

A hybrid modeling approach for parking and traffic prediction in urban simulations

Rahmatollah Beheshti · Gita Sukthankar

Received: 1 March 2013 / Accepted: 28 November 2013
© Springer-Verlag London 2014

Abstract Urban simulations are an important tool for analyzing many policy questions relating to the usage of public space, roads, and communal transportation; they can be used to predict the long-term impact of new construction projects, traffic restrictions, and zoning laws. However, it is unwise to rely upon predictions from a single model since each technique possesses different strengths and weaknesses and can be highly sensitive to the choice of parameters and initial conditions. In this article, we describe a hybrid approach for combining agent-based and stochastic simulations (Markov chain Monte Carlo, MCMC) to improve the accuracy and reduce the variance of long-term predictions. In our proposed approach, the agent-based model is used to bootstrap the proposal distribution for the MCMC estimator. To demonstrate the applicability of our modeling technique, this article presents a case study describing the usage of our hybrid simulation method for forecasting transportation patterns and parking lot utilization on a large university campus. A comparison of our simulation results against an independently collected dataset reveals that our hybrid approach accurately predicts parking lot usage and performs significantly better than other comparable modeling techniques. Developing novel architectures for combining the predictions of agent-based models can produce insights that are different than simply selecting the best model.

Keywords Agent-based modeling · Markov chain Monte Carlo (MCMC) · Transportation simulation · Urban modeling

1 Introduction

Benenson et al. (2004) present two motivations for defining urban agents as a distinct group within the general class of autonomous agents:

1. urban agents often have a high degree of mobility resulting in rapidly changing spatial relationships.
2. to succeed, urban agents require a strong capability to perceive and adapt to the evolving urban environment shaped by neighboring agents.

Urban simulation is a particularly fertile domain for research in agent-based simulation since it requires modeling a large number of interdependent agents making sequential decisions within a small region. Agent-based models have been used specifically to recreate urban environments for a wide variety of domains including the following: (1) civil and environmental transportation analysis (Jin and Jie 2012), (2) geographic information systems (GIS) for visualizing patterns and trends in spatial areas (Jordan et al. 2012), and (3) archaeological studies of land site usage in ancient civilizations (Kohler et al. 2012).

Although these urban simulations do not necessarily have to model low-level physical interactions, incorporating spatial information and heterogeneity into agent-based models can improve our ability to draw conclusions about the behavior of complex systems in realistic environments, which may be different from conclusions drawn with artificial environments (Brown et al. 2005). With the inclusion of GIS to represent a spatially, georeferenced

R. Beheshti (✉) · G. Sukthankar
Department of EECS, University of Central Florida,
Orlando, FL 32816, USA
e-mail: beheshti@knights.ucf.edu

G. Sukthankar
e-mail: gitars@eeecs.ucf.edu

environment, the impact of human behavior patterns can be linked to specific spatial locations and when used correctly can provide a powerful tool for policy-makers and the public to understand the potential consequences of their decisions (Gimblett 2002).

Yet, modelers attempting to analyze a complex urban region face a similar problem to the six blind men touching an elephant, who describe the whole elephant based on touching it. Since none of the men can feel more than a single small part of the elephant—the tail, the ear, a tusk, the belly, the trunk, and the leg—they each bring back a different report.¹ In the same way, different modeling techniques are very likely to produce slightly different answers to the same question. This phenomenon poses problems when urban simulations are used to influence important public policy debates, regulatory decisions, and to guide resource allocation. For instance, the public debate about human influence on climate change has been shaped by a disproportionate level of discussion about minor discrepancies between predictions, while the general trend consensus between models has remained largely ignored (Edwards 1999).

In this article, we introduce a novel architecture for combining two powerful modeling techniques: agent-based models (ABM) and Markov chain Monte Carlo (MCMC) estimators. Although both of these methods have a long history of practical usage (summarized in the next two sections), they have weaknesses as well. ABMs can be used to simulate very complex social phenomena, but constructing easily reproducible agent-based models is difficult due to the possibility of emergent behaviors and lack of formal representation. According to Wilensky and Rand (2007), many ABMs, with the exception of a few classic models, have never been replicated by anyone but the original developer. It is difficult to bring mathematical analysis tools to bear on the problem, so instead, models are typically studied through empirical simulation studies (Hinkelmann et al. 2011). Yet, the results of the simulation study can vary considerably by changing the range, or even the step size, of just one or two variables (Niazi et al. 2009).

On the other hand, the MCMC simulation process can be described by a relatively simple set of mathematical equations and a resampling procedure; this methodology is sometimes referred as the most powerful idea in computational statistics (Press et al. 2007). The aim of the process is to approximate the posterior distribution of the model parameters based on the observed data. However, the selection of the proposal distribution can have a significant impact on model convergence. In cases where the proposal

distribution is far from the desired posterior distribution, the algorithm may converge to a poor local minimum or require a long time to achieve convergence (Gilks et al. 1995). The nearer the proposal distribution is to the target distribution, the better the performance of the MCMC algorithm (Mengersen and Tweedie 1996). The reader can find more details about the role of proposal distribution in Andrieu et al. (2003).

To address this problem, we constructed an agent-based model to generate simulated data which is then used to initialize the proposal distribution of the MCMC. The combination of the two models, agent-based and MCMC, produces a more accurate result than either of the parent models and facilitates the MCMC convergence. An additional benefit is that manipulating the operation of an agent-based model can empower researchers with better intuitions about the reasons behind emerging group phenomena rather than merely observing the unfolding of a stochastic process (Oakes 2008). To demonstrate the strengths of this approach, we present a case study on modeling and predicting transportation patterns and parking lot usage on a large university campus.

This article is organized as follows. In the next section, we provide a brief overview of the related work in using agent-based models to guide public policy decisions. Section 3 introduces the MCMC family of methods. Then, we summarize the key elements of our urban simulation for forecasting transportation patterns and parking lot utilization in Sect. 4. Section 5 presents our hybrid approach for unifying agent-based and MCMC models. We conclude the article by presenting a detailed comparison of our method versus other similar modeling techniques.

2 Related work

Agent-based modeling has been used to analyze a variety of complex public policy-related scenarios, including climate change negotiations (Gerst et al. 2012), water management decisions (López-Paredes et al. 2005), and financial regulatory governance (Streit and Borenstein 2009). In general, ABMs are good at modeling a diverse population of rational, self-interested agents, allowing interesting social questions to be explored in simulation before enacting new laws. For instance, Garlick and Chli (2009) studied the effects of social influence and curfews on civil violence by creating an agent-based model that simulated the interactions between the police force and the community. Some social simulations explicitly model network interactions between agents; this is particularly useful when studying influence propagation (Maghami and Sukthankar 2012) and the self-repairing properties of insurgent terrorist networks (Ilachinski 2012). Social

¹ The parable of the blind men and the elephant appears in a number of religions originating from the Indian subcontinent.

choice mechanisms can be studied using agent-based simulations as well as by game theory; for instance, Verella and Wardak (2008) examined the effects of external stimuli on collective opinion formation, in the context of voting decisions.

Alternatively, interactions between agents can be governed by a combination of spatial and social constraints; in these social systems, the behavior of the agents is strongly affected by other agents in their local physical neighborhood, which is easy to simulate, but often difficult to predict analytically. Examples of systems possessing these characteristics include traffic and crowd evacuation simulations, which are heavily influenced by geographic considerations. To compare the effectiveness of simultaneous and staged evacuation strategies in different road network structures, Chen (2003) designed an agent-based simulation that shows the collective behaviors resulting from the interactions of individual vehicles during an evacuation. Human behavioral data can be added to the emergency evacuation and egress model to build a more realistic, and consistent agent-based model was done by Pan et al. (2007).

In contrast to crowd evacuation scenarios which are often used to prepare for unique disaster situations, traffic simulations are designed to characterize the effects of repetitive behaviors. Klügl and Bazzan (2012) outline five advantages agent-based methodologies have over other types of traffic-related simulations including the following: (1) ease of modeling bottom-up decision-making, (2) capacity for imbuing entities with learning and adaptive behavior, and (3) simplicity of generating a population with heterogeneous behaviors. Also, it is often feasible to gather survey and GPS data to verify the predictions of traffic simulations.

Based on detailed trip survey data from seven Traffic Analysis Zones (TAZs) in Ottawa, Canada, Jin and White (2012) present an agent-based model for analyzing the influence of neighborhood design on daily trip patterns. In Dia (2002), results obtained from a behavioral survey of driving behaviors were used to identify and fit a series of agent behavior parameters defining driver characteristics, knowledge, and preferences; the authors also present a case study implementing a simple agent-based route choice decision model within a microscopic traffic simulation tool. However, neither of those works presents a systematic evaluation of different modeling techniques through comparison with independently collected data. In our research, a physical path planning system for modeling driving and walking is used to supplement the activity-based microsimulation that governs agent behavior selection. The model is seeded with a combination of demographic information and survey data, and compared against independently collected results. A detailed review of the

applications of agent-based modeling specially in modeling traffic and transportation patterns can be found in Chen and Cheng (2010).

ABMs have been successfully employed in a variety of water management tasks (Barreteau et al. 2012; Gaillard et al. 2012). Water management, an important aspect of urban management, is affected by geography, weather patterns, and human behavior, and is additionally complicated by interdependencies between communities that share the same watershed area. López-Paredes et al. (2005) introduced an agent-based simulator called FIRMABAR for integrated freshwater assessment of the Valladolid metropolitan area. The simulator provides the policy-makers with a tool to evaluate alternative water policies in different scenarios.

Similar simulations can be used to study the combined impact of climate change and human behavior on sustainable ecosystems. Hailegiorgis et al. (2010) presented an agent-based system for modeling interactions between climate change and conflict among herders in east Africa. ENGAGE is an agent-based model that was introduced by Gerst et al. (2012) to simulate the impact of locally heterogeneous policy preferences and constituent choice on climate change negotiation at the international level. A review of related work in this area can be found in Balbi and Giupponi (2009).

In summary, agent-based models can be used to illuminate policy-makers on the ramifications of complex environmental and infrastructure decisions. In the next section, we describe the second key component of our approach, MCMC.

3 Markov chain Monte Carlo

Markov chain Monte Carlo is a family of methods principally used to perform Bayesian inference with stochastic simulation. The aim of the process is to approximate the posterior distribution of the model parameters based on the observed data. By using Monte Carlo simulations to perform the high-dimensional integration necessary to calculate marginal and posterior distributions, algorithms such as Metropolis–Hastings (MH) can make the Bayesian inference process tractable (Metropolis et al. 1953). The MH algorithm is the oldest and perhaps most commonly used of these methods. The basic procedure is as follows:

- Select a proposal distribution Q (also known as the proposal transition matrix)
- Initialize the starting point, x_0
- Do
- Generate a candidate point x_c , according to the probability $Q(x_c|x_i)$

- Calculate the acceptance probability according to

$$\alpha(x_i, x_c) = \min\left(1, \frac{\pi(x_c)q(x_i|x_c)}{\pi(x_i)q(x_c|x_i)}\right) \quad (1)$$

- Choose $x_{i+1} = x_c$ with probability α , $x_{i+1} = x_i$ with probability $(1 - \alpha)$

Effectively, MCMC allows us to draw samples from a distribution $\pi(x)$ without having to know its normalization. With these samples, it is possible to compute any quantity of interest about the distribution of x , such as means, confidence regions, or covariance.

Markov chain Monte Carlo has been successfully used in a wide variety of scientific (Liu et al. 2011) and engineering modeling applications (Liu et al. 2012). MCMC can also be applied as part of the model fitting process in social prediction problems. For instance, Cauchemez et al. (2004) use a Bayesian MCMC approach to examine the main characteristics that affect influenza disease transmission between households. Similarly, the effect of spatial influences on geopolitical conflicts has been modeled using an MCMC formulation in which the likelihood of war involvement for each nation is conditioned on the decisions of proximate states (Ward and Gleditsch 2002). In our work, MCMC is used as a simulation technique, and the sample set used to characterize the posterior distribution is simply compared against the output of other simulation techniques, rather than used to perform Bayesian inference over model parameters.

Our research focuses on improving the performance of the Metropolis-Hastings (MH) algorithm which is relatively sensitive to the initial proposal distribution. It is because of this sensitivity that researchers sometimes opt to use alternative MCMC algorithms, such as Gibbs sampling (Geman and Geman 1984). Our proposed method is a variation on the idea of using suboptimal inference and learning algorithms to generate data-driven proposal distributions for the MH algorithm (Andrieu et al. 2003). An alternate approach for creating MCMC proposal distributions was introduced by Eaton and Murphy (2007) who employed dynamic programming to create a proposal distribution for MCMC in the space of directed acyclic graphs. They showed that this hybrid technique converges to the posterior faster than other methods, resulting in a more accurate structure learning of graphical models and higher predictive likelihoods on test data.

De Freitas et al. (2001) introduced two different methods to overcome the problem of finding a good proposal distribution. In the first approach, a mixture of two kernels is used to drive the search process: (1) a variational kernel to broadly explore the problem domain and locate regions of high-probability and (2) a Metropolis kernel to explore the local regions. One drawback with this method is that finding a good variational kernel can be difficult to do.

To combat this issue, the authors proposed a second technique called adaptive MCMC in which the proposal distribution is updated at run-time based on the behavior of Markov chain; in this article, we benchmark our proposed method against adaptive MCMC. Our approach solves the problem of identifying a good proposal distribution for MCMC by constructing one from samples generated by our agent-based model. Adaptive methods generally seek to construct a better proposal distribution by combining stochastic approximation and MCMC (Andrieu and Moulines 2006). One issue with this class of adaptive techniques is that they often rely on certain mathematical assumptions being valid and thus can only be used in a limited set of conditions unlike our technique.

4 Urban simulation

In this section, we describe the development process for our activity-based microsimulation, including the agent-based model, survey data collection, activity profile generation, path planning, and simulation system; see Beheshti and Sukthankar (2012) for additional details on the data collection and model fitting procedures. For our urban region, we selected the University of Central Florida main campus, which is one the largest academic institutions in the USA with almost 59,000 students and 10,567 staff. It is adjacent to the Central Florida Research Park which is home to 116 companies with approximately 9,500 employees. The presence of nearby businesses and existence of commuters traveling between multiple UCF campuses give rise to a social system with a diverse and complex set of transportation patterns.

4.1 Data collection

To simplify the data collection process, our initial study focused solely on modeling student transportation, dining, and building occupancy patterns. About 1003 students responded to our online survey posted on KwikSurveys that was advertised on various campus email lists. The questions on the survey were grouped into six different categories, related to possible places that could be visited on the main campus:

1. Daily attendance patterns, including the days and times that the participant arrives at and departs from the main campus
2. Initial location, either the dorm (for on-campus students) or the entrance that was used to enter the campus (for commuting students)
3. Visitation frequency for on-campus dining locations
4. Usage patterns for recreation and athletic facilities

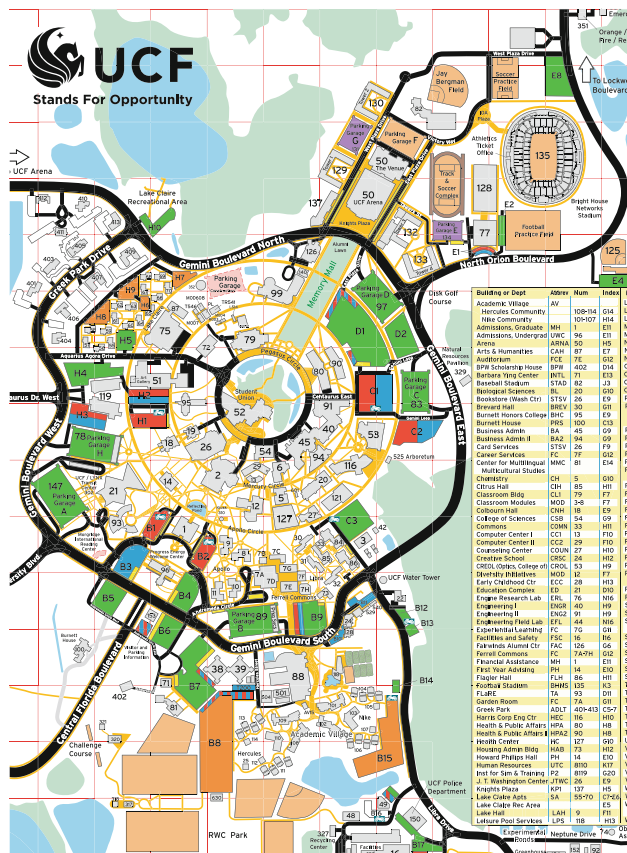


Fig. 1 The map used in the simulation. Gray spots are buildings; black lines show the campus roads; and yellow lines indicate the walkways. Parking lots are marked in green (student), blue (staff), and red (faculty) (color figure online)

5. Usage of administrative and other miscellaneous locations
6. Frequency of parking lot and shuttle stop usage.

For categories three through six, students were specifically queried about their visitation frequencies. For these questions, responses included one of the following: *never*, *rarely*, *once a month*, *several times in a month*, *once a week*, *several times in a week*, and *every day*.

In addition to the survey data, our agent-based simulation used publicly available statistics about UCF² and the main campus building map.³ A graph of the campus paths and roads was created from the main campus building map. The set of nodes in the graph is the union of the locations in the survey, plus the junctions between the streets and pathways. The edges of this graph represent the roads and walkways among the nodes. The weights of the edges show the distance between the connecting nodes. Each node and edge has a tag. This tag for the nodes indicates whether

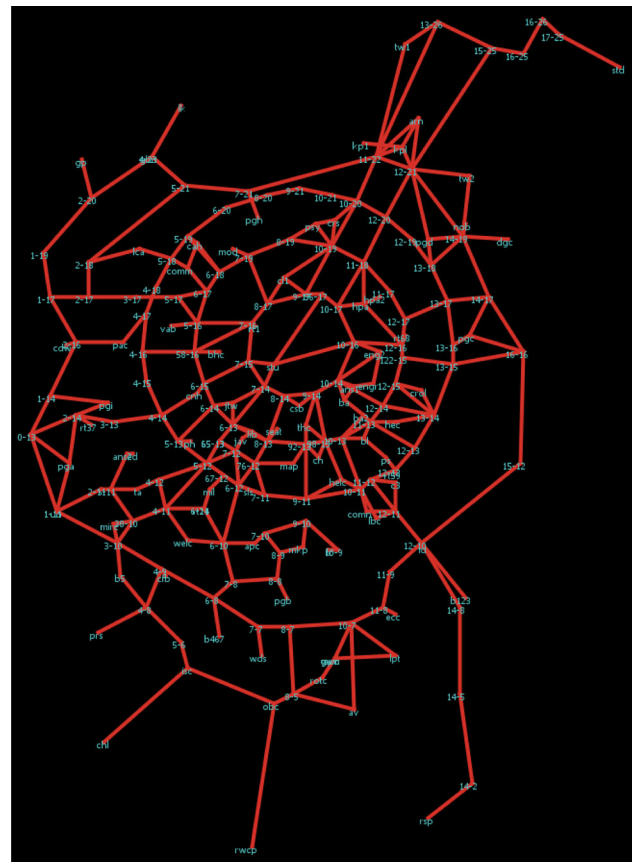


Fig. 2 The corresponding graph to the map in Fig. 1. The nodes represent different locations on campus, and the edges show the paths between them

they are a location of interest on the map or merely a junction. Figure 1 shows a snapshot of the map, and Fig. 2 shows the corresponding path planning graph.

4.2 Agent-based model

To perform transportation forecasting on the UCF campus, we created an agent-based model for simulating the common activities (transportation, dining, recreation, and building occupancy) performed by the 47,000 students on the main campus. This number refers to the total number of students on the UCF main campus. Each agent in the model represents an individual student and has a unique set of parameters that govern his/her activity profile. An agent's defining parameters are as follows: *entrance*, *dormitory*, *department*, *class building*, *arrive*, *depart*, *lunch*, *dinner*, *beverage*, *recreation* and *wellness*, *parking*, *shuttle*, and *miscellaneous*. The first four parameters designate the single (most common) value of the agents' entry point to the campus, housing situation, home department, and main class building. Note that, we did not explicitly represent the students' class schedules in the model. Even though this

² <http://www.iroffice.ucf.edu/character/current.html>.

³ <http://map.ucf.edu/printable/>.

would have improved the fidelity of the model, we felt that addition would not generalize well to other types of urban models. *Arrive* and *depart* are lists showing the times the agent enters the campus and leaves it. The remaining parameters are lists of locations for the agent's dining, recreation, and commuting. Additionally, each parameter that includes a location has another matching parameter that shows the time or frequency of visiting that location.

Rather than directly mapping the survey data to simulated entities that match the exact preferences of one of the survey respondents, we attempt to learn a general model of the population by fitting a statistical distribution to the answers of every question. For those questions that were related to the time of visiting a location (e.g., campus arrival and departure times), a Gaussian distribution was used to create a continuous distribution of arrival and departure times for the population of agents. For those questions where the respondents provided frequencies (e.g., how often campus dining locations were visited), we evaluated the performance of several discrete distributions and selected the Poisson distribution as offering the best fit for most of the questions.

After fitting the Poisson distribution on the qualitative data, a mapping function is used to work with the values obtained. This function maps the qualitative frequencies to exact dates and times. Each term, from *rarely* to *everyday*, is treated separately. For instance, the term *rarely* is mapped to a random day in a 60-day period.

4.3 Activity-oriented microsimulation

When the simulation commences, all the agents are initialized with parameters that remain constant over the lifetime of the agent and are used to create daily activity profiles. Our simulation is implemented in the Netlogo Wilensky (1999) environment.⁴

In this environment, time is discrete and simulated by ticks where a tick is one unit of time. In our model, one tick represents one hour of activity in the real world. When the model starts, each agent runs within a loop. The loop continues until the simulation is stopped. Figure 3 shows the runtime process by which an agent activity profile is generated. In this loop, whenever it is determined that the agent should be somewhere on campus, it goes to the enable (visible) state; otherwise, it goes to the disable (hidden) state.

Based on the agent's parameters, the activity profile generator determines what should an agent do and where should be at every time (tick). If sampling the agent's profile indicates that it should be on campus, then the

```

switch current-time-status:
  case entrance-time
    if live-off-campus then
      enter-campus //go to one of the entrances
      go-to-parking-or-shuttle-stop
    end if
  case on-campus-time
    if should-go-somewhere then
      go-to-destination
    else
      stay-at-department
    end if
  case return-time
    if live-off-campus then
      go-to-parking-or-shuttle-stop
      leave-campus //go to one of the entrances
    else
      go-to-dorm
    end if
  case not-on-campus
    disable

```

Fig. 3 Runtime generation of agent activity profiles

function compares the current time with the possible activity times produced by the mapping function that maps frequencies from the agent's distribution model to specific times and dates. If a match is found, then the agent opts to travel to that location. Otherwise, the agent remains at its department as its default place. On the other hand, if the profile generator determines that the agent should not be on campus, then the agent goes to (or remains in) the disabled state.

Various constraints are checked before an agent decides to go to a place. These constraints ensure the consistency of the whole model with the real-world facts. The main consistency checks are summarized below:

- *daily schedule*: Whenever an agent's model generates a date and time for visiting a location on campus, it checks the agent's arrival and departure times for that day. Campus activities that fall outside those boundaries are eliminated.
- *activity overlap*: Whenever the agent's model generates trips that overlap in time, requiring the agent to be in multiple places at once, one of the overlapping tasks is shifted to a later time.
- *campus constraints*: Known information about the operation hours of administrative offices, classroom buildings, and shuttle transportation is incorporated into the simulation. If the agent's model generates trips that violate the known operation hours, those trips are discarded.

A shortest-path graph algorithm is used to choose the path that an agent should traverse between its start and end positions. To speed-up the model, an all pairs shortest-path graph algorithm computes all of the shortest paths. A slightly modified version of Floyd–Warshall algorithm (Floyd 1962) was used for this purpose. All path planning

⁴ The complete code of this model can be accessed at this link: <http://code.google.com/p/ucf-abm/>.

Table 1 The parameter settings of ABM experiments

Parameter	Value
Agents	47,000
Days	100
Time range	07:00–24:00

occurs at initialization; candidate paths are stored in a look-up table to be accessed later. The time complexity of Floyd–Warshall algorithm is $\theta(n^3)$. The parameter values used for all of the experiments are listed in Table 1.

5 Method

To evaluate the performance of our hybrid model, we compared the performance of our model against several other ABM and MCMC variants. Figure 4 shows the relationship between the different methods in a schematic way.

5.1 MCMC

To benchmark the performance of our improved hybrid MCMC model (ABM-MCMC), we created a MCMC simulation with a standard proposal distribution (MCMC).

Our MCMC simulation uses the Metropolis-Hastings algorithm and is implemented with the `mhsample` function in the MATLAB Statistics toolbox.

Rather than creating one large monolithic simulation of the entire urban system to explore a variety of scenarios, here, MCMC is used to directly to forecast specific questions of interest, such as estimating the number of cars entering the parking lots at different times of a day. One can envision this as a two-dimensional diagram with the horizontal axis corresponding to the time of a day, and the vertical one showing the number of cars entering a specific parking lot. The survey data from the questions about the attendance pattern and frequency of parking lot usage are used to initialize the MCMC model. Observations for the Bayesian inference process are simply obtained based on the results of the survey data for a simulation period of 90 days. Imagine that based on the survey data, a student respondent enters the campus everyday before 9 am, leaves at 5 pm, and reports his general usage of parking lot A as being at a frequency of once a week. In this case, the expectation is that the student would have occupied Lot A twelve times ($90/7$) during the simulation period, so a corresponding number of samples tagged with the reported time range are produced and added to the input observation data.

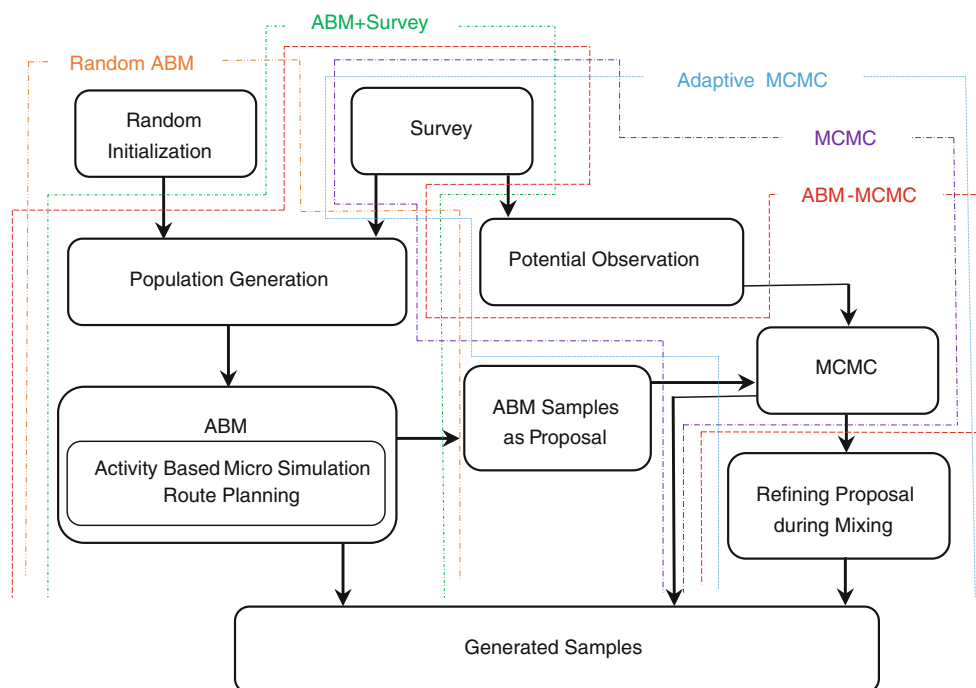


Fig. 4 A flowchart showing the relationship between the various modeling methods. In the proposed hybrid model, *ABM-MCMC*, the ABM is used to bootstrap the MCMC proposal distribution. In *ABM Random*, the agent-based model is initialized with a student population possessing randomly generated, but plausible schedules.

In *ABM+Survey*, the survey data are used to create the distributions for generating agent activity profiles. MCMC employs the MH algorithm with a standard proposal distribution, and in *Adaptive MCMC*, the proposal distribution is refined during the mixing process

Table 2 The parameter settings for MCMC

Parameter	α	Burn-in	Omission Rate	Chains
Value	[1 1 1]	$1e + 4$	2 of 3	2

Table 2 shows the parameter settings for MCMC used in the experiments. The burn-in value refers to the number of values that are discarded before the actual samples for the Markov chain are generated. In order to remove the correlation between the nearby samples, sometimes, the samples are not gathered sequentially. The number of samples that are thrown away determines the omission rate. Here, two out of three samples are omitted. The last parameter in the table shows the number of Markov chains that are created.

5.2 ABM-MCMC

Similar to the notation that is used in Axtell (2003), we can describe the relationship between the ABM and MCMC models mathematically. The state of each agent i in the agent-based model can be denoted by the vector $x_i \in X_i$ that can assume values in the range $X_i \subseteq \mathbb{R}^{n_i}$. The state space of the whole population can be designated as follows: $X = X_1 \times X_2 \times \dots \times X_n$, and the state of the model at time t as $x(t) \equiv (x_1(t), x_2(t), \dots, x_n(t)) \in X, x_i \in X_i$.

After convergence of the MCMC process, the following condition will hold: $x(t) \simeq x(t+1)$. Here, X can be designed to be same as the variable whose distribution we are seeking using MCMC. The intuition is that the world state, X , assumes the shape of target distribution. By designing an appropriate agent-based model, this variable will be close to the sought-after target distribution.

While we have enough samples from a variable x , it is easy to compute its probability distribution function (PDF). In this case, the samples drawn from the agent-based model are used to determine the PDF of the proposal distribution. In our experiments, we assign a probability value to each point x proportional to its number of occurrences in the population domain of the agent-based model: $q(\alpha) = P(x_i = \alpha)$.

In our proposed unification of ABM of MCMC, the input proposal distribution, $q(x)$, for MCMC is derived using the above assumptions. The samples that are produced by the ABM can be used to construct the proposal distribution in the MCMC. It is worth noting that there are other nonmathematical alternatives for combining the two methods. For instance, it would be straightforward to simply directly use MCMC as an embedded component to model regions of the simulation where the total occupancy is of more interest than the exact agent position.

In case of our case study, the final goal of the campus modeling problem was to reach to a model describing the

transformation patterns of students. Accordingly, the desired distribution should represent the time and location of students. This information was retrieved from the agent-based model by recording the x and y coordinates of agents at each hour (tick) for 90 days. A Dirichlet distribution is used as the unnormalized distribution, $\pi(x)$. The general PDF of the Dirichlet distribution can be expressed as follows:

$$f(x_1, \dots, x_{k-1}; \alpha_1, \dots, \alpha_k) = \frac{1}{B(\alpha)} \prod_{i=1}^{k-1} x_i^{\alpha_i-1}$$

Three variables, x , y , and *time*, form the three-dimensional support of the applied Dirichlet distribution used by our model. Hence, k in above formula is equal to three, and x_1 , x_2 , and x_3 correspond to x , y and *time*. The α values are simply assumed to be equal to one. The proposal probability of each vector, containing x , y and *time* values, is equal to the number of times the vector exists in the dataset divided by the total number of records, under the assumption that the agent-based model has produced evenly distributed samples from the population domain. The `MCmultinomdirichlet` function in R is used to implement the proposed method; this function generates a sample from the posterior distribution of a multinomial likelihood with a Dirichlet prior.

5.3 Adaptive MCMC

We benchmark our proposed hybrid method against a technique known as adaptive MCMC (De Freitas et al. 2001) in which the proposal distribution is updated at runtime based on the behavior of Markov chain. For this method, the Metropolis-Hastings algorithm from the MCMC toolbox for MATLAB (Laine 2013) was used. Our MCMC model assumes that the unnormalized distribution is of the form of a Poisson distribution, the same as our ABM model. For the proposal distribution, a Gaussian is used. The MCMC attempts to find the most likely value of the mean of the Poisson distribution (λ in $\frac{\lambda^x e^{-\lambda}}{x!}$).

6 Results

To evaluate the performance of the agent-based model under different initialization conditions, we examined the transportation forecasts produced by the simulation, both through visualization and by comparing the predictions against a dataset collected by the UCF Parking Services office. The occupancy percentage of UCF student parking garages (shown on the horizontal axis) predicted by every modeling method is compared. Figure 5a–c show the absolute differences between the forecasts for each

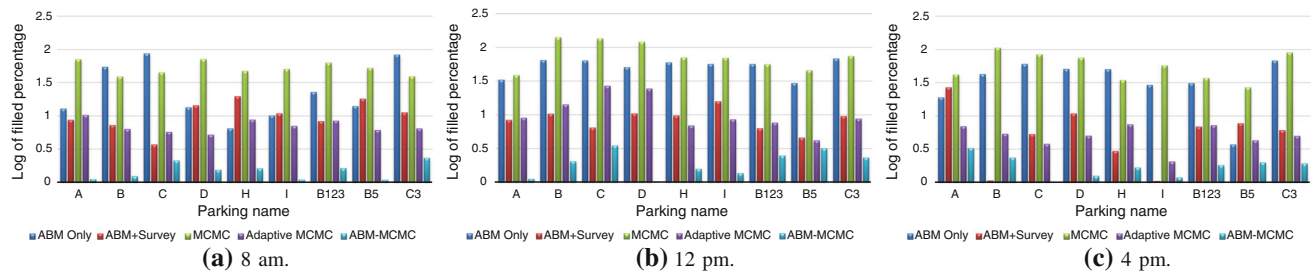


Fig. 5 The absolute difference (plotted on a log-scale) between the average occupancy percentage of the campus parking lots (shown on the horizontal axis) as predicted by different modeling methods compared to the UCF Parking Services data. Our proposed method

(ABM-MCMC), shown at the far right, yields consistently better estimates of parking lot utilization with a close to zero absolute difference

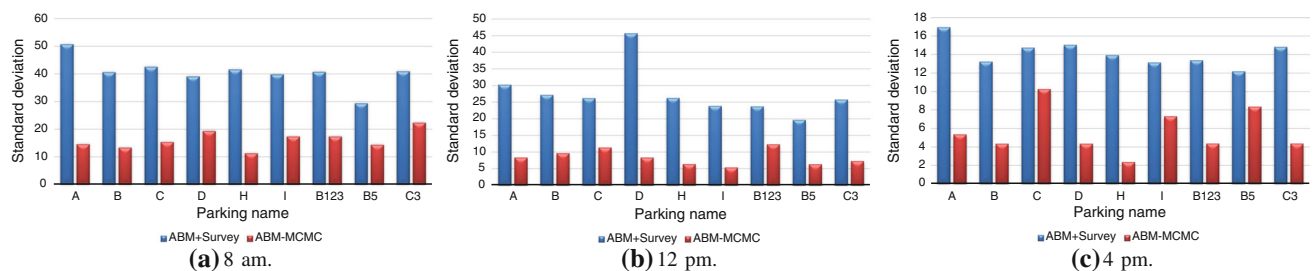


Fig. 6 The standard deviation of predicted values for parking usage, obtained by the agent-based model (ABM+Survey) and the proposed hybrid approach (ABM-MCMC) for 20 separate runs. The proposed method yields a reduction in prediction variance

modeling technique and the parking service data (closer to zero is better). Note that, our hybrid method (shown at the far right) consistently produces the best estimates, improving upon both its parent techniques. The stability of results obtained by different modeling methods is a concern; in many agent-based models, small changes in initial conditions can result in large changes in the final prediction. Figure 6 shows the standard deviations obtained from 20 runs of the ABM+Survey and the proposed ABM-MCMC methods. Note that, using the MCMC estimator reduces the variance of the raw ABM model, resulting in more consistent predictions.

The agent-based portion of our model can be used to create useful visualizations to provide intuitions about the students' transportation patterns. One of the common questions often asked by policy-makers is the density of humans at various locations. Figure 7 shows the probability of being in a location on campus for the students at large. In this figure, darker circles show more populated areas. In addition to the spots and buildings on campus, the traffic on the streets and walkways can be also predicted by our method. Some obvious facts that can be easily verified by a domain expert are also observed in this set of results. For instance, as on most university campuses, the student union is the most frequently visited place since it is the venue for most events and many dining locations. The wide drivable boulevard that surrounds the campus

dominates the road usage as it is the only way that can be used by cars and shuttles to reach most points on campus.

7 Conclusion

Hybrid models are a powerful strategy for reconciling the predictions of multiple models to present a unified picture to policy-makers, while retaining the diversity and flexibility of multiple approaches. This article introduces a new hybrid modeling method for combining agent-based models with MCMC. We demonstrate that the proposed method for initializing the MCMC proposal distribution with ABM data significantly reduces the prediction error over standard MCMC and also improves upon the ABM alone. We hypothesize that the combined ABM-MCMC finds a more general model of the posterior distribution than the ABM alone. Although agent-based models are often difficult to formally specify and reproduce exactly, the contribution of the ABM can be entirely quantified by the single proposal distribution, which makes it possible to reproduce the results without replicating the entire ABM. In a case study, we demonstrate that our method can be used to accurately model and forecasting transportation patterns in a large urban area. However, the value of the ABM-MCMC modeling approach lies not only its ability to more accurately predict variables such as parking lot occupancy, but

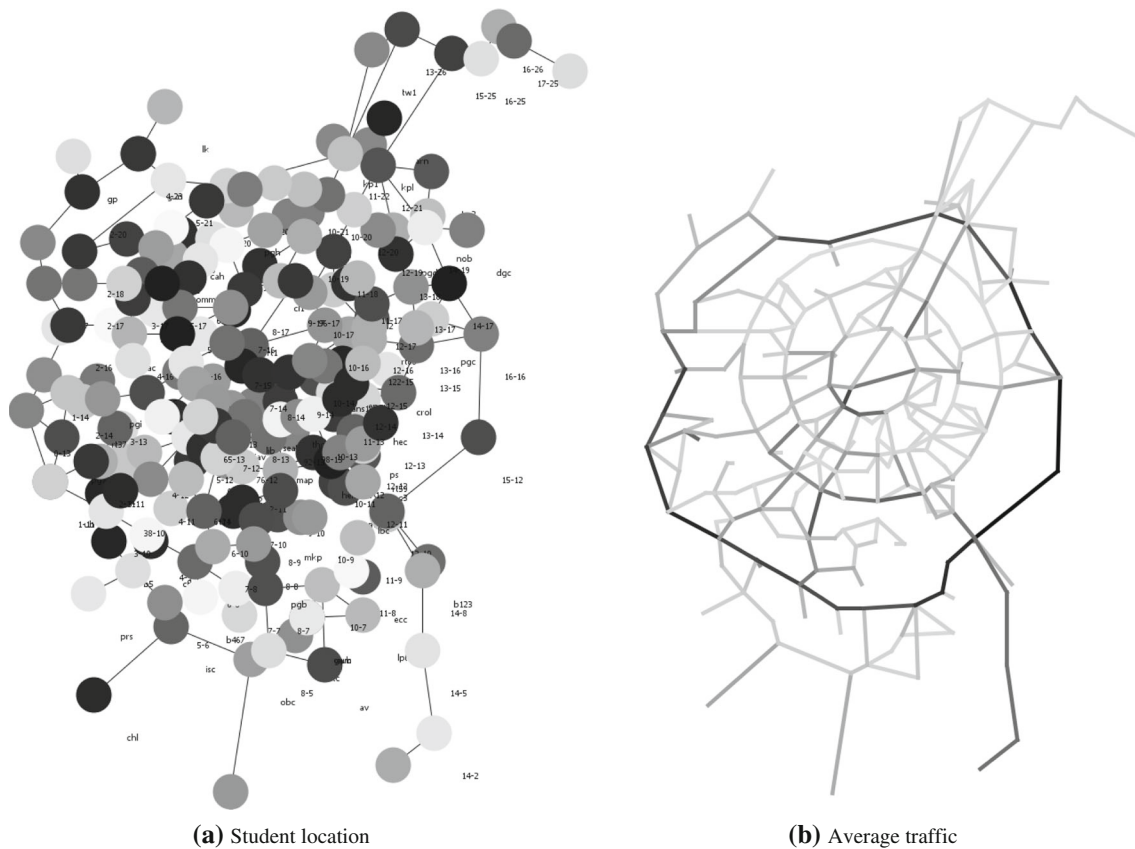


Fig. 7 Visualizations produced by our agent-based modeling system of the probability of a student being at a certain location (*left*) and the average traffic passing the streets and walkways on the campus (*right*)

also in the simplicity of the proposal distribution representation.

Transportation modeling and analysis can play a key role in facilitating effective urban planning, enabling public policy-makers to tackle difficult resource allocation decisions such as judging the relative merits of bicycle facility enhancements, pedestrian-oriented development, increases to public transit, and HOV lanes. University communities are constantly weighing the benefits of new construction: Is it better to add a parking lot or increase on-campus housing? Yet, no single model tells the whole story, and wise analysts guard themselves from avoidable failures by examining the forecasts of many models across a wide range of parameters. The human modeler always remains an important part of the equation, and hence, it is important to provide the user with flexible tools and easy-to-understand representations. Our hybrid approach provides a simple way to relate the predictions of multiple models through proposal distributions; we envision that these distributions can serve as a common exchange format between models. The proposal distributions produced by the ABM are more complicated than the ones commonly used to initialize MCMCs, but are still relatively simple to express. By directly visualizing the proposal distribution,

the modeler can gain an understanding of the effects of changing the parameters of the ABM.

One simple improvement that we are planning to make in the future is to add faculty/staff into our agent-based model; this was not a priority initially since previous work has shown that faculty/staff activity profiles have a much lower entropy and are inherently easier to predict than student profiles (Eagle and Pentland 2006). Supplementing the simulation with additional information about semester class scheduling is likely to yield the largest forecasting improvement at the cost of making the simulation less applicable to other urban modeling problems. A large amount of class attendance and scheduling information is collected by the university and could be added to the simulation without requiring additional survey efforts.

Acknowledgments This research was supported in part by National Science Foundation award IIS-0845159.

References

- Andrieu C, Moulines É (2006) On the ergodicity properties of some adaptive MCMC algorithms. *Ann Appl Probab* 16(3):1462–1505
- Andrieu C, De Freitas N, Doucet A, Jordan M (2003) An introduction to MCMC for machine learning. *Mach Learn* 50(1):5–43

- Axtell R (2003) Economics as distributed computation. In: Meeting the challenge of social problems via agent-based simulation. Springer, Japan, pp 3–23
- Balbi S, Giupponi C (2009) Reviewing agent-based modelling of socio-ecosystems: a methodology for the analysis of climate change adaptation and sustainability. University Ca'Foscari of Venice, Dept of Economics Research Paper Series (15_09), Working Paper No. 15/WP/2009
- Barreteau O, Sauquet E, Riaux J, Gailliard N, Barbier R (2012) Agent based simulation of drought management policy in practice. In: International Workshop on Agent-based Modeling for Policy Engineering (AMPLE 2012), European Conference on AI, pp 29–44
- Beheshti R, Sukthankar G (2012) Extracting agent-based models of human transportation patterns. In: Proceedings of the ASE/IEEE international conference on social informatics, Washington, pp 157–164
- Benenson I, Torrens P, Europe W, Portugali J (2004) Geosimulation: automata-based modeling of urban phenomena. *Environ Plan B Plan Design* 31(4):589–613
- Brown D, Riolo R, Robinson D, North M, Rand W (2005) Spatial process and data models: toward integration of agent-based models and GIS. *J Geogr Syst* 7(1):25–47
- Cauchemez S, Carrat F, Viboud C, Valleron AJ, Bolle PY (2004) A Bayesian MCMC approach to study transmission of influenza: application to household longitudinal data. *Stat Med* 23(22):3469–3487
- Chen B, Cheng HH (2010) A review of the applications of agent technology in traffic and transportation systems. *IEEE Trans Intell Trans Syst* 11(2):485–497
- Chen X (2003) Agent-based simulation of evacuation strategies under different road network structures. University Consortium of Geographic Information Science
- De Freitas N, Højen-Sørensen P, Jordan M, Russell S (2001) Variational MCMC. In: Proceedings of the conference on uncertainty in artificial intelligence, Morgan Kaufmann Publishers Inc., pp 120–127
- Dia H (2002) An agent-based approach to modelling driver route choice behaviour under the influence of real-time information. *Trans Res Part C Emerg Technol* 10(5-6):331–349
- Eagle N, Pentland A (2006) Reality mining: sensing complex social systems. *Pervasive Ubiquitous Comput* 10:255–368
- Eaton D, Murphy K (2007) Bayesian structure learning using dynamic programming and MCMC. In: Proceedings of the conference on uncertainty in artificial intelligence (UAI2007), pp 101–108
- Edwards P (1999) Global climate science, uncertainty, and politics: Data-laden models, model-filtered data. *Sci Cult* 8:437–472
- Floyd RW (1962) Algorithm 97: shortest path. *Commun ACM* 5(6):345
- Gailliard N, Olivier B, Audrey RF (2012) A conceptual model of participatory policy making in practice: water governance and boundary workers. In: International workshop on agent-based modeling for policy engineering (AMPLE 2012), European Conference on AI, pp 90–104
- Garlick M, Chli M (2009) The effect of social influence and curfews on civil violence. In: Proceedings of the international conference on autonomous agents and multiagent systems, pp 1335–1336
- Geman S, Geman D (1984) Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans Pattern Anal Mach Intell* (6):721–741
- Gerst M, Wang P, Roventini A, Fagiolo G, Dosi G, Howarth R, Borsuk M (2012) Agent-based modeling of climate policy: An introduction to the ENGAGE multi-level model framework. *Environ Modell Softw*
- Gilks W, Richardson S, Spiegelhalter D (1995) Markov Chain Monte Carlo in practice: interdisciplinary statistics, vol 2. Chapman & Hall/CRC, Boca Raton
- Gimblett H (2002) Integrating geographic information systems and agent-based modeling techniques for simulating social and ecological processes. Oxford University Press, Oxford
- Hailegiorgis AB, Kennedy WG, Roleau M, Bassett J, Coletti M, Balan G, Gulden T (2010) An agent based model of climate change and conflict among pastoralists in east Africa. In: Proceedings of the international congress on environmental modelling and software
- Hinkelmann F, Murrugarra D, Jarrah AS, Laubenbacher R (2011) A mathematical framework for agent based models of complex biological networks. *Bull Math Biol* 73(7):1583–1602
- Ilachinski A (2012) Modelling insurgent and terrorist networks as self-organised complex adaptive systems. *Int J Parallel Emergent Distrib Syst* 27(1):45–77
- Jin X, Jie L (2012) A study of multi-agent based models for urban intelligent transport systems. *Int J Adv Comput Technol* 4(6):126–134
- Jin X, White R (2012) An agent-based model of the influence of neighbourhood design on daily trip patterns. *Comput Environ Urban Syst* 36(5):398–411
- Jordan R, Birkin M, Evans A (2012) Agent-based modelling of residential mobility, housing choice and regeneration. In: Agent-based models of geographical systems, Springer, Netherlands, pp 511–524
- Klügl F, Bazzan AL (2012) Agent-based modeling and simulation. *AI Magaz* 33(3):29
- Kohler TA, Bocinsky RK, Cockburn D, Crabtree SA, Varien MD, Kolm KE, Smith S, Ortman SG, Kobti Z (2012) Modelling prehispanic Pueblo societies in their ecosystems. *Ecol Modell* 241:30–41
- Laine M (2013) MCMC toolbox for Matlab, Finnish meteorological institute. Retrieved from: <http://helios.fmi.fi/lainema/mcmc/>
- Liu R, Tao J, Shi N, He X (2011) Bayesian analysis of the patterns of biological susceptibility via reversible jump MCMC sampling. *Comput Stat Data Anal* 55(3):1498–1508
- Liu Y, Wang Q, Liu J, Wark T (2012) MCMC-based indoor localization with a smart phone and sparse WiFi access points. In: IEEE international conference on pervasive computing and communications workshops, pp 247–252
- López-Paredes A, Saurí D, Galán JM (2005) Urban water management with artificial societies of agents: The FIRMABAR simulator. *Simulation* 81(3):189–199
- Maghami M, Sukthankar G (2012) Identifying influential agents for advertising in multi-agent markets. In: Proceedings of international conference on autonomous agents and multi-agent systems, Valencia, Spain, pp 687–694
- Mengersen K, Tweedie R (1996) Rates of convergence of the Hastings and Metropolis algorithms. *Ann Stat* 24(1):101–121
- Metropolis N, Rosenbluth A, Rosenbluth M, Teller A, Teller E (1953) Equation of state calculations by fast computing machines. *J Chem Phys* 21:1087–1093
- Niazi MA, Hussain A, Kolberg M (2009) Verification & validation of agent based simulations using the VOMAS (Virtual Overlay Multi-agent System) approach. In: MAS&S at Multi-Agent Logics, Languages, and Organisations Federated Workshops (MALLOW)
- Oakes J (2008) Invited commentary: rescuing Robinson Crusoe. *Am J Epidemiol* 8(1):9–12
- Pan X, Han CS, Dauber K, Law KH (2007) A multi-agent based framework for the simulation of human and social behaviors during emergency evacuations. *AI Soc* 22(2):113–132
- Press W, Teukolsky S, Vetterling W, Flannery B (2007) Numerical recipes 3rd Edition: the art of scientific computing. Cambridge University Press, Cambridge
- Streit RE, Borenstein D (2009) An agent-based simulation model for analyzing the governance of the Brazilian financial system. *Expert Syst Appl* 36(9):11,489–11,501

- Verella J, Wardak A (2008) Modeling public opinion and voting as a complex system with agent-based simulations. In: IEEE Syst Inform Eng Design Symp, pp 261–266
- Ward MD, Gleditsch KS (2002) Location, location, location: An MCMC approach to modeling the spatial context of war and peace. *Polit Anal* 10(3):244–260
- Wilensky U (1999) NetLogo. Evanston, IL: Center for connected learning and computer-based modeling, Northwestern University. Retrieved from: <http://ccl.northwestern.edu/netlogo/>
- Wilensky U, Rand W (2007) Making models match: replicating an agent-based model. *J Artif Soc Soc Simulation* 10(4):2