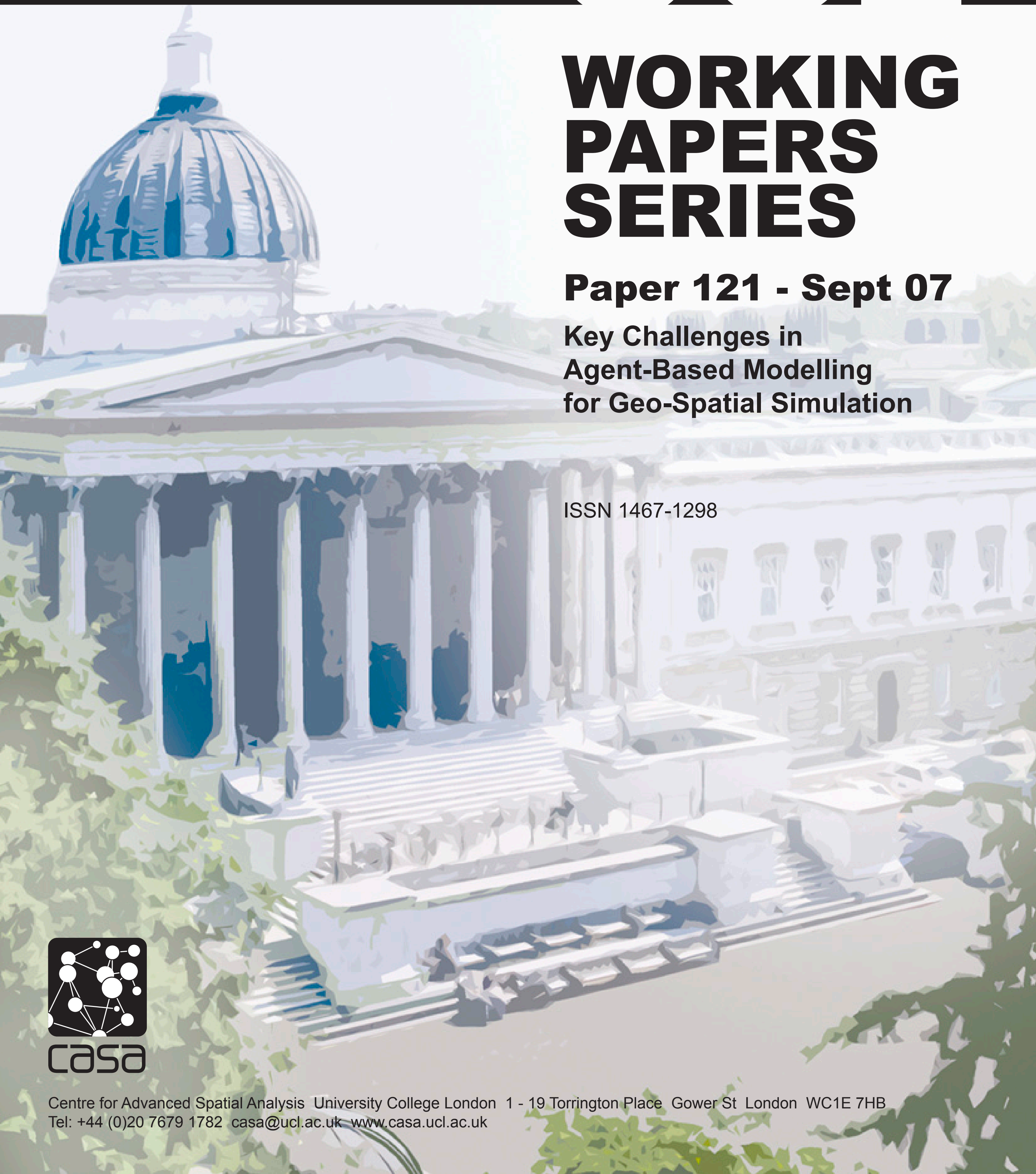
UCL CENTRE FOR ADVANCED SPATIAL ANALYSIS

UCL



WORKING PAPERS SERIES

Paper 121 - Sept 07

Key Challenges in Agent-Based Modelling for Geo-Spatial Simulation

ISSN 1467-1298

casa

Centre for Advanced Spatial Analysis University College London 1 - 19 Torrington Place Gower St London WC1E 7HB
Tel: +44 (0)20 7679 1782 casa@ucl.ac.uk www.casa.ucl.ac.uk

Key Challenges in Agent-Based Modelling for Geo-Spatial Simulation[1]

Andrew Crooks, Christian Castle, and Michael Batty

Centre for Advanced Spatial Analysis (CASA)
University College London
1-19 Torrington Place, London WC1E 6BT

Telephone: (44) (0) 207 679 1782
Fax: (44) (0) 207 813 2843
Email: andrew.crooks@ucl.ac.uk,
c.castle@ucl.ac.uk, m.batty@ucl.ac.uk

Abstract

Agent-based modelling (ABM) is fast becoming the dominant paradigm in social simulation due primarily to a worldview that suggests that complex systems emerge from the bottom-up, are highly decentralised, and are composed of a multitude of heterogeneous objects called agents. These agents act with some purpose and their interaction, usually through time and space, generates emergent order, often at higher levels than those at which such agents operate. ABM however raises as many challenges as it seeks to resolve. It is the purpose of this paper to catalogue these challenges and to illustrate them using three somewhat different agent-based models applied to city systems. The seven challenges we pose involve: the purpose for which the model is built, the extent to which the model is rooted in independent theory, the extent to which the model can be replicated, the ways the model might be verified, calibrated and validated, the way model dynamics are represented in terms of agent interactions, the extent to which the model is operational, and the way the model can be communicated and shared with others. Once catalogued, we then illustrate these challenges with a pedestrian model for emergency evacuation in central London, a hypothetical model of residential segregation tuned to London data which elaborates the standard Schelling (1971) model, and an agent-based residential location built according to spatial interactions principles, calibrated to trip data for Greater London. The ambiguities posed by this new style of modelling are drawn out as conclusions.

---

[1] This paper was first presented at **Geocomputation 2007**, National Centre for Geocomputation (NCG), National University of Ireland, Maynooth, Co. Kildare, Ireland, from 3-5 September 2007.

1. Introduction

Cities are constantly changing and evolving through time and across geographical scales where activities and features change from the split second decision involving local movements such as people walking, the development of land over months and years, the migration of peoples over decades, to the rise and fall of cultures and civilizations over eons. These sorts of problem which involve location and mobility have recently been articulated in much more disaggregate terms then hitherto with their system components or 'objects' being conceived of as agents where their movement takes place on a backcloth or in an environment composed of points, areas and networks. Such simulations began with automata used to grow systems in cell-like fashion but they have quickly evolved into models of mobile agents, interacting with one another in a landscape or environment usually constituted in traditional cellular terms (Batty, 2005) The big difference between these new approaches and the more aggregate, static conceptions and representations that they seek to complement, if not replace, is that they facilitate the exploration of system processes at the level of their constituent elements.

The development of these ideas is not without its problems and this paper will seek to identify these, posing them as key challenges to be addressed in fashioning these models to make them scientifically relevant and policy applicable (Axelrod, in press). We begin by posing seven key challenges and then illustrating these with three agent-based models coupled to geographic information systems (GIS). The challenges that we identify range across the spectrum of theory to practice, hypothesis to application, beginning with the purpose for which the model is intended, and then focusing on the extent to which independent theory lies behind the model. Fifty years ago when models became part of the scientific lexicon, it was always assumed that a model either represented an articulation of some prior theory or schema to elicit such theory inductively from data. In fact models are not often embedded in theory *per se* and thus the quest in grounding any model in theory is often to extract its essence from its operational articulation. In classical science, a model acts as some kind of experimental focus for testing a theory and the need to replicate the success or otherwise of the experiment in independent situations was regarded as the hallmark of science. This is something that is now much more uncertain with few attempts at such classical replication, at least in the social sciences. It throws the question of validation, verification and calibration into the melting pot, thus constituting another important challenge to developing good simulation.

We then explore the way agents are represented in such models, focussing on the need to identify multiple agents all acting either passively or actively in relation to one another and their environment. ABM deals with models that reflect processes of decision which agents make with respect to their location and in this context, these processes are usually temporally dynamic. This breaks with the notion that such models are parsimonious and in principle testable in the watertight manner of traditional science. We then illustrate the extent to which models need to be operational and this involves the way they are articulated in terms of software and the extent to which they are applicable to real situations and real data. It is entirely possible to develop ABM using pencil and paper; indeed the first such models were developed in this way a generation or more ago (Schelling, 2006) but now there are many variants which reflect different degrees of operationality. We conclude with perhaps one of the most important challenge – the need to share, communicate and disseminate not only model results with others but the understanding of such models as well as their operation. In this we are aided by advances in computation, particularly visualization and networked communications.

This sets the scene for illustrating these seven challenges with respect to geo-spatial simulation models which we have developed to operational status. Castle (2006) has been working with a pedestrian model of evacuation dynamics for the complex transport interchange at Kings-Cross / St. Pancras in central London which relates to the genus of such models of which the social forces model developed and popularised by Helbing and Molnar (1995) is typical. Crooks (2006, 2007) has been working with a Schelling-type residential segregation model at a coarser spatial scale which deals with processes of residential mobility over longer time periods. His model is hypothetical but tuned to London data. Batty has been working with a land use transportation model of Greater London built along traditional spatial interaction principles with many large zones at the census tract (ward) scale and this has been generalised to treat each trip maker as an agent. This focus of this model is on the journey to work where employees are allocated to residential zones according to the gravitational hypothesis. The scale is greater than the first two but the dynamics is more like that of the pedestrian model dealing with movements that take place over short periods of time.

Agent-based models have been developed for a diverse range of applications. To give a sense of the range, we note this diversity as: archaeological reconstruction of ancient civilisations (Axtell et al., 2002); understanding processes involving national identity and state formation (Cederman, 2001); biological models of infectious diseases (Eidelson and Lustick, 2004); growth of bacterial colonies (Krawczyk et al., 2003); company size and growth rate distributions (Axtell, 1999); price variations within stock-market trading (Bak et al., 1999); voting behaviours in elections (Kollman et al., 1992); spatial patterns of unemployment (Topa, 2001); and social networks of terrorist groups (North et al., 2004. These examples lie on a continuum, from minimalist models for academic research based upon idealised assumptions which are invariably used pedagogically or test clear and simple hypotheses, to large scale commercial decision support systems based upon real-world data. In many of these however, the representation of agents is critical and in several of these applications, the number and type of agents as well as their location in space and time is an important problem which raises many difficult challenges. Unlike earlier mathematical models of urban phenomena say, these kinds of models are much more generic, fashioned according to ABM principles but not embodied as a single modelling type to be tested widely on different applications but more tailored to a particular task in hand. This makes an enormous difference to their replicability and the way they are calibrated and applied as we will see a little later.

Despite the many advantages of geo-spatial agent-based models as a tool for simulating the micro-diversity of their systems of interest, their emergent properties, and their process dynamics (see Castle and Crooks, 2006), such models did not begin to feature prominently in social simulation and GI science until the mid-1990s after Epstein and Axtell (1996) demonstrated that the notion of modelling individuals making up an idealised society in space could be extended to growing entire artificial cities. The use of ABM for experimenting and exploring geographical phenomena however is still in its infancy (see Brown et al., 2005; Parker, 2005; Benenson and Torrens, 2004; Gimblett, 2002 for sample applications) and thus our focus here is on identifying key challenges to the development of such models. This makes their applicability somewhat different to the previous generations of models to spatial and urban systems, and we should stress that modellers should consider these challenges before they embark on such simulations. We will now outline these seven challenges, prior to showing how we are handling them in our own applications which we will present in the final section.

2. Key Challenges

The structure of a typical agent-based model composed of agents/objects/components which interact with each other and with their environment(s) is well known (see Castle and Crooks, 2006). Such models are usually considered as forming a miniature laboratory where the attributes and behaviour of agents, and the environment in which they are housed, can be altered, experimented with, where their repercussions are observed over the course of multiple simulation runs. The ability to simulate the individual actions of many diverse agents and measure the resulting system behaviour and outcomes over time means that agent-based models can be useful tools for studying the effects on processes that operate at multiple scales and organisational levels (Brown, 2006). Our models here roughly approximate the notion of generative social science articulated particularly by Epstein (2007) which proposes that models should be 'grown' within such simulation laboratories, thus explicitly rooting such models in temporal dynamics.

The seven challenges that we see as important to their development involve the following: the purpose for which the model is built, the extent to which the model is rooted in independent theory, the extent to which the model can be replicated, the way the model might be verified, calibrated and validated, the way model dynamics are represented in terms of agent interactions, the extent to which the model is operational, and the way the model can be communicated and shared with others. We do not consider this to be an exhaustive list but it is a beginning. Such challenges have been identified before (see particularly Axelrod, in press), but here we will address them in turn, first identifying each major issue and then demonstrating them where appropriate with actual applications.

2.1 The Purpose of the Model

Fifty years ago when computer models were first constructed for urban systems, these were always predicated on the notion that they were to be used for testing the impacts of urban plans and policies rather than scientific understanding *per se*. The argument went as follows: given a good theory, a model would be constructed which would then be validated and if acceptable, used in policy making. This rather tight loop has been relaxed in the last two decades and now models are built to explore all stages of the theory-practice continuum. This has largely occurred because the certainty of science has come under fire. The idea that the computer represents the scientist's laboratory is an attractive notion but when it comes to control of the inputs and parameters, most social systems cannot be represented in a form that guarantees any measure of closure. The difficulties and failures of earlier generations of urban model for example, bear testament to the fact that however good the fit of the model is to reality and to theory, there always seem to be features that are missing.

ABM relaxes all these assumptions and most of the social science simulations that we noted above do not focus these models on policy applications. In short, as ABM is generic, it is more a style of modelling which is largely independent of theory and practice and thus the purpose of any particular model will depend on issues that are often beyond the generic principles of ABM. In fact, whether or not ABM is appropriate for the theory and its applications, for the policies involved or for the design of systems that the model might be built to inform, cannot be guessed in advance. Thus only when we broach particular problems and develop particular models, do these issues become clear.

Frequently in ABM, the actual purpose and position in this scientific process is unclear largely due to the changing conceptions of how to do science and also the fact that agent-based models deal with systems that are complex, open-ended, hence emergent and thus exhibit novelty and surprise. However a model is only as useful as the purpose for which it was constructed and for agent-based models, this needs to be clear. A model has to be built at the right level of description for every phenomenon, judiciously using the right amount of detail for the model to serve its purpose (Couclelis, 2002). This remains more art than a science (Axelrod, in press) but of course, this is using the term science in its narrow sense for there is as much art in science as science in art. The purpose of agent-based models range from the explanatory to the predictive (see Castle and Crooks, 2006) with prescriptive and design models of increasing importance, although as we have noted, there is less focus on policy and prescription with this style of simulation than in previous, more aggregate modelling.

2.2 Theory and Model

Models should be based on theory and the traditional role of a model in the social science is as a translation of theory into a form whereby it can be tested and refined. In this sense, a computer model provides a computer laboratory for virtual experimentation, and hence a vehicle for refining theory through 'what if' style experiments and sensitivity testing. In fact as scientific method has blurred from this classical tradition, then increasingly models are being used to develop theory. In fact, the term theory has fallen out of favour in many contexts as models themselves contain theories. Our concern here however is that the theoretical implications of many agent-based models remain implicit and hidden, often covered by a thick veil of *ad hoc* assumptions about structure and process as well as a veneer of software interfacing. In many models, it is hard to figure out what they are for as they are simply additional applications of some simple structure which is tweaked for the local context and application. Domain knowledge is often lacking and increasingly agent-based models are being considered generic, independent of any particular field or application, and hence subject to use for any purpose that arises in a pragmatic way. In short, the scientific standards of the past are often buried in *ad hoc* model development.

We do not believe that ‘theory’ should necessarily be independent of ‘model’ for we are well aware that new styles of model now embrace theory in quite a different manner from those hitherto. But we do consider that in developing good models, there needs to be recognition that many styles of theorising and thinking must be brought to bear on model construction. For example, in our understanding of urban spatial structure, there is a long heritage of location theory in the urban and industrial economics domains with this theory being reflected in equilibrium micro-economics of the individual and the firm. This theory has produced many important insights and in any agent model of residential development, say, we might expect such issues to be reflected. The style of theorising in micro-economics is quite different from its embodiment using ABM but we would consider such theory essential to the structure of such models. In the same fashion, the notion of bringing large groups of researchers with different interests into the development of such a model, polling their insights and resources across the web, is also important to the development of large scale models, and this represents quite a different source of theory and understanding from the traditional practice whereby single individuals develop theory in a more formal, considered manner. In short, what we are saying is that the domain of model and theory is now considerably wider than at any time in the past and ABM must respond to such complexity.

2.3 Replication and Experiment

It is a canon of scientific inquiry that a theory that withstands the test of time is more likely to inform our understanding than one that can be easily refuted. In short, this is the inductive hypothesis that suggests that the more confirming instances of a theory, the stronger it becomes. This is the quest for generalisation in that a theory that works for one set of circumstances should work for any other and as long as the theory is not refuted, it remains ‘true’. Of course, all this has been turned on its head by the notion that theories cannot be confirmed but only falsified and that even a theory that withstands many confirmatory instances and applications has no greater probability of being true than any other. It only takes one falsification to sink it.

Nevertheless, our intuition suggests no matter how wrong this might be, our confidence in a model (or its theory) always increases the more confirming instances we have of its successful application. To pursue this, we need to replicate the model in independent situations. This is rarely done in the social sciences largely because of the difficulties in controlling all the variables that pertain to a particular situation, thus making it almost impossible to ensure comparability in terms of applications. A more limited quest which we discuss below is to make sure that the model can be verified in different laboratory situations, with different software for example, rather than different data. This is a much lesser quest. A reverse form of replication involves testing different kinds of models or different variants of the same model, for example in different software systems, on a standard data base. Axtell et al. (1996) refer to this process as 'docking' where it is clear that we have much greater control over the model than the data; hence keeping the data fixed and varying the model does give some insight into the robustness of each model. In fact in cases where rather similar models have been compared on standard data sets as in the case of the large scale study of urban land use transportation models (based on spatial interaction) conducted in the late 1980s, it was found that so many idiosyncratic decisions made by the modellers with respect to their different models and in terms of the way the data was configured and defined in different places, made comparisons almost impossible (Webster et al., 1988). And this was for parsimonious models of a rather narrow genus where applications were quite standard in terms of the data required.

2.4 Verification, Calibration, and Validation

If an agent-based model is developed to the point where it can be used to generate outcomes/results, there are several different tests that can be made and each of these involves a key challenge. Some models do not reach the point where they are set up as simulations for their purpose may be simply to articulate a way of thinking about the problem, even in terms of heterogeneous agents, environments and their interactions. However most ABM will embrace issues of testing and these can conveniently be identified as verification, calibration and validation. We will deal with these in turn.

Verification and validation are often confused in terms of their terminology but here we will define verification as the process of testing whether or not the logic of the model is acceptable. This is as a much a matter of testing the logic of the model through its computer programme as testing its formal logic. It involves checking that the model behaves as expected which is something that is often taken for granted. It is sometimes referred to as testing the 'inner validity' of the model (Brown, 2006; Axelrod, in press) but we will not use this phraseology here as it tends to confuse verification (which does not involve any external data) with validation. Validation relates to the extent that the model adequately represents the system being modelled (Casti, 1997) and in this sense, it involves the goodness of fit of the model to data. However, the validity of a model should not be thought of as binary event (i.e. a model cannot simply be classified as valid or invalid); a model can have a certain degree of validity (Law and Kelton, 1991) which of course is encapsulated by various measures of fit. Validity can thus be ascertained by comparing the output of the model with comparable data collected from a real-world system using various statistics over which there is usually quite intense debate. The question of what best statistics to use in model fitting is something that has dominated the literature on models of land cover, for example. Moreover, there are also more qualitative evaluation of model validity that might be made. For example, Mandelbrot (1983) argues that good models which generate spatial or physical predictions that can be mapped or visualised must 'look right'. Axelrod (in press) suggests that to understand the output of an agent-based model, it is often necessary to evaluate the details of a specific simulation 'history' and this too is usually a qualitative matter.

Calibration involves fine-tuning the model to a particular context and this means establishing a unique set of parameters that dimension the model to its data. This is not validation *per se* but calibration can often involve validation because the parameters are often chosen so that performance of the model is optimal in some way, in terms of some criterion of goodness of fit, for example. This is a large subject area and suffice it to say, many if not most agent-based models suffer from a lack of uniqueness in parameter estimation at this stage.

Concerns have been raised pertaining to verification and validation by numerous researchers (e.g. Batty and Torrens, 2005; Parker et al., 2002). Batty and Torrens (2005) write that with respect to developing traditional models, two rules have been taken as central to the process of developing good models in the social sciences. The first is the rule of parsimony – Occam's razor – which suggests that a better model is one which can explain the same phenomena with a lesser number of intellectual constructs. The second principle relates to independence in verification. A theory which is induced using one set of data needs to be validated against another independent set, and this obviously relates to our earlier discussion about replication. While it is sometimes possible to achieve this with traditional models, this is not the case for models developed using ABM principles, particularly where this involves human systems which evolve over time. Modellers are embracing increasingly diverse and richer model structures containing large numbers of parameters. Often with traditional models, parsimony is reflected in the linkage of dependent and independent variables while agent-based models have multiple causes which display heterogeneity of processes that are impossible to observe in their entirety (Batty and Torrens, 2005). Thus these new model structures are never likely to be validated in any complete sense against data; they are too rich and data needed to test them too poor (Batty et al., 2006).

2.5 Agent Representation, Aggregation and Dynamics

In spatial systems, what constitutes an agent is a critical issue in that the term can be applied to any aggregation of objects at any spatial scale and across different time horizons. Moreover it need not be restricted to human objects but might pertain to any object that exists in space and/or time. A slightly more restrictive definition of agents has been adopted in some spatial models and we adhere to this here in that we consider spatial agent-based models to deal with agents that have some form of mobility (Batty, 2005). Agents that do not move such as cells in cellular automata we would not define as agents in this context. Some of these issues of representation are clarified in the examples that we introduce below, particularly in the way we represent their applications and outcomes using various forms of graphic.

The scale of agents is also an issue as the finer the scale, the less ambiguous the definition, although we appreciate that this is contentious. This means that there are greater difficulties in specifying rules for defining agents which are aggregations of lower level units – i.e. groups within a human population, or defining abstracted agents such as a forest or a farmer or a city which pertain to models that in themselves are generic. In particular as we aggregate, we can unwittingly change the kinds of processes that agents enable, the kinds of mobility intrinsic to their location, and the scale at which they exist. It is thus more and more difficult to define relevant processes as these too are aggregations of lower level routines and behaviours for aggregation can confuse our identification of coherent patterns that make sense in terms of basic human decision-making.

Another issue involves the sheer number of agents and the sheer number of attributes and processes that they are engaged with. Like all systems that deal with interactions and networks, the size of the computation usually rises as the square of the number of agents, if not faster, and there are always limits on our ability to deal with such exponentiation. Sampling is often a favourite strategy to deal with multitudes but we must be cautious about proposing models that seek to represent behaviour at its most elemental level and then simplifying this back through taking samples. Sampling is not a well developed art in ABM as yet. Moreover choices are necessary in terms of the number of agents and processes which are reflected in the software used, the computational time involved, and of course the ability to get data that matches the specification of the model. In general, most agent-based models are tested against a fraction of data that could be applied to them in that many implicit explicit assumptions about behaviours cannot be observed as data does not exist. This reflects the issues about validation and calibration which we have already noted above as our fourth challenge.

2.6 Operational Modelling

Making agent-based models operational means moving them to the point where they are configured as simulation models and running them so that they might produce outcomes. In the past, most models have been programmed from scratch and although this keeps the application in touch with theory, it makes the ability to generalise the model to other situations, to replicate the model that is, difficult as the study previously referred to by Webster et al. (1988) indicated. What has happened with ABM is that because this implies a generic approach, various software are now evolving that like GIS, are being used to enable such generic applications. As always, the extent to which generic software can be well-tuned to specific situations will vary dependent on the application and its complexity, and besides the advantages of consistency and modularity that such software enables, it is always limited in its applicability.

In terms of ABM as in other areas of simulation and representation, such software enables modellers to adapt it to their problem context, implementing their model through high level scripting, for example, which the software usually allows. This opens up models to a wider community of scholars than hitherto but it also forces modellers without the skills or resources to develop their own models from scratch to meet constraints posed by the software. This can be a key problem when limits posed by the software on the numbers and representation of agents occur. Nevertheless, the development of agent-based models can be greatly facilitated through the use of simulation/modelling systems such as Swarm, Repast, NetLogo, OBEUS, etc. (see Castle and Crooks, 2006). They provide reliable templates for the design, implementation and visualisation of agent-based models, allowing modellers to focus on research (i.e. building models), rather than building fundamental tools necessary to run a computer simulation (Tobias and Hofmann, 2004; Railsback *et al.*, in press).

In particular, the use of simulation/modelling systems can reduce the burden modellers face programming parts of a simulation that are not content-specific (e.g. the Graphical User Interface (GUI), data import-export, visualisation/display of the model). It also increases the reliability and efficiency of the model because complex parts have been created and optimised by professional developers as standardised simulation-modelling functions. Additionally, the object-oriented paradigm allows the integration of additional functionality from libraries not provided by the simulation/modelling system, extending the capabilities of these toolkits. Of particular interest here is the integration of functionality from GIS software libraries (e.g. OpenMap, GeoTools, ESRI's ArcGIS, etc.) which provide ABM toolkits with greater data management and spatial analytical capabilities required for geospatial modelling. Castle and Crooks (2006) provide a comprehensive review of ABM simulation/modelling systems capable of creating geospatial agent-based models.

One feature that we have not yet seen in this field although it is possible that there are examples, is the design of models that take modules from a very wide library of components and literally throw these together in different combinations to make different types of model. The notion that with many different modules, a large array of model variants can be easily constructed enhances the use of theory in that a solution space of models types can be explored, literally. This has been done for solutions to individual models in terms of their parameter space but not really for different variants of models themselves. It is thus an attractive way forward. Repast, which we illustrate below, potentially has the components to enable this and we speculate that such diversity of modelling may well mark the next advance in this field.

2.7 Sharing and Dissemination of the Model

The last challenge involves how we might communicate and share agent-based models with all those who we seek to influence and who we and they believe that such modelling will inform their activities. In the past before the development of intensive and all pervasive computation, communicating models was mainly through discussion, simplification and visualisation, through pedagogy in all its various forms. Clearly visualisation is one of the keys to such sharing in that with digital models, their structure is easily amenable to visualisation. Of course spatial outcomes can be mapped and this is a key medium for dissemination as well as for validation and other aspects of the simulation process. But model structures can be described visually while the process of running the model, calibrating it, examining its inputs and outputs can be presented visually even while the model is running.

A good example of the power of such sharing is embodied in the current model-building capability within the GIS software ArcGIS (Maguire, 2005). The ability to drag and drop various modules – albeit only overlay (map) layers – and utilise various simple functionality – again albeit only map algebra type calculations – does offer an interesting way of involving those who are not expert in simulation in model construction. In fact much of the software that is now being evolved not only communicates and shares the modelling process and its outcomes with various non-expert participants but also enables non-experts to participate in the actual model construction. ABM in particular offers this possibility in contrast to earlier styles where the model was wrapped around a narrower professional expertise.

The other face of this revolution is the development of procedures for disseminating this kind of visualisation and model building process to whoever has an internet connection. Again an excellent example of this is the development of simple pedagogic software online. ABM is in the forefront of this as many simple examples such as the Schelling (1971) model indicate. In fact, a good example of this is on one of our own web sites where we simply took an example from the NetLogo site and embedded it in a web page (see http://www.genesis.ucl.ac.uk/model.html). The development of online laboratories – collaboratories for example – where model building and users engage in mutual and shared development activities although their infancy are very much on the horizon. The MOSES model at Leeds is a good example of the potential of this kind of activity (Birkin et al., 2006). The development of web site where many users develop agent-based models such as **NewTies** (http://www.new-ties.org/) is another good example of how this field is developing into a more sharing mode where collaboratories hold out great promise for new advances in the field of social simulation.

3. More General Challenges

To conclude our catalogue of challenges, we will briefly focus on more general issues in creating spatially explicit agent-based models before presenting various examples. While GIS is a particularly useful medium for representing model input and output of a geospatial nature, GIS are not well suited to dynamic modelling (Goodchild, 2005; Maguire, 2005) such as ABM. In particular, there are problems of representing time (Langran, 1992; Peuquet, 2005) and change within GIS (Longley *et al.*, 2005). To address these problems, numerous authors have explored linking (through coupling or integration/embedding) a GIS with a simulation/modelling system purposely built, and therefore better suited to supporting the requirements of ABM (e.g. Westervelt, 2002, Brown *et al.*, 2005).

ABM focuses on the individual, and thus the progress currently being made in the use of disaggregate data is an essential determinant of their applicability (e.g. Benenson *et al.*, 2002). Increased computer power and storage capacity has made individual-level modelling more practical recently. An example can clearly be seen in the evolution of pedestrian modelling (see Galea and Gwynne, 2006) where there has been a concerted movement from aggregate to individual level modelling. However limitations still remain when modelling large systems. For example, large and refined datasets of high-resolution information now exist for initialising agent-based models for urban simulations. For instance in the UK, there are now excellent databases on land parcels and associated land-uses (OS MasterMap Address Layer 2®), and road segment data available (OS MasterMap® Integrated Transport Network™ Layer). Current GIS are capable of encoding these datasets into forms that provide the foundations for such simulations along with providing spatial methods for relating these objects based on their proximity, intersection, adjacency or visibility to each other.

One major stumbling block is that there is potentially too much detail in these data for the current generation of computers to deal with when application to entire cities rather than just small areas are made. Thus agent-based models have the potential to suffer from similar limitations to those of the first generation of urban models developed in the 1960s

(Lee, 1973). However this can be overcome by considering the level of abstraction needed to examine the phenomena of interest for all the available detail is rarely needed. Or a series of smaller models could be created by examining specific aspects of the system. Second there is the lack of personal individualised data for the present and the past. For example, in the UK, the smallest measure of individual data from the Census is the Output Area which contains around 125 households. Sometimes access to more personal data can be obtained from commercial sources (see Benenson et al., 2002) or synthetic populations can be generated through micro-simulation techniques (Birkin et al., 2006) but dynamic, individualised data is in general a major problem which will continue to influence the development of such models in the foreseeable future.

4. Applications

4.1 A Pedestrian Model for Emergency Evacuation

The first of our models is based on simulating pedestrians exiting a subway station that is one of the busiest intersections in the London transport network. In particular the intersection of train lines and other transport modes are central to the new Eurostar station at King's Cross St. Pancras and projected visitor movements during 2012 Olympic Games. In 1987, the interchange was devastated by a major fire with considerable loss of life and the current redevelopment is still taking into account the safety recommendations identified by the Fennel (1988) investigation. In addition, the station must cope with the projected increase in future passenger demand from the new Eurostar terminal which will open in late 2007 and the projected 105,000 people will use the station during the morning peak (7-10am) during the Olympics. The model that was built is part of the appraisal by Camden Primary Health Care Trust who are responsible for the allocation and positioning of key emergency functions and facilities (e.g. ambulance loading point(s), casualty clearing station(s) to which the injured can be taken, etc.) in the event of a future emergency incident within the Underground station complex. The aim of the model is to predict the likely evacuation dynamic (i.e. total evacuation time, usage of evacuation routes, conditions experienced by passenger e.g. crowd density) given different very short term evacuation scenarios, future travel demand, and short term fluctuations in passenger use at different times of day and week (peak/off-peak and weekday/weekend).

With this well defined policy context, the model (which we refer to as the King's Cross Pedestrian Evacuation Model – KXPEM) is built on theory of movement associated with evacuation (Castle 2007b), and various other pedestrian evacuation models that have been developed (Castle and Longley, in press). Models such as this have been built by several researchers and are quite widely applied. In fact they are the closest of all agent-based models to traditional scientific models which are testable against data and are capable of being replicated in different situations. This is largely because of the simplified behaviour of the agents (e.g. optimised utility – least cost path), although the context of the problem is such that cultural and institutional differences, differences in geometric construction of facilities, and the standard emergency practice makes each application unique. Data is also difficult to obtain, especially flow and interaction data rather than routine count data of traffic volumes which is fairly basic. Developing such models with standard ABM software is quite possible and in this context KXPEM has been developed using the Repast framework as we note below, and as is clear from the GUI which is shown in Figure 1. Sharing the model is essential and although the application is still in desktop form, several groups of people involved in the Kings Cross evacuation scenarios, are involved in the model design and use. In short this is as good an example of an ABM tuned to a real problem as there currently exists.

The model was programmed in Java which is used in relation to the agent-based simulation/modelling toolkit Repast. Our comprehensive assessment of existing software (Castle and Longley, in press) identitfied this as a viable option rather than using an off the shelf packages such as Legion, buildingExodus, etc., so that we thoroughly understand the requirements of a pedestrian evacuation model. The conceptual model of KXPEM was based on theory and principles of pedestrian evacuation modelling identified by Castle (2007b) where our focus on issues such as geometric enclosure representation, occupant and enclosure perspective, speed and direction of occupant movement, and the behavioural perspective of pedestrians, are all critical factors that have influenced the design of this application. As pedestrian movement involves representing a relatively well-behaved but heterogeneous population of occupants, a coarse grain network model of movement was deemed inappropriate, whilst it was deemed unnecessary (given the purpose of the model) and impractical (due to limitations of current ABM software) to develop a model using continuous space. The enclosure representation is thus based on a regular latticed representation (i.e. 50cm by 50cm cells).

Each pedestrian is defined as a movable object of which groups are definable in terms of age, gender, and passenger type, thus permitting a heterogeneous population of passengers. Furthermore, KXPEM permits either a global or individual perspective of the station layout as pedestrians can be defined as either irregular or regular passengers. The calculation of a regular passenger's exit route-choice is based on the assumption of prior knowledge of the station layout, and habitual use. Conversely, occasional passengers are programmed with limited knowledge of the station layout. Their exit route choice is calculated based on the assumption they will follow emergency signage to exit the station. The speed at which a pedestrian can walk is dependent upon their available space (i.e. density of pedestrians within a local area), as well conditions of their local environment (e.g. surface terrain) and characteristics of the individual (e.g. age and gender). Four secondary data sources for pedestrian walking speed as a function of available space and surface terrain, have been incorporated into the model to explore their effect on simulation outcomes: these come from Hankin and Wright, (1958); Ando et al. (1988); Fruin (1971); Predtechenskii and Milinskii (1978) and are based on widely agreed observed movement patterns. Exit route-choice and way-finding are defined by cost-surfaces, akin to many other pedestrian evacuation models that adopt a regular lattice approach to enclosure representation.

Based on the assessment criteria of pedestrian evacuation models (Castle, 2007b), KXPEM adopts a rule-based approach to simulate occupant behaviour. Evacuee decision making is separated into pre-evacuation (e.g. length of time required to perceive and investigate an evacuation cue or alarm before initiating movement) and evacuation components (e.g. the effects of crowding and prior knowledge of the structure upon route choice and walking speed). The rules that determine the response of each pedestrian when confronted with a decision are a combination of deterministic and stochastic responses based on information derived from literature (Castle, 2007a).

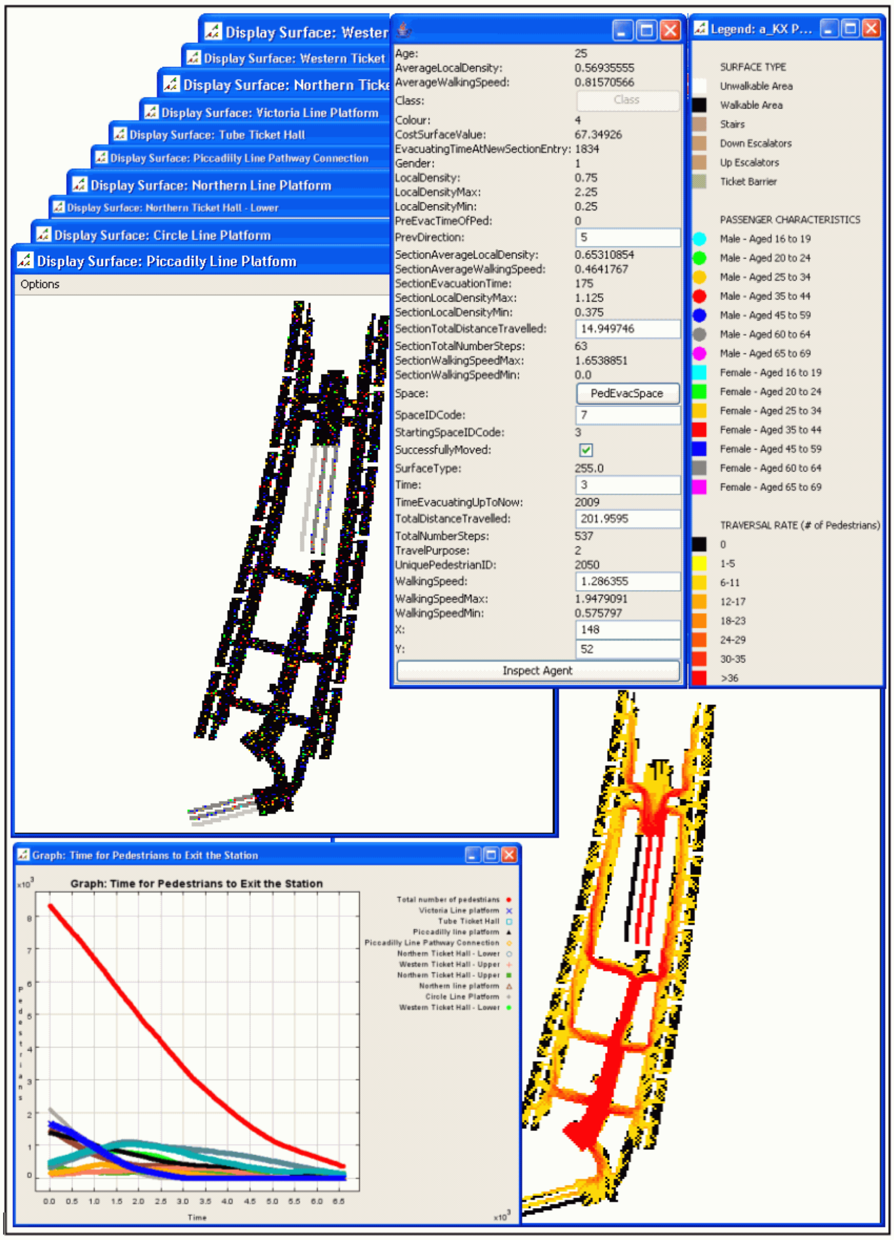


Figure 1: The Graphical User Interface of KXPEM

Illustrating the Starting Location of Pedestrians on the Piccadilly Line Platform (top left), Exit Time Profiles for each Section of the Station (bottom left), the Accumulative Exit Path of Pedestrians from the Piccadilly Line Platform (bottom right), and the Parameter Values Used.

The calibration of KXPEM was an intensive process. For example, Computer Aided Drawing (CAD) floor plans of King's Cross Underground station were used to ground the model with respect to the accuracy of enclosure's layout and therefore its capacity. In line with the models purpose, cost surfaces were developed to explore three evacuation scenarios. Two scenarios are defined by the UK Health and Safety Executive (HSE, 1996) to assist in the design and analysis of escape capacity from railway stations. Specifically, these are first, a train on fire at a platform, and second, a fire within a station structure. The third scenario permits the simulation of pedestrians from the station without incident. In particular, extensive periods of observation were made of pedestrian movement at the station in order to calibrate the cost surfaces used to specify direction of pedestrian movement and route-choice. In addition, surveys at King's Cross Underground station were used to determine passenger volume and characteristics (e.g. age, gender and passenger type at different times of the day and week), in order to specify these parameters within the model. In terms of enabling the model to be built and used, an advisory panel was set up to facilitate its development, in particular to gain access to necessary information, often not in the public domain (survey data and CAD floor plans, for instance), and to advise on the development and calibration of the model. These included the British Transport Police, Camden Council, the Health Protection Agency; London Underground Limited, the Metropolitan Police Service, Network Rail, Transport for London, and the Camden Primary Care Trust.

Typical of most ABM and related simulations, the development of KXPEM was an iterative process where model verification involved many iterations of the system. The final model was the 35th version where each version represented a major progression and backup point. A descriptive log of each programming progression was kept in case the author needed to reverse any changes. In total, over one hundred programming iterations were made. Unit testing was undertaken after every adjustment to the programming code. As the model became more complex, small and quick changes to the code took several hours to verify. Unit testing was achieved through Eclipse IDE debug mode, print lines, and by visual analysis of the model. Following this meticulous regime of verification, confidence was gained in the model, specifically in terms of model processes taking place at the right time, and each process occurring in the manner in which it was intended. KXPEM was designed and developed for predictive purposes but information regarding past evacuation drills at the station and detailed empirical data on passenger flow was largely absent. In this sense, the model was not validated in the traditional manner against real-world data, and it is thus better suited for exploratory purposes at present. The visualisations shown in Figure 1 not only reveal the process of building and testing the model but the kind of outputs that non-expert users can relate to.

4.2 A Model of Residential Segregation

Our second application involves an extension to Schelling's (1971) classic model in which individuals with very mild preferences to live amongst their own kind generate highly segregated districts when they move in response to any mismatch between their preferences and the configuration of their own and different types in their immediate neighbourhood. The purpose of this model is to explore the impact of space and geometry on such a process and in this sense, it is simply a pedagogic demonstration, a hypothetical demonstration of how individuals react to one another with respect to their preferences. The model deals with more than two groups of individuals, thus showing how segregation can occur in much more realistic systems than in the simpler system composed of two types of individual used by Schelling (1971). In so far as the model is grounded in theory, this is in the notion that individual action does not lead to any collective welfare, quite the opposite and it is an example of how unfettered and uncoordinated actions at the individual level lead to unexpected outcomes that are collectively undesirable, mild preferences revealing what are often quite wrongly judged to be extreme preferences. Although dimensioned to characteristics of populations in Greater London, the model is not capable of validation in any strict sense. Its theoretical basis is commonsensical, and the model is uncluttered with additional variables that might affect segregation – for example, how economic factors may contribute to racial segregation based systematic income differences across groups as well as price and quality of life arising from lot size and other amenities.

The model departs from other models which either explore or extend Schelling's original insights (e.g. Bruch and Mare, 2005; Fossett and Senft, 2004; Laurie and Jaggi, 2003; Omer, 2005; O'Sullivan et al., 2003) which are all based on the regular partitioning of space (e.g. cells or polygons) to represent the location of households (Benenson et al., 2002). The focus here is on how different conceptions of spatial organisation affect the process of segregation with the model allowing agents to move anywhere within the urban environment (i.e. movement is not restricted to discrete cells or areas). The model explores how segregation in space emerges as agents move to new locations, and how segregated areas grow and decline over time. In this sense, it makes Schelling's model much more explicitly geographical than any other applications to date but it is easy to replicate and is an ideal basis for experimentation. The fact that it can be demonstrated using a whole range of media from pencil and paper to a variety of types of computation – on the desktop, the web etc., illustrates its pedagogic quality and the ease with which the model can be shared amongst non-experts as a demonstration of how complex, unexpected, and surprising patterns emerge from simple foundations.

In GIS terms, the model is comprised of two vector layers – the urban environment represented as a series of polygons, and four types of agents (red, blue, green and white) represented as points. It is the information held within fields of the environment layer that is used to create the agents. The distribution of four types (ethnic groups, say) of agent as observed through aggregate census population counts form the initial starting conditions of the model. Figure 2A represents four wards in the City of London each with their own 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 data held in the fields as in Figure 2B. Note that the underlying colour of the polygon (ward) always represents the *predominant* social group in the area. This model is designed to work on many different geographical scales (e.g. boroughs, wards, output areas, and OS MasterMap TOIDs) without the need for model reconfiguration as we show in Figure 3. This was considered important as most socioeconomic data comes in this format, for example, census and geo-demographic data. This functionality was created so that the model could be easily replicated in other areas in the quest to allow the modeller to see if the same rules can be applied to different areas and at different scales. Replicability is one of the key challenges we identify above and using modular software such as Repast in which the model is scripted, enables such flexibility in application.

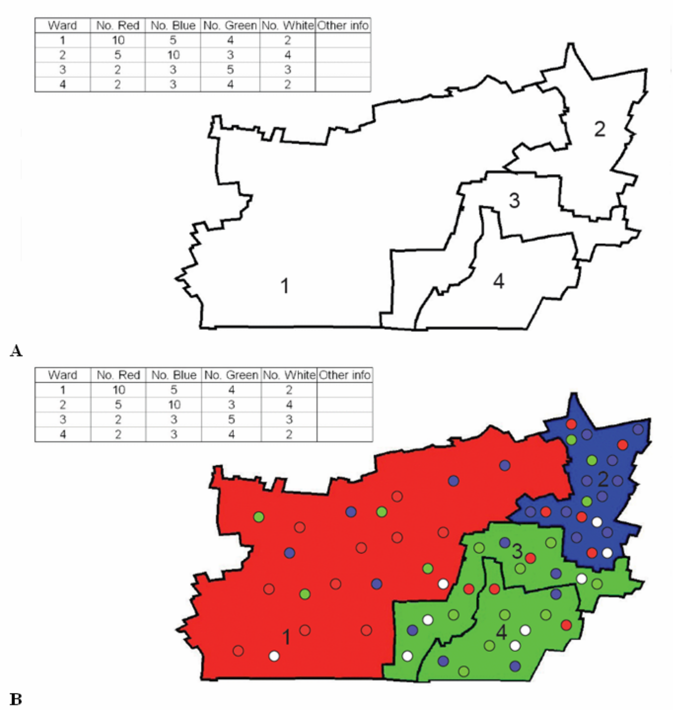


Figure 2: Reading in the Data and Creating the Agents

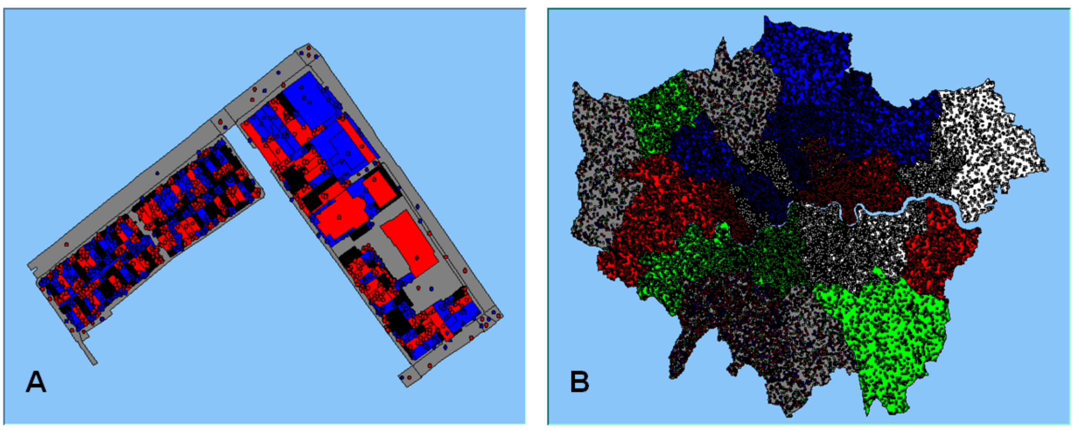


Figure 3: Spatial Representation within the Model

A: A Street Section. B: London Composed of Boroughs. Agents Are Shown As Dots.

In this model, agents only move if they find themselves in a minority which we have set as being less than 50% of the same kind in their area (neighbourhood). While Schelling's original segregation model is an excellent example explaining residential dynamics, there are limitations. First, reality is much more complex and for this reason, the model has been extended in several ways, particularly in defining more than two groups. For example within London, there are numerous types of ethnic or socioeconomic group and thus this extension explores the impact of four different types of agents (although the model can permit any number) defined as white, red, blue or green. Each agent has a preference related to residential contact (co-residence) with members of each other group. Second, not only are we interested in how patterns of segregation evolve over time but how this pattern changes with the introduction of new agents and the death of older agents, and thus the model allows for the addition and removal of agents which has an important effect on the pattern of segregation seen within urban areas.

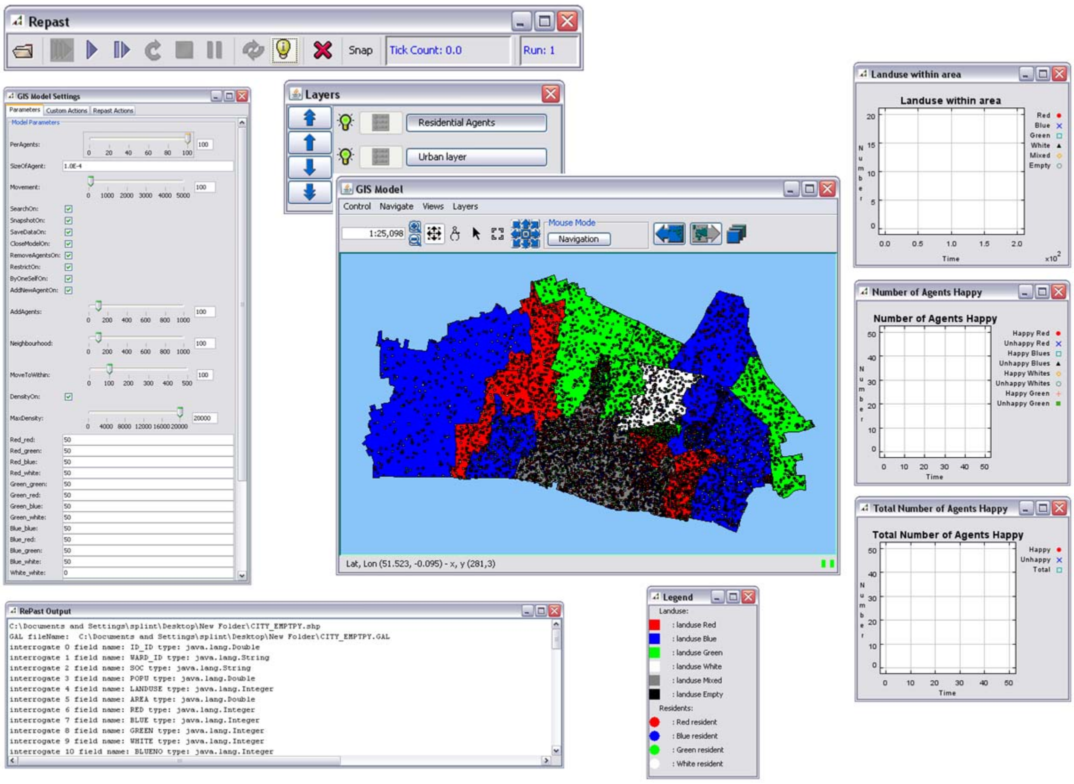


Figure 4: Segregation Model User Interface

Showing Segregation Dimensioned to the Geography of Wards in the City of London.

In Figure 4, we highlight the graphical user interface to the model. Clockwise from the top left is the control bar for the simulation, the GIS display which shows the agents and the urban environment (i.e. wards in the City of London), graphs for aggregate outputs, a legend for interpretation of the GIS interface, model output in the form of text, and the model parameters. We are not able to validate the model *per se*, except through testing its plausibility in commonsense terms but we can verify the structure. This is achieved by building the model iteratively similar to Castle's (2007a) approach, each step extending the basic model, providing greater realism and functionality. At each step, unit testing was carried out, executing the computer programme after each modification of the code to check that a mistake in the computer programme (a 'bug') had not been introduced. This permitted the identification of unexpected outcomes of the model itself as opposed to errors in the code.

Once the model was verified, a series of experiments were carried out in order to test the sensitivity of the model and to highlight the effect of the underlying model assumptions. Simulations involved between 1000 to 14000 agents. This exploration provided a detailed understanding of the implications of each assumption but also allowed one to evaluate the logic behind the model. This included the influence of the size of neighbourhoods, the influence of geographical features and the degree to which segregation changes when agent preferences for neighbourhood composition change. These explorations showed that geometry of an area can act as a physical barrier to segregation and that by increasing agents' preferences to reside by a specific group, marked segregation can emerge but not in a linear progression. A distinct shift in the degree of segregation occurs when agent preferences increase from 40% to 50% of their own type. As with the more 'traditional' segregation models, this model also highlights how with mild tastes and preferences to locate amongst 'like' demographic groups, segregation will emerge. Adding agents and removing agents from an existing population alters existing patterns but for new groups entering the system, they must have either low tolerances for other groups or be willing to live by themselves in order to become established. The model illustrates how small minority groups cluster in areas and how these clusters remain persistent over time, outcomes which are well beyond what Schelling showed in his initial model.

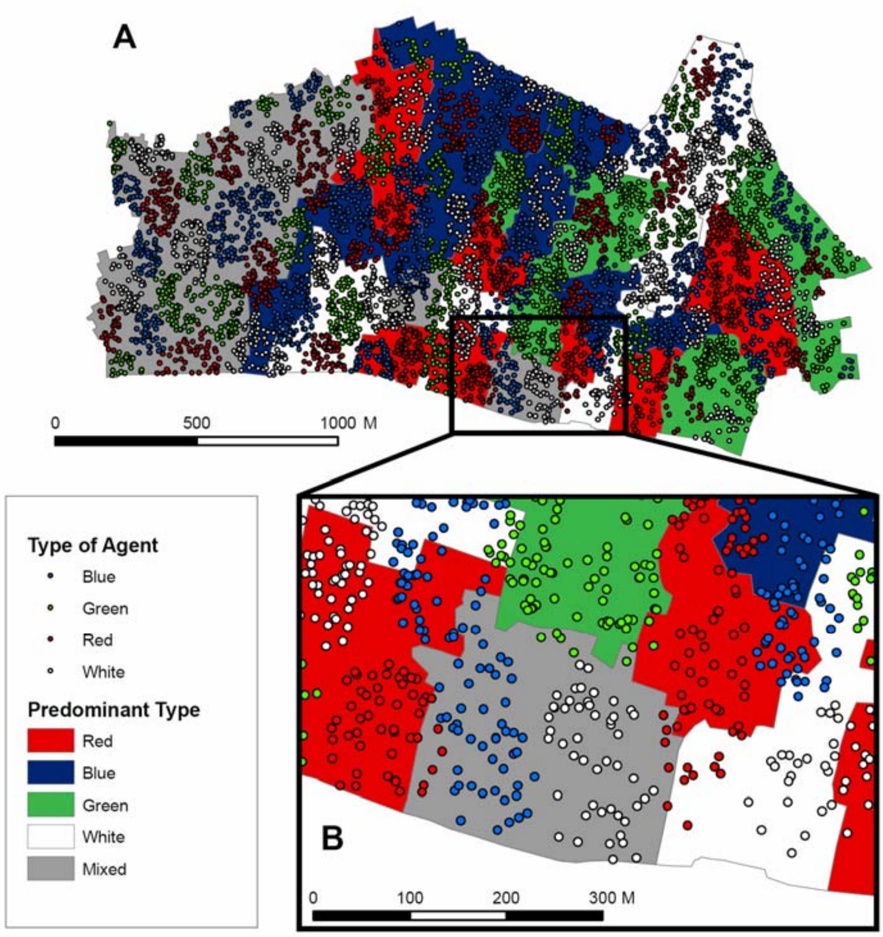


Figure 5: Segregation within Areas and across Boundaries

A: The Entire Area, B: A Zoomed in Section of A

In Figure 5A, we show a representative simulation outcome, where all agents are satisfied with their current neighbourhood locations. While areas may have a predominant type of one agent within them (e.g. a polygon shaded red, say, has more red agents than any other type), there are areas where there are equal numbers of two or more groups (grey areas). However closer inspection of these mixed areas in Figure 5B reveals distinct micro clusters of different types of agents. Moreover it is also clear that clusters do not stop at boundaries but cross them as well and these clusters would be lost if we were only to consider aggregate level data without the ability of agents to move in free space. Finally this model like KXPEM, is highly visual and in some respects is more modular in its construction. It has been put together using as much open source software as possible, built around Repast but using GeoTools and OpenMap as well as being coupled to ArcGIS in terms of its inputs and some outputs. It is still only a desktop application but its results are being disseminated across the web which provides a good example of this pedagogy (see www.casa.ucl.ac.uk/andrew/phd/). It is not designed for policy applications *per se* although policy is a clear consequence of such thinking. It is a 'classic tool to think with', part of the growing arsenal of techniques and tools useful for informed discussion of urban problems.

4.3 A Residential Spatial Interaction Model

Our third application involves a more traditional model of spatial interaction which is articulated at the level of small zones in terms of employment and population aggregates. Such models allocate employment associated with small zones to residential locations, often the same set of small zones, through simulating interactions, in this specific case, in the form of the journey to work. The logic of interaction is based on the well-established gravitational hypotheses where the flow from employment site to residential location is inversely proportional to some measure of the impedance – distance or travel cost between these origins and destinations, and directly proportional to some measure of attraction or size of each two locations. The model we are building is part of an integrated assessment of climate change impacts on the Greater London area and as such it is designed to provide small scale population estimates for long term (2050 and 2100 scenarios) interfacing between higher scale climate predictions which are factored into the regional employment and population estimates from an environmental input-output model and lower scale models related primarily to flooding and environmental risk (Dawson et al., 2007). The model is also designed to predict trips in four modes of transport – car, bus, rail and tube – which each have a considerable impact on how the pattern of residential population might adjust in terms of the way people might travel in the future.

The purpose of this model is clear and the theory on which it is based very well-established as a corner stone of classical social physics. It is also has strong links with urban economic theory and the operation of land markets through the trade-off between accessibility and the attraction of different locations. This sort of model has been replicated many times and because it is comparatively parsimonious, it can be fitted at the aggregate level to available data. Flow matrices for each mode of travel represent the key data for validation and calibration is accomplished through tuning the model to reproduce the known trip lengths for each mode. The key problems with such models relate to the fact that the heterogeneity of location is not represented other than in the distinctions between the small zones used in predicting aggregate trips. In short, although there are over 600 zones defined for the model (see Figure 6 where we show the overall data input sequence and work trips from one zone), the fine grained spatial detail which characterises each place is not picked up whatsoever and there is considerable variation within each zone. To reflect this, the model must be further disaggregated to the point where such detail is relevant and this probably means that each household or trip making entity needs to be represented separately. In short of the 4 million work trips in greater London, each of these needs to be represented in terms of location but without sacrificing the aggregate conservation properties of the spatial interaction models which enable realistic totals to be predicted.

What we have done is divide each of the 633 zones into a very fine grid which laid across greater London is of approximate dimension 1000 x 800 (each cell being about 60m x 60m). We then randomly allocate the known employment and residential totals, individual by individual, to these grid squares but constrained by the actual pattern of development in each zone. Non-residential and employment cells are thereby excluded and this produces a much more accurate pattern of trip making. As the distances used in the crude model are crow-fly/airline, this detail affects every individual who makes a trip. We are currently developing ways of building in real distance/cost surfaces into the model so that we can move beyond straight-line distances. Each one of the four million trips is then individually simulated in terms of choosing a residential location in an appropriate zone and as each trip is identified in terms of its origin, running the model at this individual levels conserves the total activity at each origin. In some respects the model is not unlike a disaggregate travel demand model built around discrete choice principles except that in this case, all the heterogeneity and choice is loaded on the spatial variation in locations.

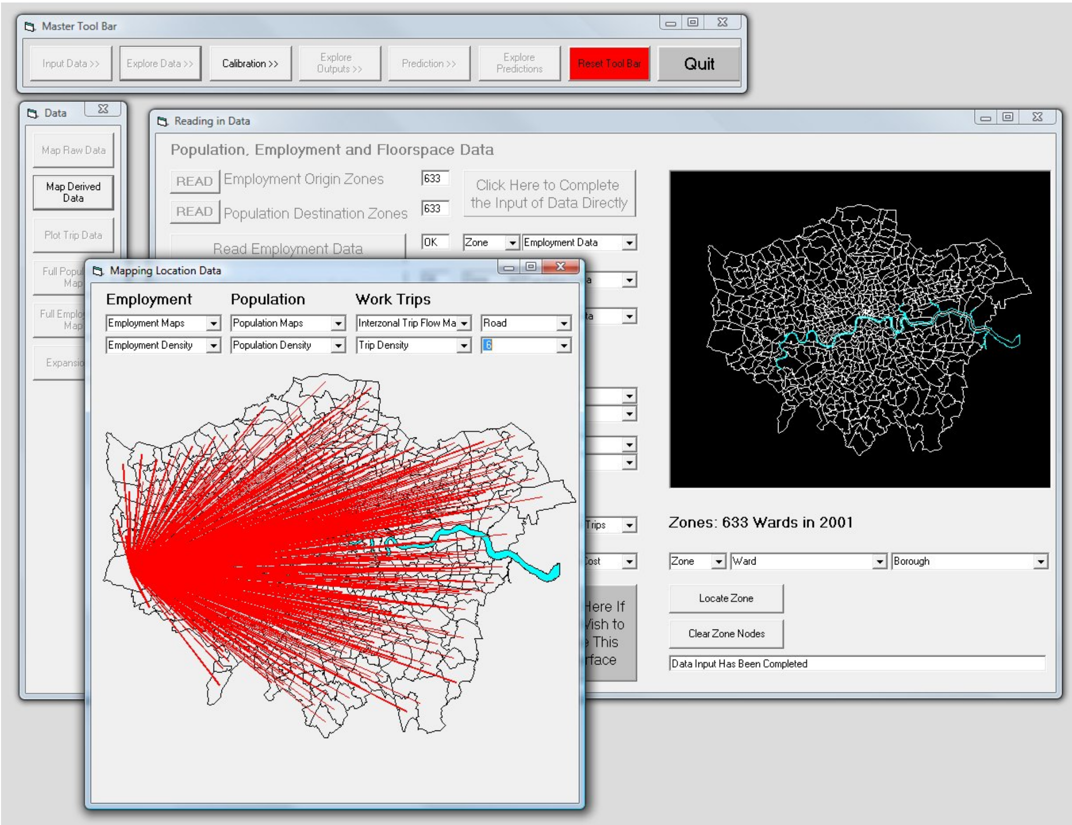


Figure 6: Inputting and Exploring the Model's Data

The Work Trips from Heathrow to the Other 632 Destinations are Shown along with the Zone Map and Drop-Down Menus Enabling the User to Interrogate the Data

In another sense, this conception of agent-based models breaks our rule that the agents are mobile, at least in the sense of an agent moving purposively. Agents move in that they make trips but they do so in entirely routine fashion, and in so far as a dynamics exists, it is simply in the way people respond to fixed locations and spatial impedances which are unvarying. This model is in fact a simulation of a static equilibrium, although the equilibrium is composed of individual agents which when aggregated to small zones meet certain conservation constraints. The model is also operational is a somewhat different way from the previous two. The model is not programmed in a particular ABM package like Repast but in a standard language – in this case in Visual Basic. Its graphics interface uses the components of Visual Studio and its maps are produced using various graphics primitives available in VB and as Windows APIs. The model is in fact programmed to show how the process of model construction is ordered and there are many graphics which enable the user to view data and predictions, and even to input various scenarios from the desktop as we show in Figure 7. But although the various windows resemble those of Repast, the similarity ends there as the model cannot be assembled in any other than the given order. It builds on earlier systems developed by Batty (1992). In short the model cannot be configured as a set of re-usable components which would involve a much higher level of coding and thus it is much more like an interactive graphics interface to a traditional batch process.

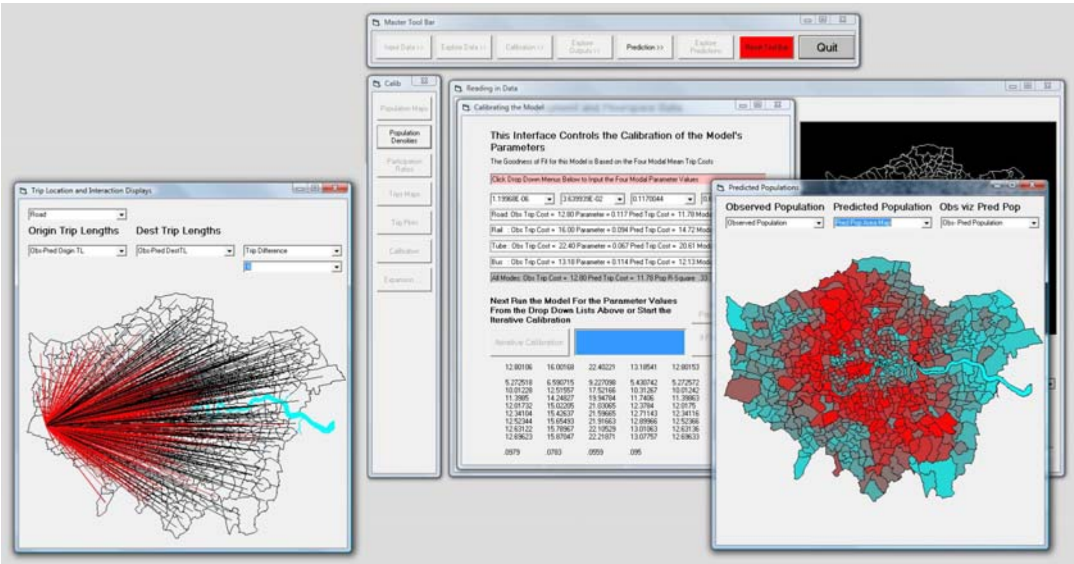


Figure 7: Calibrating the Model.

The Calibration Window, Aggregate Observed and Predicted Work Trips from Heathrow, and Predicted Populations are Shown.

However the model can be communicated and shared with others through its graphic interface but unlike the Repast models, we have not yet sought to embed its output in an online system. The main form of communication comes from enabling users to develop their own scenarios and in Figure 8, we show one part of this interface for inputting new levels of employment in each small zone of the system. As yet we have not integrated detailed transport networks into the model for these are configured within ArcGIS and at present, we do not foresee anything other than a loose coupling of software packages. It is unlikely that we will embed the model within a wider GIS as the sequence of software required for the entire integrated assessment is based on very different programming and data systems. In this sense, the model is more traditional in structure and there is no intention of developing it to the point where there are reusable software components. In a sense, this places the model into those which are specific rather than generic thus illustrating that decisions about software interact with questions of purpose, sharing, and communication.

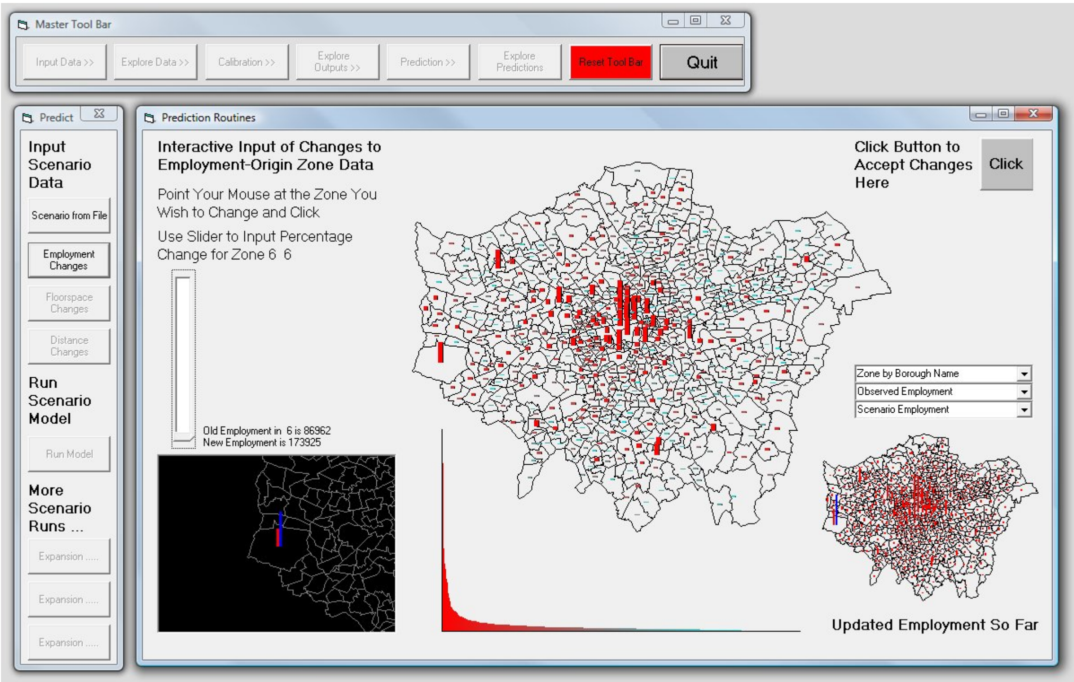


Figure 8: The Interface Enabling the User to Add or Subtract Employment from Any of the 633 Zones in Composing a Future Scenario to Test

5. Conclusions and Next Steps

These models demonstrate how the representation of individuals, through simple rules governing their behaviour and interaction at the micro-scale, can result in recognisable patterns at the macro-scale. The models apply different theories and concepts, highlighting how ideas pertaining to urban phenomena can easily be abstracted within agent-based models, helping further our understanding of how cities operate. Furthermore, these models help laminate the importance of incorporating space when modelling urban systems. Notwithstanding their potential, this class of geospatial models more than any developed hitherto raise challenges for the field that directly face the issue about the changing scientific method which is being forced by the development of computation and highly decentralised views of how spatial systems actually work.

The challenges we have identified here are not new for they pertain to all science which seeks to hypothesise the workings of a real system in the quest to develop both better understanding and tools for manipulating it, in silico, so to speak. The major challenge however which emerges from this discussion is the fact that agent-based models can be much more arbitrary than the models they both complement and replace. What is urgently required is some consensus about ways in which ABM can be structured so that major pitfalls are avoided. It is all too easy to develop models whose components seem plausible but are in fact rooted in false intuitions and unwarranted assumptions. Much of this relates to the goal of science which traditionally has been simplification but is changing to embrace possibilities for theories that are too rich to test but essential for coping with the evident complexity of the systems under scrutiny. Cities tend to be key exemplars of the dilemmas faced in such modelling, and the clear but short conclusion of this paper is that all such models should come under a much greater degree of scrutiny than any hitherto to avoid the sins of arbitrariness that plague a world where there are almost as many models and modellers. This represents a continuing challenge.

6. References

Ando, K., Ota, H. and Oki, T. (1988), Forecasting the Flow of People, *Railway Research Review*, 45: 8-14.

Axelrod, R. (in press), Advancing the Art of Simulation in the Social Sciences, in Reynard, J.-P. (Ed.) *Handbook of Research on Nature Inspired Computing for Economy and Management*, Idea Group, Hershey, PA.

Axtell, R., Axelrod, R., Epstein, J. M., and Cohen, M. D. (1996), Aligning Simulation Models: A Case Study and Results, *Computational and Mathematical Organization Theory*, 1(1): 123-141.

Axtell, R. L. (1999), The Emergence of Firms in a Population of Agents: Local Increasing Returns, Unstable Nash Equilibrium, and Power Law Size Distributions, Brookings Centre on Social and Economic Dynamics, *Working Paper No. 3*, Washington DC.

Axtell, R. L., Epstein, J. M., Dean, J. S., Gumerman, G. J., Sherlund, A. C., Hamburger, J., Chakravarty, S., Hammond, R., Parker, J. and Parker, M. (2002), Population Growth and Collapse in a Multivalent Model of the Kenta Anastasia in Long House Valley, *Proceedings of the National Academy of Sciences of the United States of America (PNAS)*, 99(3): 7275-7279.

Bak, P., Adzuki, M. and Subic, M. (1999), Price Variations in a Stock Market with Many Agents, Yale University Cowles Foundation, *Discussion Paper 1132*, Available at http://cowles.econ.yale.edu/P/cd/d11a/d1132.pdf

Batty, M. (1992), Urban Modeling in Computer-Graphic and Geographic Information System Environments, *Environment and Planning B*, 19: 663-685

Batty, M. (2005), *Cities and Complexity: Understanding Cities with Cellular Automata, Agent-Based Models, and Fractals*, The MIT Press, Cambridge, MA.

Batty, M., and Torrens, P. M. (2005), Modelling and Prediction in a Complex World, *Futures*, 37 (7): 745-766.

Batty, M., Steadman, P. and Xian, Y. (2006), Visualization in Spatial Modelling, in J. Portugali (Editor) *Complex Artificial Environments*, Springer, Berlin, 49-70.

Benenson, I. and Torrens, P. M. (2004), *Geosimulation: Automata-Based Modelling of Urban Phenomena*, John Wiley & Sons, London.

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: Planning and Design*, 29(4): 491-512.

Birkin, M., Turner, A. and Wu, B. (2006), A Synthetic Demographic Model of the UK Population: Methods, Progress and Problems, *Regional Science Association International British and Irish Section, 36th Annual Conference*, The Royal Hotel, St Helier, Jersey, Channel Islands.

Brown, D. G. (2006), Agent-Based Models, in Gist, H. (ed.) *The Earths Changing Land: An Encyclopaedia of Land-Use and Land-Cover Change*, Greenwood Publishing Group, Westport, CN, pp. 7-13.

Brown, D. G., Riolo, R., Robinson, D. T., North, M. and Rand, W. (2005), Spatial Process and Data Models: Toward Integration of Agent-Based Models and GIS, *Journal of Geographical Systems*, 7(1): 25-47.

Bruch, E. E. and Mare, R. D. (2005), *Neighbourhood Choice and Neighbourhood Change*, California Centre for Population Research University of California – Los Angeles, Los Angeles, CA.

Casti, J. L. (1997), *Would-Be-Worlds: How Simulation is Changing the Frontiers of Science*, John Wiley & Sons, New York.

Castle, C. J. E. (2007a), *Agent-Based Modelling of Pedestrian Evacuation: A Study of London's King's Cross Underground Station*, PhD Thesis, University College London.

Castle, C. J. E. (2007b), Guidelines to Assess Pedestrian Evacuation Software Applications, Centre for Advanced Spatial Analysis (University College London), *Working Paper 115*, London, Available at

http://www.casa.ucl.ac.uk/working_papers/paper115.pdf .
Castle, C. J. E. and Longley, P. A. (in press), 'Emergency Pedestrian Evacuation Analysis: A Review and Interpretation of Software Applications', in Sui, D. Z. and Cutter, S. L. (eds.), *Geospatial Technologies and Homeland Security: Research Frontiers and Challenges*, Springer, Berlin.
Castle, C. E. (2006), Using Repast to Develop a Prototype Agent-Based Pedestrian Evacuation Model, *Proceedings of the Agent 2006 Conference on Social Agents: Results and Prospects*, Chicago, IL.
Castle, C. J .E. and Crooks, A. T. (2006), Principles and Concepts of Agent-Based Modelling for Developing Geospatial Simulations, Centre for Advanced Spatial Analysis (University College London), *Working Paper 110*, London.
Cederman, L.-E. (2001), Agent-Based Modelling in Political Science, *The Political Methodologist*, 10(1): 16-22.
Couclelis, H. (2002), Modelling Frameworks, Paradigms, and Approaches, in Clarke, K. C., Parks, B. E. and Crane, M. P. (eds.), *Geographic Information Systems and Environmental Modelling*, Prentice Hall, London.
Crooks, A. T. (2006), Exploring Cities Using Agent-Based Models and GIS, *Proceedings of the Agent 2006 Conference on Social Agents: Results and Prospects*, University of Chicago and Argonne National Laboratory, Chicago, IL.
Crooks, A. T. (2007), *Experimenting with Cities: Utilizing Agent-Based Models and GIS to Explore Urban Dynamics*, PhD Thesis, University College London.
Dawson, R. J., Hall, J. W., Barr, S., Batty, M., Bristow, A., Carney, S., Evans, S., Ford, A., Köhler, J., Tight, M., Walsh, C. (2007) A Blueprint for the Integrated Assessment of Climate Change in Cities, *Tyndall Centre Working Paper 104*, University of East Anglia, Norwich, UK. Available at
http://www.tyndall.ac.uk/publications/working_papers/twp104.pdf
Eidelson, B. M. and Lustick, I. (2004), VIR-POX: An Agent-Based Analysis of Smallpox Preparedness and Response Policy, *Journal of Artificial Societies and Social Simulation*, 7(3).
Epstein, J. M. (2007), *Generative Social Science*, Princeton University Press, Princeton, NJ.
Epstein, J. M. and Axtell, R. (1996), *Growing Artificial Societies: Social Science from the Bottom Up*, MIT Press, Cambridge, MA.
Fennell, D. (1988), *Investigation into the King's Cross Underground Fire*, HMSO, London.
Fossett, M. and Senft, R. (2004), SIMSEG and Generative Models: A Typology of Model-Generated Segregation Patterns, *Proceedings of the Agent 2004 Conference on Social Dynamics: Interaction, Reflexivity and Emergence*, Chicago, IL, pp. 39-78.
Fruin, J. J. (1971), *Pedestrian Planning and Design*, Metropolitan Association of Urban Designers and Environmental Planners, New York.
Galea, E. R. and Gwynne, S. (2006), *Principles and Practice of Evacuation Modelling* (7th Edition), Fire Safety Engineering Group - University of Greenwich, London.
Gimblett, H. R. (2002), *Integrating Geographic Information Systems and Agent-based Modelling Techniques for Simulating Social and Ecological Processes*, Oxford University Press, Oxford, UK.

Goodchild, M. F. (2005), GIS, Spatial Analysis, and Modelling Overview, in Maguire, D. J., Batty, M. and Goodchild M. F. (eds.), *GIS, Spatial Analysis and Modelling*, ESRI Press, Redlands, CA, pp. 1-17.

Hankin, B. D. and Wright, R. A. (1958), Passenger Flow in Subways, *Operational Research Quarterly*, 9(2): 81-88.

Helbing, D. and Molnár, P. (1995), Social Force Model for Pedestrian Dynamics, *Physical Review E*, 51: 4282 – 4286.

HSE (1996), *Railway Safety Principles and Guidance: Part 2 - Section B - Guidance on Stations*, Health & Safety Executive, London.

Kollman, K., Miller, J. H. and Page, S. E. (1992), Adaptive Parties in Spatial Elections, *The American Political Science Review*, 86(4): 929-937.

Krawczyk, K., DeWine, W. and Yuen, D. A. (2003), Nonlinear Development of Bacterial Colony Modelled with Cellular Automata and Agent Objects, *International Journal of Modern Physics C*, 14(10): 1385-1404.

Langran, G. (1992), *Time in Geographic Information Systems*, Taylor and Francis, London.

Laurie, A. J. and Jaggi, N. K. (2003), Role of 'Vision' in Neighbourhood Racial Segregation: A Variant of the Schelling Segregation Model, *Urban Studies*, 40(13): 2687–2704.

Law, A. M. and Kelton, D. (1991), *Simulation Modelling and Analysis* (2nd Edition), McGraw-Hill, New York.

Lee, D. B. (1973), Requiem for Large-Scale Models, *Journal of the American Institute of Planners*, (39): 163-178.

Longley, P. A., Goodchild, M. F., Maguire, D. J. and Rind, D. W. (2005), *Geographical Information Systems and Science*, John Wiley & Sons, New York.

Maguire, D. J. (2005), Towards a GIS Platform for Spatial Analysis and Modelling, in Maguire, D. J., Batty, M. and Goodchild M. F. (eds.), *GIS, Spatial Analysis and Modelling*, ESRI Press, Redlands, California, CA, pp. 19-39.

Mandelbrot, B. B (1983), *The Fractal Geometry of Nature*, W.H. Freeman, New York.

Martin, L. L. (1999), The Contributions of Rational Choice: A Defence of Pluralism, *International Security,* 24(2): 74-83.

North, M., Macal, C. M. and Vos, J. R. (2004), Terrorist Organization Modelling, *North American Association for Computational Social and Organizational Science Conference 2004*, Pittsburgh, PA.

O'Sullivan, D., MacGill, J. and Yu, C. (2003), Agent-Based Residential Segregation: A Hierarchically Structured Spatial Model, *Proceedings of Agent 2003 Conference on Challenges in Social Simulation*, The University of Chicago, Chicago, IL, pp. 493-507.

Omer, I. (2005), How Ethnicity Influences Residential Distributions: An Agent-Based Simulation, *Environment and Planning B: Planning and Design*, 32(5): 657-672.

Parker, D. C. (2005), Integration of Geographic Information Systems and Agent-based Models of Land Use: Challenges and Prospects, in Maguire, D. J., Batty, M. and Goodchild M. F. (eds.), *GIS, Spatial Analysis and Modelling*, ESRI Press, Redlands, CA, pp. 403-422.

Parker, D. C., Berger, T. and Manson, S. M. (2002), Agent-Based Models of Land-Use/Land-Cover Change: *Report and Review of an International Workshop*, Bloomington, IN, Available at http://www.indiana.edu/~act/focus1/FinalABM11.7.02.pdf

Peuquet, D. J. (2005), Time in GIS and Geographical Databases, in Longley, P. A., Goodchild, M. F., Maguire, D. J. and Rhind, D. W. (eds.), *Geographical Information Systems: Principles, Techniques, Management and Applications* (Abridged Edition), John Wiley & Sons, New York.

Predtechenskii, V. M. and Milinskii, A. I. (1978), *Planning for Foot Traffic Flow in Buildings*, Amerind, New Delhi.

Railsback, S. F., Lytinen, S. L. and Jackson, S. K. (in press), Agent-Based Simulation Platforms: Review and Development Recommendations, *Simulation*, forthcoming.

Schelling, T. C. (1971), Dynamic Models of Segregation, *Journal of Mathematical Sociology*, 1: 143-186.

Schelling, T. C. (2006), Some Fun, Thirty-Five Years Ago, in Tesfatsion, L. and Judd, K. L. (ed.) *Handbook of Computational Economics*, Elsevier, Amsterdam, Netherlands, pp. 1639-1644.

Tobias, R. and Hofmann, C. (2004), Evaluation of Free Java-libraries for Social-Scientific Agent Based Simulation, *Journal of Artificial Societies and Social Simulation*, 7(1).

Topa, G. (2001), Social Interactions, Local Spillovers and Unemployment, *Review of Economic Studies*, 68: 261-295.

Webster, F. V., Bly, P. H. and Paulley, N. J. (eds.) (1988), *Urban Land-Use and Transport Interaction: Policies and Models*, Report of the International Study Group on Land-Use/Transport Interaction (ISGLUTI), Avebury, Aldershot, UK

Westervelt, J. D. (2002), Geographic Information Systems and Agent-based Modelling, in Gimblett, H.R. (ed.) *Integrating Geographic Information Systems and Agent-Based Modelling Techniques for Simulating Social and Ecological Processes*, Oxford University Press, Oxford, UK, pp. 83-104.

7. Acknowledgements

We gratefully acknowledge the Economic and Social Research Council (ESRC) (http://www.esrc.ac.uk) for their support through the Collaborative Awards in Science and Engineering (CASE) studentships (PTA-033-2004-00034 and PTA-033-2003-00008) and our sponsors Camden Primary Care Trust, and the Greater London Authority Economics Unit respectively. NERC, through the Tyndall Centre for Climate Change (http://www.tyndall.ac.uk/), funded the development of the residential location model for Greater London.