Brief introductory guide to agent-based modeling and an illustration from urban health research

Breve guia introdutório de modelagem baseada em agentes e uma ilustração a partir da pesquisa em saúde urbana

Breve guía introductoria de modelación basada en agentes y una ilustración a partir de la investigación en salud urbana

Amy H. Auchincloss [1]
Leandro Martin Totaro Garcia [2]

[1] School of Public Health, Drexel University, Philadelphia, U.S.A.
[2] Faculdade de Saúde Pública, Universidade de São Paulo, São Paulo, Brasil.

Correspondence
A. H. Auchincloss
School of Public Health, Drexel University.
3215 Market Street, Philadelphia/PA – 19104, U.S.A.
aha27@drexel.edu

Abstract

There is growing interest among urban health researchers in addressing complex problems using conceptual and computation models from the field of complex systems. Agent-based modeling (ABM) is one computational modeling tool that has received a lot of interest. However, many researchers remain unfamiliar with developing and carrying out an ABM, hindering the understanding and application of it. This paper first presents a brief introductory guide to carrying out a simple agent-based model. Then, the method is illustrated by discussing a previously developed agent-based model, which explored inequalities in diet in the context of urban residential segregation.

Computer Simulation; Epidemiologic Methods; Systems Theory; Urban Health

Resumo

Há um interesse crescente entre os pesquisadores de saúde urbana em trabalhar com problemas complexos utilizando modelos conceituais e computacionais do campo de sistemas complexos. A modelagem baseada em agentes (MBA) é uma ferramenta computacional de modelagem que tem recebido crescente interesse. No entanto, vários pesquisadores ainda não se sentem familiarizados com o desenvolvimento e a execução de uma MBA, dificultando a sua aplicação e compreensão. Este artigo primeiramente apresenta um breve guia introdutório para executar um simples modelo baseado em agentes. Em seguida, o método é ilustrado discutindo um modelo baseado em agente previamente desenvolvido, que explora as desigualdades na dieta no contexto da segregação residencial urbana.

Simulação por Computador; Métodos Epidemiológicos; Teoria de Sistemas; Saúde Urbana

Introduction

Among urban health researchers, there is growing interest in conceptualizing complex problems using a system framework [1] and in using systems modeling tools to explore how components of a complex problem interact, are sustained or changed, and ultimately identify areas for intervention [2,3]. In particular, system simulation approaches are useful tools for understanding processes and structures involved in complex problems, identifying high-leverage points in the system and evaluating hypothetical interventions [1] – an exercise that would be impossible to do by collecting and analyzing real-world data.

One tool that has been increasingly used to examine urban health issues is agent-based modeling (ABM) [4,5]. Agents are given traits and initial behavior rules that organize their actions and interactions. Stochasticity can be included in the assignment of agent characteristics and in determining which agents interact and how agents obtain information and make decisions. The model is run over time and repeated numerous times, to obtain a distribution of possible outcomes for the specified system. The microentities, referred to as "agents", are anything that alters its behavior in response to input from other agents and the environment [6].

ABM is able to accommodate high heterogeneity in agent characteristics and interactions between agents and environments, as well as features like dynamics, feedbacks and adaptation, which are impossible to represent in traditional statistical models [7,8]. Agents can be defined at multiple levels, including individuals or group of individuals (e.g., families, institutions, policy-making bodies etc.). Research questions that require significant heterogeneity within and between agents and diverse spatial and relational elements are well-suited to ABM [9]. In urban health research, simulations can be used to explore dynamic scenarios involving diverse entities and settings such as the built and social environment, city agencies, legislative bodies, health services, individual residents and families. Some agent-based models include detailed data and strive for high realism [4] while others are abstract [5,10].

Despite the ABM suitability to research complex problems in urban health, it is a new tool to many researchers. One important barrier to foster ABM adoption among researchers is their unfamiliarity with steps needed to carry out the modeling. Therefore, the purpose of this paper is to provide a very brief introductory guide to carrying out a simple agent-based model. We then use a previously constructed model [11] to illustrate the steps one can take when building a simple model. This is only a brief guide; before starting a computational model, it is recommended that readers refer to comprehensive guides [9,12,13,14].

Modeling guide

Conceptual model

As in all research endeavors, first the investigator must define the question(s) of interest. To outline the question(s), researchers rely on mental models encompassing components and mechanisms relevant to the topic of interest. The problem is that these models usually remain implicit, along with their assumptions, internal consistency and logical consequences [15]. Therefore, to define the research question(s), the initial step is to construct an explicit conceptual model. At the first stage, this should be a broad conceptual model characterizing the general problem and some specific features related to it. Then, one can identify where the significant gaps in knowledge are, and a relatively simple aspect of the problem to explore in depth. In the second stage, the researcher articulates a more specific, narrower conceptual model around this relatively simple aspect. At this stage, one works to identify the key elements that may be most important to the question(s) and think about dynamic processes and feedbacks that may play an important role.

There are three important points to highlight about constructing conceptual models. First, conceptual models can be based on either theory or empirical data, or both. The researcher may construct a new theoretical model, or explore and extend someone else's model. Second, the conceptual model is a prerequisite for computational models, but by itself has high value and can be a product for those unwilling to undertake computational modeling. Third, a common mistake among researchers starting in ABM is to try to write a computational model that addresses many elements within the broad conceptual model identified before. It is important to keep in mind that all models are analogies of real systems, and so they will fail to represent reality [16]. Good models balance simplicity and adequate representation, incorporating enough key elements and processes and ignoring those that are not directly relevant.

After defining the specific research question(s), we need to choose the most suitable tool to carry out the work. Not all questions posed within a system framework need to be answered using a systems science tool; they may

be better answered with statistical methods or qualitative approaches. Moreover, ABM is not the only tool for modeling dynamic, complex systems. Other systems science tools, such as systems dynamics, may be preferable and more appropriate [9,17].

Computational model

• Model objective, plan for experiments, and outcome assessment

Modeling is an iterative process of using a conceptual model to plan and execute the computational model, and then potentially rethinking the conceptual model. The iterative process of modeling is often where the most valuable insight occurs, rather than in the "final results". Even though modeling is an iterative process, investigators still need to begin with a clear model objective. From there, investigators must plan the simulation study, including a preliminary plan for setting up and testing experimental conditions, and how the outcome will be assessed. This is also the time to plan what types of entities will fill the system and the temporal and spatial extent of the model.

• Agents and characteristics

Agents and their characteristics should be specific to the needs of the research question(s). Select few agents and the minimum characteristics required to address the question(s). Agents do not need to appear to be "real". For example, agents representing humans do not require specification of age, sex, race etc. unless those characteristics are involved in processes or decisions that will be modeled. In a simulation framework, there is a limitless range of options, thus researchers need to curb the enthusiasm for modeling numerous types of actors and characteristics. Adding a lot of detail does not necessarily result in better insight and can make very difficult to execute, test, and interpret the model.

• The world

The simulated "world" does not need to represent the real world; instead, it must represent the simulation space that is most appropriate to the specific question(s) being asked. If mapping to a local "real" geography is important, most programming environments allow users to import Geographic Information System (GIS) layers as inputs to replicate an actual urban space or configure a generic abstract space.

• Defining agent objectives

The researcher must define the main objectives of the agents, thinking through the processes that are essential to answering the research question(s) and choose to ignore the rest.

• Defining agent behavior rules – utility functions

Agents may be required to take action and/or make decisions in response to single stimuli or weigh multiple criteria. Utility functions are aides for decision making when factoring in multiple criteria and allowing each agent to rank options and make a choice. Theory and empirical research can be incorporated in decision-making rules, drawing in particular from the fields of economics, cognitive science, neuroscience and computation science [18,19]. Typically, there is a gap in the data/theory that inform decision making in the specific contexts we want to model. Thus, researchers may not have strong conceptual justification for a particular utility function and instead choose one that has been widely used and that provides reasonable results. Deciding on the specification of the utility function can be difficult and ultimately one will need to test sensitivity to the functional forms and inputs.

• Defining agent behavior rules – randomness

In ABM, randomness can be included in the construction of each dimension. Researchers usually add randomness to the utility function itself, in order to represent uncertainty they have about a particular equation and the parameters within it, as well as to represent bounded rationality [18]. Bounded rationality refers to the fact that decision making is not a perfectly rationale procedure. Decisions are made with incomplete information or even for reasons unknown to the actor making the decision.

• Setting schedules

The model is run over time steps. Model activity can be mapped to a timeframe in the real world; however, real timeframes are not required and may make little sense in an abstract model. For some models, there may be activities that occur in fixed time intervals or triggered when particular situations transpire.

• **Dynamics and feedbacks**

A key advantage of complex systems simulations – including ABM – is the ability to incorporate dynamics and feedbacks within the model, which may be important to the process being studied. Researchers should be deliberate about incorporating dynamics and feedbacks. Dynamics allow changes over time to agent characteristics or decision rules, in ways that could affect the process under study. For example, some questions involve lifecycle processes, where deaths and births are important to include in the model for equilibrium or to explore how much information, traits, and risks are passed from one generation to another. Feedbacks can be represented as responses to structural features (the structure of the world/environment, which could be exogenously imposed) and/or behavioral conditions (how behaviors are altered by other behaviors, often an endogenous process) [20]. Feedbacks are typically most interesting when represented in both structural and behavioral processes, as they can generate changes/new behaviors at both the agent level and the system at large [12,20]. Implementing many dynamics and feedbacks into the early model stage will make it impossible to interpret and verify the system [9], so it is recommended to start small and expand.

• **Results – stochasticity**

In ABM, stochasticity is part of many steps (initialization, behavior rules etc.), thus, it is important to run the model multiple times to obtain the distribution of outcomes and then summarize results across multiple runs. Note that, in ABM, this is not assessing how well the model fits the intended system or observed world; it is only assessing the impact of stochasticity embedded in the model [14]. Tools can be used to determine the number of runs necessary to generate a representative result (for example, the Simulation Parameter Analysis R Toolkit Application package developed for R [21]). Absent tools, it is reasonable to try 10 or 30 runs on a particular scenario and evaluate the magnitude of the uncertainty across runs.

• **Results – displays and interpretation**

ABM outputs are different from those generated from statistical analysis. The main outputs are the evolution of the system and its components (process outputs) and a summary of the "final" state (summary outputs). Process outputs are displayed in graphics or tables representing the system's variables at each time step (or lightly summarized over multiple time steps), as well as visual representations of the system in action. Process data are especially useful for exploring and interpreting the system's behavior, structure and emergence. Because any one-time step is representative of the system, in order to obtain a summary of the "final" state of the model, researchers can summarize the data of a representative/relevant interval at the end of the model run (e.g., averaging the outcome for the final 20% of the run). Due to the uncertainty of data inputs and modeling process, agent-based models are not prediction models and outputs should not be interpreted as precise estimates. Interpret results qualitatively rather than quantitatively. Look for strong patterns and large differences between summary outcomes across experiments; small differences are usually not worth noting.

• **Verification, calibration, external validation**

During the modeling process, some procedures must be done to achieve the most useful and reliable model possible. Given that ABM can reveal counterintuitive processes, evaluating and testing models can be difficult. Unexpected results that appear interesting may be due to errors in computer programming or high dependence on initial choices or small variations in stochastic processes involved with strong positive or negative feedbacks. For this reason, researchers need to work to internally validate (verification and calibration) and externally validate the model:

(a) Verification. Verification is the process of checking that the computer code correctly implements the model formulation, i.e., if it does what it was planned to do [14,22]. There are diverse strategies to verify from one or few lines of code to the whole program and it is recommended to use them continuously during coding, making easier finding and fixing mistakes. Many of the processes are standard practice for quality control when writing computer code and some are specific to ABM [12,22].

(b) Calibration. Calibration is the process of tuning model parameters to align with basic patterns observed in the real system being modeled [14,23]. Calibration can aim for a qualitative match or a close, quantitative match. Qualitative matches align the parameters with literature on the topic. This method is typically used when no calibration data exist or the model is abstract. Close match calibration is often chosen when particular parameters are very important and strongly affect the model results, the parameters are thought to have reasonably independent effects on the model, and good alignment data ex-

ist [14]. In this case, the researcher needs to identify relevant empirical data, define a plausible range of parameter values and set criteria for evaluating how good the match is.

(c) External validation. A simulation model is only an approximation of the target system, thus the work of external validation builds a case for the model's truthfulness and usefulness under certain conditions [16,24]. The external validation step can include evaluating the validity of the theories being used and how well the model incorporates them [24,25], examining the appropriateness of underlying assumptions and the suitability of the model to the purpose. Model outputs can be compared to either empirical data (coming from sources other than those used during calibration) or aligning the model with patterns observed in the real world and checking if the model captures the most important systems features [14,26].

- **Programming environment**

ABM can be done in any language, but object-oriented programming languages are preferred (Python, C++, Java etc.). Commonly used interfaces/libraries are RePast (http://repast.source forge.net), NetLogo (http://ccl.northwestern.edu/netlogo), and AnyLogic (http://www.any logic.com). A detailed and commented list of programming environments is presented by Kravari & Bassiliades [27].

- **Protocols for designing, executing, and communicating the model**

Some protocols and standards were developed in order to increase transparency of ABM, reduce criticism that models are irreproducible, and provide a language that the scientific community can use to evaluate model validity. The most frequently used protocol is the ODD [28,29] (and ODD+D [30]), which includes elements to make explicit the Overview, Design concepts and Details of the model. The example below does not explicitly follow this protocol, but we included a number of its components.

Illustration of an agent-based model from urban health research

Conceptual model

An income differential in diet quality has been observed in numerous studies illustrating that lower income is generally associated with worse dietary profiles [31,32]. Thus, diet quality has been identified as a key factor in socio-economic in-

equalities in obesity and diet-related illnesses. There are three prevalent theories of dietary inequality and the variety of explanations highlights that inequalities likely occur within a complex system of interrelated processes that are not well understood:

(a) Spatial inequality and access resulting from residential segregation by income and race/ethnicity. Within many urban areas in the U.S., minority and low-income neighborhoods have significantly fewer venues for purchasing healthy foods as compared with high-income neighborhoods [33,34].

(b) Individual or group preferences that are patterned by income. High-income households prefer healthy foods so choose to live in areas with healthy food stores, while low-income households prefer unhealthy foods and choose to live in areas without them [35].

(c) Monetary constraints. Healthier fresh fruits and vegetables cost more than packaged foods. Low-income households do not have the means to purchase healthier foods whereas higher income households do [36].

Model objective and plan for experiments

We used ABM to explore the role that urban segregation can play in shaping dietary behaviors and to suggest policy levers that may be used to counter its effects. The model allowed us to focus on how location and household incomes and preferences interact over time to influence store availability and supply of healthy foods and hence have the capacity to affect income differences in healthy eating. See Figure 1 for a conceptual sketch of core features included in the

Figure 1

Conceptual sketch of core features to include in the model. The sketch illustrates structural feedback between households and stores and behavioral feedback within households and stores (arrows around them). Households choice depends on their income, proximity to stores and food preferences. On the other hand, stores influence households' diet habit.
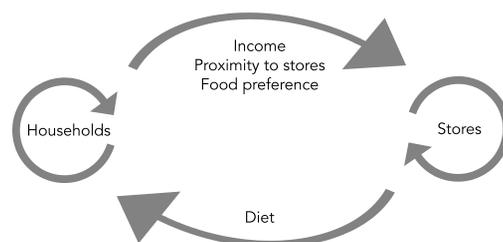
model. We imposed several extreme scenarios for economic residential segregation and spatial clustering of healthy food stores (for details, see Table 1 at Auchincloss et al. [11]). Then, we identified which particular scenario showed income differentials in diet that have been observed in previous empirical studies in the U.S., where higher incomes generally have better diet than low incomes [31,32]. Then we used the selected scenario to run experiments that explored whether pricing and preference factors were capable of reducing income differentials in diet generated by segregation.

Agents and characteristics

Only two types of agents were included: households and food stores [11].
(a) Household agents were differentiated by where they live, income and food preferences.

Table 1

Functions for weighting and scoring the inputs for utility.

Dimension	Justification and description for scoring	Weight *	High-income	Low-income
Price **	Price score. In our model, both low and high-income households preferred a cheap store – thus all households scored a cheap store as 1. To account for the fact that rich households were comparatively insensitive to price, rich households in our model scored an expensive store as 0.8 (*i.e.*, close to 1), while poor households scored it 0.1 (*i.e.*, close to 0)	0.2		
	Prices at the food store are cheap		1.0	1.0
	Price at the food store are expensive		0.8	0.1
Distance **	Both low- and high-income households preferred a cheap store, thus, all households scored a cheap store as 1. Both high- and low-income households equally preferred a closer store but, because travel was more of an obstacle for low-income households, poor households gave a distant store a lower score than a high-income household. The distance score was 1 – (number of grid cells between residence and store) / 30 for low-income households, and 1 – (number of grid cells between residence and store) / 130 for rich households	0.5		
	Distance between household and store = 0		1.0	1.0
	Distance between household and store > 0		$1 - d / 130$	$1 - d / 30$
Habit	Households remembered the store(s) they visited during the past five time steps. They gave a score of 1 (highest) to the store they last shopped at and decreased the score by 1/5 for each model step since they last visited that store	0.1		
	Same store that household shopped at on previous time step		1.0	1.0
	Different store (steps since = number of time steps since the household last shopped at store)		$1 - steps / 5$	$1 - since / 5$
Food preference	Preference scores in the utility just reflected the household's preference: households scored a healthy store with their own household preference; an unhealthy store got a score of 1 – household preference. Household food preference was a continuous attribute 0 to (preference for unhealthy food = 0, preference for healthy food = 1). Preference was either randomly assigned or assigned by household income in our "preference experiments" (unhealthy preference was assigned the lower end of the preference range [0.0-0.6] and healthy preference the upper end of the range [0.4-1.0] with some overlap between possible preferences)	0.2		
	Store sells healthy food		Preference	Preference
	Store sells unhealthy food		1 - Preference	1 - Preference

* Weighting parameters were used in the main results reported in the Auchincloss et al. [11]. Sensitivity was tested to alternative weights;
** Scores for distance and price were iteratively selected using the "random scenario" and assuming that high-income households should spend more on food and travel at least as far or farther than low-income households.

These characteristics were most relevant to food shopping behaviors we wanted to explore. Income was a proxy for other elements of socioeconomic status and it was an important trait in this model due to our interest in economic segregation. We randomly classified households into either low or high income, with 50% of households assigned to the low-income category. We ignored the middle-income category in order to keep the model simple and improve interpretation. Food preferences was a proxy for a range of personal attitudes and psychological factors – and to some extent cultural contexts – that could influence decisions around diet.
(b) Stores were assigned a location, a type of food (unhealthy or healthy; at initialization 50% of stores sell healthy foods) and average price for food (either inexpensive or expensive; 50% of stores sell inexpensive foods).

The world

Our question was abstract and not grounded in a specific city, thus the world did not require GIS layers or data that grounded it to a particular context. However, we needed to measure distance/proximity between agent locations and allow for clustering, thus, our model required a world with a measureable grid space. We chose a small grid space (50x50 grid) and each cell in the grid contained one household, thus 2,500 households in the world. At baseline, stores filled 2% of the grid cells, thus 50 stores (each store shared its cell with a household). In our model, stores made decisions after counting the number of customers. Thus, we needed a sufficient number of households to generate customers shopping at stores and the size of the world needed to be large enough to not skew results due to small samples/distances. We specified the space as toroidal, meaning that the world is a continuous space projection, so that boundaries would not present problems when agents calculated distances between themselves and the stores [37].

Agent behavior objectives

Households' objective is to select a store and shop for food, measured by which food store is selected. Stores' objective is to attract customers, measured by number of customers per period who selected the store.

Household behavior

At each time step, each household selected a store to shop. A time step was roughly conceived to represent about every 2-3 days as that frequency corresponded to food shopping frequency in empiric studies [38]. However, the duration of the model did not literally translate to human months or years. In our model, the frequency of shopping did not change across households and time, because that was not central to our research question.

Household utility score

We needed households to choose which store to shop at by ranking the stores on dimensions via a utility function, described in the Equation 1. The dimensions selected for this model are not universal; rather, they were selected due to their relevance for the question we posed. The four dimensions for ranking stores were price of food at the store, distance to the store, the stores that household shopped at previously (household's habitual shopping behavior) and the household's preference for healthy foods. Justification for each of these dimensions is included in the supplementary data for the original paper [11]. We selected a utility function that was able to balance each dimension, such that a low score in one dimension would not affect scores in other dimensions. We ended up using an additive form of the Cobb-Douglas function [39] that utilized both scores and weights. We did not use the multiplicative form because a low score in one dimension would make it difficult for a household to choose that store, even if the other scores were very high. We normalized the weights so they add to 1.0; thus, they have meaning only relative to each other. In Equation 1, i is the household, k is the dimension and ε is random noise (random variable, $\mu = 0$, $\sigma = 0.05$) to represent bounded rationality [18]:

$$Utility_i = \left(\sum_{k=1}^{4} score_{ki} \text{ x } weight_k \right) + \varepsilon_i \quad \text{[Equation 1]}$$

In our model, scores for price and distance were allowed to vary by household income, because we wanted to match existing evidence that high-income households pay more for food and travel farther than low-income households. Each score was on a scale from 0 to 1, where 1 was the most preferred score. Table 1 shows details on the scoring and brief justification for the choices we made. Weights were constant parameters for all households and did not vary by household attributes. The values for weights were determined through iteratively testing and changing model rules to adhere to calibration criteria: high-income households should spend more on food [40] and travel at least as far as low-income households [41]. Sensitivity to alternate

weighting and scoring for the utility function and size and household/store density of the grid were examined (see *Verification, Calibration, External Validation*).

Store behavior – food store sub-model

Stores were able to change the type of food they sold, but store prices remained fixed throughout the experiment. We devised a simple way to proxy dynamic processes in store behaviors in order to test the effect of stronger feedbacks between households and stores and to allow the household choice set to be slightly more dynamic. This "move-out/move-in" sub-model allowed low-performing stores to close. In locations without a store for a certain period (180 time steps), a new store could move into the old store's location, either selling the same food type as the old one or changing food type. We preferred a this simple "move-out/move-in" sub-model for the following reasons: our model was not focused on store location decision making, we knew that we had imperfect information for modeling this process, and it would take a lot of effort to construct a retail site selection sub-model.

Results – outcome measure and display

Main results are not reported here but the reader can find them in the original paper [11]. Figure 2 is an example of the display from one simulation. The primary outcome measure was the income differential in diet (diet of high-income households minus diet of low-income households). Absolute diet values for high- and low-income households were secondary outcomes. A simplifying assumption was used to derive each household's diet: if the household shopped at a healthy food store, they ate healthier food and had a better diet. Diet was summarized as the average proportion of times the household shopped at a healthy food store (i.e., diet of 0.5 meant they shopped at healthy food stores half of the time, diet values close to 0 meant they infrequently shopped at healthy food stores). Figure 3 is an example of how results can be summarized. Because uncertainty and randomness was built into agent initialization (e.g., agent location and attribute assignment) as well as store behaviors and households' selection of which store to go to, each experiment was run 60 times. From this, we obtained the distribution of outcomes and then summarized as the median and the 5th to 95th simulation percentile. Experimental results were summarized by averaging diet for the final 20% of the run of the model.

Verification, calibration, external validation

The model was simple and very abstract, not intended to have high external validity or be highly realistic or quantitatively calibrated to data. As a tool for explaining observable phenomena and stimulating questions, this model had reasonable face validity. The calibration stage used observational studies and survey data from government and industry sources to guide agent decision-making rules for generating plausible behaviors. Agent behaviors were tested against available data to reflect intuitive and known behaviors, such as high-income households spending more on food [40] and traveling at least as far or farther than low-income households [41]. Verification and calibration included testing sensitivity to alternate weighting and scoring for the utility function and size and household/store density of the grid. Figure 3 shows sensitivity summaries from the store behavior sub-model. The plot shows sensitivity to various assumptions in the store sub-model ("move-out/move-in" and changes in the type of foods are sold at the store; scenario #4 was used for the base scenario reported in the Auchincloss et al. original manuscript [11]).

Programming environment

The model was developed using an old Java version of Repast (version 3.30, http://repast.sourceforge.net). Additional libraries and code were from the Center for the Study of Complex Systems at the University of Michigan (http://www.cscs.umich.edu) and were written in Java using a Windows/Eclipse environment (Figure 4).

Important limitations – low dynamics and feedbacks

We briefly discuss here three limitations of the model. First, we did not envision or execute a full representation of the processes that result in income differentials in diet. However, our stylization of the world is not itself a shortcoming and roughly matched our objective. Second, we did not "generate" changes in the built environment. Rather, we tested how the contexts shape behaviors by exogenously imposing various segregation contexts. This strategy was not elegant, but is not a significant shortcoming. There were two reasons we did not "generate" the spatial sorting and neighborhood segregation. First, the point of the model was to see what happens in segregated contexts, not to generate the context. Second, generating segregation is an ongoing investigation topic that a number of researchers have taken on and requires a lot of

Figure 2

This four-panel figure illustrates the types of displays one can get from one run of a model. The scenario shown is where poor households were segregated from wealthy households and poor households were near stores with healthier foods. Panel 2a is a snapshot of the grid (world) were agents interact. Households are squares, stores are diamonds. Colors and shading map to select agent characteristics. Panels 2b,is an example of output that can assist with verification and validadtion: average number of customers at expensive stores (top line is total customers, middle line is high income, bottom line is low income). Panel 2c is a secondary outcome: healthier and unhealthy stores (top line is total stores, middle line is unhealthy, bottom line is healthier). Panel 2d is the main outcome: average proportion of times the household shopped at a helathier food store (on left side of plot top line is low income, bottom line is high income)[11].

2a)
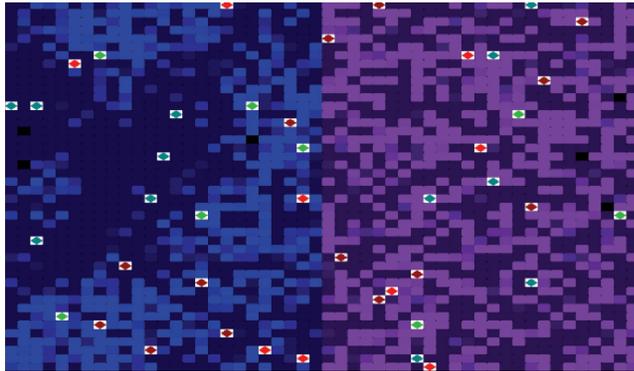
2b)
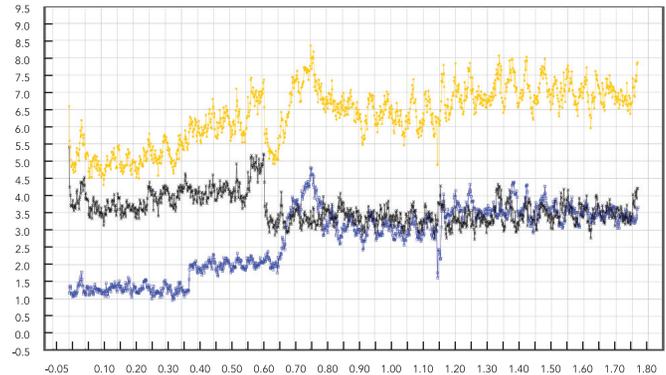
2c)
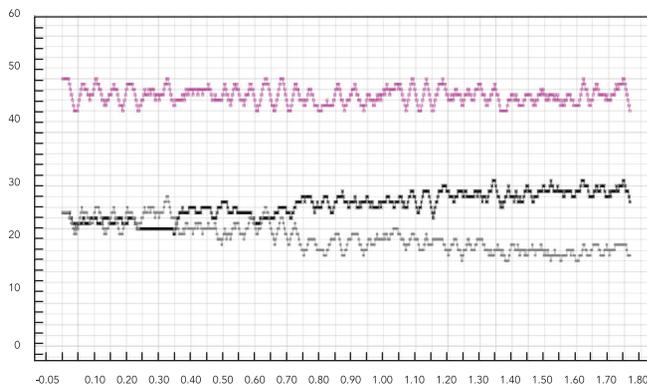
2d)
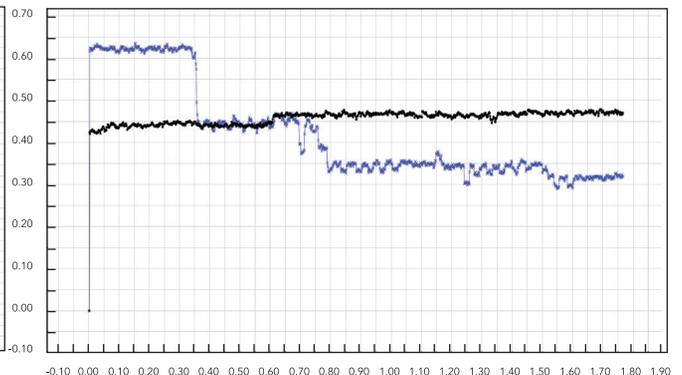
effort [42]. The third limitation is that our model did not fully exploit the opportunity to model dynamics and feedback processes. We consider this a significant shortcoming. Feedbacks were primarily structural (not behavioral) and there was no formal learning and adaptation. For example, households reacted to their environment based largely on static attributes (their income, location, and preference) and only a few dynamics (habitual/past behavior and distance to store, which was somewhat dynamic due to move-out/in store sub-model). Store agents exhibited only slightly more feedbacks: customer volume determined the probability of moving-out/in, which also enabled probabilities of changing the store's food type.

Figure 3

Example of summary data from different scenarios. The x-axis show five store behavior algorithms that used various probabilities for store move-out/move in and changes in the type of foods are sold at the stores. The purpose of this plot was to illustrate how changes in store dynamics influence diet. Each algorithm was run 60 times to obtain the distribution of outcomes and then summarized as the median (symbol in the graph) and the 5th to 95th simulation percentile (bar in the graph). The right y-axis represents the average proportion of times that households shopped at a healthy food store. The left y-axis represents the difference in proportions between diet of high- and low-income households. Diet was derived from the average proportion of times the household shopped at a healthy food store [11].
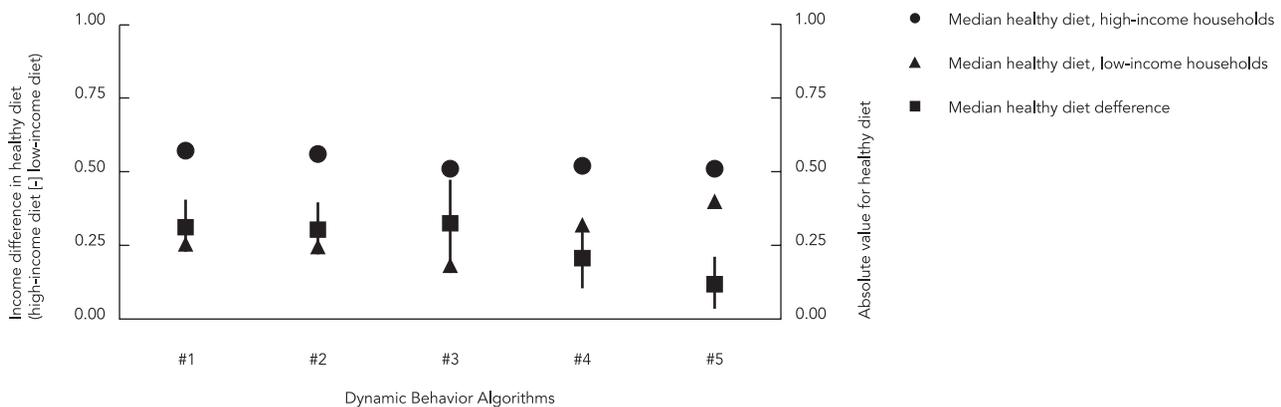
No dynamics   1   Static scenario: stores never close

2   Stores close but never change type of food

3   Urban decay scenario: stores close and do not re-open

4   Stores close and 10% of the time change the type of

High dynamics

Challenges and opportunities for modeling

Systems modeling tools are still new for urban health researchers, but could be applied to a diverse series of questions, such as: Under what conditions do particular urban problems change over time and why?; To what extent are interactions and feedbacks within and between entities shaping particular urban problems?; Under what conditions are we most likely to see unintended consequences to a planned intervention?. However, operationalizing these types of questions in a computational model will be difficult for many researchers. Modeling requires a large investment of time in computer programming and requires new ways of thinking. What follows are a few reminders when undertaking this work.

Focus on dynamics and feedbacks

Complex systems models require training ourselves to think differently. The earliest stages of model conception are difficult. Despite wanting to go beyond traditional linear thinking, it can be challenging to envision how multiple influences and pathways are more than independent correlations among components and focus on feedbacks and interdependence between entities rather than direct causal linkages.

Complex systems does not mean complex computational models

Given the limitless range of options available in ABM, the beginning modeler must ask a narrow question, work to establish a clear model purpose, and ignore processes that are not directly

Figure 4

Outline of high-level description of the model presented in Auchincloss et al. [11].

INITIALIZATION

Set parameters: either user defined, or read from a parameter file

Create the grid for households and stores

Place households:

    Make the specified number of households

    Place each one on a random open square until grid is full

    Assign household income using input settings

Place stores:

    Make the specified number of stores

    Place each one on a random square with no store already placed there (OK if square has a household)

    If we're clustering all stores, make sure the store is placed in the cluster region

    Assign price and quality based on settings

For each store:

    Reset variables

For each household:

    Reset variables

    Randomly determine initial quality preference based on the specified parameter

EACH TIME STEP

For each household:

    Calculate utility for all stores (multiply scores by weights, sum, and add in random noise)

    Choose the store with highest utility

    Update statistics based on store chosen (total quality, distance traveled, money spent)

Every 15 time steps:

    Stores calculate and reset statistics (profit, # customers high and low income, # turns high and low quality)

    The model calculates all statistics (average health, distance, price, quality preference) by averaging all households and stores' statistics

If stores are not static:

    Every 30 steps, choose a store to close:

        Choose cheap vs. expensive randomly based on the proportion of stores that are initially specified as cheap

If we are allowing a random store to close:

    Choose whether a random (3% chance) or the worst performing (97% chance) store will close, otherwise choose the worst performing store to close:

        If random, select a random store of chosen price that has been open more than 180 time steps

        If worst performing:

            Find the store with least number of customers

            If that store has been open more than 180 time steps, select it

            Otherwise, no store closes.

    Tell the selected store to close: set store's closure variable to true and reset its counters

For each store:

    Keep track of number of turns that a store location has not had an open store and has had an open store

    If store location has not had an open store > 180 time steps:

    Store opens

    Store assumes a food store type opposite of the previous store

        Store assumes a food store price that is the same as the previous store (or if systematic price differential is activated, then store assumes food store price that is linked to the food store type)

Households calculate and reset statistics (health, average money spent, average distance traveled)

---

relevant. Those new to ABM will be surprised to find that a very simple question becomes very complicated to operationalize.

Remain vigilant about deterministic modeling

The researcher needs to constantly check that (s)he is not establishing conditions or behavior rules that essentially already verify the hypothesis of interest. For example, if one wants to explore income differentials in diet but our "base" models fix expensive stores as having healthy foods, then one would essentially pre-determine an income differential in the base model for all scenarios.

Take a sensible approach to assessing reliability and validity

Do not become preoccupied with calibration and validation. Creating reliable and valid models is a difficult undertaking and should be approached sensibly. Many researchers spend most of their time and energy on calibration and validation and no time and energy remain for expanding on the science and exploring the most important questions.

Do not overpromise results

Due to the stochastic nature of micro-processes, ABM is not appropriate for detailed prediction and outputs should not be interpreted as precise estimates.

Recognize that complex systems computation models are not for every purpose and every audience

Not all questions posed within a system framework need to be answered using a computation model. Even if the questions require a computational model, the type of product may not meet the researcher's needs to make it worth the effort. First, the greatest value from modeling often comes from the modeling process itself rather than from the final model and its outputs [12]. Second, model results allow for a qualitative interpretation that may not be satisfying to some audiences. Third, ABM results can be difficult to summarize and communicate, especially to audiences unaccustomed to interpreting simulations and ABM. For example, caveats need to be mentioned such as results are conditional on a confluence of other factors and on inputs and algorithms programmed into the model. Empirical research analyses also requires strong caveats/assumptions. However, because agent-based models are constructed under fully simulated conditions, some audiences will discount the value of findings from ABM.

In sum, conceptual and computational models of complex systems forces us to carefully identify problems and processes that are likely impacted by dynamics and feedbacks that we typically ignore. The process of envisioning these models can propel us to think more realistically about complex mechanisms and perhaps think more creatively about potential solutions. ABM is a new computational tool for urban health researchers to use to address seemingly intractable urban health problems. Researchers will need to evaluate for themselves whether it is a promising tool for their own research question.

Resumen

Existe un interés creciente entre los estudiosos de la salud urbana en trabajar con problemas complejos, utilizando modelos conceptuales y computacionales del campo de sistemas complejos. La modelación basada en agentes (MBA) es una herramienta de modelación computacional que suscita cada vez más interés. Sin embargo, varios estudiosos todavía no se encuentran familiarizados con el desarrollo e implementación de un MBA, lo que dificulta su aplicación y comprensión. En este artículo se ofrece inicialmente una breve guía introductoria para llevar a cabo un simple modelo basado en agentes. De esta manera el método se ilustra discutiendo un modelo basado en agentes, desarrollado previamente, que explora las desigualdades en la dieta en un contexto de segregación residencial urbana.

Simulación por Computador; Métodos Epidemiológicos; Teoría de Sistemas; Salud Urbana

References

1. Sterman JD. Learning from evidence in a complex world. Am J Public Health 2006; 96:505-14.
2. Rydin Y, Bleahu A, Davies M, Davila JD, Friel S, De Grandis G, et al. Shaping cities for health: complexity and the planning of urban environments in the 21st century. Lancet 2012; 379:2079-108.
3. Diez Roux AV. Conceptual approaches to the study of health disparities. Annu Rev Public Health 2012; 33:41-58.
4. Barrett CL, Eubank SG, Smith JP. If smallpox strikes portland... Sci Am 2005; 292:54-62.
5. Yang Y, Diez Roux AV, Auchincloss AH, Rodriguez DA, Brown DG. A spatial agent-based model for the simulation of adults' daily walking within a city. Am J Prev Med 2011; 40:353-61.
6. Bonabeau E. Agent-based modeling: methods and techniques for simulating human systems. Proc Natl Acad Sci U S A 2002; 99 Suppl 3:7280-7.
7. Macy MW, Willer R. From factors to actors: computational sociology and agent-based modeling. Annu Rev Sociol 2002; 28:143-66.
8. Auchincloss AH, Diez Roux AV. A new tool for epidemiology: the usefulness of dynamic-agent models in understanding place effects on health. Am J Epidemiol 2008; 168:1-8.
9. Grimm V, Railsback SF. Individual-based modeling and ecology. Princeton: Princeton University Press; 2005.
10. Axelrod R. Advancing the art of simulation in the social sciences. In: Conte R, Hegselmann R, Terna P, editors. Simulating social phenomena. Berlin: Springer; 1997. p. 21-40.
11. Auchincloss AH, Riolo RL, Brown DG, Cook J, Roux AVD. An agent-based model of income inequalities in diet in the context of residential segregation. Am J Prev Med 2011; 40:303-11.
12. Miller JH, Page SE. Complex adaptive systems: an introduction to computational models of social life. Princeton: Princeton University Press; 2007.
13. Epstein JM, Axtell R. Growing artificial societies: social science from the bottom up. Washington DC: Brookings Institution Press; 1996.

14. Railsback SF, Grimm V. Agent-based and individual-based modeling: a practical introduction. Princeton: Princeton University Press; 2012.

15. Epstein JM. Why model? J Artif Soc Soc Simul 2008; 11:12.

16. Rykiel Jr. EJ. Testing ecological models: the meaning of validation. Ecol Modell 1996; 90:229-44.

17. Sterman JD. Business dynamics: systems thinking and modeling for a complex world. Boston: Irwin McGraw-Hill; 2000.

18. Simon HA. Rationality in psychology and economics. J Bus 1986; 59(4 Pt. 2):S209-24.

19. Orr MG, Plaut DC. Complex systems and health behavior change: insights from cognitive science. Am J Health Behav 2014; 38:404-13.

20. Martinez-Moyano IJ, Macal CM. Exploring feedback and endogeneity in agent-based models. In: Proceedings of the 2013 Winter Simulation Conference. Simulation: Making Decisions In A Complex World. Washington DC: IEEE Press; 2013. p. 1637-48.

21. Alden K, Read M, Andrews PS, Timmis J, Coles M. spartan: simulation parameter analysis R toolkit application. R J 2014; 6:63-80.

22. Gilbert GN, Troitzsch KG. Simulation for the social scientist. 2nd Ed. Maidenhead: Open University Press; 2005.

23. Vanni T, Karnon J, Madan J, White RG, Edmunds WJ, Foss AM, et al. Calibrating models in economic evaluation: a seven-step approach. Pharmacoeconomics 2011; 29:35-49.

24. Oreskes N. Evaluation (not validation) of quantitative models. Environ Health Perspect 1998; 106 Suppl 6:1453-60.

25. Sterman JD. All models are wrong: reflections on becoming a systems scientist. Syst Dyn Rev 2002; 18:501-31.

26. Ghorbani A, Dijkema G, Schrauwen N. Structuring qualitative data for agent-based modelling. J Artif Soc Soc Simul 2015; 18:2.

27. Kravari K, Bassiliades N. A survey of agent platforms. J Artif Soc Soc Simul 2015; 18(1). http://jasss.soc.surrey.ac.uk/18/1/11.html.

28. Grimm V, Berger U, Bastiansen F, Eliassen S, Ginot V, Giske J, et al. A standard protocol for describing individual-based and agent-based models. Ecol Modell 2006; 198:115-26.

29. Grimm V, Berger U, DeAngelis DL, Polhill JG, Giske J, Railsback SF. The ODD protocol: a review and first update. Ecol Modell 2010; 221:2760-8.

30. Müller B, Bohn F, Dreßler G, Groeneveld J, Klassert C, Martin R, et al. Describing human decisions in agent-based models – ODD+D, an extension of the ODD protocol. Environ Model Softw 2013; 48:37-48.

31. Sobal J, Stunkard AJ. Socioeconomic status and obesity: a review of the literature. Psychol Bull 1989; 105:260-75.

32. Beydoun MA, Wang Y. Do nutrition knowledge and beliefs modify the association of socio-economic factors and diet quality among US adults. Prev Med 2008; 46:145-53.

33. Zenk SN, Schulz AJ, Israel BA, James SA, Bao S, Wilson ML. Fruit and vegetable access differs by community racial composition and socioeconomic position in Detroit, Michigan. Ethn Dis 2006; 16:275-80.

34. Moore LV, Diez Roux AV. Associations of neighborhood characteristics with the location and type of food stores. Am J Public Health 2006; 96:325-31.

35. Schwanen T, Mokhtarian PL. What affects commute mode choice: neighborhood physical structure or preferences toward neighborhoods? J Transp Geogr 2005; 13:83-99.

36. Drewnowski A, Darmon N. The economics of obesity: dietary energy density and energy cost. Am J Clin Nutr 2005; 82(1 Suppl):265S-73S.

37. Ingram DR. An evaluation of procedures utilised in nearest-neighbour analysis. Geogr Ann Ser B 1978; 60:65-70.

38. Yoo S, Baranowski T, Missaghian M, Baranowski J, Cullen K, Fisher JO, et al. Food-purchasing patterns for home: a grocery store-intercept survey. Public Health Nutr 2006; 9:384-93.

39. Cobb CW, Douglas PH. A theory of production. Am Econ Rev 1928; 18:139-65.

40. Jekanowski MD, Binkley JK. Food spending varies across the United States. Food Rev 2000; 23:38-51.

41. Dunkley B, Helling A, Sawicki DS. Accessibility versus scale: examining the tradeoffs in grocery stores. Journal of Planning Education Research 2004; 23:387-401.

42. Huang QX, Parker DC, Filatova T, Sun SP. A review of urban residential choice models using agent-based modeling. Environ Plann B Plann Des 2014; 41:661-89.