Ballas, D., R. Kingston, and J. Stillwell, 2004, Using a Spatial Microsimulation Decision Support System for Policy Scenario Analysis, In: Van Leeuwen, J.P. and H.J.P. Timmermans (eds.) *Recent Advances in Design & Decision Support Systems in Architecture and Urban Planning*, Dordrecht: Kluwer Academic Publishers, ISBN: 1-4020-2408-8, p. 177-191.

Using a Spatial Microsimulation Decision Support System for Policy Scenario Analysis

Dimitris Ballas[1], Richard Kingston[2], and John Stillwell[3]
[1] *Department of Geography, University of Sheffield, England*
[2] *School of Planning and Landscape, University of Manchester, England*
[3] *School of Geography, University of Leeds, England*

Keywords: Spatial Microsimulation, Spatial Decision Support Systems, Geotools.

Abstract: This paper discusses the potential of a spatial microsimulation-based decision support system for policy analysis. 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. A large dynamic spatial micro-simulation model is being constructed for the population of Leeds (approximately 715,000 individuals) based on spatial microsimulation techniques in conjunction with a range of data, including 2001 Census data for Output Areas and sample data from the British Household Panel Survey. The project has three main aims 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. The paper reports progress in meeting the above aims and outlines the associated difficulties and data issues. One of the significant advantages of the spatial microsimulation approach adopted by this project is that it enables the user to query any combination of variables that is deemed desirable for policy analysis. The paper will illustrate the software tool being developed in the context of this project that is capable of carrying out queries of this type and of mapping their results. The decision support tool is being developed to support policy-makers concerned with urban regeneration and neighbourhood renewal.

1. 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. Spatial 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 the combinatorial optimisation spatial microsimulation technique that has been adopted. Section 3 describes the spatial decision support Micro-MaPPAS system, section 4 presents some policy relevant simulation outputs and section 5 offers some concluding comments.

2. 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.

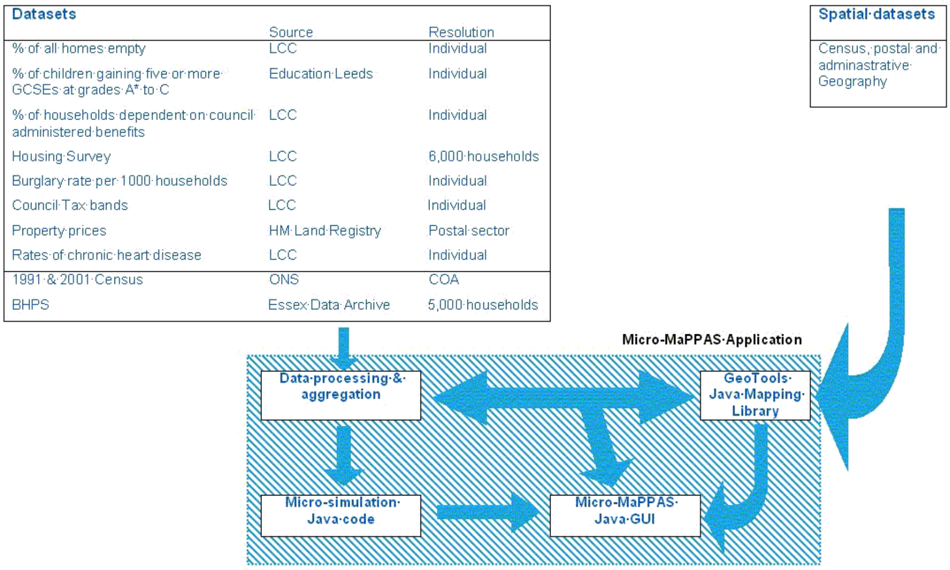


Figure 1. Data sets and system architecture of the Micro-MAPPAS project

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[1] (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. Ballas *et al.* (2004) describe the combinatorial optimisation method in some detail with the use of illustrative examples. The remainder of this paper describes the various modules of the spatial decision support system and provides some policy relevant simulation outputs.

[1] 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).

3. 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.

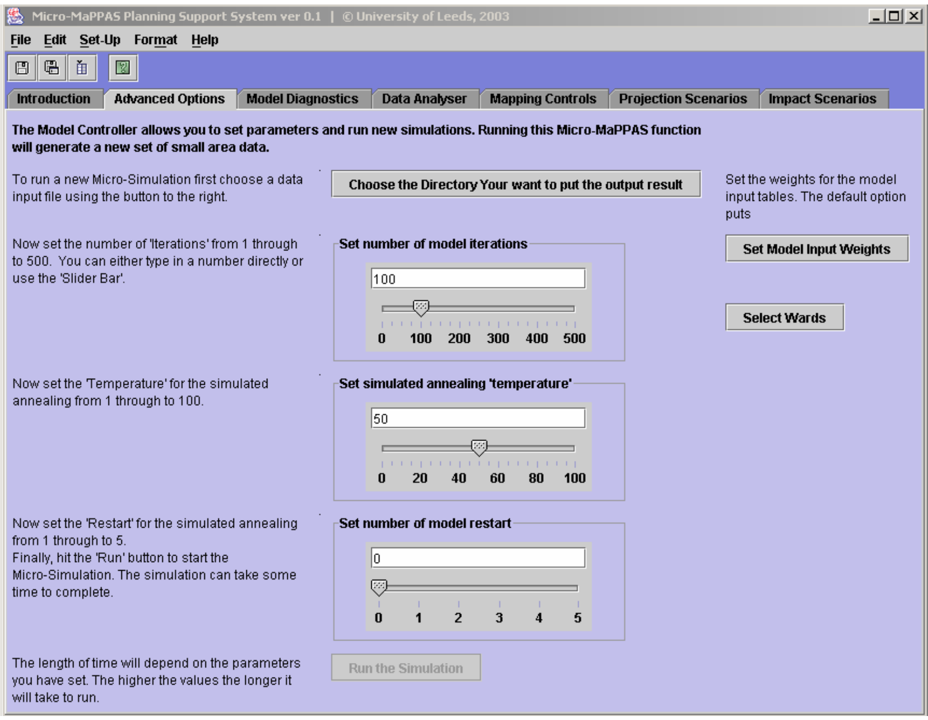


Figure 2. Model controller interface

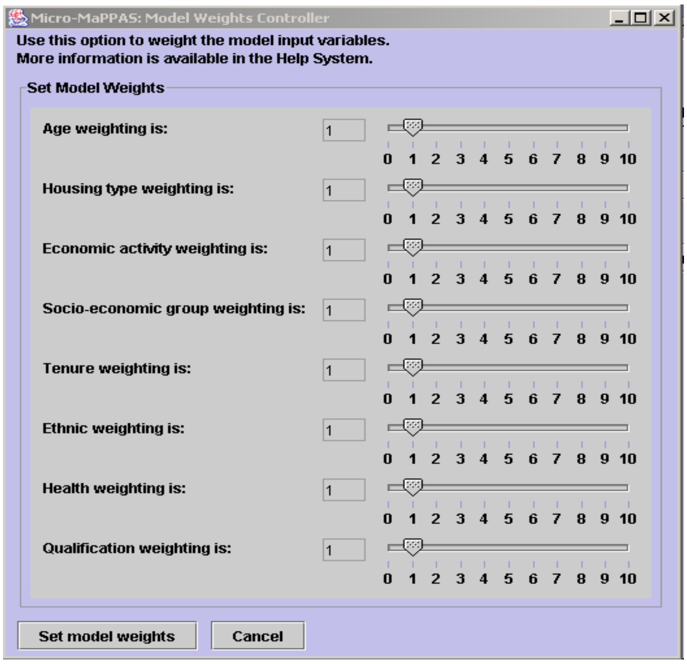
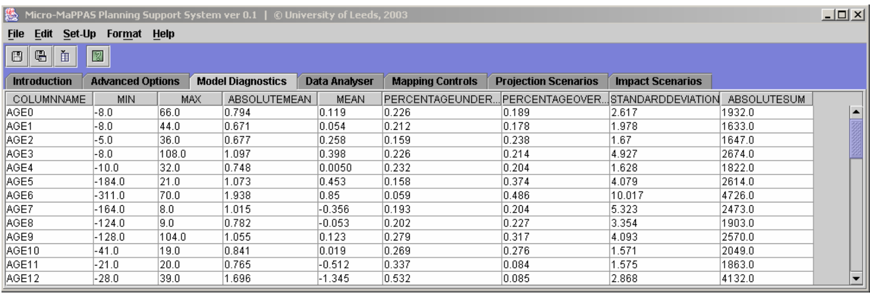


Figure 3. Weights controller interface

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.


Micro-MaPPAS Planning Support System ver 0.1 | © University of Leeds, 2003

File Edit Set-Up Format Help

Introduction | Advanced Options | Model Diagnostics | Data Analyser | Mapping Controls | Projection Scenarios | Impact Scenarios

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.

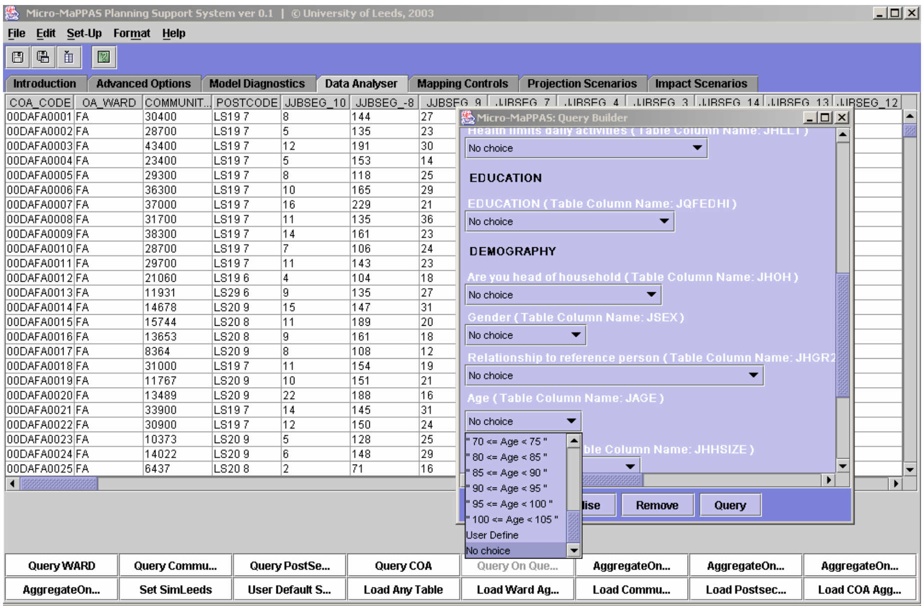


Figure 5. Data analyser and query interface with age drop down menu in use

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.

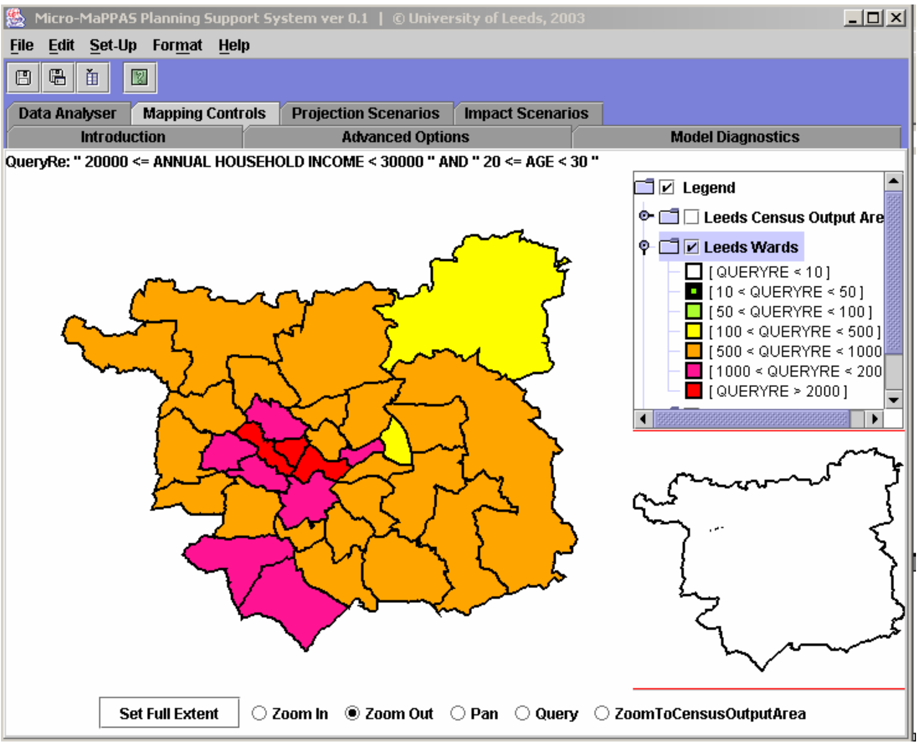


Figure 6: The results of the query as shown in the mapping controller

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.

4. SIMULATION OUTPUTS AND POLICY RELEVANCE

The MicroMaPPAS system has been designed to create sets of simulations for 2001 as the base period. Three examples of sub-group distributions generated by the MicroMaPPAS query are presented by way of illustration. These examples are all produced from the simulation that gives equal weight to all the tables used in the simulation. The first two queries provide more detailed information about the locations of disadvantaged groups. Figure 7 show the distribution of female lone parents with dependent children, where the household income is less than £10,000 per year. The query has been carried out at the scale of community areas and the distribution demonstrates the concentration of this particular subgroup in areas to the east of the city centre (Chapeltown, Burmantofts, Harehills and Seacroft South) and to the south (Beeston Hill, Belle Isle and Middelton).

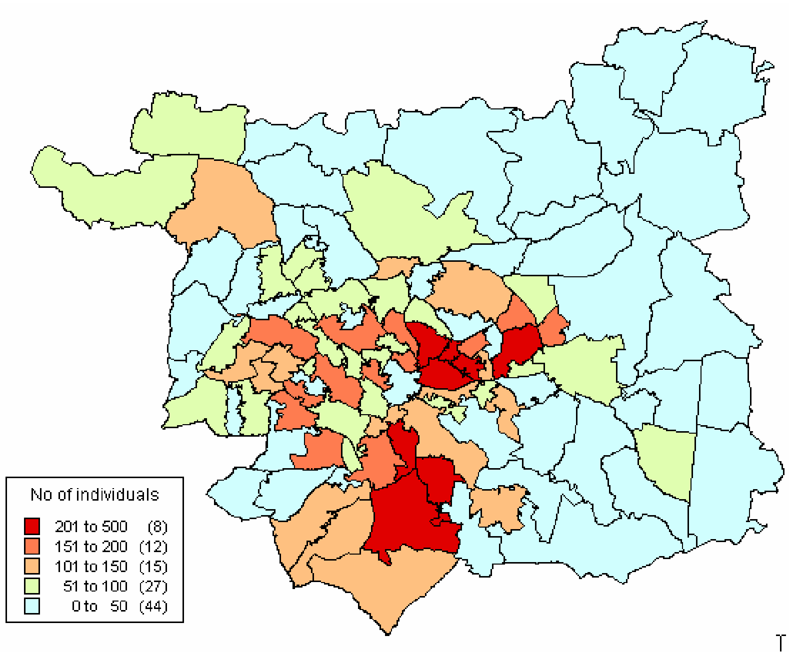


Figure 7. Simulated distribution of female lone parents with dependent children in households with low income by community area, 2001

The second query identifies children aged 0 to 15 in households where income is less than £10,000. In Figure 8, the largest numbers are again found to the east of the city centre in community areas like Halton Moor, Harehills, Gipton North, as well as Seacroft South, with pockets also in Cottingley and Belle Isle in the south

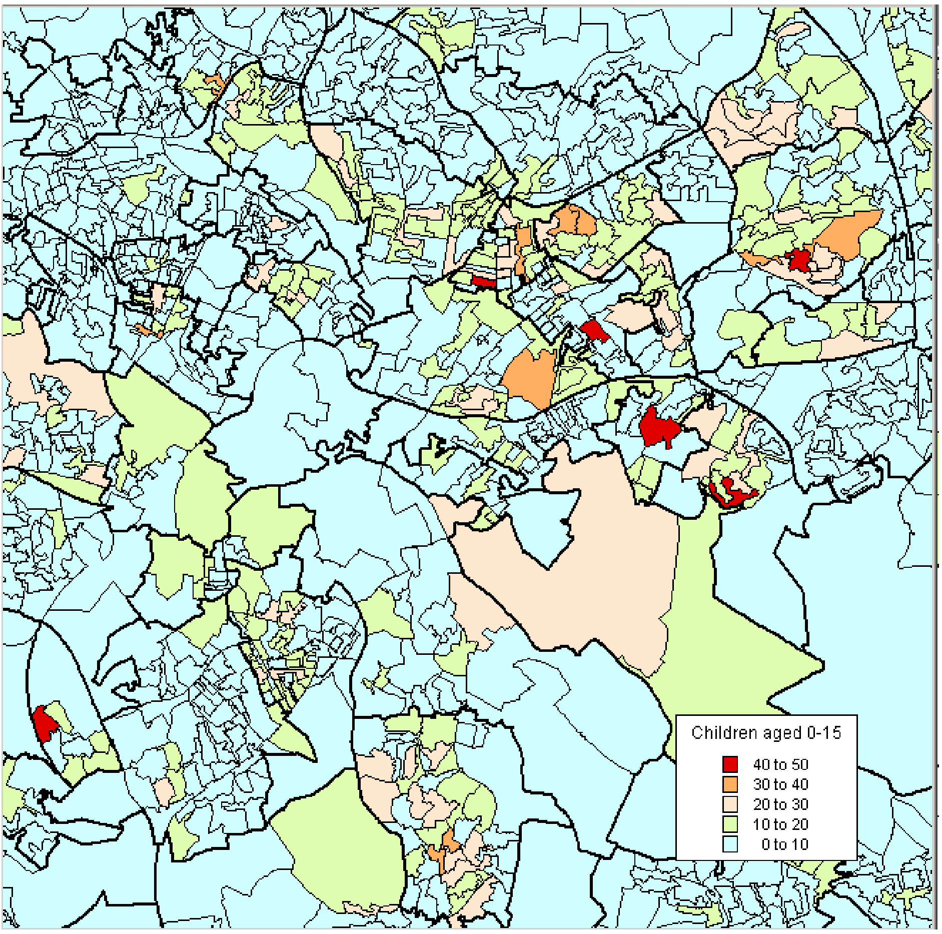


Figure 8. Simulated distribution of children in households with low income by output area, 2001

The third example involves individuals at the other end of the social spectrum. Figure 9 shows the distribution by output area across the whole of the metropolitan district of those individuals living in two person households with no dependent children and with a household income in excess of £50,000. These are the so-called DINKies and their spatial incidence is observed across the north of the district, particularly in the outer suburbs of Harewood and Arthington and Pool in the north and Calverley in the west, but also in pockets in the south.

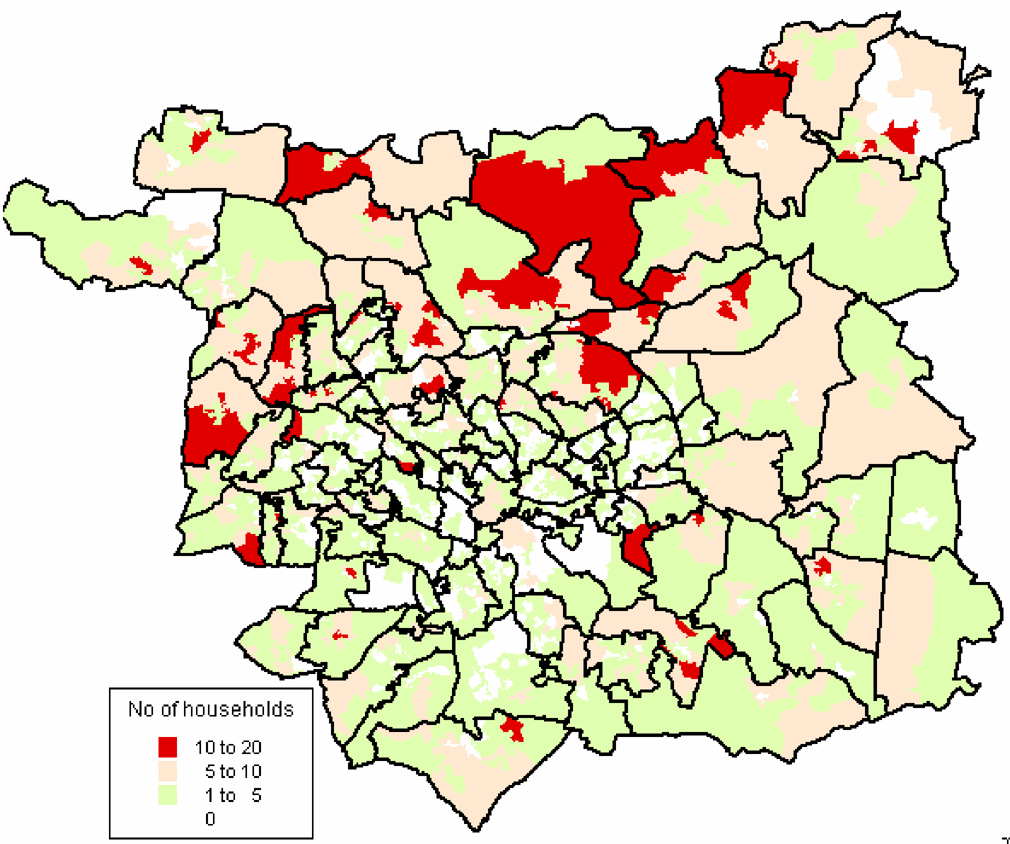


Figure 9. Simulated distribution of two person households without dependent children with high household income, 2001

In addition to modelling the population in 2001, microsimulation can be used to predict the characteristics of the Leeds population in future years. The method we adopt is to build a crude projection model for aggregate ward populations and then to use these totals to constrain more detailed projections generated by the MicroMaPPAS model. In other words, we apply the MicroMaPPAS reweighting methodology to readjust the weights of BHPS households so that they fit small area constraint data in any selected year.

It is important to recognise that, currently, there is no provision of small area projections by a single agency in Britain. The ONS is responsible for creating sub-national population projections for England but these are only produced for local authority areas and on a relatively infrequent basis. A comparison of the projected population for Leeds in 2001 from the latest (1996-based) set of ONS sub-national projections (ONS, 1999) with the mid-year estimate for 2001 indicates an over-projection of nearly 16,000 persons overall. Figure 10 shows the differences between projected counts and estimates for males, females and persons; most of the over-projection has occurred for males and females (to a lesser extent) aged 35 to 54 and these differences offset under-projections of females in their 20s.

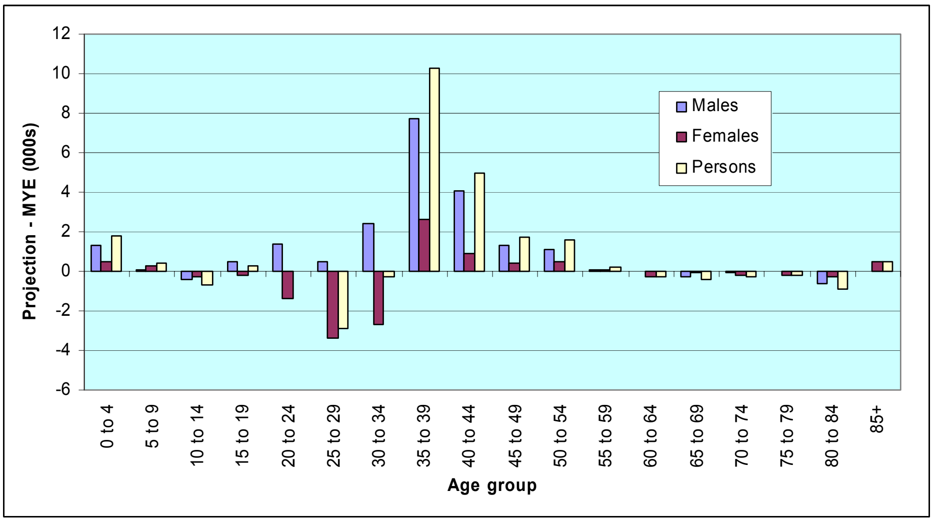


Figure 10. A comparison of 1996-based projections and mid-year estimates for Leeds in 2001

The lessons we take from the comparison are to recognize that even short term projections for relatively large areas may be error-prone and it might not be wise to use the 1996-based district projections as constraints. Consequently, the first step in our methodology is to project the numbers of households and individuals for every year into 2021 by ward within the district independently. In order to do this, we make the simple assumption that the annual rate of change between 1991 and 2001 will continue until 2021. Therefore, we calculate for each ward the annual rates of change between 1991 and 2001 for households and individuals and apply these to successive years after 2001 to give ward-based population totals up to 2021.

The projection of the disaggregated counts of the population in future years according to marital status, socio-economic group, number of cars owned, *et cetera*, is undertaken by applying annual rates of change between 1991 and 2001 in exactly the same way as with the aggregate populations. The projection of these counts is necessary given that these variables are used as constraints in the simulated annealing household reweighting procedure. So, for example, having projected the car ownership characteristics by household, the next step involves calculating the proportions of all car ownership categories in each ward. These proportions are then applied to the projected numbers of households by ward in each future year. In this way, we ensure that the sum of all cars by household categories adds up to the aggregate household projection. The same method can be applied for all other household (e.g tenure) and individual (e.g. ethnic group) variables.

The above discussion implicitly assumes that the 1991 Census recorded accurately the populations living in Leeds wards. However, over 2% of the population was missed overall in 1991 and this underenumeration did not occur uniformly across all areas or age-sex groups (ONS, 2003). Further, the 1991 Census did not record the number of students, which is quite large in some electoral wards such as Headingley. In order to deal with these problems, the following strategies were adopted. To tackle the problem of the undercount in 1991, the ward populations in 1991 can be readjusted on the basis of alternative assumptions on the extent of the undercount. For instance, if it is assumed that the Leeds population in 1991 was underestimated by 4%, the 1991 population numbers can be increased by this rate and the projection procedure described above is applied using the annual rates of change recomputed on this basis. A reasonable solution to the problem of not counting the students in the 1991 Census is to estimate their numbers on the basis of 2001 proportions. For instance, according to the 2001 Census, students in Headingley comprise 54.5% of the total population and this proportion can be added to the published 1991 population total.

5. 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. We also presented some simulation outputs which demonstrate the policy relevance of the system. We believe that Micro-MaPPAS 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-MaPPAS 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). The authors would like to thank Jianhui Jin at the School of Geography, University of Leeds for his contribution to the development of the computer code for the Micro-MaPPAS system.

REFERENCES

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. 2003, Microsimulation and Regional Science: 30 years of spatial microsimulation of populations, paper presented at the *50th Annual North American Meeting of the Regional Science Association International*, Philadelphia, USA, 19-22 November 2003.

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 Applications*, 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

Clarke, G.P. 1996, Microsimulation: an introduction, in G.P. Clarke (ed.) *Microsimulation for Urban and Regional Policy Analysis,* Pion, London, 1-9

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

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

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

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.

Orcutt, G.H. 1957, A new type of socio-economic system, *The review of economics and statistics*, *39*, 116-123

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.