Environment and Planning A 2013, volume 45, pages 532 – 552

doi:10.1068/a44153

Visually-Driven Urban Simulation: exploring fast and slow change in residential location

Michael Batty
Centre for Advanced Spatial Analysis (CASA), University College London,
90 Tottenham Court Road, London W1T 4TJ, England; e-mail: m.batty@ucl.ac.uk
Received 22 March 2011; in revised form 5 May 2012

Abstract. A large-scale residential-location model of the Greater London region is being developed in which all stages of the model-building process—from data input, analysis through calibration to prediction—are rapid to execute and accessible in a visual and immediate fashion. The model is structured to distribute trips across competing modes of transport from employment to population locations. It is cast in an entropy-maximising framework which has been extended to measure actual components of energy—travel costs, free energy, and unusable energy (entropy itself)—and these provide indicators for examining future scenarios based on changing the costs of travel in the metro region. Although the model is comparatively static, we interpret its predictions in terms of fast and slow processes—'fast' relating to changes in transport modes, and 'slow' relating to changes in location. After developing and explaining the model using appropriate visual analytics, a scenario in which road-travel costs double is tested: this shows that mode switching is considerably more significant than shifts in location—which are minimal.

Keywords: spatial interaction, residential location, visual analytics, desktop simulation, energy cost, entropy measures, fast and slow dynamics, scenario testing

Introduction: the logic of simulation

Urban simulation models appeared in the mid-1950s as computers were first used for large-scale transactions processing in business and government. These models were developed against a background of belief that good and accurate predictions could be made for systems as complex as cities in terms of the impacts of urban growth and new transportation infrastructure. Many salutary lessons can be drawn from this experience. In short, the early models were judged to be either 'too simple', or 'not simple enough' (Brewer, 1973; Lee, 1973) and so began a long period of reflection, extension, and reworking of model structures in the quest to make such models more applicable and relevant to policy making. Although progress has been made, many problems remain (Timmermans, 2006).

Two key themes have dominated model development since then. Urban models by their nature tend to treat the city system 'comprehensively' and there has been a long line of models based on ever more detailed urban representations and functions. This has involved the representation of markets which balance the simulated demand and supply of urban activities at ever finer scales of disaggregation. Such models now fashion explicit links to quasi-independent transportation models or embed these models directly within their structure. They still remain largely cross-sectional and equilibrium seeking, but tend to simulate urban change between two or more points in time, often using combinations of model types, ranging from spatial interaction to microsimulation. These models still appeal to the tradition of being 'large scale' in that they are complicated to set up, take time to run, and are the product of teams of analysts rather than the work of individuals. Their continued development revolves around the fact that policy makers demand the kind of numerical detail that such models are able to supply. Good reviews of the state of the art are provided by Timmermans (2006), Hunt et al (2005), and Iacono et al (2008).

In contrast to this tradition of making simple models more complex, there has been a less organised quest to develop models that are simpler than their predecessors. This has proceeded by decoupling submodels and developing these in more detail or by fashioning their different elements into individual models that are used as elements in a toolbox of techniques. Many planning-support systems are constructed in this fashion (see Brail, 2008), although the quest to develop simpler models is dwarfed by the wider trend of extending models to embrace new information technologies and richer sources of data. Perhaps a clearer way of impressing this difference is to adopt Bankes's (1993) distinction between 'consolidative' and 'exploratory' models (or modelling styles). The consolidative style tends to focus on the models that might ultimately provide accurate or focused predictions, in contrast to exploratory models that will never do this but are used to define salient characteristics and to 'inform' the debate. The large-scale tradition in urban modelling very definitely relies upon the consolidative approach, whereas the notion of using simpler models over and over again accords to the exploratory approach.

The model presented here is clearly in this newer exploratory mode but it originates from the earlier large-scale modelling tradition. It is based on simulating interactions from work to home through a residential-location model, with four transport modes based on road, heavy rail, tube and light rail, and bus networks. It sits squarely in the tradition of aggregate spatial interaction modelling as a singly (origin), semi-destination-constrained model where flows of workers to residential zones across the four competing modal networks determine the population locating in each of the destination zones. In this sense, it is both an interaction and a location model. This version was first developed as one stage in an integrated assessment of climate change for the London region, based on a series of coupled models, beginning with a national–regional input–output model (Hall, 2009) whose employment forecasts were then scaled to small areas of the urban region, feeding the residential location model, the subject of this paper. Predicted residential populations were then reduced to a finer spatial scale using a model reflecting physical constraints on land development, reminiscent of more physically based urban development models in the cellular automata tradition (Batty, 2009). These predictions were tested with respect to the flood risk derived from hydrological models geared to account for sea-level rises in the Thames and its estuary, consistent with 50-year and 100-year forecasts from the UK Climate Impacts Programme (Dawson et al, 2009).

The model is developed in a way that makes its use in assessing the impact of abrupt changes in energy/travel costs immediate, where such immediacy is a major requirement for communicating modelling outcomes to a range of stakeholders who are nonexpert in the particular model design. To this end, the model system is visually driven so that the greatest amount of information about the model and its predictions can be communicated as effectively as possible to diverse audiences. I first define a series of strict criteria that the model must meet. I then examine the implicit dynamics of the model, which is a cross-sectional equilibrium structure, before presenting its derivation using entropy/utility maximisation. This sets up a method for consistently calibrating, validating, and evaluating the model where the focus is on how the model handles energy use in the urban system. All these methods are then embodied in the visually driven interface, presented as a series of shapshots of how the model is implemented. I then examine the impact of abrupt and rapid change in energy costs, which are encapsulated in fast changes in interaction patterns through mode shift and slower changes in location patterns reflecting redistribution of the population. Finally, I evaluate these changes using changes in the energy–entropy balance.

Requirements for urban simulation

Rapid visually driven predictions in dialogue with stakeholders

The kind of predictions that this model must address are posed as the outcomes to 'what if' types of question. These assume either an immediate impact on the system of interest, or a reaction that takes place over an unspecified time; that is for very short or very long time horizons where the outcomes in either case assume that the city system will adjust to some equilibrium state. Accurate predictions are not the goal of this kind of model, for the predictions may take many years to realise in terms of generating a long-term equilibrium. This, in fact, can never occur due to unforeseen changes and adaptations that will take place on the trajectory towards this state. In this sense, the model predictions are designed to inform the debate and engender learning amongst the stakeholders. As Epstein (2008) so cogently argues, this style of model can "discipline the dialogue about options and make unavoidable judgements more considered." This is quite consistent with evaluating predictions for 50 or 100 years which embody significant impacts due to climate change, but the model is also capable of examining much more rapid change and its consequences.

The model must be capable of being used over and over again so that a dialogue can be maintained between model builders and users. This puts an upper limit on the time required to run the model which must be in an environment that generates predictions in a matter of seconds, suggesting desktop or web-based media in which outcomes can be communicated through visual analytics, all of which might be represented in different dimensions and through various animations. These requirements are essential to a diverse community of stakeholders. First, in using the model for integrated assessment by chaining different models together across different spatial and temporal scales, visual media make it easier to communicate model structures and outcomes to other scientists with different disciplinary and professional expertise. Our focus then extends to stakeholders involved in policy making of various sorts—from those trained in cognate professional and scientific disciplines to the informed public-at-large.

Models whose outcomes can be generated quickly and disseminated rapidly must also be capable of being reconfigured to embrace different features of the problem-solving context that become important during analysis. This suggests that these kinds of models should be modular. Although the model structure developed here is relatively simple, without any modularity per se, the manner in which its outcomes can be communicated involve extensive modularity with respect to the toolkit of visual analytics. Modularity is also essential in integrating different model types into sequences of predictions that are coupled over different spatial scales, this again reinforcing the need for a common medium of communication between different models and model builders. Visualisation is by far the most effective way in which to communicate different model outcomes, thus posing additional requirements about the need for rapid and quickly repeatable model runs that can be generated in situ in the presence of relevant scientists, decision makers, and stakeholders. All this implies, fast, simple, visual, and accessible models.

Dynamics and comparative statics: equilibrium in terms of fast and slow change

The argument that cities should be treated as equilibrium structures is based on wide agreement that most cities display a similar generic spatial structure and morphology. Such structures also appear to persist over decades, giving power to Harris's (1970) point that such clear evidence of an equilibrium should provide the prime focus for simulation. Models that do not reflect an explicit dynamics simulate what is observed at a cross section in time. They make the assumption that, whenever a prediction is made, the outcomes from the model reflect the fact that the system will have moved to a new equilibrium within the given time period. In his model for Pittsburgh, Lowry (1964) referred to this as an "instant metropolis" (page 39) and suggested that forecasts made with such models must be seen "as 'quasi-predictions' of the emerging spatial structure" (page iv). In contrast, the alternate view is that cities are forever in disequilibrium and thus simulation must focus not on replicating a static urban structure but on changes to this structure, thus reflecting dynamic processes that unfold in time, and destroying any equilibrium assumed to exist at a more aggregate level. In this respect, dynamic models tend to be more complex than cross-sectional ones, in that processes of change are integral to the model design.

Equilibrium models can deal with both short-term and long-term change if the intricate dynamics of the way in which this change works itself out in the city system does not need to be explicit. Very long-term changes, over periods of fifty years or more, as, for example, changes which pertain to climate change, lead to a new equilibrium that is clearly only one from a multitude of future states. The long-term outcome that is predicted is purely notional in that the sheer scale of adaptation that would take place between the current and future prediction dates would be such as to destroy any idea that this outcome would ever take place. In these instances, forecasting with such models simply informs the debate about the long-term future. Very short-term change, however, shows what might happen immediately if the prediction could be borne out assuming no other constraints on the outcome. But this too is unlikely for there are many constraints that only become explicit when an outcome is emerging. Adaptation usually happens even in the very short term and it is often unclear how this works itself out.

There is a distinction between slow, medium, and fast processes of change in urban systems (Wegener et al, 1986). The slowest changes relate to infrastructure, particularly transportation networks and the built environment; less slow changes relate to demographic, economic, and related processes, and the fastest changes to mobility, ranging from local migration to flows on many different scales of network. This continuum can be further elaborated from changes in physical structures, including land use, which are slow, to changes in population and labour markets through migration, which are faster. In terms of spatial interaction–location models, Wilson (2008) identifies fast change in interactions—in this case, the journey to work, which is a diurnal cycle—in contrast to population change in terms of the supply of housing, which is more likely to take place over years. Spatial interaction–location models are usually formulated in cross-sectional terms and, when used in a predictive context, it is assumed that the flows generate change immediately whereas the ultimate locational redistribution takes longer to work itself out. In fact, this process of working out is implicit and the ultimate equilibrium that occurs is a product of both fast and slow processes with no explicit time scale. The assumption is that the predicted outcomes would take place if all other conditions were kept the same, thus representing an ultimate steady state which would only occur under idealised conditions of no other change.

If the model is constructed for simulating changes in demand, then the new equilibrium that results is one that assumes that demand is met with entirely elastic supply. We know that this will never be the case in real systems and this is thus another way in which predictions made with such equilibrium models represent an idealised future state. In most instances, changes in demand will be moderated by supply and the ultimate equilibrium will be composed of a complex process of demand and supply adapting to one another and to other exogenous constraints. It is in this sense, then, that predictions from this model are to be used in wider processes of planning support to inform the debate and to pose immediate answers to 'what if' types of question.

The residential-location model

Specification and derivation using entropy maximising

The model is cast in the most parsimonious form possible, where flows between workplaces (called origins) and residential areas (called destinations) are explained as a function of strictly physical quantities. That is, the variables that we wish to model—flows (or trips)—are measured in terms of persons, explained entirely with respect to physical quantities determined by the size and scale of the system itself. This is phrased in terms of the technological limits on how people are able to interact, which relate ultimately to the geometry of the system, albeit expressed in units of cost of travel, and in terms of the land area associated with these flows. The model is in the tradition of spatial interaction (Wilson, 1970) but will be expressed in an explicit energetic framework.

Flows defined as T_{ij}^k are movements from origin zones $i, 1, 2, \ldots, I$ to destination zones $j, 1, 2, \ldots, J$ with respect to the mode of travel $k, 1, 2, \ldots K$. The numbers of zones is 633 for both origins and destinations, in contrast to a handful of modes, four in all, comprising road, heavy rail, tube and light rail, and bus. The model is derived in terms of the density of trips T_{ij}^k / A_j destined for a particular zone j with residential land area A but expressed it in terms of trip volumes. The model is subject to two physical constraints. The first is based on the total cost of travel by each mode C^k defined as

$$\sum_i \sum_j T_{ij}^k c_{ij}^k = C_i, \tag{1}$$

where c_{ij}^k is the energy expended, measured in terms of travel costs using the modal technology k. The second constraint is on the origin activity, measured as the number of jobs E_i which provides the overall dimensioning of person activities in the system

$$\sum_j \sum_k T_{ij}^k = E_i. \tag{2}$$

The total number of trips, T, is fixed implicitly by equation (2), which can be written explicitly as

$$\sum_i \sum_j \sum_k T_{ij}^k = \sum_i E_i = T. \tag{3}$$

It is worth noting the particular structure of this model. The modal costs in equation (1) are constrained so that each mode is distinct in terms of the energy it uses, whereas this is not the case when the trips are summed across modes with respect to their origins. This implies that the model simulates competition between modes—an essential criterion for handling mode switching. As the basic model is a singly constrained spatial interaction model, besides the flow matrix, the main predictor from the model is activity destined for each residential location, which is working population P_j derived as

$$\sum_i \sum_k T_{ij}^k = P_j. \tag{4}$$

Other volumes might be predicted, such as employment and population by mode at origins and destinations.

To derive the model, the well-established method of defining and maximising the entropy S of the distribution associated with $\{T_{ij}^k\}$ is followed. In fact, a more consistent definition is used for entropy than that of Wilson (1970), which is the discrete approximation to the continuous form, proposed by Batty (1974; 2010), given as

$$\begin{aligned} S &= -\sum_i \sum_j \sum_k T_{ij}^k \ln \frac{T_{ij}^k}{A_j} \\ &= -\sum_i \sum_j \sum_k T_{ij}^k \ln T_{ij}^k + \sum_i \sum_j \sum_k T_{ij}^k \ln A_j , \end{aligned} \tag{5}$$

Next, a maximisation of this entropy is performed by forming a Lagrangian L using equations (1) to (3):

$$\begin{aligned} L = -\sum_i \sum_j \sum_k T_{ij}^k \ln T_{ij}^k + \sum_i \sum_j \sum_k T_{ij}^k \ln A_j \\ + \sum_i \lambda_i \sum_j \sum_k (T_{ij}^k - E_i) + \lambda^k \sum_i \sum_j \sum_k (T_{ij}^k c_{ij}^k - C^k), \end{aligned} \tag{6}$$

which is set equal to zero for the maximisation condition. This leads to

$$\frac{\partial L}{\partial T_{ij}^k} = -\ln T_{ij}^k - 1 + \ln A_j - \lambda_i + \lambda^k c_{ij}^k = 0, \tag{7}$$

from which the model can be easily derived. Note that the -1 term is incorporated in the multipliers λ_i without loss of generality, and new variables are not defined. The model can be stated in log and then in normal form as

$$\left.\begin{aligned} \ln T_{ij}^k &= -\lambda_i + \ln A_j - \lambda^k c_{ij}^k, \\ T_{ij}^k &= \exp(-\lambda_i) A_j \exp(-\lambda^k c_{ij}^k). \end{aligned}\right\} \tag{8}$$

There are two properties worth noting. First, an interesting form can be produced for equation (2) if the model in equation (8) is substituted into this constraint. From this, the value of λ_i is derived as

$$\lambda_i = \ln \left[\frac{\sum_j \sum_k A_j \exp(-\lambda^k c_{ij}^k)}{E_i} \right]. \tag{9}$$

Equation (9) is a log-sum accessibility, a measure which is useful in consumer analysis, particularly related to transportation and, as we shall see, is related directly to the free energy in the system. An equivalent expression, however, cannot easily be produced for the modal cost parameters λ^k. Second, if any two modes are compared from equation (8), then these produce a particularly simple form of competition. Taking the ratio of the relevant model equations for, say, $k = 1$ and $k = 2$, then

$$\frac{T_{ij}^{k=1}}{T_{ij}^{k=2}} = \frac{\exp(-\lambda^{k=1} c_{ij}^{k=1})}{\exp(-\lambda^{k=2} c_{ij}^{k=2})}, \tag{10}$$

implying that modal split is, in logarithmic form, a direct function of the ratio of the relevant costs of travel $\lambda^{k=1} c_{ij}^{k=1} / \lambda^{k=2} c_{ij}^{k=2}$.

In some versions of the model, constraints on destination activities—in short, on population densities—have been imposed and this turns the model from an origin to an origin–semi-destination-constrained model, in which the following constraint is imposed:

$$\sum_i \sum_k T_{ij}^k \leqslant P_j^{\max}, \tag{11}$$

where $P_j^{\max}$ is the maximum residential population allowed in zone j. Only a subset of zones are so constrained, for in many this constraint is purely notional in that there is so much space that it is unlikely that the constraint would be breached. However, in inner and denser areas, equation (11) can be critical. If this is breached, a new parameter, λ_j, must to be introduced to ensure that equation (11) is met and then the model needs to be reformulated, solved, and iterated in a different fashion (Batty, 1976).

What might appear at first sight as a relatively simple residential location model is in fact a good deal more intricate. The use of spatial entropy in equation (5) effectively turns the model into one predicting population density rather than population counts, in that zonal land area appears explicitly in the model in equation (8). Moreover, there is strong competition between transport modes as reflected in the fact that modal split is predicted by the model in equation (10) and this changes the focus from location to distribution. The fact that the predictions can be constrained by a destination constraint, as in equation (11), also enables the model to be 'forced' to meet capacity limits. In summary, this is a very different kind of model from most of those used in previous land-use–transport models, where the formulations tend to mirror the simpler, singly constrained, models. The formulation is being further extended to embody house prices and wages in a successor model, to which readers are referred (Batty et al, 2011).

Calibration, validation, and evaluation

There are several ways of determining the parameter values, all of which revolve around the fact that the two key constraints in equations (1) and (2) must be met. We first write the model in its full form, making explicit the origin constraint as

$$T_{ij}^k = E_i \frac{A_j \exp(-\lambda^k c_{ij}^k)}{\sum_j A_j \sum_k \exp(-\lambda^k c_{ij}^k)}, \tag{12}$$

from which it is easy to see that the modal parameters λ^k can be found, starting with some reasonable estimates, such as $\lambda^k = 1.5/C^k$, and then checking how close the predicted value of these costs are to the observed. The iteration is then performed by changing the values of these parameters with respect to the differences between the predicted and observed costs until convergence. This is akin to solution of the model using maximum likelihood or by actually maximising the entropy equation directly.

There are a number of indicators that can be generated from the model that pertain to the energy used in spatial interaction. First, the entropy from equation (5) is written by substituting equation (8) into the standard form as

$$\begin{aligned} S &= -\sum_i \sum_j \sum_k T_{ij}^k (-\lambda_i + \ln A_j - \lambda^k c_{ij}^k) + \sum_i \sum_j \sum_k T_{ij}^k \ln A_j \\ &= \sum_i \lambda_i E_i + \sum_k \lambda^k C^k \\ &= \ln \sum_j \sum_k A_j \exp(-\lambda^k c_{ij}^k) + \sum_k \lambda^k C^k, \end{aligned} \tag{13}$$

where it is clear that the land-area term cancels from the maximisation, indicating that this is purely for purposes of dimensioning the distribution. Entropy S has a structure that associates it with unusable energy in the system, which is assumed to be equal to the actual energy C less the free energy F. Equation (13) is not quite in this form as the values of the parameters from the maximisation are assumed to be negative and of course in terms of the normalising constraint on origins, this is likely to be positive. Hence equation (13) might be interpreted as $S = -F + C$, from which it is clear that total energy $C = S + F$. Thus usable energy is equal to unusable energy (entropy) plus free energy (Atkins, 1994), linking to more formal analogies between urban structure and statistical thermodynamics (Morphet, 2010; Wilson, 2009).

The real value in thinking in terms of different measures of energy becomes significant when changes in the input variables—specifically costs and employment—are made. If only changes in travel costs are assumed, let us say each travel cost for each mode can change by an increment or decrement Δ_{ij}^k as $c_{ij}^k(2) = c_{ij}^k(1) + \Delta_{ij}^k$, then the change in entropy can be computed as a change between free and actual energy, that is, $\Delta S = \Delta C - \Delta F$. These quantities reflect the scale of the system and if costs are doubled, say, then there can be dramatic adjustment in scale which with redistribution can change these measures substantially.

Using the above definitions, the actual computation can be written as

$$\begin{aligned}\Delta S &= S(2) - S(1) \\ &= \ln\sum_j\sum_k A_j \exp[-\lambda^k c_{ij}^k(2)] - \ln\sum_j\sum_k A_j \exp[-\lambda^k c_{ij}^k(1)] \\ &\quad + \sum_k \lambda^k C^k(2) - \sum_k \lambda^k C^k(1),\end{aligned} \tag{14}$$

where equation (14) can be further simplified to

$$\Delta S = \ln\frac{\sum_j\sum_k A_j \exp(-\lambda^k[c_{ij}^k + \Delta_{ij}^k])}{\sum_j\sum_k A_j \exp(-\lambda^k c_{ij}^k)} + \sum_k \lambda^k \left[\frac{\sum_i\sum_j T_{ij}^k(2)(c_{ij}^k + \Delta_{ij}^k)}{\sum_i\sum_k T_{ij}^k(1)c_{ij}^k}\right]. \tag{15}$$

The free-energy term, which is the first on the right-hand side of equation (15), is reminiscent of consumer surplus and there is a degree of intuitive sense in this derivation. The crucial issue is to examine actual changes in the three overall measures—ΔS entropy, ΔF free energy, and ΔC actual energy—and all these are used in the analysis below to show how changes in travel costs have repercussions on accessibility benefits as well as on the total amount of energy used or distance travelled.

This formulation which expresses energy in two varieties—'unusable' entropy S and 'usable' free energy F—is the usual thermodynamic interpretation (Atkins, 1994). If energy C is increased by adding to transport cost, where it assumed the energy is 'imported' from outside the system, then this will be distributed between usable and unusable, and both may increase because trips are forced onto lower cost modes, but with these lower costs being much greater than the costs incurred on the changed mode. However, it is the relative distribution that is important. If entropy remains the same but energy increases, then this means that free energy increases at the same rate as total energy and the relative distribution of flows does not change. If entropy decreases, then this means that the system becomes more concentrated and thus the different components of this entropy and free energy can be examined as the system changes. When these changes are examined, it must be noted that increases in energy are measured in total trips. If, for example, total costs of energy double, then the average costs will not react in the same way because the model will redistribute individual flows that are associated with these costs. Care is thus needed when interpreting these measures.

The visually driven interface

Principles for visualisation

The general principle which is ascribed to here is to put as much information generated by the model as possible into the display device used to communicate the model's data and predictions as well as its implementation. The display device is essentially the desktop, possibly the desktop linked to the Internet through a web browser. There are three key stages in the model-building process that these models are constructed around: first, exploration of the model's input data; second, the calibration or fine-tuning of the model to these data (as well as their validation and verification); and third, the generation of model outcomes as predictions. In each of these sequential stages, the same graphics tools are used to display information visually in the form of maps, flows, networks, tree diagrams, and so on, as is illustrated below.

The interface is organised as one main window controlling key operations and outputs, and two kinds of toolbar: first, the main toolbar, which strings each stage of the modelling process together from data to prediction; and a second toolbar that is launched when graphical outputs are required. The main toolbar begins with data input, its normalisation, and then data exploration which launches the second toolbar, central to the process of exploratory spatial data analysis. Then the particular model variant can be chosen and this leads immediately to the launch of a window in which the model is fine tuned or calibrated to observed statistics, followed by the second toolbar from which the model outcomes at calibration can be explored visually in terms of their goodness of fit. Finally, predictions with the model can be activated from the main toolbar: scenarios can be either imported from file or constructed on-the-fly—which involves altering locational data on the screen. Once the scenario is built, predictions are generated and once again explored through the second toolbar, which provides similar graphical display capability as at the data-input and calibration stages.

The graphics tools which are accessed through the second toolbar mainly display media in the form of 2D thematic maps, 'desire lines' recording interaction from origins and destinations in proportion to their flow volumes, histograms showing activity volumes, tree maps that display the hierarchy of activities in proportion to their volume in small areas and their aggregates, and scatter graphs of trips and travel costs. All these data can be displayed as either counts or densities, and there are several derived maps that are built from comparing one activity with another in ratio form. To enrich the analysis, all the data produced by the model in map, flow, or histogram form have been enabled to be exported on-the-fly to Google Earth, where they can be compared against network data such as roads and rail lines and a variety of raster-based data such as topographic and climate layers. Data are exported to Google Earth in XML format, which enables KML files to be constructed from the vector data that the model generates. This is accomplished in real time when the model is running, so the user can display data in 3D, flying through it to gain a rich and detailed impression of how the data used and generated by the model compare with other data features of the region input to Google Earth.

Figure 1 illustrates the basic template—the main window and the two toolbars where the window is the panel that displays the data input. From this, the user can explore the data numerically, query the map for the location of zones, and get some sense of the correctness and the dimensionality of the data in context. On launching the model, the splash screen first occupies this window and, once the input has been displayed, a window controlling the

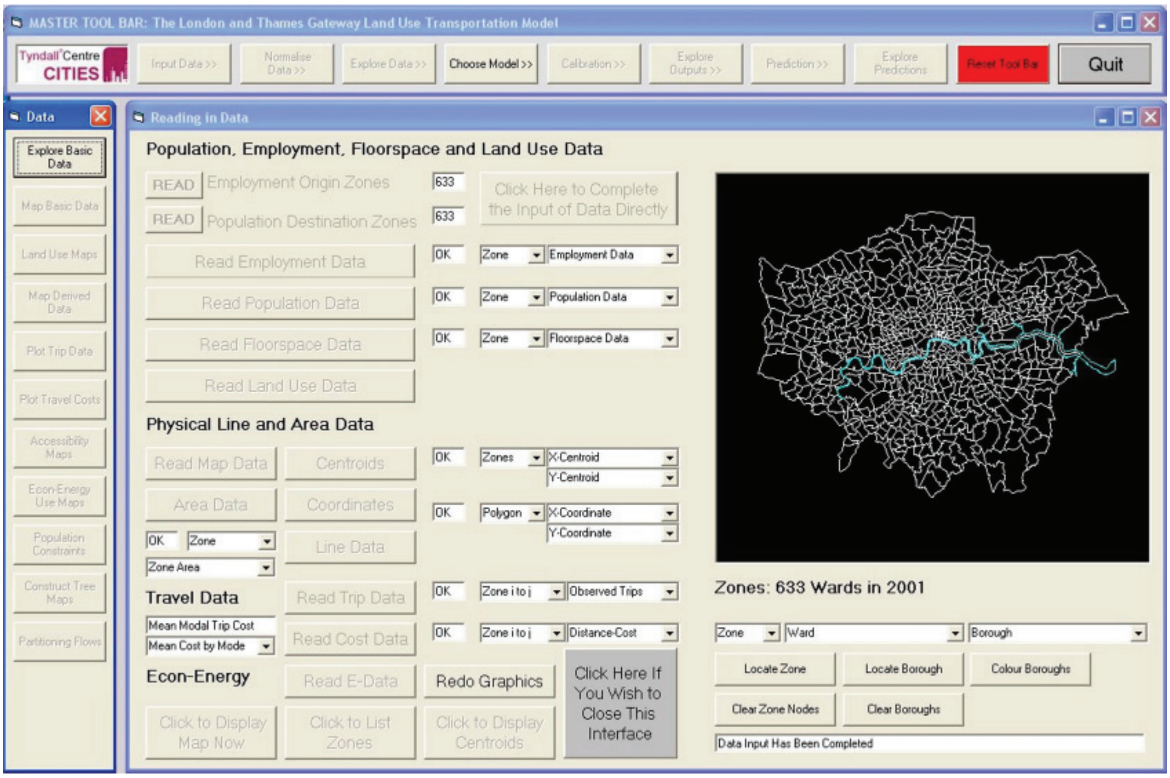

Figure 1. [In colour online.] Windows comprising the basic interactive model template.

normalisation of the model data and thence the choice of the model and its calibration are launched within the template, always leaving the location map to the right onscreen to keep the user orientated. Once calibration is finished, the main window is refreshed to enter the stage of defining and thence generating predictions, which are illustrated in the next section.

Exploratory data analysis, calibration, and the generation of scenarios

The main toolbar drives and directs the user to the sequence of stages defining the model-building process, from data through to prediction, but the second toolbar controls the graphics and is launched at each of the three key stages. I examine the tools that are available at the data-analysis stage, but these in fact are similar to those used for the model's calibrated and scenario predictions at the second and third stages. The toolbar contains ten key display types, seven of which are maps of various kinds: histograms showing activity volumes by location, thematic maps showing the same volumes by area, and flow maps showing different flow volumes from any origin to all destinations or vice versa. These displays can also be queried with respect to individual locations and individual flows in interactive form although, for the most part, the data are completely mapped each time a map is drawn. One key distinction is between counts and densities, so, for example, population (P_j) is plotted as an absolute count compared with its density (P_j/A_j).

The first button enables the user to query individual location and flow data for population, employment, and for trip data (T_{ij}^k) from each origin or destination over any mode of travel k. The second button enables the user to plot these as complete maps, which can then be exported for display within Google Earth. Buttons displaying each land use as a thematic map, and then derived data (eg, activity rates P_j/E_j), follow and then the user can plot trip data (eg, T_{ij}^k/E_iP_j) against travel costs (c_{ij}^k) as scatter graphs. Buttons activating maps as cost surfaces from any origin to all destinations for any mode, followed by detailed accessibility surface maps based on potential and consumer surplus, again for specific origins or destinations by mode, can then be displayed. The final maps reflect wage and house-price data not utilised in the model version reported here, and then population constraints are displayed in terms of land availability. Last, the user can plot various data as tree maps which effectively represent the volumes of activity in each zone tagged to their higher level unit—the borough in this case—displayed as proportional rectangles.

Figure 2 shows a typical collage of these displays including population density, employment counts, road trips from Heathrow Airport, travel costs, accessibilities, and the tree map for residential land area. These maps can be used to learn about the region in terms of its structure which, from figure 2, is clearly monocentric with respect to employment densities and counts. Note how the congestion-charge zone is picked out in terms of the road accessibility from Heathrow, but is not featured in the travel-cost map from the centre of town (Charing Cross) for the tube network. To supplement this visualisation, the input data can be exported to Google Earth, and figure 3 shows how employment counts can be plotted as histograms, population density as a thematic map layer, and the flow data over the road from Heathrow in 3D, successively updating this visualisation as the user continues to generate the data from various displays produced from the model.

The same learning cycle can be initiated with respect to comparing predictions from the calibration to the data using similar map layers, as well as direct comparisons of deviations between observed and predicted activities. Similar displays are available for exploring the impact of various activity and interaction-network changes. However, first the travel-cost data must be normalised and then there is the option of choosing the attractor for the model which is a function of land area A_j. In this model, the raw variable as implied in equations (8) and (12) above is used. The calibration is initiated using an iterative method to ensure that the trip lengths in equation (1) are reproduced, and a damped Newton–Raphson

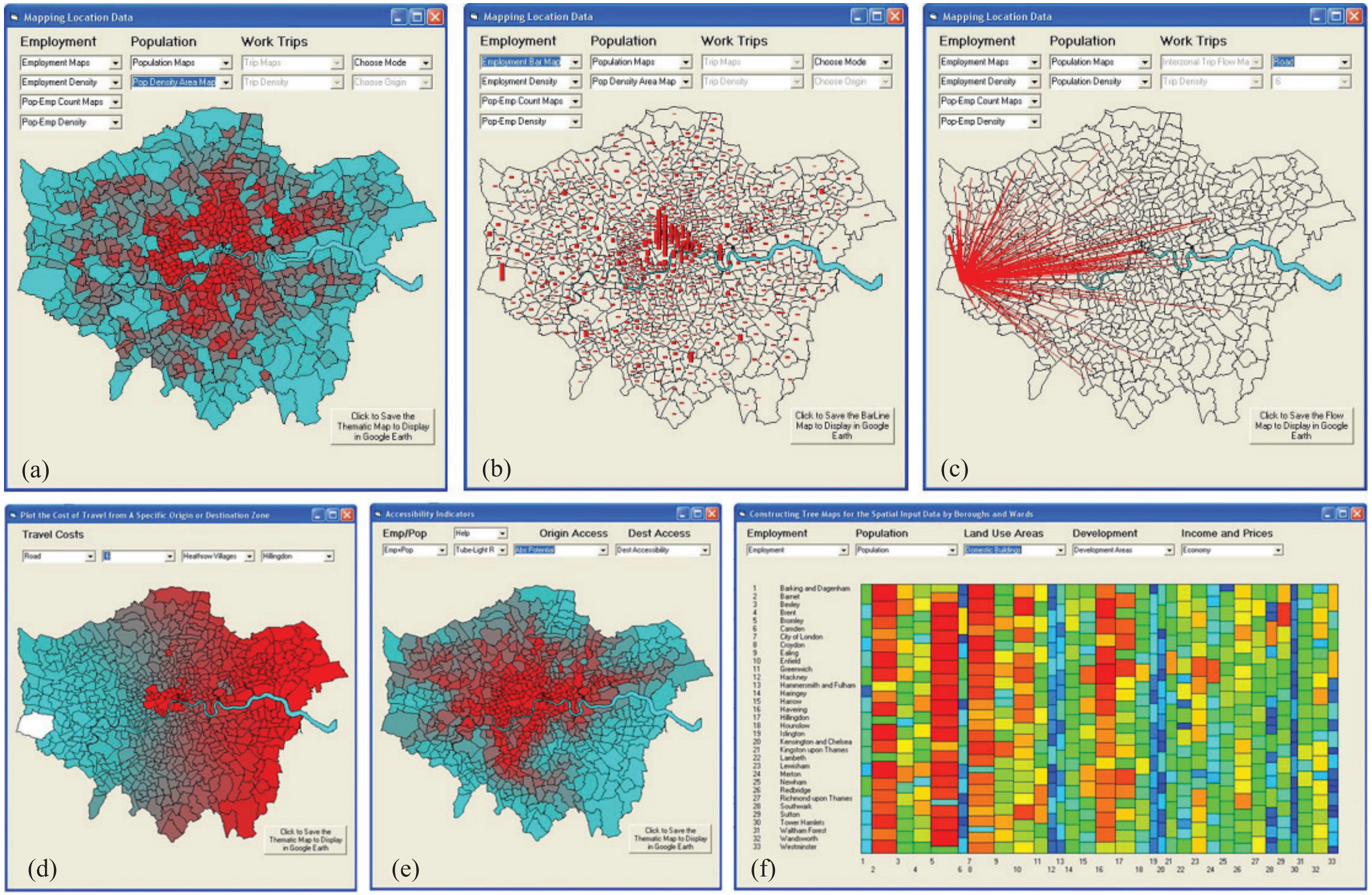

Figure 2. [In colour on line.] Small multiples of graphic output from exploration of the model data: (a) population density; (b) employment counts; (c) road trips from Zone 6 Heathrow; (d) travel costs from Heathrow; (e) accessibility potential on the tube system; (f) tree map showing residential areas at borough and ward levels.

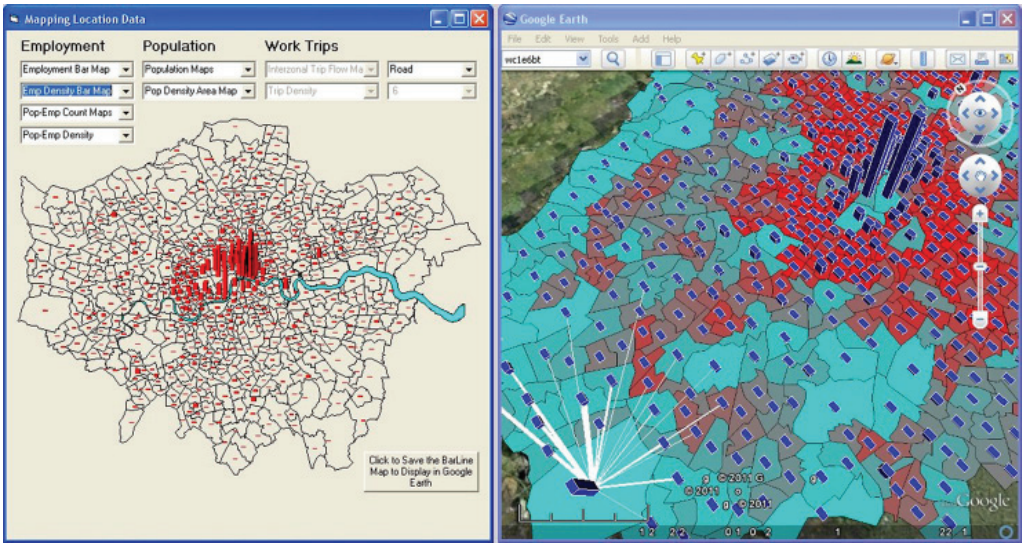

Figure 3. [In colour online.] Visualising thematic map layers, flows, and histograms using Google Earth as an external viewer linked to the desktop interface.

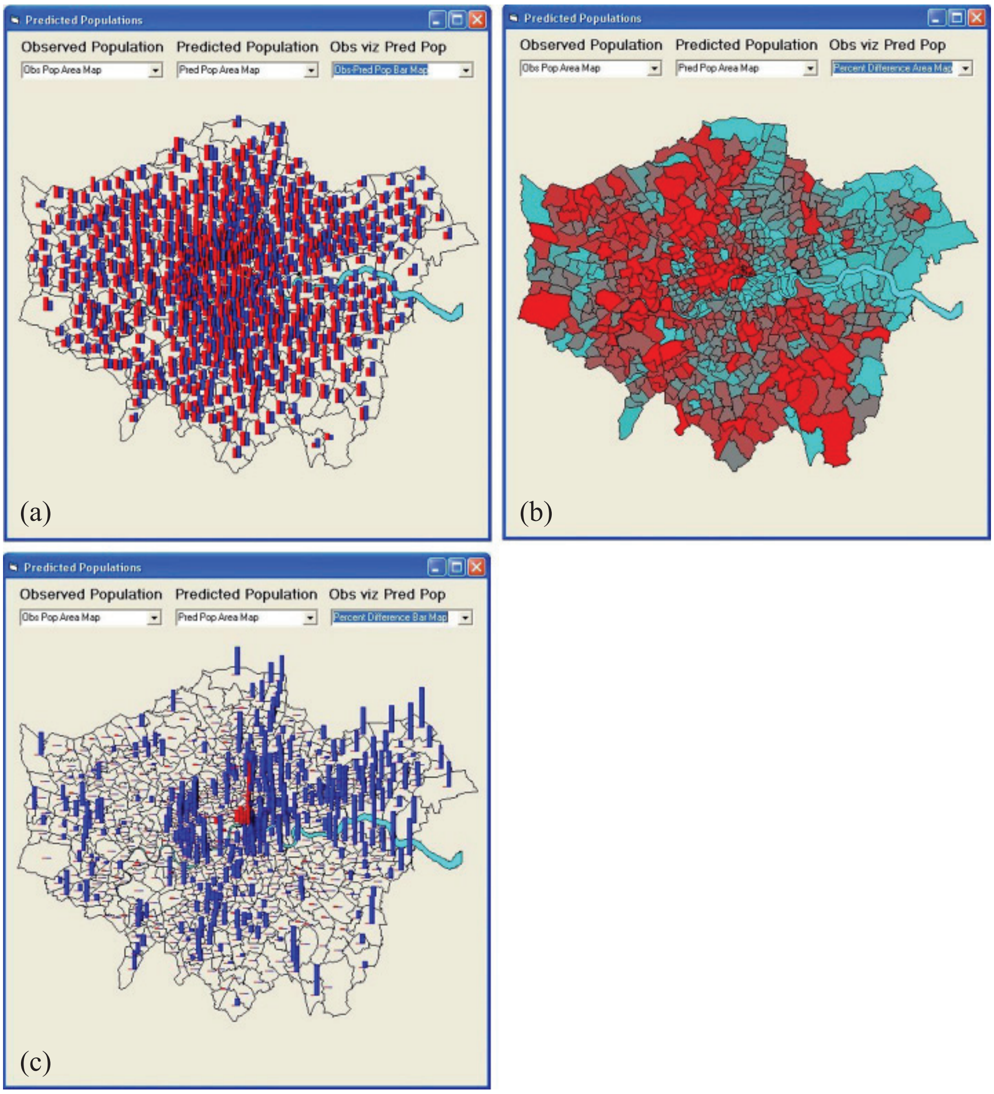

Figure 4. [In colour online.] Predictions of residential population from the model: (a) predicted and observed populations; (b) percentage differences (red—overprediction, blue—underprediction); (c) bar graph of percentage differences (red—overprediction, blue—underprediction).

(hill-climbing) method is used to ensure that this convergence takes place as quickly as possible (see Batty, 1976). The model fit is quite modest in that some 62% of the variation in the population and 43% in terms of the overall trip matrix are explained but, as this paper is just to introduce the framework, in figure 4 the simplest possible unconstrained version of the model is shown, not the variant that generates the best fit.

The last stage in this process involves testing the scenarios which are all framed in terms of changes to the input variables. Users can import new data files which contain these scenarios, or can develop them directly in the visual interface activated in the main window. One such screen is shown in figure 5, which illustrates how the user can add to the network by drawing a new transport route—in this case, a line from west London (Heathrow Airport) to central London (Kings Cross), which, when input, enables the shortest routes to be recalculated for that mode. Changes defining a scenario can be input for locations and modes and, in this way, each scenario to be tested is assembled. It is then possible to generate the relevant predictions and explore these using the second toolbar, which activates the graphics. This provides three possible sets of comparisons between observed data, calibrations, and scenario predictions.

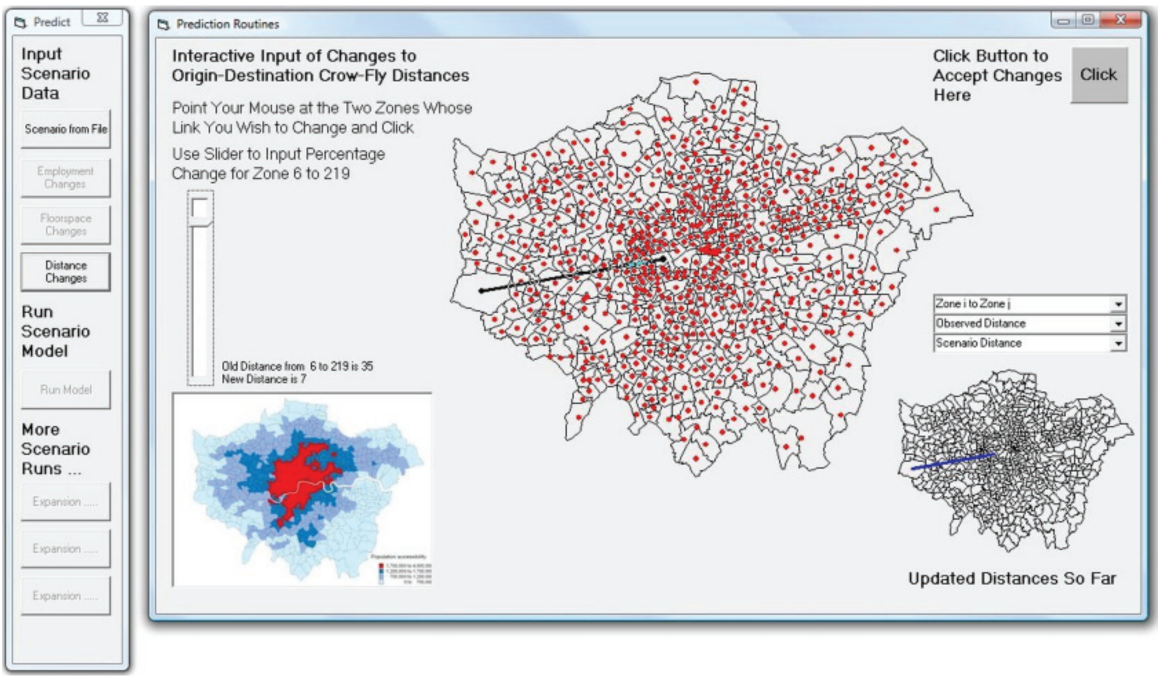

Figure 5. [In colour online.] Building scenarios on-the-fly: inputting a new heavy rail line from the Airport at Heathrow to the West End.

Slow and fast change: the impact of urban energy costs

Although this genus of model essentially simulates a world in equilibrium, using the model to predict future change implies a variety of dynamics that are assumed to work themselves out completely by the time the new equilibrium is established. Lowry's (1964) original idea of the 'instant metropolis' was predicated on the basis that contained within the existing equilibrium were emergent structures that would be revealed when the model was calibrated but would only truly show themselves when predictions to a future state were made. Thus the notion of comparing the calibrated against a future state would be one of comparing the implied equilibrium of the present (not the same as the actual present) with a future equilibrium once new changes embodied in the scenario had worked themselves out. The meaning of these predictions in terms of simulating relevant changes thus turns on the processes that are implicit in the internal dynamics of such models.

There is clearly an array of different dynamics involved in any change in location and trip-making decisions. For trip making, if costs increase, the response is likely to be rapid in that trip makers will switch to lower cost modes. If the capacity of the more cost-effective mode is limited, switching to such a mode might increase congestion, thereby increasing costs to the point where the original mode switch moves into reverse or spills over onto other networks. This process might take time to work itself out, but it is likely to be a lot faster than other types of change. Changing locations clearly takes longer as there can be no immediacy in making a residential move and changes in jobs (which are not part of this model) have a longer dynamics. The real issue is that those effects which are second order, third order, and so on are longer term adaptations that we know little or nothing about because they are often hidden by other changes.

It is possible to differentiate strictly between changes in modal split and changes in location. Changes in both, however, ultimately translate themselves into changes in built infrastructures which tend to be slow in contrast to fast changes that involve people using the same infrastructure but in different ways. In fact, the model was originally designed to examine the impact of very long-term changes in climate—specifically, sea-level rise—on the locational pattern of population over the next 100 years. There will be substantial adaptation to, and indeed mitigation of, these effects over this period which would clearly lead to a future state very different from the equilibrium that this model would predict.

An equilibrium model is, in fact, required so that these other inevitable changes can be filtered out. Such models are classic 'what-if' types of instrument in that they are used to pose and answer questions of the kind 'assuming everything else remains the same and X changes, what is the effect on the system of interest?'. In short, the model can be used to generate the causal chains exposed by this usage. Although the model can define how much change is due to changes in interaction versus location, the balance cannot in fact be attributable to anything other than that the entire model contains many such causal chains. Changes in infrastructure are harder to gauge because individual switches in transport mode, route, and location in response to such changes will clearly take place over much longer intervals. The key example here, in fact, is changes in cost, not infrastructure, and I examine one from many such possibilities. I examine the impact of a doubling of the cost of road transport relative to all other modes. The cost of road travel is increased uniformly over the system, doubling the unit cost and keeping the costs of the other three modes—heavy rail, tube and light rail, and bus—constant. Formally, the change from state (1) to state (2) in transport cost for the car mode $k = 1$ is written as $c_{ij}^{k=1}(2) = 2c_{ij}^{k=1}(1) = c_{ij}^{k=1}(1) + c_{ij}^{k=1}(1)$. If this is substituted into the modal split comparator, equation (10), the relative shift in trips between the changed mode $k = 1$ and any other mode $k \neq 1$ is a simple function of the previous time step, that is

$$\frac{T_{ij}^{k=1}(2)}{T_{ij}^{k\neq 1}(2)} = \frac{T_{ij}^{k=1}(1)}{T_{ij}^{k\neq 1}(1)} \exp[-\lambda^{k=1} c_{ij}^{k=1}(1)], \tag{16}$$

where it is clear that, as the modal cost gets greater, the percentage shift from that mode also increases. This is logical given that trips decline exponentially with travel costs.

If costs are doubled in this manner, the model predicts shifts in all modes as travellers seek to travel on more cost-effective routes and as the model is singly constrained, there will also be shifts with respect to their residential locations. There are two key indicators: first, the total average travel costs, which we would expect to rise for road travel; and second, modal split. These statistics are presented in table 1, where it is clear that the overall average trip costs rise by 17% of which by far the largest component is the increase in road-trip costs—by 27%. Rail and tube only rise between 2% and 3% while the average cost of bus travel drops by slightly less than 2%. These changes are almost the inverse of shifts in modal split, where

Table 1. Changes in average trip costs and modal split.

Mode	Mean trip cost			Modal share		
	observed[a]	predicted	percentage difference	observed	predicted	percentage difference
Road	38.668	49.157	27.124	0.389	0.210	–45.899
Heavy rail	77.780	79.591	2.328	0.122	0.165	34.997
Tube and light rail	59.662	61.196	2.570	0.331	0.400	20.988
Bus	14.659	14.428	–1.576	0.158	0.224	41.926
All modes	47.600	55.621	16.851			

[a]The observed and calibrated mean trip costs and modal shares are the same, as the model is calibrated to meet these constraints.

car ridership decreases by 46%, in contrast to bus transport, which increases by some 42%. Heavy-rail and tube and light rail ridership also increase substantially, by 35% and 21%, respectively. Such big shifts might be expected to be associated with big shifts in residential location activity, which are examined below. In fact, it should be noted that such large shifts would not actually occur: they would require increases in rail infrastructure and a massive extension to the bus fleet. However, they are indicative of the pressures in the system and, in this sense, consistent with the use of equilibrium models.

Changes in trip volumes between the existing and new states (1) and (2) lead directly to changes in activity at residential destinations through equation (4). In difference terms, these changes are

$$\sum_i \sum_k T_{ij}^k(2) - \sum_i \sum_k T_{ij}^k(1) = P_j(2) - P_j(1), \tag{17}$$

where it is clear that the total number of trips is conserved as $\sum_{ijk} T_{ij}^k(2) = \sum_{ijk} T_{ij}^k(1)$. The sum of the differences between residential activities across all locations is zero; that is,

$$\sum_j P_j(2) - \sum_j P_j(1) = 0. \tag{18}$$

In short, changes in costs simply lead to a redistribution of existing activities, and the new equilibrium predicted by the model is composed of these locational changes and the shifts in mode split shown in table 1.

Two rather graphic illustrations of these locational shifts can be computed. First, the absolute proportion of all population moving is computed as

$$\Phi = 100 \frac{\sum_j |P_j(2) - P_j(1)|}{\sum_j P_j(q)}, \tag{19}$$

and the system can also be partitioned at any point into two sets of zones Z_1 and Z_2, where the entire set of zones is $Z = Z_1 \cup Z_2$:

$$\sum_{j \in Z_1} [P_j(2) - P_j(1)] = -\sum_{j \in Z_2} [P_j(2) - P_j(1)] . \tag{20}$$

Equation (20) means that the system can be partitioned into any two sets of zones to examine the flow from one to the other. In this way, it is possible to examine whether the locational constraints are spatially biased towards any locations in the system—specifically, in this case, towards either the inner city or the outer suburbs (Batty et al, 1974).

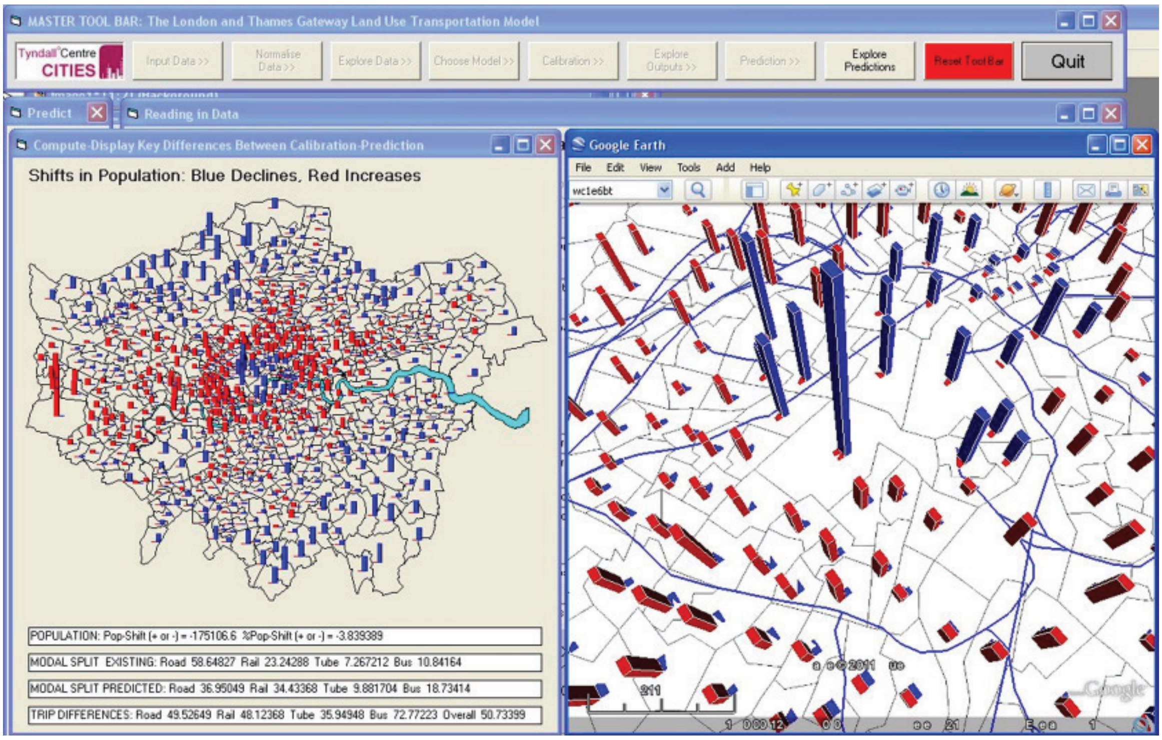

Figure 6. [In colour online.] The impact of a doubling of road travel costs for private car on location (red is increase in population, blue is decrease).

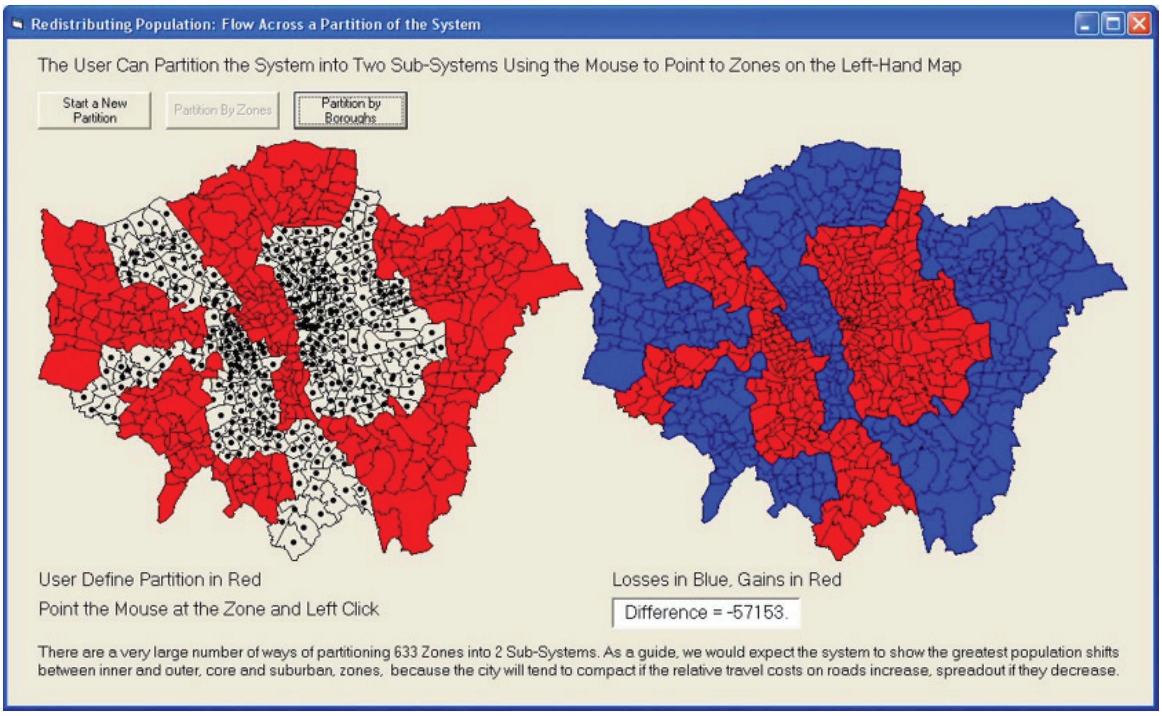

Figure 7. [In colour online.] A noncontiguous partition of the system leading to population relocation.

The most surprising prediction from the model is that the percentage of the working population shifting residential locations, Φ, is only 2.4%, which involves some 110 736 persons. This is extremely low and it is a measure of the resilience of the system to changing transport costs. In fact, overall costs might rise substantially but actual second-order costs due to potential shifts in residential location are likely to be much less than might be expected. To an extent, this result is simply indicative of the fact that there are many more degrees of freedom with respect to potential changes in interactions than in locations. Figure 6 shows the pattern of these shifts in location; it is clear that the city tends to compact slightly with loss of population from many of the suburban areas, with the exception of the relatively vibrant western corridor. This is probably due to the configuration of employment in the west, in and around Heathrow, and the relatively prosperous belt of commuters with good access to transport infrastructure in southwest London. Figure 7 shows a noncontiguous partition of the system, that leads to a high flow of population across these boundaries. The user can choose any partition by clicking on the zones at the fine or broad scale: wards or boroughs—and in terms of wards, there are 633! possible combinations to consider. Clearly, some intuition about the workings of the model and the structure of the spatial system is required to use this tool.

The energy equations introduced earlier also provide a useful if somewhat oblique perspective on these results. The units in which energy and entropy are measured bear no resemblance to the units in which the data for the model are input, which is minutes of travel time. This is because the entropy measures in equations (13) to (15) are computed in terms of total trips, not probabilities, as formulated in traditional versions of these models (Wilson, 1970). Moreover, as equation (13) makes clear, travel costs by mode are normalised by the appropriate travel parameter and then summed to produce a composite cost. However, the relative weighting of these measures gives some sense of how the system changes from the first state (1) to the scenario state (2). All these values are shown in table 2.

Table 2: Changes in entropy, energy, and costs.

Energy value	Calibrated–observed state (1)	Scenario state (2)	Percentage change
Entropy S	4550379	4551015	0.014
Free energy F	9243178	50642320	448
Total costs of trips $\sum_{ijk} \lambda^k C^k$	4692799	46091305	882
road trip costs $\lambda^1 C^1$	3657600	1654163	–55
heavy rail trip costs $\lambda^2 C^2$	3142503	6351686	102
tube and light rail trip costs $\lambda^3 C^3$	1343076	10268510	665
bus trip costs $\lambda^4 C^4$	1100000	32367970	2843

The critical issue is that the total energy in cost terms massively increases due to the external imposition of the 100% change in road costs and this leads to an equivalent increase in free energy while the entropy increases only slightly, which implies a slight decrease in the concentration of all trips across the system. The scale of these increases in energy and its free component is largely due to the fact that when the road costs increase by 100%, there is a dramatic redistribution of trips onto other modes that increases total trip costs on these modes massively. Clearly, the travel costs by mode sum to the total costs and this can be seen in the disaggregations. The disaggregation of the entropy equation by different modes has not been explored because this is distorted by the fact that the free-energy equation cannot be so broken up as it contains the coupling mechanism needed to ensure the model acts as one. However, this is an emerging and active area of research, all the more important because of our current concern for energy costs, in terms of problems of resource depletion and climate change. The substantive interpretations of energy in these entropy-maximising models have remained dormant since their inception some forty years ago. Only recently has there been any effort to ground these concepts in real measurements (Batty, 2010) and the reader is referred to the papers by Wilson (2009) and Morphet (2010).

To complete the picture, I examine the changes in accessibility that are occasioned by this 100% rise in the cost of travel by road. Figure 8 shows changes in accessibility based on computing the standard log-sum term, which is the first component of the entropy in equation (13), and is applied this to the accessibility of origins, meaning that this is accessibility to the location of employment. The change equation for each destination can be restated as follows:

$$\Delta V_j^k = \ln \frac{\sum_j A_j \exp[-\lambda^k c_{ij}^k(2)]}{\sum_j A_j \exp[-\lambda^k c_{ij}^k(1)]}, \tag{21}$$

but it is clear that no changes are communicated from mode to mode through this form of accessibility: for three of the modes, the travel costs are the same for both the *before-(1)* and *after-(2)* states. However, we can examine changes in the road accessibility, and figure 8 shows the before, after, and ratio of these two sets of accessibilities as computed from equation (21) with $k = 1$. These two accessibility surfaces are mapped in rank form; that is, the highest accessibilities are the darkest colour (red) and the lowest the lightest (blue). The surface tends to contract a little between the two states: that is, the surface tends to draw itself close to the centre but a better illustration is the ratio of the two surfaces as shown in figure 8(c). This shows that there is a relative loss of accessibility in the southwest, and the congestion-charge area becomes much more accessible because if costs are increased uniformly across

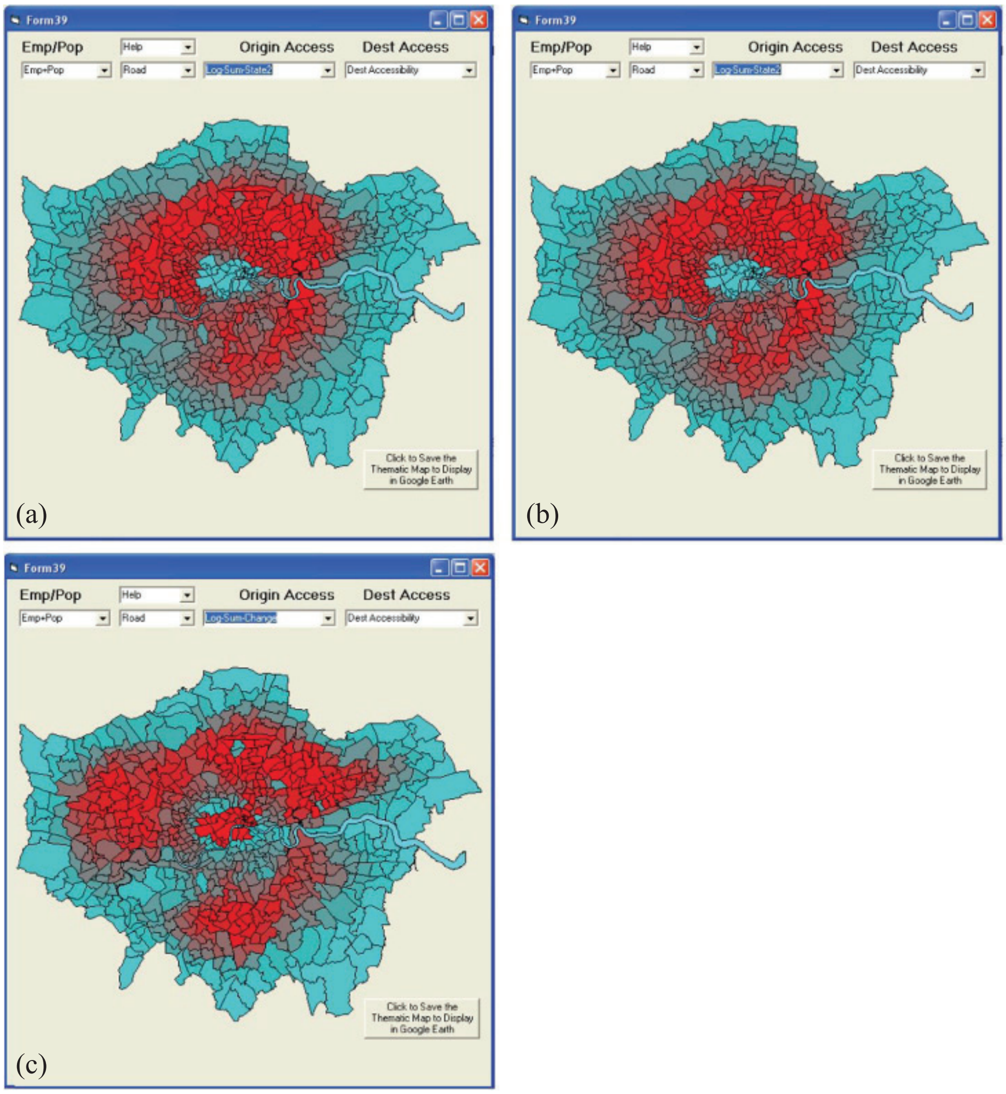

Figure 8. [In colour online.] Before (a), and after (b) accessibility by road, and their ratio (c).

the board, then the higher cost areas (central London) become more attractive even though they still have the highest unit costs.

Next steps in model development

The model is currently being extended in terms of the number of sectors modelled and the number of zones defining the size of the urban region. The criterion for visually accessible, rapid, operation of the model is still a major objective in its development, but to achieve these changes in scale, a slimmed-down desktop version of the model and a much faster web-based version built in state of the art software, using the ECLIPSE Integrated Development Environment (http://www.eclipse.org/) are being developed. The expansion of the residential model to link with retail and local services location models mirrors developments of more integrated models elsewhere (Batty, 2009; Echenique, 2004). These include disaggregation by activity type as well as by mode and will be interfaced with various capacity constraints on location and on the transport network. The new model includes assignment of trips to the various networks and the assessment of related capacity constraints reflecting cost of transport.

The biggest potential change to the structure presented in this paper is in terms of the residential location model. Wegener (2008) argues that each sector modelled in the urban system is likely to be subject to very different economic dynamics. Retailing, he argues, is a process of rapid response to changes in demand and supply which can be seen largely in terms of travel costs and accessibility, while residential location is based much more on the trade-off between house prices and travel costs, which depends on wages. There is a version of the current model in which the constraints on travel costs are replaced with a budget equation, linking incomes (wages) to travel cost, house prices, and expenditures on retailing, and this constitutes the basis for an extended model, built for several sectors and for a much larger region than the (current) metropolitan area. This new model effectively includes an economic model running in parallel to the material model in that trip flows can be interpreted as money flows, balancing the local economy in yet another way (Batty et al, 2011).

When changes to travel costs are examined in the current model, the shift in population locations is of an order of magnitude less than the shifts in modal split. Casual knowledge of urban systems and the ways in which people react to such changing costs suggests that the order of the locational shift—some 2.5% of the working population—is too low. This shift is almost the first order change that might take place although the model makes no such distinctions. We have already experimented a little with the new model, which gives much larger shifts as travel costs are directly related to house prices. If travel costs increase by 100% on road journeys, then shift in population increases some 12%. This is directly due to the fact that, as travel costs increase for road users, they have less to spend on housing and consequently seek cheaper houses. In the extended model, this kind of structure is central to locational decision making and thus the impact of changing costs will be much more realistic. In future applications, we propose to explore many different but related scenarios so that we can examine sensitivity to changing travel costs as well as actual impacts and, as the model can be run rapidly, hundreds of such scenarios will be generated (Batty et al, 2011).

Our last foray into future developments concerns the underpinning of this and similar models using the entropy–energy framework. This needs substantially more effort in making the proper connections and interpretations between city systems and the ways in which energy flows through the spatial fabric. One of the problems in measuring such energies is that the entropy-maximising framework is, for historical reasons, difficult to dimension and the roles of the travel-cost parameters and partition (normalisation) functions need to be worked out consistently for such coupled systems. These are all issues under active development in the context of the extended model, in which the impact of changes in energy will continue to be a central motivation.

Acknowledgements. Steve Evans was instrumental in assembling the locational data and Stuart Barr and Ali Ford developed the transportation-network data and analysis. Richard Milton programmed the interface to Google Earth. This project was part-funded by the Tyndall Centre for Climate Change Phase 2: Cities Research Theme, 2006–2008.

References

Atkins P W 1994 *The 2nd Law: Energy Chaos and Form* (W H Freeman, New York)

Bankes S, 1993, “Exploratory modeling for policy analysis” *Operations Research* **41** 435–449

Batty M, 1974, “Spatial entropy” *Geographical Analysis* **6** 1–31

Batty M, 1976, *Urban Modelling* (Cambridge University Press, Cambridge)

Batty M, 2009, “Urban modeling”, in *International Encyclopaedia of Human Geography* Volume 12, Eds R Kitchin, N Thrift, (Elsevier, Oxford) pp 51–58

Batty M, 2010 “Space, scale and scaling in entropy-maximising” *Geographical Analysis* **42** 395–421

Batty M, Hall P G, Starkie D N M, 1974, “The impact of fares-free public transport upon urban land use and activity patterns” *Transportation Research Forum* **15** 347–353

Batty M, Smith D, Reades J, Johansson A, Serras, J, Vargas-Ruiz C, 2011, “Visually-intelligible land use transportation models for the rapid assessment of urban futures”, WP 163, Centre for Advanced Spatial Analysis, University College London, http://www.bartlett.ucl.ac.uk/casa/pdf/paper163.pdf

Brail R K (Ed.), 2008 *Planning Support Systems for Cities and Regions* (Lincoln Institute of Land Policy, Cambridge, MA)

Brewer G D, 1973 *Politicians Bureaucrats and the Consultant* (Basic Books, New York)

Dawson R, Hall J, Barr S, Batty M, Bristow A, Carney S, Dagoumas A, Evans S, Ford A, Harwatt H, Kohler J, Tight M, Walsh C, Zanni A, 2009, “A blueprint for the integrated assessment of climate change in cities”, in *Green Citynomics* Ed. K Tang (Greenleaf Publishing, Chippenham, Wilts) pp 32–51

Echenique M H, 2004, “Econometric models of land use and transportation”, in *Handbook of Transport Geography and Spatial Systems* Eds D A Hensher, K J Button, K E Haynes, P R Stopher (Pergamon, Amsterdam) pp 185–202

Epstein J M, 2008 “Why model?” *Journal of Artificial Societies and Social Simulation* **11** no 4, 12, http://jasss.soc.surrey.ac.uk/11/4/12.html

Hall J W (Ed.), 2009 *Engineering Cities* Department of Civil and Geomatic Engineering, Newcastle University

Harris B, 1970, “Change and equilibrium in the urban system” *Highway Research Record* number 309, 24–33

Hunt J D, Miller E J, Kriger D S, 2005, “Current operational urban land-use-transport modelling frameworks: a review” *Transport Reviews* **25** 329–376

Iacono I, Levinson D, El-Geneidy A, 2008, “Models of transportation and land use change: a guide to the territory” *Journal of Planning Literature* **22** 323–340

Lee D B, 1973, “Requiem for large-scale models” *Journal of the American Institute of Planners* **39** 163–178

Lowry I S, 1964 *A Model of Metropolis* RM-4035-RC, (The Rand Corporation, Santa Monica, CA)

Morphet R 2010, “Thermodynamic potentials and phase change for transport systems” WP 156, Centre for Advanced Spatial Analysis, University College London, http://www.casa.ucl.ac.uk/working_papers/paper156.pdf

Timmermans H, 2006, “The saga of integrated land use and transport modelling: how many more dreams before we wake up?” in *Moving Through Nets* Ed. K W Axhausen (Elsevier, Oxford) 219–248

Wegener M, 2008, “Multi-scale spatial models: linking macro to micro” paper presented to the Centre for Advanced Spatial Analysis, University College London, http://www.spiekermann-wegener.de/pro/pdf/MW_CASA_090108.pdf

Wegener M, Gnad F, Vannahme M, 1986 “The Time Scale of Urban Change”, in *Advances in Urban Systems Modelling* Eds B Hutchinson, M Batty (North-Holland, Amsterdam) pp 175–197

Wilson A G, 1970 *Entropy in Urban and Regional Modelling* (Pion, London)

Wilson A G, 2008, “Boltzmann, Lotka and Volterra and spatial structural evolution: an integrated methodology for some dynamical systems” *Journal of the Royal Society Interface* **5** 865–871
Wilson A G, 2009, “The ‘thermodynamics’ of the city: evolution and complexity science in urban modelling”, in *Complexity and Spatial Networks* Eds A Reggiani, P Nijkamp (Springer, Berlin) pp 11–31