves. In both examples, the behavior of an agent (i.e., the choices made by the farmer) affects the outcome. How people respond to incentives is not simple, however, so a major concern in ABM/SES analyses relates to the lack of a clear protocol that guides, even probabilistically, the choices that the agent makes.

Agent-based models of land use and cover change (ABM/LUCCs) study changes in land use and land cover caused by the impact of human interventions on their environments. **Table 1** restates results from work by Luus et al. (2013) on the framework for representation in ABM/LUCCs. For each set of ecological characteristics of interest, an appropriate model is chosen that matches the application. Accurate simulation of the ecological process is the goal of the study. When a modeler is determining which type of representation is desired (in terms of the level of resolution and the inclusion of relevant variables), the choice depends on information about the availability of data, the timescale, the type of interactions, and the influence of human decision making. The modeling capability is driven by the detail and accuracy of such information. When the process changes slowly or there is insufficient information, a static approach might be used to model the

Table 1 Representation of ecological processes in agent-based models of land use and cover change

Ecological characteristics or function	Model approach	Primary application
Not spatially dependent Process has longer timescales Inadequate data	Static representation	Exploratory
Processes change in an ordinal or nominal manner Assumption that changes are reversible Extensive quantitative data not available	Transition rules	Exploratory Participatory Predictive
Model generates empirical outcomes Explanatory variables available	Regression	Participatory Predictive
Highly dependent upon population dynamics Respond to human behavior or environmental changes over space and time	Individual based	Participatory Predictive
Influences on ecosystem due to natural competition of species	Equilibrium based	Participatory Predictive

process, usually as an equilibrium. More often, modelers seek a dynamic representation, generally to support decision making.

There are four types of dynamic representations: transition rules–based, regression-based, individual-based, and equilibrium-based. The transition rule approach is most suitable when there is a lack of quantitative data. Rules-based transitions are created to define the state of the process over time, and the process is often potentially reversible. Regression-based models use explanatory variables to generate estimates of outcomes of an ecological process. Both the individual- and equilibrium-based approaches model long-term effects on the basis of complex interactions between species.

The primary applications of ABM/LUCCs fall into three categories, which correspond to whether the simulation focuses on micro-level behavior (often exploratory in nature), macro-level behavior (often predictive in nature), or both. Exploratory applications are theoretical studies of coupled human–environmental systems, whereas participatory applications involve many stakeholders and have a goal of gaining predictive insight into the mechanisms of the process.

3.2. Coupling Models

Ecological models may be represented in ABMs by either coupling or integrating submodels. When submodels are coupled, for example, in describing predator–prey dynamics with separate population dynamics for each species, it is important to establish feedback between the submodel components at the correct temporal scales, as coupling of systems gives rise to issues of temporal and spatial resolution. In numerical analysis, stability criteria are established to assess the accuracy of and dependency on the temporal and spatial scales. Use of a scale that is too coarse tends to reduce fidelity; however, a scale that is too fine increases computational costs and may also affect fidelity.

Linkages between submodels can entail loose coupling, tight coupling, or integration of the submodels into one system (Antle et al. 2001). Parker et al. (2008) describe three ways of linking SES models with ABMs: (*a*) one-way linkage for natural science inputs, i.e., naturescience, (*b*) a chain of multiple one-way linkages with natural system inputs but possibly differing output models, and (*c*) two-way linkage with endogenous variables resulting from interactions between the social and environmental systems. Most ABM/SES models exhibit chained one-way linkages. Two-way linkage models allow feedback between, say, human and environmental systems, thereby modeling a nonlinear interaction in which changes in one submodel can cause abrupt transitions in the other.

Geographic information systems (GISs) have been coupled with and integrated into ABMs. Crooks (2010) implemented an ABM loosely coupled with a vector GIS to model residential segregation. In that model, the GIS data provided a geometry to represent both agents and objects in the environment, reportedly resulting in a more real environment. Although an obvious advantage to coupling models is that one can study more realistic scenarios, doing so usually comes at a higher computational cost and raises more complex issues of model calibration.

3.3. Summary

This section has presented an overview of ABMs used in ecology. We have outlined the main issues that one must face when modeling the complex interactions of humans with their environment or those between multiple species. High-capacity computing has been increasingly important in pushing research boundaries forward in most fields of science, and ABMs represent one important computer-intensive tool in the modern science of ecology. Key future developments include the development of statistical methodology to (*a*) assess the fit between parameterized ABMs and real-world data, (*b*) establish stability criteria for spatial and temporal scales, and (*c*) allow researchers to make formal uncertainty statements regarding results from ABM analyses.

4. BIOLOGY AND EPIDEMIOLOGY

ABMs have proven to be a useful tool for studying a variety of biological phenomena, including molecular biology and, frequently, disease outbreaks. These simulations involve the adaptation of agents to various interventions and treatments over time, as examined in detail by Berry et al. (2002).

One particular area of ABMs related to disease outbreaks is that of bioterrorism. Burke et al. (2006) used ABMs to simulate strategies for controlling a smallpox epidemic. Their work modeled different civic arrangements and social structures (see **Figure 1**), but they focused their analysis mainly on determining the effectiveness of different public health response scenarios.

The parameters for the model governing transmission rates and contact rates were calibrated based on historical data describing smallpox outbreaks in Europe during 1950–1971. They compared simulations under an absence of public health response to the epidemic to simulations under eight different response scenarios. The public health response scenarios included contact tracing (identifying individuals likely to have been in contact with infected individuals), adjusting contact rates to reflect social distancing during the outbreak, adjusting both the preemptive and reactive vaccination rates, and implementing school closures. Both the no-response and response scenarios were simulated on communities of 6,000 and 50,000 individuals.

Burke et al. (2006) found that, although the simulated epidemics were subject to a high degree of variation, contact tracing and prompt vaccination of the household and work or school contacts of infected individuals were most effective in containing the epidemic. More generally, this study demonstrated the utility of ABMs in crafting policy for outbreak response.

Cisse et al. (2013) proposed an ABM for the spread of bilharzia (also known as schistosomiasis). They developed their model by first randomly distributing agents and randomizing their movement patterns. They then showed that this approach produces results similar to those from mathematical models based on coupled ordinary differential equations (ODEs). Next, they added a spatial component to the model to assess the impact of spatial heterogeneity on the basic model and found that the additional spatial component influenced strategies for controlling the epidemic.

This exercise demonstrates the utility of ABMs in the context of epidemic modeling, as one can mimic standard compartmental ODE models while including a spatial component to more fully examine the dynamics of the spread of a disease. This combination enables more accurate predictions and evaluations of interventions by explicitly accounting for spatial heterogeneity.

4.1. Hierarchical Modeling

Hooten & Wikle (2010) introduced Bayesian hierarchical models into ABM research. Their technique does not apply to all ABM formulations (of which there are many), but it is useful for spatiotemporal processes with fairly simple structures. Their motivating application was the spread of rabies in raccoon populations in Connecticut between 1991 and 1995. They used a gridded map to represent the townships in the state of Connecticut, where the presence or absence of rabies was indicated by a binary random variable that had a distribution that depended on the rabies status (presence or absence) in the neighboring townships at the preceding time period. Their representation also included covariates (which could also vary in time).

Let $\mathbf{u}_t = (u(1, t), u(2, t), \ldots, u(m, t))'$ denote the binary vector showing the presence or absence of rabies at time t for each of the m townships, and let $\mathbf{X}_t = (\mathbf{x}(1, t), \mathbf{x}(2, t), \ldots, \mathbf{x}(m, t))$ denote a matrix of corresponding covariates, such as population density and adjacency to the Connecticut river, among others. Define the neighborhood for township i by $\mathcal{N}_i$; this is a set of townships. The basic model for the spread of the disease is as follows:

$$u_{i,t}|\mathbf{u}_{\mathcal{N}_i,t-1} \sim [u_{i,t}|h(\mathbf{u}_{\mathcal{N}_i,t-1}, \mathbf{x}_{\mathcal{N}_i,t-1})],$$

where $h(\cdot, \cdot)$ is a very general updating function, the subscript $\mathcal{N}_i$ indicates the townships relevant to the disease spread at township i, and the bracket notation indicates that the presence or absence of rabies is a random variable with parameters that depend on the conditioning within the bracket. The only substantive difference between this model and a Gaussian state space model is that the random variables in this model need not be Gaussian (which generally precludes closed-form solutions, putting this within the realm of ABM simulation).

This model is flexible, and it allows disease spread to be anisotropic (e.g., directional along the Connecticut river). It also enables one to make probabilistic statements about the posterior probability of disease in a particular township, but doing so usually requires Markov chain Monte Carlo (MCMC) simulation. This methodology does not apply to all ABMs (e.g., genetic algorithms or the evacuation of a building), but when it does apply, it permits more explicitly statistical inference on the behavior of the ABM.

4.2. Summary

In biology and epidemiology, ABMs are particularly useful in the simulation of "what if" scenarios, allowing one to better understand the dynamics of processes of interest. In particular, these models can help inform effective treatment and intervention strategies.

5. RECENT DEVELOPMENTS

Several areas related to ABMs are currently being studied. Much of the current work in these areas addresses issues of how to (*a*) quantitatively compare multiple models and (*b*) leverage the information produced by ABMs in a more efficient manner.

5.1. Model Reduction

A growing area of interest is the ability to develop a reduced or simplified version of an existing ABM. The objective is to maintain the robustness of the simulation while reducing computational burden and possibly decreasing model run time.

Zou et al. (2012) performed some of the earliest work in model reduction for ABMs. They presented a computer-assisted approach to find a compromise between a highly detailed agent-level model and a macroscopic population-level model. They applied this methodology to a civil violence model based on the work of Epstein (2002). This exercise presents an important balance to consider in terms of the marginal information lost in decreasing the detail of a model, in contrast to the benefit of increasing its efficiency. Zou et al. (2012) proposed a reduced model in the form of a system of stochastic differential equations.

Additional work on model reduction of ABMs has been explored by researchers at RTI International (Heard et al. 2014). They explore the reduction process on an ABM for a heroin market, presented in Hoffer et al. (2009), which was based on research detailed in Hoffer (2005). To reduce simulation complexity, they explore altering the timescale of the model, and they identify elements that can be approximated by regressions and by draws from conditional probability distributions. Whereas the full model involved multiple agent types (heroin customer, street heroin dealer, street heroin broker, private heroin dealer, police officer, and homeless person), the reduced model focused on customer outcomes, allowing some of the explicit roles of other agents to be eliminated from the model. Summary statistics for customer outcomes, including money, drug inventory, addiction level, and drug concentration were identified as useful measures for comparing the fidelity of the reduced model to the full model. Although changing the timescale significantly altered many of the dynamics of the model, the process approximation yielded a model that had a 43% reduction in run time and that produced results in agreement with the full model in terms of customer behavior.

A key component in model reduction is selecting an optimal subset of variables and observables from the ABM simulation. The optimal subset is one that provides the greatest agreement with the full model for features of interest, and the process for identifying these variables is application specific. Although certain approaches, such as principal component regression (Cangelosi & Goriely 2007) for model output against model inputs, can be used to select the variables for model reduction, the topic of model reduction for ABMs requires further investigation.

5.2. Statistical Inference

The power of ABMs for many complex systems gives them the potential to be useful tools in performing statistical inference. But, because many ABMs have a complicated structure, likelihood functions are nearly always intractable. Hence, inference must be performed in a likelihood-free context. There are two main tools that allow us to perform such inference utilizing ABMs: emulators and approximate Bayesian computation (ABC).

Emulators are statistical approximations of computer simulations. They are valuable because (*a*) they can incorporate detailed functions from ABMs and (*b*) they allow one to make predictions of ABM output based on a limited number of simulations. This ability is particularly advantageous for ABMs with long run times whose results need to be processed in real time (e.g., in weather forecasting).

Bijak et al. (2013) examine the use of Gaussian process (GP) emulators in conjunction with agent-based models in the context of modeling the population of the United Kingdom. They take a set of simulations from their Semi-Artificial Model of Population (SAMP) and, using the predefined vector of inputs, define a GP emulator on the basis of the ABM output. Estimates for model calibration parameters and parameters of interest for the system being modeled are obtained using Bayesian methods via MCMC simulations.

The work of Bijak et al. (2013) identifies the value of coupling ABMs with emulators in the area of demography. They demonstrate the ability of emulators to both locate plausible regions, in terms of fit, within the parameter space of an ABM and quantify uncertainty in model predictions. The approach builds on the rich statistical literature on emulators, notably the work of O'Hagan (2006), Kennedy & O'Hagan (2001), and Higdon et al. (2008).

Another emerging tool for inference utilizing ABMs is ABC, a technique based on the work of Pritchard et al. (1999). This approach has been proposed to estimate parameter distributions as well as for model specification for ABMs (Heard 2014). The objective of ABC is to infer a parameter θ on the basis of some observed data x_0 when likelihood functions are intractable. The ABC method compares simulated data with observed data using some discrepancy measure and then uses this measure as a criterion for estimating the posterior distribution of θ. Several authors have proposed extensions to the basic ABC algorithm that increase the efficiency of simulating the approximate posterior distribution of θ. These extensions include approaches based on MCMC

(Marjoram et al. 2003), sequential Monte Carlo (Del Moral et al. 2012), and regression adjustments (Blum & Francois 2010).

Suppose the parameter θ corresponds to some uncertain quantity related to the function of the system that the ABM represents. The basic approach for using ABC for statistical inference in this case is to sample some parameter value θ' from a specified distribution. Using θ' as an input setting in the ABM, one then simulates a sample trajectory. The value θ' is accepted as a draw from the approximate posterior if the simulated data are sufficiently close to the observed data using the specified discrepancy measure.

Although the ABC approach is straightforward to implement, it has the disadvantage of requiring significantly more ABM simulations than the emulator approach. As a result, this approach may not be feasible for ABMs with substantial run times.

6. DISCUSSION

We have examined a variety of areas in which ABMs are being used for analysis. ABM techniques within particular fields are developing and becoming more specialized. Methodology for creating reduced versions of ABMs is currently being explored and improved; the development of such methodology will have value for researchers in many fields of study. Tools for statistical inference using ABMs are emerging, and, although there is still much room for further investigation, the importance of this topic has been demonstrated. As the implementation of ABMs continues to expand, the associated methodology will develop and will continue to incorporate techniques from a variety of disciplines.

SUMMARY POINT

1. The class of problems to which ABMs are being applied is growing. Advancement of ABM techniques will occur in conjunction with the development of statistical computation.

FUTURE ISSUE

1. Future research will examine joint diagnostics for ABMs and ABC.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

The authors would like to acknowledge the Statistical and Applied Mathematical Sciences Institute (SAMSI) for helping to coordinate our research efforts.

LITERATURE CITED

Antle JM, Capalbo SM, Elliott ET, Hunt HW, Mooney S, et al. 2001. Research needs for understanding and predicting the behavior of managed ecosystems: lessons from the study of agroecosystems. *Ecosystem* 4(8):723–35

Axelrod R. 1997. *The Complexity of Cooperation: Agent-Based Models of Conflict and Cooperation*. Princeton, NJ: Princeton Univ. Press

Berry BJL, Kiel LD, Elliott E. 2002. Adaptive agents, intelligence, and emergent human organization: capturing complexity through agent-based modeling. *PNAS* 99:7187–88

Bijak J, Hilton J, Silverman E, Cao VD. 2013. Reforging the wedding ring: exploring a semi-artificial model of population for the United Kingdom with Gaussian process emulators. *Demogr. Res.* 29:729–66

Blum MGB, Francois O. 2010. Non-linear regression models for Approximate Bayesian Computation. *Stat. Comput.* 20(1):63–73

Burke SD, Epstein JM, Cummings DAT, Parker JI, Cline KC, et al. 2006. Individual-based computational modeling of smallpox epidemic control strategies. *Acad. Emerg. Med.* 13:1142–49

Cangelosi R, Goriely A. 2007. Component retention in principal component analysis with application to cDNA microarray data. *Biol. Direct* 2(2):1–21

Chen S-H, Chang C-L, Du Y-R. 2009. Agent-based economic models and econometrics. *Knowl. Eng. Rev.* 27:1–46

Cisse PA, Dembele JM, Lo M, Cambier C. 2013. Assessing the spatial impact on an agent-based modeling of epidemic control: case of schistosomiasis. *Complex Sci.* 126:58–69

Crooks AT. 2010. Constructing and implementing an agent-based model of residential segregation through vector GIS. *Int. J. Geogr. Inf. Sci.* 24(5):661–75

Dawid H, Reimann M, Bullnheimer B. 2001. To innovate or not to innovate? *IEEE Trans. Evol. Comput.* 5:471–81

Del Moral P, Doucet A, Jasra A. 2012. An adaptive sequential Monte Carlo method for approximate Bayesian computation. *Stat. Comput.* 22(5):1009–20

Duffy J. 2001. Learning to speculate: experiments with artificial and real agents. *J. Econ. Dyn.* 25:295–319

Epstein JM. 2001. Learning to be thoughtless: Social norms and individual computation. *Comput. Econ.* 18(1):9–24

Epstein JM. 2002. Modeling civil violence: an agent-based computational approach. *PNAS* 99(Suppl. 3):7243–50

Epstein JM, Axtell R. 1996. *Growing Artificial Societies: Social Science from the Bottom Up*. Washington, DC: Brooking Inst.

Filatova T, Verburg PH, Parker DC, Stannard CA. 2013. Spatial agent-based models for socio-ecological systems: challenges and prospects. *Environ. Model. Softw.* 45:1–7

Gardner M. 1970. Mathematical games: the fantastic combinations of John Conway's new solitaire game "Life." *Sci. Am.* 222:120–23

Gintis H. 2000. *Game Theory Evolving*. Princeton, NJ: Princeton Univ. Press

Grimm V, Berger U, Bastiansen F, Eliassen S, Ginot V, et al. 2006. A standard protocol for describing individual-based and agent-based models. *Ecol. Model.* 198:115–26

Grimm V, Berger U, DeAngelis DL, Polhill JG, Giske J, Railsback SF. 2010. The ODD protocol: a review and first update. *Ecol. Model.* 221(23):2760–68

Heard D. 2014. *Statistical inference utilizing agent-based models.* PhD Thesis. Duke University, Durham, N.C.

Heard D, Bobashev GV, Morris RJ. 2014. Reducing the complexity of an agent-based local heroin market model. *PLOS ONE* 9:e100263

Higdon D, Gattiker J, Williams B, Rightley M. 2008. Computer model calibration using high-dimensional output. *J. Am. Stat. Assoc.* 103:570–83

Hoffer LD. 2005. *Junkie Business: The Evolution and Operation of a Heroin Dealing Network (Case Studies on Contemporary Social Issues)*. Beverly, MA: Wadsworth

Hoffer LD, Bobashev GV, Morris RJ. 2009. Researching a local heroin market as a complex adaptive system. *Am. J. Community Psychol.* 44:273–86

Hommes CH. 2002. Modeling the stylized facts in finance through simple nonlinear adaptive systems. *PNAS* 99(3):7221–28

Hooten MB, Wikle CK. 2010. Statistical agent-based models for discrete spatio-temporal systems. *J. Am. Stat. Assoc.* 105:236–48

Izumi K, Ueda K. 2001. Phase transition in a foreign exchange market: analysis based on an artificial market approach. *IEEE Trans. Evol. Comput.* 5:456–70

Kennedy MC, O'Hagan A. 2001. Bayesian calibration of computer models. *J. R. Stat. Soc. B* 63(3):425–64
LeBaron B. 2001. Empirical regularities from interacting long and short horizon investors in an agent-based stock market. *IEEE Trans. Evol. Comput.* 5:442–55
Lorek H, Sonnenschein M. 1999. Modelling and simulation software to support individual-oriented ecological modelling. *Ecol. Model.* 115:199–216
Luus KA, Robinson DT, Deadman PJ. 2013. Representing ecological processes in agent-based models of land use and cover change. *J. Land Use Sci.* 8:175–198
Lux T, Marchesi M. 2000. Volatility clustering in financial markets: a micro-simulation of interacting agents. *Int. J. Theor. Appl. Finance* 3:675–702
Marjoram P, Molitor J, Plagnol V, Tavare S. 2003. Markov chain Monte Carlo without likelihoods. *PNAS* 100:15324–28
O'Hagan A. 2006. Bayesian analysis of computer code outputs: a tutorial. *Reliab. Eng. Syst. Saf.* 91(10–11):1290–300
Parker DC, Hessl A, Davis SC. 2008. Complexity, land-use modeling, and the human dimension: fundamental challenges for mapping unknown outcome spaces. *Geoforum* 39(2):789e804
Polhill JG, Parker D, Brown D, Grimm V. 2008. Using the ODD protocol for describing three agent-based social simulation models of land-use change. *J. Artif. Soc. Soc. Simul.* 11(2):3
Pritchard JK, Seielstad MT, Perez-Lezaun A, Feldman MW. 1999. Population growth of human Y chromosomes: a study of Y chromosome microsatellites. *Mol. Biol. Evol.* 16:1791–98
Tassier T, Menczer F. 2001. Emerging small-world referral networks in evolutionary labor markets. *IEEE Trans. Evol. Comput.* 5:482–92
Tesfatsion L. 2002. Agent-based computational economics: growing economies from the bottom up. *Artif. Life* 8(1):55–82
Wilhite A. 2001. Bilateral trade and small-world networks. *Comput. Econ.* 18(1):49–64
Wolfram S. 1983. Statistical mechanics of cellular automata. *Rev. Mod. Phys.* 55(3):601–44
Zou Y, Fonoberov VA, Fonoberova M, Mezic I, Kevrekidis IG. 2012. Model reduction for agent-based social simulation: coarse-graining a civil violence model. *Phys. Rev. E* 85(6):066106