*Environment and Planning B: Planning and Design* 2013, volume 40, pages 987–1002

doi:10.1068/b4006mb

# SIMULACRA: fast land-use–transportation models for the rapid assessment of urban futures

**Michael Batty, Camilo Vargas**
Centre for Advanced Spatial Analysis (CASA), University College London,
90 Tottenham Court Road, London W1N 6TR, England;
e-mail: m.batty@ucl.ac.uk, camilo.ruiz@ucl.ac.uk

**Duncan Smith¶**
LSE Cities, London School of Economics, Tower 2, 8th Floor, Clement's Inn,
London WC2A 2AE, England; e-mail: d.smith2@lse.ac.uk

**Joan Serras**
Centre for Advanced Spatial Analysis (CASA), University College London,
90 Tottenham Court Road, London W1N 6TR, England; e-mail: j.serras@ucl.ac.uk

**Jon Reades¶**
Department of Geography, King's College London, K7.37 Strand Campus,
London WC2R 2LS, England; e-mail: jonathan.reades@kcl.ac.uk

**Anders Johansson¶**
Department of Civil Engineering, University of Bristol, Queen's Building, University Walk,
Clifton, Bristol BS8 1TR, England; e-mail: A.Johansson@bristol.ac.uk
Received 19 January 2012; in revised form 25 October 2013

**Abstract.** We are building a series of fast, visually accessible, cross-sectional, hence static urban models for large metropolitan areas that will enable us to rapidly test many different scenarios pertaining to both short-term and long-term urban futures. We call this framework SIMULACRA which is a forum for developing many different model variants which can be finely tuned to different problem contexts and future scenarios. The models are multisector, dealing with residential, retail/service, and employment location, are highly disaggregate, and subject to constraints on land availability and transport capacities. They have an explicit urban economic focus around transport costs, incomes, and house prices and thus encapsulate simple market-clearing mechanisms. Here we will briefly outline this class of models, paying particular attention to their structure and the way physical flows and locations are mirrored by economic flows in terms of costs and prices. Several versions of the model now exist, but we will focus, first, on the simplest 'one-window' desktop pilot version with the most obvious graphical interface; and, second, on a much more elaborated framework developed for web access, extensible to web service architectures and other related services. To demonstrate its flexibility and intelligibility, we define the various interfaces and demonstrate how the aggregate model can be calibrated to the wider London region to which it is applied. We will demonstrate the model, albeit briefly with respect to the rapid assessment of different urban futures—"what-if" scenarios, based on the impact of new London airports in the Thames Estuary. The key feature of this entire project is that the model and its variants can be run in a matter of seconds, thus entirely changing the traditional dialogue associated with their use and experimentation.

**Keywords: Land Use Transport Interaction (LUTI) models, economic flows, trip distribution, rapid visualisation, web-based portals, Greater and Outer London**

¶ Honorary Research Fellows, CASA.

**Early history, interfaces, stakeholder requirements**

The original genus of land-use–transportation models emerged in the early 1960s hard on the heels of transportation modelling. These models were relatively large scale in terms of their data and computer requirements and stretched the limits of both of these kinds of resources to the point where many early attempts were never completed. The urban system was also articulated as being in equilibrium with the models being comparative static or equilibrium seeking in the language of those times, notwithstanding the dilemma of simulating systems that were intrinsically dynamic. This tension between statics and dynamics is endemic to the field and has remained at the cutting edge of urban simulation ever since. The field has now bifurcated into two newer traditions: into various theoretical dynamics associated with equilibrium approaches such as those being developed by Wilson (2008), and into more physically based land-development models mirrored around cellular automata and agent-based models (Batty, 2008a; Heppenstall et al, 2012). Their comparative static precursors, however, have not stood still as other papers in this theme issue illustrate. They continue to be the most operational and practical of all urban models, but they have undergone much further disaggregation into activity types with the addition of an incremental dynamics to update their static equilibria. In fact, some of the newer large-scale models such as UrbanSim (Waddell et al, 2003), IRPUD (Wegener, 2014), and the models emanating from Echenique's group (Echenique et al, 2013; Jin et al, 2013) are moving fast towards dynamic agent-based approaches and are now intrinsically wedded to notions of system modularity.

The intrinsic complexity of the earliest models was immediately appreciated and attempts were made even in the very first land-use–transport model developed for the CATS study in the mid-1950s to disseminate this complexity through graphical outputs (Plummer, 1990). SYMAP graphs and maps were popular but, until the advent of desktop computers, there was no immediacy and little interaction between model users and builders other than at the level of the occasional demonstration. Rudimentary interactive graphics was explored in the early 1980s through online access to urban models running on minicomputers which by the late 1980s were simulated on workstations (Batty, 1992), but it was not until the last decade that computers reached the point where truly interactive processing could be developed (Batty, 2008b). During this evolution, our collective view of the role of models in the planning and design of city systems has radically changed. Fifty years ago there was a sense in which both model builders and stakeholders regarded models as providing predictions which could be used with some confidence to help figure out the impact of their plans in rather definite ways with a high degree of certainty. This confidence is now widely regarded as having been misplaced and the role of most models is now to inform, steer, and focus dialogue (Epstein, 2008), notwithstanding the continuing practical plea for some measures of certainty about the future. Interaction between model builders and stakeholders has thus become the name of the game and it is in this context that the models here are being developed.

This modelling effort began in 2007 with the construction of a residential location model for the Greater London Authority (GLA) region which was part of an integrated assessment of climate change, largely sea-level rise over the next 100 years. This model was used to make predictions of future populations by small area which historically have been in many locations with severe flood risk. The model is prefaced by an input–output model which drives employment growth and is tailed by a cellular-automata-like urban development model which distributes population (from ward level) to 50 m grid squares (Dawson et al, 2009). The current model to be presented here extends to a much larger region. It began with the GLA area which was originally modelled for 33 boroughs, moving then to a full model based on 633 zones (wards), and is now extended to 1767 wards covering a region from Reading in the west to Southend in the east and from Luton in the north to Gatwick–Crawley in the south. This is what is referred to as the Outer Metropolitan Area which has a population of about 14 million in comparison with the GLA area that has around 8 million.

Here we will first sketch the framework for these models which we call SIMULACRA, outlining the particular model in formal terms, and then illustrating how the model can be extended to embrace a rudimentary structure for the urban economy. We then focus on the visual interactive interface, showing the desktop pilot and then sketching the full version which runs in a web-based environment. This sets the scene for a summary of the potential applications of the model which can be implemented extremely rapidly as part of a wider process of planning support. The key rationale which we restate here is that these models can be run rapidly on the desktop and within the web portal displaying predictions in a matter of seconds. This enables many different alternative futures to be assessed equally rapidly and the challenge thus shifts from the examination of a handful of scenarios to an iterative process of planning support that enables users and policy makers to 'evolve' different scenarios in ways that enable them to continually improve these possible futures.

## The SIMULACRA models

### Cross-sectional structures and many-sector models

The key difference between our approach to urban modelling and that which has dominated past practice is that we are no longer building a single model, but a framework in which we can generate many different variants of a generic model structure rather quickly. SIMULACRA is our suite of models that are initially static models, simulating more than one sector of the urban system, but at only one cross-section in time, yet having the clear potential for extension to deal with increments of time which are encapsulated within the assumed equilibrium. We call the suite SIMULACRA after Baudrillard (1995 [1981]) who defined the term as "copies of things that no longer have an original". It is implicit in our philosophy of modelling that this is not the real thing but a copy of our own abstractions, perhaps one more stage removed from our older perceptions of what urban models are really about. Here the term is also an acronym that unpacks in several ways, one of which is SIMulating Urban Landuse As Commercial and Residential Activities. The current model deals with four sectors: workplaces defined by employment $E_i$, residential location defined by population $P_j$, shopping defined by retail employment $R_k$, and what we define as local industries generated endogenously in the system that we call internal employment $M_l$.The subscripts $i$, $j$, $k$, $l$ refer to zones of the system that pertain to all $n = 1767$ for each activity.

The model links these four activity types through spatial interactions—the journey from work to home defined by trips $T_{ij}$ linking employment to population, trips $S_{jk}$ from residential areas to shopping centres, and through implicit industrial linkages measured as accessibilities to employment and to commercial activities defined as $A_l$. These activities can be disaggregated in any way and extensions to such classifications are obvious and straightforward. The GLA model that preceded this current one is the first in this suite of models based only on the employment and residential sectors but disaggregated into four travel modes. The current model is disaggregated into different employment and population (household) types, but the modes have been aggregated to two types: public and private. In all cases the activity volumes and their interactions are subject to capacity constraints, but currently these are invoked only for the residential sector. Extending these to deal with trip volumes is a major task in taking the current model forward, but only the most aggregate version will be presented here.

**The simplest** SIMULACRA **model**

We first simulate the flow of work trips from workplace origins $i$ to residential destinations $j$ using a singly constrained spatial interaction model defined as:

$$T_{ij} = E_i \frac{L_j \exp(-\lambda c_{ij})}{\sum_j L_j \exp(-\lambda c_{ij})}, \qquad \text{where} \sum_j T_{ij} = E_i . \tag{1}$$

$L_j$ is the residential land area which acts as an attractor, $c_{ij}$ is the travel cost from origin zone $i$ to destination zone $j$, and $\lambda$ is a fraction of the travel cost (distance) parameter that is approximately equal to twice the inverse of the mean trip cost in the system. The working population at zone $j$ can be predicted by summing equation (1) over zone $i$ and then scaling this by the activity rate $\alpha$ which converts employment into population as:

$$P_j = \alpha \sum_i T_{ij} . \tag{2}$$

Residential population is also subject to a capacity constraint $P_j^{\max}$ which if invoked, that is if $P_j > P_j^{\max}$, leads to a cycling of equations (1) and (2) and the introduction of weights in the residential attractor to ensure that these constraints are met. The residential sector is connected to the retail sector using a similar spatial interaction model which simulates the trips between destinations of the population $j$ and origins of retail employment $k$:

$$S_{jk} = \beta P_j \frac{F_k \exp(-\varphi c_{jk})}{\sum_k F_k \exp(-\varphi c_{jk})}, \qquad \text{where} \sum_k S_{jk} = \beta P_j . \tag{3}$$

$F_k$ is retail floorspace in $k$, $\beta$ is a population-serving ratio which converts population into the demand for retail jobs, and $\varphi$ is a friction of travel cost parameter similar to $\lambda$ with the same interpretation. $S_{jk}$ is in fact the spatial demand for retail jobs rather than shopping trips, but suitable conversion factors can be employed, albeit at a rather crude aggregate level. Retail employment is thus predicted as:

$$R_k = \sum_j S_{jk} . \tag{4}$$

Internal employment $M_l$ which forms the local industry sector is predicted using a rather different type of model, more akin to those developed by Putnam (2007 [1983]) amongst others. This model can be stated as:

$$M_l = MK \left\{ \gamma \frac{O_l}{\sum_z O_z} + (1-\gamma) \frac{A_l}{\sum_z A_z} \right\}, \qquad \text{where} \sum_l M_l = M, \quad 0 \leqslant \gamma \leqslant 1 , \tag{5}$$

and the constant of proportionality $K$ is defined so that total internal employment $M$ is conserved. $O_l$ is the total floorspace associated with all commercial/office employment including retail floorspace and $A_l$ is the accessibility to total floorspace $\Phi_l$ associated with all employment, which in turn is defined as:

$$A_l = \Phi_j \exp(-\eta c_{lj}) . \tag{6}$$

In equations (1)–(6), the parameters $\lambda, \varphi, \gamma, \eta$ have obvious meanings and are estimated to ensure dimensional consistency and maximise the goodness of fit. Note that the model as presented is in its aggregate form, but it is a straightforward matter to disaggregate equations (1)–(6) with respect to any number of classes or types and to add modal split. A version of the GLA model where modal choice is represented explicitly is given in a related paper (Batty, 2013).

The causal chain from total employment to population to retail and then internal employment is the one first developed by Lowry (1964) but there are many ways in which these three submodels—equations (1)–(2), (3)–(4), and (5)–(6)—can be stitched together and balanced through iteration. In fact, the model we are working with solves the equations once in the order given, with predicted population being the driving force for retailing. These equations can also be solved as three separate submodels, but some modest coupling is useful as this ensures that the predictive power of the model is potentially greater than three individual models. Of course, total employment is also predicted from the model as $E_i' = X_i + M_i + R_i$ and this can form the driver for an iterative loop from equation (1) to equation (6). These equations can also be solved in any order and, if constraints are invoked, these can be resolved either through an inner iteration or as part of the outer activities balancing loop. These many possibilities have never been explored in models of this kind, largely because computer resources were not available when these models were first devised and their subsequent history of ever more disaggregation has meant that their aggregate equivalents have not been thoroughly explored. We will not do this here, but it is an essential goal of the wider project. The employment model has been specified in various ways and a full discussion using regression structures to augment equation (5) is detailed by Smith et al, (2013) for a region even wider than that used here. A block diagram of the model structure is presented in figure 1 and this encapsulates the various linkages that are implied in the way the previous set of model equations are coupled and solved. It shows how the various submodels are coupled and how aggregate predictions of population and employment are interfaced with the model as exogenous inputs that drive and control the total activities allocated by the spatial simulation models.

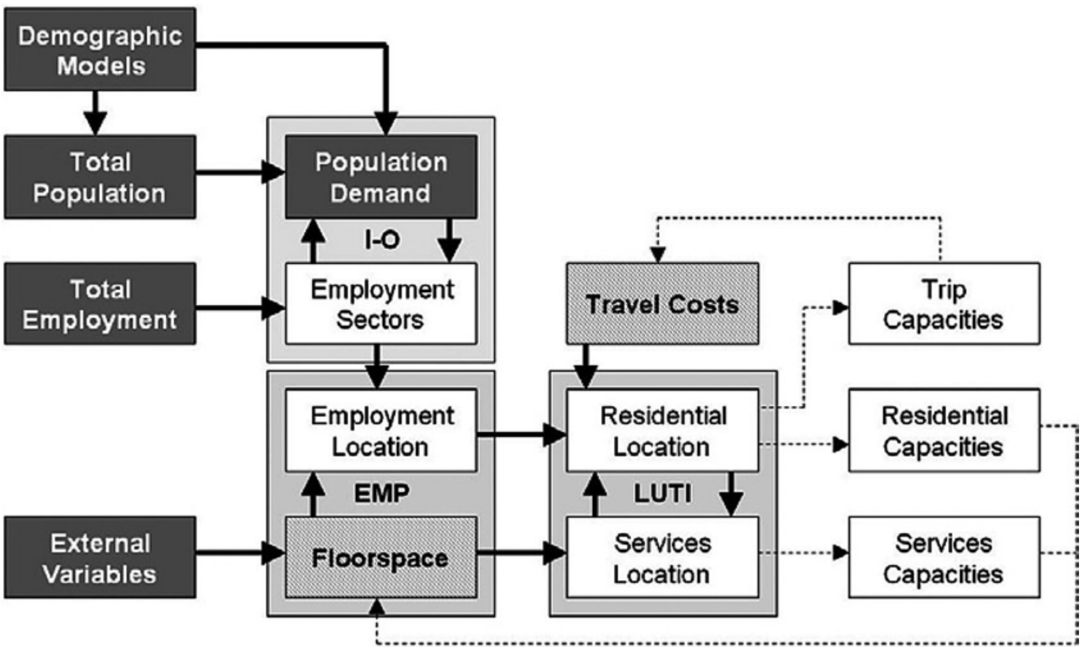



**Figure 1:** The SIMULACRA model structure. Aggregate external totals are coloured dark grey; external zonal variables are coloured light grey; predicted variables are coloured white.

One last point. Two of these three models—the residential and retail location models—are similar in structure. That each simulates a rather different process suggests that they should be sufficiently different to capture the salient elements of their respective subsystems. This we refer to as Wegener's (2008) principle, which is based on his plausible argument that, as retailing is a rapid response activity, the spatial interaction model used here is appropriate whereas the residential location model is not, largely because housing costs and budgets are not factored into the model. The model dealing with internal employment is appropriate in that it contains factors dealing with land supply in terms of floorspace. To meet his critique, we can extend this model to deal with the local urban economy and, to this end, we will digress into a parallel formulation which leads to a more appropriate residential location model.

## Adding the urban economy

### Incomes, house prices, and travel costs

It is possible to consider flows of people in this model as flows of money. Wages are earned at employment locations and money is spent on travel from work to home and then on housing at the residential end of such a trip. Consumer goods are generated by the population, purchased at the retail end of the trip, and consumed at the residential origin. Here our model will not embrace consumer spending on retail goods, but simply consider monies spent on housing which we will integrate into a form which leads to a more appropriate residential location model for which we have the requisite data. We have income per head $y_j$ and also average house price $\rho_j$ at place of residence $j$, but we do not have wages $w_i$ per head or in total $W_i$ at places of employment. Noting, however, that we have flows of workers, trips $T_{ij}$, the money flow from $i$ to $j$ can thus be calculated as $T_{ij}y_j$ where we can work out both total wages $W_i$ and total income $Y_j$ at the household end as:

$$\sum_j T_{ij} y_j = W_i \,, \qquad \text{and} \sum_i T_{ij} y_j = \alpha^{-1} P_j y_j = Y_j \,. \tag{7}$$

It is thus easy to show that:

$$\sum_i W_i = W = \sum_j Y_j = Y \,, \tag{8}$$

where $Y$ is total income in the system, balancing with total wages $W$ at all times.

We need to link wages and incomes to the total spent on transport and housing. At the household end we have good data on the spending in these categories by income $y_j$ and from these we are able to derive excellent regression models that enable us to generate monies spent on housing $h_j$ and monies spent on travel $c_j$. We simply state the regressions with no further explanation as $t_j = -134.81 + 0.388y_j$, and $r^2 = 0.99$, and $h_j = 67.50X + 0.029y_j$, $r^2 = 0.60$, translating to $t_i = -134.81 + 0.3877\left(\sum_j T_{ij} y_j / E_i\right)$, and noting that all these parameters are significant at the 5% level. We can then factor back in the same way and generate monies spent on housing and transport at employment locations $i$. To balance budgets we need to ensure that average house prices $\rho_j$ and travel costs $c_{ij}$ sum to the total monies $\Pi$ and $T$ available and thus we must ensure that:

$$\sum_i E_j h_j = \Pi = \sum_i \sum_j T_{ij} \rho_j = \sum_j \alpha^{-1} P_j h_j \,, \tag{9}$$

and

$$\sum_i E_i t_i = T = \sum_i \sum_j T_{ij} c_{ij} = C = \sum_j \alpha^{-1} P_j t_j \,. \tag{10}$$

We are now in a position to formulate the new model.

### The residential location model

The model is based on the idea that workers have monies to spend on housing and transport which vary according to the wages they receive at their place of work. This conditions the probability of their journeying to some different location to live, the assumption being that the smaller the difference between their available monies for transport and housing and the cost of travel to that place and the cost of housing there, the greater the probability that they will locate there. This replaces the classic negative exponential travel cost function. Strictly we can formulate the constraint associated with travel as a difference or variance $\sigma^2$ between these two sets of costs. Then, the system must satisfy the constraint:

$$\sum_i \sum_j T_{ij} [(h_i + t_i) - (c_{ij} + \rho_j)]^2 = \sigma^2. \tag{11}$$

The model that is generated from this constraint and which is the alternative residential location model in the current model variant, suitably disaggregated by household group and employment type when required, is:

$$T_{ij} = E_i \frac{A_j \exp(-\lambda[(h_i + t_i)] - (c_{ij} + \rho_j)]^2}{\sum_j A_j \exp(-\lambda[(h_i + t_i)] - (c_{ij} + \rho_j)]^2}. \tag{12}$$

This is subject to the usual origin constraints, generating population from equation (2) with equation (12) replacing equation (1). The model is also subject to the density constraint on population capacity at the destination end of the trip that we indicated earlier after equation (2).

We need to note the calibration of the model structure in equations (1)–(6) with either variant of the residential model based on equations (1) or (12). Essentially the parameters $\lambda, \varphi, \eta$ relate to constraints on travel costs or variances which are consistent with their derivation using entropy maximising or maximum likelihood. These can be approximated from continuous equivalents of the submodels, but strictly some iterative scheme is required because the models are coupled and because the parameter $\gamma$ is not related to any formal constraint equation. In this paper we will not explore this calibration process in any detail other than to note it because our focus is more on the framework and its interface to the planning support system that sustains it.

## Visual templates

### The desktop model

Our previous residential location model for the 633 zones comprising the GLA metro region was configured as a visually accessible interactive desktop application (Batty, 2013). The user can interrogate this model at every stage from the initial stage of data exploration and analysis, through calibration, and thence into evaluating predicted impacts on location and interaction as part of a wider set of 'what-if' style scenarios. This model is based on multiple windows being launched in a systematic way through a toolbar sequence that drives the model input–calibration–prediction processes. In this version of the SIMULACRA model we decided to construct a much simpler desktop pilot which essentially is configured within 'one window' which in turn is divided into different frames. This simpler interface makes the model much faster to run and the data input or predictions are much easier to explore and comprehend. We decided to develop this in parallel with a more elaborate version built generically in state-of-the-art software. This will be the main workhorse here as we have disaggregated activities and interactions to specify much greater detail in the population and employment sectors and also we are extending the number of activity sectors that the model will deal with to embrace education, health, and various types of leisure activity

The 'one-window' desktop version is driven from a simple toolbar which contains the key buttons controlling the sequence of stages. The main frame contains a map window while other frames relate to model settings, parameter values, and model outputs. Despite the fact that as much information is packed onto the screen as possible, various numerical outputs are routinely produced for offline analysis. The screen is thus organised in one window with the various frames configured in this window as we show in figure 2. Model functions form the commands that drive the modelling process, while once the data are read in and normalised, activity totals for the region as well as the zonal map are displayed to give the

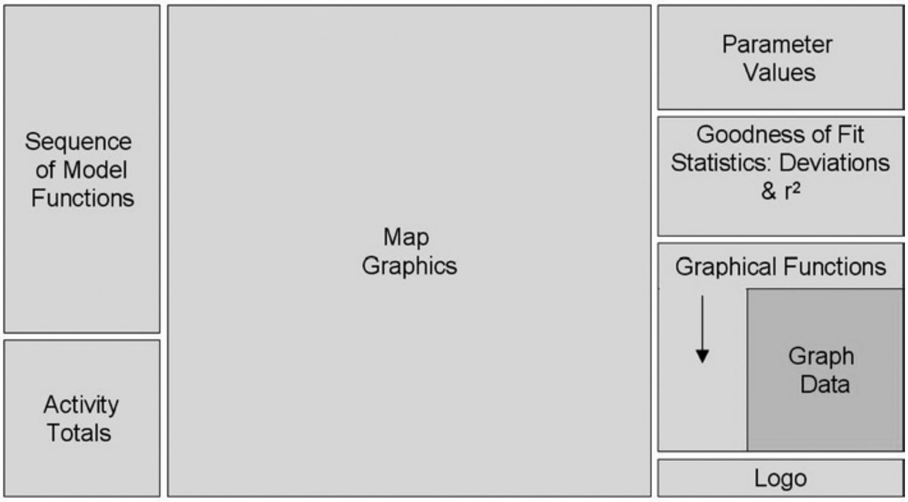



**Figure 2.** Organisation of the 'one-window' desktop pilot interface.

user some immediate sense of the scale of the region. Parameter values for the three models are then fixed using sliders whose default values are $\lambda = 2/\overline{C}, \varphi = \eta = 2/\overline{S}$, and $\gamma = 0.5$, where $\overline{C}$ and $\overline{S}$ and are the mean trip travel costs defined as $\overline{C} = \sum_i \sum_j T_{ij} c_{ij} / \sum_i \sum_j T_{ij}$ and $\overline{S} = \sum_j \sum_k S_{jk} c_{jk} / \sum_j \sum_k S_{jk}$, and $\gamma$ is the parameter moderating the weight of floorspace supply and accessibility in the simulation of internal employment. These can be varied by the user but, once fixed, the model is run, constraints invoked if required, and then the goodness-of-fit statistics are computed. These are displayed in appropriate panels.

The user is then invited to explore the predictions through a series of graphic functions enabling activity count and/or density data to be mapped in thematic or histogram form. Deviations between observed and predicted activities can also be explored in map form, with the user launching each choice of map sequentially within the map window. Figure 3 provides a typical example of the process at the point where the model has been run and goodness of fits produced. The user is also informed of the time taken to run each stage of the model as it completes. The fastest we have been able to run the model in figure 2 is 6 s on a PC 64 Bit 3.07 GHz with the overall process taking some 25 s. This is comparable with running the model under Virtual Fusion on an iMac with 3.2 GHz which gives 7 s and 29 s. On the Vaio VGN-SZ 1.32 GHz laptop used originally to demonstrate the model it takes 24 s for a model run and 90 s overall, but this machine is now at least 7 years old. However for a model with 1767 zones, this is an order of magnitude faster than anything we have come across hitherto, notwithstanding the dearth of experience in running these styles of aggregate model in recent years.

**The web-based model**

The SIMULACRA web-based model has been designed following a multitier architecture (Fowler, 2003) for enterprise applications and is currently in its development phase. Unlike most web applications, this model is intended to be run on the user's own machine, a dedicated server, or a cluster of computers. In the basic design principles of the multitier architecture, the components related to the user interface, the functional process logic, and the data storage are maintained as independent tiers. Through the separation of the components into different tiers, the solution allows any of the modules to be upgraded or replaced independently as requirements change without affecting other modules. The multitier architecture clearly separates the internal logic of the underlying urban models from the user-interface components. This makes the implementation of a range of simulations fast and intuitive and ensures that the user interface contains familiar desktop and web technologies; at the same time, this

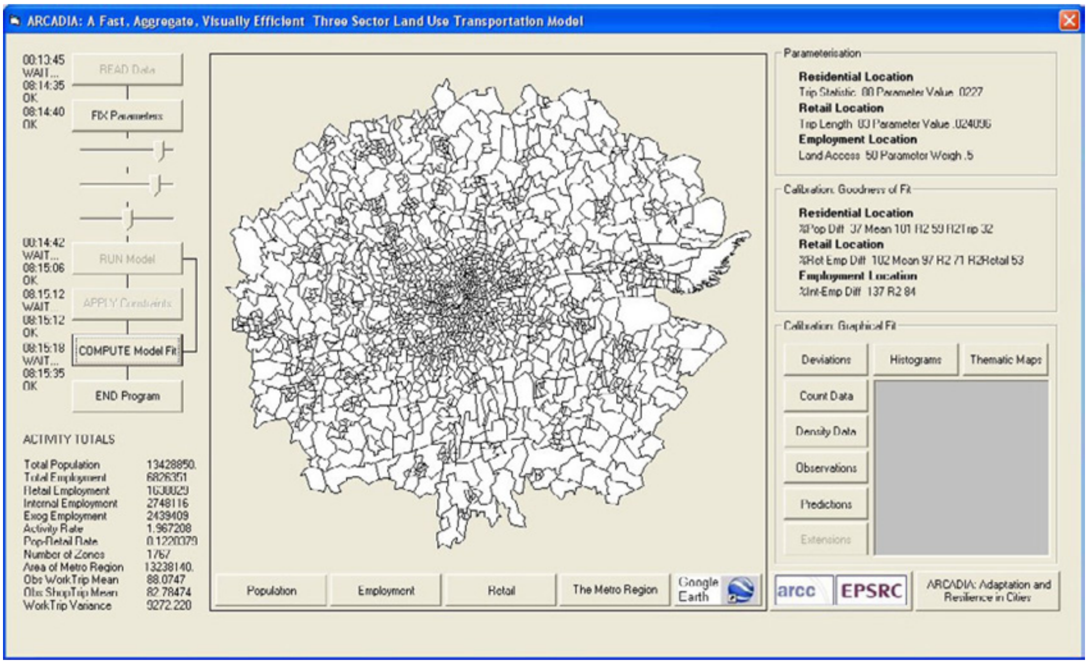



**Figure 3.** [In colour online.] A complete run of the desktop pilot model.

enables us to embed common subsystems like GIS, various simulation models, management tools, databases, and other external data sources into the framework.

Figure 4 defines the three main tiers of the architecture of the model. The individual tiers of the tool include a data tier composed of a relational database and/or a distributed system file, a logic tier that describes the underlying logic of the urban simulation models, and a presentation tier that forms the user interface. The data tier is composed using a relational database implemented in PostgreSQL (http://www.postgresql.org/) with PostGIS (http://postgis.net/) and an instance of a Hadoop Distributed File System (http://hadoop.apache.org/). The database is designed as a snowflake scheme following the principles of dimensional modelling for data warehousing (Kimball, 2002).

The logic tier is concerned with the state of the data that the urban models depend upon at a given instance and the rules driving the mechanics of the SIMULACRA model. In this tier the object-oriented domain model is the set of classes that represents the entities that define the structure of the actual world entities and perform the associated modelling processes. This is formulated as a series of classes that we do not have time to detail here but essentially fall under the umbrella of spatial interaction models. However, the variety of object-oriented design patterns (Gamma et al, 1995) used for SIMULACRA development are sufficiently generic that a range of alternative urban modelling techniques could be implemented in the future.

The logic tier is written entirely in JAVA and its entry point is a JAVA Servlet class. The Servlet class receives all requests from the client side and performs generic functions that direct the execution of an appropriate model from the SIMULACRA suite of models. However, some portions are built on Apache Hadoop (http://hadoop.apache.org/) distributed computation project in order to be able to run the models on a networked cluster of computers. In the current version of SIMULACRA, the JAVA Servlet class is hosted in an Apache Tomcat (http://tomcat.apache.org/) web application server. However, the design of the Servlet class leaves the possibility of hosting the JAVA Servlet in different Servlet containers, and develops a RESTful web service (Richardson and Ruby, 2008). The logic tier and the data-tier relational database component interact via SQL with Open Geospatial Consortium

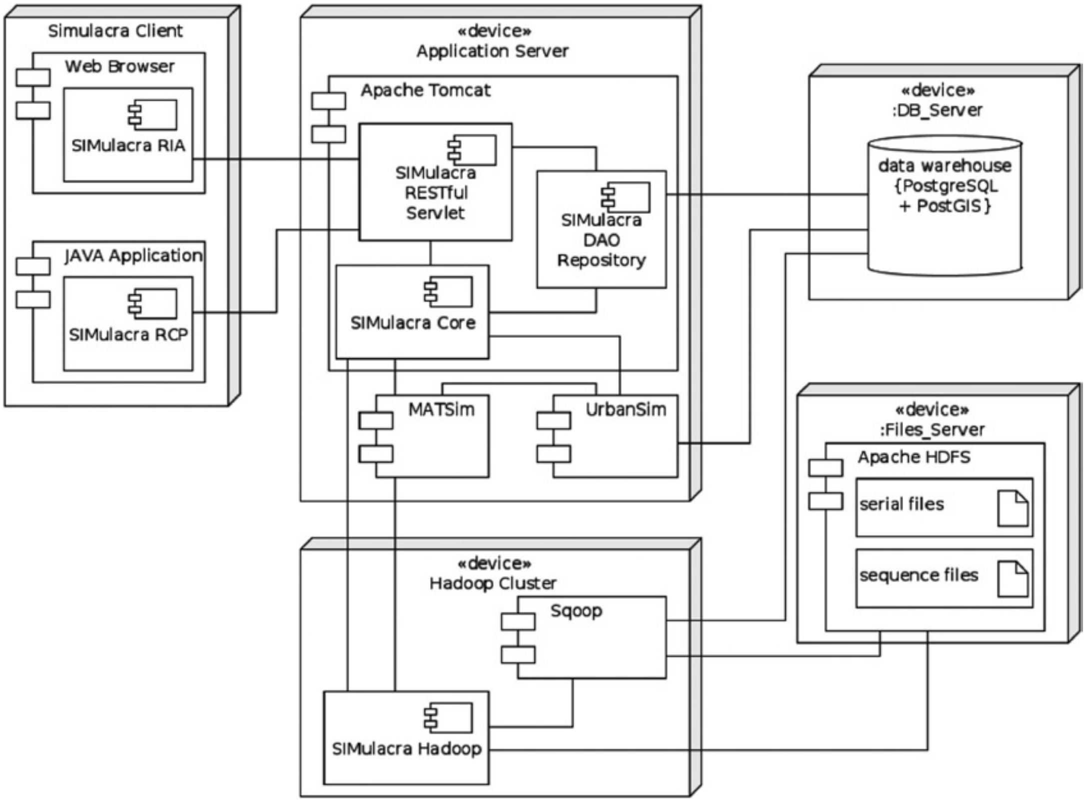



**Figure 4.** The three-tier architecture for the full (RCP–RIA) model.

(http://www.opengeospatial.org/) standard spatial extensions over a JDBC network protocol. The logic tier and the data-tier Hadoop Distributed File System instance interact over an HTTP interface via Apache Hadoop.

The focus of the presentation tier is usability based on interactivity, flexibility, and robustness. Throughout the SIMULACRA project, there has been a consistent emphasis on developing both a Rich Client Platform (RCP) and Rich Internet Application (RIA) for end users wishing to interact with the models' outputs.

Two discrete user-interface solutions have been developed to provide a means of interacting with the full range of urban models in an intuitive manner. These are the SIMULACRA RCP and the SIMULACRA RIA. Both the RCP and RIA provide a dashboard screen that visualises outputs in a range of formats and provides the user with the necessary tools to interact with the underlying logic tier to generate new outputs. This provides a rich user experience, which encourages experimentation and helps the end user to quickly develop an essential understanding of the dynamics of the underlying model. SIMULACRA RCP and RIA are based on libraries that are 100% client-side implementations. They follow a Model–View–Controller architectural style where the client makes GET and POST AJAX requests to different processes on the logic tier. The client interacts with the logic tier via JSON-encoded documents over an HTTP application protocol. SIMULACRA RCP is written entirely in JAVA using Eclipse RCP, GeoTools, and NASA World Wind SDK which we show below. SIMULACRA RIA is written in HTML5, CSS, and Javascript using RequireJS, Twitter Boostrap, JQuery, underscore.js, Backbone.js, Google Maps, and Google Charts which generates the outputs that we show in the next section.

## Calibration, predictions, applications

### Fitting the model

The desktop and RCP–RIA models are currently coordinated, reproducing the same outputs and we will report preliminary results. Table 1 presents parameter values, deviation statistics, and coefficients of variation for the three submodels decoupled from one another and for the house-price–travel cost variant of the residential location model, sketched out above. The models have been calibrated by trial and error, but we know the results shown can be massively improved once the models are better tuned to the data and formulated with their appropriate constraints. Nevertheless, for the standard three submodels the results are acceptable. The residential submodel does not perform as well as the other two and when we use the price–cost variant, this performance deteriorates further. We are currently exploring this model formulation offline through extensive data analysis of costs and prices in the region, and we intend to produce a much more realistic specification once we have identified the appropriate trip making behaviour with respect to these costs.

**Table 1.** Goodness of fit to the base year (2006) calibration.

| Location model | Residential | Residential price–cost | Retail | Internal employment |
|---|---|---|---|---|
| Parameter | 0.268 | 0.0001 | 0.278 | 0.278 |
| Percentage deviations | 39 | 42 | 104 | 134 |
| $r^2$—locations | 0.55 | 0.61 | 0.68 | 0.84 |
| $r^2$—interactions | 0.34 | 0.10 | 0.55 | na [a] |

[a] na—not applicable

In figure 5 we show several map variants with respect to the performance of the desktop pilot where we plot a sample of outputs based on deviations between observed and predicted counts and densities, histograms of counts and densities, and standard thematic maps. The key point to note is the speed at which these model runs and plots can be presented; in essence, the desktop model is suitable for use within a normal process of stakeholder dialogue where users and model builders can cluster around the model and explore many, many variants in a matter of hours, enabling rapid feedback concerning model results and the impacts of different scenarios. In figures 6 and 7 we show similar outputs from the RCP and RIA models, respectively, where more information is generated in terms of the display. This is akin to the previous 633-zone desktop model that we built for the GLA except that the run and display speeds are dramatically faster. For both models, immediacy of response is impressive.

However these models are the simplest possible, too simple perhaps, notwithstanding the fact that there is considerable research still to do on their overall coupling and hence convergence properties. We have disaggregated this model by two transport modes—public and private, which is an aggregation of the modes used in our GLA model (Batty, 2011; 2013). This doubles running times of the simplest model because each submodel is essentially replicated for the two modes that are made explicit. We have then disaggregated the demographic–residential sector into five population/household types and we have reconfigured the employment sectors so that we have five different employment types—in the pilot model there are three—retailing, internal employment, and exogenous employment. This increases running times by a factor of at least 5 again and thus a conservative estimate of the disaggregated model is that it takes at least 20 times as long to run. We are thus using the RCP–RIA framework for all future development although we will continue to mirror the full model with the desktop pilot as long as this remains useful to overall development, demonstration, and dialogue.

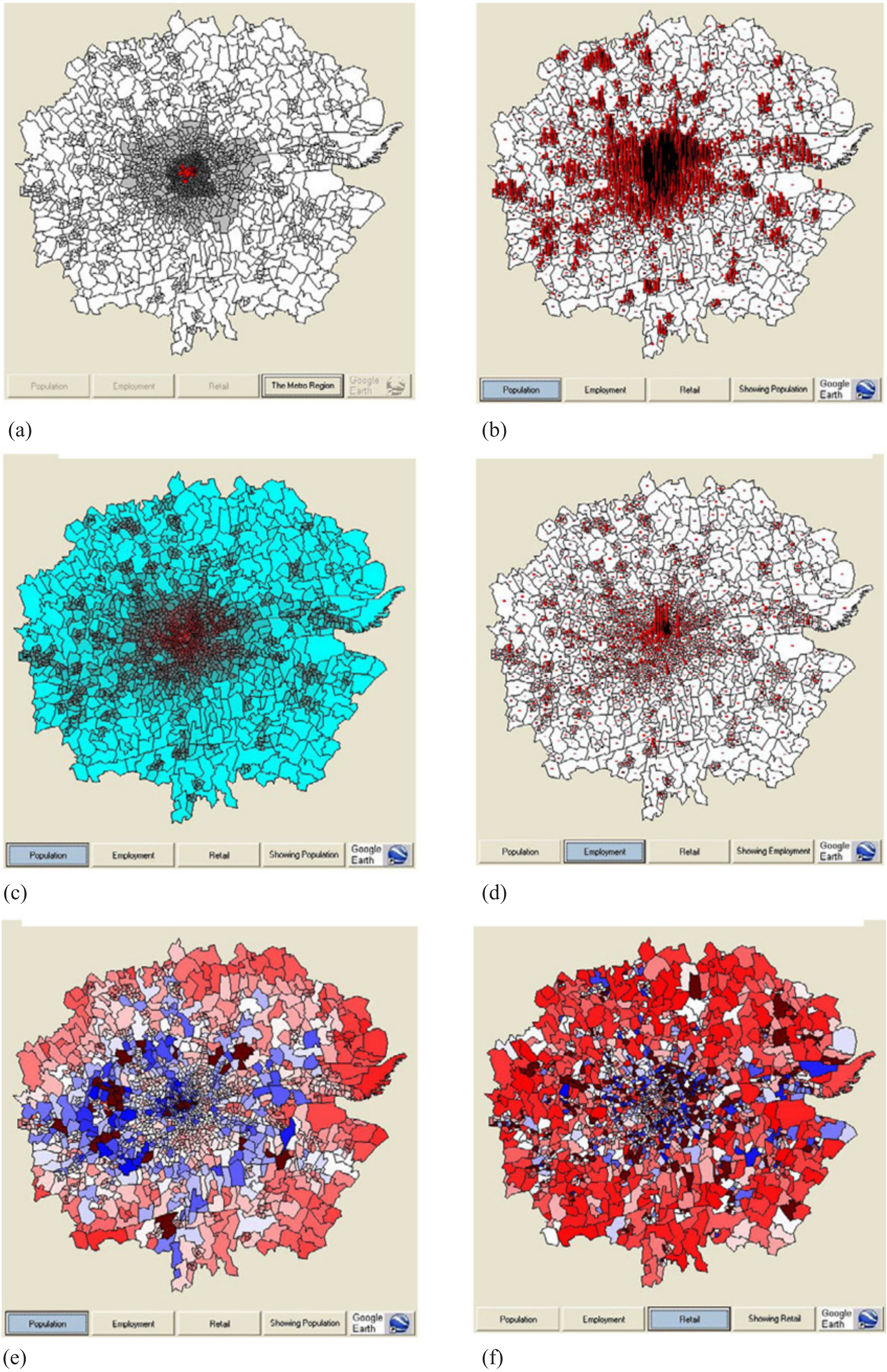
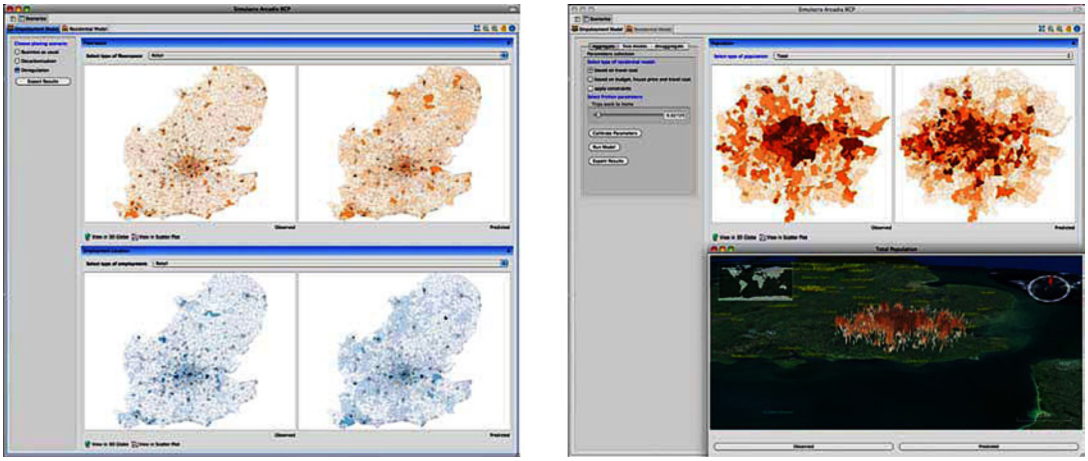
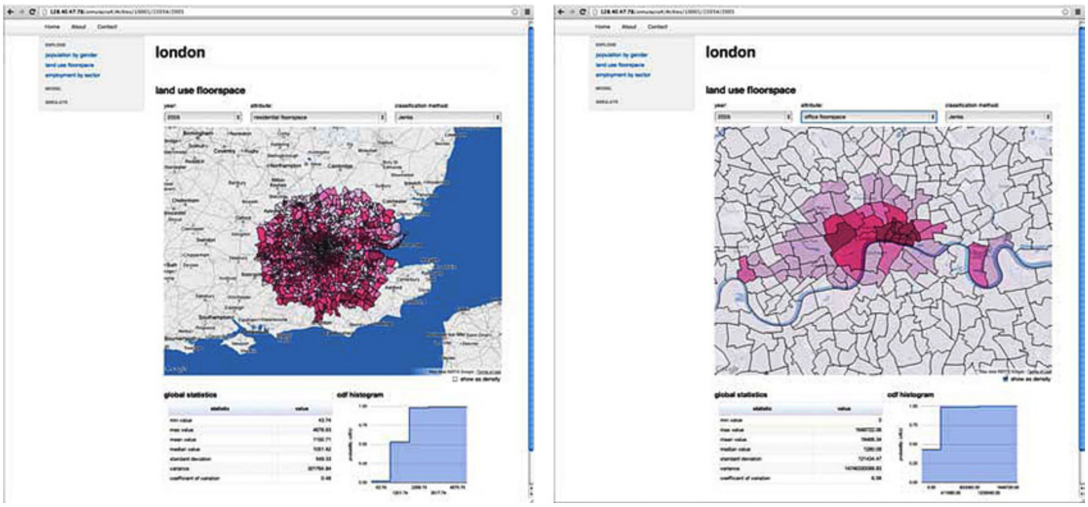



**Figure 5.** [In colour online.] Sample data and model outputs from the desktop pilot: (a) the Wider London Region, (b) observed population counts, (c) predicted population densities, (d) observed employment counts, (e) percentage population (observed−predicted) deviations, (f) percentage retailing (observed −predicted) deviations.



(a) (b)

**Figure 6.** [In colour online.] Sample data and model outputs from the SIMULACRA Rich Client Platform. (a) Thematic map outputs for the employment model in the Wider South East England Region, and (b) the model Outer Metropolitan Region with display in 3D using NASA World Wind.




(a) (b)

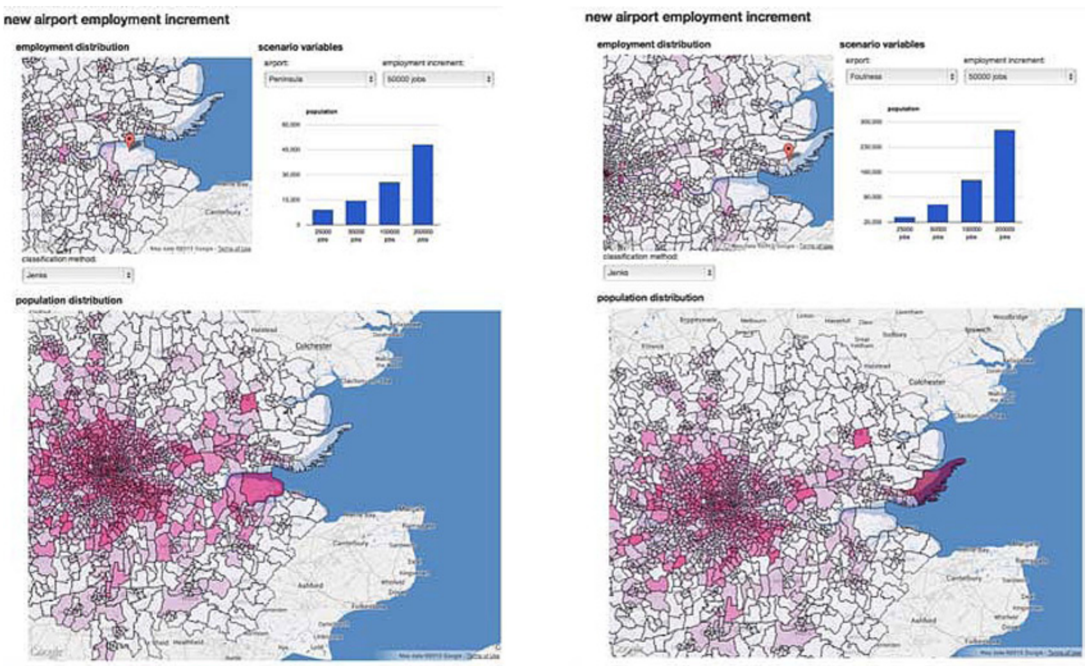
**Figure 7.** [In colour online.] Sample data and model outputs from the SIMULACRA Rich Internet Application: (a) thematic map outputs for the commercial floorspace in the model Outer Metropolitan Region and (b) the same floorspace data scaled to Central London and Docklands.

**Rapid assessment of alternative plans**

We already have some experience of using our previous GLA model for such rapid assessment, using the model directly with stakeholder groups. In fact, urban models of any kind are increasingly being ported into contexts where immediacy and visual intelligibility are key criteria and this can often only be illustrated by hands-on demos. We have also illustrated how we have been evaluating scenarios involving changing energy costs which we have explored extensively in our previous GLA model (Batty, 2013).

The requirement to evaluate scenarios rapidly requires an evaluation function within the modelling framework that has not yet been built but is under construction. Part of this is based on the fact that the model is now underpinned by prices and travel costs. We are extending this to deal with expenditure on consumer goods and we will be building a land-supply component into the various models that reflects rents (as well as house prices which are already a part of the residential location model). Because we are able to simulate the flow of money as well as people within the model, we can assess scenarios that do not only deal with changes to the physical structure of the metropolis in terms of land supply (constraints), transport infrastructure, and changes to the provision of floorspace but also with changes in travel costs, wages, house prices, and such like. In fact, these kinds of model are most appropriate (as indeed are many if not all models) to deal with 'what-if'-type scenarios of which we can generate many variants. As the model can be run rapidly and the evaluation of scenarios is immediate, we need to produce a framework for the consistent generation of such scenarios so that we can engage in considered choice of what to test and relate these using appropriate sensitivity analyses. These are features of more general planning support systems that are slowly being exploited despite the somewhat tortuous path to their development (Batty, 2008b).

To conclude our description of the way we have developed this class of models, and to further demonstrate the kinds of outputs that are possible, we have examined the impact of two new sites in the Thames Estuary for a new international airport to deal with the congestion at Heathrow and other London airports and to prepare for much more air traffic in the medium-term future, notwithstanding arguments about policies pertaining to a low-carbon future. Sites for such an airport have been explored for over forty years, but we are now in a position where, almost for the first time, we have models that are capable of generating many different possibilities almost immediately and in a context where designers and policy makers can begin to evolve potential scenarios through processes of systematic trial and error. In figure 8 we show two locations for the airport in the RCP interface. These are located on the south and north banks of the Thames: first on the Isle of Grain and second at Maplin Sands. We have input three levels of employment to define these airports and all the model does is to provide predictions of the impact of these jobs on population, but what is immediately clear is that the impact of these in terms of residential location is much wider than the immediate airport areas themselves. None of this is unexpected but, in the future, we will produce a battery of indicators to be added to this kind of scenario testing and thus




(a) (b)

**Figure 8.** [In colour online.] Visualising the impacts of new employment on residential population in the vicinity of two proposed sites for a new London airport: (a) Isle of Grain, (b) Maplin Sands.

provide a process where decision makers and designers can explore in intricate detail the impacts of these 'what-if' scenarios. Figure 8 in fact indicates that, in terms of visualisation, these models are extremely well integrated with exploratory data analyses which include many innovative kinds of visualisation.

**Key challenges, future developments**

We have listed the key development trajectories for this suite of models, but in conclusion we need to critically explore the extent to which this genus of comparative–static equilibrium models of urban structure provides a path worth following. As we have noted, there are many aspects of such models that have never been explored; particularly their coupling and convergence properties. This, we believe, is worth doing as much because of the light it will shine on coupling models in general and on the perpetuation of errors as on the appropriateness of the particular model structures to be further developed here. These models do not come into their own until they are disaggregated and thus this is a work in progress, as much informing readers about what we intend to do and how we will do it, as reporting work that has already been finished.

However, there are some very important developments in comparative static models that we intend to pursue. First, these models have an intrinsic dynamic structure in that we can partition activities into movers and stayers, activity that is inert and that which has the potential for change. This has never been done before and it represents a new way of embedding dynamics into model structure. Some hints as to how to do this were given a generation ago by one of the authors (Batty, 1986), but such developments are now possible because we have the power to continually experiment with model structures, something that was quite impossible before the current advances in computation which have essentially abolished or at least dramatically changed the limits on such exploration. Just as the whole process of planning support is being refashioned due to the use of online tools that provide immediate feedback, our ability to explore large, relatively realistic models in a rapid manner is providing a new dimension to this science. In the immediate future we will develop the SIMULACRA models in three ways: first, through disaggregation which poses challenges for running time: second, through the addition of explicit indicators which can be grafted onto this framework in the manner developed for the Propolis consortium project in 2004 (http://www.ltcon.fi/Propolis/index.htm); and, last but not least, the development of an intrinsic internal dynamics to comparative static models that promises to address the obvious concern that cities are never and never will be in equilibrium.

**References**

Batty M, 1986, "Technical issues in urban model development: a review of linear and non-linear model structures", in *Advances in Urban Systems Modelling* Eds B G Hutchinson, M Batty, (North-Holland, Amsterdam) pp 133–162

Batty M, 1992, "Urban modeling in computer-graphic and geographic information system environments" *Environment and Planning B: Planning and Design* **19** 663–685

Batty M, 2008a, "Fifty years of urban modelling: macro statics to micro dynamics", in *The Dynamics of Complex Urban Systems: An Interdisciplinary Approach* Eds S Albeverio, D Andrey, P Giordano, A Vancheri (Physica, Heidelberg) pp 1–20

Batty M, 2008b, "Planning support systems: progress, predictions, and speculations on the shape of things to come", in *Planning Support Systems for Cities and Regions* Ed. R K Brail (Lincoln Institute of Land Policy, Cambridge, MA) pp 3–30

Batty M, 2011, "Dynamics of land-use change and growth", in *Computer Modelling for Sustainable Urban Design: Physical Principles, Methods and Applications* Ed. D Robinson (Earthscan, London) pp 237–258

Batty M, 2013, "Visually-Driven Urban Simulation: exploring fast and slow change in residential location" *Environment and Planning A* **45** 532–552

Baudrillard J, 1995 [1981] *Simulacra and Simulation* (University of Michigan Press, Ann Arbor, MI)
Dawson R et al, 2009, "A blueprint for the integrated assessment of climate change in cities", in *Green Citynomics: The Urban War Against Climate Change* Ed. K Tang (Greenleaf, Chippenham, Glos) pp 32–51
Echenique M H, Grinevich V, Hargreaves A J, Zachariadis V, 2013, "LUISA: a land-use interaction with social accounting model; presentation and enhanced calibration method" *Environment and Planning B: Planning and Design* **40** 1003–1026
Epstein J M, 2008, "Why model?" *Journal of Artificial Societies and Social Simulation* **11(**4) 12, http://jassssocsurreyacuk/11/4/12html
Fowler M, 2003 *Patterns of Enterprise Application Architecture* (Addison-Wesley, Boston, MA)
Gamma E, Helm R, Johnson R, Vlissides J, 1995 *Design Patterns* (Addison-Wesley, Boston, MA)
Heppenstall A J, Crooks A T, See L M, Batty M (Eds) 2012 *Agent-based Models of Geographical Systems* (Springer, Berlin)
Jin Y, Echenique M, Hargreaves A, 2013, "A recursive spatial equilibrium model for planning large-scale urban change" *Environment and Planning B: Planning and Design* **40** 1027–1050
Kimball, R 2002 *The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling* 2nd edition (John Wiley, New York)
Lowry I S, 1964 *A Model of Metropolis* memorandum RM-4035-RC (The Rand Corporation, Santa Monica, CA)
Plummer A V, 1990, "The Chicago Area Transportation Study: creating the first plan (1955–1962): a narrative", http://wwwcatsmpocom/pub/reports/history/cats_1954-62pdf
Putnam S H, 2007 [1983] *Integrated Urban Models: Analysis of Transportation and Land Use* (Routledge, London)
Richardson L, Ruby S, 2008 *RESTful Web Services* (O'Reilly Media, Inc., San Francisco, CA)
Smith D A, Vargas-Ruiz C, Batty M, 2013 "Simulating the spatial distribution of employment in large cities: with applications to Greater London", in *Employment Location in Cities and Regions: Models and Applications* Eds F Pagliara, M de Bok, D Simmonds, A Wilson (Springer, Berlin) pp 79–106
Waddell P, Borning A, Noth M, Freier N, Becke M, Ulfarsson G, 2003, "Microsimulation of urban development and location choices: design and implementation of urbanism" *Networks and Spatial Economics* **3** (1) 43–67
Wegener M, 2008, "Multi-scale spatial models: linking macro to micro", http://wwwspiekermann-wegenerde/pro/pdf/MW_CASA_090108pdf
Wegener M, 2014, "Land-use transport interaction models", in *Handbook of Regional Science* Eds M M Fischer, P Nijkamp (Springer, Heidelberg) pp 741–758
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