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

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

1. Introduction

Cities are constantly changing 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. Problems such as these which involve location and mobility have recently been articulated in much more disaggregate terms then hitherto with components of objects of these systems being conceived of as agents with movement taking place on a backcloth or in an environment composed of points, areas and networks. In particular, automata approaches are being used to grow systems while agent-based modelling (ABM) has become the key way of representing objects of interest (Batty, 2005). The big difference between these new approaches and the more aggregate, static conceptions and representations that they seek to 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 in ways that make them scientifically relevant and policy applicable. We begin by posing five key challenges and then illustrating these with several agent-based models coupled to geographic information systems (GIS): namely i) two models of the residential segregation and the location of residents and businesses within an urban system dimensioned on inner London, ii) a model of pedestrian evacuation from the King's Cross/St. Pancras underground transport hub which is subject to extensive redesign at present and for which emergency evacuation is a key design requirement, and iii) a macro structure for land use transportation in which models of population and employment movement and location are configured using agents.

These models have been designed to address key issues that agent-based modelling faces in general and geospatial agent-based models in particular.

2. The Development of Agent-Based Models

Agent-based models have been developed for a diverse range of applications, such 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 (Krzysztof et al., 2005); 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); social networks of terrorist groups (North et al., 2004), to name but a few. These examples can be constructed as lying on a continuum, from minimalist models for academic research based upon idealised assumptions, 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

Despite the many advantages of as a tool for simulation which enable the representation of the micro-diversity of systems of interest, their emergent properties and their process dynamics (see Castle and Crooks, 2006), Agent-based models did not begin to feature prominently in GI science research 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 five key challenges to the development of such models, which make their applicability somewhat different to the previous applications of models to spatial and urban systems. We outline these challenges in the next section and demonstrate how we are handling them in our own applications which we will present in the final section.

3. Key Challenges

The structure of a typical ABM composed of agents which interact with each other and with their environment are 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 and the repercussions observed over the course of multiple simulation runs. The ability to simulate individual actions of many diverse agents and measure the resulting system behaviour and outcomes over time (e.g. changes in patterns of pedestrian emergency egress), 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). The five challenges that we see as important to their development involve the following: the purpose of the model, the dependence of the model on theory, the representation of agents and their dynamics, calibration, validation and verification of the model against theory and data, and the development of operational models through software. We do not consider this to be an exhaustive list but it is a beginning.

3.1 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. Frequently in ABM, the actual purpose and position in this 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). The purpose of agent-based models range from explanatory to predictive (see Castle and Crooks, 2006) with prescription and design models based on agents of increasing importance.

3.2 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 such as agent-based 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. However our concern here 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 arise in an ad hoc way. In short, the scientific standards of the past are often buried in ad hoc model development.

3.3 Agent Representation, Aggregation and Dynamics

In spatial systems, what constitutes an agent is a critical issue in that the term can 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. We illustrate some issues of representation in the examples we deal with later as shown in Figure 1.

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 generic agents such as 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. As we aggregate, it is more and more difficult to define relevant processes as these too are aggregations of lower level routine and behaviours.

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 interaction and networks, the size of the computation usually rises as the square of the number of agents. 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 level of 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 tested as data does not exist. This reflects the issues about validation and calibration noted below as our fourth challenge.

3.4 Calibration, Verification and Validation

Calibration involves fine-tuning the model to a particular context and this means establishing a unique set of parameters that dimensions the model to the local context. This is not validation 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.

Once developed, verification relates to testing the logic of the model through its computer programme. This involves checking that the model behaves as expected; this is often referred to as 'inner validity' of the model (Brown, 2006; Axelrod, in press). Validation relates to the extent that it adequately represents the system being modelled (Casti, 1997) and in this sense, this involves the goodness of fit of the model to data. However, the validity of a model should not be thought of as binary (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). Validity can be ascertained by comparing the output of the model with comparable data collected from a real-world system. For example, to understand the output of an agent-based model it is often necessary to evaluate the details of a specific simulation 'history' (Axelrod, in press).

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. 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, it was 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 their entirety against data; they are too rich and data needed to test them too poor (Batty et al., 2006).

3.5 Operational Modelling

In terms of ABM as in other areas of simulation and representation, various generic software have been developed which enables modellers to adapt their problem context to the software in question, implementing their model through high level scripting which the software allows. This opens up models to a wider community of scholars than hitherto but it also forces modelers without the skills or resources to develop their own models from scratch to meet constraints posed by the software. This can be 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 (e.g. Swarm, Repast, NetLogo, OBEUS). 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 toolkits can reduce the burden modellers face programming parts of a simulation that are not content-specific (e.g. a 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 toolkit, 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[1]. We illustrate some of our interfaces using both standard ABM software such as Repast and more native developments in our own examples shown in Figure 2.

The remainder of this section will focus on the more general challenges to creating spatially explicit agent-based models. While GIS is a particularly useful medium for

[1] Castle and Crooks, 2006 provide a comprehensive review of ABM simulation/modelling systems capable of creating geospatial agent-based models

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, progress is clearly being made in the use of disaggregated data (e.g. Benenson *et al.*, 2002). Increased computer power and storage capacity has made individual-level modelling more practical in recent times. An example of which 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 United Kingdom, there is a database on land parcels and associated land-uses (OS MasterMap Address Layer 2®[2]), and road segment data is available (OS MasterMap® Integrated Transport Network™ Layer[3]). Current GIS are capable of encoding these datasets into the foundations of a simulation along with providing 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 for the current generation of computers to deal with when applying this study to the whole of a city rather than just a small area. Thus agent-based models have the potential to suffer from similar limitations of the first generation of urban models developed in the 1960s (Lee, 1973). However this can be overcome by considering what level of abstraction is needed to examine the phenomena of interest (for example, is 'all the detail needed?'). Or a series of smaller models could be created examining specific aspects of the system. Secondly there is the lack of personal data both 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 attained (see Benenson *et al.*, 2002) or synthetic population can be generated through micro-simulation techniques (Birkin et al., 2006).

4. Applications

We are actively researching these challenges facing ABM through our current applications of spatially explicit agent-based models at varying geographical scales, Crooks (2006) explores the importance of space and geometry on the processes of the segregation, and the location of residents and businesses within an urban system. Castle (2006) focuses on fine scale pedestrian evacuation from King's Cross St. Pancras underground station in the event of an emergency incident. Batty is working on embedding agent representations into traditional cross sectional land use transport models where individuals tagged to different areas move through routine trip-making. The

[2] http://www.ordnancesurvey.co.uk/oswebsite/products/osmastermap/layers/addresslayer2/

[3] http://www.ordnancesurvey.co.uk/oswebsite/products/osmastermap/layers/itn/

models show how geospatial agent-based models can be built to operate over different time scales, where in the case of the pedestrian model, time can be considered in units of seconds and minutes while in the land use transport model, movements occur over the hour or day. In the segregation model, time can be seen to operate over months and years. The models use both raster and vector based geometries to represent space, and can be related directly to the 'real' world places. Both highlight how agents interact with their environment and with each other and how aggregate patterns emerge. We give a taste of the applications we will display and demo at the meeting in Figures 1 and 2.

5. Conclusion

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.

6. References

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, Heresy, USA.

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

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, Cowles Foundation: Discussion Paper 1132, Available at http://cowles.econ.yale.edu/P/cd/d11a/d1132.pdf

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

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

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

Benenson, I., Omer, I. and Hanta, 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.

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

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, 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.

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

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, USA.

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, England.

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, IL, USA.

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. and Axtell, R. (1996), *Growing Artificial Societies: Social Science from the Bottom Up*, MIT Press, Cambridge, USA.

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, California.

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

Krzysztof, K., DeWine, W. and Yuen, D.A. (2005), Nonlinear Development of Bacterial Colony Modelled with Cellular Automata & Agent Objects, *International Journal of Modern Physics C*, 14: 1385-1404.

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

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

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 and Sons, USA.

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, USA.

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, USA.

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, USA, 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, USA, 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 and Sons, USA.

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

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.

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, 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.

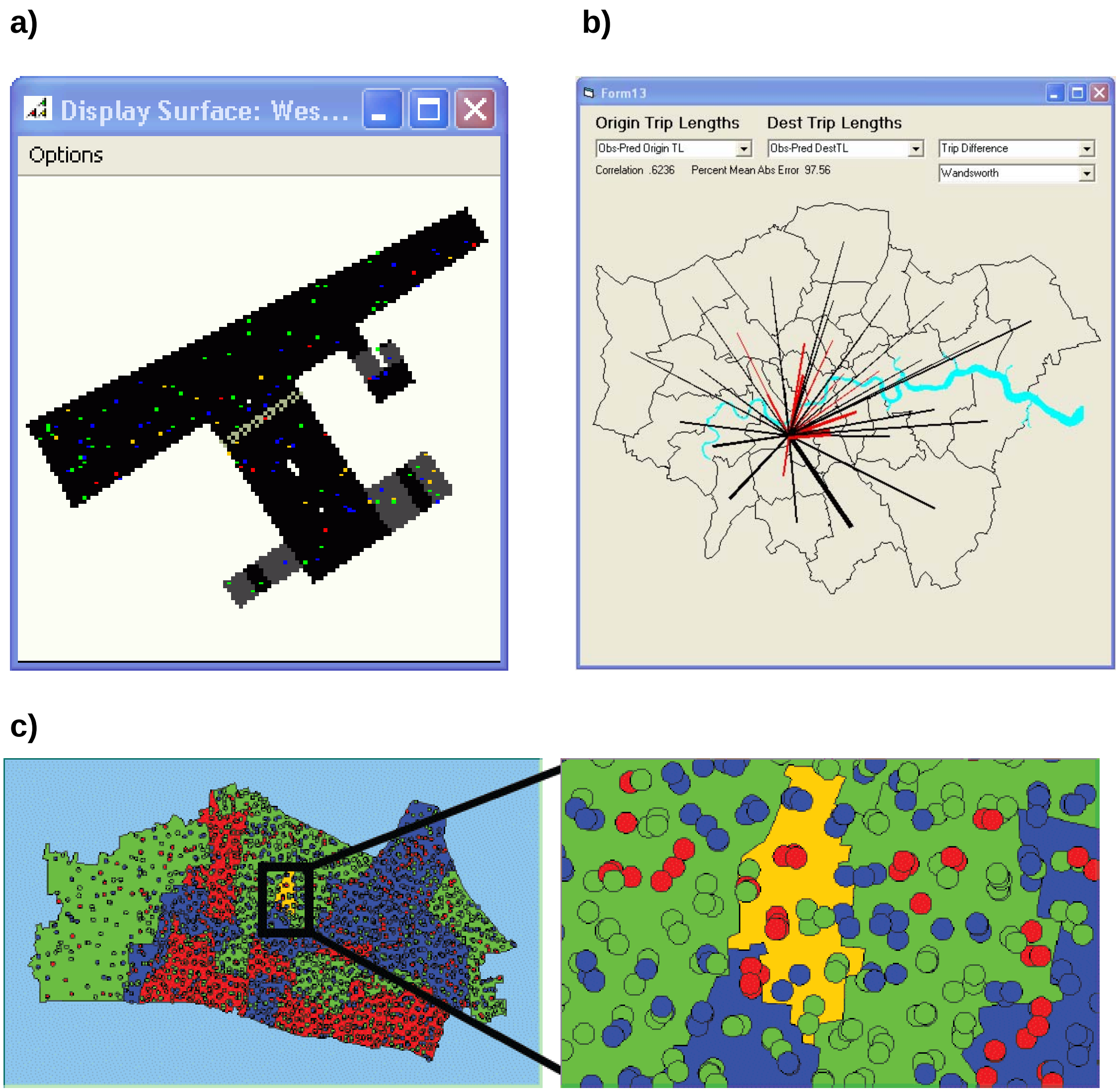

Figure 1: Agent Representation as Typical Model Outputs
a) The King's Cross Pedestrian Evacuation Model, b) the Greater London Transport Agent-Based Model and b) The Inner London Residential Segregation Models

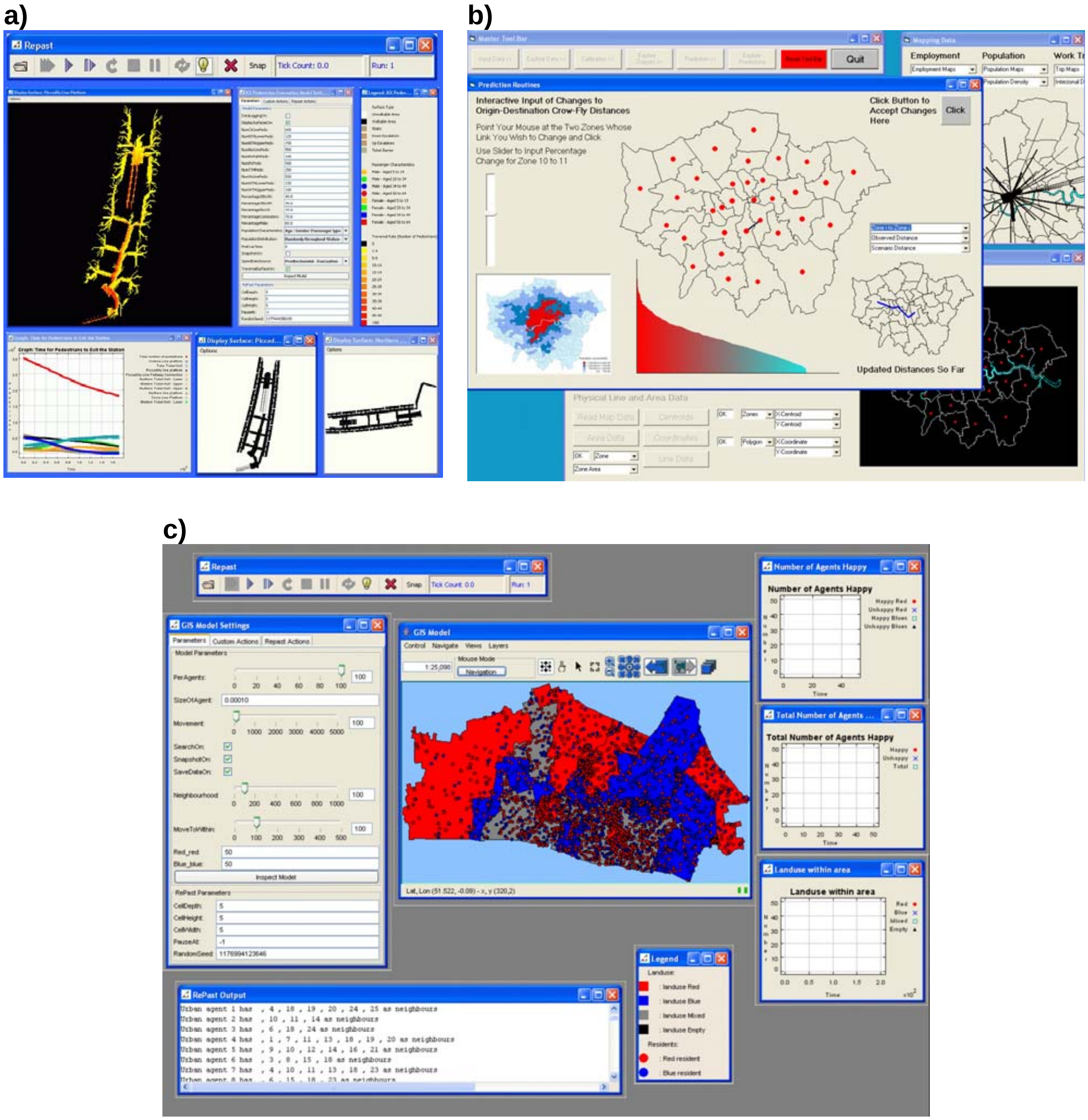

Figure 2: Graphical User Interfaces of the Agent-Based Models
a) The Kings Cross Model, b) The GLA Land Use Transport Model, and c) The Residential Segregation Model