Building a Spatial Microsimulation Decision Support System

Dimitris Ballas[1], Richard Kingston[2], John Stillwell[3] and Jianhui Jin[3]

[1] Department of Geography, University of Sheffield, Winter Street, Sheffield S10 2TN, England
e-mail: d.ballas@sheffield.ac.uk

[2] School of Planning and Landscape, University of Manchester, Manchester, England
e-mail: richard.p.kingston@man.ac.uk

[3] School of Geography, University of Leeds, Leeds LS2 9JT, England;
e-mail: john@geography.leeds.ac.uk; j.jin@geography.leeds.ac.uk;

SUMMARY

This paper presents MicroMaPPAS (Micro-simulation Modelling and Predictive Policy Analysis System), which is a Spatial Decision Support Systems (SDSS) under development for Leeds City Council. The innovative feature of this system is the use of spatial microsimulation techniques for the enhancement of local policy decision making. The paper addresses the relevant data issues and technical aspects of the linkage of spatial microsimulation modelling frameworks to SDSS and deals with the wider implications that such a linkage may have to local governance procedures. It will therefore be of interest to local government policy makers and practitioners as well as to researchers interested in the prospects of policy simulation models for the enhancement of local decision making and democracy.

KEYWORDS: *spatial microsimulation, spatial decision support systems, GeoTools*

INTRODUCTION

This paper presents a spatial microsimulation-based decision support system. Traditionally, confidentiality concerns have been the main reason why demographic and socio-economic data on individuals, despite being collected from censuses and surveys, have not been available for researchers. Microsimulation is a methodology that attempts to estimate the demographic and socio-economic characteristics of human behaviour of individual people or households (Clarke, 1996). Ballas and Clarke (2003) have recently provided a detailed review of the development of the methodology from Orcutt's original work in the 1950s (Orcutt, 1957).

In this paper we present a spatial decision support system for Leeds City Council. The system is based on a microsimulation model which is capable of constructing a list of 715 thousand individuals within households along with their associated attributes for any point in time, past or future. There are various different ways of calibrating the model but the results are particularly valuable because they combine data from different sources to provide estimates of the probabilities that individuals or households will have particular characteristics and thus create new population cross-classifications unavailable from published sources. So, for example, it becomes possible to identify individuals with the characteristics of being aged 18, a lone parent, unemployed and living in private accommodation in an area prone to high levels of crime. Alternatively, households can be identified in the outer suburbs that contain five persons and have a head of household who is a professional working in another city and earning over £50,000 per year. Once the long list of individuals and their attributes has been simulated, the individuals and households (and the attributes which they possess) can be aggregated to any geographical scale which is deemed appropriate such as output areas, wards or postal sectors for example, or more specific areas designed for policy implementation such as regeneration areas.

The system presented in this paper utilises spatial microsimulation techniques to provide a spatial decision support tool for local council officers. In particular, the system can be used to describe current conditions and issues in neighbourhoods, predict future trends in the composition and health of neighbourhoods and conduct modelling and predictive analysis to measure the likely impact of policy interventions at the local level. The overall aims of the project underpinning the system presented here has are as follows: (i) to develop a static microsimulation model to describe current conditions in Leeds; (ii) to enable the performance of '*What if?*' analysis on a range of policy scenarios; and (iii) to develop a dynamic microsimulation model to predict future conditions in Leeds under different policy scenarios. This paper reports progress in meeting the above aims and outlines the associated difficulties and data issues. In particular, section 2 briefly describes the data used and introduces the combinatorial optimisation spatial microsimulation technique that has been adopted. Section 3 describes the spatial decision support MICROMaPPAS system and section 4 offers some concluding comments.

DATA AND METHODS

The spatial decision support system presented in this paper utilises a spatial microsimulation model which links a wide range of data sets, including 2001 Census data for Output Areas and sample data from the British Household Panel Survey. The framework for the spatial decision support system presented in this paper is illustrated in Figure 1.

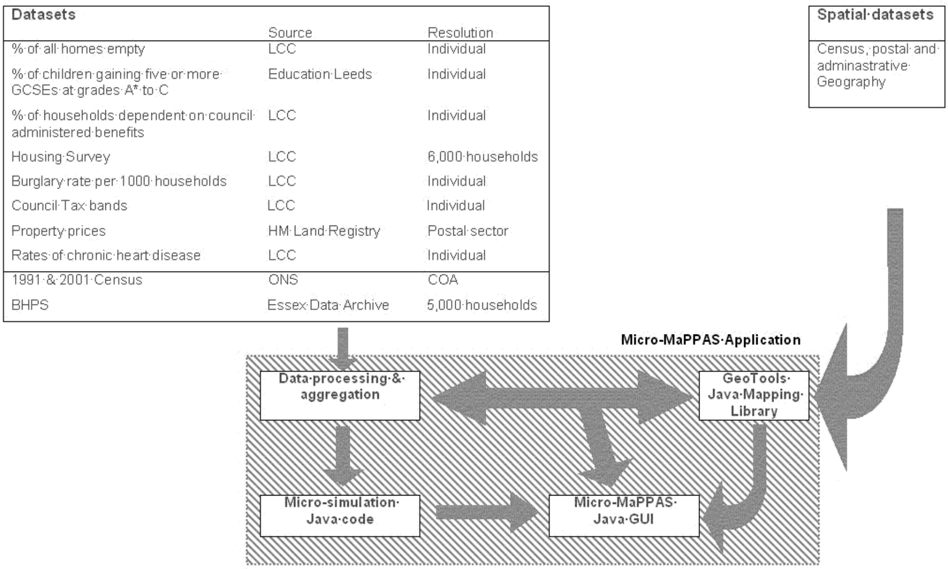


Figure 1: Data sets and system architecture of the Micro-MAPPAS project

The data sets described in figure 1 are linked on the basis of a spatial microsimulation model which implements a combinatorial optimisation approach to generating spatially disaggregated population and household microdata sets at output area (OA) level for the metropolitan district of Leeds as defined in 2001. In particular, the modelling exercise involves the construction of micro-level population using existing 2001 Census Key Statistics (KS) tables for population and household characteristics of all 2,439 OAs in Leeds together with sample data from the British Household Panel Survey (BHPS), a national microdata set of household characteristics. The BHPS contains data on different variables (such as income) for households and their occupants that can be used to derive estimates of 'new' variables for

OAs. The technique we have adopted is known as 'simulated annealing' and is distinguished from other methods such as iterative proportional fitting (Ballas, 2001; Norman, 1999; Ballas and Clarke, 2000). Simulated annealing involves reweighting the microdata sample from the BHPS so that it fits OA data for Leeds from the Census. In the first instance, the BHPS microdata set has been reweighted to estimate its parent population at the micro-spatial scale. The BHPS provides a detailed record for a sample of households and all of their occupants. The reweighting method can enable the sampling of this universe of records to find the set of household records that best matches the population described in the KS tables for each OA.

The actual procedure works as follows. First, a series of Census (KS) tables that describe the small area of interest must be selected. The next step is to identify the records of the BHPS microdata that best match these tables. However, there are a vast number of possible sets of households that can be drawn from the BHPS sample. Clearly, it would be impractical to exhaustively consider all possible sets so this is where the simulated annealing[57] (Ballas *et al.,* 1999; 2003) is used to find a set that fits the target tables well. The Micro-MaPPAS simulation model builds on a previous computer software known as SimLeeds (Ballas, 2001) and uses the 10th wave of the BHPS to provide a detailed record for a sample of households and all of their occupants.

A simple example can be described for clarification. Let's assume that, according to the Census, in a particular OA there are 100 households, of which 60 are owner occupiers, 10 are renting from a local authority (LA) or housing association (HA), and the remaining 30 are privately rented. The simulated annealing procedure would select a combination of BHPS households that would have tenure characteristics as close as possible to the actual data. An exact match would be possible if the tenure KS table is the only 'constraint' in the procedure. However, the purpose of using a combinatorial optimisation technique is to select households that match several KS table constraints. Let's assume that we introduce the number of cars by household KS table as a further constraint and that our OA contains 50 households with 1 car, 20 households with 2 or more cars; and 30 households with no car. The aim of the annealing would now be to find a set of 100 BHPS households that best fit both tenure and car ownership constraints. To do this, an initial random sample of records is selected from the BHPS until sufficient households are represented (i.e. if there are 100 households in the OA, then 100 households will be selected at random). These records are used to create tables that match the selected target KS tables. An initial random selection of 100 BHPS households could result in the distribution described in the first row of Table 1. The total absolute error of 88 is the sum of the differences between the simulated and the actual Census values on the bottom row.

[57] Annealing is a physical process in which a solid material is first melted in a heat bath by increasing the temperature to a maximum value at which point all particles of the solid have high energies and the freedom to randomly arrange themselves in the liquid phase. This is then followed by a cooling phase, in which the temperature of the heat bath is slowly lowered. The particles of the material attempt to arrange themselves in a low energy state during the cooling phase. When the maximum temperature is sufficiently high and the cooling is carried out sufficiently slowly then all the particles of the material eventually arrange themselves in a state of high density and minimum energy (Kirkpatrick et al. 1983; Dowsland, 1993; Pham and Karaboga, 2000; Van Laarhoven and Aarts, 1987). In geography, simulated annealing has been applied in various contexts for different problems (see for instance Alvanides, 2000; Ballas, 2001; Openshaw and Rao, 1995; Openshaw and Schmidt, 1996; Williamson *et al.*, 1998).

	Household car ownership characteristics			Household tenure characteristics		
	1 car	**2+ cars**	**No car**	**Owner-occupier**	**LA/HA rented**	**Other**
Simulation	27	24	49	39	17	44
Census	50	20	30	60	10	30
Absolute error	23	4	19	21	7	14

Table 1: Calculating the absolute error

The task in the simulated annealing procedure is to minimise the total absolute error. In order to do so a record in the originally selected household set is selected at random and replaced with one chosen at random from the universe of records. The error is recalculated and the change in error (Δe) is calculated. If Δe is less than zero then there has been an improvement and the change is accepted; if not, then exp(-Δe/t) is compared to a random number between 0 and 1. If it is greater than the random number, then the change is accepted; otherwise the change is rejected and reversed. So, in the above example, a randomly selected household out of the 100 originally sampled would be swapped with another household. Let's assume that a LA/HA renting household is swapped with an owner occupier household from the BHPS. This would result in the reduction of error by 1, as the new owner occupier simulated total would be 40 (instead of 39) and therefore closer to the actual Census total (60 households) and the LA/HA rented total would be 16 (instead of 17), which is closer to the Census actual total of 10. This household swap would therefore be accepted. Conversely, if the change increased the error it would be rejected and another household selected. It is evident from this example that the simulated annealing model is a very computationally intensive, particularly when several table constraints are introduced.

THE MICRO-MaPPAS SYSTEM

The Micro-MaPPAS software is written in the Java programming language version 1.4, which means that it can be installed and operated on any computer system and platform. A default set of simulations generated for OAs is loaded when the system is booted up. The structure of the Micro-MaPPAS system is illustrated in Figure 1. Through the graphical user interface (GUI), the user has access to various modules: the model controller, the model diagnostics, the data analyser, the mapping controller and the scenario builders.

First, the *model controller* module allows the user to set the parameters for a small area simulation. The user can set the temperature, the number of model iterations and the number of restarts (Figure 2) and also apply weights to input tables (Figure 3) using slider bars. A simulation can take several hours to run if results are required for all OAs in Leeds and if relatively high temperature and large number of iterations or restarts are selected. It is also possible to run the simulation model for OAs contained within individual wards or community areas.

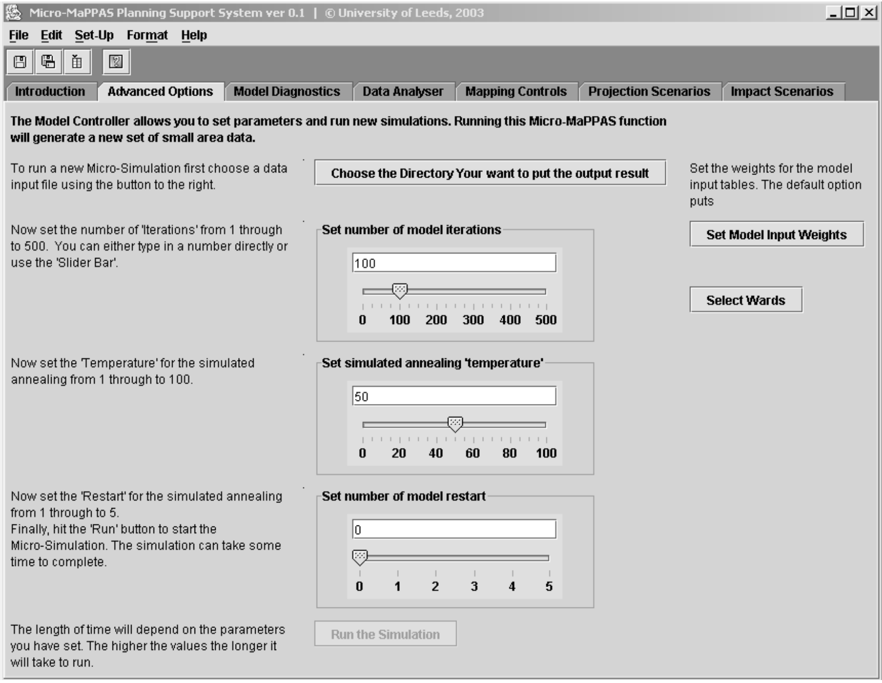


Figure 2: Model controller interface

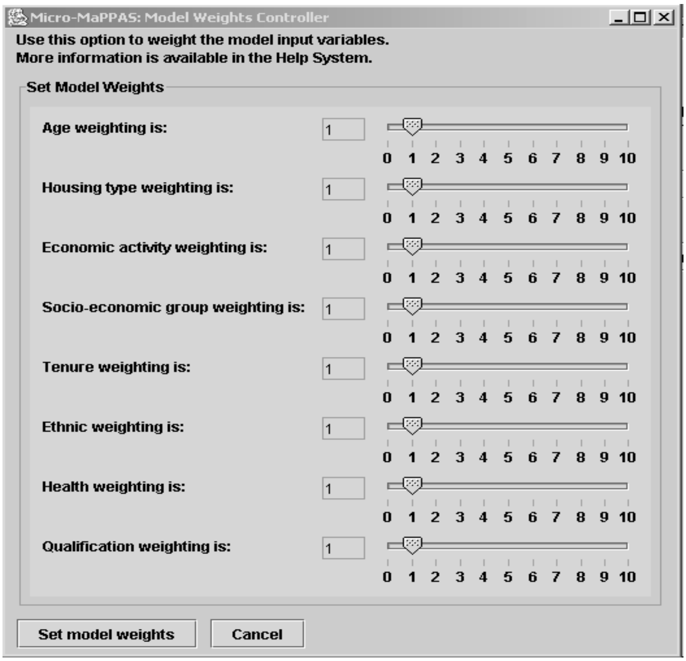
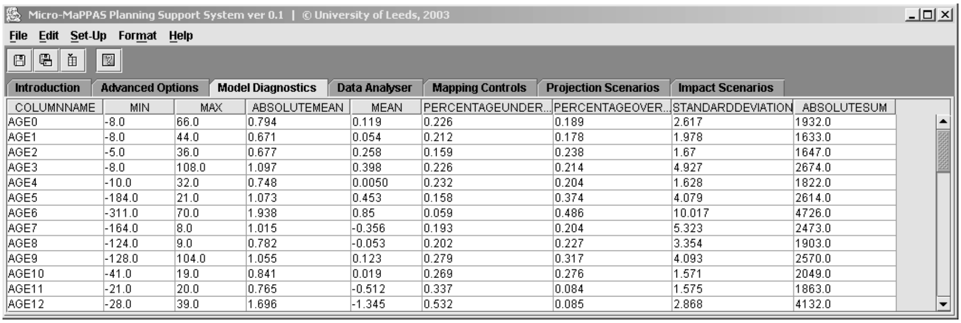


Figure 3: Weights controller interface

The *model diagnostics* module provides details of the accuracy of the microsimulation. It compares the simulated data for OAs with the actual census variables and produces a set of basic statistics (minimum, maximum, absolute mean, mean, standard deviation) and also the percentages of values that have been over-predicted and under-predicted. Each simulation generated has a corresponding model diagnostics table and Figure 4 illustrates the statistics associated with the single year of age variables from 0 to 12. The mean error is lowest (closest to zero) for those aged 10 and highest for those aged 12, for example.



COLUMNNAME	MIN	MAX	ABSOLUTEMEAN	MEAN	PERCENTAGEUNDER...	PERCENTAGEOVER...	STANDARDDEVIATION	ABSOLUTESUM
AGE0	-8.0	66.0	0.794	0.119	0.226	0.189	2.617	1932.0
AGE1	-8.0	44.0	0.671	0.054	0.212	0.178	1.978	1633.0
AGE2	-5.0	36.0	0.677	0.258	0.159	0.238	1.67	1647.0
AGE3	-8.0	108.0	1.097	0.398	0.226	0.214	4.927	2674.0
AGE4	-10.0	32.0	0.748	0.0050	0.232	0.204	1.628	1822.0
AGE5	-184.0	21.0	1.073	0.453	0.158	0.374	4.079	2614.0
AGE6	-311.0	70.0	1.938	0.85	0.059	0.486	10.017	4726.0
AGE7	-164.0	8.0	1.015	-0.356	0.193	0.204	5.323	2473.0
AGE8	-124.0	9.0	0.782	-0.053	0.202	0.227	3.354	1903.0
AGE9	-128.0	104.0	1.055	0.123	0.279	0.317	4.093	2570.0
AGE10	-41.0	19.0	0.841	0.019	0.269	0.276	1.571	2049.0
AGE11	-21.0	20.0	0.765	-0.512	0.337	0.084	1.575	1863.0
AGE12	-28.0	39.0	1.696	-1.345	0.532	0.085	2.868	4132.0

Figure 4: Model diagnostics table for simulated data for OAs

It should be noted that the model generates a simulated set of data of individuals but these are never visible through the interface. The simulated data are aggregated to OAs and the *data analyser* module provides a table view of this information (Figure 5). Below the table are a number of buttons that enable the user to run queries on the OA data to select the information required but also to aggregate the data to another spatial scale if required. The query builder interface is shown in Figure 5 and, once a query has been constructed, the results are returned to a table in the bottom half of the data analyser window. As an example, a query is undertaken which selects the number of individuals in each ward aged between 20 and 30 whose household income is between £20,000 and £30,000. The results of the query are shown in the bottom section of the data analyser and these can then be mapped.

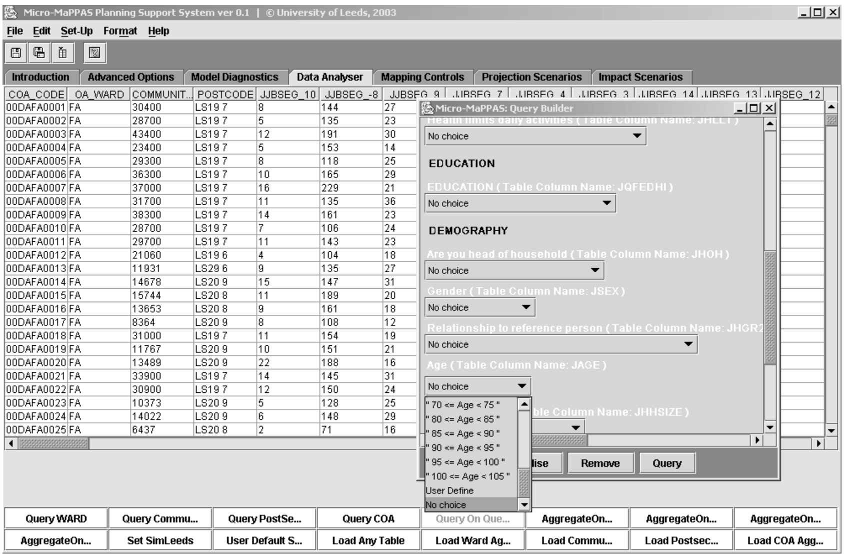


Figure 5: Data analyser and query interface with age drop down menu in use

Further, the *mapping controls* module allows the user to select a variable from a query and map the results at any of the geographical scales of OA, community area, ward or postal sector. Figure 6 illustrates the mapping of the query relating to persons aged 20-30 with incomes of £20-30,000. Mapping functions include panning and zooming and symbology editing.

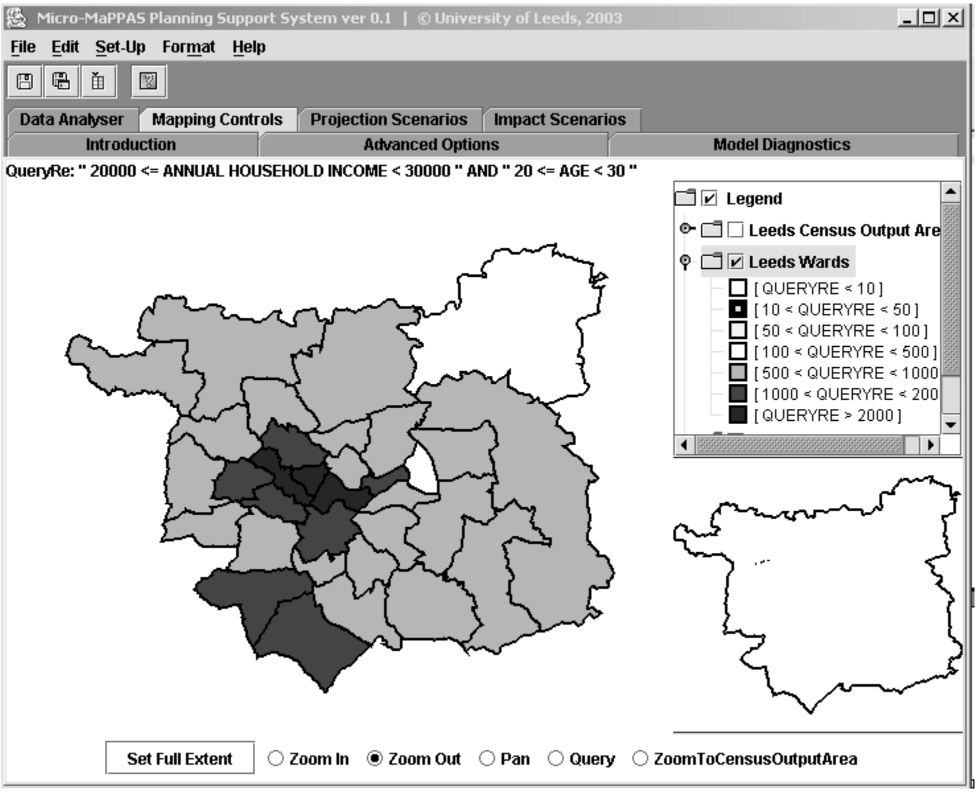


Figure 6: The results of the query as shown in the mapping controller

Finally, one of our immediate priorities is to render the system capable of designing and running simulations for future scenarios based upon different assumptions about population change in the future, and also to undertake some evaluation of 'what if' scenarios. Thus, we are currently developing two additional modules: *projection scenarios* and *impact scenarios* modules.

CONCLUDING COMMENTS

It has long been argued that spatial microsimulation has a great potential for socio-economic impact assessment (see for instance, Ballas and Clarke, 2001a) and for the geographical analysis of the impacts of social policies (Ballas and Clarke, 2001b; Ballas *et al.*, 2003). In this paper we demonstrated this potential further by presenting the first attempt to link spatial microsimulation modelling frameworks to Spatial Decision Support Systems (SDSS). In particular, we presented Micro-MaPPAS, which is a system that added spatial decision making capabilities to a spatial microsimulation model. We believe that systems such as Micro-MaPPAS can play a very important role in the on-going debates on the role of potential of new technologies to promote local democracy and electronic decision-making. This paper addressed the relevant data issues and technical aspects of the linkage of spatial microsimulation modelling frameworks to SDSS and described the capabilities of the Micro-MaPPAS system. The latter has a great potential for local policy analysis and may have wider implications for local governance procedures. It also demonstrates the prospects of policy simulation models for the enhancement of local decision making and democracy.

Further, it can be argued that systems such as Micro-MaPPASa model such as *SimLeeds* developed in JAVA, which is a platform independent programming language, can be put on the World Wide Web and linked to Virtual Decision-Making Environments (VDMEs). The latter are Internet World Wide Web based systems that allow the general public to explore 'real world' problems and become more involved in the public participation processes of the planning system (Kingston *et al.,* 2000). Systems such as Micro-MaPPAS can potentially be used not only to provide information on the possible consequences and the local multiplier effects of major policy changes but also to inform the general public about these and to enhance, in this way, the public participation in policy making procedures (Ballas, Kingston and Stillwell, 2003).

ACKNOWLEDGEMENTS:

The research presented in this paper was funded by Leeds City Council. All Census and digitised boundary data used in this paper are Crown Copyright. The BHPS data were obtained from the UK Data Archive (University of Essex).

BIBLIOGRAPHY

Alvanides S, (2000) Zone design methods for application in human Geography, *Unpublished PhD Thesis*, School of Geography, University of Leeds, Leeds.

Ballas, D. (2001) A spatial microsimulation approach to local labour market policy analysis, *Unpublished PhD Thesis*, School of Geography, University of Leeds, Leeds.

Ballas, D., Clarke, G.P. (2000), GIS and microsimulation for local labour market policy analysis, *Computers, Environment and Urban Systems, 24*, 305-330

Ballas, D. and Clarke, G. P. (2001a), Towards local implications of major job transformations in the city: a spatial microsimulation approach, *Geographical Analysis* 33, 291-311

Ballas, D., Clarke, G.P. (2001b), Modelling the local impacts of national social policies: a spatial microsimulation approach, *Environment and Planning C: Government and Policy, 19,* 587 – 606

Ballas, D., Clarke, G. P. and Turton, I. (1999) Exploring microsimulation methodologies for the estimation of household attributes, Paper presented at the 4th International Conference on GeoComputation, Fredericksburg, Virginia, USA, 25-28 July.

Ballas, D., Clarke, G. P. and Turton, I. (2003) A spatial microsimulation model for social policy micro-spatial analysis, in Boots, B., Okabe, A. and Thomas, R. (eds.) *Modelling Geographical Systems: Statistical and Computational Applic ations*, Kluwer, Dordrecht, pp. 143-168.

Ballas, D, Kingston R, Stillwell, J, (2003), *Public participation in local policy-making with the use of GIS-based microsimulation*, poster presented at the 2nd annual Public Participation GIS Conference, Portland State University, Portland, USA, 20-22 July 2003

Dowsland, K. (1993) Simulated annealing, in Reeves, C. (ed) *Modern Heuristic Techniques for Combinatorial Problems*, Blackwell, Oxford.

Kirkpatrick, S, Gelatt, C. D. Jr. and Vecchi, M. P. (1983) Optimization by simulated annealing, *Science*, 220: 671-680.

Kingston, R., Carver, S., Evans, A. and Turton, I. (2000) Web-Based Public Participation Geographical Information Systems: an aid to Local Environmental Decision-Making, *Computers, Environment and Urban Systems*, 24, 109-125

Norman, P. (1999), *Putting Iterative Proportional Fitting on the Researcher's Desk*, working paper 99/03, School of Geography, University of Leeds, Leeds LS2 9JT

Openshaw, S. and Rao, L. (1995) Algorithms for reengineering 1991 Census geography, *Environment and Planning A*, 27: 425-446.

Openshaw, S. and Schmidt, J. (1996) Parallel simulated annealing and genetic algorithms for re-engineering zoning systems, *Geographical Systems*, 3: 201-220.

Pham, D.T. and Karaboga, D. (2000) *Intelligent Optimisation Techniques: Genetic Algorithms, Tabu Search, Simulated Annealing and Neural Networks*, Springer, London.

Van Laarhoven, P.J.M. and Aarts, E.H.L. (1987) *Simulated Annealing: Theory and Applications*, Kluwer, Dordrecht.

Williamson, P., Birkin, M. and Rees, P. (1998) The estimation of population microdata by using data from small area statistics and samples of anonymised records, *Environment and Planning A*, 30: 785-816.