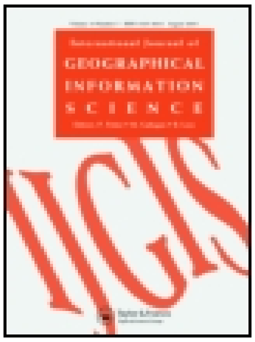

International Journal of Geographical Information Science

ISSN: 1365-8816 (Print) 1362-3087 (Online) Journal homepage: http://www.tandfonline.com/loi/tgis20

Constructing and implementing an agent-based model of residential segregation through vector GIS

A. T. Crooks

International Journal of Geographical Information Science
Vol. 24, No. 5, May 2010, 661–675

Constructing and implementing an agent-based model of residential segregation through vector GIS

A.T. Crooks*

Department of Computational Social Science, George Mason University, Fairfax, VA, USA

(*Received 19 February 2009; final version received 9 November 2009*)

In this article, we present a geographically explicit agent-based model (ABM), loosely coupled with vector geographical information systems (GISs), which explicitly captures and uses geometric data and socioeconomic attributes in the simulation process. The ability to represent the urban environment as a series of points, lines, and polygons not only allows one to represent a range of different-sized features such as buildings or larger areas portrayed as the urban environment but is a move away from many ABMs utilizing GIS that are rooted in grid-based structures. We apply this model to the study of residential segregation, specifically creating a Schelling (1971) type of model within a hypothetical cityscape, thus demonstrating how this approach can be used for linking vector-based GIS and agent-based modeling. A selection of simulation experiments are presented, highlighting the inner workings of the model and how aggregate patterns of segregation can emerge from the mild tastes and preferences of individual agents interacting locally over time. Furthermore, the article suggests how this model could be extended and demonstrates the importance of explicit geographical space in the modeling process.

Keywords: agent-based modeling; GIS; residential segregation

1. Introduction

Many of the applications linking geographical information systems (GISs) and agent-based models (ABMs) focus on representing space as a series of discrete cells (e.g. Gimblett *et al.* 2002), and while these ABMs have provided valuable insights into urban phenomena as they can capture geographic detail, they miss geometric detail. This area is critical to good applications but is barely touched upon in the literature (Batty 2005). There is a need to move away from the cellular representation of cities and begin to incorporate the details of the geometry and geography of the real city (see Xie and Batty 2003, Stanilov 2009 for more detailed discussions). The ability to represent the world as a series of points, lines, and polygons permits the inclusion of geometry in the modeling process, thereby allowing for different sizes of features such as houses and roads, for example, to be more realistically portrayed and letting one explore how these features might affect the simulation outcomes depending on the processes being modeled.

This article presents an ABM loosely coupled with a vector GIS, which explicitly captures and uses geometric data and related attributes in the simulation process. To highlight this, the model is applied to the study of segregation. In the remainder of this article,

*Email: acrooks2@gmu.edu

ISSN 1365-8816 print/ISSN 1362-3087 online

DOI: 10.1080/13658810903569572
http://www.informaworld.com

first, there is a brief description of segregation and a discussion of ABMs that have been created to study and explore this phenomenon. Second, the basic model is introduced focusing on the underlying mechanisms and the use of vector-based GIS data. Third, results from various experiments are presented, highlighting not only the inner workings of the model but also how the creation of such geographically explicit models helps with our understanding of such phenomena. This is then followed by discussion and conclusions concerning this particular modeling approach with respect to the study of urban phenomena.

2. Background and literature review

Segregation is all too clear in most urban areas, where there are clear clusters of economic and residential groups based on ethnicity, social class, and so on. While we are able to quantify the degree of segregation within neighborhoods (e.g. Reardon and O'Sullivan 2004), such analysis tells us little about the behavior that leads to or from particular outcomes. Without this knowledge, trying to prevent such a process or phenomenon becomes challenging. One might think that individuals must have strong preferences for these racially or economically homogeneous neighborhoods to emerge. However, empirical evidence suggests that individuals do not have strong racial preferences but have rather mild preferences (e.g. Clark 1991). To find clear examples of this segregation process taking place is difficult, because it only becomes noticeable when it is clearly underway, and by then a detailed chronology becomes impossible to reconstruct (Batty *et al*. 2004). To understand this behavior, we have to examine how the process of individual choice leads to these outcomes.

Schelling (1971) demonstrated that segregation could emerge through mild preferences to locate amongst like demographic or economic activity groups; although subsequent researchers have endorsed his conclusions (see Clark 1991), his work has also received criticism (e.g. Massey and Denton 1993). Nevertheless, this does not undermine Schelling's central insight: marked segregation can arise from rather mild individual preferences for living amongst one's own kind. Not only is the model one of the best known ABMs, but it has additionally continued to inspire theory and research into the segregation phenomena (e.g. Fossett 2006).

2.1. Agent-based models of segregation

Many ABMs have been inspired by Schelling's (1971) model or can be seen as extensions to his original insights; in this section, we will briefly explore some of these. Various neighborhood shapes (e.g. Flache and Hegselmann 2001) and sizes have been investigated to explore their impact on segregation outcomes. For example, with larger cell neighborhoods, models took longer to stabilize and more extreme patterns of segregation would arise (e.g. O'Sullivan *et al.* 2003, Fossett and Waren 2005). Others have included preferences for neighborhood status and housing quality, and differing levels of socioeconomic inequality within and between ethnic populations (see Fossett and Senft 2004), or incorporated income and cultural preferences for neighborhoods (e.g. Bruch 2006). Benenson *et al*. (2002), for example, used individual census records and GIS data representing streets and buildings to explore ethnic residential segregation in the Yaffo area of Tel Aviv, the model itself consisting of two interacting layers – one representing agents located on a physical environment layer representing streets and buildings. Each house is converted into a Voronoi polygon and the agents' residential behavior is affected by the ethnic composition of the neighborhood defined using these polygons.

The examples presented above can be viewed on a continuum between abstract demonstrations to real-world applications. Each one brings something new to the basic insights Schelling first presented. However, many are dependent on regular grids where each cell is often used to represent a single home, with one agent being allowed to occupy the cell at any one time (a common feature of ABMs using cell space; however, there are models that allow more than one agent per cell, such as Benenson *et al*. [2002], but these are rare). It is often argued that this is unrealistic, especially within cities; for example, within a block of flats there can be numerous people but their geographical footprint would be the same and would be missed by restricting one agent to one cell. Most of these models employ a featureless plain, paying little attention to physical barriers. Noonan (2005) showed empirically how physical barriers (such as parks, railroads, and major roads) have impacts on neighborhoods. Below, we present how individual entities can be created and located within space where movement is not restricted by cells as our model will not contain any cells. Furthermore, more than one agent can be located in the same area, and clusters of residential groups can bridge different areas.

3. A vector-based geographically explicit segregation model

There are many types of segregation, a product of many factors, but the model presented in this article only explores one such hypothesis, that of Schelling's (1971) – where many agents with mild tastes and preferences locate amongst like social groups, segregation will emerge. The purpose of this article is to simply extend this well-known model so we can explore the impact of space and geometry; it is a pedagogic demonstration which illustrates a way of thinking about modeling the built environment in the particular context of segregation. This section will outline the basic model, while further details of the model, including the source code, a complete description, data files used in the simulations, animations of simulation runs, and additional models can be found at http://www.casa.ucl.ac.uk/abm/segregation.

The model itself is loosely coupled to GIS,[1] especially the vector data structure (as model inputs and outputs are ESRI shapefiles[2]), which is written in Java and extends a number of basic operating classes from the RepastJ library, an open source Java-based agent-based toolkit (Repast 2008). Repast is primarily used for its ability to display model information, scheduling of model actions, and the importation of GIS vector data (in the form of ESRI shapefiles), along with its recording change classes. The model additionally utilizes other Java-based libraries, namely the JTS Topology Suite (JTS 2008), which provides general 2D-GIS functions such as line intersection and buffering algorithms, and OpenMap (2008), which provides a simple GIS display with pan and zoom, and query functions with respect to GIS layers.

Within the model, we consider agents as virtual households with the ability to search the virtual world and make residential choices. These agents possess an ethnic status which we denote, for example, as red and blue with these labels of course being arbitrary. These households have preferences for co-ethnic contact specified in terms of the percentage of co-ethnic households found in their 'neighborhood' in which the household lives or to where it is considering moving. Preferences can be the same or different between different ethnic groups. Unlike many models exploring segregation, these households are not restricted to discrete cells and can move to areas that are already occupied by other agents.

Translating GIS methods into agents and their environment, the model is composed of two vector layers – the urban environment that is represented as a series of polygons created directly from the ESRI shapefiles, and agents that are represented as points. It is the information held within fields of the environment layer that is used to create the point agents, but also the environment layer defines the boundary of the world which the agents occupy using the spatial analysis operation known as union. The distribution of the types of point agents (representing ethnic groups, say) as observed through aggregate census population counts forms the initial starting conditions for the model. Such counts or other attributes in the environment layer could also be used to represent capacity constraints, such as maximum number of agents in an area if desired. For example, Figure 1a represents four areas dimensioned to wards in the City of London, each with their own hypothetical attribute information stored in a data table where each row relates to a specific ward (e.g. Ward 1 has a population of ten red, five blue, four green, and two white agents). The model reads this data and creates an environment polygon for each ward and for the desired agent population based on the data held in the fields (Figure 1b). Note that the underlying color of the polygon (ward) always represents the predominant social group in the area (accomplished by counting the number of agents of different types within each polygon). The agents are initially randomly placed within each polygon. This provides a city landscape that is integrated at initialization. However, these agents could be placed in precise locations if they were known (see Crooks 2007 for further details). The basic model is designed to work on many different geographical scales and areas (e.g. boroughs and output areas) without the need for its reconfiguration of the model code (i.e. the geometry of the environment is not hard-coded into the model but relies on the shapefile used). This was considered important as most socioeconomic data, for example, census and geo-demographic data, come in this format. This functionality was created so that the model could be easily replicated in other areas in the quest to allow the modeler to see if the same rules can be applied to different areas and at different scales.

The ability to represent the urban system as a series of spatial objects – points, lines, and polygons each with a spatial reference describing the location of the object rather than just as a series of cells – leads to conceptual problems in defining neighborhoods. Furthermore, it makes definitions of the ways in which agents move and search their environment difficult. To overcome these problems, the model relies on a series of spatial analysis operations, specifically buffering, union, line intersection, and point(s)-in-polygon analysis utilizing the JTS library. It is to these that we now turn.

Unlike the case of cellular space models where neighborhoods are often calculated using Moore or von Neumann neighborhoods or variations of these (Batty 2005), representing

Ward	No. red	No. blue	No. green	No. white	Other info
1	10	5	4	2	
2	5	10	3	4	
3	2	3	5	3	
4	2	3	4	2	

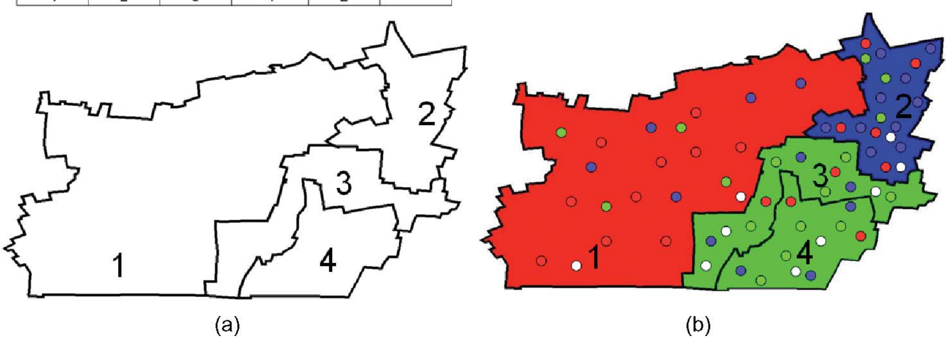

Figure 1. Populating the model with agents: reading in the data and creating the environment and the agents.

agents as points means different tools are needed to calculate neighborhoods, specifically when incorporating physical boundaries (e.g. rivers and motorways) into the modeling process. For the point agents, buffers are created to calculate neighborhoods. This involves the creation of a circular region around the point. The radius of the circle is defined by the analyst using euclidean distance. Figure 2a highlights how a geographical feature (such as a river) is incorporated into the model when calculating neighborhoods for point agents. Within Figure 2a, the black circle represents the agent of interest. This agent wants to know which agents are within a specified distance and in the same geographical area. A buffer polygon is created at this specified distance based on the centroid of the agent. However, in this case, the buffer crosses the river, which is stored in the urban environment layer, resulting in a multipart polygon (one part on each side of the river) being formed. In such a case, the agent calculates which part of the multipart polygon it is within through a point-in-polygon operation and only counts agents in the same multipart polygon as itself. Therefore, the black agents on the other side of the river are not classed as neighbors, although they are within the buffered region (light gray line). The gray agents, who are on the same side of the river as the agent of interest, are within the agents' defined multipart polygon buffer region and are classed as neighbors. However, if the two regions were connected, the agents on both sides of the river would be considered neighbors as the buffer would only be a single-part polygon as demonstrated in Figure 2b.

As the agents are represented as a separate layer to the environment they reside in, one needs to relate objects in one layer to objects in another. This is achieved through a point-in-polygon analysis. This allows the model to determine whether or not a given agent lies inside a specific polygon, which allows for more aggregate statistics to be computed such as population counts. These counts are derived bottom-up, through the interactions of agents observed and modeled at micro scale.

The above GIS operations allow an ABM to be created where spatial and geometric relationships are explicitly incorporated into the simulation. Each type of agent knows its position and can use the operations such as buffering and point in polygon to find out more about its neighborhood and the urban environment it resides within (Figure 3). Once the

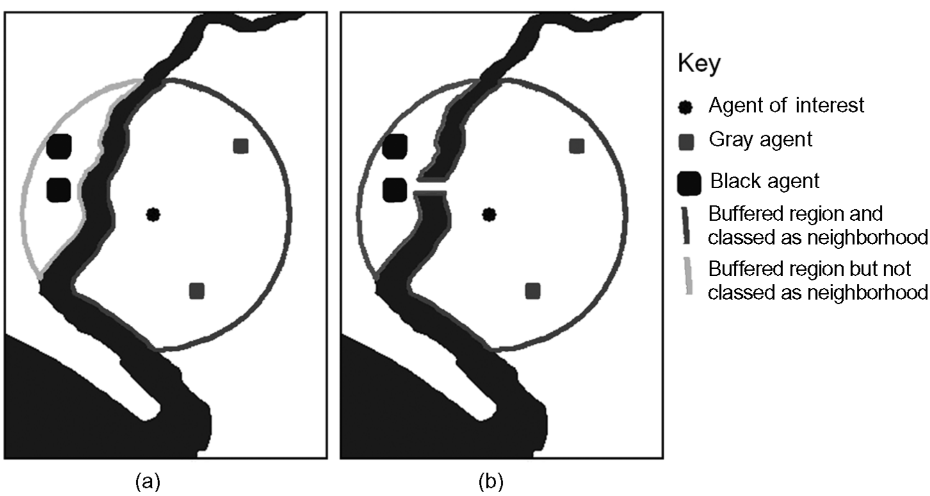

Figure 2. (a) Defining neighborhoods with the inclusion of geographical features (constrained buffer); (b) defining neighborhoods where the two areas are connected.

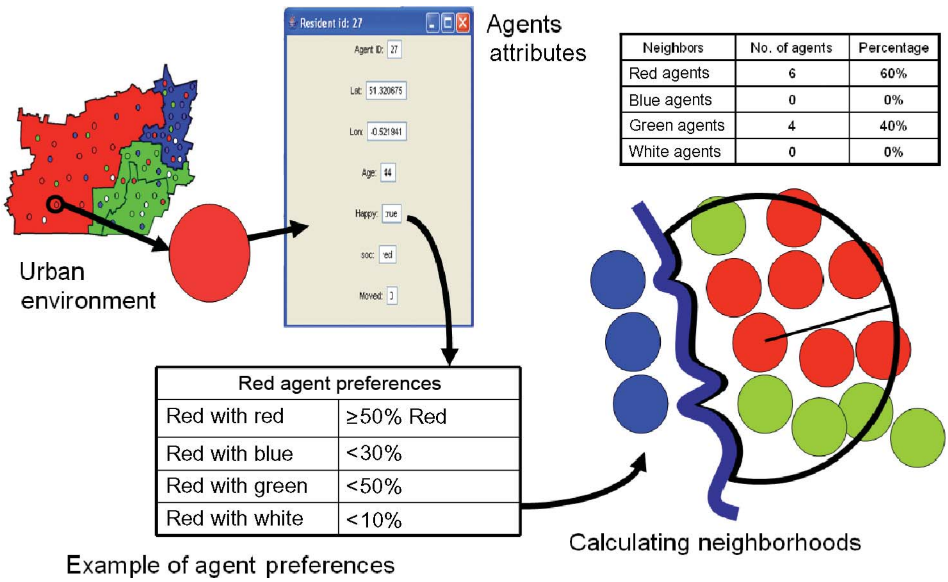

Figure 3. Basic model structure.

environment and the agents have been created, each agent uses its neighborhood function to query the surrounding neighbors, calculating if it is currently satisfied with its current neighborhood environment, taking into account physical features of the urban environment. Figure 3 highlights this process, where an agent is selected at random and it 'evaluates' the ethnic mix in its immediate neighborhood. If the agent is satisfied based on its preferences, it does nothing. On the contrary, if it is dissatisfied with its current neighborhood, the agent moves to the nearest location where its preferences are met.

As with other segregation models, the time frame within the model is purely hypothetical but could be considered as yearly intervals. At each time step (iteration), all the agents are given the option to move if they are dissatisfied with their current neighborhood configuration. The movement process involves two stages. In the first instance, the agent randomly searches its immediate area for a given number of moves, each time moving to a new location up to 100 m from its previous location, calculating the neighborhood composition using the buffering mechanism. If the agent is still dissatisfied after a given number of random movements, it moves to a location within 100 m of its nearest neighbor of the same type based on the euclidean distance from its initial location (this was done to speed up the searching process and assumes agents want to locate nearby other agents of the same type). Once the agent has moved to a location near to its nearest agent of the same type, it locally searches this new area for a neighborhood composition that satisfies its needs. If the agent cannot be satisfied, it moves to the location of its next nearest agent and so on, until all the area has been searched. If the agent is still not satisfied, it is removed from the system.

Once all the agents have had the option to move, the model advances one iteration and again all agents who are dissatisfied with their neighborhood have the option to move. This process continues until all the agents are satisfied with their current neighborhood configuration or the model is stopped by the user. While such a search criterion is abstract and simple and does not reflect the complex decision-making process of residents (such as the associated cost of moving), the model does capture the dynamics within neighborhoods, such as how people moving in and out of areas affect the overall composition of that area. Furthermore, it demonstrates how with bounded rationality, agents (i.e. agents do not know what all the potential neighborhood configurations are) locate in areas where their preferences are met, minimizing the distance traveled.

4. Results from selected experiments

While numerous experiments, variations, or extensions of the model are possible, the following subsections highlight how certain assumptions affect the outcome of the model and how changes can easily be made to the model. Not only do these demonstrate the structure of the model but they emphasize the rationale as to why such features were included in the model (for other experiments such as density constraints, the reader is referred to the website). Each simulation was run multiple times and the results and subsequent discussion present average outcomes.

4.1. The role of preferences

As with Schelling's (1971) model, agents only have preferences for their own group and it is this preference that causes agents to seek out different areas in the city. However, Clark (1991) demonstrated that preferences for specific compositions of neighborhoods were varied among cities. This section will, therefore, explore how the degree of segregation changes because of different preferences and explores how the preferences of individuals for their own group influences the degree of segregation seen within an area.

The only model parameter that changes within the simulation is the agents' preferences for the percentage of their same type to be located within their neighborhood. The world the agents occupy is a 1.5 km^2 polygon, which could be considered as representing a small cityscape. One could imagine this as the checkerboard that Schelling originally used. However, neither agents' neighborhoods nor their movement was restricted to a cell-based environment and multiple agents can occupy one area. As with Schelling's original model, we have equal numbers of two types of agents, 2000 of each color, placed randomly within the area.

Figure 4 highlights the typical final patterns of segregation that emerge from different preference parameters for neighborhood composition where the radius was set at 100 m. As the percentage of neighbors of their same type increases, the pattern of segregation becomes more noticeable. It is only when preferences rise above 80% that agents are forced to leave the system as a result of their preferences being unable to be matched. Increasing the percentage of neighbors of the same type within the agents' neighborhood, more agents are forced to move at least once during the course of a simulation. For example, when preferences are low (e.g. $\leq$20%), little movement occurs. However, as the preference for a minimum neighborhood increases, so does the total number of agents that move (e.g. $\geq$40%), along with the degree to which neighborhoods are segregated, as highlighted in Figure 4.

4.2. The effect of different neighborhood sizes

Neighborhoods mean different things to different people. Some may perceive a neighborhood as houses that are directly associated with their home (e.g. Benenson *et al*. 2002), whereas others may consider it as a street, or a collection of streets. As already stated, neighborhoods within the model are calculated using a buffer at a

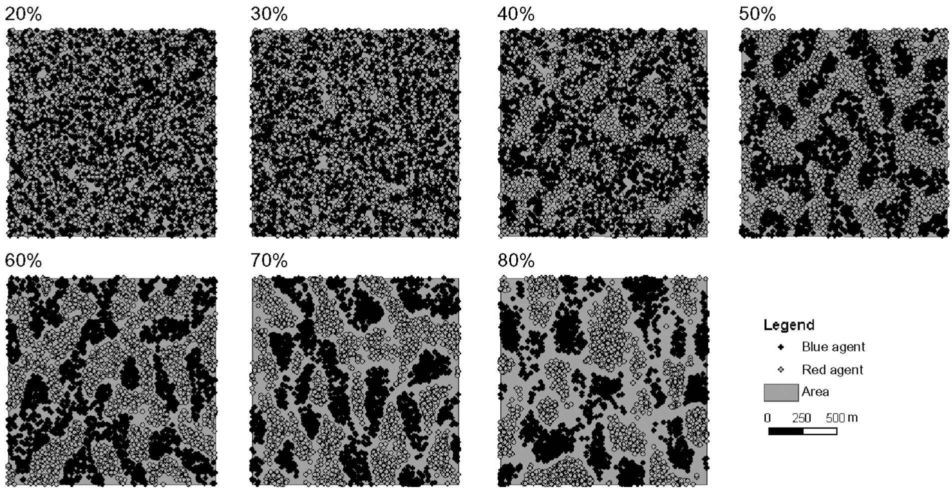

Figure 4. Typical patterns of segregation with different percentage preferences for neighborhood composition.

specified radius around the agent. To test the influence of neighborhood size on the resulting pattern of segregation that emerges, various neighborhood sizes were tested, ranging from a radius of 50 to 1000 m. All other parameters within the simulations were kept the same. The smallest neighborhood would only encompass the agents' immediate neighbors, for if an agent were to be dissatisfied with the area, this would reflect the composition of the agent's immediate neighbors. For larger neighborhoods, agents would consider larger areas, with the agent not considering its immediate neighbors *per se* but its overall neighborhood composition (which would therefore include on average a greater number of agents).

Typical patterns of segregation resulting from different neighborhood sizes are presented in Figure 5, which clearly shows the influence of neighborhood size on the outcome of the pattern of segregation. Smaller neighborhoods result in small segregated areas and larger neighborhoods result in larger segregated areas in proportion to the size of the buffer used. As with the O'Sullivan *et al.* (2003) model, when neighborhoods became larger, the simulation took longer to stabilize and more extreme patterns of segregation emerged. Another feature of the model is that at the boundary of these neighborhoods, agents appear to be more clustered than in the center of the neighborhood, which is an artifact of the search criteria (see Section 3): when agents are trying to find a location where their preferences are met they in the first instance search their immediate area.

4.3. The impact of geographical features

Areas within cities are bounded by features such as highways, railway lines, and rivers that can act as boundaries between residential groups. The model presented here was designed to explore the effect these features have on the outcome of a particular simulation. Thus the segregation model is not only capable of exploring segregation but also examining the effect geographical features have on the pattern of segregation that emerges. Within the model presented in this article, we consider all geographical features as obstacles, but depending on

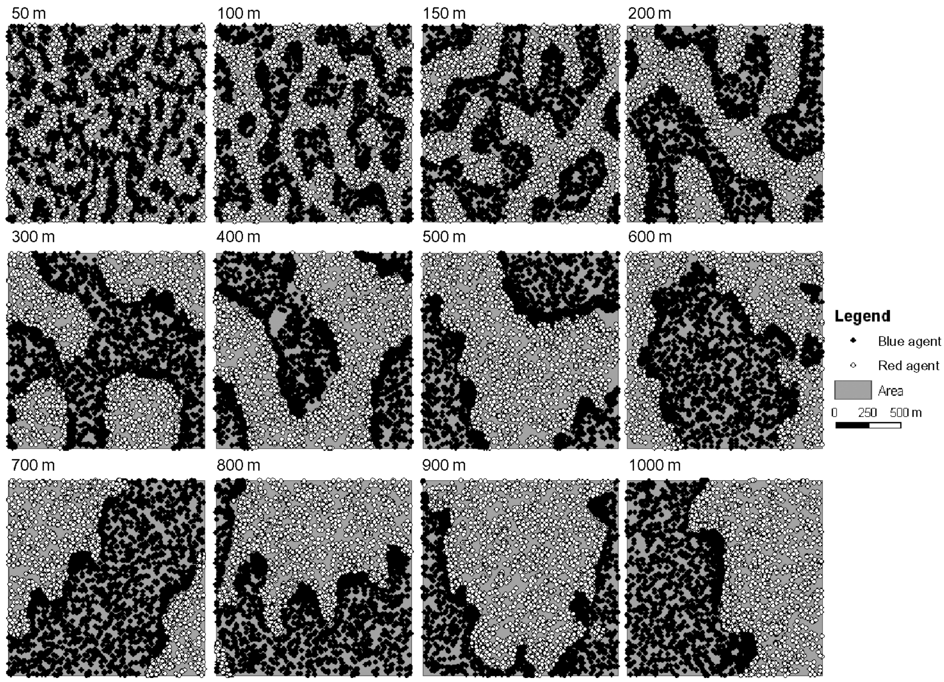

Figure 5. Example patterns of segregation when all agents are satisfied for different neighborhood sizes.

the context geographical features may not act as barriers; to include such contexts within the model, one would need to code which features act as barriers or not and include this in the neighborhood calculation. In this section, we demonstrate how neighborhoods can be influenced by the geometric features of the urban environment and how this impacts on the pattern of segregation. To achieve this, the segregation model will be compared to a variation that does not include geographical features (geometry) when neighborhoods are being calculated.

An arbitrary area was used to represent the urban space as shown in Figure 6. The dark gray area represents locations that agents can inhabit, whereas the light gray area represents areas where the agents cannot be located and could be considered void areas, such as water features. The void spaces act as barriers in the neighborhood calculations of the basic segregation model, where buffers are created and constrained by geographical features. As illustrated in Figure 6, agents located directly opposite to each other but separated by void space would not be considered as neighbors when calculating neighborhoods (in contrast to the variation of the model that does not include geometry when calculating neighborhoods). Agents are randomly placed at the start of the simulation and all want to be in a neighborhood where 50% or more of its neighbors are of the same type. Neighborhood size was set to 200 m to allow for agents on the fringes of geographical features to consider as neighbors agents on the other side of these fringes where geographical features are not considered in neighborhood calculations.

Figure 7 shows two final patterns of segregation, one from the basic segregation model and one from the model where geometry was not included (the extent is the same as shown in Figure 6). The patterns are similar, suggesting that the influence of geometry in

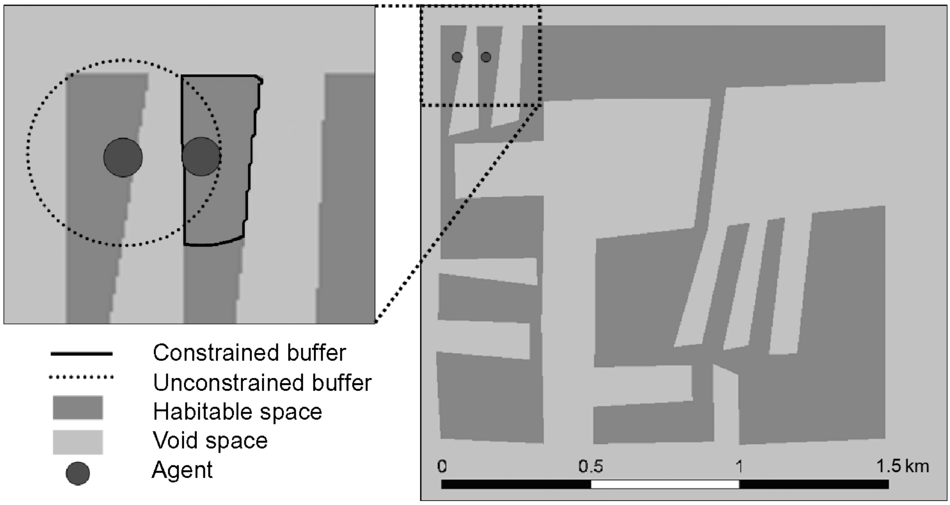

Figure 6. How geographical features impact on the pattern of segregation: an example of constrained and unconstrained buffers used in neighborhood calculations.

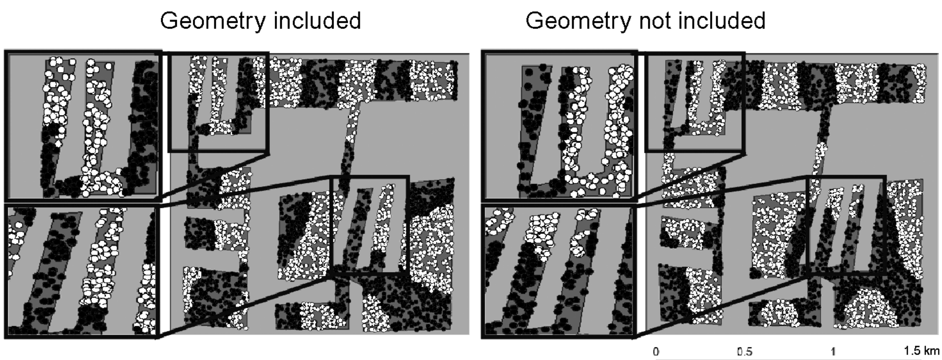

Figure 7. Final patterns of segregation when all agents are satisfied for different geometric and non-geometric segregation models.

neighborhood calculations is quite subtle. The influence of geometry is most obvious at the fringes of geometric boundaries, specifically the top left and center of the images as shown in the two zoomed-in areas of Figure 7. In the segregation model where geometry is not included in neighborhood calculations, one can clearly see that agents are considering as neighbors agents across the void on opposing spits, and thus agents of different types are located in proximity to each other although each others' preferences are met. Where geometry is considered, it can be seen that agents of different types are not located in the same areas as they would not consider agents on the opposite side of the void as neighbors and thus their preferences could potentially not be met. This forces them to locate in more homogeneous neighborhoods. From these simulations, one can see the effect that geometry has on neighborhood formation, which is not normally explicitly included in Schelling-type models in particular or ABMs in general.

4.4. Small minority populations

The simulations presented above have presumed 50/50 populations, which is rarely the case within cities. Often within populations, there are small minority groups that cluster in specific areas of the city. The ability to model more than two groups thus allows one to explore differences between the numbers of dominant and subdominant groups within a population and we will extend the model this way here.

Here, we explore a population of 4200, composed of 37% red, 34% blue, 24% green, and 5% white agents in seven areas, where the white agents are only located in one area.[3] This experiment was designed to explore how this minority population group can change and cluster over the course of a simulation run as agents search for neighborhoods where 50% or more of their neighbors are of the same types as themselves. Figure 8 highlights a typical simulation run with the majority of the white population staying and clustering in the area they originated in. The number and type of predominant social groups in the area are on average the same as when the model was first initialized.

4.5. Addition and removal of agents

All cities and regions change by both growth and decline. However, as there is no mechanism for population turnover within Schelling's basic model, households are 'immortal' and thus a satisfied household can reside in the same location for ever (Fossett and Waren 2005). The previous simulations were designed to study how established groups participate in constructing the city's social-spatial pattern. This section highlights how an additional dynamic process can be added to the more traditional

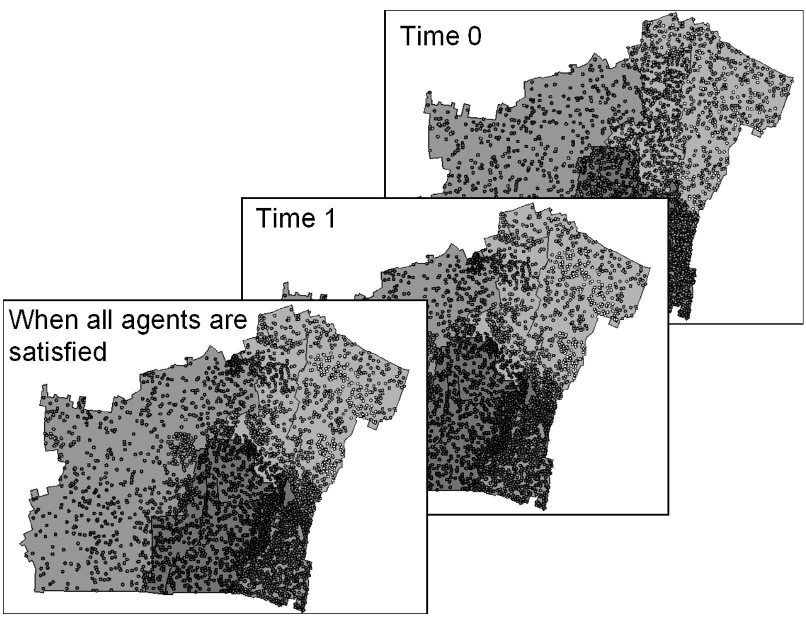

Figure 8. A typical simulation portraying the initial, first, and last iteration until all agents are satisfied when only 5% of the population is white.

Schelling-type model, through the addition and removal of agents, and how an area's social-spatial pattern might change through such a process. This could be considered as immigration and aging and the death of population in urban areas.

The model itself varies from others presented thus far in a number of ways. First, all the agents are given a new attribute: 'age'. At the end of each step, the agent's age increases by one. In the previous versions of the model, agents were only removed from the system if they could not find a suitable neighborhood. In this version, agents are also removed from the system when their age reaches 50. The second variation to the model is the addition of 100 new agents at the end of each step. These new agents are given a random social class and an age of 0 and placed randomly in the urban area when first created. The new agent then evaluates its neighborhood, and if dissatisfied it moves to an area where its preferences are met. If the new agent cannot become satisfied with the area, it is removed from the system.

The initial population was 700 agents (390 of type red [56%] and 310 of type blue [44%]) with ages randomly assigned values between 0 and 50. These agents were spread over several polygons. Each agent wanted to be in an area where 50% or more of its neighbors are of the same type. By the 100th iteration, the percentage of both red and blue agents becomes approximately equal and remains roughly constant. However, the degree to which one group dominates one area varies over the course of each simulation as agents are added and removed (Figure 9). As agents are added and removed from the system, the predominant social group of each area changes as can be seen from the underlying polygon color in Figure 9. Additionally, clusters of groups do not stop at boundaries of areas (Figure 9). So while the aggregate data suggest that the area is predominantly of one type, clusters of distinct groups appear throughout the area and cross the boundaries between areas. These clusters would have been lost by purely using aggregate information.

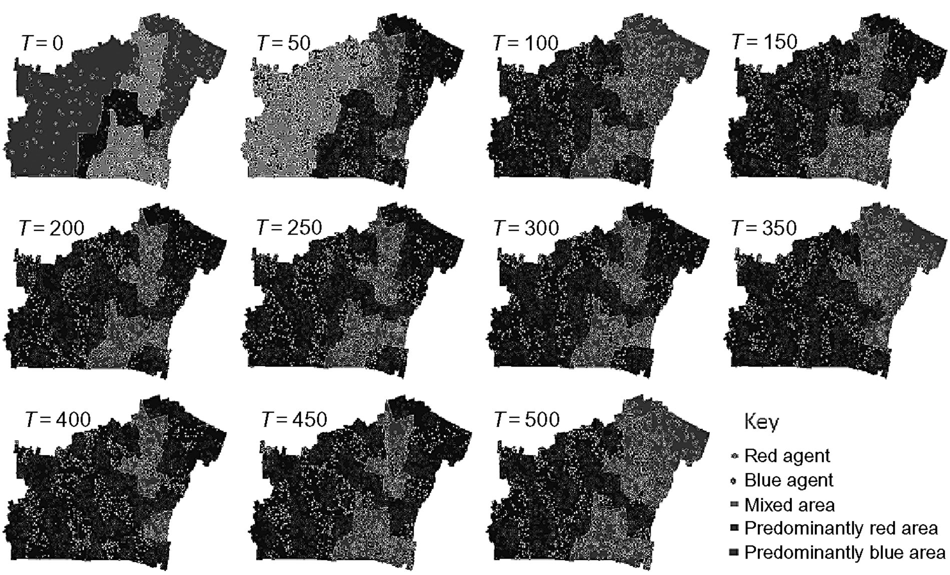

Figure 9. The effect of adding and removing agents to the system and the resulting patterns of segregation.

5. Conclusions

The model presented here demonstrates how the representation of individuals through simple rules governing their behavior and interaction over space and time at the micro scale can result in recognizable patterns emerging at the macro scale. The model departs from other typical models of segregation based on Schelling's ideas in a number of ways. To reiterate, most ABMs exploring segregation use a regular partition of space, or polygons, to represent the location of households. However, to date little research has been carried out on how the geometry of the environment affects the model outcomes. The model presented here is tightly bound to the vector GIS data model, thus resolving the lack of geometry, and the inability to represent objects of various shapes and sizes. We have demonstrated that the environment and agents can be derived from GIS features by using the coordinate representation of each feature, which allows them to move freely within the urban environment. Agents' movement is not restricted to discrete empty cells or areas. In addition, most relationships between agents can be evaluated within vector GISs using standard overlay operators such as point in polygon, buffering, intersection, and so on, making it possible to determine where agents are situated in relation to other agents and their environment. More specifically, neighborhood rules are available for evaluating adjacency, distance, and so on.

The experiments highlight not only how individual actions can lead to more aggregate patterns' emerging but also how agents can be linked to geographic locations and how geometry can be incorporated directly into the simulation process. Furthermore, the experiments can be considered as sensitivity tests of the model and highlight the effects of the underlying model assumptions on the simulation outcomes. This exploration provides a detailed understanding of the implications of each assumption but also allows one to evaluate the logic behind the model. This includes the influence of the size of neighborhoods, the influence of geographical features, and the degree to which segregation changes when agent preferences for neighborhood composition change. These explorations show that the geometry of an area can act as a physical barrier to segregation and that by increasing agents' preferences to reside in a specific group, marked segregation can emerge but not in a linear progression. The model illustrates how small minority groups cluster in areas and how these clusters remain persistent over time when agents are added and removed, outcomes which are well beyond what Schelling showed in his initial model. Furthermore, the model raises the importance of incorporating space and geometry when modeling urban systems. Additionally, the approach the model takes allows us to relate closely to 'real' urban form, while many other ABMs use stylized forms to represent the urban environment.

The analyses of the results from the simulations in Sections 4.4 and 4.5 demonstrate an important issue relating to the scale of analysis of model results, especially for the segregation phenomena. In particular, as we aggregate, we can unwittingly lose patterns that the agents enable us to explore. Aggregation can thus confuse our identification of coherent patterns that make sense in terms of basic human decision making. For example, in the simulations, distinct micro-clusters of different types of agents can be seen as a result of agents' decisions to locate in areas based on their mild tastes and preferences to be among similar agents. Moreover, it is also clear that clusters do not stop at boundaries but cross them as well. These clusters would be lost if we were only to consider aggregate-level data (e.g. total number of agents of different types in an area) and we might assume that the area is perfectly mixed, which is not the case. The model presented here was purposely kept simple, mainly to explore how space and geometry impact on Schelling's segregation model and to highlight how this approach can be used to study urban phenomena. It is envisaged that the models can easily be extended by others if so desired from the provision of source code. For example, one could include extra variables in the agent's choice of location, such as economic preferences about an area or income (e.g. Crooks 2007). With regard to making the models more operational, one could use fine-scale data sets to represent the built environment (e.g. the United Kingdoms (UK) MasterMap products), and population data could be synthetically micro-simulated to populate such a model. Ethnicity data could be obtained for individual addresses from the surnames and forenames of the UK electoral role (see Mateos *et al*. 2007) and used to link the micro-simulated data to actual buildings (as both would contain ethnicity information). From such data, detailed empirical analysis could be carried out for neighborhood preference values, sizes, and so on. Such data could be used to calibrate the model, in the sense that calibration of transition rules such as neighborhood preferences could be based on values that closely correspond to the actual values seen within the system.

In summary, we have outlined the model implementation, making explicit the components of the model and the key mechanisms that drive the model findings. Clear description of how the model is implemented along with the source code helps with verification of the model, thus furthering our ability to model urban systems. The model provides the essential ingredients for cumulative scientific inquiry with a clearly specified model that facilitates replication and extension, which is the key mission of traditional science. Furthermore, the model highlights how theories and concepts pertaining to urban phenomena can easily be abstracted within geographically explicit ABMs, helping further our understanding of how processes within cities operate and thus raising the importance of incorporating space and geometry when modeling urban systems.

Notes

1. For a discussion on advantages and disadvantages of coupling approaches the reader is referred to Castle and Crooks (2006).
2. A proprietary but widely available vector file format from ESRI.
3. The background color of the polygons represents the predominant social group of the area. For example, if the polygon is shaded red, there are more red agents in this area than any other type of agent, while a polygon shaded gray has equal numbers of at least two types of agents, for example, red and blue.

References

Batty, M., 2005. *Cities and complexity: understanding cities with cellular automata, agent-based models, and fractals*. Cambridge, MA: The MIT Press.

Batty, M., Barros, J., and Alves, S., Jr., 2004. *Cities: continuity, transformation and emergence*. London: Centre for Advanced Spatial Analysis, University College London, Working Paper 72.

Benenson, I., Omer, I., and Hatna, E., 2002. Entity-based modelling of urban residential dynamics: the case of Yaffo, Tel Aviv. *Environment and Planning B*, 29 (4), 491–512.

Bruch, E., 2006. Residential mobility, income inequality, and race/ethnic segregation in Los Angeles. *In*: *Population association of America 2006 annual meeting program*, Los Angeles, CA.

Castle, C.J.E. and Crooks, A.T., 2006. *Principles and concepts of agent-based modelling for developing geospatial simulations*. London: Centre for Advanced Spatial Analysis, University College London, Working Paper 110.

Clark, W.A.V., 1991. Residential preferences and neighbourhood racial segregation: a test of the Schelling segregation model. *Demography*, 28 (1), 1–19.

Crooks, A.T., 2007. *Experimenting with cities: utilizing agent-based models and GIS to explore urban dynamics*. Thesis (PhD). University College London.

Flache, A. and Hegselmann, R., 2001. Do irregular grids make a difference? Relaxing the spatial regularity assumption in cellular models of social dynamics. *Journal of Artificial Societies and*

Social Simulation, 4 (4). Available from: http://jasss.soc.surrey.ac.uk/4/4/6.html [Accessed 28 December 2009].

Fossett, M., 2006. Ethnic preferences, social distance dynamics, and residential segregation: theoretical explorations using simulation analysis. *Journal of Mathematical Sociology*, 30 (3–4), 185–274.

Fossett, M. and Senft, R., 2004. SIMSEG and generative models: a typology of model-generated segregation patterns. *In*: C.M. Macal, D. Sallach, and M.J. North, eds. *Proceedings of the agent 2004 conference on social dynamics: interaction, reflexivity and emergence*, 7–9 October 2004, Chicago. Argonne National Laboratory, 39–78.

Fossett, M. and Waren, W., 2005. Overlooked implications of ethnic preferences for residential segregation in agent-based models. *Urban Studies*, 42 (11), 1893–1917.

Gimblett, H.R., Richards, M.T., and Itami, R.M., 2002. Simulating wildland recreation use and conflicting spatial interactions using rule-driven intelligent agents. *In*: H.R. Gimblett, ed. *Integrating geographic information systems and agent-based modelling techniques for simulating social and ecological processes*. Oxford: Oxford University Press, 211–243.

JTS, 2008. *JTS Topology Suite* [online]. Available from: http://www.vividsolutions.com/jts/ [Accessed 19 December 2008].

Massey, D.S. and Denton, N.A., 1993. *American apartheid segregation and the making of the underclass*. Cambridge, MA: Harvard University Press.

Mateos, P., Webber, R., and Longley, P.A., 2007. *The cultural, ethnic and linguistic classification of populations and neighbourhoods using personal names*. London: Centre for Advanced Spatial Analysis, University College London, Working Paper 116.

Noonan, D.S., 2005. Neighbours, barriers and urban environments: are things 'different on the other side of the tracks'? *Urban Studies*, 42 (10), 1817–1835.

OpenMap, 2008. OpenMap^{TM} – *Open systems mapping technology* [online]. Available from: http://openmap.bbn.com/ [Accessed 19 December 2008].

O'Sullivan, D., MacGill, J., and Yu, C., 2003. Agent-based residential segregation: a hierarchically structured spatial model. *In*: C.M. Macal, M.J. North, and D. Sallach, eds. *Proceedings of agent 2003 conference on challenges in social simulation*, 7–9 October 2003, Chicago. Argonne National Laboratory, 493–507.

Reardon, S.F. and O'Sullivan, D., 2004. Measures of spatial segregation. *Sociological Methodology*, 34 (1), 121–162.

Repast, 2008. *Repast – Recursive Porous Agent Simulation Toolkit: overview* [online]. Available from: http://repast.sourceforge.net/repast_3/index.html [Accessed 19 December 2008].

Schelling, T.C., 1971. Dynamic models of segregation. *Journal of Mathematical Sociolog*y, 1 (1), 143–186.

Stanilov, K., 2009. Typo-morphology and object-based automata: methodological advances towards a more accurate modelling of urban growth patterns. *In*: *Proceedings of the 11th international conference on computers in urban planning and urban management*, Hong Kong, 16–18 June 2009. Hong Kong: The University of Hong Kong.

Xie, Y. and Batty, M., 2003. *Integrated urban evolutionary modelling*. London: Centre for Advanced Spatial Analysis, University College London, Working Paper 68.