NSF/ESRC Agenda Setting Workshop on Agent-Based Modeling of Complex Spatial Systems: April 14-16, 2007

Evaluating Agent-Based Spatial Models

Michael Batty, CASA, UCL

I am interested but worried about the development of this relatively new class of models which tend to fight against very long standing principles in science that seek for simplicity in abstraction and application. Agent-based models (ABM) have developed for spatial systems through advances in computing that enable distinct objects to be defined with respect to their behaviours which in turn suggest that much more highly disaggregate systems can be represented than hitherto. The fact that ABMs are defined with respect to behaviour implicitly means that such models are temporally dynamic and this too increases the richness of these models in terms of the variability of the systems that are being simulated. Spatial data too is being acquired at ever finer spatial and temporal scales and is making possible the development of richer models structures of which ABMs are typical.

I consider that the workshop should spend time discussing various types of ABMs, defining different types at different levels of object specification, scale, and spatio-temporal disaggregation. Not all ABMs are equated with large scale data requirements that are hard to meet and thus there may be model types that do not conflict with the hallowed canons of parsimony that have defined the scientific method and the model-building process hitherto. Moreover we need to consider ABMs that are not designed to mirror a reality in terms of religiously simulating an observed system. Some models are designed for much more general purposes in terms of defining baselines, structuring data, and enabling hypothesis generation, and thus we need to be clear about the purpose for which particular models are being built.

My own view is that most ABMs are being developed for purposes that are similar to those for which models in the social science have traditionally been constructed: for making predictions that inform policy in the future, although there is a distinct subgroup of such models that are focussed on the past and these tend to break this symmetry[1]. In fact, where ABMs do not work very well in terms of their generating predictions that are close to the reality they seek to simulate, the purpose for which they are initially defined tends to change to less ambitious aims. What has clearly happened however is that as new types of models such as ABMs have emerged over the last 20 years, the whole process of confronting models with reality through observational data, has been elaborated. Here we will refer to the process of matching the model against reality and theory as 'evaluation' which will include different aspects of this confrontation which are called verification, validation, calibration, confirmation, falsification, prediction, and accreditation.

[1] van der Leeuw, S. E. (2005) Why model? **Cybernetics and Systems**, **35**, 117-128.

Accreditation we can probably dispense with as it involves the extent to which the model comes from credible sources, although there is always the prospect of models emerging from outside the scientific establishment.

If a model is designed to predict some reality, then the only unique test that is acceptable is one in which the model's predictions are compared with observations that are entirely independent of the information used in constructing the model in the first place[2]. If a model is fitted to an existing set of data which contains variables that are defined as inputs and outputs with the model being based on associations between these inputs and outputs, the fact that the model predicts the outputs correctly from the inputs is not an acceptable test of the model's quality as the data used is not independent. The only acceptable test would be for the model to be used on another set of data which is judged to be independent of the data set in question. In terms of spatio-temporal models, this data set should be at a different time in a different space. Even models that are fitted to a space at time *t* and generate predictions for the same space at time *t+1* are unacceptable for the data is not completely independent. In fact in spatio-temporal systems, this is an extremely hard criterion to follow and it can damn entire classes of model in terms of determining their predictive abilities. Although this is often confused with verifying a model, we call this process of testing a model's predictions against an independent set of observations validation. The process of fitting a model to the data set around which is has been designed is in fact called calibration which is akin to statistical estimation in that this is the way the unknown parameters are determined which fine tune the model to a real situation. In fact, the process of validation and calibration might be one and the same in that what this implies is that at least two sets of data are needed, one on which to estimate the model and one on which to gauge its predictions from these estimations. Sometimes a single data set is partitioned into two, with estimation occurring on one and validation on the other. Sometimes it is said that the model is trained to the first subset so that when trained it can be used to predict the second subset of data.

The process of verification is defined here as one in which the model's assumptions and construction is checked for consistency and plausibility. A good definition by Miser and Quade[3] is "…the process by which the analyst assures himself and others that the actual model constructed is indeed the one he intended to build". They also contrast with the definitions of validation as "…the process by which the analyst assures himself and others that the model is a representation of the phenomenon being modelled that is adequate for the purposes of the study…".In short, verification consists in generating some sense in which the model is consistent with theory, produces logical results, and is

---

[2] Manson, S. M. (2007) Challenges in Evaluating Models of Geographic Complexity, **Environment and Planning B**, **34**; forthcoming. http://www.envplan.com/contents.cgi?journal=B&volume=forthcoming

[3] Miser, H. J. and Quade, E. S. (1988) Validation, In Hugh J. Miser and Edward S. Quade (Editors) **Handbook of Systems Analysis: Craft Issues and Procedural Choices**, North Holland, New York, 527-565; quoted in Hodges, J. S. (1991) Six (or So) Things You Can Do with a Bad Model, **Operations Research**, **39**, 355-365.

appropriate to the problem being explained. These issues also appear during validation but it is possible to verify a model without validating it and vice versa.

In validation which involves comparing model outputs and independent observations, model predictions might be confirmed or falsified. In fact, this process is always ambiguous for there is always uncertainty in the data which confronts the model, and there is always uncertainty in calibration. In short, predictions are never perfect and thus there are important value judgements to be made in the process of validation. For ABMs, the balance of calibration, verification and validation needs to be explored in considerable detail for models which lack parsimony as most ABMs do, contain assumptions and processes that cannot be compared against data, and thus the balance between verification and validation is different from those model processes and outputs which an be so compared with data. In discussion, I hope we will introduce various examples that illustrate these difficulties. From my own work and some of my associates, I will show how models of local movement where actual behaviour can be hypothesised quite plausibly but rarely observed in terms of individual movements defy full validation, and require less than best strategies to be designed for their evaluation.[4]

Many of the models that we will discuss at this workshop cannot be validated in the traditional sense and this requires us to be very clear as to the purposes for which they are constructed. Hodges and Dewar[5] define several distinct reasons for building models that cannot be validated and these revolve around notions of using model for 'book-keeping', for selling an idea, for training, for management, for communication, for strengthening arguments, and for generating new theory and hypotheses. I am particularly intrigued by the class of agent based models that deal with disaster management such as those dealing with crowding where the models are potentially near to full validation but the circumstances of such validation may not be repeatable – i.e. disasters – and are certainly not ethically desirable to be repeated. These events are also conditioned by extreme control and this makes their observation unstable through time. Helbing's work[6] on the disasters at the Hajj, the Pilgrimage to Makkah, count amongst these examples. I would be fascinated in what the balance of opinion on these questions is amongst our expert group as I imagine that this will depend as much in disciplinary perspective and on the extent to which we consider our models as being central to policy-making.

---

[4] Batty, M., Desyllas, J., and Duxbury, E. (2003) Safety in Numbers? Modelling Crowds and Designing Control for the Notting Hill Carnival, **Urban Studies**, **40**, 1573-1590.

[5] Hodges, J. S., and Dewar, J. A. (1992) Is It You or Your Model Talking? A Framework for Model Validation, RAND Corporation, **R-4114**, Santa Monica, CA.

[6] Helbing, D., Johansson, A., and Al-Abideen, H. Z. (2007) The Dynamics of Crowd Disasters; http://arxiv.org/pdf/physics/0701203; and http://www.trafficforum.org/crowdturbulence