MOSES: Modelling and Simulation for e-Social Science

Mark Birkin[1], Haibo Chen[2], Martin Clarke[1], Justin Keen[3], Phil Rees[1], Jie Xu[4]

[1] School of Geography, University of Leeds, Leeds LS2 9JT, UK.

[2] Institute for Transport Studies, University of Leeds

[3] Nuffield Institute for Healthcare Studies, University of Leeds

[4] School of Computing, University of Leeds

Email address of corresponding author: m.h.birkin@leeds.ac.uk

Abstract. The objective of the project is to develop representation of the entire UK population as individuals and households, together with a package of modelling tools which allows specific research and policy questions to be addressed. Our project seeks to combine methodological advances within core social science domains, particularly geography, with the development and application of a sophisticated computational infrastructure for decision support systems using the latest technology for Grid computing. The advances which we will seek to achieve include the creation of a dynamic, real-time, individually-based demographic forecasting model; for defined policy scenarios, to facilitate integration of data and reporting services, including GIS, with modelling, forecasting and optimisation tools, based on a secure grid services architecture; to use hybrid agent-based simulations to articulate the connections between individual level and structural change in social systems; and to provide high level capability for the articulation of unique evidence-based user scenarios for social research and policy analysis.

1. Background

This paper describes an e-Social Science research programme at the University of Leeds with a specific focus on Modelling and Simulation (MOSES – Modelling and Simulation for e-Social Science). The cornerstone of the programme is the creation of a dynamic simulation model of the UK population, represented as a series of richly disaggregated individuals and households. We aim to use the power of e-Science technologies to deliver a complete representation of the population which draws on attributes from a diverse portfolio of databases. The simulation model will be applied to address research questions in three social science domains, relating to healthcare policy and practice, transport & environmental sustainability, and the business impacts of socio-demographic change.

Social modelling and simulation has had a somewhat chequered history as both an academic and applied discipline, ever since Lee's (1973) damning early critique. And yet the value of scenario-based approaches to policy development, evaluation, planning and research remains widely recognised (e.g. Masser et al, 1992; Ravetz, 2000). The advent of e-Social Science provides an obvious impetus for a re-evaluation. Modelling approaches provide an effective means for the integration of data from diverse sources; they provide the means by which social processes can be not only hypothecated, but rounded in hard evidence (evidence-based social science); they are of direct value not just to researchers but also to policy-makers and commercial users; and industrial-strength simulations are dependent on high performance computational technology.

2 Methodology

2.1 Base populations

The objective of the baseline modelling component is to represent the UK population at the level of individuals and households for the base year 2001. In other words, we wish to create a database of 24.6 million households and 59 million people which simulates the UK population. The starting point for the modelling process is the 2001 Sample of Anonymised Records. The SARs are individual selections from the 2001 census records which are anonymised through the incorporation of high level spatial codes, so that each individual can be located only to the detail of a Local Authority Area or Government Office Region. The integrity of the data is not compromised through any other form of randomisation or corruption.

2.2 Forecasting

The objective of this stream of the research is to provide enhanced long-term value of the results by ensuring that the results are updateable rather than instantaneously obsolete, and to provide a suitable backcloth for further policy testing and evaluation.

The key research question in this section is how to link the future population of an area in (say) five years time, to its current population. There are two basic approaches to this question, based on dynamics and comparative statics. Dynamic approaches rely on the application of models and rules relating to demographic and economic change, which are of varying complexity e.g. ageing, probability of marriage, parenthood, entry into the labour market, and so on. Comparative statics is an extension of the synthetic approach to population creation. Aggregate variables are projected on the basis of expected trends, and a completely new future population is created, which can then be differenced from the current population in the assessment of change. (Again a good recent example is the work of Ballas et al, 2004).

Complete projections of the UK population will be provided at five year intervals from 2006 to 2031. We also plan to generate two long range scenarios (2051, 2101) for use in scenarios involving long-term trends (environmental change, super-ageing of the population).

2.3 Incorporation of new variables

In this part of the research, the objective is to extend the coverage of the population model through the addition of data from third party sources. We will use specific examples drawn from the domains of health, business and transport.

The process of merging micro-databases is well-established for individual applications (e.g. Orcutt et al, 1986). The procedures simply demand a knowledge of the common features of two databases, such that a linkage can be established. For example, if a survey database shows that people vote according to their age, occupation and family status, and age, occupation and family status are all contained within a micro-database, then the addition of likely voting behaviour to the micro-database is straightforward. In developing a general methodology, however, a number of problems will need to be overcome, regarding the general description of the database structures and content (meta-data), consistencies in the timing and definition of variables, and the spatial resolution at which data is captured, stored and analysed.

3 Applications

In the substantive part of this research, the robustness of this methodology will be tested through the incorporation of information from a variety of sources.

3.1 Business Applications

In this application area we propose building a model that sits on top of the individual and household microsimulation model and simulates the effects of a number of critical personal financial service events and scenarios to examine their potential impact both at a national but also, importantly, at a local level. The events we propose exploring relate to the increased level of personal indebtedness in the UK. Latest Bank of England estimates suggest that personal indebtedness have reached £1 trillion, equivalent to annual GDP. Several key factors come in to play looking towards the future: for example, ‘the pensions timebomb’ and relating to this the increased use of Equity Release Products to generate annuity incomes; the reduction of inter-generational transmission of wealth; potential crash in house prices; potential rise in interest rates; and increases in the rate of household formation (more smaller households).

We propose building a simulation model that would explore the interdependencies of these potential events over the next decade. We believe that the impacts of a fall in house prices/Increase in interest rates will have substantively different impacts in different regions/localities of the UK that the simulation model should be able to detect and predict.

3.2 Transport Applications

This research stream aims to explore the environmental consequences of demographic change, and in particular the impact of the evolution of populations and their social and economic behaviour on the sustainability of urban environments. For a variety of UK cities, scenarios for demographic and social change will be used to simulate and forecast traffic change and thus the impacts on congestion and emissions episodes and hotspots which have significant effects on health.

3.3 Health care applications

In this research strand, we will use Grid technology to combine population-based data held on different servers across Bradford, which it would otherwise be impractical to merge. Specifically, we will seek to investigate differences in access to services for people with different clinical conditions, e.g. diabetes, coronary heart disease. The objective of this part of the project would be to merge individual patient data with synthetic demographic data from the micro-population database. Although the demographic data is very fine-grained its synthetic character provides a basis for highly realistic modelling without prejudicing the confidentiality of individual patients. We see this technology as a potentially high value resource for the anonymisation of high value but sensitive National Health Service data resources. These data will be cross-tabulated with demographic data, for example from the General Household Survey, to establish relationships between household status and morbidity. The simulation modelling capability will also be used to forecast future changes in patient morbidity.

3.4 Generic social science applications

We are now in a position to add an applications layer to our description of the modelling process. In order to build problem-focused applications on top of the microsimulation model, two types of inputs are required – data about individual characteristics and behaviours (morbidity, propensity to own a personal pension, preferred mode of transport) and information relating to infrastructure and service provision (hospital treatment rates, house price data, trip cost by mode).

There seems no reason why a general model of this type might not be applied to a wide range of problems. For example, a user with an interest in crime patterns might access data on propensities to commit crimes (or the likelihood of falling victim to crime) together with intelligence relating to the crimes reported to various local police forces. This could lead to a model which allows the effectiveness of crime prevention to be benchmarked.

References

Ballas, D, Rossiter, D, Thomas, B, Clarke, G P and Dorling D (2004), *Geography matters: simulating the local impacts of national social policies*, Joseph Rowntree Foundation, York.

Lee, D.B. (1973) Requiem for Large Scale Urban Modelling, *Journal of the American Institute of Planners*, 39, 163-178.

Masser I. Suiden O., Wegener M. (1992) *The geography of Europe's futures*, Belhaven, London.

Orcutt, G.H., Mertz J., Quinke, H. (eds) (1986), *Microanalytic Simulation Models to Support Social and Financial Policy*, North-Holland, Amsterdam

Williamson, P. (1996), Community care policies for the elderly, 1981 and 1991: a microsimulation approach, in G.P. Clarke (ed.) *Microsimulation for Urban and Regional Policy Analysis*, Pion, London, 64-87

Zedlewski, S. (1990) The development of the Dynamic Simulation of Income Model (DYNASIM), in Lewis, G. and Michel, S. (eds.) Microsimulation Techniques for Tax and Transfer Analysis. Washington, D.C.: The Urban Institute Press, 109-136.