

Statistical Challenges in Agent-Based Modeling

David L. Banks^a and Mevin B. Hooten^{b,c}

^aDepartment of Statistical Science, Duke University, Durham, NC; ^bDepartment of Fish, Wildlife, and Conservation Biology, U.S. Geological Survey, Colorado Cooperative Fish and Wildlife Research Unit, Fort Collins, CO; ^cDepartment of Statistics, Colorado State University, Fort Collins, CO

ABSTRACT

Agent-based models (ABMs) are popular in many research communities, but few statisticians have contributed to their theoretical development. They are models like any other models we study, but in general, we are still learning how to fit ABMs to data and how to make quantified statements of uncertainty about the outputs of an ABM. ABM validation is also an underdeveloped area that is ripe for new statistical developments. In what follows, we lay out the research space and encourage statisticians to address the many research issues in the ABM ambit.

ARTICLE HISTORY

Received January 2021
Accepted March 2021

KEYWORDS

Emulator; Individual-based model; Mechanistic model

1. Introduction

Agent-based models (ABMs) are simulators used to study interactions among artificial entities (agents) that are governed by rules. The goal with ABMs is often to understand emergent behavior based on a set of explicit rules describing how elements of a system interact. Often complex phenomena arise from simple rules.

ABMs are used in many different disciplines. Examples include:

- Weather forecasting, in which the agents could be viewed as cubic kilometers of atmosphere that exchange pressure, temperature and humidity with their neighbors according to standard physical laws (Simmonds, Gómez, and Ledezma 2019). One can initialize the agents with specific values, and then watch a weather front roll across a continent.
- Transportation, in which agents are vehicles that traverse a road way, obeying rules that determine spacing, speed, destination, and route choice. TRANSIMS is a famous early ABM. The goal is to identify points of high congestion, or to perform “what if?” experiments that study the impact of, for example, a lane closure.
- Sociology, in which agents are “people” with various characteristics (e.g., age, gender, and education) who form friendship networks according to theories of homophily or heterophily (Snijders 1996). The goal might be to estimate the amount of homophily that leads to networks whose connectivity patterns resemble those of human networks.
- Epidemiology, in which infected agents interact with each other to spread a virtual disease. The goal could be to determine what kinds of public health response (handwashing campaigns, social distancing, school closure, and quarantine) are effective for diseases with specific infectiousness and incubation period (Perez and Dragicevic 2009; Frias-Martinez, Williamson, and Frias-Martinez 2011).

ABMs have become a standard tool in many fields, but their use is still nascent in statistics (Hooten, Wikle, and Schwob 2020). They are models like many others that we use, but statisticians have not yet agreed on a theory and methodology for estimating their parameters from data or for making quantified statements of uncertainty about ABM predictions.

There are, of course, a few exceptions. Hooten and Wikle (2010) developed an ABM to infer the probability of rabies in Connecticut townships, and Hunter, Mac Namee, and Kelleher (2018) use an ABM to describe H1N1 spread. But these are isolated studies that make special assumptions and that address relatively simple situations.

1.1. History of ABMs

The ABM perspective dates back to the 1940s, with work on cellular automata by John von Neumann and Stanislaw Ulam (Von Neumann et al. 1966). The most famous of these automata was J. H. Conway’s Game of Life (Gardner 1970). The agents are squares on a sheet of graph paper in the Game of Life. The squares may be colored either white or black and may change color from one time step to another. A black square with fewer than two black neighbors (out of eight) turns white in the next time step; a black square with two or three black neighbors stays black; a black square with more than three black neighbors turns white; and any white cell with three black neighbors turns black. This rule set generated complex and sometimes persistent patterns, which has made the game a longstanding object of mathematical study (Berlekamp, Conway, and Guy 2004).

The next step in the evolution of ABMs was the development of theory for interactive particle systems. Prominent probabilists involved in this work include Frank Spitzer and David Griffeath, who used methods from statistical mechanics to study system dynamics and phenomena such as phase changes (Griffeath 1993).

In the 1990s, ABMs broke away from mathematics and probability, becoming widely used in biology, ecology, economics, epidemiology, military strategy, political science, sociology, and transportation studies. This popularity was the result of many factors. One factor was that computation became both faster and more accepted as a scientific tool of inquiry. The increased computational power made it possible to simulate realistically complex scenarios. Another factor was that the reasoning process that underlies the creation of an ABM enabled domain experts to build models in a new way, without having to master sophisticated mathematics. A third factor was that several high-profile applications grabbed the imagination of those research communities.

Looking ahead, it is clear that ABMs will remain a popular tool across a wide range of scientific fields. They are also becoming accepted as a public policy tool, in large part because of their ability to perform “what if?” experiments. For example, the TRANSIMS ABM enables city planners to model the impact of a road closure, and possibly foresee unintended consequences. Such tools are being discussed at the national level in Sweden (Eliasson and Taymaz 1992) and in regional urban planning (Gilbert 2019; Li and Liu 2008, chap. 7) Chen and Zhan (2014) used them to evaluate urban evacuation plans, and Klabunde and Willekens (2016) used them to model international migrants.

Modern and complex ABMs typically have all or most of the following features:

- Many agents, often with differentiated roles and thus different rule sets.
- Rules, which may be complex. Sometimes the rules are heuristic, sometimes an ABM is based on randomized rules.
- Learning and adaptation. Agents learn about their environment (including other agents). Axelrod’s Prisoner’s Dilemma competition is an example (Axelrod 1980).
- An interaction space. This defines which agents affect each other—usually this is a model of propinquity, but for auctions it is a star-graph.
- A nonagent environment. This may include initial conditions, and/or background processes.

To illustrate these ideas, the following section describes three examples of ABMs. Following that is a discussion of the inferential strategies that the statistical community might apply to ABMs, and we close with a few observations.

2. Examples

We review three case studies in what follows. These case studies illustrate ideas and issues that arise in ABM applications.

2.1. Evacuating a Building

For safety purposes, civil engineers must study how long it takes a building with a specific design to be evacuated in an emergency. ABMs are a standard tool (Pelechano and Malkawi 2008; Zheng, Zhong, and Liu 2009). A procedure may involve 100 ABM simulations of the evacuation and result in a histogram of the time needed for the last agent to exit the building.

In this example, the interaction space is the virtually rendered architecture of the building being studied. The agents are “people” who are randomly distributed within the building. As an example of differentiated agent roles, some of the people may be “fire marshals” who direct others to safe stairwells. The rules for most agents can be quite simple: (i) When the alarm goes off, exit by the closest stairwell. (ii) If the stairs are blocked, try the next nearest stairwell. (iii) If a fire marshal directs you to a specific stairwell, use it. (iv) Move at a prespecified rate (perhaps different for different agents). Fire marshal rules could be specified similarly and comparably simple.

The nonagent environment in this example are such things as conditions that prevent a faster agent from walking through a slower agent on the stairs, and how to handle bottlenecks at doorways. Learning occurs when an agent discovers that a stairwell is blocked, or receives instruction from a fire marshal.

In this kind of simulation, one can model actions second-by-second and include a great deal of detail, such as the location of office furniture and agents with mobility challenges. However, as a general rule, one should avoid overcomplicating ABMs with unnecessarily granular detail.

2.2. Growing Artificial Societies

Sociologists became excited about ABMs in part because of an influential book, *Growing Artificial Societies: Social Science from the Bottom Up* (Epstein and Axtell 1996). The environment is a “sugarscape,” which is a Cartesian plane on which a resource, “sugar,” grows at a fixed rate at the lattice points. The agents are virtual people who stand at lattice points and consume the sugar until it is gone; then they search for a new lattice point at which to harvest sugar.

As an example, here are the first three of 17 rules used in Epstein and Axtell (1996).

1. *Sugarscape Growback*: At each lattice position, sugar grows back at a rate of α per time interval up to the capacity of that position.
2. *Agent Movement*: Look out as far as vision permits in each of the four lattice directions, north, south, east, and west:
 - Considering only unoccupied lattice positions, find the nearest position producing maximum welfare;
 - Move to the new position;
 - Collect all the resources at that location.
3. *Agent Mating*:
 - Select a neighboring agent at random;
 - If the neighboring agent is of the opposite sex *and* if both agents are fertile *and* at least one of the agents has an empty neighboring site, *then* a newborn is produced by crossing over the parents’ genetic and cultural characteristics;
 - Repeat for all neighbors.

Note that the first rule includes a tunable parameter; there are many such in the full rule list, and these are common in ABMs.

The first two rules govern search and movement; they generate simulated migratory patterns that mimic those seen in hunter-gatherer populations. The third rule for mating and reproduction produce age pyramids and population dynamics similar to those seen in humans.

Epstein and Axtell (1996) introduce a second resource, “spice,” and it, with additional rules, creates realistic barter economies. Ultimately, a total of 17 rules enable war, epidemics, division of labor, religion, and other rich behaviors. This kicked off a groundswell of interest in the sociological community, and now these ABM tools are used for many purposes, notably the study of social networks (Snijders 1996) and illegal drug markets (Dray et al. 2008; Romano, Lomax, and Richmond 2009).

In this application, the agents are differentiated (male/female, buyers/sellers), the rules are surprisingly simple, agents adapt to their environment in various ways, such as by movement and consumption. The interaction space is determined by the distance to nearby agents, and the nonagent environment is the sugarscape. This ABM has a compelling balance between simple rules and complex behavior.

2.3. Disease Spread and Epidemiology

A common application of ABMs is to study disease spread (Perez and Dragicevic 2009; Frias-Martinez, Williamson, and Frias-Martinez 2011). These are generally complementary to differential equation models and other alternatives (Bobashev et al. 2007; Hefley et al. 2017), and often enable greater realism and accuracy (Ajelli et al. 2010).

The general strategy is straightforward. One uses or models data on transportation flows within the geographic unit of interest (e.g., nursing home, city, nation, or world). The analyst picks disease parameters (e.g., age-specific infectiousness, incubation period, and mortality rate), and then seeds the community with an infected person, who travels within the geography, interacting with people in realistic ways. Often the interest is in the peak number of ill people, the duration of the outbreak, and the effect of mitigating measures, such as closing schools or social distancing.

In this example, the agents generally take different roles (e.g., children, physicians, people with more or less exposure). The rules mostly govern movement. The interactions are opportunities to spread the infection, behavioral adaptation includes such things as social distancing or going to a hospital, and the environment is the geography. Some ABM simulations are quite detailed, but generally the important questions can be addressed with relatively simple models.

The disease application indicates one of the limitations of ABMs. There were several ABMs for the 2014 ebola outbreak in Liberia (Merler et al. 2015; Siettos et al. 2015; Venkatramanan et al. 2018). Generally, such models did well in predicting how many villages would experience an outbreak in the next week, but they were poor at predicting which villages would suffer. This is reasonable, because a common transmission pattern in Liberia was someone working in Monrovia would start to feel ill, and then travel to their village so that family could care for them. The ABM could predict how many new cases would occur

in the city, but it was not able to forecast the home village of the infected person.

3. Inferential Strategies

Stochastic ABMs are models just like other statistical models but, with some exceptions, we cannot write down their likelihood functions. Therefore, we often have little theoretical guidance in estimating the parameters needed to fit the model to data, or in making quantified statements of uncertainty about model forecasts.

Sometimes ABM likelihoods are accessible, as we describe in what follows. But when the likelihood function is intractable, statisticians generally must use one of two tools. The first is statistical emulation, and the second is Approximate Bayesian Computation (ABC).

3.1. Tractable ABMs: Spread of Epidemiological and Ecological Processes

Many statistical models used in wildlife ecology to study population dynamics are individual-based and tractable using conventional statistical methods (i.e., maximum likelihood or Bayesian methods). For example, capture–recapture models are mature and provide inference about population abundance, survival, recruitment, and movement (e.g., King et al. 2009; Hooten, Wikle, and Schwob 2020). These models use data sources that arise from tracking (or relocating) animals over time to understand various demographic rates that can aid in the management and conservation of species. Most capture–recapture models are simple enough that they can be fit to data using conventional statistical methods. This is because they often lack components that allow for interaction among individuals, and hence individuals are assumed to be conditionally independent, which facilitates a hierarchical model structure that is amenable to implementation.

Like the capture–recapture class of models, the rapidly growing field of animal movement modeling treats dynamic individual-level processes explicitly (Hooten et al. 2017). Individual-based animal movement models could be referred to as ABMs, yet rarely are (but see Hooten et al. 2010b). In fact, population level inference using animal movement models often proceeds by treating individual-level parameters as random effects (Hooten et al. 2016). To formally scale individual-level processes up to the population level, many approaches use PDEs (Wikle 2003; Hooten and Wikle 2008; Lu et al. 2019). While these approaches to modeling collections of individuals have mechanistic mathematical underpinnings and have been simplified to be computationally tractable, they are often specified in terms of an Eulerian approximation to an underlying Lagrangian ABM for individual movement (Turchin 1998).

Despite their sophistication, most population-level models based on individual-level animal movement processes do not account for interactions among individuals beyond density dependence. For example, Lu et al. (2019) extended a spatio-temporal population model developed by Williams et al. (2017) to describe the colonization of Glacier Bay by sea otters that

incorporated regulated population growth based on density dependence. More complex interactions among individuals are the basis for realistic ABMs and have been accommodated in models for animal movement (e.g., Scharf et al. 2016, 2018). These models, often referred to as “collective” movement models (e.g., Couzin et al. 2005; Strandburg-Peshkin et al. 2015; Russell et al. 2017) have been used for a variety of applications but are not usually treated in a formal statistical framework. Moreover, connecting other important ecological and epidemiological processes with movement models may be challenging in conventional statistical frameworks.

Motivated by Smith et al. (2002), who employed a cellular automata model for the spread of rabies in raccoon populations throughout Connecticut, USA, Hooten and Wikle (2010) developed a statistical ABM for the spread of ecological and epidemiological processes. Instead of using a PDE as the basis for characterizing movement, their approach relied on fundamental dispersal processes of raccoons on a regular grid with resolution matching the approximate size of Connecticut townships. However, rather than use deterministic cellular automata, they specified the entire model in a hierarchical Bayesian framework that linked movement probabilities to a gradient surface based on spatial covariates to understand barriers to and corridors for disease spread. They fit the model using conventional Bayesian computational methods such as MCMC. In fact, Broms et al. (2016) generalized the statistical ABM to account for imperfect detection and used it to model the spread of an invasive bird in South Africa, fitting the model with JAGS (Plummer et al. 2003).

At the heart of these statistical ABMs is a mixture Bernoulli process for binary data (Wikle and Hooten 2015). For example, following Broms et al. (2016), to model the observed presence or absence of a species $y_{ij}(t)$ in region i during period j at time t , they accounted for the probability $p_{ij}(t)$ of detecting a species when it is present, by letting

$$y_{ij}(t) \sim \begin{cases} \text{Bern}(p_{ij}(t)) & \text{if } z_i(t) = 1 \\ 0 & \text{if } z_i(t) = 0 \end{cases}, \quad (1)$$

where $z_i(t)$ represents the true latent occupancy process of the species. This type of measurement process is commonly used in the so-called occupancy models (e.g., Hoeting, Leecaster, and Bowden 2000; MacKenzie et al. 2002; Johnson et al. 2013). A conditional binary model is often specified for the latent processes $z_i(t)$ so that the probability of occupancy depends on a set of spatially referenced predictor variables as in typical species distribution models (Hefley and Hooten 2016). Hooten and Wikle (2010) and Broms et al. (2016) treated the areal units as “agents” and modeled the process using a dynamic conditional specification such as

$$z_i(t) \sim \begin{cases} \text{Bern}(\phi_i(t)) & \text{if } z_i(t-1) = 1 \\ \text{Bern}(\theta_i(t)) & \text{if } z_i(t-1) = 0 \text{ but } z_{i'}(t-1) = 1 \\ & \text{for } i' \text{ a neighbor of } i \\ \text{Bern}(\psi_i(t)) & \text{if } z_i(t-1) = 0 \text{ and } z_{i'}(t-1) = 0 \\ & \text{for all } i' \text{ a neighbor of } i \end{cases} \quad (2)$$

where the probability $\phi_i(t)$ represents persistence or survival of the species or disease, $\theta_i(t)$ represents neighborhood-based dispersal probability, and $\psi_i(t)$ represents long-distance dispersal

probability. Hooten and Wikle (2010) specified the persistence and long-distance dispersal probabilities to be homogeneous, but allowed for heterogeneous short-distance dispersal processes by allowing $\theta_{i,t}$ to be the cumulative probability of immigrating individuals (or disease). This first requires an understanding of movement from the individual perspective and then aggregates that process to assess incoming individuals to new regions. Using similar principles of disease spread, Hooten, Anderson, and Waller (2010a) modeled influenza dynamics in the continental U.S. based on multiple mechanisms of human movement.

The ABM examples described in this section were all developed to facilitate fitting statistical models to data. However, such accommodations limit their potential realism to describe more complicated processes. Thus, we often seek inference in cases where it is computationally possible to simulate the process, but not trivial to express a joint or even conditional probability model for the data directly. This puts the statistician in an awkward position because they have access to a generative model for the data (as one does in a maximum likelihood or Bayesian setting), but they cannot evaluate a PDF or PMF associated with the model. In these cases, we must rely on some form of approximation that makes statistical inference possible. Either we must approximate the process itself which includes mechanisms in the ABM (i.e., the emulator approach), or we approximate the a Bayesian posterior distribution using an MCMC algorithm that substitutes a suitable distance metric for the likelihood (i.e., the ABC approach), or both.

3.2. Emulators

Emulators are approximations (or surrogates) to complex simulation models, either deterministic or stochastic (Gramacy 2020). Generally, emulators use Gaussian processes or treed Gaussian processes as the model (Gramacy and Lee 2008). These are Bayesian procedures first proposed by Kennedy and O’Hagan (2001), but have been subsequently extended and developed by many others (Bayarri et al. 2007; Higdon et al. 2008; Hooten, Anderson, and Waller 2010a; Smith 2013).

In our case, the ABM is the complex simulator that we seek to emulate. Farah (2014) examined such emulators in the context of an ABM for the spread of H1N1 influenza. Heard (2014) and Heard, Bobashev, and Morris (2014) compared emulators and ABC in the context of ABMs used to describe HIV spread and drug markets. They found that emulators were generally easier to implement and more robust.

Suppose the ABM takes input $\mathbf{x}_i$ and produces (potentially random) outputs $\mathbf{Y}_i$. The ABM contains parameters (e.g., the growth rate in the sugarscape and the rules that determine agent movement) that we denote by $\boldsymbol{\theta}$. Then the simulation output is modeled statistically using the emulator $\eta(\mathbf{x}_i, \boldsymbol{\theta})$ as follows:

$$\mathbf{Y}(\mathbf{x}_i) = \eta(\mathbf{x}_i, \boldsymbol{\theta}) + \delta(\mathbf{x}_i) + \epsilon(\mathbf{x}_i)$$

where the random term $\delta(\mathbf{x}_i)$ accounts for the discrepancy between the emulator and the simulator and $\epsilon(\mathbf{x}_i)$ is noise.

One goal is tune the parameter $\boldsymbol{\theta}$ to find $\boldsymbol{\theta}^*$ that best calibrates the emulator to the simulator. This purpose is to calibrate the model using data assimilation, a statistical concept. A second goal is to estimate the discrepancy function; this indicates where

the emulator performs poorly and may need more elaboration. This aspect enables one to infer credible regions and prediction intervals. Additional goals might include analogues of variable selection, estimation of variable importance, and insight into model dynamics.

The Kennedy-O'Hagan approach scales all inputs to the unit hypercube. Then the Bayesian version uses a Gaussian process to model the unknown function $\eta(\cdot, \cdot)$. In much of the related work, the Gaussian process is specified to have a constant mean function and a product covariance with power exponential form (Higdon et al. 2008)

$$\text{cov}[(\mathbf{x}, \boldsymbol{\theta}), (\mathbf{x}', \boldsymbol{\theta}')] = \lambda_{\eta}^{-1} R((\mathbf{x}, \boldsymbol{\theta}), (\mathbf{x}', \boldsymbol{\rho}_{\eta}))$$

where λ_{η} controls the marginal precision of $\eta(\cdot, \cdot)$ and $\boldsymbol{\rho}_{\eta}$ controls the strength of dependency in each component of $\mathbf{x}$ and $\boldsymbol{\theta}$. It is often useful to add a little white noise to the covariance model to account for small numerical fluctuations (from, say, adaptive meshing or convergence tolerances). The formulation of the prior is completed by specifying independent priors for the parameters controlling $\eta(\cdot, \cdot)$: the μ , λ_{η} , and $\boldsymbol{\rho}_{\eta}$.

A similar Gaussian process model is used for the discrepancy term $\delta(\mathbf{x})$. This has mean zero and covariance function

$$\text{cov}(\mathbf{x}, \mathbf{x}') = \lambda_{\delta} R((\mathbf{x}, \mathbf{x}'); \boldsymbol{\rho}_{\delta}).$$

As before, there are technical details regarding the specification of the covariance function.

This emulator structure enables one to use Markov chain Monte Carlo (MCMC) sampling to learn the posterior distributions for critical ABM output. Specifically, one obtains posterior distributions for

- the $\eta(\mathbf{x}, \boldsymbol{\theta})$, which is the hard-to-know implicit function calculated by the ABM;
- the optimal calibration parameter $\boldsymbol{\theta}^*$;
- the calibrated emulator $\eta(\mathbf{x}, \boldsymbol{\theta}^*)$;
- the physical system $\mathbf{Y}(\mathbf{x})$; and
- the discrepancy function $\delta(\mathbf{x})$.

The last, of course, is the most interesting from the standpoint of model validation. When and where this discrepancy function is large points out missing mechanisms or poor approximations.

Iterated application of this method permits successive estimation of the discrepancy function. One winds up with an approximation to the ABM in terms of simple functions, and the accuracy of the approximation can often be improved to the degree required by making the Gaussian process model more elaborate (e.g., by using a treed Gaussian process; Gramacy and Lee 2008).

In some ABM applications, one can go further in calibrating the ABM to reality. For example, in the case of weather forecasting, one has an ABM that predicts wind, humidity, and temperature in the future. Because we get to observe the weather on the following day, we can use essentially the same strategy outlined above to tune the emulator to the simulator to let us exploit the simulator to characterize real world data.

The Kennedy-O'Hagan approach is specified and implemented based on the second-moment dependence, but other emulator approaches have been developed based on the first-order structure (Hooten, Anderson, and Waller 2010a). Those

approaches using heterogeneous mean functions to mimic ABM processes may be faster and simpler to implement in high-dimensional settings.

3.3. Approximate Bayesian Computation

ABC is an alternative to emulators when the likelihood function is unavailable. First proposed in Rubin (1984), it did not achieve modern form until Tavaré et al. (1997). Although ABC antedates emulators, in many ways the methodology seems less mature.

The core idea in ABC for ABMs is that one has a sample $\mathbf{X}$ of data from the process that the ABM is modeling; e.g., meteorological data, in the context of a ABM that forecasts weather. The ABM has parameters $\boldsymbol{\theta}$ and the analyst has a prior $\pi(\boldsymbol{\theta})$ on those.

One then draws a value of $\boldsymbol{\theta}$ from the prior, and conditions the ABM on that draw. Next, one generates a sample $\mathbf{X}(\boldsymbol{\theta})$ from that data. For some appropriate metric $d(\cdot, \cdot)$ and tolerance γ , the analyst checks whether $d(\mathbf{X}, \mathbf{X}(\boldsymbol{\theta})) < \gamma$. If it is within the tolerance, then $\boldsymbol{\theta}$ is accepted; if not, it is rejected. Then a new $\boldsymbol{\theta}$ is drawn and the process repeats. The histogram of accepted values is approximately the posterior distribution that is wanted.

There are many reasons why this approach is problematic for ABMs. Often one does not have much (or any) data $\mathbf{X}$, as was the case with the Sugarscape example (Epstein and Axtell 1996). When there are data, they can be highly multivariate (e.g., temperature and rainfall measurements for all weather stations in a state). In that case, one must contend with the Curse of Dimensionality, or find some appropriate low-dimensional summary of the data, which in turn reduces the accuracy of the approximation. It is not obvious how to choose the metric or how to select the tolerance (Robert et al. 2011). For complex ABMs, it is difficult to develop a meaningful prior. Also, ABC is susceptible to pathologies (Berger et al. 2010).

ABC researchers have strategies for addressing all of these concerns, and ABC may be the right tool for some ABM applications. But, ABC could also more difficult to implement and less reliable than emulator methodology.

4. Model Validation

A key concern with ABMs is validating the simulation. Is the ABM elaborate enough to capture the full range of behaviors that are needed?

There are five main approaches to validation, not all of which are possible in some applications. We refer to the first as physics-based validation, in which the ABM encodes well-understood principles that govern emerging behavior. An example would be an ABM simulation of an n -body problem, in which objects in space interact according to gravitational attraction (Trenti and Hut 2008). If the physics are correct and complete, then the ABM is valid.

The second approach might be termed "intelligent design." Thoughtful experts encode all the factors that they believe are relevant, and hope they have not overlooked any important interactions. This is the most common method of validation; it is used by Hoffer, Bobashev, and Morris (2009) to simulate a drug

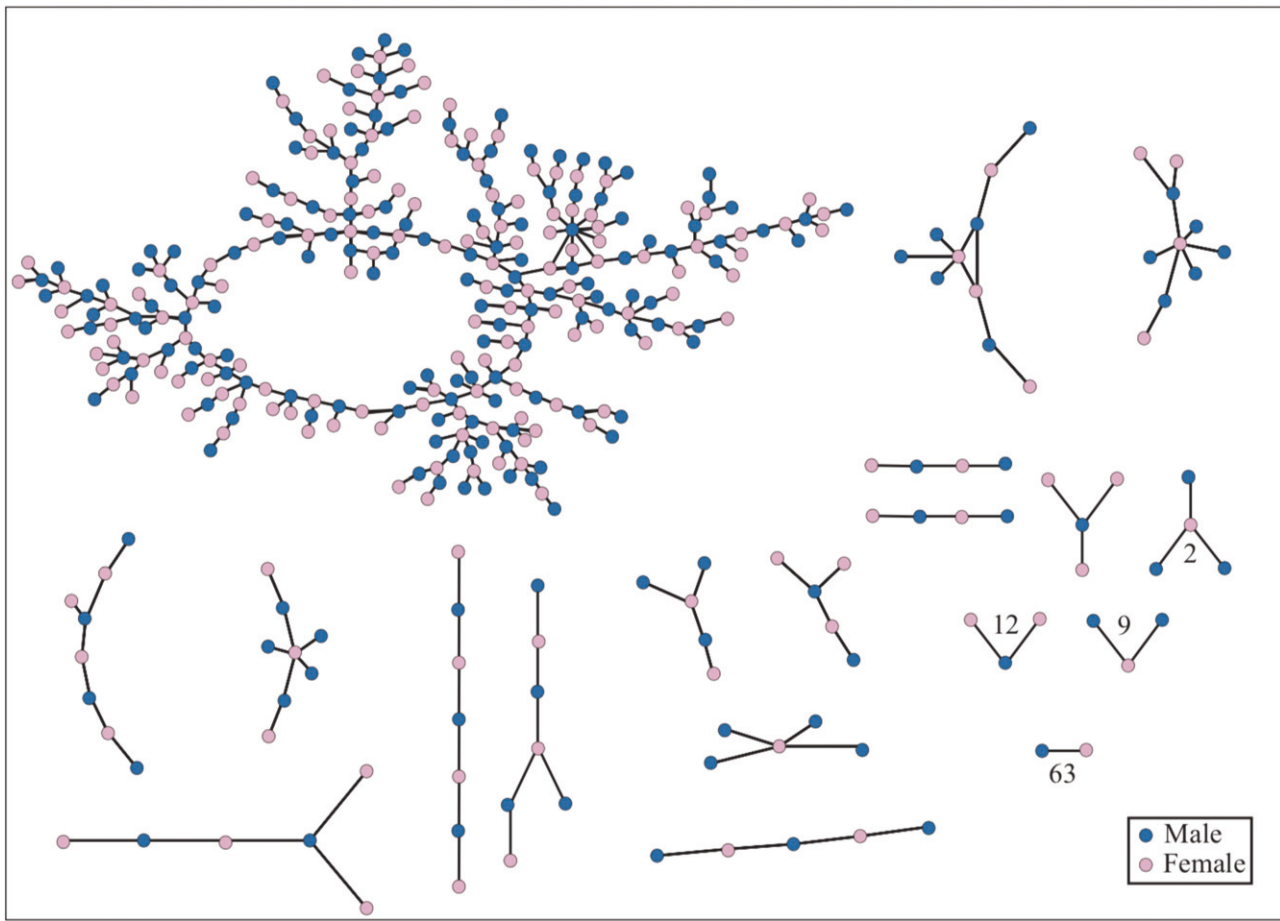


Figure 1. A network showing sexual relationships at an anonymized high school over the course of a year.

market and by Farah (2014) to describe disease spread. It may be about as reliable as software coding in some cases, because the thought processes are similar.

The third approach is face validity, and it is a true validation protocol. The designer runs the ABM with preselected inputs, ideally inputs for which there are known real-world outcomes, and evaluates the plausibility of the outputs. Such ABM testing is commonly used for battlefield simulation by the Defense Modeling and Simulation Office and the Naval Postgraduate School, among many others (Cioppa, Lucas, and Sanchez 2004).

The fourth method is more rare. The analyst compares the results from the ABM to predictions from another model (which may be an independently coded ABM, or some other forecast). One example is to compare an ABM model for disease spread to, say, the Kermack-McKendrick model, which employs a set of coupled ordinary differential equations to describe an epidemic curve (Brauer 2005). If these agree, the ABM seems sound.

Finally, in some cases, it is possible to compare the ABM outputs to real-world observations, as in weather forecasting (Simmonds, Gómez, and Ledezma 2019). This is the strongest form of validation, but it may still be unreliable if the analyst intends to use the ABM for inputs not previously seen (e.g., predicting weather outcomes from climate change).

As an example of the difficulty of model validation, consider the high-school social network of romantic relationships studied by Bearman, Moody, and Stovel (2004) shown in Figure 1.

These sociologists found that racial and smoking homophily predicted relationships, as did gender heterophily.

However, when they coded an ABM to simulate similar networks, conditioning on the number of edges and the estimated parameters in their logit model, the outputs looked like tangled balls of yarn rather than the filamentary structure seen in Figure 1. Then they realized that their data contain very few 4-cycles (a swap, in which two couples exchange partners). When they reran their ABM with an additional condition to exclude 4-cycles, the filamentary structure emerged. This, of course, does not prove that their ABM is correct, but illustrates the challenges of validation.

As a final comment on model validation, it is important to consider the intrinsic dimension of the ABM. Some analysts create elaborate ABMs, cluttered with code that is irrelevant to the emergent behavior. But that is not as great a problem as an ABM that omits structure needed to simulate the phenomenon, as with the initial ABM in Bearman, Moody, and Stovel (2004).

Every ABM has a set of parameters. The intrinsic parameters are those that control the outputs, and the rest affect the output to a lesser degree. For example, in an ABM for disease spread, it may be that the ABM models transmission in homes, at work, at the grocery store, at the laundry, and so forth. But it is plausible that the parameters driving transmission at the laundry are largely irrelevant to predicting the course of the epidemic.

Using methods in Banks and Olszewski (1997), one can use principal components regression to estimate the local intrinsic

dimension of an ABM. This approach can inform the analyst about whether the ABM is underparameterized or overparameterized.

5. Conclusions

We have reviewed ABMs and their inherent statistical challenges as a call to action for the statistical community. ABMs are used in many disciplines, but few statisticians are addressing their theoretical foundations. If we are to stay relevant for many domains, we need to address this research topic.

ABMs will be used for the foreseeable future. They are easy to program, and domain experts can embed their beliefs in straightforward ways. But as we suggest, those reasons are insufficient. In most cases, researchers do not know how to best fit ABMs to data, or how to make quantified statements of uncertainty about their predictions. And in general, we do not have an agreed upon procedure for validating ABMs.

Most ABMs do not have tractable likelihood functions. In those cases, statisticians have limited tools, including emulators and ABC. Emulators have been popular for implementing ABMs and other mechanistic process models, but ABC and related methods may be helpful in certain cases (Hooten, Wikle, and Schwob 2020).

Acknowledgments

The authors thank the Statistics and Applied Mathematics Sciences Institute (SAMSI) program on Model Uncertainty: Mathematical and Statistical (MUMS). Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Funding

This research was funded by NSF DEB 1927177 and NSF DMS 1614392.

References

- Ajelli, M., Gonçalves, B., Balcan, D., Colizza, V., Hu, H., Ramasco, J. J., Merler, S., and Vespignani, A. (2010), "Comparing Large-scale Computational Approaches to Epidemic Modeling: Agent-based Versus Structured Metapopulation Models," *BMC Infectious Diseases*, 10, 190, 1–13. [237]
- Axelrod, R. (1980), "Effective Choice in the Prisoner's Dilemma," *Journal of Conflict Resolution*, 24, 3–25. [236]
- Banks, D., and Olszewski, R. (1997), "Estimating Local Dimensionality," in *Proceedings of the Statistical Computing Section of the American Statistical Association*, ASA. [240]
- Bayarri, M. J., Berger, J. O., Paulo, R., Sacks, J., Cafeo, J. A., Cavendish, J., Lin, C.-H., and Tu, J. (2007), "A Framework for Validation of Computer Models," *Technometrics*, 49, 138–154. [238]
- Bearman, P. S., Moody, J., and Stovel, K. (2004), "Chains of Affection: The Structure of Adolescent Romantic and Sexual Networks," *American Journal of Sociology*, 110, 44–91. [240]
- Berger, J. O., Fienberg, S. E., Raftery, A. E., and Robert, C. P. (2010), "Incoherent Phylogeographic Inference," *Proceedings of the National Academy of Sciences*, 107, pp. E157–E157. [239]
- Berlekamp, E. R., Conway, J. H., and Guy, R. K. (2004), *Winning Ways for Your Mathematical Plays* (Vol. 4), Oxford, UK: AK Peters/CRC Press. [235]
- Bobashev, G. V., Goedecke, D. M., Yu, F., and Epstein, J. M. (2007), "A Hybrid Epidemic Model: Combining the Advantages of Agent-based and Equation-based Approaches," in 2007 Winter Simulation Conference, 1532–1537. IEEE. [237]
- Brauer, F. (2005), "The Kermack–McKendrick Epidemic Model Revisited," *Mathematical Biosciences*, 198, 119–131. [240]
- Broms, K. M., Hooten, M. B., Johnson, D. S., Altwegg, R., and Conquest, L. L. (2016), "Dynamic Occupancy Models for Explicit Colonization Processes," *Ecology*, 97, 194–204. [238]
- Chen, X., and Zhan, F. B. (2014), "Agent-based Modeling and Simulation of Urban Evacuation: Relative Effectiveness of Simultaneous and Staged Evacuation Strategies," *Agent-Based Modeling and Simulation*, 78–96. [236]
- Cioppa, T. M., Lucas, T. W., and Sanchez, S. M. (2004), "Military Applications of Agent-based Simulations," in *Proceedings of the 2004 Winter Simulation Conference*, 2004 (Vol. 1). IEEE, Thousand Oaks CA: Sage. [240]
- Couzin, I. D., Krause, J., Franks, N. R., and Levin, S. A. (2005), "Effective Leadership and Decision-making in Animal Groups on the Move," *Nature*, 433, 513–516. [238]
- Dray, A., Mazerolle, L., Perez, P., and Ritter, A. (2008), "Drug Law Enforcement in an Agent-based Model: Simulating the Disruption to Street-level Drug Markets," in *Artificial Crime Analysis Systems: Using Computer Simulations and Geographic Information Systems*, eds. L. Liu and J. Eck, 352–371. Hershey, PA: IGI Global. [237]
- Eliasson, G., and Taymaz, E. (1992), "The Limits of Policy Making: An analysis of the Consequences of Boundedly Rational Government Using the Swedish Micro-to-macro Model (MOSES)," Technical report, IUI Working Paper. [236]
- Epstein, J. M., and Axtell, R. (1996), *Growing Artificial Societies: Social Science from the Bottom Up*, Cambridge, MA: Brookings Institution Press. [236,237,239]
- Farah, F., Birrel, P., Conti, S., Agnes, D. D. (2014), "Bayesian Emulation and Calibration of a Dynamic Epidemic Model for A/H1N1 Influenza," *Journal of the American Statistical Association*, 109, 1398–1411. [238,240]
- Frias-Martinez, E., Williamson, G., and Frias-Martinez, V. (2011), "An Agent-based Model of Epidemic Spread Using Human Mobility and Social Network Information," 57–64. IEEE. [235,237]
- Gardner, M. (1970), "Mathematical Games-The Fantastic Combinations of John Conway's New Solitaire Game, Life, 1970," *Scientific American*, October, pp. 120–123. [235]
- Gilbert, N. (2019), *Agent-Based Models*, Thousand Oaks, CA: Sage Publications, Incorporated. [236]
- Gramacy, R. B. (2020), *Surrogates: Gaussian Process Modeling, Design, and Optimization for the Applied Sciences*, Boca Raton: CRC Press. [238]
- Gramacy, R. B., and Lee, H. K. H. (2008), "Bayesian Treed Gaussian Process Models With an Application to Computer Modeling," *Journal of the American Statistical Association*, 103, :1119–1130. [238,239]
- Griffeath, D. (1993), "Frank Spitzer's Pioneering Work on Interacting Particle Systems," *The Annals of Probability*, 21, 608–621. [235]
- Heard, D. (2014), "Statistical Inference Utilizing Agent Based Models" (doctoral dissertation). Duke University. [238]
- Heard, D., Bobashev, G. V., and Morris, R. J. (2014), "Reducing the Complexity of an Agent-based Local Heroin Market Model," *PLoS One*, 9, e102263. [238]
- Hefley, T. J., and Hooten, M. B. (2016), "Hierarchical Species Distribution Models," *Current Landscape Ecology Reports*, 1, 87–97. [238]
- Hefley, T. J., Hooten, M. B., Russell, R. E., Walsh, D. P., and Powell, J. A. (2017), "When Mechanism Matters: Bayesian Forecasting Using Models of Ecological Diffusion," *Ecology Letters*, 20, 640–650. [237]
- Higdon, D., Gattiker, J., Williams, B., and Rightley, M. (2008), "Computer Model Calibration Using High-dimensional Output," *Journal of the American Statistical Association*, 103, 570–583. [238,239]
- Hoeting, J. A., Leecaster, M., and Bowden, D. (2000), "An Improved Model for Spatially Correlated Binary Responses," *Journal of Agricultural, Biological, and Environmental Statistics*, 5, 102–114. [238]
- Hoffer, L. D., Bobashev, G., and Morris, R. J. (2009), "Researching a Local Heroin Market as a Complex Adaptive System," *American Journal of Community Psychology*, 44, 273. [239]
- Hooten, M., Wikle, C., and Schwob, M. (2020), "Statistical Implementations of Agent-based Demographic Models," *International Statistical Review*, 88, 441–461. [235,237,241]
- Hooten, M. B., Anderson, J., and Waller, L. A. (2010a), "Assessing North American Influenza Dynamics With a Statistical SIRS Model," *Spatial and Spatio-Temporal Epidemiology*, 1, 177–185. [238,239]

- Hooten, M. B., Buderman, F. E., Brost, B. M., Hanks, E. M., and Ivan, J. S. (2016), "Hierarchical Animal Movement Models for Population-level Inference," *Environmetrics*, 27, 322–333. [237]
- Hooten, M. B., Johnson, D. S., Hanks, E. M., and Lowry, J. H. (2010b), "Agent-based Inference for Animal Movement and Selection," *Journal of Agricultural, Biological and Environmental Statistics*, 15, 523–538. [237]
- Hooten, M. B., Johnson, D. S., McClintock, B. T., and Morales, J. M. (2017), *Animal Movement: Statistical Models for Telemetry Data*, Boca Raton, FL: CRC Press. [237]
- Hooten, M. B., and Wikle, C. K. (2008), "A Hierarchical Bayesian Non-linear Spatio-temporal Model for the Spread of Invasive Species With Application to the Eurasian Collared-Dove," *Environmental and Ecological Statistics*, 15, 59–70. [237]
- Hooten, M. B., and Wikle, C. K. (2010), "Statistical Agent-based Models for Discrete Spatio-temporal Systems," *Journal of the American Statistical Association*, 105, 236–248. [235,238]
- Hunter, E., Mac Namee, B., and Kelleher, J. (2018), "An Open-data-driven Agent-based Model to Simulate Infectious Disease Outbreaks," *PLoS One*, 13, e0208775. [235]
- Johnson, D. S., Conn, P. B., Hooten, M. B., Ray, J. C., and Pond, B. A. (2013), "Spatial Occupancy Models for Large Data Sets," *Ecology*, 94, 801–808. [238]
- Kennedy, M. C., and O'Hagan, A. (2001), "Bayesian Calibration of Computer Models," *Journal of the Royal Statistical Society, Series B*, 63, 425–464. [238]
- King, R., Morgan, B., Gimenez, O., and Brooks, S. (2009), *Bayesian Analysis for Population Ecology*, Boca Raton, FL: CRC Press. [237]
- Klabunde, A., and Willekens, F. (2016), "Decision-making in Agent-based Models of Migration: State of the Art and Challenges," *European Journal of Population*, 32, 73–97. [236]
- Li, X., and Liu, X. (2008), "Embedding Sustainable Development Strategies in Agent-based Models for Use as a Planning Tool," *International Journal of Geographical Information Science*, 22, 21–45. [236]
- Lu, X., Williams, P. J., Hooten, M. B., Powell, J. A., Womble, J. N., and Bower, M. R. (2019), "Nonlinear Reaction–diffusion Process Models Improve Inference for Population Dynamics," *Environmetrics*, 31, e2604. [237]
- MacKenzie, D. I., Nichols, J. D., Lachman, G. B., Droege, S., Andrew Royle, J., and Langtimm, C. A. (2002), "Estimating Site Occupancy Rates When Detection Probabilities Are Less Than One," *Ecology*, 83, 2248–2255. [238]
- Merler, S., Ajelli, M., Fumanelli, L., Gomes, M. F., Piontti, A. P. y., Rossi, L., Chao, D. L. I., Longini, Jr, M., Halloran, M. E., and Vespignani, A. (2015), "Spatiotemporal Spread of the 2014 Outbreak of Ebola Virus Disease in Liberia and the Effectiveness of Non-pharmaceutical Interventions: A Computational Modelling Analysis," *The Lancet Infectious Diseases*, 15, 204–211. [237]
- Pelechano, N., and Malkawi, A. (2008), "Evacuation Simulation Models: Challenges in Modeling High Rise Building Evacuation With Cellular Automata Approaches," *Automation in Construction*, 17, 377–385. [236]
- Perez, L., and Dragicevic, S. (2009), "An Agent-based Approach for Modeling Dynamics of Contagious Disease Spread," *International Journal of Health Geographics*, 8, 50. [235,237]
- Plummer, M., et al. (2003), "JAGS: A Program for Analysis of Bayesian Graphical Models Using Gibbs Sampling," in *Proceedings of the 3rd International Workshop on Distributed Statistical Computing* (Vol. 124). Berlin, Germany: Springer. pp. 1–10. [238]
- Robert, C. P., Cornuet, J.-M., Marin, J.-M., and Pillai, N. S. (2011), "Lack of Confidence in Approximate Bayesian Computation Model Choice," *Proceedings of the National Academy of Sciences*, 108, 15112–15117. [239]
- Romano, D. M., Lomax, L., and Richmond, P. (2009), "NARCSim an agent-based Illegal Drug Market Simulation," in 2009 International IEEE Consumer Electronics Society's Games Innovations Conference, Thousand Oaks, CA: Sage. pp. 101–108. IEEE. [237]
- Rubin, D. B. (1984), "Bayesianly Justifiable and Relevant Frequency Calculations for the Applied Statistician," *The Annals of Statistics*, 12, 1151–1172. [239]
- Russell, J. C., Hanks, E. M., Modlmeier, A. P., and Hughes, D. P. (2017), "Modeling Collective Animal Movement Through Interactions in Behavioral States," *Journal of Agricultural, Biological and Environmental Statistics*, 22, 313–334. [238]
- Scharf, H. R., Hooten, M. B., Fosdick, B. K., Johnson, D. S., London, J. M., Durban, J. W. (2016), "Dynamic Social Networks Based on Movement," *The Annals of Applied Statistics*, 10, 2182–2202. [238]
- Scharf, H. R., Hooten, M. B., Johnson, D. S., and Durban, J. W. (2018), "Process Convolution Approaches for Modeling Interacting Trajectories," *Environmetrics*, 29, e2487. [238]
- Siettos, C., Anastassopoulou, C., Russo, L., Grigoras, C., and Mylonakis, E. (2015), "Modeling the 2014 Ebola Virus Epidemic-agent-based Simulations, Temporal Analysis and Future Predictions for Liberia and Sierra Leone," *PLoS Currents*, 7. [237]
- Simmonds, J., Gómez, J. A., and Ledezma, A. (2019), "The Role of Agent-based Modeling and Multi-agent Systems in Flood-based Hydrological Problems: A Brief Review," *Journal of Water and Climate Change*, 11, 1580–1602. [235,240]
- Smith, D. L., Lucey, B., Waller, L. A., Childs, J. E., and Real, L. A. (2002), "Predicting the Spatial Dynamics of Rabies Epidemics on Heterogeneous Landscapes," *Proceedings of the National Academy of Sciences*, 99, 3668–3672. [238]
- Smith, R. (2013), *Uncertainty Quantification: Theory, Implementation, and Applications* (Vol. 12). Philadelphia, PA: SIAM. [238]
- Snijders, T. A. (1996), "Stochastic Actor-oriented Models for Network Change," *Journal of Mathematical Sociology*, 21, 149–172. [235,237]
- Strandburg-Peshkin, A., Farine, D. R., Couzin, I. D., and Crofoot, M. C. (2015), "Shared Decision-making Drives Collective Movement in Wild Baboons," *Science*, 348, 1358–1361. [238]
- Tavaré, S., Balding, D. J., Griffiths, R. C., and Donnelly, P. (1997), "Inferring Coalescence Times from DNA Sequence Data," *Genetics*, 145, 505–518. [239]
- Trenti, M., and Hut, P. (2008), "N-body Simulations (Gravitational)," *Scholarpedia*, 3, 3930. [239]
- Turchin, P. (1998), *Quantitative Analysis of Movement*, New York, NY: Sinauer Associates. [237]
- Venkatramanan, S., Lewis, B., Chen, J., Higdon, D., Vullikanti, A., and Marathe, M. (2018), "Using Data-driven Agent-based Models for Forecasting Emerging Infectious Diseases," *Epidemics*, 22, 43–49. [237]
- Von Neumann, J., Burks, A. W. et al. (1966), "Theory of Self-reproducing Automata," *IEEE Transactions on Neural Networks*, 5, 3–14. [235]
- Wikle, C., and Hooten, M. (2015), "Hierarchical Agent-based Spatio-temporal Dynamic Models for Discrete Valued Data," in *Handbook of Discrete-Valued Time Series*, chapter 16. Boca Raton, FL: Chapman and Hall/CRC Press. [238]
- Wikle, C. K. (2003), "Hierarchical Bayesian Models for Predicting the Spread of Ecological Processes," *Ecology*, 84, 1382–1394. [237]
- Williams, P. J., Hooten, M. B., Womble, J. N., Esslinger, G. G., Bower, M. R., and Hefley, T. J. (2017), "An Integrated Data Model to Estimate Spatiotemporal Occupancy, Abundance, and Colonization Dynamics," *Ecology*, 98, 328–336. [237]
- Zheng, X., Zhong, T., and Liu, M. (2009), "Modeling Crowd Evacuation of a Building Based on Seven Methodological Approaches," *Building and Environment*, 44, 437–445. [236]