

Article

“Space, the Final Frontier”: How Good are Agent-Based Models at Simulating Individuals and Space in Cities?

Alison Heppenstall ^{1,*}, Nick Malleson ¹ and Andrew Crooks ²

¹ School of Geography, University of Leeds, Leeds LS2 9JT, UK; n.s.malleson@leeds.ac.uk

² Computational Social Science Program, George Mason University, 4400 University Drive, MS 6B2, Fairfax, VA 22030, USA; acrooks2@gmu.edu

* Correspondence: a.j.heppenstall@leeds.ac.uk; Tel.: +44-113-343-3361; Fax: +44-113-343-3308

Academic Editors: Koen H. van Dam and Rémy Courdier

Received: 5 December 2015; Accepted: 19 January 2016; Published: 26 January 2016

Abstract: Cities are complex systems, comprising of many interacting parts. How we simulate and understand causality in urban systems is continually evolving. Over the last decade the agent-based modeling (ABM) paradigm has provided a new lens for understanding the effects of interactions of individuals and how through such interactions macro structures emerge, both in the social and physical environment of cities. However, such a paradigm has been hindered due to computational power and a lack of large fine scale datasets. Within the last few years we have witnessed a massive increase in computational processing power and storage, combined with the onset of Big Data. Today geographers find themselves in a data rich era. We now have access to a variety of data sources (e.g., social media, mobile phone data, *etc.*) that tells us how, and when, individuals are using urban spaces. These data raise several questions: can we effectively use them to understand and model cities as complex entities? How well have ABM approaches lent themselves to simulating the dynamics of urban processes? What has been, or will be, the influence of Big Data on increasing our ability to understand and simulate cities? What is the appropriate level of spatial analysis and time frame to model urban phenomena? Within this paper we discuss these questions using several examples of ABM applied to urban geography to begin a dialogue about the utility of ABM for urban modeling. The arguments that the paper raises are applicable across the wider research environment where researchers are considering using this approach.

Keywords: cities; agent-based modeling; big data; crime; retail; space; simulation

1. Introduction

By 2050 the United Nations (UN) predicts that around 66% of the world’s population will be living in urban areas [1]. The greatest future expansion is expected to be in cities in developing regions. These cities will face significant challenges in meeting the needs of its residents such as the provision of adequate housing, energy, health care, education, transport and employment. Planners have traditionally used aggregate models such as spatial interaction models (e.g., [2]) for formulating policies and plans for the design and growth of cities. However, these have several drawbacks due to their aggregate treatment of individuals and their lack of dynamics and behavioral realism [3,4]. This has led to an increased interest in using individual-based approaches from Geocomputation such as cellular automata [5] and agent-based modeling [6] for improving our understanding of the processes and dynamics within cities, and in particular simulating how cities may grow in the future.

Simulating the dynamics and processes of cities has been a focus of quantitative geographers for over 50 years. The earliest attempts by researchers to simulate and understand the inner mechanisms

of these systems came from the application of formal mathematical tools. These included Alonso's [7] bid rent model, Hagerstrand's [8] diffusion model and gravity models of spatial interaction put forward by Wilson [9] and Fotheringham and O'Kelly [10]. These models had one central common intellectual caveat: in order to say something useful about spatial systems, analysis must take place at the aggregate level. Of course, this meant that any interesting variation or noise that could be indicative of meaningful processes within the city was "aggregated" out. In essence, parsimony won out and cities were distilled down into homogeneous units whereby it was virtually impossible to say anything meaningful about any of the inner workings or micro dynamics [4].

The beginnings of digital computation began to emerge in the 1950s and led to the creation of "virtual laboratories" that enabled experiments to be performed on synthetic cities [6,11]. This led to a new perspective on how social systems should be modeled, based largely on concepts drawn from other disciplines such as physics and biology. These new ideas changed our understanding about how urban systems were organized and how they should be modeled. However, a lack of data and computing power meant that most of this work was focused at the aggregate level [3]. For reasons of simplicity and parsimony, it was believed that homogeneity in populations was the best way to account for behavior in space and time [12].

Increases in computer processing power and data storage, coupled with the new availability of varied detailed data sources, opened up the study of cities to approaches such as cellular automata (CA) and agent-based modeling (ABM). Initially within the geographical modeling community, many urban problems—such as traffic [13] and urban growth [14]—were tackled from the CA perspective. Under this paradigm, the world is represented as a series of cells that possess individual "states". These can change based on transition rules and the states of the surrounding cells [15]. However, it was quickly realized that the fixed and homogeneous nature of the cells in a CA reduces the ability of the method to model phenomena that require features like mobility and autonomy in the individuals. Many researchers started to adapt CA models to make them more representative of real systems by, for example, giving the entities the ability to move around the environment and embedding heterogeneous (rather than homogeneous) transition functions to reflect individual behavior.

Batty [16] was one of the earliest to recognize the potential offered from looking at the individual/agent perspective for simulating the evolution of cities. However, the first example of an ABM in a geographical journal was that of Bura [17]. This model explored the evolution of settlements and laid the foundations to more complex urban growth models [18]. However, these early models suffered from a lack of detailed data and computational power to simulate more than a few thousand agents operating complex rules.

Interest in how well ABM can simulate urban dynamics has been partly led by changes in how geographers view cities. Cities are now seen as complex systems with researchers fully aware of the need to be able to simulate different components, their interconnections, behavior, feedback and emergence (e.g., [19,20]). While successes can be seen in modeling components of the system, for example in pedestrian modeling (e.g., [21]), traffic (e.g., [22]), and residential dynamics (e.g., [23]), a critical review of this area highlights weaknesses both in the nature of the application and more fundamentally in the methodology itself.

This paper is primarily concerned with examining the ability of agent-based models (ABMs) to handle space and individuals in cities. With respect to modeling the different components of cities, where are we doing well and, perhaps more pertinently, what challenges remain? Replicating human behavior is an oft lamented problem in simulating the dynamics and process at the micro-level in cities—is it possible and pragmatic to model everyone's behavior in a city (just because we can)? What part does Big Data have to play? And finally, what other problems do we still have to tackle firstly as geographers and then as modelers?

The focus of this paper is purely on urban simulation. Other closely related subjects, for example land use models are comprehensively reviewed by [24,25]. From a techniques perspective, we are strictly focusing on ABM. Interested readers are directed to [15,16] for a detailed review on applications

of CA. Section 2 presents an overview of how our view of cities has changed with Section 3 setting out how agents are potentially an excellent methodological match for simulating cities. Section 4 then makes use of a multitude of examples to critique how ABM has been used in city simulation, focusing in particular on difficulties in simulating behavior and calibration and validation issues. Section 5 introduces Big Data examining what it promises and its false hopes. Section 6 concludes the paper with a discussion of all the main questions posed.

2. Cities as Complex Entities

It has long been recognized that cities are complex systems. For example, Jacobs [26] wrote that cities are essentially “people systems” and hence should not be planned as mere networks of brick and mortar. Simply stated, complex systems can be characterized by a small number of rules or laws, applied at a local level and among many entities. They are capable of generating complex global phenomena such as collective behaviors, extensive spatial patterns, hierarchies, *etc.*, which are manifested in such a way that the actions of the parts do not simply sum to the activity of the whole [27]. By viewing cities as complex adaptive systems, much like natural systems, we gain unique perspectives into their inner-workings. By observing phenomena through local-level interactions we can move from the top-down studying of cities to that of a bottom-up approach which allows us to witness the emergence of previously unexpected macroscopic phenomenon (e.g., land rents) from individual level interactions (e.g., people competing for land). Recently, Batty [28] has argued that the scientific study of cities is changing “from thinking of ‘cities as machines’ to ‘cities as organisms’” which is evidenced by the explosion of ABMs inspired by complexity thinking ([6], also the focus of this special issue).

For example, cities are hierarchic, composed of interrelated subsystems (parts within parts) in which each subsystem is interdependent but can connect to a number of other subsystems. Such subsystems are able to self-organize without higher level direction. In economics for example, national and global markets evolve from locally interacting agents all pursuing their own goals. On a city level one could consider town centers, or sub-centers as such as those shown in Figure 1A. If we look at the connections between these elements we have a hierarchy as shown in Figure 1B. This argument echoes those of near-decomposability [29]: here a system (*i.e.*, the city) has subsystem components interacting among themselves “in clusters or subgraphs, and interactions among subsystems being relatively weaker or fewer but not negligible” ([30], p. 134). It is not just the town centers, *etc.* that are connected but also urban processes. A simple selection of these are shown in Figure 1C.

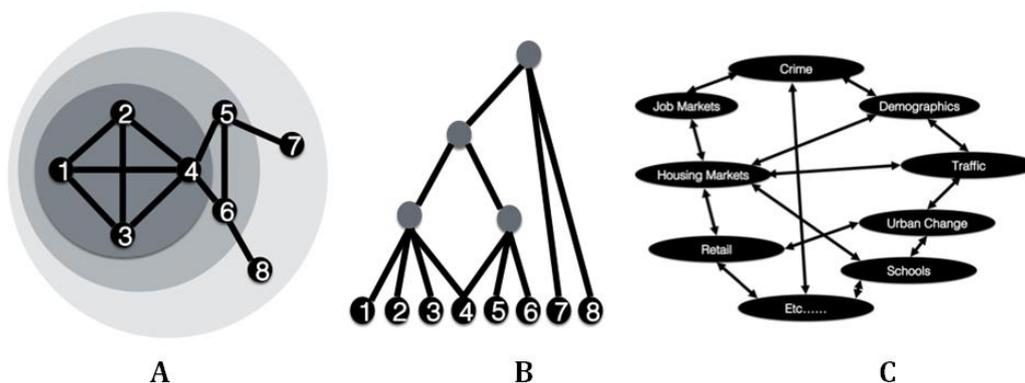


Figure 1. (A) System structure; (B) System hierarchy; and (C) Related subsystems/processes (adapted from [31]).

This systems approach to understanding cities is not new (see [31]). For example Christaller’s [32] Central Place Theory noted the hierarchal structures of villages, towns and cities. Simon [29] argues that hierarchy is a fundamental property of how a complex system holds itself together: “hierarchal organization from the bottom-up is essential for evolving systems and that hierarchical structures are

the way nature and society develop robust and resilient structures” ([31], p. 23). But these subsystems do not operate in isolation. In the short term they might appear to be independent from the rest of the system, but in the long run they are dependent on the aggregate system behavior.

Regardless of the subsystem that we are exploring, from a geographical science/urban modeling perspective there are three main aspects that we need to consider if we are to study cities from the bottom-up. We need to identify: the appropriate spatial scale; the appropriate temporal scale; and the behavioral process of the actors [24].

Determining the most appropriate spatial and temporal scales (we will address behavior in the following section), is a problematic area. Individuals in a city react over very different time periods and exert influence at very different spatial scales. Consider movement around a city for example: the behavior of a pedestrian will vary on a second-by-second basis and exert a spatial influence of only a few meters. A firm, on the other hand, might decide to relocate every 25 years and search for new premises over an area covering tens of kilometers. How should these vastly different spatial and temporal scales be rationalized in a single model?

Liu and Andersson [33] categorize land-use change processes into “slow”, “medium” and “fast”. Slow processes are classed as those on a 3–5 years time scale, for example, as affecting the physical structure of the city (industrial, residential and transport construction). Medium time processes are classed as those relating to economic, demographic and technological changes, which affect the use of physical structures and occur over months. Fast processes are classed as those that occur with a regularity less than one year, such as the mobility of labor, goods and information. All three affect the location of both residents and businesses, and all could be represented within the ABM using a series of internal clocks.

The issue of appropriate spatial and temporal scales are also pertinent when tied into the notion of dynamic complexity [34]. Here, “processes evoked are intricate and the phenomena that are produced are complex across many scales” ([35], p. 139). It is our challenge to capture and simulate these complex processes in order to better understand how cities grow and evolve.

3. Simulating Individual Behavior in the City

In the introduction it was noted that until relatively recently, for reasons of simplicity and parsimony, aggregate homogenized populations were used in attempts to simulate and understand geographical systems. While there has always been “resistance” to the notion that social scientists should “search for some atomic element or unit of representation that characterizes the geography of a place” ([12], p. 2), there has been a discernable shift from building simulations at the aggregate level to focusing instead on the individual level. This is due, in part, to the advent of Big Data and new computational methods (discussed below).

This change in simulation approach is developing in tandem with how we understand cities. Cities are now being viewed as organisms that are a product of networks, comprised of individual heterogeneous actors, interconnected at multiple levels [31]. When viewing a city in this way, the emphasis is on modeling to capture new emergent properties arising when the individual components of a complex system interact. In this context, we can see why the notion of an agent (representing any component of a geographical system; humans, buildings, organizations) has become such an attractive option.

ABM can be defined as the study of systems that possess “interacting autonomous entities, each with dynamic behavior and heterogeneous characteristics” ([36], p. 39). With agents, we can represent individual artifacts at multiple levels, give them the ability to learn, evolve, and make decisions adaptively in both space and time (see [37] for a detailed introduction to ABM). This unique ability of ABM to represent individual decision-making and interactions allows the researcher to examine complex systems, such as cities, that are defined by heterogeneity, feedbacks through interactions, and adaptations. Two of these characteristics of ABM are worth looking at more closely in context of cities; their ability to represent (i) behavior and (ii) feedback from interactions.

Humans, with their complex and adaptive behavior, are notoriously difficult to model [38]. Traditional methods to model human behavior were often based around rational choice theory [39], where it is assumed that humans behave in ways to maximize their benefits or minimize their costs. However, people rarely meet the requirements of rational choice models [40] in the sense that most, if not all people have limited cognitive abilities and limited time to make decisions [29]. Through the ability to model heterogeneity within ABMs we can capture the uniqueness that makes us human, in the sense that all humans have diverse personality traits (e.g., motivation, emotion, risk avoidance) and complex psychology [41]. We also know that human behavior is influenced by others (e.g., [42]) through their social networks which can introduce positive and negative feedbacks into the system. These properties again can be captured through the agent's heterogeneity and active status. As O'Sullivan comments of ABM:

“The appeal is undeniable: it appears obvious that individual-level decision-making is the fundamental driver of social systems. . .” ([43], p. 113)

In comparison to “traditional” statistical methods (e.g., utility maximization, spatial interaction modeling), ABM has several advantages. In brief these are:

- (i) Implicit representation of individual micro-dynamics—statistical models can only represent these interactions if the population is homogeneous or has coordinated or coherent interactions;
- (ii) Representation of potentially multiple spatial relationships;
- (iii) The structure of most ABM platforms are generally flexible enough to incorporate equations, statistical techniques, *etc.*, whereas the converse often is not true.

Use of ABMs therefore gives rise to a more heterogeneous approach that one hopes will reflect the richness and diversity of reality [12]. However, ABM is not without its criticisms. Whilst there has been a great advancement in standardizing model descriptions (such as the Overview, Design concepts and Details (ODD) protocol [44]), and carefully considered multi-level model evaluation (e.g., pattern-orientated modeling [45]), there remains significant work to be done in the areas of calibration and validation as a whole, embedding behavioral modeling and linking models and processes across scales. The extent to which these are addressed in city simulation is discussed below with reference to case-studies.

4. ABM for City Simulation

From a methodological standpoint, the ideological framework of ABM can be readily applied to a complex system such as a city. However, there are a number of critical issues that can be seen as dangerously prohibitive to successful implementation; these are calibration and validation, linking together processes at different scales, and simulation of behavior. Within this section, we pull examples from the published literature to assess how well ABM handles these specific challenges.

Table 1 presents a summary of the studies that will be used within the discussion below. Within the table we provide information about the application domain, the spatial scale(s) that they operate within, what are the main agents under investigation, their temporal resolution, how is behavior represented and whether or not they discuss explicitly verification, validation, and calibration. These studies were chosen due to being representative of other studies in their area, but also focused on real world locations. In this sense they use actual geographic information in the simulation. We do this because the role of modeling ranges across the spectrum from abstract models which are used for the development of theory to models whose purpose is for explanation and projection [46]. Thus how one should evaluate models with respect to verification, validation and calibration will vary. For example with abstract models such as [47], which explores urban creativity and development, often the verification and validation strategy is theoretical comparisons to others works and the ability to replicate existing patterns. Such styles of models are in contrast to models based on real world locations which attempt to explain and project what might happen in the future. For such models the validation ranges from qualitative to quantitative agreement with the emergent “macro-structures” through various goodness of fit measures [48].

Table 1. Summary of the studies used in the discussion.

Author	Application	Entity	Behavior	Spatial Scale	Temporal Scale	Verification (Y/N)	Validation (Y/N)	Calibration (Y/N)
[49]	Public Event	Individuals	Mathematical	Neighborhood	Seconds	N	N	Y
[50]	Riots	Individuals	Mathematical	Neighborhood	Seconds	Y	N	N
[51]	Indoor Movement	Individuals	Mathematical	Indoor Scene	Seconds	Y	Y	N
[52]	Disease propagation	Individuals	Mathematical	City	Minutes	Y	Y	Y
[53]	Disease propagation & urban traffic	Individuals	Mathematical	City	Seconds	Y	Y	Y
[54]	Crime	Individuals	Cognitive Framework	Neighborhood	Minutes	Y	Y	Y
[55]	Crime	Individuals	Mathematical	City	Hours	Y		N
[22]	Traffic	Individuals	Mathematical	City Center	Seconds	N	N	N
[56]	Flooding	Individuals	Mathematical	Town	Minutes	N	Y/N	N
[57]	Retail	Individuals	Mathematical	City	Days	N	Y	Y
[23]	Residential Location	Individuals	Mathematical	Neighborhood	Years	N	N	Y
[58]	Informal Settlement Growth	Households	Mathematical	Neighborhood	Days	Y	Y	N
[59]	Regeneration	Households	Mathematical	Neighborhood	Years	N	Y	Y
[60]	Urban Shrinkage	Households	Mathematical	City	Years	N	Y	Y
[61]	Urban Growth	Institutions & Developers	Mathematical	Region	Years	N	N	N
[18]	City Systems	City	Mathematical	Countries & Continents	Years	Y	Y	Y

As can be seen in Table 1, one question that preoccupies researchers is “how to select the most representative scale for an application in terms of agents (entities), spatial and temporal scales?” Fortunately, the underlying rationale for using ABMs is the notion of complexity, which focuses on a “bottom-up” approach to modeling cities and provides a ready solution with the emphasis on representing the smallest individual unit of interest. The examples in Table 1 present a range of agent representation from households to entire cities or institutions. The choice of agents depends on the problem being investigated be it people [54], facilities (e.g., petrol stations [57]) or cities [18]).

With respect to modeling micro-scale processes, perhaps the most successful applications of ABMs have been applied to pedestrian and traffic modeling (see [62] for a review). Here, the ability of ABM to handle dynamic complexity can be readily exploited. Early pedestrian models used rules based on the dynamics of physical systems (e.g., molecules of gas) [63], before moving to more advanced swarm algorithms for routing (e.g., [49]). Today, even richer cognitive pedestrian models are being developed (e.g., [50]). Such models operate over small areas such as indoor areas, city blocks or small neighborhoods [49–51]. The temporal resolution is often in seconds (or part of) and the models represent 10s to 1000s of agents. When researchers are modeling larger areas or large number of agents, often the temporal resolution increases to minutes (e.g., [52,54]) or hours (e.g., [55]) or agent interactions are simplified (e.g., [53]). Often this is done as a compromise between computational resources (*i.e.*, clock cycles) and the phenomena being modeled. For example, [52] argue that representing time in minutes is appropriate because, although symptoms of cholera takes several hours to show after exposure, capturing and understanding the movement and activities of individuals within the refugee camp is a vital aspect of predicting how the disease will spread.

Bridging the gap between micro and macro are a set of models that look at larger areas; ranging in scale from neighborhoods within cities, to whole cities themselves. Perhaps the most disaggregate ABMs within this category are that of traffic models such as [22,53] who model individual travel patterns, while others explore residential locational decisions at the household level over years (e.g., [23]). Often in such models, the temporal unit of analysis is months or years (*i.e.*, it does not make sense to model residential choice at a second by second basis because there will rarely be any change in behavior from one iteration to another which adds unnecessarily to the computational cost).

There are very few ABMs that attempt to model entire cities or counties. For example [60] model urban shrinkage in Leipzig, Germany, where individual households are simulated to model intra-urban migration and residential vacancy rates over years. Other models such as [61] explore urban growth along the West Yellow River Corridor of China. In this model the agents represent developers, conservationists and regional-planners. Individual households are not represented for the sake of modeling simplicity. A similar simplification is made in the SimPop family of models (see [18,64]). These look at systems of cities, whereby the growth of interconnected cities is modeled through treating each city as an agent.

Each of the applications have a central commonality: the phenomena under investigation drives the temporal clock of the agents. While this is entirely appropriate for many applications with well documented phenomena, what is not clear is whether ABM researchers are also analyzing their systems to (i) uncover which processes are the most important to consider at which scales and (ii) link the drivers of different processes at different spatial scales together.

One of the hallmarks of ABM is its ability to capture and model human behavior—ironically, this is also one of the areas that ABM is heavily criticized within. In Table 1 we have classified how models represent behavior in 2 ways; either as a mathematical approach or through the use of cognitive frameworks (following the typology set out by Kennedy [38]). The mathematical approach centers on the custom coding of behaviors within the simulation, such as using random number generators to select a predefined possible choice (e.g., to buy or sell) [65]. But, as noted above, human behavior needs to be understood in the context of individual preferences, memory, environment, *etc.* This has led to the development of more nuanced methods. For example, threshold-based rules have been incorporated—*i.e.*, when an environmental parameter exceeds a predetermined threshold this triggers a specific agent behavior, for

example, moving to a new location within a neighborhood [66]. These approaches have great utility when the behavior is both well understood and documented (as seen in Table 1).

A second approach to modeling human behavior uses conceptual cognitive frameworks. Within such models, instead of using thresholds, more abstract concepts such as beliefs, desires, and intentions (BDI, [67]) or physical, emotional, cognitive, and social factors (PECS, [68]) are imbued in individual agents. Both the BDI and PECS frameworks have been successively applied to modeling human behavior in a number of applications such as what drives people to crime—for example see [69] for an a-spatial model that uses the BDI framework and [54] for a geographically explicit model that uses the PECS framework). These conceptual cognitive frameworks and mathematical approaches for representing behavior can be considered as a rule based systems and are often applied to tens to millions of agents.

A third approach is through the use of cognitive architectures, (e.g., Soar [70] and ACT-R [71]). Cognitive architectures focus on abstract or theoretical cognition of one agent at a time with a stronger emphasis on artificial intelligence compared to the previous two approaches. Such an approach is rarely applied to more than one or two agents therefore their utility for large scale geographically explicit models is currently limited. The authors are unaware of any applications of this architecture to geographically explicit models.

This brief review raises important issues. If the main selling point of agents are their ability to replicate realistic behavior, why is this not reflected in the application of these frameworks? For example, frameworks such as PECS and BDI can be applied to a specific empirically grounded situations that produce specific behaviors which in turn can be compared with empirical data. However, there is still a strong tendency to build rule sets that by their nature can only ever support abstract behavior. For applications where the behaviors are not well documented or understood (*i.e.*, the product of emergent behavior), it is entirely possible that we are overlooking the key elements that drive social systems. Our understanding of the individual-level drivers behind complex geographical systems is deepening, a corresponding effort is therefore needed to improve the behavioral realism in our individual-based models, the entities that are driving these systems. This cannot be achieved by simply adding in more rules and probabilities to our models, instead we need to build better tools to identify these important behaviors and drivers that can feed directly into representation of behavior in agents. In the era of Big Data, perhaps we should also be considering the notion that in the very near future, we will be in possession of so much individual level behavioral data that these behavioral frameworks will become obsolete! It is important to note here that this is not a notion that we subscribe to, but it does have interesting parallels with the “end of theory” [72] arguments that were briefly put forward as “big data” became popular.

Calibration and validation are the processes by which we first fine tune the dynamics of our model, and secondly test its performance on an independent data set. Whilst much has been written of the principles (e.g., [73,74]) and statistics/pattern identification [45] that we should abide by, much of the process of calibrating and validating ABMs remains a dark art. Certainly this is an area within ABM that has not developed as quickly (in terms of rigor and metrics) as the rest of the discipline and for this reason is openly criticized by both proponents and critics [75–79]. A large part of the problem arises due to the nature of ABM applications. As ABM can be applied to almost any scale and represent any entity, a “one size fits all” approach will not work. Methods and metrics are required that can readily identify patterns and processes at different spatial and temporal scales. Added into this, we need an abundance of data to allow calibrate and validate potentially thousands of heterogeneous agents operating distinctly individual rule sets.

It is clear through the review undertaken that, although evaluation is attempted to some extent in most models, there are no published guidelines or standard approaches that researchers can draw on to evaluate an ABM. As already noted, the type of evaluation is very much dependent on the nature of the application; it is no surprise that there is no consensus about how one should verify, calibrate or validate a model. For example, in the brief review in Table 1, calibration of individual parameters

was either not mentioned at all (e.g., [22]), based on relevant literature (e.g., [52]) or data analysis (e.g., [57]). Validation was discussed through engagement with stakeholders (e.g., [56]), structural validation [55] or [54] used experiments to validate the output of the model. In some papers, no explicit mention of validation is given but it is clear that model results were compared to real world data (e.g., [23,60]), whilst others clearly do discuss the role of validation and verification but do not explicitly mention them in context of the results (e.g., [58]). Others say validation will be a future task (e.g., [61]). However, one area within ABM where progress has been made is within pedestrian modeling which is basking in a proliferation of data due to motion capture techniques [21,51], but there still remains room for improvement, especially in unknown conditions such as riots [50]. Another area where validation is performed with more rigor is within land-use changed models. Here results are validated by comparison at the smallest unit (normally pixel-by-pixel). However, this form of analysis is restricted to raster formats of data and as evidenced above, validation of models using vector data is less developed [80]. While much work has been carried out on validation with respect to the similarity of model outputs to macro structures, little attention has been paid to those models that exhibit not only quantitative agreement with empirical macro structures but also quantitative agreement with empirical micro-structures (*i.e.*, the agents being modeled [48]). As Robert Axtell is quoted as saying in [81] "... there is a large research program to be done over the next 20 years, or even 100 years, for building good high-fidelity models of human behavior and interactions."

Through innovations in modeling methodologies, such as ABM, and rapid increases in the availability of computer and storage facilities, it is now possible to model systems at an extremely fine level of detail and capture complex phenomena in a way that is not possible with traditional aggregate approaches. However, the availability of essential data required for theorizing, calibrating, and validating such models lags behind their new capabilities. For example, research discussed above [57] was able to capture system dynamics at an extremely high resolution, but lacked essential data required to substantiate the behavioral assumptions embedded in the agents. With the advent of Big Data, modelers potentially have access to a wealth of information that might go some way to resolving the critical issues of model evaluation and accurate behavioral simulation [82].

5. Big Data

The term "Big Data" is misleading. It not only refers to data sets that are unusually large in volume, but those that also exhibit other properties that distinguish them from "traditional" datasets. These include velocity (data that are generated rapidly and might only be relevant for a short amount of time), variety (data that are stored in various formats and hold diverse pieces information), veracity (there are uncertainties around bias, noise and the level of representation) as well as any number of new "V's" (validity, value, volatility, *etc.*) that go beyond Laney's [83] original three. For a more complete breakdown of the key features, see [84]. Considering these features is important because new methods are required in order to derive insight from the data.

In the general public and academic discourse, Big Data has become a prevalent topic as highlighted in Figure 2. New digital information is emerging about aspects of people's lives that previously went undocumented. These data can be broadly broken down into the following categories [84]:

- **Automated** data are those that are collected covertly/discretely and often by a third party. These include records of: individual movement (e.g., travel cards, automatic number plate recognition systems, pedestrian flow counters); websites visited; consumer behavior (e.g., spending on credit/debit cards, loyalty card schemes); environmental conditions (e.g., air quality, light/sound levels); health (e.g., life tracking, activity monitoring); and a wealth of others.
- **Volunteered** data are those that are donated freely by individual users (this assumes, of course, that contributors are aware that their contributions will be public). These include: messages posted to social media services like Facebook, Twitter and foursquare; contributions to collaborative sites such as blogs, wikis, discussions, and OpenStreetMap; and uploaded media (e.g., photos and videos).

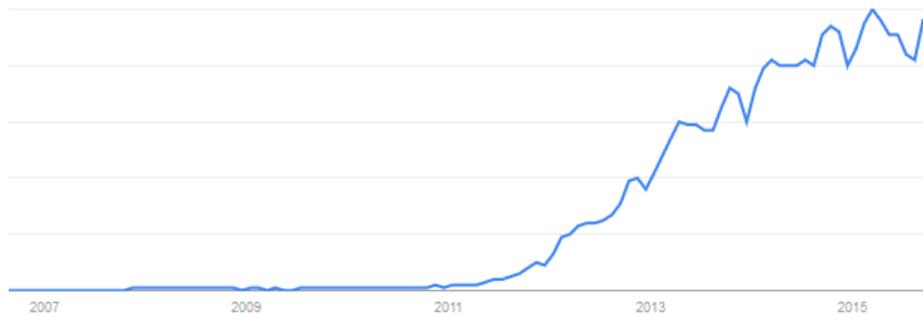


Figure 2. The “search interest” on Google in the phrase “big data” over time. Actual numbers are relative. Source: Google Trends [85].

Whether collected automatically or volunteered, these new data potentially contain a wealth of information to inform spatial models of urban processes. This does, of course, depend on resolving issues of bias, noise and generalization. Section 3 outlined two areas in which the need for the improvement of ABMs is particularly strong: understanding behavior; and calibration and validation. In both of these cases the issues often center around data availability and hence “big” sources have the potential for innovation in urban modeling. The proliferation of Internet-enabled devices such as smart phones has enabled individuals and third-party organizations to begin to capture digital information about aspects of peoples’ lives that have historically gone undocumented [82]. This “datafication” [86] might not only include an individual’s precise location in time and space, but also their thoughts, feelings, moods, and behaviors. Furthermore, this information is usually observed rather than reported, which overcomes some limitations associated with traditional activity-based surveys (for example see [87] for a discussion about recall and bias issues with time-activity diaries).

Recent examples of the use of Big Data to inform quantitative geographical research are broad. They include: the development of new area/neighborhood profiles using social media data [88–90]; estimates of the mobile population at risk of crime [91]; the identification of “important” places in peoples’ lives from mobile telephone data [92]; the detection and delineating of events [93,94]; analysis of regular mobility patterns [95,96]; classification of areas based on their Twitter temporal profile [97]; and a wealth of others. However, examples applied in the context of urban modeling, let alone ABM specifically, are much scarcer. That said, early attempts at using these new data to improve urban models are proving fruitful. For example, research has been pioneered [98,99] that uses social media data to calibrate models of human flows to museums and shopping centers respectively. The most relevant example is [100], where crowd-sourced data are used to update a road network in response to damage by an earthquake. An agent-based model is subsequently used to model the movements of individuals through the city towards aid centers. There are clear avenues for further integration of “big” data into studies such as this to meet some of the challenges posed earlier. For example, geo-located tweets might provide information about real activities on which to build improvements into the behavior of the agents, or on which to validate model results.

Interestingly, there are few other relevant examples, particularly regarding how “big” data can be used to improve ABMs through a better understanding of behavior and calibration/validation. One explanation for the relative scarcity of Big Data applications to urban models centers on the drawbacks. Commentators rightly point to issues around bias (which groups of people are absent from the data?) and reliability (how accurate are Global Positioning System (GPS) coordinates? Can you derive a person’s intended meaning from short texts?). In addition, the “3Vs” will also play an important role. Researchers are yet to come to terms with data that are high velocity, noisy, varied, as well as often being extremely large. This renders the data much less amenable to traditional methods of analysis and necessarily increases the difficulty in deriving useful insight. National censuses, for example, are clearly structured, freely available, static, and can usually undergo analysis on a desktop computer. These limitations of these new data, and some potential solutions, will be revisited in Section 6 below.

6. Discussion

The study of cities has undergone considerable change over the past 50 years. As computers have grown in power and our ability to store and process data has increased, we can begin to create models that exist at the micro level and can be linked to processes at different spatial and temporal scales. This shift has given rise to a deeper understanding of how cities evolve and how we can potentially intervene in their development to make them smarter and more sustainable.

How well have ABM approaches lent themselves to simulating the dynamics of urban processes? While the examples above show that disaggregation to the micro level has resulted in a greater understanding and appreciation of how city processes and dynamics work at this level, it could be argued that we are no better off than we were 30–40 years ago. With a lack of processes being linked at different spatial scales (see [26]), we are often still working at one spatial level, albeit at a much higher resolution. City dynamics can be fast or slow depending on the scale of analysis [101]. It is clear therefore that to simulate the essence of cities we need to be able to capture and link these dynamics. It is here where ABM can significantly contribute. As Batty remarks: “As the scale gets finer and the agents and their cells become more like real objects, their operationality increases to the point at which substantive policy applications are possible” ([101] p. 1393). However, as evidenced by discussion of the case studies presented, there are several areas such as calibration and validation, behavior and linking processes, where further work is needed.

By simulating at the individual level, we have given ourselves a new set of problems. Foremost, if we model a city of heterogeneous individuals how should we calibrate and then validate such a model? Aggregate level models, in this respect, could be said to be more robust in that they could be more successfully calibrated and validated. However, this rather misses the point. Cities are not homogeneous and aggregate entities. They are the product of hundreds and thousands of human decisions. Therefore to be able to understand and intervene in their development we need to be able to model with certainty at this individual level [19,31]. There is some irony in the fact that the continued disaggregation of variables of cities to greater and greater detail to give sufficient heterogeneity to allow for better representation has meant that it is near impossible (at present) to rigorously validate and calibrate.

If cities are the product of an individual’s decision, then there is an argument for including more fundamentally realistic behavior in our models. Almost all of the agents in the examples presented above are powered through mathematically-based rule sets. While this is appropriate for some examples such as petrol station agents [57], humans do not operate via a probabilistic set of rules. Often, decisions are made on incomplete data and a decision can be immediately changed based on the availability of new information. Rationality is a luxurious assumption in modeling human behavior. The counter argument to this is simple; would the actions of potentially millions of heterogeneous agents with their own behavioral framework simply cancel out as we “scale up”? And how could we tease out one piece of significant behavior that can potentially shift the behavior from one state to another when we potentially have millions of agents and enumerable individual interactions? Of course, we do not always need to model an entire population (although it is possible [102]) but for those occasions that we do, would looking towards other disciplines that handle large complex systems with numerous interconnected components—such as physics and meteorology—be useful?

While there are many clear reasons to use ABM for simulating complex spatial systems, O’Sullivan *et al.* [43] raise an interesting point: “Any gain in understanding of the system resulting from the modeling process derives from our ability to analyze the model and experiment with it” ([43], p. 113). While we have the raw materials to simulate urban systems at the individual level, we lack sophisticated tools to both evaluate our models and analyze the results. Do we need to rethink how we design our experiments? Could this be an area whereby development in tools for processing and understanding Big Data could make a meaningful contribution?

“Big” data sources, particularly those centered around the use of social media, might allow urban modelers to better understand the behaviors that should imbue their agents, and might simultaneously

offer new avenues for multi-scale calibration. Pattern-orientated modeling [45] is an approach to model calibration and validation that advocates the comparison of numerous patterns produced by models and their counterparts in real (non-simulated) data. Commonly in ABM, this involves analyzing model outcomes at multiple spatial and temporal scales. For aggregate analysis this is often unproblematic, but for patterns at higher spatio-temporal resolution this can be difficult when appropriate data are less forthcoming. For example, in a model of urban movements it might be advantageous to compare the movements of individual simulated agents to those of individual real people. There are a wealth of geo-located social media and administrative/corporate “big” data that—limitations and ethical implications aside—could be used to build an accurate, individual-level dataset of urban flows on which to calibrate models. Similarly by delving deeper into these data to explore any text or multimedia attachments as well as spatial location, modelers might learn more about how people use urban spaces and use this to inform the rules that drive their agents.

However, there are serious limitations to the use of these data. Firstly, the signal to noise ratio is often very low. Unlike data derived through traditional surveys that are usually designed for a specific purpose and cleaned/verified before release, “big” data are rarely created for the exact purpose required by the researcher. It can therefore be extremely difficult to find valuable insight among a mass of irrelevant (in terms of a specific research goal) pieces of information. Proponents would argue, however, that it is advantageous to have a large set of noisy data rather than a small, precise set: “Big data, with its emphasis on comprehensive data sets and messiness, helps us to get closer to reality than did our dependence on small data and accuracy” ([86], p. 48). This philosophy has worked well for many businesses (the continued success of Google is probably the most notable example) but the level of rigor required in academic research might make noise a more difficult hurdle. However, it has been long argued that the move towards understanding larger, noisier data sets is essential [103].

A second drawback is the extent to which the data are a reliable representation of the true underlying population. Unlike traditional surveys, Big Data are not drawn from the population using reliable sampling methods (*i.e.*, random, stratified or systematic). The digital divide [104] is alive and well, so participation inequality in digital services is extremely likely. Research continues to find disparities in access and participation with digital services [105,106] along the “familiar fault lines of social inequality: class, ethnicity, gender, age, and geographic location” ([107], p. 526). There will also be clear geographic inequalities—“‘big data’ have their own geographies” [108]—which might make it difficult to build models in regions that are poorly served by contemporary Internet services. There might be a wealth of data available for models of capital cities in developed countries, but what about smaller, less technologically advanced places? That said, the drawback of participation bias does not rule out their potential usefulness. A reliable weighting scheme, for example, may go some way to smoothing the inequalities in participation.

Thirdly, even if the data are not biased and useful information can be distilled from noise, it is not always clear how accurate the results are. Inaccuracy might arise along both technical social lines. Technical accuracy relates to aspects such as the locational accuracy of point coordinates. In spatial research with social media, GPS accuracy is often overlooked but there is evidence that positional errors can reach more than 2 km [109]. “Social” inaccuracies (for want of a better word) relate to the reliability of the core meaning of an individual datum. For example, someone might create a message that includes the word “happy”, but linguistic intricacies might mean that algorithmically discerning whether that message actually reflected a happy mood is far from trivial.

Finally, there are clear ethical implications that might limit the use of “big” data in modeling. The data are of particular concern because they usually represent individuals and might be predicated on relatively weak consent. For example, users might have given consent for use of their data (e.g., the use of their mobile phone, social media contributions, credit card purchases, *etc.*) when they sign a contract, but it is not always clear that this is a properly *informed* consent. Users can theoretically agree to some terms, but it is unlikely that most really understand what they are agreeing to. For example, the Apple iTunes UK Terms and Conditions (updated 30th June 2015) contains over 18,000 words.

It is unlikely that most users have read and understood those terms. This offers some fundamental challenges to traditional ethical frameworks. Ethical justification usually consists of referring to explicit consent, public availability, or by taking strict measures to ensure privacy (e.g., [110,111]). This position is not without merit: most authors are largely in agreement that “public” data published on the internet (including through social media) are suitable for research [112–115]. This is, of course, predicated on properly informed consent. These data are already being used in business, and, as we have argued here, could make an extremely valuable contribution to academic research. However, even if academia can evidence the valuable contribution that the use of such personal data can make to society and demonstrate that the data can be handled safely—e.g., stored securely, treated sensitively and ethically, and used to produce outcomes that are ultimately for social good—some caution is paramount. Revelations about the extent of personal surveillance by some governments (e.g., the National Security Agency (NSA) leaks [116]) and media reports about the depth of knowledge of some loyalty card customers (e.g., predicting pregnancy [117]) have the potential to foster a much more cautious public attitude towards the use of personal data. Similarly, the European Union (EU) is in the process of reviewing personal data protections partly in response to the wealth of personal data that are becoming available. Rather than rushing to make the most efficient use these sources, perhaps the role for academia is to foster greater democratic ownership, understanding, and surveillance of big data, and for greater privacy protection of individuals. A more cautionary approach does not need to be to the detriment of valuable research, (see [118] for advances in privacy-preserving data mining). In fact, such an approach might help to increase public confidence in the use of their data and foster an environment in which more data become available because individuals can be confident that it is not being (or going to be) exploited.

To summarize, “big” data offer the opportunity to reveal insight into individual behavior, actions, and the use of space. Assuming that the difficulties with noise, bias and accuracy can be overcome (or at least understood) and the research is agreed to be ethically sound, then it offers great potential as a means of improving agent-based models for urban simulation. Researchers have the methodological tools and the computational infrastructure necessary to execute massive simulations, but currently lack high-quality, hi-resolution data on which to predicate agent behavior and evaluate across various scales (e.g., [45]). Overcoming these issues is fundamentally important if ABM is to play a role in assisting both practitioners and policy makers in creating sustainable and smart cities in the future.

Acknowledgments: This work has been supported by the following Economic and Social Research Council grants (grant numbers ES/L009900/1, ESRC ES/F000405/1 and ESRC ES/K004409/1)

Author Contributions: The central arguments and structure of the paper were equally devised by each of the authors through a series of discussions.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. United Nations. *World Urbanization Prospects: The 2014 Revision*; Department of Economic and Social Affairs: New York, NY, USA, 2014.
2. Batty, M. *Urban Modelling: Algorithms, Calibrations, Predictions*; Cambridge University Press: Cambridge, UK, 1976.
3. Torrens, P.M. *How Land-Use-Transportation Models Work*; Centre for Advanced Spatial Analysis (University College London): London, UK, 2000.
4. Batty, M. Fifty years of urban modelling: Macro-statics to micro-dynamics. In *The Dynamics of Complex Urban Systems: An Interdisciplinary Approach*; Albeverio, S., Andrey, D., Giordano, P., Vancheri, A., Eds.; Springer Physica-Verlag: New York, NY, USA, 2008; pp. 1–20.
5. Torrens, P.M.; O’Sullivan, D. Cellular automata and urban simulation: Where do we go from here? *Environ. Plan. B* **2001**, *28*, 163–168. [[CrossRef](#)]
6. Heppenstall, A.J.; Crooks, A.T.; Batty, M.; See, L.M. *Agent-Based Models of Geographical Systems*; Springer: New York, NY, USA, 2012.

7. Alonso, W. *Location and Land Use: Toward a General Theory of Land Rent*; Harvard University Press: Cambridge, MA, USA, 1964.
8. Hagerstrand, T. *Innovation Diffusion as a Spatial Process*; The University of Chicago Press: Chicago, IL, USA, 1967.
9. Wilson, A. *Catastrophe Theory and Bifurcation: Applications to Urban and Regional Systems*; Routledge: Oxen, UK, 1981.
10. Fotheringham, A.S.; O'Kelly, M.E. *Spatial Interaction Models: Formulations and Applications*; Springer: New York, NY, USA, 1989.
11. Gilbert, N.; Troitzsch, K.G. *Simulation for the Social Scientist*, 2nd ed.; Open University Press: Milton Keynes, UK, 2005.
12. Batty, M.; Crooks, A.T.; See, L.M.; Heppenstall, A.J. Perspectives on agent-based modelling. In *Agent-Based Models of Geographical Systems*; Heppenstall, A., Crooks, A.T., See, L.M., Batty, M., Eds.; Springer: New York, NY, USA, 2012; pp. 1–18.
13. Nagel, K.; Schreckenberg, M. A cellular automaton model for freeway traffic. *J. Phys.* **1992**, *1*, 2221–2229. [[CrossRef](#)]
14. White, R.; Engelen, G. Cellular automata and fractal urban form: A cellular modelling approach to the evolution of urban land use patterns. *Environ. Plan. A* **1993**, *25*, 1175–1199. [[CrossRef](#)]
15. Benenson, I.; Torrens, P.M. *Geosimulation: Automata-Based Modelling of Urban Phenomena*; John Wiley & Sons: London, UK, 2004.
16. Batty, M. *Cities and Complexity: Understanding Cities with Cellular Automata, Agent-Based Models, and Fractals*; The MIT Press: Cambridge, MA, USA, 2005.
17. Bura, S.; Guérin-Pace, F.; Mathian, H.; Pumain, D.; Sanders, L. Multi-agent systems and the dynamics of a settlement system. *Geogr. Anal.* **1996**, *28*, 161–178. [[CrossRef](#)]
18. Pumain, D. Multi-agent system modelling for urban systems: The series of simpop models. In *Agent-Based Models of Geographical Systems*; Heppenstall, A.J., Crooks, A.T., See, L.M., Batty, M., Eds.; Springer: New York, NY, USA, 2012; pp. 721–738.
19. O'Sullivan, D. Geographical information science: Agent-based models. *Prog. Hum. Geogr.* **2008**, *32*, 541–550. [[CrossRef](#)]
20. Torrens, P.M. Agent-based modeling and the spatial sciences. *Geogr. Compass* **2010**, *4*, 428–448. [[CrossRef](#)]
21. Torrens, P.M. Moving agent-pedestrians through space and time. *Ann. Assoc. Am. Geogr.* **2012**, *102*, 35–66. [[CrossRef](#)]
22. Manley, E.; Cheng, T.; Penn, A.; Emmonds, A. A framework for simulating large-scale complex urban traffic dynamics through hybrid agent-based modelling. *Comput. Environ. Urban Syst.* **2014**, *44*, 27–36. [[CrossRef](#)]
23. Benenson, I.; Omer, I.; Hatna, E. Entity-based modelling of urban residential dynamics: The case of yaffo, tel aviv. *Environ. Plan. B* **2002**, *29*, 491–512. [[CrossRef](#)]
24. Parker, D.C.; Manson, S.M.; Janssen, M.A.; Hoffmann, M.J.; Deadman, P. Multi-agent systems for the simulation of land-use and land-cover change: A review. *Ann. Assoc. Am. Geogr.* **2003**, *93*, 314–337. [[CrossRef](#)]
25. An, L.; Zvoleff, A.; Liu, J.; Axinn, W. Agent-based modeling in coupled human and natural systems (chans): Lessons from a comparative analysis. *Ann. Assoc. Am. Geogr.* **2014**, *104*, 723–745. [[CrossRef](#)]
26. Jacobs, J. *The Death and Life of Great American Cities*; Vintage Books: New York, NY, USA, 1961.
27. Crooks, A.T. The use of agent-based modelling for studying the social and physical environment of cities. In *Complexity and Planning: Systems, Assemblages and Simulations*; De Roo, G., Hiller, J., van Wezemael, J., Eds.; Ashgate: Burlington, VT, USA, 2012; pp. 385–408.
28. Batty, M. A generic framework for computational spatial modelling. In *Agent-Based Models of Geographical Systems*; Heppenstall, A., Crooks, A.T., See, L.M., Batty, M., Eds.; Springer: New York, NY, USA, 2012; pp. 19–50.
29. Simon, H.A. *The Sciences of the Artificial*, 3rd ed.; MIT Press: Cambridge, MA, USA, 1996.
30. Cioffi-Revilla, C. *Introduction to Computational Social Science: Principles and Applications*; Springer: New York, NY, USA, 2014.
31. Batty, M. *The New Science of Cities*; MIT Press: Cambridge, MA, USA, 2013.
32. Christaller, W. *Die Centralen*; Gustav Fischer: Jena, Germany, 1933.

33. Liu, X.; Andersson, C. Assessing the impact of temporal dynamics on land-use change modelling. *Comput. Environ. Urban Syst.* **2004**, *28*, 107–124. [[CrossRef](#)]
34. Batty, M. Cellular automata and urban form: A primer. *J. Am. Plan. Assoc.* **1997**, *63*, 266–274. [[CrossRef](#)]
35. Torrens, P.M. High-fidelity behaviors for model people on model streetscapes. *Ann. GIS* **2014**, *20*, 139–157. [[CrossRef](#)]
36. Heckbert, S.; Baynes, T.; Reeson, A. Agent-based modeling in ecological economics. *Ann. N. Y. Acad. Sci.* **2010**, *1185*, 39–53. [[CrossRef](#)] [[PubMed](#)]
37. Crooks, A.T.; Heppenstall, A. Introduction to agent-based modelling. In *Agent-Based Models of Geographical Systems*; Heppenstall, A., Crooks, A.T., See, L.M., Batty, M., Eds.; Springer: New York, NY, USA, 2012; pp. 85–108.
38. Kennedy, W. Modelling human behaviour in agent-based models. In *Agent-Based Models of Geographical Systems*; Heppenstall, A., Crooks, A.T., See, L.M., Batty, M., Eds.; Springer: New York, NY, USA, 2012; pp. 167–180.
39. Coleman, J.S. *Foundations of Social Theory*; Harvard University Press: Cambridge, MA, USA, 1990.
40. Axelrod, R. Advancing the art of simulation in the social sciences. In *Simulating Social Phenomena*; Conte, R., Hegselmann, R., Terno, P., Eds.; Springer: Berlin, Germany, 1997; pp. 21–40.
41. Bonabeau, E. Agent-based modelling: Methods and techniques for simulating human systems. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 7280–7287. [[CrossRef](#)] [[PubMed](#)]
42. Friedkin, N.E.; Johnsen, E.C. Social influence networks and opinion change. *Adv. Group Process.* **1999**, *16*, 1–29.
43. O’Sullivan, D.; Millington, J.; Perry, G.; Wainwright, J. Agent-based models—Because they are worth it? In *Agent-Based Models of Geographical Systems*; Heppenstall, A.J., Crooks, A.T., Batty, M., See, L.M., Eds.; Springer: New York, NY, USA, 2012.
44. Grimm, V.; Berger, U.; Bastiansen, F.; Eliassen, S.; Ginot, V.; Giske, J.; Goss-Custard, J.; Grand, T.; Heinz, S.; Huse, G.; et al. A standard protocol for describing individual-based and agent-based models. *Ecol. Model.* **2006**, *198*, 115–126. [[CrossRef](#)]
45. Grimm, V.; Revilla, E.; Berger, U.; Jeltsch, F.; Mooij, W.M.; Railsback, S.F.; Thulke, H.; Weiner, J.; Wiegand, T.; DeAngelis, D.L. Pattern-oriented modeling of agent-based complex systems: Lessons from ecology. *Science* **2005**, *310*, 987–991. [[CrossRef](#)] [[PubMed](#)]
46. Crooks, A.T.; Castle, C. The integration of agent-based modelling and geographical information for geospatial simulation. In *Agent-Based Models of Geographical Systems*; Heppenstall, A., Crooks, A.T., See, L.M., Batty, M., Eds.; Springer: New York, NY, USA, 2012; pp. 219–252.
47. Malik, A.; Crooks, A.; Root, H.; Swartz, M. Exploring creativity and urban development with agent-based modeling. *J. Artif. Soc. So. Simul.* **2015**, *18*, 12. [[CrossRef](#)]
48. Axtell, R.; Epstein, J.M. Agent-based modelling: Understanding our creations. *Bull. St. Fe Inst.* **1994**, *9*, 28–32.
49. Batty, M.; Desyllas, J.; Duxbury, E. Safety in numbers? Modelling crowds and designing control for the notting hill carnival. *Urban Stud.* **2003**, *40*, 1573–1590. [[CrossRef](#)]
50. Torrens, P.M.; McDaniel, A.W. Modeling geographic behavior in riotous crowds. *Ann. Assoc. Am. Geogr.* **2013**, *103*, 20–46. [[CrossRef](#)]
51. Crooks, A.T.; Croitoru, A.; Lu, X.; Wise, S.; Irvine, J.M.; Stefanidis, A. Walk this way: Improving pedestrian agent-based models through scene activity analysis. *ISPRS Int. J. Geo-Inf.* **2015**, *4*, 1627–1656. [[CrossRef](#)]
52. Crooks, A.T.; Hailegiorgis, A.B. An agent-based modeling approach applied to the spread of cholera. *Environ. Model. Softw.* **2014**, *62*, 164–177. [[CrossRef](#)]
53. Eubank, S.; Guclu, H.; Kumar, A.V.S.; Marathe, M.V.; Srinivasan, A.; Toroczkai, Z.; Wang, N. Modelling disease outbreaks in realistic urban social networks. *Nature* **2004**, *429*, 180–184. [[CrossRef](#)] [[PubMed](#)]
54. Malleson, N.; Heppenstall, A.; See, L.; Evans, A. Using an agent-based crime simulation to predict the effects of urban regeneration on individual household burglary risk. *Environ. Plan. B* **2013**, *40*, 405–426. [[CrossRef](#)]
55. Groff, E.R. Simulation for theory testing and experimentation: An example using routine activity theory and street robbery. *J. Quant. Criminol.* **2007**, *23*, 75–103. [[CrossRef](#)]
56. Dawson, R.J.; Peppe, R.; Wang, M. An agent-based model for risk-based flood incident management. *Nat. Hazards* **2011**, *59*, 167–189. [[CrossRef](#)]
57. Heppenstall, A.J.; Evans, A.J.; Birkin, M.H. Using hybrid agent-based systems to model spatially-influenced retail markets. *J. Artif. Soc. Soc. Simul.* **2006**, *9*, 2.

58. Augustijn-Beckers, E.; Flacke, J.; Retsios, B. Simulating informal settlement growth in dar es salaam, tanzania: An agent-based housing model. *Comput. Environ. Urban Syst.* **2011**, *35*, 93–103. [[CrossRef](#)]
59. Jordan, R.; Birkin, M.; Evans, A. An agent-based model of residential mobility assessing the impacts of urban regeneration policy in the easel district. *Comput. Environ. Urban Syst.* **2014**, *48*, 49–63. [[CrossRef](#)]
60. Haase, D.; Lautenbach, S.; Seppelt, R. Modeling and simulating residential mobility in a shrinking city using an agent-based approach. *Environ. Model. Softw.* **2010**, *25*, 1225–1240. [[CrossRef](#)]
61. Xie, Y.; Fan, S. Multi-city sustainable regional urban growth simulation—Msrugs: A case study along the mid-section of silk road of china. *Stoch. Environ. Res. Risk Assess.* **2014**, *28*, 829–841. [[CrossRef](#)]
62. Helbing, D.; Balmelli, S. *How to Do Agent-Based Simulations in the Future: From Modeling Social Mechanisms to Emergent Phenomena and Interactive Systems Design*; Santa Fe Institute: Santa Fe, NM, USA, 2011.
63. Helbing, D.; Molnár, P. Social force model for pedestrian dynamics. *Phys. Rev. E* **1995**, *51*, 4282–4286. [[CrossRef](#)]
64. Pumain, D.; Sanders, L. Theoretical principles in interurban simulation models: A comparison. *Environ. Plan. A* **2013**, *45*, 2243–2260. [[CrossRef](#)]
65. Gode, D.K.; Sunder, S. Allocative efficiency of markets with zero-intelligence traders: Market as a partial substitute for individual rationality. *J. Political Econ.* **1993**, *101*, 119–137. [[CrossRef](#)]
66. Crooks, A.T. Constructing and implementing an agent-based model of residential segregation through vector gis. *Int. J. GIS* **2010**, *24*, 661–675. [[CrossRef](#)]
67. Rao, A.S.; Georgeff, M.P. Modeling Rational Agents within a BDI-Architecture. In Proceedings of the Second International Conference on Principles of Knowledge Representation and Reasoning, San Mateo, CA, USA, April 1991.
68. Schmidt, B. The modelling of human behaviour: The pecc reference model. In Proceedings of the 14th European Simulation Symposium, Dresden, Germany, 23–26 October 2002.
69. Brantingham, P.; Glasser, U.; Kinney, B.; Singh, K.; Vajihollahi, M. A computational model for simulating spatial aspects of crime in urban environments. In Proceedings of the 2005 IEEE International Conference on Systems, Man and Cybernetics, Waikoloa, HI, USA, 10–12 October 2005; pp. 3667–3674.
70. Laird, J.E. *The Soar Cognitive Architecture*; The MIT Press: Cambridge, MA, USA, 2012.
71. Anderson, J.R.; Lebiere, C. *The Atomic Components of Thought*; Psychology Press: Mahwah, NJ, USA, 1998.
72. The end of theory: The Data Deluge Makes the Scientific Method Obsolete. Available online: <http://www.uvm.edu/~cmplsys/wordpress/wp-content/uploads/reading-group/pdfs/2008/anderson2008.pdf> (accessed on 23 June 2008).
73. Law, A.M. *Simulation Modelling and Analysis*, 5th ed.; McGraw-Hill: New York, NY, USA, 2015.
74. Balci, O. Verification, validation, and testing. In *Handbook of Simulation: Principles, Methodology, Advances, Applications, and Practice*; John Wiley & Sons: New York, NY, USA, 1996; pp. 335–393.
75. Lee, J.S.; Filatova, T.; Ligmann-Zielinska, A.; Hassani-Mahmooui, B.; Stonedahl, F.; Lorscheid, I.; Voinov, A.; Polhill, G.; Sun, Z.; Parker, D.C. The complexities of agent-based modeling output analysis. *J. Artif. Soc. Soc. Simul.* **2015**, *18*, 4. [[CrossRef](#)]
76. Takadama, K.; Kawai, T.; Koyama, Y. Micro- and macro-level validation in agent-based simulation: Reproduction of human-like behaviours and thinking in a sequential bargaining game. *J. Artif. Soc. Soc. Simul.* **2008**, *11*, 9.
77. Windrum, P.; Fagiolo, G.; Moneta, A. Empirical validation of agent-based models: Alternatives and prospects. *J. Artif. Soc. Soc. Simul.* **2007**, *10*, 8.
78. Smajgl, A.; Brown, D.G.; Valbuena, D.; Huigen, M.G.A. Empirical characterisation of agent behaviours in socio-ecological systems. *Environ. Model. Softw.* **2011**, *26*, 837–844. [[CrossRef](#)]
79. Moss, S. Alternative approaches to the empirical validation of agent-based models. *J. Artif. Soc. Soc. Simul.* **2008**, *11*, 5.
80. Kocabas, V.; Dragicevic, S. Agent-based model validation using bayesian networks and vector spatial data. *Environ. Plan. B* **2009**, *36*, 787–801. [[CrossRef](#)]
81. Weinberger, S. Web of war: Can computational social science help to prevent or win wars? *Nature* **2011**, *471*, 566–568. [[CrossRef](#)] [[PubMed](#)]
82. Crooks, A.T.; Pfoser, D.; Jenkins, A.; Croitoru, A.; Stefanidis, A.; Smith, D.A.; Karagiorgou, S.; Efentakis, A.; Lamprianidis, G. Crowdsourcing urban form and function. *Int. J. Geogr. Inf. Sci.* **2015**, *29*, 720–741. [[CrossRef](#)]
83. Laney, D. *3D Data Management: Controlling Data Volume, Velocity and Variety*; META Group Inc: Stamford, CT, USA, 2001.

84. Kitchin, R. Big data and human geography opportunities, challenges and risks. *Dialogues Hum. Geogr.* **2013**, *3*, 262–267. [[CrossRef](#)]
85. Google Trends. Searching for “Big data” In google trends. Available online: <https://www.google.com/trends/explore#q=big%20data> (accessed on 8 September 2015).
86. Mayer-Schönberger, V.; Cukier, K. *Big Data: A Revolution that will Transform How We Live, Work and Think*; John Murray: London, UK, 2013.
87. Crosbie, T. Using activity diaries: Some methodological lessons. *J. Res. Pract.* **2006**, *2*, 1.
88. Cranshaw, J.; Schwartz, R.; Hong, J.I.; Sadeh, N.M. The livelihoods project: Utilizing social media to understand the dynamics of a city. In Proceedings of the Sixth International AAAI Conference on Weblogs and Social Media, Dublin, Ireland, 4–7 June 2012.
89. Kling, F.; Pozdnoukhov, A. When a City Tells a Story: Urban Topic Analysis. In Proceedings of the 20th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, Redondo Beach, CA, USA, 7–9 November 2012; pp. 482–485.
90. Frias-Martinez, V.; Frias-Martinez, E. Spectral clustering for sensing urban land use using twitter activity. *Eng. Appl. Artif. Intell.* **2014**, *35*, 237–245. [[CrossRef](#)]
91. Malleson, N.; Andresen, M.A. The impact of using social media data in crime rate calculations: Shifting hot spots and changing spatial patterns. *Cartogr. Geogr. Inf. Sci.* **2015**, *42*, 112–121. [[CrossRef](#)]
92. Isaacman, S.; Becker, R.; Cáceres, R.; Kobourov, S.; Martonosi, M.; Rowland, J.; Varshavsky, A. Identifying important places in people’s lives from cellular network data. In *Pervasive Computing, Lecture Notes in Computer Science*; Lyons, K., Hightower, J., Huang, E.M., Eds.; Springer: Berlin, Germany, 2011; pp. 133–151.
93. Crooks, A.T.; Croitoru, A.; Stefanidis, A.; Radzikowski, J. #Earthquake: Twitter as a distributed sensor system. *Trans. GIS* **2013**, *17*, 124–147.
94. Croitoru, A.; Wayant, N.; Crooks, A.T.; Radzikowski, J.; Stefanidis, A. Linking cyber and physical spaces through community detection and clustering in social media feeds. *Comput. Environ. Urban Syst.* **2014**. [[CrossRef](#)]
95. Qu, Y.; Zhang, J. Regularly visited patches in human mobility. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Paris, France, 27 April–2 May 2009; pp. 395–398.
96. Birkin, M.; Harland, K.; Malleson, N.; Cross, P.; Clarke, M. An examination of personal mobility patterns in space and time using twitter. *Int. J. Agric. Environ. Inf. Syst.* **2014**, *5*, 55–72. [[CrossRef](#)]
97. Arribas-Bel, D.; Kourtit, K.; Nijkamp, P.; Steenbruggen, J. Cyber cities: Social media as a tool for understanding cities. *Appl. Spat. Anal. Policy* **2015**, *8*, 231–247. [[CrossRef](#)]
98. Lovelace, R.; Birkin, M.; Cross, P.; Clarke, M. From big noise to big data: Toward the verification of large data sets for understanding regional retail flows. *Geogr. Anal.* **2015**. [[CrossRef](#)]
99. Lovelace, R.; Birkin, M.; Malleson, N. Can Social Media Data be Useful in Spatial Modelling? A Case Study of “Museum Tweets” and Visitor Flows. In Proceedings of the 22nd Geographical Information Systems Research UK Conference, Glasgow, UK, 16–18 April 2014.
100. Wise, S. Agent-Based Modeling and Gis: Exploring Our New Tools in a Disaster Context. In Proceedings of the 22nd Geographical Information Systems Research UK Conference, Glasgow, UK, 16–18 April 2014.
101. Batty, M. Agents, cells, and cities: New representational models for simulating multiscale urban dynamics. *Environ. Plan. A* **2005**, *37*, 1373–1394. [[CrossRef](#)]
102. Epstein, J.M. Modelling to contain pandemics. *Nature* **2009**, *460*, 687. [[CrossRef](#)] [[PubMed](#)]
103. Savage, M.; Burrows, R. The coming crisis of empirical sociology. *Sociology* **2007**, *41*, 885–899. [[CrossRef](#)]
104. Yu, L. Understanding information inequality: Making sense of the literature of the information and digital divides. *J. Librariansh. Inf. Sci.* **2006**, *38*, 229–252. [[CrossRef](#)]
105. Fuchs, C. The role of income inequality in a multivariate cross-national analysis of the digital divide. *Soc. Sci. Comput. Rev.* **2008**, *27*, 41–58. [[CrossRef](#)]
106. Schradie, J. The digital production gap: The digital divide and web 2.0 collide. *Poetics* **2011**, *39*, 145–168. [[CrossRef](#)]
107. Chen, W.; Wellman, B. Minding the cyber-gap: The internet and social inequality. In *The Blackwell Companion to Social Inequalities*; Romero, M., Margolis, E., Eds.; Blackwell Publishing Ltd: London, UK, 2005; pp. 523–545.
108. Graham, M.; Shelton, T. Geography and the future of big data, big data and the future of geography. *Dialogues Hum. Geogr.* **2013**, *3*, 255–261. [[CrossRef](#)]

109. Glasgow, M.L.; Rudra, C.B.; Yoo, E.H.; Demirbas, M.; Merriman, J.; Nayak, P.; Crabtree-Ide, C.; Szpiro, A.A.; Rudra, A.; Wactawski-Wende, J.; *et al.* Using smartphones to collect time-activity data for long-term personal-level air pollution exposure assessment. *J. Expo. Sci. Environ. Epidemiol.* **2014**. [[CrossRef](#)] [[PubMed](#)]
110. Salzmann-Erikson, R.M.; Eriksson, R.H. Torrenting values, feelings, and thoughts—Cyber nursing and virtual self-care in a breast augmentation forum. *Int. J. Qual. Stud. Health Well-Being* **2011**, *6*, 7378. [[CrossRef](#)] [[PubMed](#)]
111. Ratti, C.; Pulselli, R.M.; Williams, S.; Frenchman, D. Mobile landscapes: Using location data from cell phones for urban analysis. *Environ. Plan. B* **2006**, *33*, 727–748. [[CrossRef](#)]
112. American Sociological Association. *Code of Ethics and Policies and Procedures of the Asa Committee on Professional Ethics*; American Sociological Association: Washington, DC, USA, 1999.
113. Eysenbach, G.; Till, J.E. Ethical issues in qualitative research on internet communities. *BMJ* **2001**, *323*, 1103–1105. [[CrossRef](#)] [[PubMed](#)]
114. Wilkinson, D.; Thelwall, M. Researching personal information on the public web: Methods and ethics. *Soc. Sci. Comput. Rev.* **2011**, *29*, 387–401. [[CrossRef](#)]
115. McKee, R. Ethical issues in using social media for health and health care research. *Health Policy* **2013**, *110*, 298–301. [[CrossRef](#)] [[PubMed](#)]
116. NSA Prism Program Taps in to User Data of Apple, Google and Others. Available online: <http://www.alleanzaperinternet.it/wp-content/uploads/2013/06/guardian.pdf> (accessed on 7 June 2013).
117. Duhigg, C. How companies learn your secrets. *N.Y. Times*, 2012, *16*, 1–16.
118. Aggarwal, C.; Yu, P. *Privacy-Preserving Data Mining: Models and Algorithms*; Springer: New York, NY, USA, 2008.



© 2016 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons by Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).