An Agent-based Microsimulation Framework For Investigating Residential Segregation Using Census Data

Mahdavi, B. [1, 3], D. O'Sullivan [1, 3] and P. Davis [2,3]

[1] School of Geography, Geology, and Environmental Science, The University of Auckland, New Zealand
[2] Department of Sociology, The University of Auckland, New Zealand
[3] Social Statistics Research Group, The University of Auckland, New Zealand
Email: b.mahdavi@auckland.ac.nz

Keywords: *Residential segregation, agent-based modelling, microsimulation, simulation*

EXTENDED ABSTRACT

Social and urban structures are inherently complex and multidimensional. They do not arise from a single process, but from the multifaceted interplay of different social and economic processes. This is true for racial and residential/neighbourhood segregation which is the result of many factors, and its causes are not well understood. Over years, researchers have studied and measured the role and influence of different factors implicated in residential segregation, using traditional techniques and tools, mostly in social science. These include empirical studies and statistical analyses of survey and census data. Such complex and nonlinear systems are difficult to study analytically. To better understand the complexity of the social world, building a model – a smaller, less detailed and less complex version of the real world (structure, system) – is recognized as an indispensable tool. Simulation is an excellent way of modelling which results to a better understanding of social processes and features of the social world. Simulation can be used to test theories and the result of simulation will often be the development of further theory.

The two most appropriate simulation modelling approaches for studying residential segregation are microsimulation and agent-based simulation.

Microsimulation (Gilbert and Troitzsch 1999) is based on a large random sample of population (e.g. individuals, households, firms, etc.), treated as a unit. Transition probabilities determine the chance that units will undergo particular changes as they "age". Microsimulation is not intended as an explanatory tool, but as a means of predicting the future behaviour. Aggregate statistics can be calculated and used as estimates of the future characteristics of the population. Microsimulation does not attempt to model interaction between units, nor attempt to contain the motivations or intentions of the units. A random mechanistic process (hence its historical link to stochastic processes) takes care of the unit "aging" task. Moreover, microsimulations are developed with validation ambitions.

By contrast, agent-based simulation (Gilbert and Troitzsch 1999) offers an environment of autonomous individuals (agents), where interaction between agents is permitted, and in which agents can assess information (sometimes in the light of their past experience) before taking their next action (hence its historical link to artificial intelligence). They usually do not have any validation ambitions.

None of these simulation approaches alone is sufficiently good for an empirically informed model (based on census data) with validation ambitions and with the aim of exploring linkage between the social micro (e.g. household) and macro levels (e.g. social structure expressed through demographic patterns in geographical space). Therefore, on the promise of enabling such a micro-macro exploration, and to better understand the complex and dynamic world of residential segregation, where different scenarios (in light of social theories from geography and sociology) can be tested, and validated using census data, a hybrid agent-based microsimulation framework is proposed, which can be used as a test-bed for future modelling research, addressing theoretical issues in urban geography and sociology.

The proposed framework offers a more comprehensive approach, where it will be possible to investigate the interaction between households, households and their environments, the relationship between attributes and behaviour of individuals and the reverse effects – the impact of social structure on individual households, and the global properties of social groups – hence, operationalising micro-macro linkages to improve understanding of residential segregation.

1. INTRODUCTION

The phenomenon of residential segregation is a good example of the inherently multidimensional nature of the urban spatial structure. It is the result of many factors, and moreover, "the causes of segregation are not well understood" (Bruch and Mare 2006). Figure 1 illustrates the range of various factors that play potential roles in the dynamics of distribution of population and hence of residential segregation. Some of these factors are at an individual level, including fixed no-choice factors such as ethnicity, age, and sex. Others may be related to different types of preferences. Choices may also be influenced by attributes related to schools or work places. On the other side of the spectrum, the institutional or communal forces such as government or municipality policies or different forms of (institutional) discrimination play an important role in segregation. For instance, Kempen and Özüekren (1998) expand on the role of the government, the influence of the context and other macro-developments (e.g., economics, cultural factors, demographics, and politics).

In the debate about the relative role of these forces, the consensus is that patterns of separation have a multifaceted explanation; no one factor explains those patterns (Clark 1991). In other words, "urban problems, including those relating to race and housing, should not be treated in isolation from society at large but are best understood as the result of a complex inter-relationship between many processes" (Sarre *et al.* 1989). In fact, the question of whether residential segregation causes poverty, or other socio-economic factors generate residential segregation still remains and is the subject of many studies. For instance, Small (2007) found that "while residential segregation has created conditions in which some races are more likely to live in high-poverty neighbourhoods, it is the poverty, not the racial composition, of the neighbourhoods that is significantly associated with weaker social ties".

The complexity of the system, as seen in Figure 1, is related to what is also described as the theory of self-organization. Self-organized systems are complex and the factors involved are so numerous that there is no technical way to establish a causal relationship among them, and their parts and components are interconnected in a nonlinear fashion by a complex network of feedback loops (Portugali 2000). "In certain situations external forces acting on the system do not determine/cause its behaviour, but instead trigger an internal and independent process by which the system spontaneously self-organizes itself" (ibid: 50).

Over the years, researchers have studied and measured the role and influence of different factors, using traditional techniques, methods and tools, mostly in social science. These include empirical studies and statistical analyses of survey and census data. For instance, sociologists have investigated the willingness of different ethnic groups to live in neighbourhoods of varying race-ethnic composition, using survey questions (Bruch and Mare 2006). However, complex and nonlinear systems, such as residential segregation are difficult to study analytically.

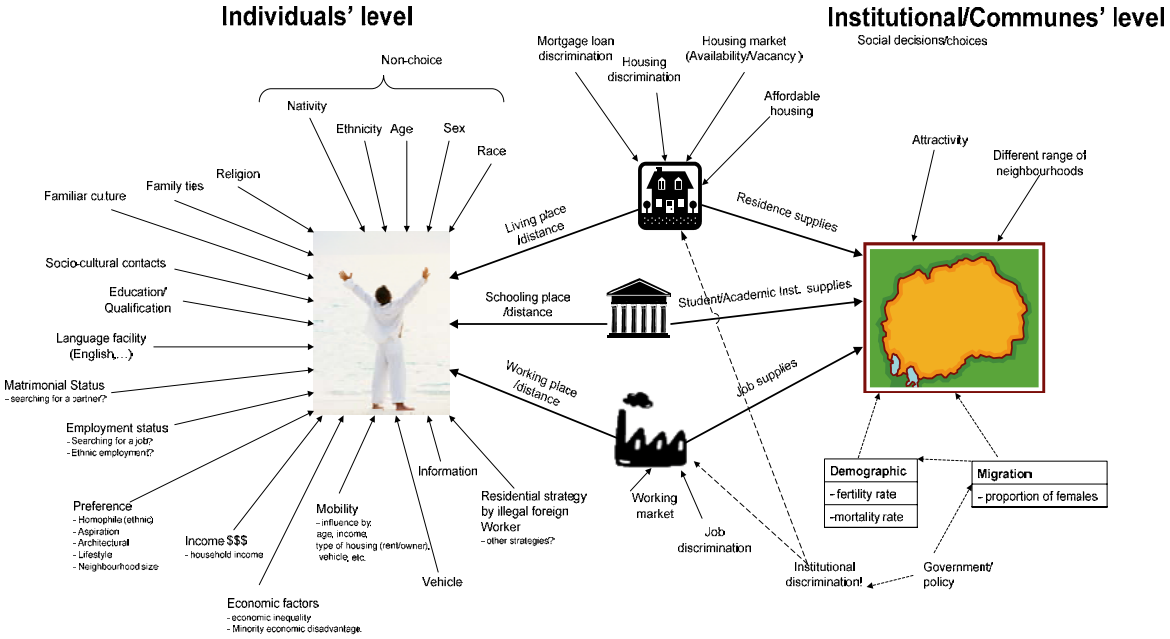

Figure 1. Factors in the segregation dynamic

On the other hand, the study of segregation using dynamic models is a more recent endeavour and promises new perspectives on this issue.

Pioneering work by Thomas Schelling (1969, 1971) presents seminal examples of an abstract study of the interactive dynamics of (discriminatory) individual choices showing that segregation can result from individual preferences. Schelling also set the foundation for one of the most influential agent-based models, "regarded as the seminal example of multi-agent simulation in social science" (Macy and Willer 2002), where it is possible to investigate how micro-level residential choice behaviour can produce aggregate-level patterns of ethnic residential segregation.

Many agent-based research simulation have employed or extended the Schelling model, including, for example, work by Bruch and Mare (2006), Laurie and Jaggi (2003), and Fossett and Waren (2005). On the other hand, Kempen and Özüekren (1998) argue that "any explanation [of residential segregation] should take a wider perspective than the traditional focus on the choices and constraints of households and individuals".

However, agent-based simulations are not the ideal tools for investigating the linkage between micro level (e.g. the relationship between attributes and behaviour of individuals) and macro level (e.g. the global properties of social groups). They "often tend strongly towards methodological individualism …. [which] is accompanied by a one-way notion of emergence: the social can emerge from the individual but not the other way round" (O'Sullivan and Haklay 2000). Moreover, they do not easily support validation.

Mirosimulations offer another possible way to approach the modelling of individual and household dynamics, and have been particularly used for policy analysis. Major attempts to simulate spatially and dynamically the population (although not for the purpose of examining residential segregation) using microsimulation include CORSIM (Caldwell et al. 1998), SVERIGE (Vencatasawmy et al. 1999), and SimBritain (Ballas et al. 2005).

Although microsimulations are developed with validation ambitions (Boman and Holm 2004), they have other limitations. They are not the best approach for explanatory purposes, but rather tools for predicting or estimating the future behaviour (hence their applications for policy analysis). Often aggregate statistics are calculated and used as estimates of the future characteristics of the population. Microsimulations neither attempt to model interaction between units, nor attempt to contain the motivations or intentions of the units (Gilbert and Troitzsch 1999). Table 1 shows a more detailed comparison between microsimulation and agent-based simulation.

Table 1. Comparison between dynamic microsimulation and agent-based simulation

Characteristic	Microsimulation	Agent-based
Origin	Stochastic process	Artificial intelligence
Main purpose	Projection	Explanation
Research approach	More quantitative, Deductive: infer from aggregate behavior to individual agents	More qualitative, Inductive: infer from individual agents behavior to aggregate behavior
Investigation focus	Aggregate, Trends	Behavioral, Emergence
Applications	Public policy, Prediction, …	Theory, Contrast, …
Results	More realistic	More explanatory
Building blocks	Micro-units (individual agents and aggregate entities)	Individual agents connected by feedback loops
Object of interest	Aggregate micro-units, Trends	(Connection between) agents' rules, Emergence
Number of possible interaction between agents	2	2+
Number of agents	Many	Fewer
Complexity of agents	Average	High
Communication between agents	~No	~Yes
Development of object of interest over time	More fixed structure ("Aging" through transition probabilities)	More flexible structure (Agents' rules interaction)
Handling of time/state	Continuous/ Discrete	Discrete/ Continuous
State update	Synchronous/ Asynchronous	Synchronous/ Asynchronous
Validation	~Yes (data-based / data-driven)	~No ("mock" data)

This paper develops the idea of a hybrid agent-based microsimulation framework where exploring linkage between the social micro (e.g. household) and macro levels (e.g. social structure expressed through demographic patterns in geographical space) is possible, and wherein census-based validation will also be possible.

Little attention has been paid to verification of simulation models informed by empirical data from census in the past. Related investigations based on census data do not use a simulation approach, or those using simulation do not exploit census data. There are a variety of reasons behind this, among them, for instance, Logan and Zhang (2004) state that there exists no consensus among social scientists about how to define an ethnic neighbourhood!

2. PROPOSED FRAMEWORK

There are significant methodological challenges faced by any study of changing patterns of residential location, particularly given the inter-, multi-, and trans-disciplinary nature of residential segregation.

Figure 2, embodies core aspects of the proposed model design and approach and illustrates an original agent-based microsimulation model which seeks to explore the linkage between macro and micro agent elements.

In contrast to the autonomous rule-based individual setting in agent-based modelling (base layer) - ideal for examining the interaction between individuals and their environments, microsimulation (middle layer) treats a sample (usually a large random sample) of the population as the unit (usually in the form of macro-unit such as households, firms, or other institutions), without attempting to model interaction between units, or to encompass the motivations or intentions of the units. A set of transition probabilities determine the chance that the unit will undergo some change during the unit "aging" process. This layer attempts to predict the future behaviour and characteristics of the population, while the purpose of the lower layer (agent-based) is to understand and explain these behaviours and characteristics.

On the top layer, we see the incarnation of the phenomenon that Portugali (2000) calls 'circular causality'. Individuals (people, households, institutions, and so on) enslaved by society/city/neighbourhood's 'order parameter' (ibid: 3) – which is similar to the 'mode of production'. "The concept of order parameters and their relationships to the individual part of the system, a relationship governed by circular causality, applies to a great variety of phenomena in society" (ibid: 56). Figure 2 also depicts micro and macro forces and their relationships. "The individuals are the parts of a … society and determine its [microscopic and] macroscopic manifestations, such as language, religion, form of government, culture, educational system, city structure and so on" (ibid: 56). These micro- and macro-scopic manifestations – which play the role of order parameters – will also determine the behaviour of individuals in turn. As can be seen, some macro components can be aggregates of smaller micro elements – although this is not necessarily the case.

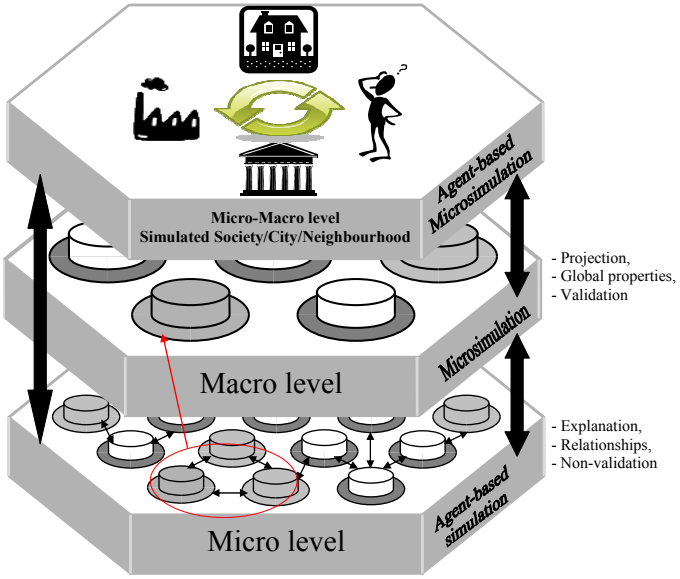

Figure 2. Core components of the hybrid agent-based microsimulation framework

As seen in the base layer, interactions between agents are possible, while the middle layer is concerned with the behaviour of more aggregated micro-units (macro), where the validation will happen.

The proposed approach offers a more comprehensive model by combining various settings, from behavioural hypotheses and relationships with regard to the facets that compromise the different scenarios need to be explored. This enables a thorough investigation; where it will be possible to investigate the interaction between households, households and their environments, the relationship between attributes and behaviour of individuals and the reverse effect – the impact of social structure on individual households, and the global properties of social groups – hence, operationalising the micro and macro linkage for a better understanding of residential segregation dynamism and its features.

O'Sullivan (2004b) argues that "determining what data are available must be one of the first questions addressed in model building". It is also emphasised that models should be exercises in exploring particular theories and that questions at the heart of complexity science such as which

theories are represented and how and why not, should be explained and then examined and evaluated in their own terms (O'Sullivan 2004a). A series of scenarios (informed by theories in geography and sociology) can be tested, in consideration with the data-led nature of the model and research. An attempt to capture this plan in a framework is depicted in Figure 3.

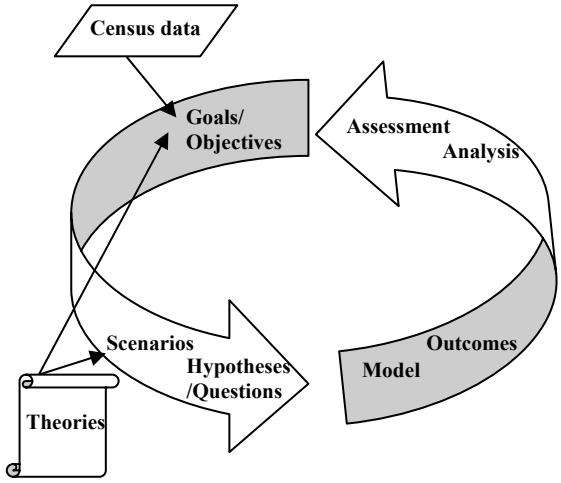

Figure 3. Multiple scenarios can be tested in data-led model based on census data

One important aspect illustrated in figure 3 is the fact that the nature of census data in the dataset, that is, the available variables, influences directly the goals and objectives, while the scenarios to be tested are subject to theories from geography and sociology. Examples of such theories include Weber's theory of the role of the individual, Schelling's theory on the role of preference and choice, and Durkheim's theory about aggregate social forces in society.

As seen in the circular feature of the methodology, analyses and assessment of the outcomes can inform other parts of the cycle, which in turn will entail adjustment or modification of scenarios and hypotheses to be re-feed again into the model for reassessment. In other words, we can learn from the result and improve their effectiveness.

Another perspective is depicted in Figure 4. This shows that execution of each scenario will take place via a preset but flexible environment of selected key variables from the census dataset, while examination can be carried out by methodical changes to parameters and through comparison of the results and advanced analysis for discerning interesting behaviours and patterns. Aggregate effects are obtained by combining individual simulations - that is, synthesising from relatively simple parts. Findings will comprise knowledge about specific scenarios (theory) gained from this investigation.

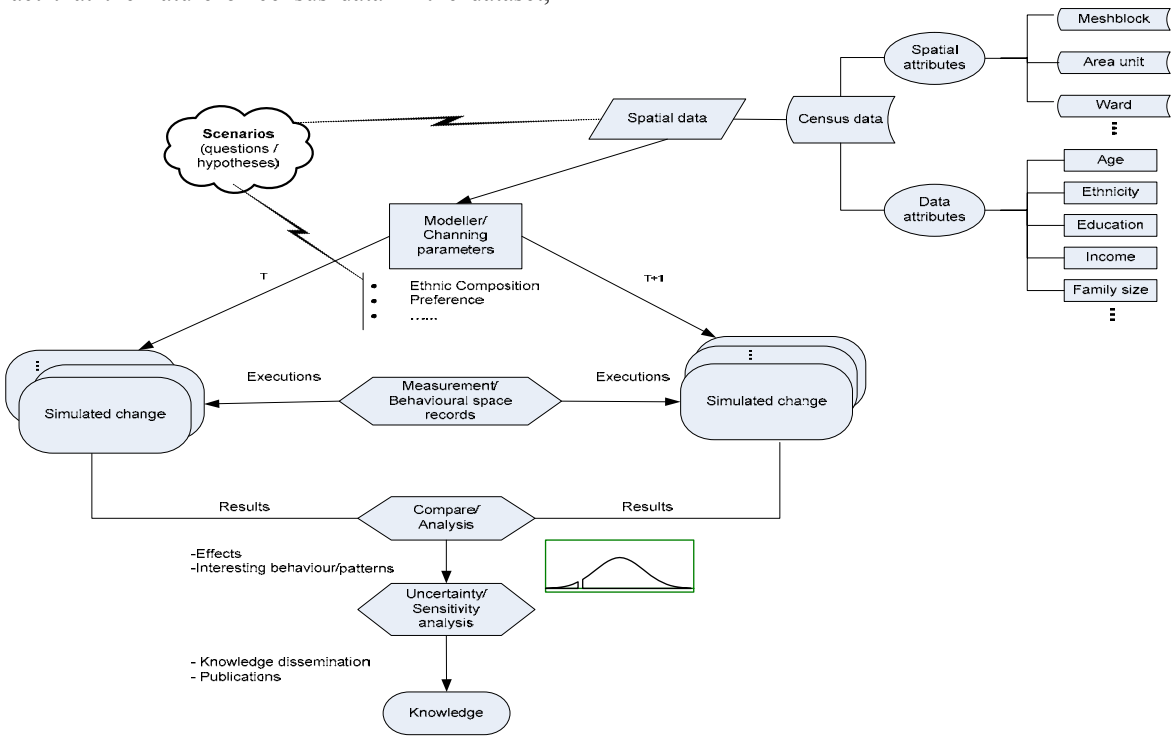

Figure 4. Executive view of the research framework

3. CONCLUSION

The causes, patterns, and consequences of urban residential segregation by ethnicity and socio-economic status are important topics in contemporary urban geography and sociology. Segregation is important because of its potential impacts on issues such as access to education and healthcare opportunities. However, there are significant methodological challenges faced by any study of changing patterns of residential location. Studies that can describe and explain the details of such processes of transformation, at the level of both individual household decision making, and of aggregate demographic shifts over time, are of great value for urban geographers and sociologists in understanding neighbourhood formation and social stratification.

Simulation is the most pertinent way to distinguish between various forces and evaluation of the intensity and character of their influence considering the complex, multi-dimensional, temporal and spatial aspects of residential segregation. This paper builds on the potential for a novel investigation of residential segregation using a hybrid agent-based microsimulation framework, informed and validated by empirical census data. The proposed framework is distinctive in its aims, objectives and methodologies from previously attempted works. It is an amalgam of agent-based model and spatial dynamic micosimulation, which can be used to investigate variety of scenarios, in particular, the dynamism between micro and macro, along with scenario analysis using census data.

The goal is to overcome shortcomings of both agent-based and microsimulation methods in regards to a more comprehensive study of interactions between micro and macro forces, along with an empirical census-based validation.

This will be promising in the context of available valuable (and expensive) census data, along with the relative lack of research on segregation (using simulation method) in the world, and particularly in New Zealand.

4. REFERENCES

Ballas, D., G. Clarke, D. Dorling, H. Eyre, B. Thomas, and D. Rossiter (2005), SimBritain: a spatial microsimulation approach to population dynamics, *Population, Space and Place*, 11, 13-34.

Boman, M., and E. Holm (2004), Multi-Agent Systems, Time Geography, and Microsimulations, In Olsson, M. and G. Sjöstedt (Eds.), Systems approaches and their application.

Bruch, E. E., and R. D. Mare (2006), Neighborhood choice and neighborhood change, *The American Journal of Sociology*, 112(3), 667.

Caldwell, S. B., G. P. Clarke, and L. A. Keister (1998), Modelling regional changes in US household income and wealth: a research agenda, *Environment and Planning C: Government and Policy*, 16(6), 707-722.

Clark, W. A. V. (1991), Residential preferences and neigborhood racial segregation: a test of the Schelling segregation model, *Demography*, 28(1), 1-9.

Fossett, M., and W. Waren (2005), Overlooked implications of ethnic preferences for residential segregation in Agent-based models, *Urban Studies*, 42(11), 1893-1917.

Gilbert, N., and K. G. Troitzsch (1999), Simulation for the social scientist: Open University Press.

Kempen, R. V., and A. S. Özüekren (1998), Ethnic segregation in cities: new forms and explanations in a dynamic world, *Urban Studies*, 35(10), 1631-1656.

Laurie, A. J., and N. K. Jaggi (2003), Role of "vision" in neighbourhood racial segregation: a variant of the Schelling segregation model, *Urban Studies*, 40(13), 2687-2704.

Logan, J. R., and W. Zhang (2004), Identifying ethnic neighbourhoods with census data, In Goodchild M. F., and D. G. Janelle (Eds.), Spatially Integrated Social Science (pp. 113-126): Oxford.

Macy, M. W., and R. Willer (2002), From factors to actors: computational sociology and agent-based modeling, *Annual Review of Sociology*, 28.

O'Sullivan, D., and M. Haklay (2000), Agent-based models and individualism: is the world agent-based?, *Environment and Planning*, 32(8), 1409-1425.

O'Sullivan, D. (2004a), Complexity science and human geography, *Transactions of the*

Institute of British Geographers, 29(3), 282-295.

O'Sullivan, D. (2004b), Too much of the wrong kind of data: implications for the practice of micro-scale spatial modelling, In Goodchild M. F. & D. Janelle (Eds.), Spatially integrated social science (pp. 95-107): Oxford University Press.

Portugali, J. (2000), Self-organization and the city: Springer.

Sarre, P., D. Phillips, and R. Skellington (1989), Ethnic minority housing: explanations and policies: Avebury.

Schelling, T. C. (1969), Models of segregation, *The American Economic Review*, 59, 488-493.

Schelling, T. C. (1971), Dynamic models of segregation, *Journal of Mathematical Sociology*, 1, 143-186.

Small, M. L. (2007), Racial differences in networks: do neighborhood conditions matter?, *Social Science Quarterly*, 88(2), 320-343.

Vencatasawmy, C. P., E. Holm, T. Rephann, J. Esko, N. Swan, M. Öhman, et al. (1999), Building a spatial microsimulation model, Paper presented at the 11th European Colloquium on Quantitative and Theoretical Geography, Durham, England.