

# A hybrid microsimulation model for a UK city population with dynamic, spatial and agent based features

Mingqing Belinda Wu

Submitted in accordance with the requirements for  
the degree of Doctor of Philosophy

School of Geography

University of Leeds

28 August 2012

The candidate confirms that the work submitted is her own, except where work which has formed part of jointly-authored publications has been included. The contribution of the candidate and the other authors to this work has been explicitly indicated below. The candidate confirms that appropriate credit has been given within the thesis where reference has been made to the work of others.

This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgements.

# Acknowledgement of Funding and Data

This PhD uses part of the work by the author when she worked as a researcher in two ESRC funded research projects Moses (Modelling and Simulation of e-Social Science, ref: RES-149-25-0034) and Genesis (Generative e-Social Science for Socio-Spatial Simulation, ref: RES-149-25-1078). The baseline population data for year 2001 used in the initialisation of the simulation are provided by Prof. Mark Birkin, who recreated the individual records using data from the 2001 Census Samples of Anonymous Records (SAR) and British Household Panel Survey (BHPS). However, this project focuses on the dynamic simulation of the population only and is not concerned with the recreation of the synthetic population.

The 2001 Census statistics used in this thesis are Crown Copyright produced by the Office for National Statistics (ONS). The data are licensed for academic use by the Economic and Social Research Council (ESRC) and the Joint Information Systems Committee (JISC) Census Programme, which funded free access to the data for researchers in the UK. The data service providers are the Census Dissemination Unit through the Manchester Information and Associated Services (MIMAS) of Manchester Computing, University of Manchester, through an interface called CASWEB. Census output is Crown copyright and is reproduced with the permission of the Controller of HMSO and the Queen's Printer for Scotland.

The boundary data used in this thesis are Crown Copyright produced by the Office for National Statistics (ONS). These data are licensed for academic use by the Economic and Social Research Council (ESRC) with the support of the Joint Information Systems Committee (JISC), which funded free access to the data for researchers in the UK. The data service provider is the Census Geography Data Unit (UK Boundary Outline and Reference

Database for Education and Research Study: UKBORDERS) via Edinburgh University Data Library (EDINA).

The data from the Health Survey for England are used in the obesity application model through the UK Data Archive (UKDA), licensed for academic use.

Chapter 6 is based on two previous journal publications (publications 1 and 2 in the list below) and Chapter 2 (publication 3) uses materials from a book chapter, where the author is the first author and the contribution of the other named authors, Mark Birkin and Phil Rees, were editorial and advisory.

Cited publications:

1. Wu, B. M., Birkin, M. H. and Rees, P. H. (2008) A spatial microsimulation model with student agents, *Journal of Computers, Environment and Urban Systems*, 32, 440-453, DOI: 10.1016/j.compenvurbsys.2008.09.013.
2. Wu, B. M., Birkin, M. H. and Rees, P. H. (2010) A dynamic microsimulation with agent elements for spatial demographic forecasting, *Journal of Social Science Computing Review*, 29 (1), 145-160, DOI: 10.1177/0894439310370113.
3. Wu, B. M. and Birkin, M. H. (2012) Moses: A dynamic spatial microsimulation model for demographic planning, in Tanton, R. and Edwards, K. (eds.) *Spatial Microsimulation: A Reference Guide for Users*, Understanding Population Trends and Processes 6, DOI 10.1007/978-94-007-4623-7\_11.

# Acknowledgements

This research would not have been possible without the help, support and advice of many very special people, for which I am deeply grateful.

First of all, I would like to thank my supervisors: Prof. Mark Birkin and Prof. Phil Rees. Thank you for the continuous support and professional insights throughout this study. Without your support, the completion of this long distance PhD programme would have not been possible. Thank you, Mark, for your efficient supervisions and decisive interventions when I drift away from the course. Thank you, Phil, for your advice, support and patience when I caught the “PhD Blues”. I will miss our supervision meetings over lunches in your office, your supervision notes posted to my home and telephone supervisions over distance.

Also thanks for my Research Support Group members: Prof. Graham Clarke, Dr. Alison Heppenstall and Dr. Paul Norman, for their support and feedback. Thanks for Graham’s prompt useful feedbacks, Alison’s chats and support over the years, Paul’s agreement to the last minute fill-in for Alison’s maternity leave. Special thanks also go to my colleagues in School of Geography, for their direct and indirect support in this project.

To my family, there will never be enough words to express how grateful I am to you all for your love and support. Special thanks for the unwavering belief in me from my husband, a sympathetic ear and continuous support and the constant injections of optimism from my son.

# Abstract

This research aims to study and understand a complex social system through the development of an individual based hybrid model of the population of Leeds, UK. It attempts to demonstrate the importance of individual based modelling and simulation tools within the scope of demographic planning, as well as in application of a variety of substantive research and planning environments.

The model adopts a hybrid modelling approach that combines the strength of two individual based modelling approaches: the first employs a dynamic spatial Micro-Simulation Model (MSM) and the second uses an Agent Based Model (ABM). This framework enables the modelling of a complex social system that is both theoretically and practically challenging. It attempts to provide a fuller picture of the population evolution through the simulation of discrete changes experienced by a large number of individuals within small areas and demonstrates heterogeneous characteristics in individuals and their behaviours reflecting not only individual demographic characteristics, but also interactions with each other and/or their local environment.

This thesis describes the modelling method, system development, results alignment and model applications, as well as discussing the limitations and future potential of this model.

# Table of Contents

<b>Title page.....</b>	<b>I</b>
<b>Acknowledgements of funding and data.....</b>	<b>II</b>
<b>Acknowledgements .....</b>	<b>IV</b>
<b>Abstract .....</b>	<b>V</b>
<b>Table of contents.....</b>	<b>VI</b>
<b>List of figures.....</b>	<b>XIV</b>
<b>List of tables.....</b>	<b>XVIII</b>
<b>Chapter 1 Introduction</b>	
1.1 Introduction .....	1
1.2 Background.....	2
1.3 Aim and objectives of the study.....	5
1.4 Structure of the thesis.....	6
1.5 Conclusion.....	7
<b>Chapter 2 A review of literature on microsimulation and related models.....</b>	
2.1 Introduction.....	9
2.2 Social models.....	10
2.2.1 <i>History of development of approaches to complex social system...</i>	<i>10</i>

2.2.2	<i>Comparison of the macroscopic and the microscopic approaches of social models</i> .....	11
2.3	Microsimulation models.....	14
2.3.1	<i>Definition of microsimulation</i> .....	14
2.3.2	<i>History of MSMs</i> .....	15
2.3.3	<i>Application areas of MSMs</i> .....	18
2.3.4	<i>Types of MSMs</i> .....	24
2.4	Cellular Automata and Agent Based Models.....	29
2.4.1	<i>Cellular Automata (CA)</i> .....	30
2.4.2	<i>Agent Based Models (ABM)</i> .....	31
2.5	Multi-level simulation models.....	40
2.6	Hybrid modelling approach .....	43
2.6.1	<i>Individual based approach</i> .....	43
2.6.2	<i>Limitations of MSM and ABM</i> .....	44
2.6.3	<i>Previous hybrid modelling approaches</i> .....	45
2.7	Conclusion.....	55
	<b>Chapter 3 System design</b> .....	<b>57</b>
3.1	Introduction.....	57
3.2	System design.....	58
3.2.1	<i>Projection model design</i> .....	58
3.2.2	<i>Population model design</i> .....	64

3.3	System development.....	69
3.3.1	<i>Programming approach and language.....</i>	69
3.3.2	<i>Architecture.....</i>	71
3.3.3	<i>Modelling approach.....</i>	74
3.4	Data selection.....	77
3.4.1	<i>Baseline data selection.....</i>	77
3.4.2	<i>Attribute selection.....</i>	78
3.4.3	<i>Data used in simulation.....</i>	82
3.4.4	<i>Data used in updating.....</i>	85
3.5	Result alignments.....	86
3.6	Conclusion.....	88
<b>Chapter 4 Development of the dynamic spatial microsimulation model</b>		
.....		<b>91</b>
4.1	Introduction.....	91
4.2	General development method.....	92
4.2.1	<i>Dynamic spatial MSM.....</i>	92
4.2.2	<i>Demographic processes.....</i>	95
4.2.3	<i>Base attributes.....</i>	96
4.2.4	<i>General processing procedures.....</i>	99
4.2.5	<i>Probabilities and dataset.....</i>	102
4.3	Mortality.....	104



4.3.1	<i>Factors</i> .....	104
4.3.2	<i>Mortality process in simulation</i> .....	111
4.3.3	<i>Mortality probability generation</i> .....	112
4.4	Fertility.....	118
4.4.1	<i>Factors</i> .....	119
4.4.2	<i>Fertility process in simulation</i> .....	121
4.4.3	<i>Fertility probability generation</i> .....	123
4.5	Health.....	126
4.5.1	<i>Factors</i> .....	127
4.5.2	<i>Fertility process in simulation</i> .....	129
4.5.3	<i>Fertility probability generation</i> .....	130
4.6	Household formation.....	132
4.6.1	<i>Factors</i> .....	132
4.6.2	<i>Household formation process in simulation</i> .....	134
4.6.3	<i>Household formation probability generation</i> .....	138
4.7	Migration.....	139
4.7.1	<i>Factors</i> .....	139
4.7.2	<i>Migration process in simulation</i> .....	140
4.7.3	<i>Migration probability generation</i> .....	140
4.8	Initial result analyses .....	142

4.8.1	<i>Initial result analysis at aggregate level of LA.....</i>	142
4.8.2	<i>Initial result analysis at small area level of wards .....</i>	145
4.8.3	<i>Initial result analysis by demographic process.....</i>	150
4.9	Conclusions.....	158
<b>Chapter 5 Further development of the MSM .....</b>		<b>162</b>
5.1	Introduction.....	162
5.2	The extension of migration process.....	164
5.2.1	<i>New migration framework.....</i>	164
5.2.2	<i>Further development under the new migration framework .....</i>	166
5.3	Fertility updating .....	172
5.4	Mortality updating .....	173
5.5	Result analyses of the revised model .....	175
5.5.1	<i>Population sex-age structure analysis in Leeds .....</i>	176
5.5.2	<i>Population sex-age structure analysis in small areas .....</i>	177
5.5.3	<i>Result analysis by demographic processes.....</i>	180
5.6	Conclusion .....	184
<b>Chapter 6 Experiments with ABM .....</b>		<b>188</b>
6.1	Introduction.....	188
6.2	Student migration experiment.....	193
6.2.1	<i>Pure MSM.....</i>	193

6.2.2	<i>Hybrid model combines MSM and ABM.....</i>	200
6.2.3	<i>Comparison of results using pure MSM and the hybrid Model.....</i>	207
6.3	Mortality experiment.....	211
6.3.1	<i>Pure MSM.....</i>	211
6.3.2	<i>Hybrid model combines MSM and ABM.....</i>	212
6.3.3	<i>Comparison of results using pure MSM and the hybrid Model... </i>	214
6.4	Conclusions .....	215
<b>Chapter 7 Model validation and alignments .....</b>		<b>218</b>
7.1	Introduction.....	218
7.2	The need of alignments for MSMs .....	220
7.2.1	<i>Issues in validating dynamic spatial MSMs.....</i>	220
7.2.2	<i>The alignment approach.....</i>	223
7.3	The alignment framework .....	225
7.4	Model validation and alignment .....	227
7.4.1	<i>Model A.....</i>	227
7.4.2	<i>Model B.....</i>	229
7.4.3	<i>Model C.....</i>	230
7.5	Analysis of alignment results .....	231
7.5.1	<i>Analysis of age-sex distribution of Leeds population.....</i>	232
7.5.2	<i>Analysis of components of change of Leeds population.....</i>	235
7.5.3	<i>Analysis of age-sex distribution of small area populations.....</i>	241

7.5.4	<i>Analysis of components of change of small area populations</i> .....	244
7.6	Conclusion.....	248
<b>Chapter 8 Health application: obesity scenarios</b> .....		<b>251</b>
8.1	Introduction .....	251
8.2	Public health planning with a focus on obesity.....	252
8.2.1	<i>Public health planning</i> .....	252
8.2.2	<i>Obesity focus in public health planning</i> .....	255
8.2.3	<i>Previous obesity research findings</i> .....	257
8.3	Obesity application model .....	259
8.3.1	<i>Modelling method</i> .....	259
8.3.2	<i>Description of the model and scenarios</i> .....	264
8.4	Analysis of results .....	268
8.4.1	<i>Obesity patterns at LA level</i> .....	268
8.4.2	<i>Obesity patterns at ward level</i> .....	272
8.4.3	<i>Obesity patterns by sub-population</i> .....	275
8.5	Conclusion .....	280
<b>Chapter 9 Conclusions</b> .....		<b>285</b>
9.1	Introduction .....	285
9.2	Summary of main findings.....	286
9.2.1	<i>Research objectives</i> .....	286

9.2.2	<i>Main findings against objectives by chapters</i> .....	287
9.3	Limitations .....	296
9.4	Future work .....	299
9.5	Conclusion .....	301
	References .....	305
	Appendix A Abbreviations.....	329

# List of figures

Figure 1.1 Thesis Structure.....	8
Figure 2.1 Historical developments of contemporary approaches to complex systems.....	11
Figure 3.1 Lexis diagram showing the age-time space used in the projection model .....	61
Figure 3.2 Illustration of differences of cohort-age, period-age and period-cohort approaches in age-time spaces .....	62
Figure 3.3 Process of the population simulation.....	65
Figure 3.4 Framework for modelling the migration process .....	68
Figure 3.5 System architecture .....	71
Figure 3.6 An illustration of program hierarchy.....	73
Figure 4.1 Age-Sex distribution in Leeds small area populations: Cookridge and Headingley in 2001.....	97
Figure 4.2 The general simulation procedure.....	102
Figure 4.3 Mortality by age and sex in Leeds Wards 2001.....	105
Figure 4.4 Mortality process.....	113
Figure 4.5 Fertility rates by age of mother in two Leeds wards, 2001.....	120
Figure 4.6 Fertility process.....	124
Figure 4.7 The general health information in census 2001 records.....	126
Figure 4.8 Health by age and sex in Leeds wards, 2001.....	129

Figure 4.9 Health change process.....	131
Figure 4.10 Household formations through marriages by age and sex in Leeds wards, 2001.....	133
Figure 4.11 The marriage and household formation process.....	135
Figure 4.12 Matching candidates in marriage process.....	137
Figure 4.13 HRP status updating.....	138
Figure 4.14 Migration process.....	141
Figure 4.15 The age–sex structure of Leeds population in 2001 and 2029 generated by the MSM compared with the ONS estimates in 2029.....	144
Figure 4.16 The age–sex structure of ward population: Cookridge and Headingley in 2001 and 2030.....	146
Figure 4.17 Location quotient analyses of ward level projections.....	148
Figure 4.18 Ward level projections of mortality in 2030.....	151
Figure 4.19 Ward level projections of household formation in 2030.....	152
Figure 4.20 Ward level projections of fertility in 2030.....	154
Figure 4.21 Ward level projections of health change in 2030 by sex and age.....	155
Figure 4.22 Ward level projections of in-migration in 2030 .....	156
Figure 4.23 Ward level projections of out-migration in 2030 .....	157
Figure 4.24 Ward level projections of net-migration in 2030 .....	158
Figure 5.1 The main processes under the new framework.....	168

Figure 5.2 Population sex-age structures of Leeds population in 2001 and 2030 generated by the revised MSM compared with the ONS estimates in 2030.....	176
Figure 5.3 Population sex-age structures of Cookridge population in 2001 and 2030 generated by the revised MSM.....	178
Figure 5.4 Population sex-age structures of Headingley population in 2001 and 2030 generated by the revised MSM.....	179
Figure 5.5 Mortality result analyses of the revised MSM: Cookridge and Headingley in 2001 and 2030.....	181
Figure 5.6 Fertility result analyses of the revised MSM: Cookridge and Headingley in 2001 and 2030 .....	182
Figure 5.7 Migration result analyses of the revised MSM: Cookridge and Headingley in 2001 and 2030.....	185
Figure 6.1 Location quotient analyses of ward level MSM projections ...	195
Figure 6.2 Migration question (Q14) in UK census form, 2001.....	198
Figure 6.3 Student migration process.....	206
Figure 6.4 Migration experiment: students in small areas 2001 and 2031.....	208
Figure 6.5 LQ analyses of Headingley projections by hybrid model: 2001 and 2031.....	210
Figure 6.6 Migration history of migrant A .....	213
Figure 6.7 Mortality experiment: deaths in small areas 2001 and 2031 ...	215
Figure 7.1 Result alignment framework.....	228



Figure 7.2 Population sex-age structures of Leeds population in 2002 and 2031 generated by Model B compared with the ONS projections (Model A) in 2031.....	233
Figure 7.3 Population sex-age structures of Leeds population in 2002 and 2031 generated by Model C compared with the ONS projections (Model A) in 2031.....	234
Figure 7.4 Age-sex structures in small area populations projected by Model B.....	242
Figure 7.5 Age-sex structures in small area populations projected by Model C.....	243
Figure 8.1 The Dahlgren-Whitehead health model.....	253
Figure 8.2 Obesity prevalence of ages 16 and over by social class, 1997–2007.....	263
Figure 8.3 Foresight obesity scenarios.....	266
Figure 8.4 Obesity in Leeds 2001: all population.....	269
Figure 8.5 Obesity in Leeds 2031: Scenarios 1, 2 and 3.....	271
Figure 8.6 Obesity rates by age and sex in Cookridge and Headingley in 2001 and 2031 (Scenarios 1, 2 and 3).....	274
Figure 8.7 Obesity by age (adult and child) in 2001 and 2031 (NSSEC1).....	276
Figure 8.8 Obesity by sex in 2001 and 2031 (NSSEC1) .....	278
Figure 8.9 Child obesity by Socio-economic status in 2001 and 2031.....	279

# List of tables

Table 2.1 Comparison of static and dynamic MSMs.....	27
Table 4.1 Life table example: Cookridge in 2001.....	116
Table 5.1 The extended migration framework.....	166
Table 5.2 Actual and assumed overall average annual rates of mortality improvement.....	174
Table 5.3 Population differences between variant mortality projections and principal projection by age, 2011-31.....	174
Table 6.1 Index of dissimilarity for student populations .....	209
Table 7.1 Components of change analyses of Leeds population.....	236
Table 7.2 Components of change of small area populations projected by Model B.....	245
Table 7.3 Components of change of small area populations projected by Model C.....	246
Table 7.4 Disaggregation impact on internal migration out in Model C.....	247
Table 8.1 Definitions of BMI.....	260
Table 9.1 Objectives achieved in chapters.....	287

## Chapter 1

# Introduction

## 1.1 Introduction

This project aims to develop a hybrid MicroSimulation Model (MSM) in order to facilitate the study of a complex social system which is the population of Leeds, a medium sized city in northern England. It attempts to demonstrate the importance of individual based modelling and simulation tools within the scope of demographic planning, as well as their applications in a variety of substantive research and planning environments.

To achieve this aim, the model adopts a novel modelling approach that combines the strength of two individual based modelling approaches: the first employs a dynamic spatial MSM and the second uses an Agent Based Model (ABM). This hybrid approach enables the modelling of discrete changes in the population of a large number of individuals living within small areas and demonstrates the heterogeneous characteristics in individuals and their behaviours. Therefore, this population model based system provides a fuller picture of the evolution of the population and a

generic basis for analysing different demographic planning or other wider social science research problems.

This chapter introduces the contextual background of this study, describes the aim and specific objectives of this study, as well as providing a short description of the structure of the thesis.

## 1.2 Background

People have always been at the centre of studies of social systems. People's movements, interactions and behaviours will inevitably have an important impact on the society and environment that they are living in. At the same time, such individual changes will also lead to an evolution of the whole population over time. It is such complexity within our social system that keeps attracting research interest. Indeed, people and societies, along with their past, present, and future magnitudes and compositions, have never failed to capture the fascination of researchers and modellers.

Advances in technologies and new tools often bring new visions to our research. Computer based models have now been extensively used to model complex social systems, not only because they can provide valuable groundwork when it is too expensive or impossible (for practical reasons) to experiment in reality, but also new research methods enabled by the capabilities of modern computers can radically transform human ability to reason systematically about complex social systems. This has become increasingly important as our world today confronts rapid and potentially profound transitions driven by social, economic, environmental and technological changes.

Traditionally the macroscopic approach has been adopted to model populations. Macroscopic models such as the cohort-component model have been widely applied in demographic studies. However, macroscopic models have limits in representing microscopic occurrences, discontinuity and heterogeneity within a system (Billari *et al.*, 2003). Modern studies of social

systems now often require detailed information about the behaviours and characteristics at the level of individual decision making units. The macroscopic approach struggles in such studies. On the other hand, the microscopic approach can provide valuable insight into the behaviours of a system under a range of conditions through the provision of rich details about individuals. As an Individual Based Model (IBM), a MSM can provide valuable information over a wide range of circumstances. Therefore it has become not only an excellent knowledge discovery tool, but also provided indispensable assistance to strategic decision making (Orcutt, 1957; van Imhoff and Post, 1998; Harding, 2007). Now, advances in computing power and analytical techniques allow great sophistication in the range of questions that the MSM may address.

However, the statistical nature of MSM limits its effectiveness in modelling the interactions and behaviours of individuals. Another alternative IBM, the Agent Based Model (ABM), can complement MSM in such areas by modelling the individual changes through interactions among individual agents and between individuals and their environment. Such interactions are driven by autonomous actions/reactions from the agents according to their built-in rules which are not necessarily equation or probability based. Thus the integration of ABM features into an MSM can help bridge knowledge and data gaps about certain behaviours of sub-populations or particular demographic processes.

Although advances in computing and the availability of statistics have now enabled us to portray the world in a more sophisticated way, this study faces some great challenges. The main challenges are two-fold: on the one hand, we are studying a complex social system; on the other, we require an effective and efficient computer model. Because social systems are complex, they tend to be very large and complicated with unclear or sometimes even unknown boundaries. There is often a more complex mix of factors and the relationships between such factors changes. Also individual actors in social systems can make decisions on their own, based on some

necessary processes (Moss, 2000). Such features make the modelling (understanding and conceptualising) more difficult.

The thesis work moved in parallel with one strand in the two ESRC projects where the author is a full-time researcher: Moses (Modelling and simulation of e-Social Science) and Genesis (Generative Simulation for the Social and Spatial Sciences). The outputs were a set of journal papers and conference papers jointly with the supervisors. However, the thesis constitutes a full documentation of the model and its experimental outputs, which are only briefly described in the journal papers. Both projects ambitiously attempt to provide a generative simulation basis for explorations in various fields within the spatial and social science dimensions, on the basis of a population microsimulation model. Such an ambitious model previously has taken a whole team decades to accomplish, e.g. DYNASIM (Orcutt 1957), CORSIM in USA (Caldwell, 1998) and APPSIM in Australia (Harding, 2007) and EUROMOD for the whole Europe (Sutherland, 2007).

This model attempts to capture both individual and spatial characteristics. This presents great challenges as described in Duley (1988) in his previous microsimulation model for updating individual and household characteristics in small areas. Such a model comes with a high volume of data and complexity within the system. The model starts with about 730,000 individuals in the year of 2001, each with a rich set of population attributes. However, over the period of the simulation, the number of calculations begins to grow exponentially. For instance, survival rates in this model are derived separately for each of 33 census wards, both genders, and 101 individual age groups, giving a total of 6,666 individual probabilities. At each time period, a survival probability is applied to each individual on the basis of age, gender and location. The model is run in annual time increments, and therefore the ageing rule for all survivors is that they become a single year older in each time step. It can be seen that for this most straightforward process in the model alone, projection of mortality over a 30 year time interval involves the estimation of roughly 200,000 individual mortality rates. The implication for the computing requirements

is huge: such large volumes of data not only require great computing resources, but also considerable computing skills are required for the data to be stored, retrieved, updated and output.

Adopting the hybrid modelling approach, this study also attempts to address the theoretical and practical issues posed by the challenges. This will be discussed in detail in following chapters. In the next section, the main research aim and a set of more specific objectives to tackle such challenges are described.

### 1.3 Aim and objectives of the study

This research aims to study a complex social system to facilitate the understanding of it through the development of a population based hybrid model that combines the strength of both the MicroSimulation Model (MSM) and the Agent Based Model (ABM). It will demonstrate the importance of individual based modelling and simulation tools within the scope of demographic planning, as well as a variety of substantive research and planning environments. To achieve the aim of the research, some more specific objectives of this study have been established:

1. To review and discuss the microscopic approach in social modelling.
2. To develop an innovative modelling framework that enables the study of population evolutions through individual changes over the period of 2001-2031.
3. To provide a complete representation of the studied population at a fine spatial scale.
4. To produce rich, detailed and robust projections of the future population.
5. To investigate scenarios of demographic related public planning.

In terms of the demographic planning, this project aims to use the power of computational simulation to get beyond traditional macroscopic approaches to demographic analysis and forecasting in terms of:

6. spatial disaggregation of population projections;
7. rich representation of heterogeneous characteristics in individuals and their behaviours and
8. scenario-based analysis of population changes.

While the next section describes the structure of the thesis, the aim and objectives of the study will be examined again in the conclusion chapter to find out how they are reflected in each chapter of the thesis.

## 1.4 Structure of the thesis

This thesis has been organised into 9 chapters. Figure 1.1 is used here to help describe the organisation of the thesis chapters.

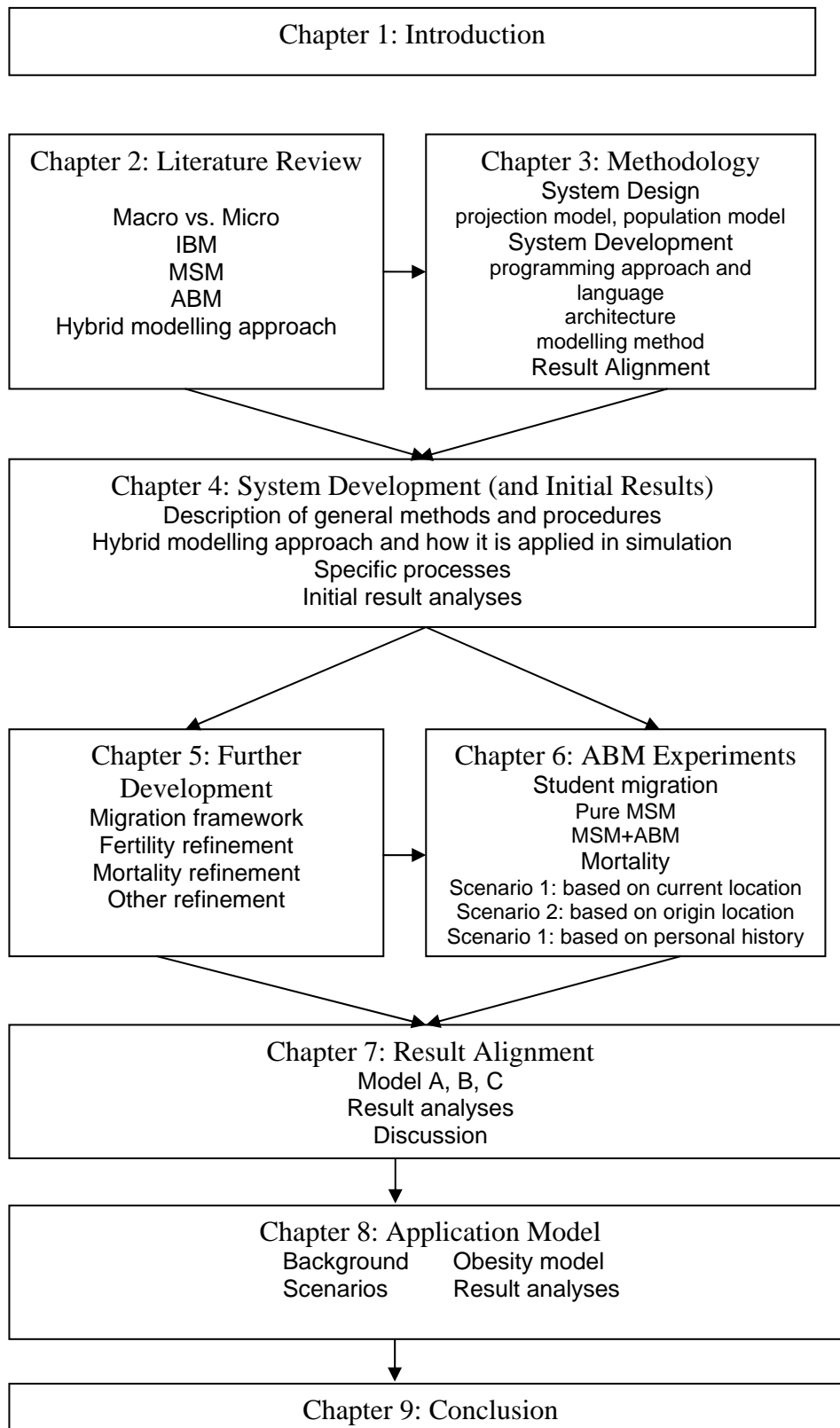
As illustrated in the diagram, Chapter 1 provides an introduction of the study with a brief introduction of the background of the study, followed by the aim and the specific objectives of the study, and finishes with a description of the structure of the thesis. Chapter 2 provides a review on current literature and models and provides a theoretical basis for the model development and methodology development. Chapter 3 describes the methods used in the processes of system design, system development and result alignment. Chapter 4 explains the system development in more detail. A description of general method used in the model development has been provided, as well as the more detailed modelling of each demographic process. Analyses have also been carried out on the initial results and a short discussion will lead to the next chapter. Chapter 5 describes the further development and refinement of the model in an attempt to address the issues arising in the initial results, while Chapter 6 explains how the ABM can strengthen the MSM through two experiments on the modelling of student



migration and mortality with agent features. Chapter 7 describes the attempt to align the model results to the official projections. The three step alignment exercises are explained, results compared to the ONS projections and the main findings are discussed in this chapter. Chapter 8 tries to demonstrate the potential use of the model by providing an application model that projects obesity in small areas. Main findings from the model have been discussed and such information can be used to assist strategic public health planning, as well as location targeted tactical interventions. Chapter 8 also discusses other potential applications of the model. Chapter 9 summaries the main findings of the study, discusses the limitations of the model with a view of future work before drawing some conclusions (Figure 1.1).

## 1.5 Conclusion

This chapter has briefly introduced the background of the study, followed by a description of the research aim and specific objectives. A description of the structure of the thesis has been provided and discussed. The next chapter, Chapter 2, will provide a review of the relevant current literature and models.



**Figure 1.1 Thesis structure**

## Chapter 2

# A review of literature on microsimulation and related models

## 2.1 Introduction

In this chapter, a review is conducted of various approaches used in modelling social systems and relevant major model developments around the world. The limitations of approaches to model individuals are discussed. In particular, the purpose, methods and outcomes of a class of models known collectively as IBMs (Individual Based Models) are discussed in Section 2.2. This chapter focuses on three important IBMs: MSMs (MicroSimulation Models), CA (Cellular Automata) models and ABMs

(Agent Based Models). In Section 2.3 the MSM is introduced and a review is conducted of models used in different public planning application areas. A classification of MSMs is also outlined in this section. Section 2.4 introduces CA models and ABMs separately and reviews ABM theory, architecture and available frameworks. Section 2.5 discusses multi-level models and Section 2.6 proposes the hybrid modelling approach for this study. Finally, Section 2.7 concludes this chapter by reflecting on the review. This chapter uses some material from Wu and Birkin (2012).

## 2.2 Social models

### 2.2.1 *History of development of approaches to complex social systems*

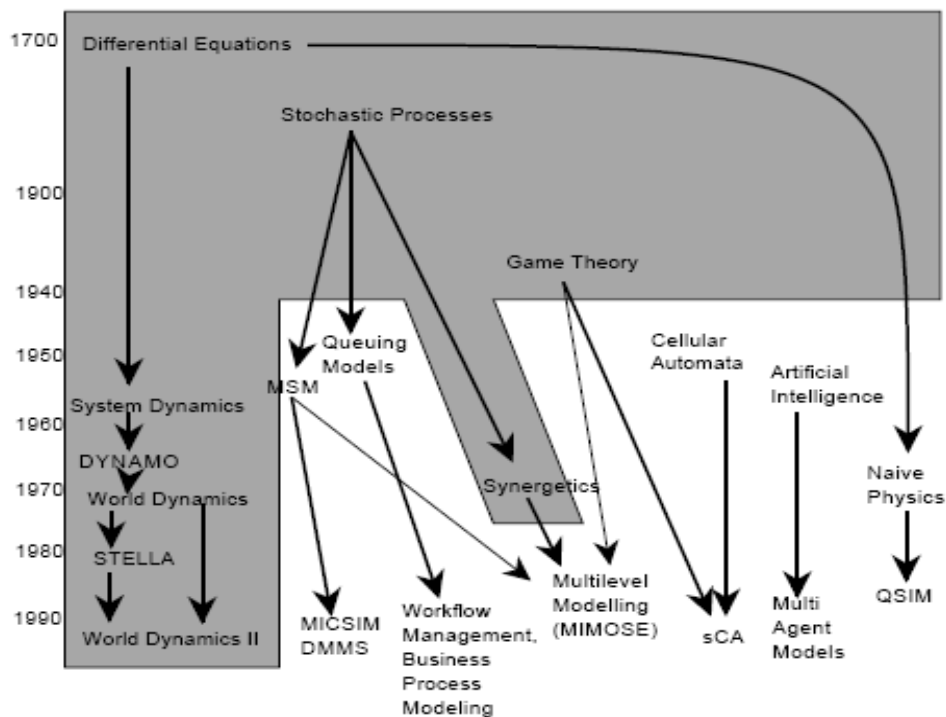
*Chambers Dictionary* (2013) defines model as “a small-scale representation of something that serves as a guide in constructing the full-scale version” or “a small-scale replica”. Models help us understand complex systems through simpler representations of the systems and they also allow us to study problems when it is difficult or impossible to do so in reality.

When computer power was more limited, traditional mathematical models attempted to analyse and predict the behaviour of the system from a set of parameters and initial conditions. With the advances in computing, computer simulations/models have become more important in modelling complex systems as they enable us to verify our knowledge about the world through various experiments in great detail on computers. Computer simulations/models also allow us to address a wider range of research questions and make more accurate predictions for the future to facilitate strategic decision-making.

Although scientists have been using models in research for a long time, computer simulation in the social sciences is still a relatively new area. Such simulations only started to be widely adopted in the 1990s. This is partly because in social science research it is more difficult to establish the truth as

in the some other pure sciences. This is because of the very nature of the subject that social science studies and the social systems tend to be very large and complicated. There is often a more complex mix of factors involved and the relationships between such factors can change quickly.

Given the challenges, researchers have been exploring new technologies and techniques in order to provide a better representation of the complexity of social systems. Gilbert and Troitzsch (2005) illustrate different types of social models as shown in Figure 2.1. The trend of moving from macro approaches (shaded area in the diagram) towards the micro approaches to obtain more individual characteristics are also demonstrated in this diagram.



**Figure 2.1: Historical development of contemporary approaches to complex systems**

Source: Gilbert and Troitzsch (2005, pp7)

### 2.2.2 *Comparison of the macroscopic and the microscopic approaches of social models*

As illustrated in Figure 2.1, social systems can be modelled at different levels. Macroscopic and microscopic are the two major levels. Generally speaking, the macroscopic approach presents a higher level of abstraction

than the microscopic one. Traditionally the macroscopic approach has been used to model social systems and macroscopic models have been widely applied in demographic studies, for instance. Macro level difference or differential equation models developed into system dynamics and world models. Macrosimulation has a longer history than the microsimulation and it has a solid basis in the tried and tested codes of practice around the world. Compared to microscopic simulators, macrosimulation has the following advantages:

- it needs fewer inputs;
- it is easier to model large-scale systems;
- it has less computation requirement and
- it is easier to calibrate.

However, modern studies now require much more disaggregated details of the social system. Quite often they demand to capture the characteristics and changes of individuals. Macroscopic simulators have limits in representing microscopic occurrences, discontinuity and heterogeneity within a system. The macroscopic approach is well situated for the large-scale network scenarios, but is less strong when the impact of microscopic events is to be evaluated. Macrosimulations do not model the interaction of individual units - but rely on the behaviour of the aggregate system as a large scale interactive system. Therefore disaggregate details are often overlooked. This limits model effectiveness in studies where individual behaviour is important. For instance, it would be almost impossible to model the walking behaviour of 50,000 pedestrians using macrosimulation approach. To model a crowd using macroscopic models we need to state numerous underlying assumptions and to implement frequent probability generations and statistical calculations.

Van Imhoff and Post (1998) point out that macro and micro simulation models are alternative ways of representing the same demographic processes. The benefits of MicroSimulation Model (MSM)s (in contrast to macroscopic modelling approaches with similar objectives) within a

demographic modelling context have been argued persuasively and eloquently by these authors. In particular, they demonstrate the richness of MSM as a device for the representation of both relationships between members of a population, and of the transitions between states within a population. They point out that:

- MSM performs better under conditions of a sizeable state space, while macro models become unmanageable when include large number of attributes/values;
- MSM is more flexible in handling interactions between individuals, which is lost in macro models;
- MSM can handle continuous covariates and
- MSM can provide much richer output.

Grimm (1999) suggests that Individual Based Model (IBM)s make more realistic assumptions than macroscopic models. Also the individual-based approach is a bottom-up approach which starts with individuals within a system, then tries to understand how the system's properties emerge from the interaction among these parts. Therefore, another advantage of IBM is that system-level phenomena such as population dynamics can be explained through the relatively simple individual changes. The microscopic approach can deal with more detail. Microscopic techniques provide insight into the behaviour of a system under a wide range of conditions. Such features make the IBMs an excellent discovery tool. They can provide valuable information over a wide range of behavioural inputs and theories and options can be modelled and tested (Krupp, 1986).

Various IBMs can facilitate the social models with individual characteristics. MSMs model interactions between the policies and individual decision making units and can therefore meet such demands. CA model social dynamics with a focus on the emergence of properties from local interactions and ABMs simulate more complex situations than the CA where the "agents" control their own actions based on their perceptions of the environment. ABMs can also provide the potential to incorporate learning over time in response to environmental constraints or changes. Due

to such features, the IBM has become more and more popular in social modelling.

Attempting to link the macro and the micro aspects of the social systems, the multilevel simulations try to model the interactions between the macroscopic level and the microscopic levels. Therefore they will be discussed in this review.

## 2.3 Microsimulation models

Brown and Harding (2002, pp.3) define the term of social modelling as “the representation of social phenomena and/or the simulation of social processes” and think that the MSM is “a pre-eminent type of social model” (pp.2). Advances in computing and statistics now allow MSMs to portray the world in a sophisticated way. A MSM can model interactions between the people and policy through simulation of distinctive behaviours and characteristics at the level of individual decision making units (Orcutt, 1957). This meets the requirements of the modern social science studies, where detailed information of the interactions between the policy and the social-economic behaviours of people are often needed.

### 2.3.1 *Definition of microsimulation*

Microsimulation has been defined as “a statistical procedure for estimating the characteristics of individuals from knowledge of the aggregate characteristics of the population to which they belong” (Johnston, 2000). However, this refers just to one set of techniques - the use of probabilities derived from aggregate data together with a random number generator to select a characteristic for an individual.

Generally speaking, a spatial MSM works to create small area microdata at a certain point of time and then generate future microdata on the basis of the initial microdata (Ballas *et al.*, 2005) . For instance, if we start with a population of entities, set  $P$ , made up of individuals  $[P^1, P^2 \dots, P^n]$  where



$n$  is the number of individuals in the population sample. Each individual has a set of attributes,  $[a_1^t, a_2^t \dots, a_m^t]$ , which describe the individual at the time  $t$ . We therefore have an  $n \times m$  array of person attributes. This array needs to be populated with reliable data or estimates (in the light of directly surveyed information). Then we update the population so that the baseline population  $[P_{a_1^t a_2^t \dots a_m^t}^1, P_{a_1^t a_2^t \dots a_m^t}^2, \dots, P_{a_1^t a_2^t \dots a_m^t}^n]$  changes to new sets with attributes/states at a point in time  $t+1, t+2 \dots$  and so on.

One of the most important advantages of microsimulation modelling is that it enables us to examine the impact of policy changes on individual decision units, as it is based on unit records. This distinguishes MSMs from the traditional mathematic models. Such models are often based on aggregated or averaged values and individual characteristics can often become blurred and can even disappear in such models. An MSM deals directly with social processes at the individual level and therefore it has been extensively used for various purposes in studies in which individual characteristics are important. In this literature review, we focus on public policy MSMs, where interactions between the public policies and individual decision making units are modelled.

### 2.3.2 *History of MSMs*

MSMs have come a long way since they were first conceived and developed to provide a new way to predict distributions of individual decision-making units. This kind of models has advantages in handling nonlinear relationships over the traditional aggregate representations, as the aggregate value tends to depend on the distribution of disaggregate values. Although the applications of MSM have spread across the whole social science spectrum, this review has found that its roots are based in two disciplines in particular: economics and geography.

### 2.3.2.1 *The economic roots of MSM*

Orcutt (1957) is one of the founders of the micro-analytical research methodology. Orcutt developed behavioural dynamic microsimulation in the form of DYNASIM (Dynamic Simulation of Income Model) in the USA (Orcutt, 1957; Orcutt *et al.*, 1976). As the name indicates, DYNASIM was used for individuals as decision-making units to understand the impacts of new public policies concerning income transfers, for example taxation and benefits (Orcutt, Merz & Quinke, 1986). However, DYNASIM has been used for a range of studies and inspired the development of many other MSMs. Among them is Caldwell's (1998) CORSIM (Cornell Microsimulation Model).

CORSIM (Caldwell, 1998) is a microsimulation model of the United States population developed in Cornell. CORSIM models large scale government programmes and is particularly strong in modelling the social security programme. There are three main units in CORSIM: the database construction modules pre-process, format and archive dataset into a database; the simulation modules simulate demographic and economic changes during the lifecycle of the base population and then the output modules produce user-specified outputs about families and individuals using information from the simulation results. Demographic and economic attributes of the base population are held in the CORSIM database. CORSIM simulates changes of each individual unit (persons and families) on a yearly step. The resulting data are validated and aligned using the available external data before projections into the future are generated. As all MSM age data using either dynamic or static ageing techniques to project data into the future, CORSIM ages the data dynamically through a set of modules:

1. Migration between states
2. Immigration to the US
3. Mortality

4. Fertility
5. Ageing
6. Marriage
7. Divorce
8. Leaving home
9. Education
10. Disability
11. Earnings
12. Housing
13. Wealth and
14. Insurance.

Due to its wide ranging modules and potential application domains, CORSIM was selected as the starting point and further inspired many MSMs around the world.

#### *2.3.2.2 The geographical roots of MSM*

Around the same time as Orcutt was developing DYNASIM, Hägerstrand (1957) in Sweden changed the fashion of social studies by linking the spatial and temporal capacities and restraints on individual behaviour. His time-space work came well after his spatial MSM work, of which his 1957 piece is the most famous. Having studied human migration for many years, he pointed out in the 1950s that the study of human beings as groups and aggregate populations masked the true nature of human patterns of movement. Hägerstrand's work on migration of individuals in central Sweden incorporated space into MSM and used a mixture of empirical "information fields" and decision rules to model inter-zonal migration. The application of heterogeneous decision rules to individuals also demonstrated similarity to one of the main features in the rule-driven individual based simulations in CA models and ABMs. Further his space-time framework recognised that activity participation has both spatial and temporal dimensions, as all human activities occur at particular places for limited durations (Hägerstrand, 1985). Hägerstrand's research has explored the conceptual basis for developments in spatial microsimulation, which is

distinguishable from other types of MSMs. The spatial microsimulation will be described later in this review (Section 2.3.4.4).

Hägerstrand's space-time geography revolutionised the study of a wide range of urban studies from city planning to social equity, as it allows the space-time measures to be rooted in the local urban environment and incorporated with individual contexts (Pred,1977). Such features caught the attention of the transport researchers particularly (Kwan, 1998; Miller, 1991). In fact, many transport MSMs also exhibit the similar features as Hägerstrand's migration studies: the individual units (vehicles/pedestrians) demonstrate the capacity of certain degrees of intelligence and parts of the simulation are driven by rules designed explicitly for them, which exhibit some of the basic features of agents within an ABM.

More examples of the two approaches of MSM and ABM have been used together in various previous studies will be discussed in more details in Section 2.6.3.4. In section 2.3.3 below, we will review the MSM application areas with a focus on public planning.

### 2.3.3 *Application areas of MSMs*

From its economic and geographic roots, MSMs have been applied in different contexts across the whole social science spectrum. Microsimulation may mean different things (people, households, traffic) to different people, but in this section we just focus on the public policy MSMs, as it is one of the objectives of this project. As a consequence of the increasing availability of large and detailed datasets and the continuous increase in computing power with less cost, major MSM developments have been seen around the world during the past few decades. Such developments include CORSIM (Caldwell, 1993) in the United States, DYNACAN (Morrison, 2003) in Canada, SVERIGE in Sweden (Rephann, 1999), PENSIM in the United Kingdom (Curry, 1996), SMILE in Ireland (Ballas *et al.*, 2005a), APPSIM in Australia (Harding, 2007) and on the scale of the whole Europe, EUROMOD (Sutherland, 2007).

Major developments such as DYNASIM, APPSIM, CORSIM and other models have been used extensively in public policy assessments. For instance, CORSIM has been used to facilitate research on income distribution and wealth accumulation; individual health conditions and health service usage; school, work, marital and child choices and their interrelationships; social mobility; individual-societal relationships; and voting behaviour (Caldwell, 1993). However, more specialised microsimulation models (sometimes derived from the major models) have also been developed to facilitate studies in different research fields as follows.

#### *2.3.3.1 Tax-Benefit MSMs*

MSMs have been widely used to assess the impact of government policies and to help reform the current social policies. Policy makers were interested in who would gain/lose from changes in the tax and benefit systems. As a result of the requirement of suitable tools to analyse the impact of such changes, many tax-benefit models have been developed.

The US model **TRIM** (Transfer Income Model) was developed and maintained by the Urban Institute. It is one of the earliest and most extensive models. TRIM simulates the major governmental tax, transfer, and health programmes that affect the US population, and can produce results at the individual, family, state, and national levels. Since the first TRIM model was made operational in 1973, TRIM models have been used to understand the potential outcomes of public policy changes such as welfare reform, tax reform, and national health care reform (Beebout and Bonina, 1973).

The UK static model **POLIMOD** utilises the microdata from the 1991 Family Expenditure Survey (FES) to demonstrate how VAT, national insurance contributions and local taxes are calculated under different assumptions. It also presents a method for incorporating entitlement to retirement pension and other non-means-tested social security benefits (Redmond *et al.*, 1998).

**STINMOD** from Australia is APPSIM's static MSM of the Australian tax and transfer systems. In **STINMOD** the rules of government programmes are applied to individuals. These results can then be aggregated to calculate outcomes for income units, families, or households. This gives flexibility in choice of analysis unit. The choice of unit in **STINMOD** depends on the purpose of the analysis and assumptions about financial relationships within units, e.g. pooling of income, sharing of expenditure (Lambert *et al.*, 1994).

**ITEP** is the US model for calculating revenue yield and incidence of federal, state and local taxes by income group. It calculates revenue yield and proposed amendments to current law. Separate incidence analyses can be done for categories of taxpayers specified by marital status, by the existence of children in the family and by age. To forecast future revenue and incidence the model relies on government or other widely respected economic projections (Ettliger and O'Hare, 1996).

**EUROMOD** is a static tax-benefit model that covers 15 European states. It provides estimates of the distributional impact of changes to personal tax and transfer policy at either the national or the European Level. Therefore **EUROMOD** can be used not only to evaluate national tax-benefit policies within a European perspective, but in evaluating policies at the level of the European Union (Sutherland, 2001).

#### 2.3.3.2 Pension MSMs

More recent interest in pension studies has promoted models such as **PRISM** (Pension and Retirement Income Simulation Model), a derivation of **DYNASIM** (Kennell and Sheils, 1990). **PRISM** was developed in the early 1980's and this dynamic MSM models income for individuals aged 25 or over from social security, earnings, assets, public and private occupational pensions and retirement savings plans. It also simulates taxation and the utilisation and financing of long term health care.

The German model **SfB3** has been used to analyse pension reforms, the effect of shortening worker hours, distributional effects of education

transfers and inter-personal redistribution in the state pension system. This model provides three models to investigate different aspects of the distribution of income: long and short term projections, cross-section and lifecycle analyses (Galler and Wagner, 1986).

**PENSIM** (Curry, 1996) has been used to simulate UK pensioners' incomes up to the year 2030 and to facilitate pension reform. The model tries to predict the future of the population rather the future of any individuals. It is operated by the UK Government's Department of Social Security (now the Department of Work and Pensions) and focuses in particular on:

- the treatment of pensioners by the social security system;
- the regulations and coverage of private pension schemes and the performance of pension funds investment portfolios;
- projected demographic movements and
- movements in aggregate variables such as unemployment and interest rates.

**LIFEMOD** (Falkingham and Hill, 1995) was developed at the London School of Economics to model the lifetime impact of a welfare state. LIFEMOD simulates individuals and does not model the characteristics of household members except children.

#### *2.3.3.3 Health MSMs*

In 1997-98, the Pharmaceutical Benefits Scheme (**PBS**) model was developed using STINMOD as a base and then added data from the National Health Survey (NHS) about usage of prescribed pharmaceuticals according to age, gender and concession cardholder status. The model simulated expenditure on pharmaceuticals by different types of households, the resultant government outlays under the PBS and the remaining out-of-pocket costs (patient co-payment contributions) to both concessional and general patients (Walker, 1998).

**LIFEMOD** has been used to examine health status over the life-course and implications for health care financing in the UK (Propper, 1995). The distributions of health care finance and its relationship to income, and the relationship between the health care receipt and need have been studied using the LIFEMOD simulations.

**DYNACAN** has been used to assess the private pension in Australia. Morrison (2003) believes that simulation of private pensions is essential for meaningful policy studies of government programmes in the context of the broader retirement income system. DYNACAN's pension module considers both own-retirement pensions and survivor pensions and projects the incidence, average levels and variation in private pensions into the future. It models the proportion of the population receiving private pensions as a function of birth-year, age, and gender. The validation of the pensions against historical data and feasibility for the quality of the projection into the future has also been briefly addressed.

The **LifePaths** modelling framework in Canada has been used to examine time use issues (Wolfson and Rowe, 1998). LifePaths is a dynamic longitudinal MSM of individuals and families. Using behavioural equations estimated using a variety of historical micro-data sources, LifePaths creates statistically representative samples consisting of complete lifetimes of individuals. The model's behavioural equations generate, at sub-annual resolution, the discrete events that together constitute an individual's life history. In addition to its longitudinal capabilities, a complete set of overlapping cohorts allow LifePaths to produce accurate and representative cross-sectional results from the year 1971 onwards.

#### 2.3.3.4 *Transport MSMs*

MSMs have been developed in transport field for various purposes. Often they can be used for both transport policy assessment and simulation of or part of a transport system. **DRACULA** (Liu *et al.*, 1995) is a traffic MSM. It can represent the progress of and interactions between individual vehicles as they pass through a road network. DRACULA provides the animation of



vehicle flows to give insight into the response of traffic to different network layouts and control strategies. DRACULA is aimed primarily at the kind of operational and time-varying problems, which essentially static models like SATURN find difficult to address. Network performance can be measured by the simulation outputs of the average travel time, speed, queue length, fuel consumption and pollutant emission over regular time periods defined by the user.

**Paramics** (Laird, 1999) is a suite of microscopic simulation modules that models a range of real world traffic and transportation problems. Paramics is fully scalable and designed to handle scenarios ranging from a single intersection, to a congested freeway or the modelling of an entire city's traffic system.

**VISSIM** models transit and traffic flow in urban areas as well as interurban motorways on a microscopic level. It is a commercial product with continuous add-ons provided by various research institutions. The traffic flow model of VISSIM is a discrete, stochastic, time step based microscopic model, with driver-vehicle-units as single entities. The model contains a psycho-physical car following model for longitudinal vehicle movement and a rule-based algorithm for lateral movements (lane changing). Car following and lane changing together form the traffic flow model and that is the kernel of VISSIM (PTV AG, 2011).

**AIMSUN** tries to reproduce real traffic conditions in an urban network which may contain both expressways and arterial routes. It is a combined discrete-continuous simulator: there are some elements of the transportation system (vehicles, detectors) whose state changes continuously over the simulated time period, while there are other elements (traffic lights, entrance points) whose state changes discretely at specific points during the simulation time. It provides detailed modelling of the traffic network: it distinguishes between different types of vehicles and drivers; it can deal with a wide range of network geometries; it can also model incidents and conflicting manoeuvres (Barceló *et al.*, 1999).

**TRANSIMS** is one part of the multi-track Travel Model Improvement Programme in U.S. This set of integrated analytical and simulation models and supporting data bases deal with individual behavioural units and proceed through several steps to estimate travel. TRANSIMS predicts trips for individual households, residents and vehicles rather than for zonal aggregations of households. A regional microsimulation executes the generated trips on the transportation network, modelling the individual vehicle interactions and predicting the transportation system performance. Motor vehicle emissions are estimated using traffic information (TRANSIMS TRAVELOGUE, 1996).

### *2.3.4 Types of MSMs*

Having reviewed an extensive set of MSMs, we are now in a position to classify and compare them. There are two main types of MSMs: static and dynamic. “static” and “dynamic” are used in different ways by authors, but we will try to identify the main differences between the two types of MSMs and provide a summary of the comparison in this section.

#### *2.3.4.1 Static MSM*

Generally speaking, static models do not have direct interactions of microanalytic units within the context of the model during the time period simulated. Static models normally are either deterministic or stochastic. In a static microsimulation, change of the demographic structure in the model is performed by static ageing techniques. Typically such techniques take a large representative sample with detailed information and apply a set of rules to the sample to generate the synthetic demographic and economic characteristics expected in the future year. Simulations can estimate the impact of a change in the future year. As the change of the demographic structure of the modelled population is performed by reweighting the age class according to external information, it is focused on what consequences of external information brings to the population and therefore it does not model the changes in population itself. A typical “What-if” Static MSM scenario would be: if there had been no poll tax in 1991, which communities

would have benefited most and which would have had to have paid more tax in other forms (Ballas *et al.*, 2005b; Gilbert and Troitzsch, 2005; Citro and Hanushek, 1991)?

Most tax-benefit MSMs are static. Examples of static microsimulations include models such as TRIM (Beebout and Bonina, 1973), POLIMOD (Redmond *et al.* 1998), STINMOD (Lambert *et al.*, 1994) and EUROMOD (Sutherland, 2001). Descriptions of such models can be found in the previous section.

Some researchers use the static MSM to synthesise populations from sample records (Birkin and Clarke, 1995; Williamson *et al.*, 1998). Such models typically attempt to fit disaggregate samples from microdata such as census records to aggregated constraints (sometimes called marginals or control totals, e.g. total males in area) from other aggregate empirical data. Important attributes from the samples are selected as the control variables. Often the Iterative Proportional Fitting procedure (IPF) is used to calculate conditional probabilities of having particular attributes and these attributes are assigned to individuals on the basis of random sampling procedures such as Monte Carlo simulation. Alternatively population reconstruction can be achieved by reweighting techniques (Ballas *et al.*, 2005b). The base population of year 2001 used in this model is recreated from a static MSM by Prof. Birkin (Birkin *et al.*, 2006; Birkin and Clarke, 1995).

#### 2.3.4.2 *Dynamic MSM*

Dynamic MSM can be considered as a technique where entities change their characteristics as a result of endogenous factors within the model. Various degrees of direct interaction between micro population units can be found in dynamic MSMs. Such interaction typically includes processes such as birth and marriage. Dynamic microanalytic models rely on an accurate knowledge of the individuals and the dynamics of such interactions. In a dynamic MSM, the updating of the demographic structure is performed by ageing the modelled population individually (by asking “yes or no” questions on birth, death, marriage) with transition probabilities according

to life tables and/or exogenous time series. Thus the changes in the population itself are modelled and the simulation in one year may affect an individual unit's characteristics in the subsequent year. A typical future-oriented "what if" Dynamic MSM Scenario would be: if the current government had raised income taxes in 1997, what would the redistributive effects have been between different socio-economic groups and between central cities and their suburbs by 2011 (Birkin *et al.*, 1996; Ballas *et al.*, 2005b; Gilbert and Troitzsch, 2005; O'Donoghue, 2001)?

Examples of dynamic MSMs include the following,

**DYNAMOD** in Australia (Harding, 2002) is the first MSM using discrete event simulation features. The model starts with the 1986 census one per cent sample (about 160,000 individuals) and ages each of those individuals on a monthly basis for up to about 60 years. Dynamic models are particularly useful for looking at the likely future or long-term impacts of government policy or current social and economic trends. Recently assets and superannuation has been added to the DYNAMOD model to facilitate the research of the likely future retirement incomes of Australians.

A major update of **DYNASIM** (Orcutt *et al.*, 1976) has resulted in DYNASIM3. This dynamic MSM has been used to study the distribution of retirement and ageing. The model ages the individual and family data by year, simulating demographic events as births, deaths, marriages and divorces and economic events as labour force participation, earnings, hours of work, disability onset, and retirement. DYNASIM3 models a wide range of topics, including Social Security coverage and benefits, pension coverage and participation, benefit payments and pension assets, as well as home and financial assets, health status and living arrangements (Favreault and Smith, 2004).

**SAGEMOD** (Zaidi, 2004) is a dynamic demographic/tax model. *It estimates the incomes* as the established practice of other MSMs (e.g. DYNAMOD, PENSIM) and also estimates a random-effects cross-sectional wage equation which included some individual wage history data with the

error components. The impact of other labour market states (unemployed, inactive, student) in previous years has been investigated on the earnings of currently employed individuals.

### 2.3.4.3 Comparison of static and dynamic MSMs

As described before, there are a couple of important difference between the static and dynamic MSMs. In Table 2.1, we try to summarise the main dimensions of such difference between the two types of MSMs.

Static and dynamic MSMs each have their own strengths. Static models are regarded as more effective at times for specific short run projection purposes because of their greater simplicity and the often lower costs associated with building such models and obtaining computer generated model solutions. Another advantage of static models is that they have very detailed programme simulations. From the view of the computation, static MSMs demand less computing resource.

**Table 2.1 Comparison of static and dynamic MSMs**

Type of MSM	Characteristics	Ageing technique	Entity Interactivity	Time	Population Change	Impact of previous step on the next
<b>Static</b>	Deterministic / Stochastic	Static ageing	No	No time element/ stocks of entities updates	No	No
<b>Dynamic</b>	Stochastic	Dynamic ageing	Possible	Change processes and events built in	Yes	Yes

In contrast, dynamic models feature more detailed and realistic population changes. There is general acceptance that dynamic models provide a more realistic representation of micro population unit behaviour. Dynamic models are also viewed as better at producing realistic long-term estimates, which account for interim changes in economic and demographic trends

(O'Donoghue, 2001). Due to the interactions / interdependencies of the updating, one limitation is that dynamic MSMs are computationally demanding, even for high-speed modern machines (Ballas *et al.*, 2005b; Gilbert and Troitzsch, 2005; Citro and Hanushek, 1991; McDonald, *et al.*, 2006).

#### 2.3.4.4 *Spatial MSMs*

Spatial MSM is a special type of MSM that simulate virtual populations in given geographical areas (Ballas *et al.*, 2005b). In a spatial MSM, local contexts can be taken into account when studying the characteristics of these populations. Such MSMs are concerned with the creation of large-scale datasets estimating the attributes of individuals within the study area and are used to analyse policy impacts on these micro units (Birkin and Clarke, 1995; Clarke, 1996). Spatial MSMs therefore have advantages over other MSMs in exploration of spatial relationships and analysis of the spatial implications of policy scenarios. A spatial MSM can be either static or dynamic.

Spatial MSM was first studied by Hägerstrand (1957) in the 1950s who introduced the spatial and temporal dimensions into social studies. Geographers such as Wilson (1967), Clarke (1996) and Birkin and Clarke (1995) extended the theoretical framework over the years. Various spatial microsimulations have been developed, including both static and dynamic microsimulations. They allow data from various sources to be linked and patterns to be explored at different spatial scale with re-aggregation or disaggregation of the data. Furthermore they allow updating and projecting, which is of particular importance in forecasting future patterns (Clarke, 1996; Ballas and Clarke, 2001).

An examples of such a model is **SVERIGE** in Sweden (Rephann, 1999). This dynamic population model is designed to study human eco-dynamics (the impact of human cultural and economic systems on the environment). Its main distinguishing characteristic is that it simulates spatial location and mobility of every individual in the data. The model took the CORSIM

model framework as a starting point, adapting behavioural modules to Swedish society and datasets. The migration module attempts to model locational transitions to an accuracy of 100m.

**SimBritain** (Ballas *et al.*, 2005c) is a dynamic simulation attempting to model the British population at different geographical scales up to the year 2021. Datasets used in this model are the 1991 UK Census Small Area Statistics (SAS) data and the British Household Panel Survey (BHPS). Microdata for all wards in Britain have been generated through re-weighting the original BHPS data. Previous data from 1971, 1981 and 1991 census Small Area Statistics (SAS) have been used for projections of a set of small area statistics. Using these three time points, a trend curve was produced allowing tables to be predicted up to 2021.

**SMILE** in Ireland (Ballas *et al.*, 2005a) is a dynamic spatial MSM designed to analyse the impact of policy change and economic development on rural areas in Ireland. The core model of SMILE is a demographic model. It simulates the basic components of population change, fertility, mortality and internal migration and projects population change at the sub-county level.

## 2.4 Cellular Automata and Agent Based Models

Cellular Automata (CAs) and Agent Based Models (ABMs) are two other important IBMs that are closely related to MSMs. The CAs and ABMs differ to the traditional policy assessment MSMs in the way that most of these models are more focused on simulations based on individual unit behaviours to model a complex system in the real world, instead of policy assessment and development as the traditional MSMs do, although CAs and ABMs can also be used for such purpose.

### *2.4.1 Cellular Automata(CA)*

CA models are discrete dynamical systems where behaviour is completely specified in terms of a local relation. As the name indicates, space in a CA model is represented as a uniform lattice of cells with local states. A uniform set of rules compute the state of a particular cell as a function of its previous state and the states of the adjacent cells and thus drives the behaviour of the system (Dijkstra, 2000).

As Von Neumann and Burks (1966) pointed out, Ulam developed the first work in which the terms “cellular space” and “automata networks” appeared. Von Neumann extended this work and proposed CA as a way to model self-reproducing biological systems. This concept did not come out of laboratory until Gardner’s article (1970) about the “Game of Life” was widely read among computer scientists. In this article the best-known example of CA, the so called “Conway’s Life” by J.H. Conway has been introduced. In the CA model of Conway’s Game of Life, cells can be in one of two states: alive or dead. Each update of the global state represents a new generation of organisms. From a theoretical point of view, its capacity of a universal Turing machine has an important indication: anything that can be computed algorithmically can be computed within Conway's Game of Life.

Langton (1991) and Toffoli and Margolus (1987) also produced important works in CA, but it is Wolfram who further extended the theoretical and experimental development of CA. Wolfram (1984) proposed a classification scheme which divided cellular automata rules into four categories: homogeneous, regular, chaotic and complex. Evolutions of the four types of CA lead to a homogeneous state, a set of separated simple stable or periodic structures, a chaotic pattern, or complex localised structures with complex behaviour, stable or periodic structures which persist for an infinite time.

Previous examples demonstrate that CA based modelling can facilitate the understanding of social dynamics, as locality, overlapping neighbourhoods, and repeated interaction are significant in dynamic social processes. The CA



approach, however, does not exclude centralised institutions in the model, either. CA based models can include those features and are particularly well suited to contribute to understanding of micro/macro relations (Schelling, 1971). However, CA based modelling tends to focus on the dynamics of elementary social interactions and requires sufficient understanding of all components to make sense. Such models therefore tend to be found in more mature sciences than in the social sciences (Kliemt, 1996).

### *2.4.2 Agent Based Models (ABM)*

Agent based modelling is an alternative approach that models social life as interactions among intelligent agents and is related to traditional sociological systems that model social processes as interactions among variables. ABMs originate from a sub-field of AI (Artificial Intelligence) and it is concerned with the behaviour of multiple agents working together to solve a given problem that is beyond the capability or knowledge of individuals (Jennings, 2000). The agents within such a system can interact with each other and the environment that they live in. Through their interactions, simple and predictable local interactions can generate familiar but unpredictable global patterns, such as the diffusion of information, emergence of norms, coordination of conventions, or participation in collective action. ABMs provide theoretical leverage where the global patterns of interest are more than the aggregation of individual attributes. The emergent pattern can be understood as a product of a bottom-up model of the micro-foundations at the relational level. Emergence is one of the properties common to most agent related researches. Any apparently organised or unexpected meta-behaviour that arises from the repeated application of simple micro-level rules, when this behaviour was not explicitly placed in those rules, can be considered as emergence. Its importance and utility is seen as a parallel to natural life that emerges out of the organised interactions of large numbers of non-living molecules without the need for any form of global controller responsible for the behaviour of every part (Russell and Norvig, 1995; Conte and Gilbert, 1995).

ABM has similar characteristics to CA. Time proceeds iteratively through discrete steps in both ABM and in CA. Like CA, agents in ABM also exist in some environment. Depending on the types of the agents, this environment can be some communication networks, or a computer's memory storage unit, or an artificial urban environment.

However, what differentiates ABMs and CAs are that agents in ABMs are free to navigate and explore their spatial environments, as their spatial behaviour is not constrained by a lattice as in CAs. Also the interaction of the agents can go beyond the neighbourhood. Agents in ABM have true mobility. Information exchange is mediated through the neighbourhood in CA models. In ABM, agents can interact with other agents as well as with their environments. Computer languages such as KQML (Knowledge Query and Manipulation Language) have been developed for such communication. CA is driven by state transitions, but the behaviour of agents in ABM are governed by different rules, depends on the goals, preferences, decisions or plans of the agents.

ABM research is a very active field. However, generally speaking, the research is still at an early stage and many important questions are still being studied or need further study. Currently, agent-related research includes its theory, communication and interaction technology, agent architecture and organization, agent language, cooperation and negotiation between the agents.

#### *2.4.2.1 Agent theories*

As Wooldridge and Jennings (1995) have pointed out, although the term of "Agent" is widely used, a single universally accepted definition has not been produced. They distinguish two general usages of the term of "Agent" in terms of a weak notion and a strong notion. In the weak notion of agency, an agent can have the following properties:

- *autonomy*: agents operate without the direct intervention of humans or others, and have some kind of control over their actions and internal state;
- *social ability*: agents interact with other agents (and possibly humans);
- *reactivity*: agents perceive their environment and respond to changes;
- *pro-activeness*: agents are able to exhibit goal-directed behaviour.

Under the stronger notion of agency, in addition to having the properties identified above, agent is either conceptualised or implemented using concepts that are more usually applied to humans, such as such as knowledge, belief, intention, and obligation. (Wooldridge and Jennings, 1995).

In 1987, Bratman's BDI (Belief-Desire-Intention) theory had an intensive influence on AI research. Since then it has been considered as one of the important theoretical foundations of DAI (Distributed Artificial Intelligence). The BDI model of human practical reasoning was developed as a way of explaining future-directed intention where the notion of intention was seen as a way of limiting time spent on deliberating about what to do by eliminating choices inconsistent with current intentions (Bratman, 1987). Rao and Georgeff (1991) have provided a formal logical description of BDI theory on the basis of a branching model of time, in which belief-desire-intention-accessible worlds are themselves branching time structures. They are particularly concerned with the notion of *realism* - the question of how an agent's beliefs about the future affect its desires and intentions.

Other important agent theories include the following:

- Kripke's (1963) Possible World Semantics that sees agents' beliefs as a set of characterised *possible worlds*;

- Moore's (1990) Knowledge and Behaviour (Knowledge & Action) logic where the main concern is what an agent needs to know in order to be able to perform some action;
- Cohen and Levesque's (1990) Intention theory has proved to be useful for reasoning about agents and in the theoretical foundations of cooperative problem solving and
- Konolige's (1986) logical interpretation model (Deduction Model of Belief) theory is a direct attempt to model the "beliefs" of symbolic AI systems, where an agent would apply its inference rules wherever possible to generate the *deductive closure* of its base beliefs under its deduction rules.

As the main purpose of the study is to study the population evolution, it is decided that a weaker notion of the agency is sufficient and the individual BDI will not be the focus of this study.

#### 2.4.2.2 Agent architecture

Agent architecture is used to describe the basic components of the agents and their functions, the association and interaction mechanism between the components, how agents act based on the perception of their internal state and external environment and the impact of agents' behaviours on their internal state and external environment. Wooldridge and Jennings (1995) classified the present agent architecture into the following three broad categories: deliberative architecture, reactive architecture and hybrid architecture.

**Deliberative Architecture** is that the agent has an explicit symbolic model of the world and its logic of decision-making is based on pattern matching and symbol operation. The cognitive component of such agent is composed of a planning module and a model of the world. The assumption here is that the cognitive function of the agent can be modularised. That is, we can study the different cognitive functions separately and then assemble them together to form the intelligent autonomous agent (Jennings, 1992).

The complexity of symbolic computation algorithms means a lot of computation time even when the simplest theorems are used. Therefore it leads to questions such as: How can such a structure in a certain period of time be translated into an accurate and appropriate symbolic description of the real world? How in a certain period of time does the agent use such information for reasoning and decision-making? In addition, Brooks (1991) thinks that the level of human intelligence is too complicated to decompose it correctly. Even when decomposition is correct, we still do not know the interface between these elements.

**Reactive Architecture** does not use symbolic models of the world or a complex symbolic reasoning system. When the internal and external environment of an agent meets certain conditions, the agent would act accordingly. The assumption of Reaction Architecture is that the complexity of the agent can be a reflection of the complexity of operating environment, rather than a reflection of agents' complex interior design. The test showed that in the real world, reactive agents are superior than the deliberative agents when dealing with a limited number of tasks. However, reactive agents have some problems when handling tasks that require knowledge of the world, as such knowledge can only be obtained through reasoning or memory, rather than by the perception. Also the reactive agents have poor adaptability and are often without learning ability. Each act of the agents needs to be coded separately and this can be problematic in terms of the extendibility of the system. Reactive Agents currently dominate mainstream distributed systems (Brooks, 1991).

**Hybrid Architecture** is the third category. Neither a completely deliberative nor completely reactive approach is appropriate in a complex system. Many researchers suggest combine the two together. One example of the hybrid architecture is Georgeff and Lansky's (1987) PRS (Procedure Reasoning System) structure. PRS is an information-desire-target system. PRS has a plan library as well as explicitly expressed beliefs, desires and goals. PRS plan library contains incomplete plans, as Knowledge Areas (KAs). KAs are linked to the activation states that will decide when the KAs

to be activated. KAs can be activated by targets or by data, therefore, KAs can be deliberative or reactive.

This study chose to use the hybrid approach, as the deliberate architecture allows compatibility with the probability driven microsimulation architecture, while reactive architecture is useful during interactions between the agents themselves, as well as between agent and the environment.

#### *2.4.2.3 Agent application areas*

Multi-agent systems offer strong models for representing real-world environments with an appropriate degree of complexity and dynamism. Agent-based simulation is characterised by the intersection of three scientific fields, namely agent-based computing, the social sciences, and computer simulation (Luck *et al.*, 2003).

ABMs have been used in a wide range of studies in representing large scale complexity with individual heterogeneous behaviour. It has been used to study the Internet and other communication networks, in distributed computing studies, in artificial neural network researches and even in animation and interactive multimedia. ABMs have been built for subjects in Ecology and Biology, from fish, birds, insects to mammals. However, only recently has this approach become popular in social science. Only recently have there been developments in modelling humans and artificial societies in disciplines such as Anthropology and Sociology. Various models have been developed to assist studies in Economics, traffic and vehicle simulations, urban planning, pedestrian research and emergency and evacuation studies (Reynolds, 1999; Batty and Densham, 1996; Schelhorn, 1999; Helbing, 2000; Raney and Nagel, 2003). ABMs normally are composed of multiple agents, so we will focus on Multi Agent Systems (MASs) in this review.

Rouchier (2001) identified the main application areas of MASs as follows:

1. MASs allow distribution of the problems or problem solving and this feature makes it very flexible in problem solving. As an alternative to centralised problem solving, distributed agents cooperate to solve problems efficiently. Sometimes it may be the only way to solve social science problems.
2. MASs provide testbeds for theories about local behaviours through artificial universes.
3. Synthetic Worlds/Artificial Societies made up by agents can be used to describe specific interaction mechanisms and analyse their impact at a global level in the system.
4. MASs can be considered as Collective Robotics, in which tasks can be divided into specific goals for each robot. The accomplishment of the task can then be achieved through completion of all the small tasks.
5. Genetic Programme Design: MASs can also be seen as a very efficient modular way to programme. Genetics will probably be the new science that deals with collective action and interaction.

Social science systems are often large scale systems without clear boundaries in between. Such complexity of social systems requires sophisticated methods for understanding and predicting what can happen. Features of MASs as described above make them useful for facilitating social science studies.

#### *2.4.2.4 ABM frameworks*

There are a variety of ABM frameworks. The major ones with a focus on MAS are listed below.

**Swarm** (Minar *et al.*, 1996) is probably the most famous agent-based development environment. Many later ABM framework developments borrow ideas from Swarm. It has been used in all areas of science and there are a couple of extensions of it such as Kenge and MAML (Multi-Agent Modelling Language). The original development platform was Objective-C, but it also has a java version. Since SWARM has been used in many

disciplines, the standard version is not focused on a specific field of application.

Swarm has been used as a standard toolkit by researchers and brought significant benefit to the researchers in building and testing simulations without starting from scratch. One of the earliest demonstrations produced using Swarm is called Heatbugs. In this model multiple agents act as Heatbugs. Each tries to find a spot on the lattice with their ideal temperature and diffuses a small amount of heat into the environment they are leaving. After running the simulation for some time, the bugs forms into groups of different sizes where the heat given off by bugs in the group keep each other at their ideal temperature. There are many MASs that have been developed using this framework. Swarm is one of the earliest and most used agent frameworks.

**Repast** (REcursive Porous Agent Simulation Toolkit) was developed at the University of Chicago and operates in much the same way as the Swarm libraries (Robertson, 2003). Compare to Swarm, it is more focused on social science applications. This set of JAVA libraries allows programmers to build simulation environments (e.g. regular lattices), create agents in social networks, collect data from simulations automatically and build user interfaces easily. It has a model-building application, Evolver, to let you build the infrastructure of your model by dragging-and-dropping components in a GUI (Graphic User Interface). Repast has been used extensively in the social microsimulation modelling and it has integral statistical libraries and GIS connections. Generally speaking Repast can be considered as a social simulation framework that is more focused on network dynamics and more comprehensive agents.

Like Swarm, Repast also provides demonstrations such as Heatbugs, but unlike Swarm, Repast is fully object-oriented and Repast J is developed in pure JAVA to allow ease in development. Repast 3 also provides a lot of useful built-in functions, including:



1. a fully concurrent discrete event scheduler that allows both sequential and parallel discrete event operations;
2. simulation results logging and graphing tools;
3. an automated Monte Carlo simulation framework;
4. a range of two-dimensional agent environments and visualizations;
5. libraries for genetic algorithms, neural networks, random number generation, and specialized mathematics;
6. integrated geographical information systems (GIS) support.

Another feature of Repast is that Repast is available on virtually all modern computing platforms including Windows, Mac OS, and Linux and it provides platform support for both personal computers and large-scale scientific computing clusters (Collier *et al.*, 2003).

**StarLogo** (Colella *et al.*, 2001) was developed at the MIT Media Lab, originally to explore parallel simulations on personal computers and help children learn simulation and complex systems. It can be used as a CA or agent development framework and is based on the Logo language for robotic ‘turtles’. It allows you to set up individual agents and their environment on a grid-based system, has basic attributes and resources built into turtles and the environment they live in, and a simple scripting language for more complex development. It has two variants: StarLogoT and NetLogo.

**Ascape** (Parker, 2000) can be either a CA or agent frameworks, as integrated agent framework that will run models on a grid or other topologies. Ascape is a very flexible tool developed in JAVA. It is designed explicitly for social science simulation and famous for Sugarscape development environment at the Brookings Institution. Like Repast, Ascape also has built-in statistical libraries and GIS connections, but it concentrates on the ability to create simple models easily.

**Aglets** (Clements *et al.*, 1997) was developed by IBM. Aglets includes both a complete Java mobile agent platform, with a stand-alone server called Tahiti, and a library that allows developer to build mobile agents and to

embed the Aglets technology in their applications. This robust agent-based simulation development environment is mostly used for intelligent agent design rather than social science simulation. However, it should be able to be used for the latter purpose.

**JADE** (Java Agent DEvelopment Framework) was developed at the Telecom Italia Lab (Moraitis and Spanoudakis, 2004). JADE also focuses on intelligent agents rather than social simulation. It claims to comply with the FIPA specifications through a set of tools that supports the debugging and deployment phase. The agent platform can be distributed across machines and the configuration can be controlled via a remote GUI.

Due to the large scale and the complexity that our model in this study is attempting to model, although RePast has been considered to be one possibility, it has been decided that a model developed from the scratch will allow us more flexibility in terms of the functionalities required by the hybrid model, as well as making the most of the computing power in simulation efficiently without complex utility packages such as charting.

## 2.5 Multi-level Simulation models

As macrosimulations and microsimulations have different features as described previously, multi-level simulations attempt to link the two types of system together in order that the interactions on different levels within a system can be captured.

In a typical MSM, the impact between the policy and population is only focused on one direction and the impact of population changes on the public policy is not modelled. While multi-level simulations model the interactions in both directions. Normally in a multi-level simulation, a master equation or a system of differential equations for the time dependent distribution of populations will be derived from individual transition probabilities in order to model the studied system at the macroscopic level as well (Gilbert and Troitzsch, 2005).

MIMOSE (MIcro and Multilevel MOdelling SoftwarE) consists of a model description language and an experimental frame for the simulation of models. MIMOSE considers special demands of modelling in social science such as birth and death processes and allows micro and multilevel models. Its approach of three-level treatment (macro, micro and cross levels) is very useful to readers interested in multi-level approaches. The modelling approach is tested on different simulation models, including the assessment of the sequence of military interventions and democratisation in Thailand (Möhring, 1996).

Norwegian researchers developed an integrated Micro-Macro Computable General Equilibrium (CGE) model that is composed of a detailed micro-econometric simulation model of labour supply. CGE attempts to explore impact and interactions between heterogeneous individual labour supply behaviours and the policy problems of fiscal sustainability and welfare state caused by ageing (Aaberge *et al.*, 2003). In this simulation model, the microsimulation captures the heterogeneity in individual labour supply behaviour and an accurate description of the personal tax system. The CGE model provides a detailed account of the endogenous changes in other tax bases and government expenditures. This model demonstrates the importance of both general equilibrium effects and individual behaviour. The relative importance of the macro or micro simulation seems to vary depending on the point in time and the policy environment in the model.

The Public Service Human Resources Management Agency of Canada simulates personnel models: one at the macro level and another at the micro level (Treasury Board of Canada Secretariat, 2000). In an HR environment, the software of ITHINK has been used to simulate scenarios at the macro level based on current and historical patterns of movement and transactions (e.g. recruitments, promotions, separations and retirements). While microsimulation using PERSIM offers the view of the individual career paths to allow the analyst to look at how individuals move through the organization, how policy alternatives affect these movements and how the effects vary among individuals even under the same set of assumptions.

This combined macro-micro model therefore generates many simulations under the same set of assumptions and captures the macro results of each simulation. The user can then compare an individual simulation to the expected average and analyse the statistical variation of the average results.

Magne *et al.* (2000) use a hybrid macro-micro simulator to model the traffic flow. This simulator models a large transportation network at the macroscopic level with selected sections at the microscopic level. The compatibility problems introduced by simulations at two different modelling levels have been resolved by ensuring that the microscopic model will have the same macroscopic characteristics as the macroscopic one. Both the macroscopic view of the whole transportation network and some detailed microscopic information of traffic flow on certain sections of the network can be obtained from this system.

As a population model, MicMac was developed in the attempt to provide a bridge between aggregate projections of cohorts (Mac) and projections of the life courses of individual cohort members (Mic). Mac focuses on transitions among functional states by age and sex to produce cohort biographies, while Mic addresses demographic events and other life transitions at the individual level, using a multistate model to produce individual biographies. Thus it tries to bridge the gap within the traditional macroscopic age-sex based population projections by providing the cohort heterogeneity or intra-cohort variation (Willekens, 2005). Both Mic and Mac are both based on transition rates that depend on covariates. The problem with time-varying covariates is that the parametric estimation of the transition rates is complicated or even impossible. This is particularly problematic in Mic. However, to achieve the consistency between the projections of Mic and Mac, the cohort transition rates in Mac are set similar to the expected values of the individual transition rates of Mic (van der Gaag *et al.*, 2005). This, combined with the cohort based approach, may reduce the effectiveness in modelling the individual characteristics.

Some studies of population health/behaviours also tried the multi level approach to capture environmental impact on the health outcomes/behaviour

in areas. Such models often consider cases that can be explained by various individual factors (e.g. age and sex), then introduce area factors (e.g. deprivation) to explain the remaining cases. The new prevalence estimates are then estimated on the basis of such individual and area factors (Moon *et al.* 2007, Pearce and Witten, 2010, Twigg and Moon, 2002). However, such estimates remain as the zonal average that cannot truly reflect the individual characteristics. More importantly, if the zoning systems used in the area factors are different, then the results can result in the wrong spatial patterns. The individual based MSM, however, can overcome such problems by linking data from various sources by a common individual variable (Smith *et al.*, 2010).

## 2.6 Hybrid modelling approach

### *2.6.1 Individual based approach*

IBMs study systems at individual levels and try to capture the individual features. In their originally conceived forms, Microsimulation Models (MSMs), Cellular Automata (CAs) and Agent Based Models (ABMs) may be regarded as the three important types of IBMs.

Although MSMs, CAs and ABMs each have different focus, they all model the studied system at individual levels and there is some common ground among the three approaches. First, all three approaches are simulations based on the global consequences of local interactions of members of a population. Unlike the aggregated models that often overlook the details at more refined level, they provide a more effective and natural way to handle individual behaviours. Second, these three approaches all track the characteristics of each individual through time, in contrast to traditional modelling techniques where the characteristics of the population are averaged together. Finally the emergence of the global phenomena through local interactions in all IBMs offers more than changes that are simulated on the basis of average data for the whole population in the traditional models.

## *2.6.2 Limitations of MSM and ABM*

As discussed previously, MSM and ABM both have many valuable features. However, they each also have limitations. In the following section, the limitations of MSM and ABM will be discussed separately to reveal the potential of bring the two approaches together to complement each other to address the limitations.

### *2.6.2.1 Limitations of MSM*

Over the years, microsimulation has been proved to be a successful approach modelling social science, especially in facilitating the public policy making and development. Large scale MSMs enable us to explore the interaction between policy changes and narrowly defined ranges of individuals or demographic groups, yet retain the heterogeneity in the population as revealed in the large household surveys. The capability of MSMs to replicate the complex policy structures also allows us to forecast the outcomes of policy changes and ‘what if’ scenarios.

The main criticised points of MSM include:

- MSMs require large datasets with high quality;
- MSM developments are normally computing intensive;
- large scale MSMs can take a long time and considerable effort to accomplish;
- MSM only models one-direction interactions: the impact of the policy on the individuals, but not the impact of individuals on the policy;
- MSMs are less strong in modelling interactions and behaviours and
- it is difficult to validate MSMs (Krupp, 1986; Williamson, 1999; Citro and Hanushek, 1991; O’Donoghue, 2001; Gilbert and Troitzsch, 2005).

### 2.6.2.2 *Limitations of ABM*

With the advance in computing, the first three limitations have been improved greatly and new technologies: ABM can provide the capability for behaviour modelling and allow us to study the interaction at both macro and micro levels, as well as interactions in both directions. However, despite the usefulness of the ABM as described in previous discussion, being a relatively new technology, sometimes it is found that it lacks more refined and well-established theories and concepts as demonstrated in MSM (Gilbert and Troitzsch, 2005; Conte *et al.*, 1998). ABM is also known as hard to validate. Many applications of agent systems to public or social policy domains involve the development of alternative scenarios to facilitate decision-making. However, there is no formal theory of scenarios and scenario analysis that tell us how to construct scenarios, how many scenarios to construct and how to reason between and across their outcomes.

Despite the work that remains to be done, agent based social simulation can provide insight into the structure and effects of policies and norms and can assist in understanding and exploring interaction patterns and behaviours where appropriate and possible (Luck *et al.*, 2003).

### 2.6.3 *Previous hybrid modelling approaches*

To address the limitations of the individual based models, this study proposes a hybrid modelling approach that brings the strength of the MSM and ABM together. The main reasons for the proposal of such a hybrid approach include:

- MSM and ABM complement each other;
- geography provides a bridge to link the MSM and ABM;
- previous attempts using hybrid approaches have resulted in fruitful outcomes;
- the hybrid approach may provide a new angle to view classical problems.

In the following sections, the features of MSM, ABM and some previous hybrid modelling approaches have been discussed and main findings have been summarised.

#### 2.6.3.1 *MSM and Geography*

Geographical information plays an important role in effecting social progress and welfare. Given the nature of social systems, it would not be complete without considering the spatial impact in a policy microsimulation. When assessing the impact of the policy changes on individuals, many studies have identified that the outcomes do vary spatially (Clarke, 1996; Wu and Hine, 2003). Indeed there is a need to estimate the geographical impacts as well as the socio-economic impacts of policies.

From a means-ends perspective, De Man (1988) suggests that means-ends relationships often have a spatial context, as means need to be employed somewhere. He further argues that information has three dimensions:

1. theme, content or attribute; 2
2. space or location and
3. time.

Many information systems explicitly or implicitly deal with two out of the three. De Man finds typical decision support systems may have mechanisms to represent time and strong attribute manipulation tools, but have no spatial representation. GIS (Geographical Information Systems) however, captures all three dimensions of the information. GIS provide mechanisms for attribute and spatial data management and often also enables temporal data management. He argues that GIS are heavily utilized by technical personnel and it can assist an extensive range of management decisions. While offering the addition of the spatial dimension needed for many managerial analyses, both qualitative and quantitative differences can be presented in GIS, compared to conventional information systems.

Ballas *et al.* (2005b) suggest that there are close linkages between the social



policy and geographical patterns. They argue that social policies can be seen as alternatives to area-based policies and in some instances, spatial impacts of social policies can even be validated through the respective impacts of area-based policy studies. Although area-based policies have a geographical impact by definition, there has been very limited analysis of the spatial impacts of policies that were not designed to have a geographical impact. The authors suggest that spatial microsimulation can also be used for the design of proactive geographically oriented social policies. Such models have demonstrated the strength of the combination of the MSM and GI science. A selection of examples of spatial MSMs has been discussed here.

#### 2.6.3.2 *ABM and Geography*

Torrens and Benenson (2005) proposed a new paradigm for integrating GIS and agent based simulation called Geographic Automata Systems (GAS). This system takes advantage of the formalism of automata theory and GIS to unite cellular automata and multi-agent systems techniques and provides a spatial approach to bottom-up modeling of complex geographic systems that are comprised of infrastructure and human objects. In this framework, geographic phenomena as a whole are considered as the outcomes of the collective dynamics of multiple animate and inanimate geographic automata. Geography serves as the binding force in merging a CA and an ABM (which are popularly confused in the geographic literature). Therefore automata become uniquely geographical, fusing CA and ABM but extending the concept to incorporate notions from GIS and GAS.

Murphy (1995) believes that the evolution of GIS as a decision support system relies on improvements in technology, the creation of new analysis tools, and increased understanding of the interaction between decision support tools and the decision maker. He also points out that particularly fruitful areas may come from the use of artificial intelligence approaches for alternative representation of decision domains and knowledge. He thinks cooperation between the disciplines will be particularly beneficial in areas such as data quality, uncertainty representation and issues related to the management and sharing of large time-reliant and source-dependent data.

Thus, a rewarding exchange may be possible between GI Science and decision support system research streams relating to the management, representation, and interpretation of complex multi-dimensional knowledge.

Gonçalves *et al.* (2004) suggest that GIS and ABM address space in different perspectives: GIS models geographic space and ABM models the behaviour of intelligent agents within geographic space. Gonçalves *et al.* propose a conceptual framework for integrating these different perspectives in the context of modelling and simulation of complex dynamic systems. They suggest that GIS enables the definition of a geographic region to be related with the phenomena in that region, but GIS tools do not seem to be appropriate to study dynamic phenomena in an area. Most ABM tools that use geographic information are not coupled with GIS. However, the simulation of the human behaviour with mobility in geographic space and intelligent behaviour has increased in the recent decades, which has led to a special interest in the integration of agent based models (mainly ABM) and GIS.

The authors proposed that in the hybrid model, ABM can be used to model the intelligent behaviour of entities, e.g. behaviour of people, animals, enterprises, while GIS can be used to model geographic space. Intelligent agents move and reason within this environment. The authors also point out that GIS are already extensively used by people from the natural sciences, civil engineering, territory management authorities, urban planning. Therefore there is no point not to give them what they already know plus the agents.

### 2.6.3.3 *MSM and ABM*

It is generally agreed that microsimulation models (MSMs) provide important and effective tools for modelling in social science. With recent advance in computing and data production, building a MSM has become less expensive than before, as the first three limitations listed above have therefore been improved significantly (Holm *et al.* 1996).

Traditionally dynamic microsimulations are agreed to be better at modelling individual interactions and providing a more realistic representation of micro unit behaviour than the static microsimulations. However, Davidsson (2000) has pointed out that such models have two main limitations in comparison with ABM: such models do not justify the behaviours of each individual in terms of individual preferences, decisions and plans, and interactions between individuals are not modelled in the simulation. He believes that ABM is well suited for the simulation of a large number of heterogeneous individuals with different behaviours and ABM is an ideal simulation method for the social sciences.

Although there are limited examples of using ABM approach in demographic models, Billari *et al.* (2002) consider ABM as a promising approach to help improve our understanding of demographic behaviours. ABM can provide an alternative means to study demographic processes as the outcome of interacting agents. By focusing on the dynamics of the population instead of equilibria, ABM is better suited for modelling demographic processes. The question whether those dynamic changes are due to compositional changes or changing behaviour rules of the individual agents can also be studied within the framework of ABM. It is also useful in explaining individual behaviours by taking into account both micro and macrofactors. For instance, agents in an artificial society can follow their built-in rules while subject to the macro behaviour rules (e.g. conventions, institutions). The explanation of behaviour is based on simple propositions about individual behaviour, but can produce complex situations and feedback at the macro level (Billari *et al.*, 2002).

ABMs are especially helpful in terms of behaviour modelling in a complex system. Social systems are complex because they are composed of different entities having non-linear interactions. ABM can model the structure of the intelligent entities in such complex dynamic systems. In these systems, agents act and interact on behalf of different entities. ABMs have advantages in:

1. modelling the intelligent behaviour of individuals by itself or in society

and

2. improving efficiency by distributing the control of the computation by multiple simpler units evolving through their interactions (Jennings, 2000).

Axelrod (2005) suggested that ABMs can bring different disciplines together. This is because ABMs has several distinctive features:

1. They can address fundamental problems in different disciplines. Axelrod used computer tournaments as an example. Computer models are established on the basis of the strategic understanding of theorists from many disciplines and have a wide range of applications even beyond the social science.
2. They facilitate interdisciplinary collaboration. Social science is multi-disciplinary and social models often need to involve different disciplines. For instance a sustainability model would involve environmental, social, economic, and other disciplinary considerations.
3. They provide a useful multidisciplinary tool when the mathematics is intractable. Taking the evolution of genes as an example, Axelrod pointed out that agent-based modelling could easily simulate the evolutionary effects of genes where application of mathematical equations is difficult.
4. They can reveal unity across disciplines. For instance, Axelrod found that the ABM about military alignments could successfully predict strategic alignments of computer companies.

From an interdisciplinary perspective, David *et al.* (2004) also point out that ABM based social simulation originates in the intersection of the social and the computer sciences and this interdisciplinary character has encouraged collaborations from scientific fields. They also suggest that the wide interpretative scope of the theory of agent and the advances in computer capability have enlarged the communicative and interpretative room for

ABM to interchange between different scientific fields and model interdisciplinary complex systems.

Conte *et al.* (1998) believe that the potential of the computational study of social phenomena has not been fully exploited and cooperation between ABM and social simulation would strengthen this. They believe such cooperation will benefit both ABM and social science development. More specifically, they believe that ABM is likely to:

1. profit from the more refined and well-established theories, concepts and models of social organizations and institutions developed within the social sciences;
2. adopt the more dynamic approach shared by the social scientists using computer simulation;
3. acknowledge the importance of theory-driven computational studies even in addressing applied objectives and
4. import an approach to computer simulation from the social sciences which sees it as a tool for making and testing theories, rather than applications.

And social scientists in their turn are likely to:

1. give up both the static view of the agent as proposed by some rationality theories, and the behavioural view as proposed by theories of social learning,
2. refine their view of the agent and start to conceive of it as a computable although complex entity,
3. discover the role of the mind as a necessary intermediate between social structures and social behaviours and
4. familiarise themselves with more sophisticated agent architectures.

#### 2.6.3.4 *A hybrid approach to bring together MSM, ABM and GI Science*

As discussed in previous section, attempts to bring MSM and ABM together (Caldwell, 1998; Rephann, 1999) or ABM and GI Science together (Torrens and Benenson, 2005) or MSM and GI Science together (Ballas, 2005b; Holm *et al.*, 1996) have proved to be successful.

Given the characteristics of the agent based technology and the important geographical factors in social policies, researchers including Boman and Holm (2004) have promoted using the combination of different paradigms of MSM, ABM and time geography. Boman and Holm (2004) argue that time geography provides a perspective to help unify the paradigm of ABM as developed within computer science and the paradigm of MSM as developed within the social sciences. Time and space have important impact on human activities in any social system. The authors suggest that time geography provides an alternative perspective on agents since it emphasises the importance of concurrent micro-level representation of agents and their relations to other agents. Time geography can also introduce a conceptual framework for analysing social micro-level interaction in time-space in MSM and ABM.

Boman and Holm (2004) attempted to unite the paradigms through defining them and reasoning about their central concepts. They found that all three methodologies emphasise individual representation and computational solution. However, many MSM only apply a fairly aggregated and disconnected representation of individual behaviour, while ABM can provide the capacity to model individual adaptive behaviours and emergence of such behaviours. On the other hand, MSM are developed with high estimation and validation ambitions, close to observables that facilitate empirical tests.

Boman and Holm's argument for a MSM-time geographic approach is that aggregation prior to analysis and modelling of trajectories over the state space of individuals with several attributes distorts not only individual but

also aggregate results. Individual trajectory interaction and constraints need to be modelled individually to reflect the whole picture. As discussed in earlier, Hägerstrand's space-time geography revolutionised the study of a wide range of urban studies from city planning to social equity, as it allows the space-time measures to be rooted in the local urban environment and incorporated with individual contexts (Pred, 1977). In particular, such features caught the attention of the transport researchers (Kwan, 1998; Miller, 1991). In fact, many individual transport MSMs also exhibit the similar features as Hägerstrand's migration studies where individual migration are driven by rules designed explicitly for them. The individual units (vehicles/pedestrians) demonstrate the capacity of certain degrees of intelligence and parts of the simulation are also rule-driven, which exhibit some of the basic features of agents within an ABM.

For instance, the internationally applied and tested commercial transport MSM, VISSIM, is a multi-modal MSM that includes real interactions between pedestrians and vehicles. The behaviour based model contains a psycho-physical car following model for longitudinal vehicle movement and a rule-based algorithm for lateral movements (lane changing) (Fellendorf and Vortisch, 2010). Another example is a car park selection study using another popular commercial choice, Paramics. Each car in the MSM tries to find a car park according to the minimum journey cost, subject to the availability of the car park spaces. As each car originates from different location and travels by different route, the car driver needs to interact with the environment to update the car's statuses or even to react according to different rules, e.g. a previously available space may not remain due to congestion and the car may need to be re-assigned to another car park (Sykes *et al.*, 2010). The agent-like behaviours demonstrated by the individual units in the above MSMs when they interact with the environment and sometimes even with other individuals/units in the model further confirm the close alliance between MSM and ABM.

With the advances in computing, a comprehensive MSM of the whole urban system can be attempted. After years of team effort, the ILUTE (Integrated

Land Use, Transportation, Environment) model now tries to simulate the evolution of an integrated urban system over time, where activities of individual agents/objects are modelled, including people, transportation networks, the built environment, even the economy and the job market (Salvini and Miller, 2005). A variety of modelling methods have been used within ILUTE to capture agent/object behaviours including the rule-based models and learning models, as well as hybrids of various approaches. In ILUTE, individual behaviours are demonstrated by agents/objects called decision-making units. In the model, agents/objects such as households and families can make decisions directly by themselves or they can collaborate to form ad-hoc decision-making units. The application of distributed intelligence, individual based rules and learning abilities in this MSM demonstrated a strong similarity of an ABM.

From the above review, a strong alliance has been found between the MSM and ABM techniques through their link with the geography, especially Hägerstrand's space-time research. Therefore developments based on a synthesis of the three paradigms can offer a great potential for substantial advance of systems analysis methodology. Boman and Holm (2004) believe it gives a new angle to classical problems where we need to:

1. achieve consistency with the world outside a defined core system boundary;
2. simultaneously represent processes on different spatial and temporal scales;
3. enable agents to concurrently obey internal and external rules, and
4. integrate observable and postulated behaviour while preserving achievability of endogenous emergence.

On the other hand, some research suggest that ABM has the potential to provide a link between Macro and Micro as the multi-level models aspires to (section 2.5). At the micro level, agents interact locally and such interactions eventually give rise to the emergence (aggregated patterns that are not modelled intentionally, e.g. norms) at the macro level. Also agents' ability to obey different rules make them versatile to observe the rules at the



external Macro level as well as their own internal Micro rules, if the two do not conflict (Sawyer, 2003; Gilbert, 2002).

## 2.7 Conclusion

The literature review has suggested that MSM provides a powerful approach in modelling social science and has a particular importance in public policy modelling studies. MSMs have been widely used in a range of application domains and major developments of MSMs have been experienced all over the world in the past decades.

Although MSMs have limitations such as huge data requirement and computing intensity, resource and time cost, recent advances in computing and data production have greatly strengthened such points. More importantly, new technologies such as ABM are naturally complementary for traditional MSMs. One advantage of using ABM is that it allows us to model these systems not only using traditional mathematics and statistics, but also using behavioural information. The flexibility of ABM can also help us to achieve consistency outside a defined core system boundary. The usage of ABM also enables us to generate the emergence of global complexity from relatively simple local actions. Hence the computing requirement can be reduced.

Geographical factors have important impacts on human activities and therefore it is important to model the social system with its local context. Also the geographical information framework provides a bridge to link MSM and ABM together and provides and the hybrid approach may provide an alternative way to study social problems.

As previously discussed, success of attempts of hybrid approaches in modelling and simulating social systems provide the basis for the unification of MSM and ABM. A hybrid approach may offer a great potential for substantial advance in modelling the social systems. In this study, we aim to provide a hybrid model that combines the strength of both

MSM and ABM to provide the capacity to model heterogeneous movements, interactions and behaviours of a large number of individuals within a complex social system at a fine spatial scale. Dynamic changes are captured within the model by simulating the discrete steps when individuals going through six important demographic transitions.

## Chapter 3

# Methodology

## 3.1. Introduction

This chapter discusses various methods used to build the individual based demographic model in this thesis. Methods used in System Design, System Development, Data Selection and Results Alignment of the model to external information are discussed here. In the following sections, we will describe how the system is designed and how the design is followed through during the system development. The modelling approach, population projection method, main components of the system, representation of the population in the system and the choices of the data, programming approach and programming language will also be described in this chapter.

The modelling approach is grounded in a dynamic spatial MSM that simulates discrete demographic processes at a fine spatial scale and that projects the individuals into the future from the year 2001 to 2031. At the core of the model is the Population Model, which consists of 6 modules, which model demographic processes of Ageing, Mortality, Fertility, Health

Change, Marriage and Migration. This model simulates the demographic life course for annual intervals. During the simulation of the 6 demographic processes, the household formation and dissolution processes can be observed as a consequence. The transitions of population statuses, movements and interactions of individuals are the focus in this dynamic model.

A hybrid modelling approach that brings in the agent insight has been adopted to address the need to capture the movements, interactions and behaviours of individuals in the model. This generic demographic model has been developed so that it can provide the basis for further modelling of population behaviours in different application models.

## 3.2. System Design

In this section, designs of three aspects of the model will be discussed separately: the Projection model design, the Population model design and Representation of the population in the system. At the core of the system is the Population model, which utilises the Projection model and other various system components to project the baseline population into the future.

### *3.2.1. Projection Model design*

As Rees (2009) has pointed out, “One of demography’s main contributions to societal planning is to provide projections of the future population”. Despite the simplification and uncertainty of the modelled population, the population projections still play an indispensable role in our society today, as population enters an important variable in all levels of public planning or policy making.

A population projection is a future trajectory of the population and its constituent groups based on assumptions about the drivers of change. Therefore alternative projections are often made on the basis of plausible

assumptions, and a forecast combined with policy parameters can then be chosen from these to assist decision making. There are various methods of projecting population. Traditionally macroscopic methods have been used to project populations. There are three main macroscopic projection methods: mathematical, economic and component methods (Rowland, 2003). Among them, the component method enables us to study the evolution of population through the most important components of changes in population, such as birth, death and migration.

Van Imhoff and Post (1998) provide a comprehensive comparative review of macrosimulation and microsimulation. In their paper *Microsimulation Models (MSMs) and macrosimulation models* are recognised as two alternative methods for making similar statements about the future, as both approaches are based on formal models, need assumptions about the future paths of their input variables and must always contain the time element. As the review pointed out in chapter 2, both approaches have their own advantages. However, MSMs have been considered as a more convincing recreation of a set of conditions or real-life events to analyse the behaviour of a system based on individual entities. MSMs allow the heterogeneity of the state of a system to be fully represented and stored by the most basic units. On the other hand such characteristics tend to be blurred or even invisible within macroscopic models, due to the limited disaggregation in such models. Thus MSMs can reveal and update the non-linear relationships within such disaggregation that are often invisible in macroscopic models (Orcutt 1957).

For the reasons discussed above and in chapter 2, this study uses a dynamic spatial MSM for the population and its change processes, but the model structure is close to that of a macroscopic multi-state cohort-component (MSCC) projection model, as it is the intention of this research to model the population evolution capturing the components of changes. The section below describes the multi-state cohort-component projection method briefly.

Population projections are typically based on assumptions of fundamental demographic transitions such as mortality, fertility and migration. Population projection is thus decomposed into the separate forecasting of these components of population change (Rowland, 2003). Populations of the same area at two points in time are connected by the following “components of change” relationships.

$$P^{t+n} = P^t + N^t + M^t \quad (\text{Equation 3.1})$$

where  $P^{t+n}$  is the population at time  $t+n$  years,  $P^t$  is the population at time  $t$ ,  $N^t$  is the natural increase of the population in the time interval  $t, t+n$ , and  $M^t$  is the net balance of in- and out-migration in the time interval  $t, t+n$ .

In this study the cohort-component model has been used for the population projection. This is an expansion of the component model, adding age in period-cohort form. A cohort is a group of people who have a common initial demographic characteristic, e.g. birth in the same year. A cohort contains the same population over time, e.g. the “baby boomers”. The following section will explain the cohort-component model briefly using information from Rees (2009).

In a cohort-component model, for all period-cohorts from  $x = 0$  to  $x = z$  (the last) by gender,  $g$ , the components of change can be calculated using the following equation:

$$P_{xg}^{t+n} = P_{xg}^t - D_{xg}^t + M_{in,xg}^t - M_{out,xg}^t + I_{xg}^t - E_{xg}^t$$

For the infant period-cohort (from birth to age 0):

$$P_{-1g}^{t+n} = B_{-1g}^t - D_{-1g}^t + M_{in,-1g}^t - M_{out,-1g}^t + I_{-1g}^t - E_{-1g}^t \quad (\text{Equation 3.3})$$

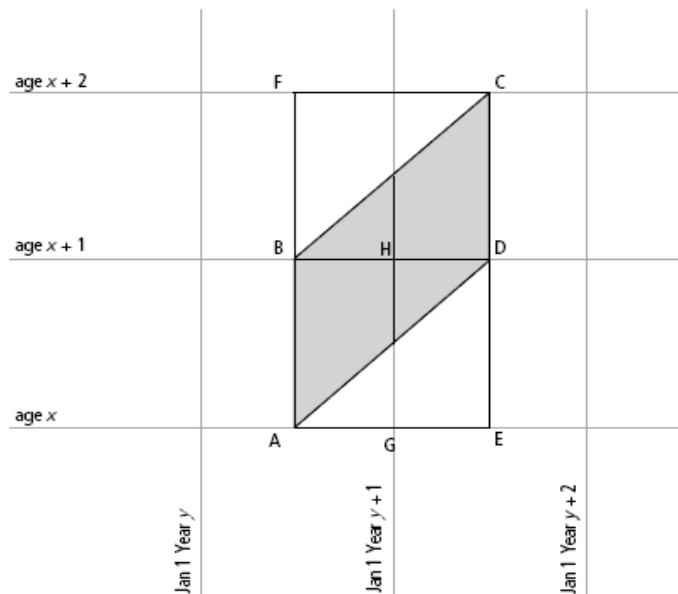
The population is transferred to the next period-cohort at the start of the next time interval:

$$P_{x+ng}^{t+n}(\text{next}) = P_{xg}^{t+n}(\text{current}) \quad (\text{Equation 3.4})$$

Adding the end populations for the last, open-ended age group, the equation becomes:

$$P_{xg}^{t+n}(\text{next}) = P_{x-ng}^{t+n}(\text{current}) + P_{xg}^{t+n}(\text{current}) \quad (\text{Equation 3.5})$$

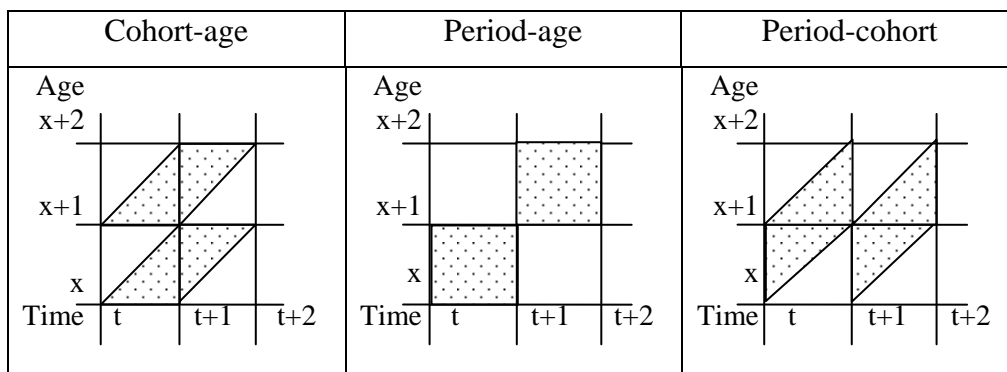
The cohort-component model also requires that equal intervals of age and time in length. It is preferable to use single years of age and annual time intervals, as was the case in the study (Rees, 2009). The cohort-component method is widely applied in population projection. The cohort-component projection is based on the past population structure by age and sex. These cohorts of population evolve through the fundamental components of changes such as births, deaths, and migration. To predict the future population, the future trajectories of the component rates are forecast (Rees, 2009). The projection process can be illustrated by a Lexis diagram (Figure 3.1).



**Figure 3.1 Lexis diagram showing the age-time space used in the projection model**  
Source: ONS (2008)

A Lexis diagram is an age-time diagram, in which age is represented in the diagram on the vertical axis and time on the horizontal axis, while the life of an individual or a birth cohort is represented by a diagonal line running from bottom left to top right. As illustrated in Figure 3.1, the line AB represents the population aged  $x$  at mid-year  $y$ . The size of this cohort one year ahead, i.e. aged  $x + 1$  at mid-year  $y + 1$ , is represented by the line DC. To calculate this population one year ahead, you need to measure or model the components of change events that occur in the shaded parallelogram ABCD.

There are three main approaches to calculate the transitional probabilities: the cohort-age, period-age and period-cohort approaches. Each models cases within different age-time spaces, as illustrated by the dots in Figure 3.2.



**Figure 3.2 Illustration of differences of cohort-age, period-age and period-cohort approaches in age-time spaces**

The cohort-age method is used to measure changes between exact ages (birthdays) and is employed in life tables. Period-cohort rather than cohort-age method is used to model the changes from one period to another for age groups and also to be consistent with the ONS approach illustrated in Figure 3.1. Age-period rates, derived from various published data, are converted into period-cohort probabilities in different ways. Details about how the period-cohort method is used in probability calculations are described in specific processes such as mortality and fertility in Chapter 4.

Multi-state projections were developed by Rogers (1975) to take into account of additional characteristics, or "states," of a population other than age and sex. Multistate projections using the cohort-component framework



were originally developed in order to account for the place of residence of population members and have since been extended to other dimensions (Lutz, *et al.*, 2004; Rees and Wilson, 1977; Rees, 1984). In Rogers' original study, populations of particular "states" (e.g. regions) are the results of the flows from and to all the other regions in the studied system. The populations of regions are thus dependent on the populations of all other regions: the out-migrations from a region are the in-migrations to other regions (Rogers, 1975; 1995). Multi-state methods are employed by many national statistical offices in their regional and local population projections.

The multistate cohort-component models disaggregate the population further by the variable of space. Therefore, changes of the populations in sub-regions include their own birth, death components and the migrants from and to other sub-regions. In macro models events are calculated for population groups. In micro models events which change the status of the individual either happen or do not happen to individuals.

Although its structure parallels that of a multistate cohort-component model, our dynamic spatial MSM adopts a bottom-up approach and is an individual based model. The baseline population is simulated at the individual level to provide disaggregated information to facilitate further population related studies or decision making for modern social problems. Individuals with a rich set of attributes are projected at the ward level to enable us to study the characteristics of the population within a local context. The model results can be easily aggregated to higher spatial scales for explorations of population patterns at various levels. The model projects each component of population change separately, but each component of change affects the others. For instance, a large number of young female migrants moving into one area will have a significant impact on the local births.

The modelling period has been set to years 2001-2031 in order to facilitate explorations of population evolution patterns, as well as the assessment of mid-term policy effects. With further assumptions about future population trends, this model can be extended to long term projections. The projections are made for successive years running from one year to the next using the

cohort component method. For each age, the starting population plus in-migrants, international immigrants and those born during the year, minus internal out-migrants, international emigrants and the number of deaths, produces the number of survivors in the population at the end of the year. As a dynamic MSM, this model uses the dynamic ageing techniques. The simulation is driven by probabilities used in different demographic transitions. Although all probabilities will reflect the impacts of age, sex and location, each probability varies according to the relevant factors such as marital status, previous health status and household types etc. due to the specific requirement of different demographic processes. The probabilities are generated on the basis of the demographic characteristics that have been revealed in the census data in year 2001. Such probabilities are then updated each year in light of the recently available information of demographic trends from 2001 to 2031 to reflect the population trends. Details of the probability generation for each process will be described in Chapter 4.

### *3.2.2. Population Model design*

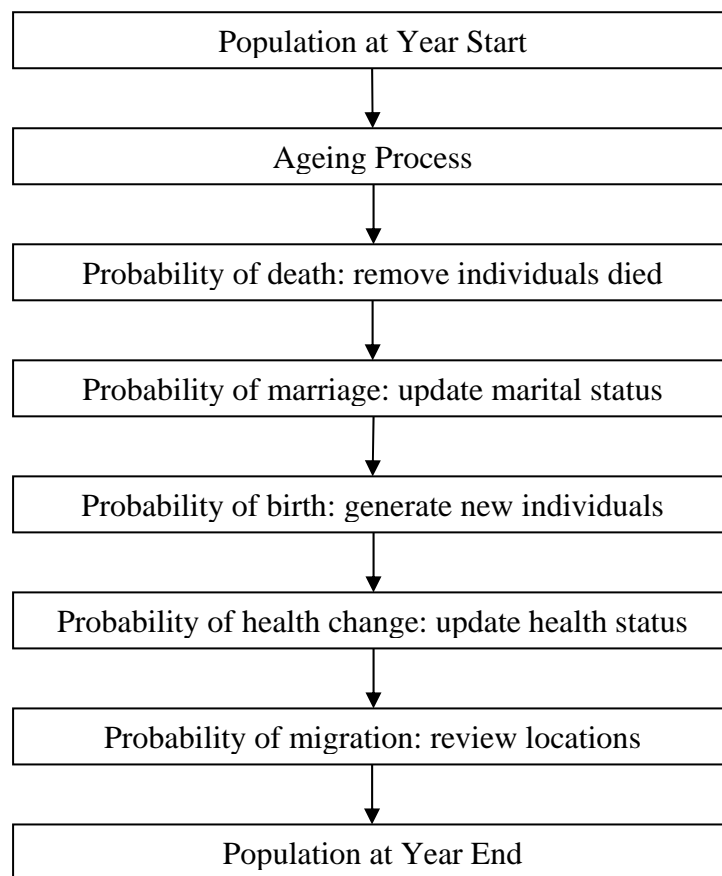
As described in section 3.2.2, the population model is based on a dynamic spatial MSM, but its structure parallels the macro MSCC projection model. The component structure allows the model to be designed as a modularised system. Modularised systems are easy to extend and maintain: system components can be reused, altered, removed or extended in the future without affecting other parts of the system. Therefore the model is designed using this principle.

At the centre of the MSM is the Population Model. The population model includes a set of population modules to deal with specific simulations of population changes in various demographic processes. This is commonly used in many major MSM developments such as Corsim and SVERIGE, which are described in Chapter 2. The Population Model design includes 6 demographic processes, namely: Ageing, Mortality, Fertility, Health Change, Marriage/Household Formation, and Migration.

The modularised design of the model provides great flexibility not only for development and maintenance, but also for implementing different projections or simulations. For example, if you are only interested in the natural change of the population, all you need to do is to switch off modules other than Ageing, Fertility and Mortality when you run the simulation model. You can then have a full run to assess the migration impact. More details of the architecture will be discussed in Section 3.3.3. In the following sections, the modelling of the demographic processes will be described in more details.

### 3.2.2.1 Demographic processes

As described before, six important demographic processes have been modelled in separate modules in the population model. The organisation of the processes/modules is illustrated in Figure 3.3.



**Figure 3.3 Process of the population simulation**

Mortality and Fertility are modelled because their obvious importance: “Matters of life and death have been a central interest of demography from its beginning” (Rowland, 2003). In a cohort-component projection, all the components of change need to be used for age groups in a period-cohort framework. Migration in the form of several streams of out-migration and in-migration (discussed in detail in Chapters 4 and 5) has a profound impact on the population changes. Together with Fertility and Mortality, they form the three basic population flows that change the population, also known as the components of change. In this study, ageing, household formation/marriage and health information are felt useful for public policy making and demographic planning as they have important impacts on decision making. Therefore these six demographic processes have been modelled in this study.

Transition probabilities for each of these events are applied discretely at one year intervals. Many different orderings are possible for this sequence of events. In this model, it is considered more logical to evaluate fertility following the formation of marriages and partnerships. Mortality is considered early in the process for practical reasons in terms of computing power, since if an individual life course is terminated subsequent processes can then be ignored. At times, different processes can be interdependent. For example, Marriage and Migration processes are often connected, as a change in marital status will frequently occur alongside the move to a new home. Migration process is modelled after the Marriage process in this model to include migration on marriage and other reasons. Such interdependency between Marriage and Migration and between other demographic processes means that each component of change can have an impact on the others, although the model projects each component of population change separately.

The six processes are modelled because of their indispensable roles in the transition of population status, people’s movements and interactions. The details of the designs and implementations of the respective six modules will be described in Chapter 4.

### 3.2.2.2. *Representation of the population in the system*

The baseline population of the model is recreated from 1% sample of the UK population contained in the Household Sample of Anonymous Records from UK Census 2001. The process of the population recreation will not be discussed in this thesis, as it is not the work of this PhD study. For details of this model, see Birkin *et al.* (2006). There are more than 723,000 individual records in total for the baseline population of Leeds at census date in year 2001. Each individual record contains a set of rich attributes, including individual, household and small area variables. Such attributes are updated, going through the six important demographic transitions.

The studied population is modelled as individuals in households. Each individual has been allocated to a specific ward of Leeds, as the model focuses on the dynamic simulation of the Leeds population at the level of ward. Although the studied population is modelled as individuals, there is interdependency between the household, individual and environment/area. The attributes of individuals, households and their environment can change due to various interactions between the individuals and other individuals, households, or the area that they live in.

For instance, during the process of Marriage, the formation of a new household will mean changes in at least one individual's location; for households, this will result in changes in both old and new households or all three households if both brides and groom moving out of existing households; for the areas that they used to /are going to live in, it will see changes in both local housing and local population. Similar changes will be experienced in any migration process. Due to this interdependency, the operation of these demographic processes of individuals also leads to the formations and dissolutions of households during course of simulation.

Compared to other demographic processes, migration is a complex process. The original design of model was to keep it simple and built on the basis of a mostly closed population (other than the student migration). The original Migration process only focused on the internal migration within Leeds.

However, after initial analysis of the results, limitations are soon found. Such issues have to be addressed by introducing internal migration to and from the rest of UK and international migration to and from the rest of world. Five migration flows are then modelled in the migration process to provide a full picture of the migration impact on Leeds population.

A new migration framework has been developed to describe the modelling of different migration flow types and relationships. The new migration framework is illustrated in Figure 3.4, but more details of the migration module will be discussed in Chapters 4, 5 and 6.

<b>From</b>		<b>to</b>					
<b>Leeds</b>		<b>Aireborough</b>	<b>...</b>	<b>Wortley</b>	<b>Rest of England</b>	<b>Rest of UK</b>	<b>Rest of the world</b>
	<b>Aireborough</b>						
	<b>Armley</b>						
	<b>...</b>						
	<b>Wortley</b>						
<b>England minus Leeds</b>	<b>Rest of England</b>						
<b>UK minus England</b>	<b>Rest of UK</b>						
<b>World minus UK</b>	<b>Rest of World</b>						
<b>Flow type</b>	<b>Flow name</b>						
	<b>intra Leeds flows</b>						
	<b>in-migration from rest of UK</b>						
	<b>immigration from the rest of the world</b>						
	<b>out-migration to rest of UK</b>						
	<b>emigration to rest of world</b>						

**Figure 3.4 Framework for modelling migration process**

To facilitate the modelling of movement, interaction and behaviour, agent technology has also been employed. Individual micro-units in the MSM are also agents at the same time. All agent functions have been developed in a separate package/module. According to the specific demands of the simulation, individuals/agents can choose to switch on/off the agent simulations that are driven by specific rules designed for various agents.

With the agent features switched off, all agents just follow the general MSM rules, just like the agents in a CA (Cellular Automata), where all individuals follow the same set of rules/probability determination processes in specific demographic processes. As such rules are applied to the whole population in whole system, not individually encapsulated, we will call these as the external rules for the agents in this thesis.

However, if the Agent features are switched on in certain processes such as student population in the Migration process, they then stop following the MSM rules and start to act and react according to individual rules that are designed according to their specific individual, household and/or area characteristics. For instance, a student agent will act towards achieving the goal of moving to the area where their fellow students live by finding out where they are. Then the agent can react to the environment such as limited accommodation vacancies to decide whether to move in or not. To differentiate from the MSM rules, we call these the internal rules, which are specifically designed to drive the interactions between the agents and interactions between the agents and the environment.

### 3.3. System Development

This section focuses on the methods used in the development of the model. The programming approach and language, system architecture and development, as well as the hybrid modelling approach used in this project will be introduced.

#### *3.3.1 Programming approach and language*

The Object-Oriented Programming (OOP) approach provides a more natural representation of the real world than the traditional approaches. The traditional programming approach views a system as a central control manipulating the data flows and therefore many of such systems also adopt a top-down approach. The OOP approach views a system as a collection of objects with their own attributes and methods encapsulated within

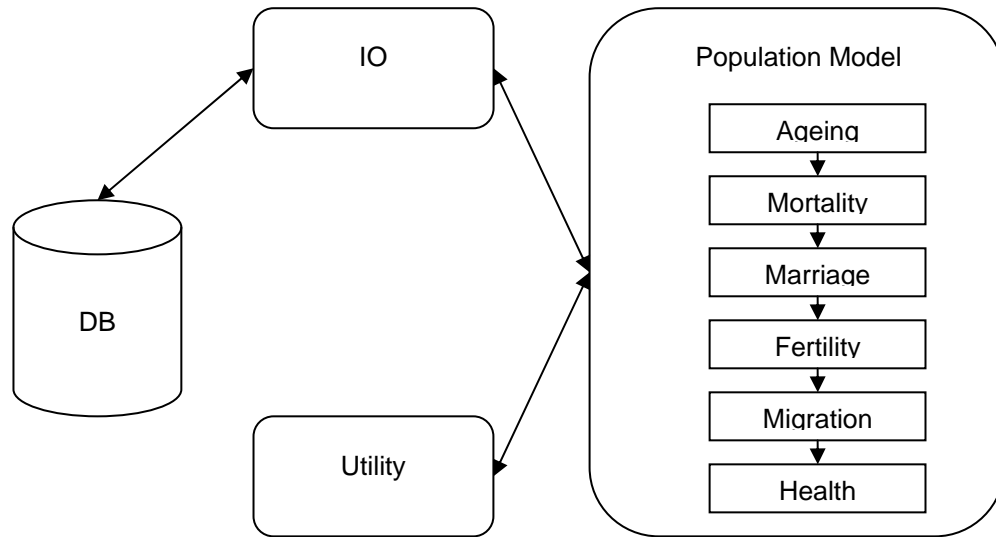
individual “shells”. They respond to the “world” according to their attributes and methods, thus the OOP naturally adopts a bottom-up approach. The approach of building the system from individual base units is a much more organic strategy for information processing. It is for this reason that many new technologies such as agent technology are naturally compatible with the OO approach. In object-oriented programs, data are represented by objects with two sections: attributes that describe what an object is and methods that decide what an object does. These attributes and methods are closely connected to the object's real world characteristics and behaviours. When a program runs, messages are passed between objects and an object responds to the received messages according to its methods and attributes. This feature makes the OO programs written in the languages such as Java much easier to understand and Java programmers don't have to unlearn bad programming habits when moving into real world projects.

Java is chosen as the development language for this project. Java is a high level programming language that can do almost anything a traditional programming language like Fortran or C can do. However Java has learned from the mistakes of its predecessors. It is considerably easier to program and to learn than those languages without giving up any of their power. Another important feature of Java for this project is its independence from platforms. Java has great portability as Java programmes are compiled to an intermediate form code called byte-code that is precisely defined and remain the same on all platforms. Therefore it boasts the feature of “Develop once, run anywhere”. Lastly Java is safer than many programming languages. It is very difficult to write a Java program that will crash your system, unlike any other language. Java is designed to allow for secure execution of code across a network, even when the source of that code is distrusted and possibly malicious. The worst that a Java program can do to a host system is to bring down the runtime environment rather than the entire system. Most importantly Java applets can be executed in an environment that prohibits them from introducing viruses, deleting or modifying files, or otherwise destroying data and crashing the host computer (Harold, 1996).



### 3.3.2 Architecture

The simplified system architecture of the model can be illustrated in the form of a diagram in Figure 3.5.

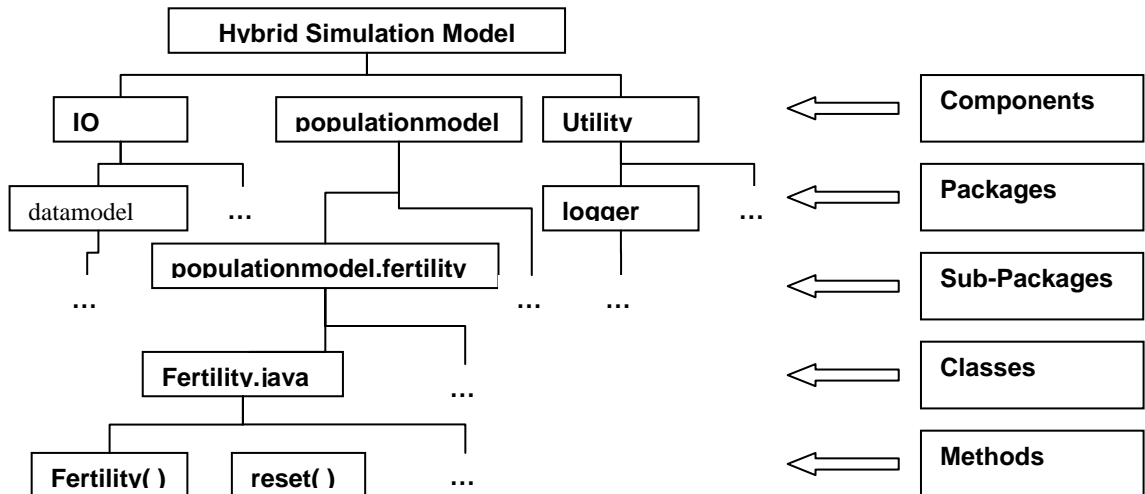


**Figure 3.5 System architecture**

At the core of the system is the Main Simulation component, labelled Population Model, which consists of 6 modules, namely Ageing, Mortality, Marriage, Fertility, Migration and Health. Due to the nature of the demographic complexity of the modules, some modules require sub-modules to deal with specific aspects of the demographic processes. For instance, Migration has different sub-modules to deal with sub-population such as students and other types of migration flows. The simulator is assisted by two components: the Input and Output (IO) component deals with the data flow between the database (DB) and the simulation model. That is, it loads the specific data from the database upon requirements from the simulator, updates and outputs various simulated results to the database. The database currently is in the form of a simple file system. The Utility component provides various tools for the simulation. For example, various probability generators and determinators, as well as a series of helper programs dealing with different tasks to ensure the program run smoothly such as preparing the data for input/output, as well as logs of events.

The modularised design of the model provides great flexibility to system runs, developments and maintenances. In the run-time environment, it allows the user to choose to switch on or off the different module(s) to suit various purposes of their studies. For instance, the user can run the simulation with only Ageing, Mortality and Fertility modules switched on to assess the natural changes only of the studied population. More importantly, such design also minimises the impact if further development of one of the modules or extension of the system in the future by simply adding on or taking off certain modules without the need to worry too much about the rest of the system. The modularised system makes the system development and maintenance much easier. In the following section, we will describe the projection model design and population model design.

As described previously in Section 3.2.3.2, we want to combine the strength of microsimulation modelling and agent based modelling to enable an individual based model with great details. Such integration of the two modelling approaches demands a flexible design of the system architecture. Individuals in the model are constructed as agents within the households, but they all obey the commands from the central control in MSM. According to the different requirements of the simulations, agents can act as “dumb” agents whose “intelligent” properties are disabled and they are just micro-units in the MSM, or they can be more “intelligent” and demonstrate their distinctive behaviours through individual built in rules according to their types. Such rules can override the MSM commands if necessary, e.g. student migrants do not migrate according to the location based migration probabilities (Section 3.2.3.2). As objects/agents are encapsulated in their own “shells” (Section 3.3.2), such changes/switches will not interfere with other objects and as the processes are modularised, changes to one process will not affect the rest processes, either. The organisation of the code reflects the system design. A diagram illustrating the hierarchy of the coding is provided here (Figure 3.6).



**Figure 3.6 An illustration of program hierarchy**

For the purposes of illustration, only part of the population model package has been represented. The four main components consist of different packages such as populationmodel, which according to the function requirement can contain further sub-packages such as populationmodel.fertility. Packages are used to organise classes for portability and efficient access. Within the packages, there are classes. In object-oriented programming, a class is a programming language construct that is used as a blueprint to create objects. A class may represent a person, place, or thing and it encapsulates the state and behaviour of the concept it represents. A class such as Fertility.java encapsulates state through data placeholders called attributes such as the variables and constants in a class; it encapsulates behavior through reusable code called methods such as Fertility( ) and reset( ). Objects of the same class inherit the same set of states and behaviours.

From the demographic view, the processes are modelled using various packages such as fertility within the population model. Individuals going through the process of fertility will use the same sets of classes such as Fertility.java, where variables such as age and marital status and methods such as reset() in the class can be used to update their statuses, e.g. to reset the probability for this individual to the one corresponding to his/her current individual, household and/or area characteristics.

### *3.3.3 Modelling approach*

The advantages of the Individual Based Models (IBMs) to model system changes from the level of individuals have been described in Chapter 2. Two different approaches, ABM and MSM, are used in this model. Based on the discussion of each of their strength and limitations previously in chapter 2, the MSM and ABM approaches are found to be complementary to each other in both theoretical and practical aspects. Details of how these methods are implemented in the model are described in the sections below.

#### *3.3.3.1 MSM*

As a static MSM only represents the population at a given time (Pudney 1994), where no characteristics of the micro units are changed during the process, a dynamic MSM is used in this study. This is mainly because a dynamic MSM cannot only update the characteristics of the micro units caused by the stimulation of endogenous factors, but can also project them over time to include demographic processes and social economic transitions, such as ageing, mortality, fertility or social and geographical mobility (O'Donoghue, 2001). On the other hand, small area differences play an indispensable role in modelling the population changes in this study and they also pose a substantial impact on various planning applications. Therefore a spatial MSM is used to model the individual changes with the local context. Although a dynamic spatial MSM is used for modelling the population and its dynamics to capture details at the individual level, the model structure parallels the macro multi-state cohort-component projection model. Populations are therefore simulated at the individuals (MSM) within small areas of ward (spatial) through various demographic processes, where characteristics of each of them are updated on the basis of transitional probabilities each step of the simulation (dynamic).

Being a widely applied instrument in studying and predicting the evolution of population, MSM is as important to the analysis of event histories as macrosimulation is to traditional aggregated demographic analysis. However, there is a gap between the micro-demographic theory and

demographic microdata and modelling techniques (Billari *et al.*, 2002). Quite often, there is an unavailability of appropriate microdata for the modelling a specific important demographic transition in an MSM, due to its high standard of requirements of the data at the individual level. Also in a traditional demography model, there is a limitation of precision in theoretical constructions and often lacks theory at the basis of the applications of statistical models and data collection at the level of individuals (Billari *et al.*, 2002).

#### 3.3.3.2 ABM

Agent-based modelling (ABM) is an alternative approach that can model individual behaviours through multiple agents. In an ABM, each agent follows their built-in rules and acts/reacts according to such rules and the knowledge gained through interactions with each other and the environment that they live in. Through such interactions, simple and predictable local interactions can generate familiar but unpredictable global patterns, e.g. the formation of norm in an artificial society. This is called the emergent property of the ABM (Russell and Norvig, 1995). With such features, ABM provides theoretical leverage to model a complex social system where the global patterns of interest are more than the aggregation of individual attributes.

Although there are limited examples of using ABM approach in demographic models, Billari *et al.* (2002) consider ABM as a promising approach to help improve our understanding of demographic behaviours by studying demographic processes as the outcome of interacting agents. By focusing on dynamics of the population instead of equilibria, ABM is better suited for modelling specific processes. Migration is a complex demographic process where interactions and behaviours play an important role (Champion *et al.*, 2002). Using ABM, individual activities and diversity of migration decisions leading to complex migration patterns can be simulated in detail. Espindola (2006) analysed the rural–urban migration using ABM, where the migration of workers is modelled as a process of social learning by imitation. As emergent properties of the model,

transitional dynamics are observed with continuous growth of the urban fraction of overall population towards equilibrium. While Loibl and Toetzer (2003) studied urban sprawl patterns through modelling suburban migration and residential area occupation. Distinctive migration behaviours of households with varying socio-economic status have been simulated in an ABM. Makowsky *et al.* (2006) build an ABM to simulate crisis-driven migration of agents within a multi-ethnic population. This study reveals that cultural networks temper an agent's security calculus, with strong social ties dampening the human security dilemma.

However, compared to a MSM, an ABM is often built without the validation ambition and the individual rule-driven simulation can slow down the simulation when applied to a large population of individuals.

#### *3.3.3.3 A hybrid approach*

Based on the above discussion of the strength and limitation of ABMs and MSMs, a hybrid modelling approach is proposed in this section to bring the strength of both approaches together and address the limitations discussed.

As described in Chapter 2, the origins of the hybrid approach can be traced from the time when the name of ABM had not been invented. Although mainly recognised for his pioneering work in spatial microsimulation of innovation diffusion and the conceptual framework for the analysis of spatial dynamics in a micro level time-space frame (Hägerstrand, 1953; Holm *et al.* 2006), Hägerstrand has also brought the ABM concepts into the spatial MSM in order to explore the relationship between social contact and migration (Hägerstrand, 1957). In Hägerstrand's spatial MSM, population and vacancies are evenly distributed in "migration fields" that are divided into cells. Then "active migrants" are distinguished from the "passive migrants". The active migrants can randomly select and move to a destination in an adjacent cell, whereas the passive migrants are stimulated by the active migrants in a way that a passive migrant chooses a new cell where an earlier migrant from the same origin is staying and the attractiveness of all earlier migrants are equal. Although not a computer

based model, the migrants in this model demonstrate the basic “agents” characteristics of being able to interact with others and the environment. Since then, more discussions of using ABM in various types of models have been published and more modern attempts of the hybrid approaches have been made which confirm the potential of integrating the two approaches in modelling complex social systems (Conte *et al.*, 1998; Axelrod, 2005; Boman and Holm, 2004).

In the current model discussed in the thesis, the MSM provide the theoretical basis with its roots of using real data and to be used for real application, providing valuable guidance for the projections. The statistical nature ensures the similarity between what it predicts and what is actually observed in the gathered data. Practically, its list processing power allows the model to simulate the detailed changes in large number of individuals with rich attributes. Most MSMs are also built with the validation ambitions and the alignment exercises have been well studied and widely practised in recent MSM research.

With great flexibility of the rule-driven simulations, the ABM provides a way to bridge the knowledge gap and data limitations when we study individual movements, interactions and behaviours. This model uses an ABM to model student migration where suitable microdata are not available and distinctive migration patterns have been found from the rest of the population. As agents can also carry their personal history and personal history sometimes can have an important impact on demographic changes, we also use ABM to explore the impact of personal migration history on mortality projections.

## 3.4 Data selection

### 3.4.1 *Baseline data selection*

#### 3.4.1.1 *Baseline population data*

As described previously in Section 2.3.4, the baseline population is recreated from the Census data. The household populations are generated

from HSAR records and communal establishment populations are generated using ISAR records using a separate MSM. The process of recreating the synthetic population is NOT part of this PhD research and the baseline population data are provided by Prof. Birkin. More details of the creation of the baseline population data can be found through his paper and other works referenced in Section 2.3.4.

#### *3.4.1.2 Spatial data*

The ward boundary data of Leeds have been downloaded from UKBORDERS (Edina, 2011), using the selection of Leeds wards from Census boundary data 2001. Individuals have been allocated the spatial coordinates according to their ward location, using the centroid of the specific ward. The level of ward is selected, as it is considered to be an appropriate level to provide sufficient small area details for planning support. At the ward level, it also allows many data available from various resources to be easily linked to the baseline population data. For instance, the small area vital statistics for births and deaths of Leeds are available at the ward level. In future development work, this can be changed to MSOA level, which becomes increasingly popular.

#### *3.4.2 Attribute selection*

A number of important attributes have been considered in terms of their influence on demographic changes of individuals and populations. These attributes have been discussed one by one here.

- Age
- Sex
- Socio-economic status
- Family/Household structure
- Marital status
- Ethnicity
- Lifestyle (smoking, alcohol, drug use, diet, exercise)
- Environmental factors



**Age:** Analysis of the age composition of populations is essential in demographic investigations. The intensity of each of the demographic processes varies significantly by age but in different ways; mortality mainly affects older people; fertility involves women in the age range 13 to 50; migration is most intense at ages from 18 to the mid 30s; general health steadily deteriorates with age (Rees, 2009). At the aggregate level, demographers are interested in the age compositions of populations and the ages when people engage in certain behaviours. Previous studies also find that the ages of individuals when they move house, get married, have babies, use certain products or services are vital to explaining social trends, targeting markets and planning for the future. Mortality, fertility, and migration processes vary profoundly and in different ways by age. Indeed, the age composition of the population is central to understanding the nature and functioning of our society (Rowland 2003).

**Sex:** Due to the biological factors, behaviour and wellbeing differences (women are generally better at looking after themselves if alone), males and females demonstrate considerable variation in their behaviours even when they are of the same age. Also the general fact that more boys are born than girls in the first place (although suffer greater mortality than girls later in different life stages) means that there is a need to model the sex dimension of the studied population from the beginning. Age and Sex are the two fundamental attributes and they play indispensable roles in all the six important demographic transitions. In fact, Age and sex composition are so important to the nature and functioning of societies that all traditional models are based on it.

**Socio-economic status:** It is obvious that socio-economic statuses of individuals have an important impact on the living standard. Its impact on individual life style, education and occupation then in turn affect demographic changes in various aspects such as health, marriage and migration. There is also a strong relationship between socio-economic status and mortality risk: people in deprivation tend to have higher risk of mortality (Marmot, 2005; Seeman *et al.*, 2004). However, Socio-Economic

Status (SES) is a general attributes which combines various attributes such as education, occupation and income in particular, which are measured in censuses or general surveys, together with standing or regard that are measured in specialist surveys.

The first two elements of SES, education and occupation, are measured in the 2001 Census. The difficulty is that they do not cover the whole population in the same way. Children may have completed only part of their education. Housewives may not have a paid occupation. So a method would need to be developed to compute an SES to each household member in the Census and in the survey. Also this attribute is more useful for working age population, therefore only covers about 75% of the population. Also different microdata often use different categories for this factor. Due to such complications, socio-economic status is not used as a general factor that applies to simulations of all demographic processes.

**Family/Household structure:** Family/household structure plays an important role in decisions on marriage, migration and fertility. For example, we are all familiar with the migrations driven by housing stress due to growth (new arrivals or kids growing up) within current households. Family structure also becomes increasingly important in people's later life. For example, there have been evidences that men living with their wives have much lower morbidity/mortality risks than single men living alone.

**Marital status:** The current marital status affects the decisions about marriage and fertility. More births are found within marriage than outside marriage. Increasingly it has become a more important risk factor in mortality, especially in people's later stage of life. Elderly people within marriage are at less risk of mortality than those whose who are not married, due to the support between them and their spouses. The cohabitating population is not separated from the singles in this model. Averaged probabilities of the cohabitated and singles are used in the demographic transitions such as fertility for women outside marriage. As decisions are based on whether an individual is married or not, the impact of not differentiating the cohabitated population is considered to be acceptable.

**Ethnicity:** Potentially ethnicity can reflect many aspects of demographic characteristics. Ethnicity has a known impact on life style, migration, fertility and household structures. For instance, studies on ethnicity and fertility reveal that South Asians tend to have bigger families than other populations in UK and poorer groups such as Bangladeshis and Pakistanis tend to have poorer health records that may relate to their dietary/lifestyle (Dubuc and Haskey, 2010; Dubuc, 2012).

**Lifestyle attributes:** Lifestyle attributes such as smoking, drinking or diet is important in terms of health, fertility and mortality. Because of their importance for being underlying many other demographic changes, these are currently being estimated for small areas by various research teams. However these characteristics would need to be imputed for the microsimulated populations.

**Environmental factors:** This model intends to introduce the environmental impact from a geographical view. Lots of demographic and socio-economic processes (probabilities, rates and transitions etc.) vary by geographical location. The model also attempts to explain such variations by including in the description of the individuals in the MSM e.g. age, sex, socio-economic position, occupation and ethnicity and it tries to include environmental influences eg: land pollution and contaminated water. However, often some of these individual or household variables cannot be measured very well and there are lots of factors that remain unobserved. For example, the high mortality risk of people in Glasgow can be explained by their socio-economic deprivation and the high smoking and drinking rates, but still not all of the mortality can be accounted for. The mortality in Glasgow is represented by the influence of the other factors which might include cultures (gangs, violence and drugs etc.) or genetics (natural selection which results in high cardiovascular mortality etc.) However, it is known facts that such demographic patterns persist in certain geographical locations (Shaw *et al.* 2006). Therefore geographical location is used in this study as a proxy for all such unobserved factors, as well as to provide a local context. Geographic codes are used as the surrogates for unobserved variables

(features for wherever good microdata are not available).

After careful consideration, age, sex, geography factors and marital status have been used as the base attributes in all demographic process simulations. As described above, age and sex are the fundamental underlying factors of any demographic changes. Geography provides useful substitute for many factors that may contribute to the demographic changes, but difficult to capture in the model. Marital status is selected because of its important role in modelling Marriage and Fertility. It is necessary to simulate the Marriage process on the basis of the current marital status and the birth patterns within and outside marriages have been found to be quite different. Therefore the four attributes are selected as the base attributes for modelling all demographic processes. However, if necessary, different attributes/combinations can be used in the specific demographic processes, according to their roles on the processes. For instance, in the Fertility process, as this model uses a female-led approach to model the fertility process, sex is not relevant and the probabilities are only applied to female populations.

Due to the limitation of computing resources, appropriate microdata and the time scale for the PhD program, other attributes are excluded.

### *3.4.3 Data used in simulation*

Once the baseline population data are in place, the specific probabilities will then need to be calculated and applied to each individual during the simulations of different demographic processes. A range of data has been used to calculate various probabilities of specific demographic transitions when modelling the demographic processes.

#### *3.4.3.1 List of main data*

The main datasets used in the model are listed here, with the description of the different spatial scales. Among them, various life tables have also been used in mortality probability calculation. Special Migration Statistics and International Passenger Survey have been used for migration probability

calculation. ONS Mid-Year Estimation and Sub-national projections, various tables in the ONS journal Population Trends have been used to update the trends in various processes. There are also various other census tables, commissioned tables and bespoke data have been used in simulating various demographic processes. The list is as follows:

- Census data (HSAR and ISAR) and BHPS data (national level)
- Vital Statistics (ward level)
- Special Migration Statistics (ward level)
- International Passenger Survey (national level)
- ONS Mid-Year Estimation and Sub-national projections (sub-national level)
- Tables in Population Trend (various)
- Other census table and life tables (various)
- Commissioned census tables and bespoke data (various)

Probabilities for specific transitions are calculated on the basis of various rates, which are derived from event counts (e.g. birth, death and migration) and counts of population at risk (e.g. women aged 15-45 in the area for fertility rates) from relevant datasets such as MYE (Mid-Year Estimates) and SNPP(Sub-National Population Projection) data. For instance, MYE mortality rates are derived from event counts from Vital Statistics and mid-year population estimates (section 3.4.3.2), i.e.:

$$MYEMortalityRate = \frac{DeathCount}{MYEPopulationEstimates} \quad (\text{Equation 3.6})$$

SNPP based rates are calculated similarly. Details of the calculation of rates and probabilities can be found in descriptions of probability generation in specific processes in Chapter 4. In the following sections, further details of the datasets used in specific demographic processes are described.

#### *3.4.3.2 Data used in Mortality process*

The main data used in the Mortality probability calculation have been listed

below:

- the reconstructed baseline population using HSAR(ONS, 2006a), ISAR(ONS, 2006b) and BHPS data (Taylor *et al.*, 2010);
- the 2001 Census Standard Tables provides population cohorts by single year of age and sex for Leeds wards (ONS, 2001a);
- the death counts for Leeds: from Vital Statistics 2001(ONS, 2001b). ONS Vital Statistics at LA level are used to generate ward level data in the way described in Rees *et al.* (2004);
- the Leeds ward level life table where Leeds population is in 5 year age bands (Rees *et al.*, 2004) and
- the Life tables for UK and local authorities, 1998 to 2002 computed from ONS life table where population is in single year of age (Rees *et al.*, 2004).

#### *3.4.3.3 Data used in Fertility process*

Naturally, for fertility process, all the datasets selected here only concern the female population. This model chose to study women aged 15 to 45+, as they are the group at most risk. The data used in the Fertility probability calculation involves the following:

- the reconstructed baseline population using HSAR (ONS, 2006a), ISAR (ONS, 2006b) and BHPS data (Taylor *et al.*, 2010)
- the 2001 Census Standard Tables provides the population by single year of age and sex for Leeds wards (ONS, 2001a)
- the fertility data for England and Wales for women both within and outside marriages (ONS, 2001c)
- ward birth counts for Leeds from the Vital statistics data for areas in England and Wales by ONS. The data we use is Table VS4: Vital statistics for wards 2001 - 2001 boundaries (ONS, 2001b).
- twin births: maternities with multiple births data (ONS, 2001d).
- sex of babies: based on Table VS4: Vital Statistics for Wards 2001 - 2001 Boundaries (ONS, 2001b).

#### *3.4.3.4 Data used in Health Change process*

- Census 2001 Standard Table: ST 016: Age and sex by general health and limiting long-term illness) (ONS, 2001e) and
- ISAR data on communal establishment populations for those aged 65+ (ONS, 2006b). Based on values of *cetype* that include four types of formal care: NHS psychiatric hospitals, NHS other hospitals, nursing homes, residential care homes and other medical and care establishment.)

#### *3.4.3.5 Data used in Marriage process*

- Population Trends 123 (ONS, 2006c), Table 5: Age at marriage by sex and previous marital status, 1991, 2001-2004, England and Wales and
- Commissioned table (ONS, 2001f) Table C0729 - Age of Married Couples England: contains age of husband and wife's age from 16 to 100+.

#### *3.4.3.6 Data used in Migration process*

- 2001 Census data: ISAR (ONS, 2006b) and HSAR (ONS, 2006a)
- 2001 Census Special Migration Statistics (CIDER, 2001)
- International Passenger Survey (national level)

### *3.4.4 Data used in updating*

The rate of change for each of the components depends on both observed historical trends in the area and on forecast national trends. The model will provide simulation results of the Leeds population for 30 years (2001–2031). The initial probability calculations are based on the most up-to-date census data in 2001. Then the probabilities are updated at an annual interval in light of the ONS assumption and data generated from ONS SNPP (Subnational Population Projections) and MYE (Mid - Year Estimates) at the local authority level during 2001-2011. Details of these will be discussed in the following sections. The longer term trends of 5 year intervals have been used from year 2011 due to the limitations of the data.

The changes of the attributes of individuals are determined by probabilities generated in different demographic processes. In the original model the Mortality, Fertility and Migration rates have been assumed to remain the same throughout the simulation period, but there is some significant variance found over the time according to ONS projection assumptions. Hence in a further version the decision was made to re-calculate the probabilities according to the newly available trends. Then the probabilities of the three components of change are updated in light of the ONS assumptions used in ONS Subnational Population Projections (SNPP) and Mid-Year Estimates (MYE) at the local authority level during 2001-2031. For instance, the changes in mortality from one year to another are applied to the individual based probabilities to update the single year of age, sex and location specific probabilities. The longer term trends at 5 year intervals have been used from year 2011 due to the limitation of data that are available at the time this model was developed. The annual variations are updated as described above, using an averaged change rate throughout the 5 year duration. SNPP and MYE data on components of change are used as the basis of updating to capture the three components of change in small area populations.

### 3.5 Result alignments

It is always difficult to validate a spatial MSM due to the lack of appropriate empirical data. As the microdata used to estimate the transition probabilities in the dynamic MSM are often subject to sampling error or lack of critical characteristics to explain heterogeneous behaviours, it is not surprising that the predictions of unaligned dynamic MSMs can drift away from benchmark aggregates such as the official population projections. On the other hand, the decision makers and some researchers often have used the official datasets/models in the processes of decision making or a study. Such users are often reluctant to accept results/models drifting too far away from the official estimates. Due to the above reasons, recently the result alignment has emerged as a crucial component of many dynamic MSM.



Nowadays “almost all existing dynamic MSMs are adjusted to align to external projections of aggregate or group variables when used for policy analysis” (Anderson, 2001).

In an MSM little attention to changes in the underpinning transition probabilities while Macro models do. Therefore alignment exercises can be used to improve the MSM inputs. The transition probabilities and rates can also arguably better estimated for aggregate populations than disaggregate, as errors either way at disaggregate level cancel out at aggregate level. However, the problem with macro top-down approach is that it generates all sorts of inconsistencies, e.g. the outputs of the model are no longer a simple function of the inputs. The macroscopic model projections are less sensitive to sub-population disaggregation and often overlook the disaggregated characteristics.

Bækgaard (2002) in his thorough investigation of alignment identified the objective of alignment as to compensate for imperfectness of data and estimation techniques. As he pointed out: not only the total output, but also the distribution of base data and output can be aligned. Alignment not only includes the calibration of processes but also the adjustment of base data caused by, e.g. sample stratification. There are two major classes of alignment: alignment of outputs and alignment of inputs.

Through alignment exercises, macroscopic feedback can be imposed on MSM aggregate results. Therefore by aligning the aggregated results to ONS projection results, it enables incorporation of the recent and future population trends revealed in ONS assumptions that have significant impact on simulation results. The attempt is to find out the assumptions underlying the most up-to-date ONS SNPP projections on the basis of year 2006 population and modify the probabilities that have been used in various demographic transitions accordingly to align the model to the ONS projections.

There are a couple of steps in our alignment exercises. A framework has been designed for this exercise. The alignment exercises are composed of 2

assessments between 3 models. Model A is the ONS aggregate projection, Model B is a naïvely disaggregated hybrid MSM and Model C adopts a full disaggregation. The first assessment is through the alignments to the ONS projections by applying the naïvely disaggregated/averaged estimations that are used in ONS projection to all populations in Leeds wards. Then the results from the modified simulation are re-aggregated and compared against ONS' aggregated results to test the consistency of the model. In the second assessment, the ONS estimations are fully disaggregated to the ward level and the results are re-aggregated and compared to test the robustness of the model. The analyses and framework specified here will be discussed in details in Chapter 7.

It should be noticed that in these alignment exercises this study only tried to match the assumptions/trends found in ONS' SNPP. However this model is not trying to recreate the results of ONS. Due to the aggregate nature of the ONS SNPP projections, such projection results do not necessarily provide the best representation of Leeds population, especially at the scale of small areas. Previous studies have recognised the limitations in subnational projections (Smith and Shahidullah, 1995; Rees *et al.*, 2001; Smith and Tayman 2003). The alignment exercises will be the first steps towards model validation. Further development of the validation method is also being considered, e.g. investigation through more empirical studies will be carried out and findings from this will be used to help refine our model so that the model provides a better representation of the Leeds population. More details of the alignment framework and model validation will be discussed in Chapter 7.

## 3.6 Conclusion

We discussed the methodology of the project in this chapter. The methods used for system design and system development, as well as how the two aspects link to each other have been explained.

The population projection method, demographic processes, main components of the system, the development and validation methods of this dynamic spatial MSM have been discussed in this chapter. The component-cohort projection method has been explained using the Lexis diagram. The six demographic processes of Ageing, Mortality, Marriage, Fertility, Migration, and Health Change and the sequence of the processes have also been introduced. The representation of individuals and the interdependency between the individual, household and environment has also been explained. To explain how such design has been implemented into the model, the development methods and data selection have been described. Also the proposed method for model validation has been described with a framework design for alignment to ONS projections. However, there is a noticeable difference: this study is only trying to align to the ONS assumption, not trying to recreate the ONS results. Due to the limitation of aggregate projections, ONS projections do not necessarily provide the best representation of the ward population in Leeds.

The hybrid modelling approach has also been described here. This is mainly because we are trying to resolve two issues that arise from the requirement of this model. This model would like to use the list processing power and the real data roots of the MSM to tackle the scale issue that this study is facing: 761 thousand individuals in Leeds (mid-2007). The scale challenge for the model arises not only from the size of the base population, but also from the richness of the attributes, probability generation and updating.

At the same time, this study attempts to capture the individual movements, interactions and behaviours of people in Leeds in the model. It is found that MSM is not a very flexible instrument to accommodate modelling of such behaviours because of its statistical nature and lack of quality data on specific behaviours. ABM is an alternative social modelling approach, where individuals are modelled as agents that move around and interact with each other and the environment according to their built-in rules. Thus, it is very flexible to model heterogeneous individuals/sub-populations where there is a knowledge gap or data limitations.

After careful consideration, it is believed that a hybrid modelling approach can combine the strength of both MSM and ABM to accommodate both aspects of requirements of the model. Adopting the hybrid approach, we have achieved the goal of both the effective handling of a large-scale individual based system, as well as providing extra flexibility to model various movements, interactions and behaviours of sub-populations in different scenarios.

## Chapter 4

# Development of the dynamic spatial microsimulation model

## 4.1 Introduction

This chapter will focus on the systematic development of the dynamic spatial MSM. The model simulates the population dynamically from 2001 to 2031. The baseline population in 2001 is provided by colleagues (Birkin *et al.*, 2006). For more details of the baseline population see chapter 3. Six important demographic processes are modelled for populations in Leeds and each process is dealt with in separate modules, namely: Mortality, Fertility, Health, Household formation, Migration and Ageing. In the following sections, this chapter outlines the general principles adopted in modelling the above processes, including discussion of the factors, model design and simulation methods shared by all modules. Then more specific development

details will be described by individual modules. Finally the initial results are analysed and further developments of the model are proposed in Chapters 5 and 6.

## 4.2 General development method

In this section, the development of the dynamic spatial MSM will be described covering the following topics:

- the modelling approach;
- factors affecting the demographic processes;
- how the processes are simulated in the model and
- probabilities and data.

### 4.2.1 *Dynamic spatial MSM*

As discussed in previous chapters, macrosimulation and MicroSimulation Models (MSM) are alternative ways of representing the same demographic processes (Van Imhoff and Post, 1998). However, the richness of MSM is a useful device for the representation of both relationships between members of a population, and of the transitions between states within a population. A dynamic spatial MSM is used to model the population and its dynamics, although the model structure parallels the macro multi-state cohort-component projection model (see Chapter 3).

There are many different interpretations of what “dynamic” means. In this study, we believe that in a dynamic MSM, entities change their characteristics as a result of endogenous/internal factors within the model. A Dynamic MSM tries to move individuals forward through time, by “updating each attribute for each micro-unit for each time interval” (Caldwell, 1990). The updating of the demographic structure is performed by dynamically ageing the modelled population individually (by asking “yes or no” questions on birth, death, migration and other transitions) using transition probabilities from life tables and/or exogenous time series at each

simulation step (see section 4.3.3.1). As the end population of the first year will be the base population for the next year, the changes from the former year are captured within the population and can affect an individual's characteristics in the subsequent year. The dynamic MSM can also be used to assess the evolution patterns of a population over a longer period than a static MSM. Certain degrees of interactions between micro population units can be found in dynamic MSMs. Such interactions typically include processes such as birth and marriage (O'Donoghue, 2001). The population is modelled at the individual level. As described previously, dynamic MSM provides more realistic representation of the studied population than static MSM.

Compared to a static MSM, the dynamic MSM developed in this study

- builds in change over time;
- has processes for introducing units into the system of interest and taking them out and
- has the capacity of reproduction and of dying.

This model dynamically ages each micro unit by modelling for each time interval the state-to-state transitions that individuals experience and by adjusting the characteristics of the households that the individuals occupy consequent on those transitions.

The spatial MSM is a special type of MSM that simulates virtual populations in given geographical areas (Ballas *et al.*, 2005). In a spatial MSM, local contexts can be taken into account when studying the characteristics of these populations. Spatial MSM therefore have advantages over other MSMs in the exploration of spatial relationships and the analysis of the spatial implications of policy scenarios. A spatial MSM can be either static or dynamic, but within a dynamic spatial MSM, both the characteristics of the individual and the context can change.

MSM is chosen to model the demographic dynamics of small area populations. A MSM is chosen also because of its roots in working with real

data and simulating real phenomena. From a computing point of view, MSM provides the capability to model a large-scale complex system with its power of list processing. This makes manipulation of large volume of data efficient, as well as providing a better basis for system scalability. Due to the advantages described above, this model has been constructed in a form of a dynamic spatial MSM. This dynamic MSM models the population as individuals in households on a fine spatial scale of wards in the attempt to capture the local characteristics. The electoral ward is the basic unit in this area model. Wards are zones within local authorities (LAs), sub-national units of government, from which representatives are elected to the LA council by eligible electors (ONS, 2011a). Wherever possible, probabilities that drive the population changes have also been calculated at the ward level to reflect the local characteristics.

As described in Chapters 2, MSM has been used extensively in facilitating public policy development and has proved to be an effective approach. This study proposes a hybrid approach that uses ABM and MSM in order to model the social complexity in a creative way to strengthen the interaction and behaviour modelling of traditional MSMs. However, as pointed out in Chapter 3, the MSM modelling approach is less effective when movement, interaction and behaviours are important and it struggles when lacking appropriate microdata for important transitions. In an Agent Based Model (ABM), behaviours of agents (micro-units) can be influenced by interactions with other agents or the environment that they live in. Individual behaviours can be based on a set of built-in decision rules rather than a probability distribution across alternative choices. A hybrid modelling approach introduces ABM techniques in various sub-modules to strengthen such areas as mentioned above. An ABM is not used for modelling the whole system for two reasons: one, ABMs are more computationally costly - they are less effective than MSM in dealing with a large volume of data; two, ABMs are useful in exploring various behaviours, but are less strong in modelling real social phenomena with the ambition to assist decision making due to the weakness within their theoretic basis.



## 4.2.2 *Demographic processes*

Mortality, Fertility and Migration are the three fundamental components of the population change. In this study, knowledge about Ageing, Household Formation/Marriage and Health information are felt to be very important, especially for public policy making and demographic planning, due to their important impact on housing, transport, health and other public service provision, for instance. Also such processes are often connected to other demographic processes, e.g. Migration and they may lead to changes in the fundamental components of change. For instance, most Household Formation/Marriage processes are intertwined with Migration. Due to the above two reasons, six demographic processes have been modelled in this study (The processes have been illustrated in Figure 3.3, Chapter 3).

This model projects each component of population change separately, but each component can also affect the others. Therefore, although each process has been developed in a separate module, modules often interact with each other. For instance, the migration process may create more/fewer marriage candidates in a small area and therefore impact on the household formation patterns.

Transition probabilities for each of these events are applied at discrete one year intervals. Many different orderings are possible for the sequence of events. Mortality is considered fairly early for practical reasons, since if an individual life course is terminated subsequent processes can then be ignored. It is considered more logical to evaluate fertility following the formation of marriages and partnerships. As Migration often interacts with other processes and is more complicated, it is modelled after the relevant demographic processes. Logically the model should allow re-ordering of the the modules with appropriate inputs and the same results would be obtained. For instance, if the mortality is modelled first and all the probabilities used in following modules are conditional on survival, then the results should not demonstrate significant difference. Even if they are not conditional on survival, the impact will not be significant except in older ages where

mortality risks begin to rise steeply with age. It is for this reason that only the fertility probabilities are adjusted in this model, as the population at risk is mainly younger females. In future, when more time is available, such simplification can be improved upon with further work. In the next section, the selection of the main attributes will be discussed.

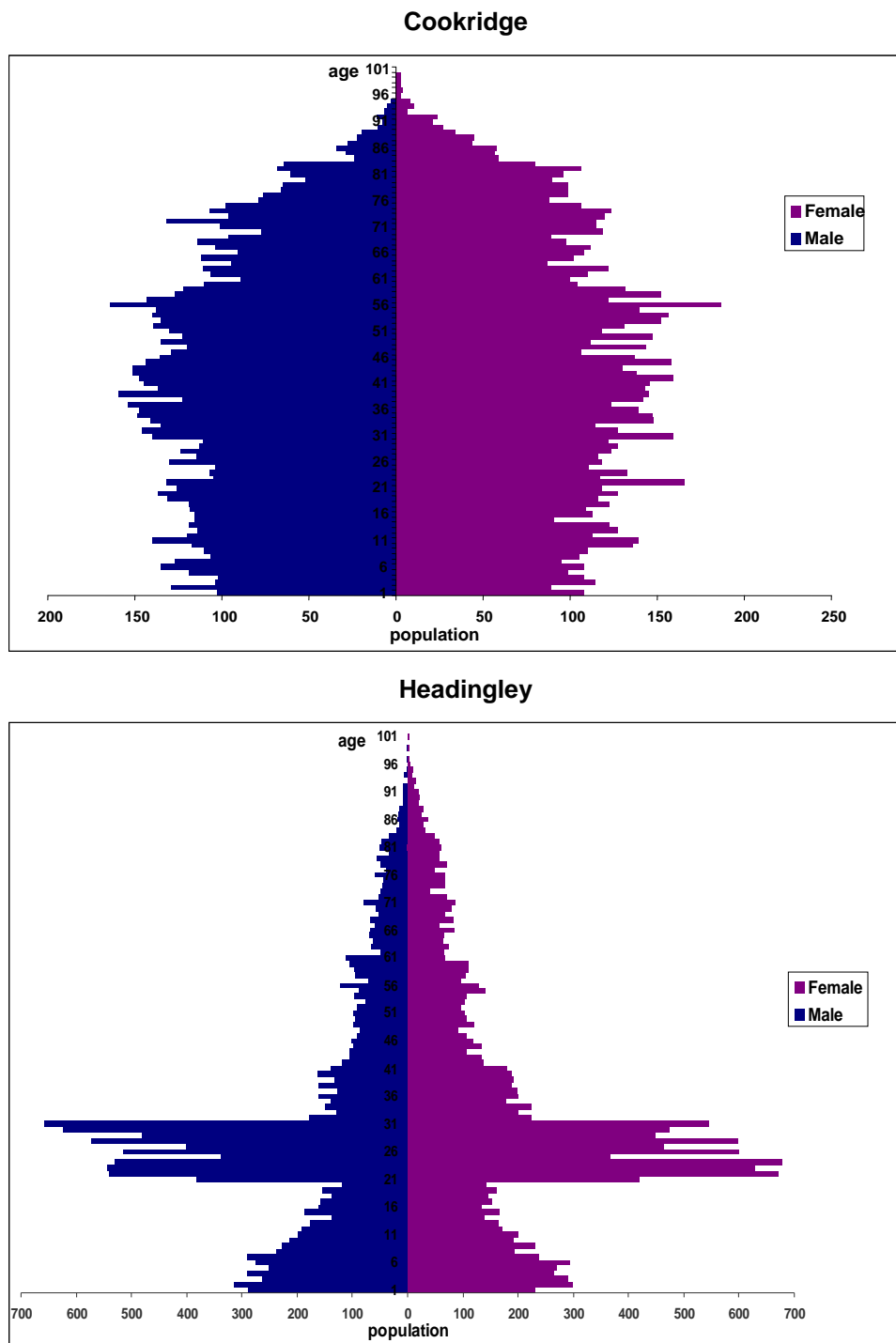
### 4.2.3 *Base Attributes*

A number of important attributes have been considered in detail in Chapter 3 in terms of their influence on demographic changes in individuals and populations. The reasons for why or why not they are selected have also been explained. In this section, only the four attributes that have been used as the base attributes in probability calculations for important demographic transitions are discussed:

- Age
- Sex
- Environmental factors
- Marital Status.

Age and Sex are the two fundamental factors and they play indispensable roles in all the six important demographic transitions. In fact, Age and Sex composition are so important to the nature and functioning of societies that all traditional social models are based on it. The population pyramid is a well used graph used to illustrate populations' age-sex structures. In a population pyramid, the numbers/percentages of males and females in each age group are represented in the graph. Two population pyramids are provided here as examples to represent the age-sex composition of the populations in Leeds wards: Cookridge and Headingley in 2001 (Figure 4.1).

From the population pyramids, we can see clearly that the population structures in small areas can be very different. Headingley is an area that contains a concentration of university students' accommodation, while Cookridge is a more established suburban area. Students tend to leave the



**Figure 4.1 Age-Sex distribution in Leeds small area populations: wards Cookridge and Headingley in 2001**

Source: Author's computation using ONS (2001g)

area upon the completion of their studies and new students move into the area. Due to the replenishment of the student migration, it is no surprise that we cannot see an ageing trend in the population in Headingley. There are more young people, especially aged 20-29 in Headingley: in fact, this age group is the largest group within this local population. There are few people

aged over 60 and even smaller numbers of people aged over 80 in Headingley. On the other hand, there are more old people in Cookridge, especially aged over 65.

As indicated in Figure 4.1, the Age-Sex composition has a fundamental impact on the population patterns, even in small areas. Therefore it is important to use them in projecting the population changes with rare exceptions. In this study, Sex is not used as a base attribute in modelling the Fertility process. This study uses a female-dominant fertility model, one in which only females are considered at risk of giving birth. The key factors that influence fertility rates are the age of women and their marital status.

**Environmental factors:** This model intends to introduce the environmental impact from a geographical view. As discussed in Chapter 3, lots of demographic and socio-economic processes (probabilities, rates and transitions) vary by geographical location. There are also attempts try to explain such variations by including in the description of the individuals in the MSM e.g. age, sex, socio-economic position, occupation and ethnicity and environmental influences. However, often some of these individual or household variables cannot be measured very well and there are lots of factors that remain unobserved. Therefore geographical location is used in this study as a proxy for all such unobserved factors, as well as to provide a local context. Area codes are used as the surrogates for unobserved variables (features for wherever good microdata are not available).

**Marital status:** Marital status is the last base attribute that has been selected to model the demographic changes. This is because of the important role it plays in processes such as Marriage and Fertility. The current marital status affects the decisions of marriage and more births are found within marriage than outside marriage. Although Marital status has been found important in influencing mortality, especially in people's later stage of life, their relationship is not pursued further, due to the time and resource limits of this project.

These four base attributes have been used in calculating various

probabilities to model important demographic transitions in the model. The processes in which they are embedded are explained in Section 4.2.4.

#### 4.2.4 *General processing procedures*

There are six important demographic processes modelled in this study, where individuals go through the changes in Ageing, Mortality, Fertility, Household formation/Marriage and Migration.

Age and sex are the fundamental factors that contribute to demographic processes (Rowland, 2003). Transition probabilities for each of the model processes vary by geographical area as a result of the differences in social, economic and environmental profile of area populations. Marital status is important in Household Formation/Marriage and Fertility processes. Therefore this model assumes that all the demographic processes modelled are functions of age, sex/marital status (in Fertility only marital status is used, due to the female-led approach) and location. Probabilities of even occurrence are computed for population groups classified by age, sex, marital status and geographical locations from suitable census or survey data, correctly tabulated. The probabilities are used to determine whether an event occurs to an individual in the model population. Therefore a change/an event occurs to an individual in a demographic process if

$$\text{Ran}(0,1) \leq P(k_2|a, s, l, k_1) \quad (\text{Equation 4.1})$$

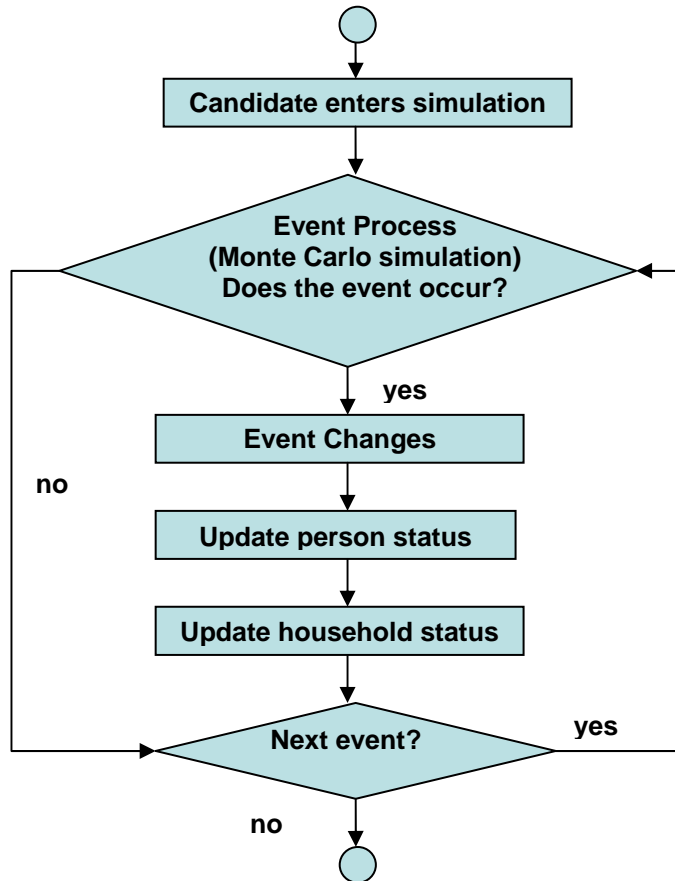
where  $\text{Ran}$  is a random number between 0 and 1,  $a$  is the age,  $s$  is the sex and  $l$  is the location of the individual and  $P(k_2|a, s, l, k_1)$  is the probability of the event occurrence for an individual of age  $a$ , sex  $s$  at location  $l$  with current characteristic of  $k_1$  to change to a characteristic  $k_2$ . For instance: in mortality,  $k_1$  might be “alive” and  $k_2$  will be “dead”; in fertility,  $k_1$  might be “no child born alive during the year” and  $k_2$  will be “child(ren) born alive during the year”.

Transition probabilities for each of the events within the six demographic processes are applied at discrete one year intervals. According to the nature of specific demographic processes, extra factors will be introduced in calculating the probabilities. For instance, the Fertility process considers women by age, marital status and location. Marital status is used here as patterns of births within marriages differ hugely from births outside marriages.

Monte Carlo simulation is adopted in this MSM, because it is very useful when it is infeasible or impossible to compute an exact result with a deterministic algorithm. Monte Carlo simulation gets its name from the resort on the southern coast of France where gamblers bet at games of chance of a random number at each draw. Monte Carlo simulation uses repeated computation of random or pseudo-random numbers. Monte Carlo simulation converts uncertainties about the relationship between input variables and output variables of a model into conditional probabilities. By randomly selecting values from inputs repeatedly, it recalculates and brings out the probability of the outputs. Apart from the Ageing process, the rest five demographic processes are modelled through Monte Carlo simulation where age, sex and location specific probabilities are applied to determine whether or not an event of demographic change will happen. For example, if you want to find out if a male candidate aged 85 living in Aireborough is going to survive this year, you need to compare the mortality probability for men aged 85 in Aireborough (0.104505) to a generated random number between 0 and 1 (0.563542) as described in Equation 4.1. As the random number is bigger than the probability, it can then be decided that this person survives this year and proceeds to determine whether other demographic changes happen to him this year. Each determination requires a new random number and a corresponding event probability for a person with certain attributes in that year. All probabilities are updated annually and any change in any factor will result in the change of the probability and in turn the simulation process and result.

The general microsimulation process in this model uses the Monte Carlo method to simulate the changes within each demographic process. Monte Carlo simulation can introduce a degree of randomness and often results from different runs vary. Various what-if scenarios and sensitivity analysis have been used to help reduce variances in results from different runs. To facilitate the estimates of the absolute responses; i.e., prediction or interpolation from the observed responses for the scenarios that have already been simulated, variance reduction techniques are needed. Most used techniques include common random numbers, antithetic variates, control variates, conditioning, stratified sampling and importance sampling (Law, 2007; Rubinstein and Kroese, 2008). The common random number method (sometimes called correlated sampling, matched streams or matched pairs) is intuitively preferred, because it reduces variability based on the principle that a random number to be generated using the same draw from a random number stream for all configurations. Due to time scale and the resource limitations, this project uses a simple method that is based on the same principle. All probabilities used in this model are of 6 decimals to minimise the variance and the same draw from the number stream generated using `Math.random()` method from Java, employed in all runs. The current control has provided reasonable variance reduction for the model results. However, given more time and resources, a more sophisticated method can be developed to better facilitate such purpose. The general process is illustrated in Figure 4.2.

When a candidate for a certain transition enters the simulation, a process will ask the question whether this transition happens or not through a Monte Carlo simulation on the basis of a specific probability. If the answer is affirmative, then it carries on processing the changes and updating the individual's status and household status (e.g. minus one household member in case of death), before move on to process the next event in the same fashion. If the answer is negative, then it goes directly to update the individual's status and household status before move to the next event.



**Figure 4.2 The general simulation procedure**

#### 4.2.5 *Probabilities and datasets*

All people will age one more year at the end of each annual cycle of simulation. Event occurrences in the other five processes are all based on specific probabilities. Finding, obtaining and pre-processing appropriate data for probability calculation is a non-trivial task. Generally speaking, datasets at three different spatial scales have been used to calculate various probabilities: England and Wales, local authority and ward level.

A method has been developed to localise the probabilities for Mortality and Fertility processes to the ward level using selected national, sub-national (for Leeds LA) and ward-based datasets. In the Mortality process, probabilities for individuals in each of 33 Leeds wards, both sex and at



single year of age from 0-100+ have been calculated using all three levels of data. There are 6666 mortality probabilities in total (age 0-100+: 101 age bands; for both sexes:  $2*101=202$ ; for 33 wards in Leeds:  $33*202=6666$ ). Similarly, the fertility probabilities are based on single year of age and marital status of the mothers. This produces 2,112 localised probabilities (age under 15-45+: 32 age bands; for marital status of single or married:  $2*32=64$ ; 33 wards in Leeds:  $64*33=2,112$ ). The sex of the babies is then determined using a national level probability. However, due to the computing demands and limitations in available data, probabilities of sub-national and national level have been used for Household formation and Health change. Household formation probabilities are based on age, sex and current marital status. Then the spouse age distributions are extracted from the commissioned table of age of married couples (ONS, 2001f). Health change probabilities using synthetic cohorts by age and sex classification are calculated with information from the HSAR (for household population) and ISAR (for communal establishment population) at the national level. Due to the complex nature of demographic processes, the input probabilities range from national to ward level.

Details of the probability generation will be described in the development of individual demographic processes. The general principle of probability application is based on Monte Carlo simulation, but various methods have been developed for computing efficiency. Details of these will also be described in individual processes.

The main datasets used in the probability calculation are listed in Chapter 3 and further details of data can also be found there.

In the following sections, each of the processes will be described. We will examine the conceptualisation of the model, the modelling method, probability generation and event determination. As Mortality is the first module to be explained, more details have been provided to illustrate the procedures such as the attribute selection.

## 4.3 Mortality

As Rowland (2003) points out “Matters of life and death have been a central interest of demography from its beginning”. This section introduces how the mortality process is modelled during the annual projection of the studied population.

### *4.3.1 Factors*

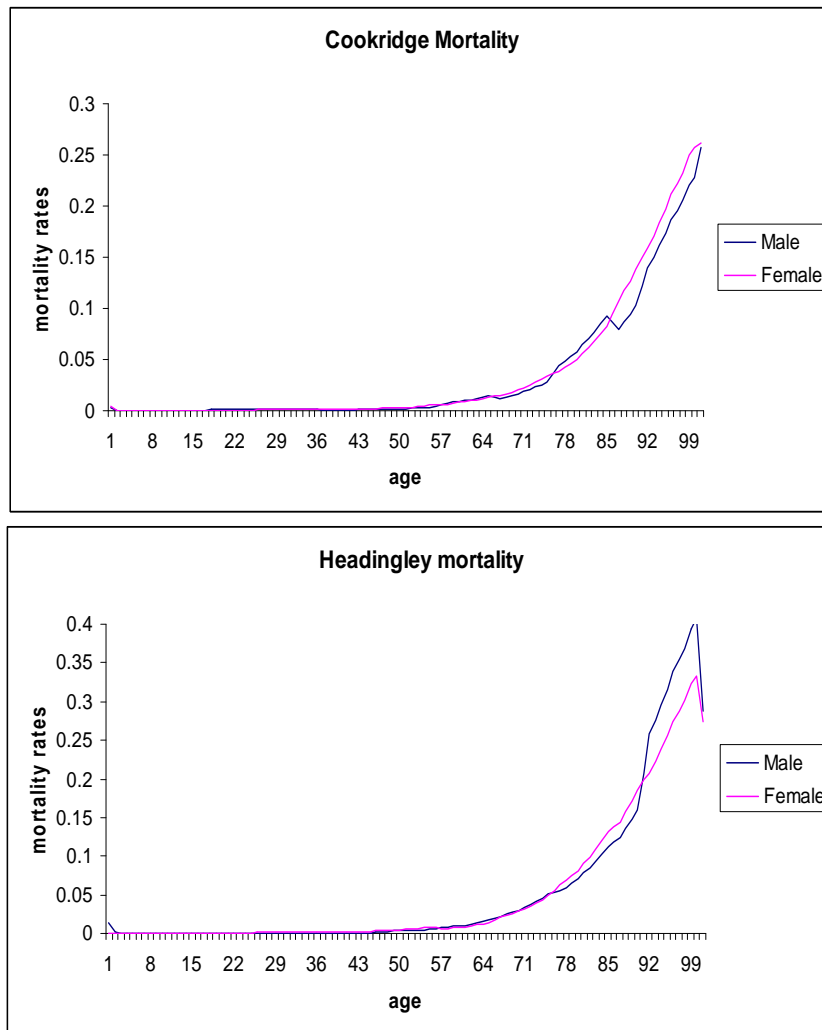
This section discusses the major factors which influence mortality risks in small area populations, drawing on the relevant literature. At the end of the discussion, decisions will be made on which factors to carry forward into the MSM.

There are a number of factors that influence the risk of mortality for individuals and populations such as:

- Age
- Sex
- Socio-economic position (degree of poverty/deprivation)
- Family context/type of household
- Ethnicity
- Life Style (smoking, alcohol, drug use, diet, exercise)
- Geographical variations.

These factors will be discussed one by one in the following section.

**Age:** Mortality risk varies systematically and significantly by age. Any model that needs to predict or project mortality must take this into account. From the age-sex mortality rate curves presented in Figure 4.3, we can clearly see the impact of age on mortality risk.



**Figure 4.3 Mortality by age and sex in Leeds Wards 2001**

Source: Author's computation using ONS (2001g, h)

From the mortality curve the mortality variance across different age groups can be observed:

- infant mortality starts high but falls after age 0 – this indicates a higher mortality at birth in the two wards (Headingley has a higher mortality risk at birth than Cookridge);
- there are very low rates before age 16 and
- mortality rates rise from age 16 onwards.

There may be departures from this picture: there is some evidence at older ages that the rise in mortality may reduce and even drop for older age groups (90+). This may indicate a selection effect: the survivors at such ages are the fittest. Or there may be errors in estimating the population at very

old ages. The sample number of very old ages in a small area is also an issue. Due to the small base population at ages 90+, one or two more deaths can change mortality rate for that age group substantially.

**Sex:** Sex is another important factor. In developed countries, the average life expectancy of women is often longer than that of males, partly because of biological factors, and partly because of behaviour differences. Lopez (1983) believes the following:

- Due to their responsibility for reproduction, females are biologically designed to survive longer than males in most species. Among humans, men have greater susceptibility to life-threatening diseases.
- Also, men are more likely to be engaged in more dangerous occupations, as well as misbehaviours. They may smoke, drink and abuse drugs more than women, too.

From Figure 4.3, we can also observe the sex impact on mortality curve.

- Men aged 85 or over have lower mortality rates and therefore have higher survival probabilities than women in Cookridge, indicating the selection effect. This suggests that women may live longer, but they may live in poorer health than men in older ages.
- The Headingley mortality curve suggests otherwise. There is more mortality in women from age 76, but the male mortality quickly increased from age 90. However, Headingley population is hugely impacted by the student migration and there may be limited samples of older people in this small area.

However, it should be noticed that the mortality rates for wards are subject to considerable variability from one year to the next. This variability shows up particular in the final ages where the numbers of people in their 90s is small and where there is a problem of poor estimation (Grayer, 2011).

**Socio-economic position:** There is a strong relationship between socio-economic status and mortality risk. It is observed that there are higher risks

of mortality for those who are more deprived. There is a strong relationship between socio-economic positions and mortality risk – the higher a person's socio-economic position, the lower the mortality risk (Valkonen, 2005).

**Family context:** This is important in later life. For example, married men living with their wives have lower risks of death than single men living alone.

**Ethnicity:** It may be possible to infer ethnic mortality rates from ethnic health information in the census, because self-reported health predicts subsequent mortality well but not perfectly (Rees *et al.*, 2009).

**Lifestyle variables:** Because of their importance these are currently being estimated for small areas by various research teams.

**Geographical variation:** The differences in regional mortality have been fairly constant for at least the past 30 years. Overall mortality was highest in the North and West of England and Wales (Britton, 1989). Spatial variation in mortality rates is often associated with social deprivation, especially social inequalities in health. However the observed geographical variations in mortality rates in England and Wales, even after taking into account levels of social deprivation and area type, show marked regional differences in all-cause mortality rates (Langford and Bentham, 1996; Boyle, 2004; Shaw *et al.*, 2008). Geographical analyses of mortality rates not only demonstrate interesting patterns and provide clues about aetiological factors at the regional level, but even at much finer spatial scale of small areas, significant spatial variations in mortality exist. An example is provided in Figure 4.3. The mortality curves of two different wards, Cookridge and Headingley, demonstrate that mortality varies at the ward scale.

It would be possible to incorporate any or all of these factors into the mortality module of the MSM. However, this poses two difficulties, the first theoretical and the second practical. The theoretical difficulty is the problem concerning the state space representation. A state space is the set of all possible states of a dynamical system; each state of the system corresponds

to a unique point in the state space (Meiss, 2007). In this model, a state space representing all factors simultaneously would be huge and the resulting multistate transitions could not be reliably estimated. We would therefore need to model the multistate transitions by chaining together transitions based on subsets of the state space. This would need decisions on dependence or independence of factors and an ordering of the chain of probabilities. This is precisely the issue that suggested the move from macrosimulation to microsimulation. We must therefore choose to use only a subset of the factors for modelling mortality. This is acceptable because the aim of the project is to show how an operational dynamic spatial MSM for a large set of small area populations can be built. It is not the aim of the thesis to explain as fully as possible the factors determining individual mortality risks.

There is a strong argument that we must include the age factor because mortality varies by several orders of magnitude across the age range. No other variable has anything like this influence. We should also include sex because there are systematic differences in the mortality experience of men and women (4 to 6 years' higher life expectancy for women than men, although recent trends in UK shows that men are catching up with women).

We can estimate socio-economic differences in the mortality risk using national datasets. The longitudinal study links together censuses and mortality records, but the effort involved is considerable and only national results can be generated.

We could include ethnic differentials in mortality risk as well because recent estimates have been made of ethnic specific mortality (Rees *et al.*, 2009). However, about 90% of the Leeds population belongs to one group only, the White British, which experiences mortality risks close to the national average. We would not gain much explanatory power therefore by adopting the ethnicity.

Lifestyle variables (smoking, obesity) have been estimated for small areas using aggregate statistical model (Moon *et al.*, 2006) or using

microsimulations (Edwards and Clarke, 2009), but integration into the current model would be challenging.

Location is a useful proxy variable for the simultaneous operation of socio-economic, ethnic, life style and environmental variables. If we know the locations of mortality, we can capture some of the effects of these factors on mortality.

We therefore develop sets of age-sex mortality rates for all Leeds wards for use in the mortality module of an MSM. This is mainly due to the following reasons:

- Age-Sex specific mortality data are available (with some computation), while for other attributes there is little information corresponding to mortality;
- Such data cover all regions and all age and sex groups (mortality differences by socio-economic status are usually only well-studied for working age men and it is hard to estimate socioeconomic mortalities/differences for other groups such as women, children or the retired people);
- The categories of the two variables, age and sex, are the most consistent, compared to other variables (e.g. the categories of socio-economic from various data).

However, age-sex specific data can only provide a basic picture of mortality distribution. A way that can represent some features of the other factors into the mortality probability generation needs to be introduced. Given the importance in the geographical variation in mortality described previously, the mortality rates for wards have been estimated to represent some of the factors. Age-sex standardised mortality ratios and life expectancies vary systematically across the UK in ways that reflect the distribution of socio-economic status, life style, ethnicity and environmental variables. The geographical variation can also reflect the influence of the other factors. These can be summarised as follows:

- South East (lower mortality) to North West (worse mortality);
- cities (worse mortality) versus outside cities (better mortality);
- particularly high mortality in areas of industrial decline (e.g. former coalfields) versus low mortality in areas of industrial vibrancy (e.g. South East) and
- particularly high mortality where life styles are health threatening (e.g. Glasgow) versus low mortality where life styles are better (e.g. London).

Leeds is in the centre of the UK and this is reflected in its mortality rates, which is about the UK average, although different small areas will demonstrate unique geographical impact from the industry influence, downtown-suburban and lifestyle differences.

So far in Section 4.3 we have discussed the factors affecting mortality and proposed that we use mortality rates broken down by age, sex and location (ward). We now discuss how to measure these rates and how to convert them into the probabilities needed in the MSM. To do this we need to introduce a mortality model, the life table. This was first proposed in the 17<sup>th</sup> century by the English statistician John Grant, then expressed mathematically by Edmond Halley in the 18<sup>th</sup> century and applied systematically to local populations in the 19<sup>th</sup> century by William Farr, who worked on the census and death registration to try to understand the epidemics of mid-19<sup>th</sup> century London (Woods, 2000). The life tables refer to the chances of dying given membership of an age group and sex at the start of the year. Period-cohort survivorship probabilities,  $S_x$ , are used in this simulation. The details will be discussed in the probability calculation section.

Here we use age-sex specific mortality rates to create full life tables for Leeds wards and then add the computation of survival probabilities, which are entered into the Mortality module of the MSM. This improves this MSM by providing a local context.



### 4.3.2 *Mortality process in simulation*

As demographic literature has pointed out that age and sex are the fundamental factors contribute to demographic processes (Rowland, 2003). A third dimension of the geography has been added in this model in the attempt to capture the local characteristics that can demonstrate a degree of the impact of other factors. This model assumes that the mortality process is a function of age, sex and location. An individual dies if

$$\text{Ran}(0,1) \leq P(S|a, s, l) \quad (\text{Equation 4.2})$$

where  $\text{Ran}$  is a random number between 0 and 1,  $a$  is the age,  $s$  is the sex and  $l$  is the location of the individual and  $P(S|a, s, l)$  is the probability of the survival  $S$  over a year applicable to an individual of age  $a$ , sex  $s$  at location  $l$ .

The process of Mortality is pretty straight forward. Each individual is tested whether they are going to survive for next year's simulation by the Mortality Determinator implemented using the Monte Carlo method. If the individual survives, then he/she moves on to next process; otherwise the record is removed from the database and the Mortality module goes on to update the household and area information, e.g. household size, total number of people in area. If the deceased is an HRP (Household Referenced Person) then a new HRP is selected.

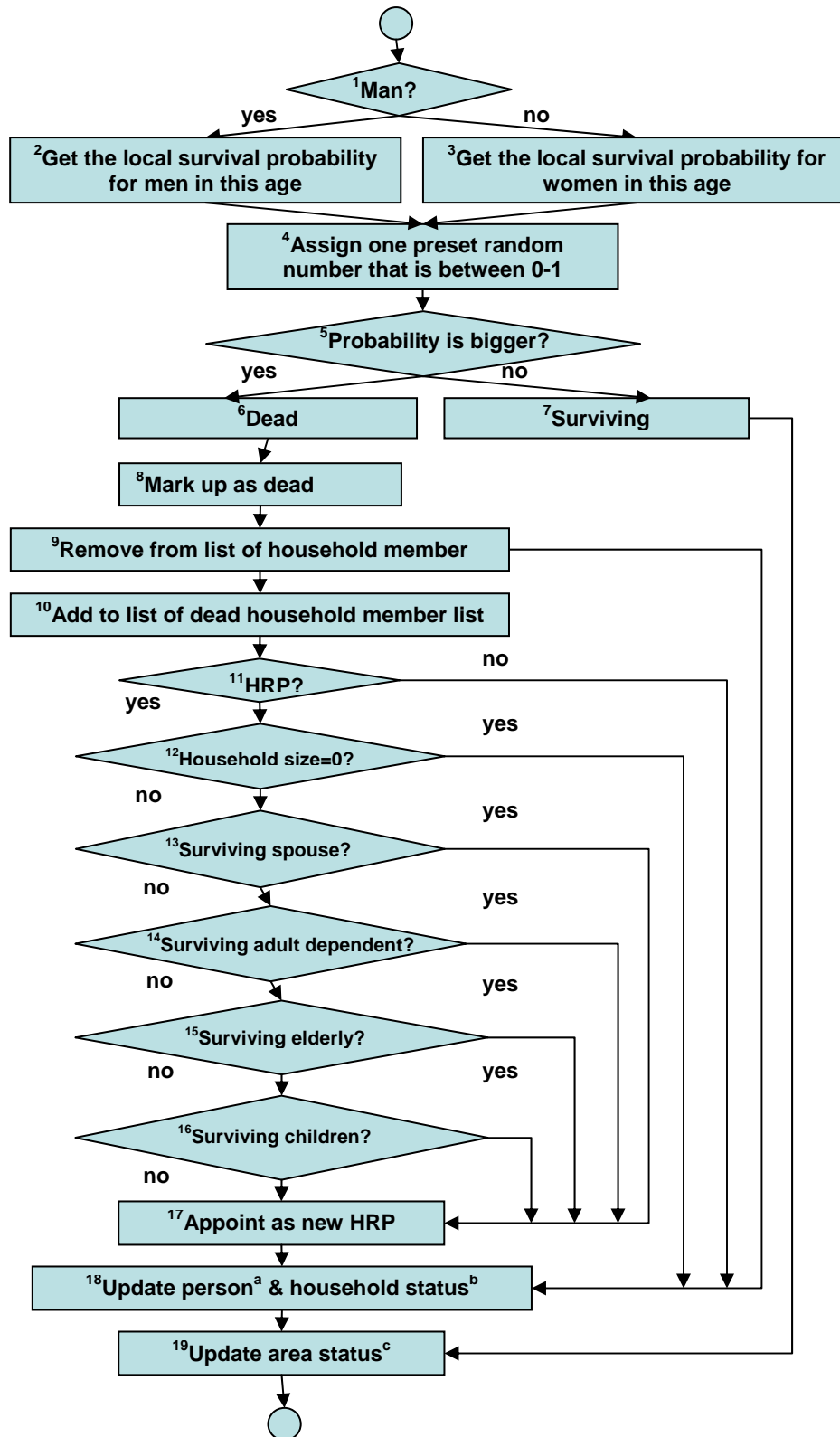
The Mortality simulation process is illustrated in Figure 4.4. Note that survival probabilities instead of mortality probabilities are used in the model. The Mortality process is straightforward. Even though simple, it demonstrates the strength of MSM in the representation of both relationships between members of a population and of the transitions between states within a population (Figure 4.4).

### 4.3.3 *Mortality probability generation*

A life table method has been developed to estimate the mortality/survival probability for an individual within specific age, sex and location (ward) characteristics.

#### 4.3.3.1 *Life table method*

The life table method has been adopted for the generation of the mortality and survival probabilities. A life table examines the toll of mortality, measuring life expectancy and the extent to which death diminishes population numbers as age increases. Although often used in mortality studies the life table also plays an important role in the studies of fertility and migration, as well as population growth and structure. A life table shows for a person at each age, the probability for them to die before their next birthday ( $q_x$ ) and can be extended to give the probability that a person in a given age group survives to the next age  $S_x$ .  $S_x$  is used in this model instead of  $q_x$  that is used in ONS projections, as the focus of the model is the change of survivors in the system. The mortality process is used to determine whether an individual survives rather than to determine who dies as illustrated in Figure 4.4. A complete life table provides information on single years of age, unlike abridged life tables which provide a summary for selected ages (Rowland, 2003). The complete life table is used in this study to provide information on age 0 to 100+ for each sex (Table 4.1 shows an example). The steps of using a life table to calculate the mortality/survival probabilities are summarised as follows (Rees, 2007):



**Figure 4.4 Mortality Process**

Note: the pre-super-script in the diagram represents the simulation step and the post-super-scripts are:

- a. the order to appoint as new HRP's are spouses, adult dependents, elderly dependents and child dependents
- b. household size, household member type and count, family roles etc.
- c. total population in area, total households in area etc.

**Step 1:** enter observed data:  ${}_1D_x$  and  ${}_1P_x$

where  ${}_1D_x$  = deaths by single year of age between ages  $x$  and  $x+1$

${}_1P_x$  = population by single year of age between ages  $x$  and  $x+1$

**Step 2:** compute mortality rates:

$${}_1M_x = {}_1D_x / {}_1P_x \quad (\text{Equation 4.3})$$

**Step 3:** compute mortality probabilities:

$${}_1q_x = {}_1M_x / (1 + 0.5 {}_1M_x) \quad (\text{Equation 4.4})$$

Except for the first age and the last, open-ended age (0.3 is the constant suggested by Rowland, 2003):

$${}_1q_x = {}_1M_x / (1 + 0.3 {}_1M_x) \quad (\text{Equation 4.5})$$

$$\text{and } {}_{\infty}q_{100} = 1.0 \quad (\text{Equation 4.6})$$

**Step 4:** compute survival probabilities (age-cohort variables):

$${}_1p_x = 1 - {}_1q_x \quad (\text{Equation 4.7})$$

$$\text{and } {}_{\infty}p_{100} = 0.0 \quad (\text{Equation 4.8})$$

**Step 5:** compute number surviving of the radix (the number of babies from each year into the lifetable stationary population and normally the radix  $l_0 = 100,000$ )

$$l_1 = l_0 * p_0 \quad (\text{Equation 4.9})$$

$$l_x = l_{x-1} * p_{x-1} \quad (\text{Equation 4.10})$$

**Step 6:** compute number not surviving from radix:

$${}_1d_x = l_x - l_{x+1} \quad (\text{Equation 4.11})$$

**Step 7:** compute life years lived in age interval:

$${}_1L_x = l_{x+1} + 0.5{}_1d_x \quad (\text{Equation 4.12})$$

$${}_1L_0 = l_1 + 0.3{}_1d_0 \quad (\text{Equation 4.13})$$

and

$${}_{\infty}L_z = l_z (1/{}_{\infty}M_z) \quad (\text{Equation 4.14})$$

where  $z$  is the age that states the last open-ended age interval

**Step 8:** Compute cumulative life years lived beyond age  $x$ :

$$T_x = T_{x+1} + L_x \quad (\text{Equation 4.15})$$

**Step 9:** Compute life expectancy:

$$e_x = T_x/l_x \quad (\text{Equation 4.16})$$

**Step 10:** Compute survivorship probabilities (period-cohort variables):

$${}_1S_x = {}_1L_{x+1}/{}_1L_x \quad (\text{Equation 4.17})$$

$${}_1S_0 = {}_1L_x/l_0 \quad (\text{Equation 4.18})$$

$$\text{and } {}_{\infty}S_z = {}_{\infty}S_{z-1} = {}_{\infty}L_z/({}_{\infty}L_z + {}_1L_z) \quad (\text{Equation 4.19})$$

Through the steps, we can see that period-age mortality rates are converted into period-cohort probabilities using a lifetable model. Having completed the above steps, the result will be a populated life table and it should look like the example in Table 4.1, the lifetable of a ward in Leeds, Cookridge. The change of the variable type from period-age to period-cohort is also indicated in Table 4.1.

**Table 4.1 Life table example: Cookridge in 2001**

Age	age interval	fraction live	Estimated Deaths	Population	Estimated Mortality rate	Mort. Prob.	Surv. Prob.	Cohort surv.	Cohort non-surv.	Life years	Rem. Life Years	Life exp.	Surv. Prob.
Variable type					Period-age	Cohort-age	Cohort-age	Cohort-age	Cohort-age	Age	Age	Age	Period-cohort
x	n	ax	nDx	nPx	nMx	nqx	np <sub>x</sub>	l <sub>x</sub>	dx	nL <sub>x</sub>	T <sub>x</sub>	e <sub>x</sub>	nS <sub>x</sub>
0	1	0.1	0.33	101	0.003300	0.003291	0.996709	100000	329	99704	7832730	78.33	0.997038
1	1	0.5	0.00	76	0.000000	0.000000	1.000000	99671	0	99671	7733026	77.59	0.999670
2	1	0.5	0.00	92	0.000000	0.000000	1.000000	99671	0	99671	7633355	76.59	1.000000
3	1	0.5	0.00	118	0.000000	0.000000	1.000000	99671	0	99671	7533684	75.59	1.000000
4	1	0.5	0.00	128	0.000000	0.000000	1.000000	99671	0	99671	7434013	74.59	1.000000
5	1	0.5	0.03	119	0.000263	0.000263	0.999737	99671	26	99658	7334342	73.59	0.999868
6	1	0.5	0.02	122	0.000201	0.000201	0.999799	99645	20	99635	7234685	72.60	0.999768
7	1	0.5	0.03	122	0.000237	0.000237	0.999763	99625	24	99613	7135050	71.62	0.999781
8	1	0.5	0.02	116	0.000199	0.000199	0.999801	99601	20	99591	7035437	70.64	0.999782
...	...	...	...	...	...	...	...	...	...	...	...	...	...
99	1	0.5	0.00	0	0.250170	0.222356	0.777644	3587	797	3188	12275	3.42	0.772270
100+	1	3.3	0.00	0	0.306903	1.000000	0.000000	2789	2789	9088	9088	3.26	0.740313

Source: Rees *et al.* (2004)

#### 4.3.3.2 Estimating localised mortality probability

Age, sex and location all have significant impact on the mortality probability differences. We need to use single year of age, sex and ward specific survivorship probabilities for the MSM. However, such life table is not available. A method needs to be developed to generate the life table at the ward level. This method takes two datasets: the ward level mortality data developed by Rees *et al.* (2004) where Leeds population is in 5 year age bands and the ONS life table at the level of England and Wales where the population is in single years of age. We use the two datasets to work out the survival probability for Leeds by single year of age, sex and ward.

A method is developed to calculate the probabilities and the simplified process is described below.

**Step 1:** Download from the ONS (2001g) the population by single year of age and sex for Leeds wards.

**Step 2:** Estimate the single year of age mortality rates for Leeds wards

$$M_x^w = M_x^n \times [D_y^w / \sum_{x \in y} (M_x^n P_x^w)] \quad (\text{Equation 4.20})$$

where

M =observed or estimated mortality rate,

D = observed deaths,

x = single year of age,

y = five year age group,

P = population

n= nation (England and Wales)

w= ward (in Leeds)

This method is approximate but robust enough for current purposes. See Grayer (2011) for a full review of methods for estimating small area lifetables. These mortality rates are computed separately for men and women. The spatial level is indicated by n and w: n means national level, w means ward level. The national mortality is modified by the ratio of observed ward deaths in a five year age group to the expected deaths found by applying national rates by single years x to the ward population by single years. The calculated ward based probabilities are then adjusted according to the observed deaths in Vital Statistics to balance observed and predicted deaths in small areas (ONS, 2001g). Thus the spatially disaggregated mortality still reflects the trends in the national lifetable. The final survival probability is applied to each individual on the basis of age, sex and location each time period.

**Step 3:** Set up the ward single year of age life tables in the same format as the UK single year of age life tables and use the results of the estimation of single year of age mortality rates.

**Step 4:** Follow the steps described in the life table method to calculate the survivorship probabilities for period-cohorts as the probabilities for the MSM.

## 4.4 Fertility

The Fertility module deals with the process where women at risk have new babies. A similar approach has been adopted as in the Mortality module. First, the Fertility module needs to determine whether a woman gives birth in the current time interval. If no birth happens, then the woman simply carries on to the next simulation module. If the birth happens, then it needs to determine whether this is a singleton or a twin birth. Triplets or higher number of multiple births are not considered in this model as it wouldn't significantly affect the results of the model. The sex of babies is determined based on national probabilities. The Fertility module then goes on to update the personal, household and area information as in the Mortality module.



### 4.4.1 *Factors*

The Fertility module goes through similar processes as with the mortality process when considering various factors that will affect the fertility rates. After careful consideration, the following three attributes are selected to generate the fertility probabilities for this microsimulation. In the following section the reasons of using them are explained.

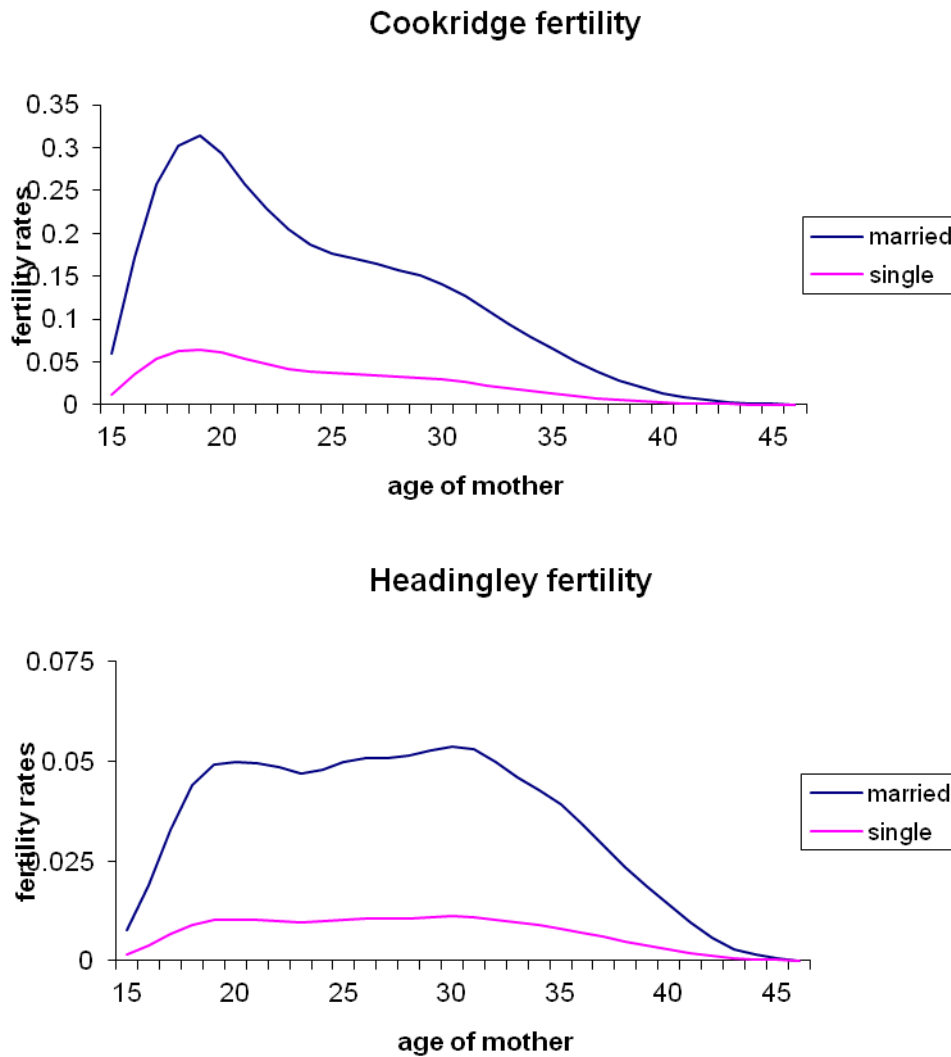
- Age
- Marital Status
- Geographical variation

**Age:** Fertility risk varies systematically and significantly by age. Women are naturally fertile from their early or mid-teens to their late forties. Natural fertility is higher for women aged 18-39 than either younger or older women. Medical techniques are extending fertility beyond age 50, but such cases are rare and may be ignored. Therefore, for these biological reasons, age has the greatest impact on fertility risk. The age-fertility rate curve is presented in Figure 4.5 to illustrate the age impact on fertility risk. Details of how the fertility rates are calculated will be described in the following sections.

Figure 4.5 shows the variation of fertility by age: most women give birth to babies between the ages of 15 to 45. In Cookridge, the peak age of fertility is around 19, particularly for married women, but in Headingley, there is a flat plateau of fertility rates between 18 and 32. But the flat plateau and low level reflects the large student population in Headingley. If a lot of women in their twenties are in full time education, they may choose to delay child births until they graduate and leave the area (Figure 4.5).

**Marital Status:** Apart from age, marital status also has an important impact on fertility. It can be seen as illustrated in Figure 4.5 that the fertility rates of married women are much higher than those of unmarried women. However, Interestingly, ONS (2007) points out that there has been a significant growth in birth outside marriage, because the greater acceptability of and rise in

cohabitation, though this is often just a stage between singleness and marriage. Over the last decade, the biggest increase in the proportion of births outside marriage has occurred among 25 to 29 year olds, from 30 per cent in 1996 to 40 per cent in 2006.



**Figure 4.5 Fertility rates by age of mother in two Leeds wards, 2001**

Source: Author's computation using ONS (2001g, h)

**Geographical variation:** Similar to the impact on mortality, geographical variation has been recognised in previous studies. Griffiths and Kirby (2000) studied the teenage conceptions under 18 and found substantial variations between home countries, regions and local authority areas. At the small area level, we can clearly see the geographical impact on fertility in Figure 4.5. Headingley has a much lower fertility rates than Cookridge, due to its large full time student population. Consequently it also has a much

lower fertility rates for women aged around 20. Armitage (1997) classified local authorities into 11 groups and 29 clusters of areas with similar socio-economic characteristics, using 1991 Census data. This study revealed that the local fertility rates reflect not only the socio-economic characteristics, but they also suggest important effects from education, occupation and location (urban/rural) factors. We have discussed how geographical variations can demonstrate impact of multiple factors in mortality factor section. Because of the difficulty in modelling all the factors, we use geographical variation to capture variations in the unobserved factors. There are two main difficulties - one practical, one theoretical. The practical difficulty is the work of estimating the relationship between the independent variable and fertility outcomes for local areas. The theoretical difficulty is constructing a model that combines the influence of a large set of variables. There are studies where these difficulties have been overcome, but combining all the knowledge in a model is a challenge.

#### 4.4.2 *Fertility process in simulation*

The Fertility process has been simulated in the same way as the Mortality process. However, in the fertility process, a female dominant approach is adopted, which is almost universally used in the fertility component of projection models. A female dominant fertility model is one in which only women contribute to the birth process. It is assumed that there will always be enough men around to ensure reproduction. The reason for this is that a complete record of each baby's mother is collected through birth registration whereas the record of the father is often incomplete. The relationships between the individuals, not only between mothers and babies, but also between them and other household members are successfully captured in the microsimulation.

Probabilities for the Fertility module have been generated using observed data. As with the Mortality module, a two stage probability generation procedure is adopted. First, a function of the age and marital status of the mother has been used to determine whether a woman gives birth during the

current time period of simulation. The process determines for each female at risk whether she is going to give birth before next year's simulation. If the woman does not give birth to baby(ies) this year, she moves on to the next module. Otherwise if the woman does give birth, then there is a need to determine if she has a single baby or twins (triplets and more are not considered in the current model). The new baby(ies) can then be added to the mother's household and a new record for each baby be added to the database. The Fertility module goes on to update the household and area information as in the Mortality module. A woman at risk gives birth if:

$$Ran(0,1) \leq P(b|a,m,l) \quad (\text{Equation 4.21})$$

where  $Ran$  is a random number between 0 and 1;  $a$  is the age of a woman;  $m$  is the marriage status and  $l$  is the location of the woman; and  $P(b|a,m,l)$  is the probability of a woman of age  $a$ , marriage status  $m$  and in location  $l$  giving birth to a baby/babies.

Twins are born if

$$Ran(0,1) \leq P(bt|b) \quad (\text{Equation 4.22})$$

where  $P(bt|b)$  is the probability of having a birth of twins, which is only available at the national level (ONS, 2001j). The sex of the baby is then determined by probabilities of male and female births (ONS, 2001g). The new individual's attributes are set as the following: age is zero, sex is determined probabilistically and the other attributes (e.g. location) are copied from the mother. In the next simulation period, the new individual is simulated along with the other individuals in the household.

This model does not differentiate between co-habitation and single households. Therefore it can overlook the enhanced stability of cohabiting couples with children. This may affect transitional probabilities that over time can lead to different household and individual populations such as in the fertility process. However, in fertility process, the birth statistics used to

generate the fertility probabilities are already based on women within or outside marriages. Therefore, fertility rates for births outside marriages are an average of both the single and cohabitated women (ONS, 2002). However, not differentiating cohabitating and single households can have a more serious impact on household formation process, as the long-term rise in cohabitation will continue and have an impact on household structures (Wilson and Stuchbury, 2010). Due to the time and resource limit on this project, this issue will be addressed in future versions of this model.

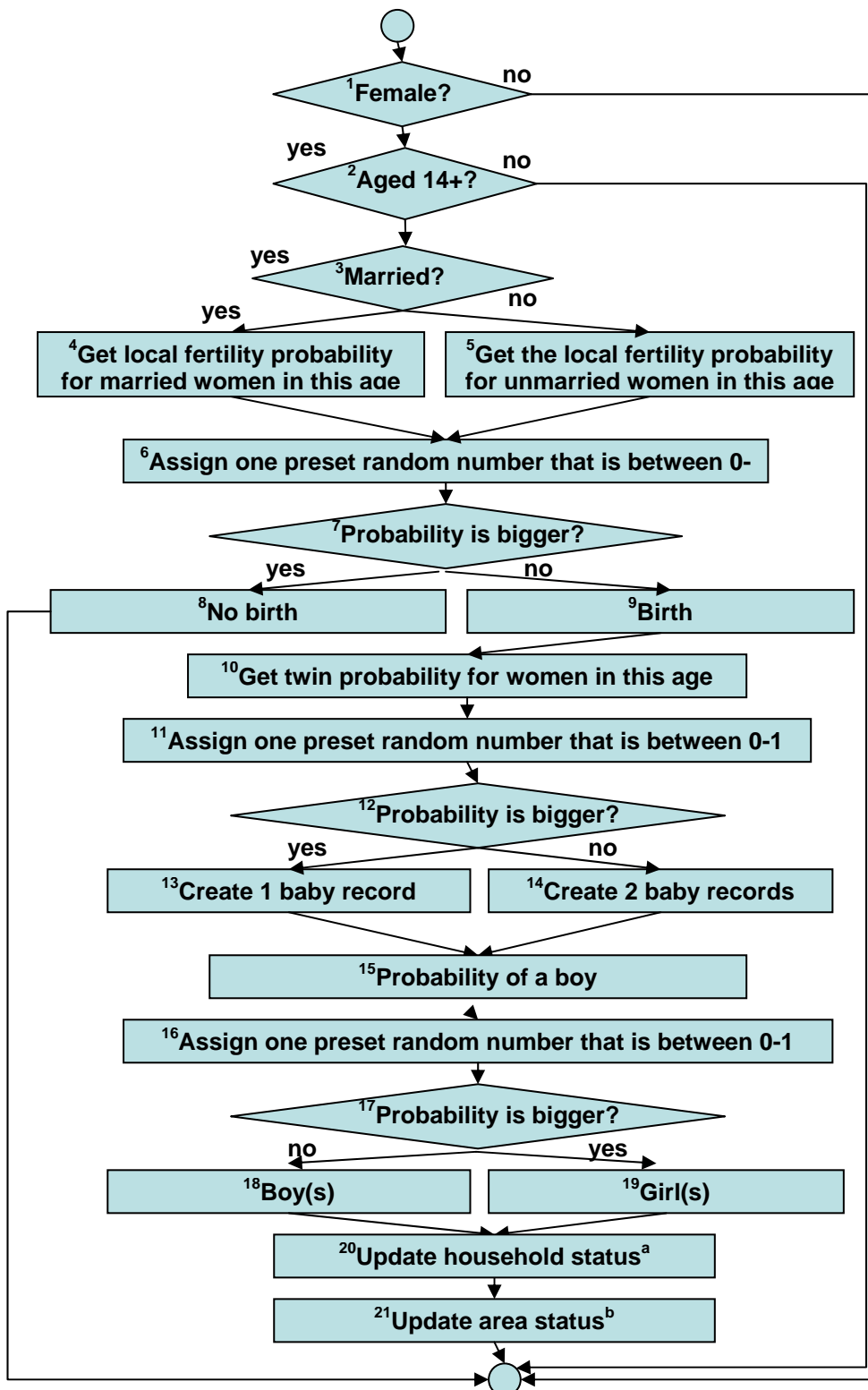
The main process of the Fertility simulation is illustrated in Figure 4.6.

Given the importance of geographical difference as described previously, localised probabilities by ward are used in this model to capture the local characteristics. The fertility probabilities, which are age, marital status and ward specific, are then used in the microsimulation. The next section describes the steps of the probability generation (Section 4.4.3).

### 4.4.3 *Fertility probability generation*

The fertility process requires the estimation of probabilities by age, location and marital status specific probabilities. As described before, a female-dominated approach is used in this study and the age range is limited to 15 to 45+ years old. The calculation steps are described below.

**Step 1:** Download the fertility data for England and Wales (E&W) for women both within (M) and outside marriages (U) from ONS (2002) and adjust age groups into single year until age 46 to get single year fertility rates  $f_x^{E\&WM}$  and  $f_x^{E\&WU}$ .



**Figure 4.6 Fertility process**

\* The pre-super-scripts in diagrams represent the simulation step and post-super-scripts are:

- a. household size, household member type and count, family roles
- b. total population in area, total households in area

**Step 2:** Compute ward fertility probability within marriages:

$$f_x^{wM} = \frac{B^w}{\sum_x f_x^{E\&WM} P_{xf}^{wM} + \sum_x f_x^{E\&WU} P_{xf}^{wU}} f_x^{E\&WM} \quad (\text{Equation 4.23})$$

where

$f_x^{E\&WM}$  is the single year of age fertility rate for women within marriages, based on the national data,

$B^w$  is the ward birth counts from Vital Statistics (ONS, 2001g)

$P_{xf}^{wM}$  and  $P_{xf}^{wU}$  are the populations of the married and unmarried women aged x in a ward

**Step 3:** To adjust the fertility rates to allow for women who die in the interval. As mortality is simulated before fertility, fertility rates are multiplied by an inverse survival probability:

$$\text{results} * f_x^{wM^1} = f_x^{wM} \frac{1}{({}_1S_x)^{1/2}} \quad (\text{Equation 4.24})$$

**Step 4:** Convert from period-age to period-cohort:

$$f_{x \rightarrow x+1}^{wM^1} = 1/2 f_x^{wM^1} + 1/2 f_{x+1}^{wM^1} \quad (\text{Equation 4.25})$$

**Step 5:** Compute ward based fertility probability for women outside marriage: This process is similar to Step 2, but we replace with data associated with women outside marriages (replace those variables with wM with wU).

**Step 6:** Download from ONS website Table PBH63A: Births: Maternities with multiple births: rates per 1,000 maternities, age of mother, 1938-2004,

a. all maternities (ONS, 2001j) and use this to generate the probabilities for twins.

**Step 7:** Again use the sex information in Table VS4: Vital Statistics for Wards 2001 - 2001 Boundaries to determine the sex of the babies (ONS, 2001g).

Following those steps 1-3, the period-age fertility rates  $f_x^{E&WM}$  and  $f_x^{E&WU}$  are converted into period-age probabilities, then the rest of the steps convert the period-age probabilities into period-cohort probabilities. Only the fertility rates are adjusted to allow for women who die in the interval and other probabilities are not adjusted, as the initial focus was the spatial populations and the three components of change of populations. As mortality only increases sharply at older ages and elderly are less likely to migrate, therefore the migration probabilities are not adjusted in this version of the model.

## 4.5 Health

Health is a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity (WHO, 1948). In this study, only the changes in the general health are simulated using the census data. In Census 2001, the general health is measured for populations by asking individuals to report their own health status. The information about this variable from the census code book (Individual Samples of Anonymised Records: ISAR) is provided in Figure 4.7 (ONS, 2006d).

### 1.31 health - General Health over last twelve months

<i>Value</i>	<i>Label</i>	<i>Percentage</i>
-9	Not applicable (student living away)	0.97%
1	Good	67.84%
2	Fairly good	21.95%
3	Not good	9.24%

This is a self-assessment of a person's general health over the 12 months before the Census. It was a new question for 2001 and was asked in all parts of the United Kingdom.

**Figure 4.7 The general health information in Census 2001 records**



Three categories are used to classify the general health status of people: Good, Fairly Good and Not Good. Individual health changes from “Good” to “Fairly Good” or from “Fairly Good” to “Not Good” are simulated in the model. Generally speaking, people’s general health deteriorates with age. With the ageing trend of UK population, health services need to know where the largest demand for health care arises. Therefore this model only focuses on the health deterioration process in the general health of ages 65 or older. There is also a simulation of moving such older people into formal care based on probabilities by age, sex and current health status. Although household structure is also important for this process, as elderly people living alone are more likely to be admitted than elderly living with others, it is difficult to calculate the two different sets of probabilities using available data. Only the most important factors are selected due to practical and theoretical difficulties discussed in Chapter 3. Also, this model only simulates the process of health deterioration, not the process of health improvement. This is mainly because of the purpose of facilitating explorations and planning of the ageing related researches and health service provision, but also because of the limitations of data, time and resources in this PhD research. Details of the modelling of the health change process are discussed in detail in the following sections.

#### *4.5.1 Factors*

Age, sex, socio-economic position, household structure, ethnicity, life style and environmental factors have all been considered for their impact on health and changes in health. However, age, sex and current health status are used. In the following section, the reasons of using the three factors will be discussed separately:

**Age:** Health risk varies systematically and significantly by age. Figure 4.8 illustrates the age impact on health status.

- There are very low rates for poor health from ages 0 to 10.

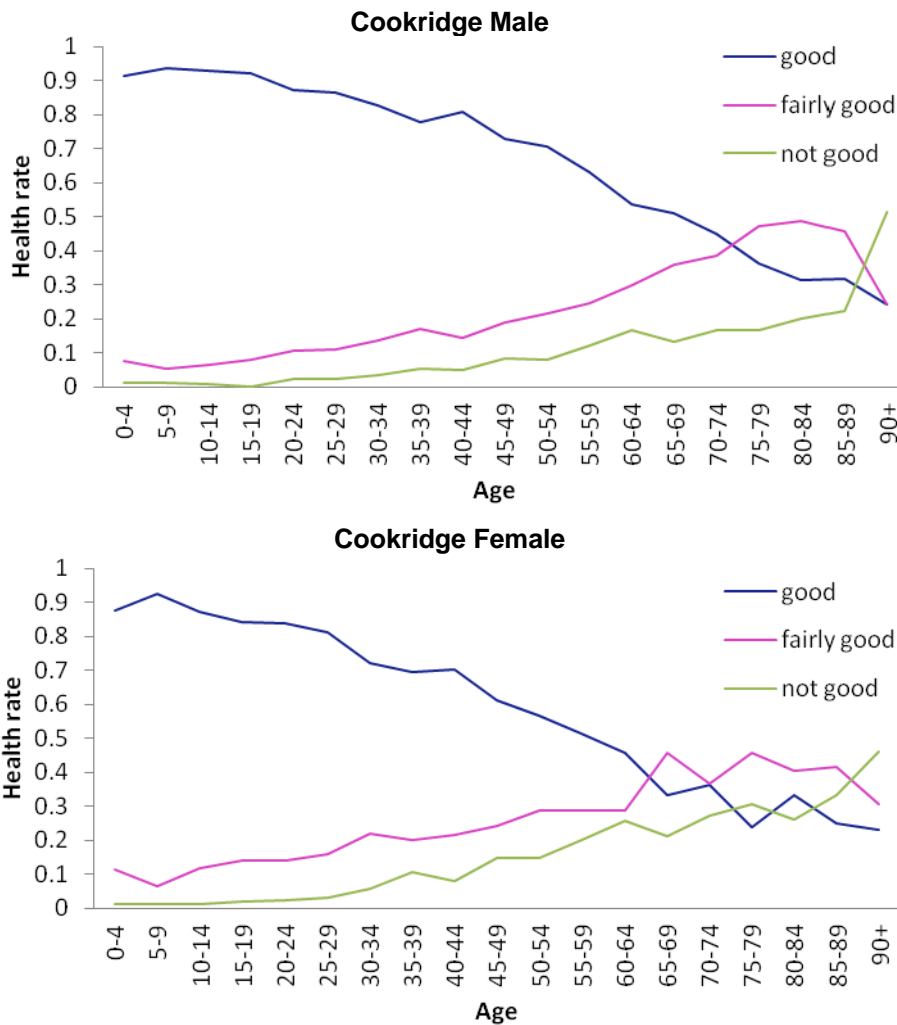
- Deterioration in general health starts in young people in their mid-teens to just before their 30s.
- Health improves slightly for people around age 40.
- Health generally starts to deteriorate faster for people aged around 60.
- There are more people in fairly good and not good health than in good health around age 75 and onwards.

As observed in Figure 4.8, people in their later stage of life generally tend to have health problems due to the ageing and may need more care than the rest of the population. This simulation of the health changes in this model therefore focuses on those who are aged 65 and over.

**Sex:** Sex is another important factor affecting health change. In developed countries, women generally live longer than men and men have greater susceptibility to life-threatening diseases. Also, men are more likely to be engaged in more dangerous occupations, as well as misbehaviours. They may smoke, drink and use drug more than women, too (Lopez 1983).

From Figure 4.8, the sex differences can also be seen in small areas. It is found that, in the old and frail ages (ages 80 and over), a substantially greater proportion of women are in fair or poor health than men.

**Current health status:** Often health history has a significant impact on changes in health. Normally people's general health gradually deteriorates over the time due to the ageing process and changes from the current health status to a worse health status. Therefore current health status has an important impact on the health changes. Based on the discussion above, it was decided that a combination of age, sex and current health was required to determine/project health changes.



**Figure 4.8 Health by age and sex in Leeds Ward, 2001**

Source: Author's computation using ONS (2001i)

#### 4.5.2 Health process in simulation

In this model the health change process is assumed to be a function of age, sex and current health status. An individual experiences a change in their general health if

$$Ran(0,1) \leq P(h_2|a, s, h_1) \quad (\text{Equation 4.26})$$

where  $Ran$  is a random number between 0 to 1,  $a$  is the age,  $s$  is the sex and  $h_1$  is the current general health status of the individual and  $P(h_2|a, s, h_1)$  is the probability that an individual of age  $a$ , sex  $s$  with a current general

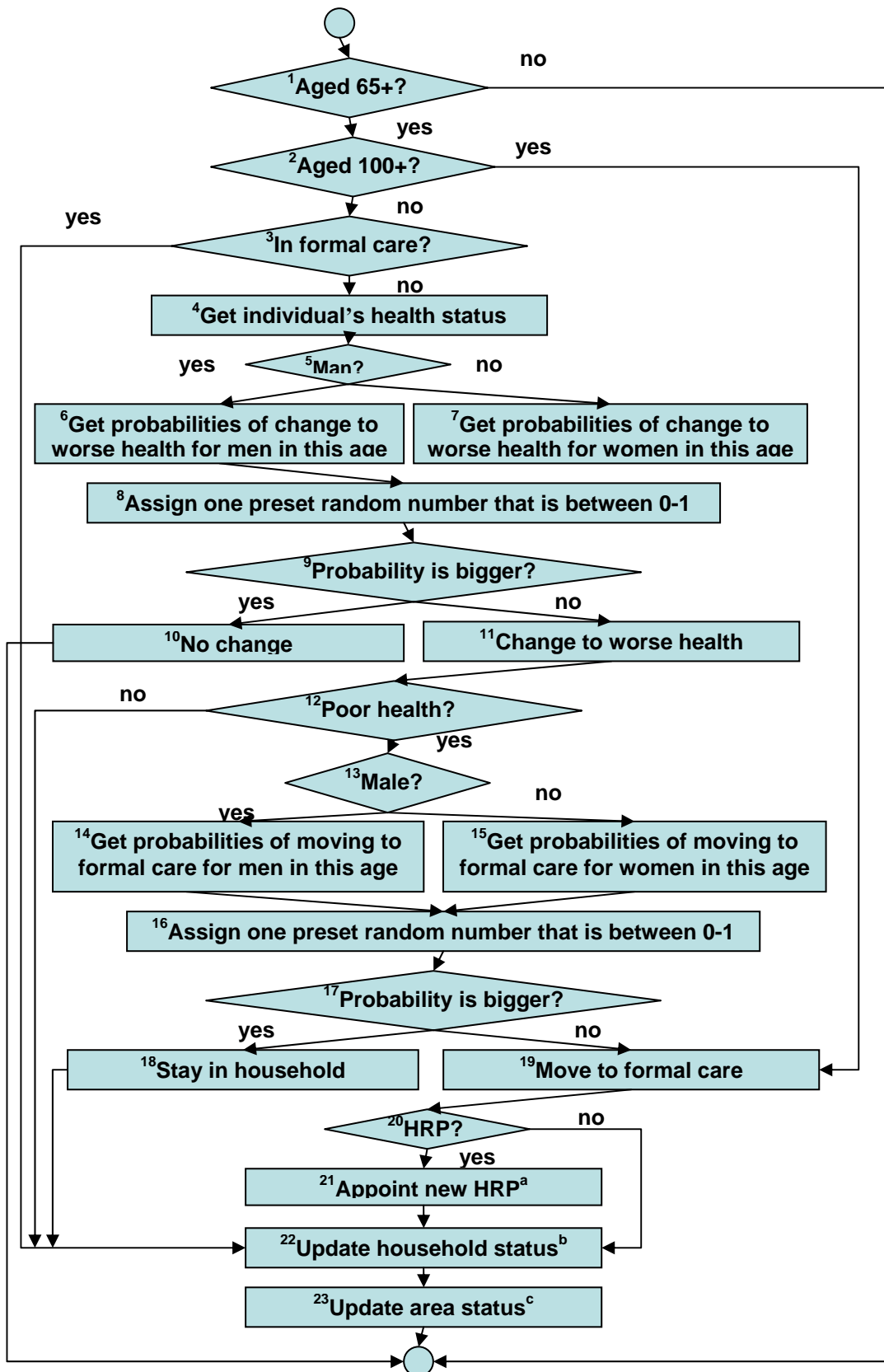
health status of  $h_1$  changes to health status  $h_2$ . The simulation of moving people aged over 65 to formal care is very similar to the process of health change, except  $h_2$  is the status of formal care and  $P(h_2|a,s,h_1)$  is the probability that an individual of age  $a$ , sex  $s$  with a current general health status of  $h_1$  changes to formal care status of  $h_2$ . Populations in formal care are moved from the normal household population to the Communal Establishment population and therefore do not participate in other normal population processes. If the person moved to formal care is from a lone household, the household is removed from the area and the Communal Establishment population gains one; if the person is living with others in a household, the household population is reduced by one in both the household and the area, while the Communal Establishment population gains one.

The simulation steps in health process have been illustrated in Figure 4.9.

### *4.5.3 Health probability generation*

Health steadily deteriorates with age, as Figure 4.8 shows. With the ageing trend of UK population, health services need to know where the largest demand for health care arises. Also due to the resource limit, this model only focuses on the health deterioration process in the general health for ages 65 and older. ISAR data from the 2001 Census are used to compute the probabilities for 5 year age bands (65-69, 70-74, 75-79, 80-84, 85-89, 90-94, 95+), by sex and health status.

This is a two stage modelling procedure: health change and formal care. Probabilities of health change, from “Good” to “Fairly Good” and from “Fairly Good” to “Not Good”, have been calculated using the synthetic cohort method, e.g. using the difference between the proportion of ages 65-69 with “Good” health and the proportion of ages 70-74 with “Fairly Good” to calculate the probability of health change from “Good” to “Fairly Good” for ages 65-69.



**Figure 4.9 Health change process**

\* the pre-super-script in diagrams represent simulation steps and the post-super-scripts are:

- a. the order to appoint as new HRP is: spouses, adult, elderly and children dependents
- b. household size, household member type and count, family roles etc.
- c. total population in area, total households in area etc.

The transition probability to move into formal care is calculated using the ISAR data. Probabilities of moving into formal care are calculated on the basis of data on those who are aged 65+ and dwell in a communal establishment of NHS psychiatric hospitals, NHS other hospitals, nursing homes, residential care homes and other medical and care establishment. The probabilities of moving into formal care are then calculated by dividing such populations by the total population aged 65+ by sex and current general health status. The probabilities can then be applied to those who are aged 65+ by respective current general health status and sex to determine whether they move into a formal care establishment or not. This use of stock data to estimate transitions is necessary because the required local transitions are not available. This model does not include the factor of household composition, but a lone elder is more likely to be “admitted” than elderly people living with support from other household members. However, as explained before, the model cannot consider all factors due to both theoretical and practical difficulties and only the most important factors are selected in modelling this transition (see Chapter 3).

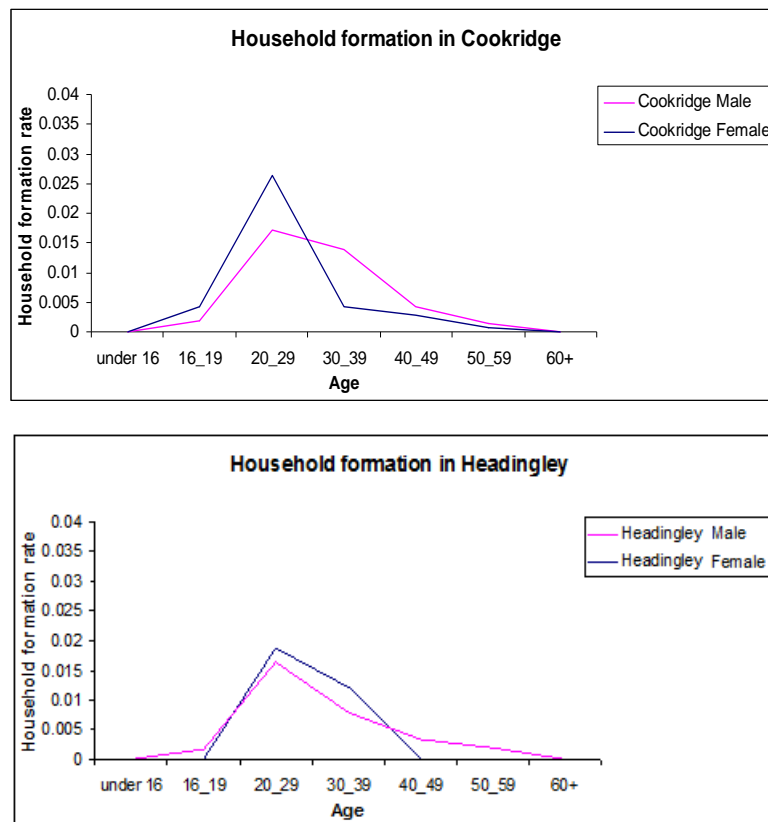
## 4.6 Household formation

### 4.6.1 *Factors*

There are various ways to form a household. The majority of households form through marriage, but there are also households formed through cohabitation or through students or other non-related people sharing accommodation. However, due to both the theoretical and practical issues as discussed in the mortality factor section, it is decided to use age, sex and geographical factors as the base attributes for this process. In the following section, the three factors will be discussed separately. Here we use marriage as the illustrative example. This model does not model cohabitation separately and individuals in the system are considered to be either married or single.

**Age:** Age is a vital factor in forming a household, even just from the legal point of view. The household formation pattern also varies systematically by age (Figure 4.10). From Figure 4.10, it can be seen that:

- the number of people getting married continuously increases from the age 16 and reaches the highest around the age of 30;
- the number of marriages then gradually decreases, by the age of 40 there is a substantial drop in the marriage rates and
- this decrease continues and there are very low rates of marriages after age 60.



**Figure 4.10 Household formations through marriages by age and sex in Leeds wards, 2001**

Source: author's computation using ONS (2001f)

**Sex:** Sex also has an important impact on household formation. Although both women and men have the highest marriage rates around age 20-29, women and men demonstrate different patterns in ages when they form their households. From Figure 4.10, it can be seen that:

- more women marry earlier than men: more women get married before the age of 30, but
- more men than women get married from age 45-60.

**Geographical variation:** As discussed in the Mortality and Fertility section, the age at marriage of people also varies geographically. Figure 4.10 also shows the variations within small areas in terms of the patterns of household formation through marriages. In the following sections, how the household formation process is simulated in the MSM will be described.

#### 4.6.2 *Household formation process in simulation*

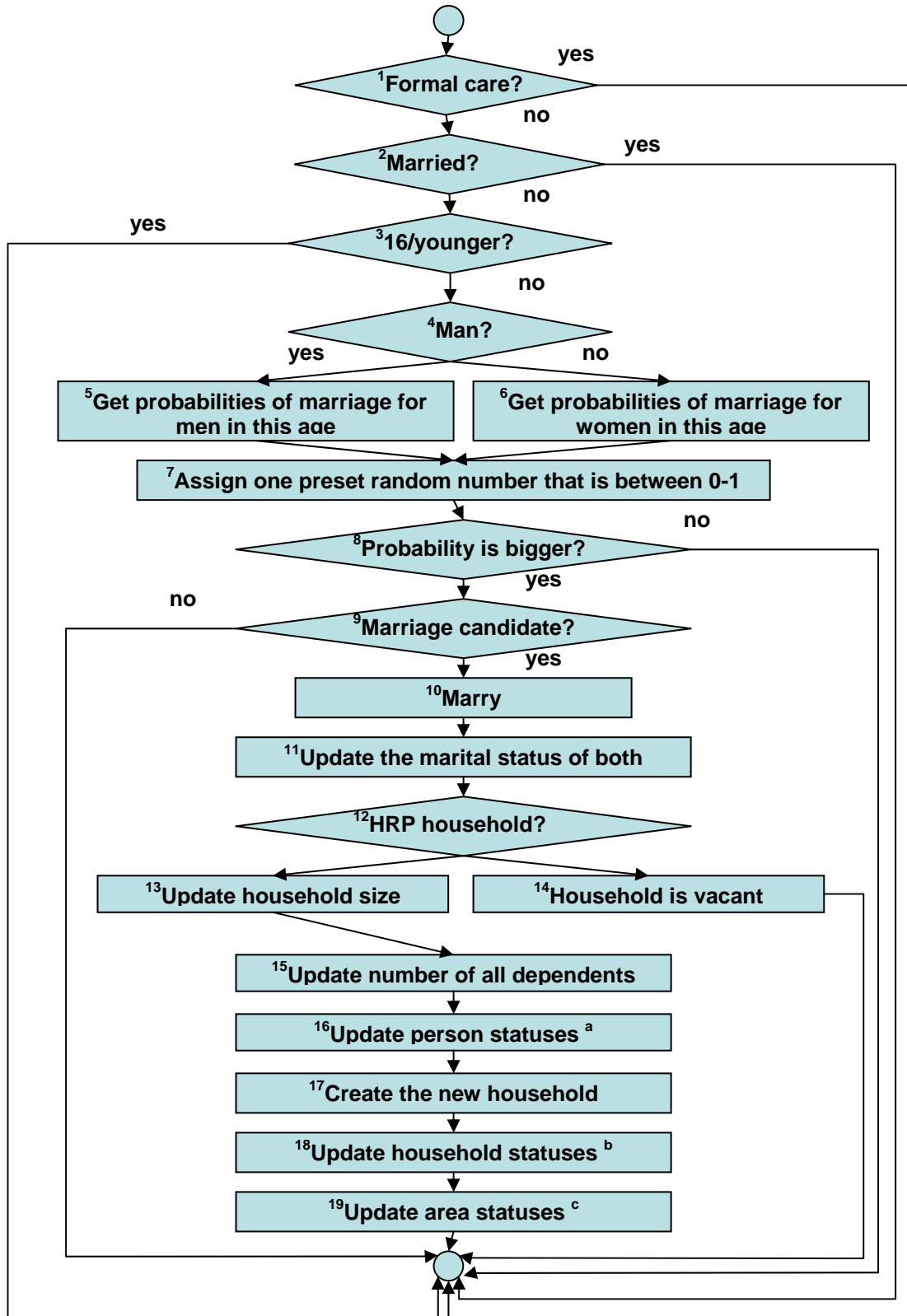
In this simplified process of household formation through marriage it is assumed that the household formation process is a function of age, sex and location. An individual gets married and forms a new household with the new spouse (and their existing household members if they are the HRP or Household Reference Person) if

$$Ran(0,1) \leq P(hf|a, s, l) \quad (\text{Equation 4.27})$$

where  $Ran(0,1)$  is a random number,  $a$  is the age,  $s$  is the sex and  $l$  is the location of the individual and  $P(hf|a, s, l)$  is the probability of household formation occurring to an individual of age  $a$ , sex  $s$  at location  $l$ . This model does not differentiate cohabitated population from the singles, so there is a limitation on individual household structure modelling. However, this will not have much impact on processes as reported in this thesis, which are modelled independent of household structure.

The simplified household formation process is illustrated using the marriage process as an example as in Figure 4.11.





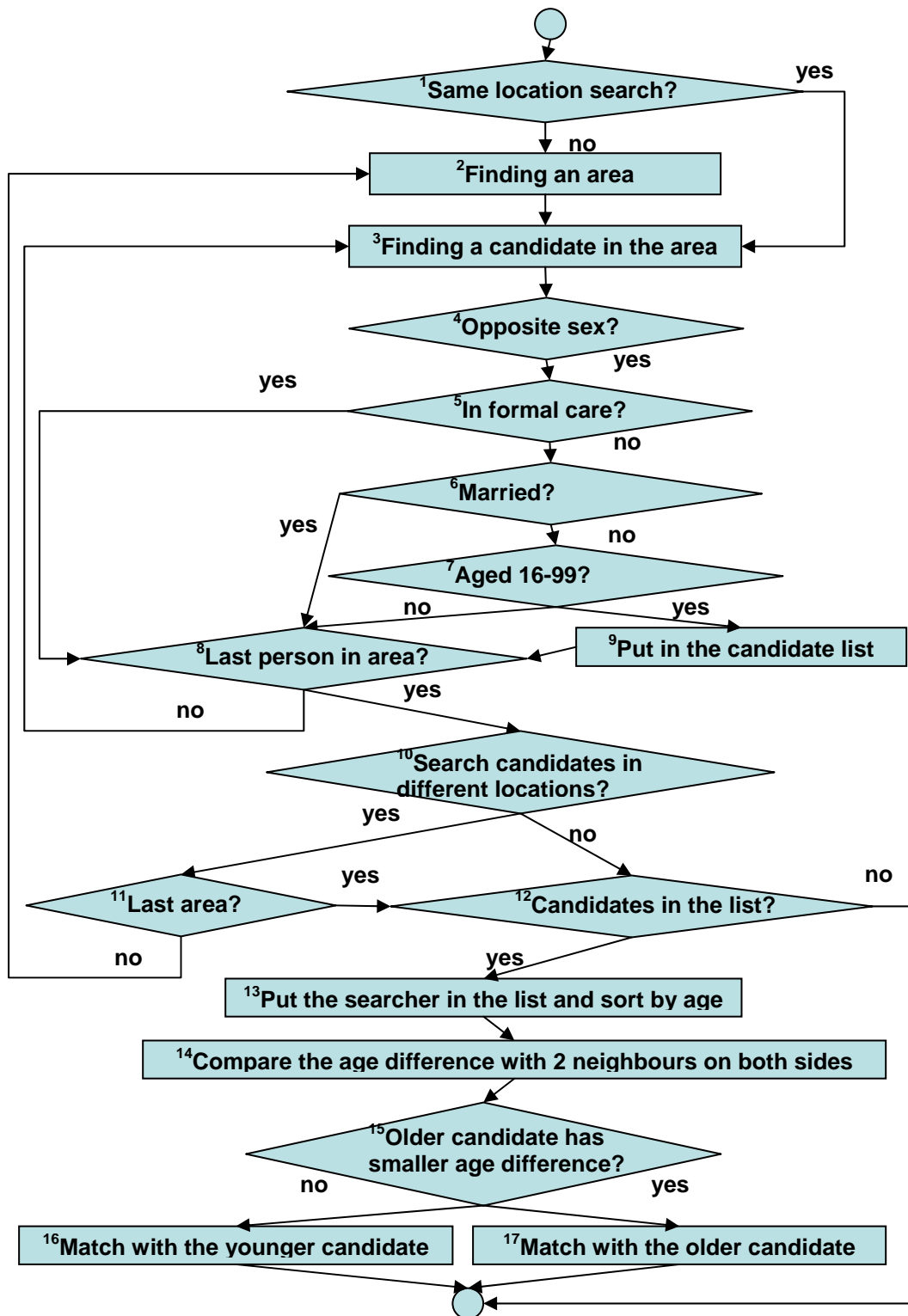
**Figure 4.11 The marriage and household formation process**

Notes: the pre-super-scripts in diagrams represent the simulation steps and the post-super-scripts are:

- a. HRP status, current and previous locations etc.
- b. household size, household member type and count, family roles etc.
- c. total population in area, total households in area etc.

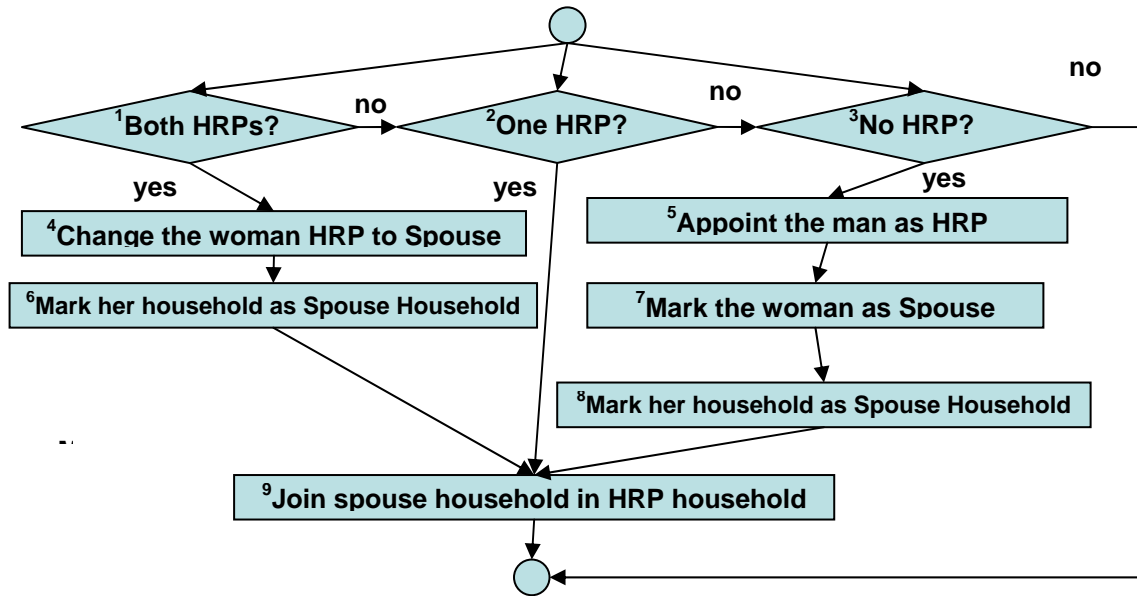
Other types of household formation can be generated through similar processes. First a set of marriage candidates needs to be generated based on a set of marriage probabilities by age, sex and location. Once an individual is available to get married, the program then tries to find a suitable candidate with the opposite sex and a suitable age (this model does not study the household formation of same-sex couples). The space within which a partner is sought is randomly decided to be either within the same location or another different location as the candidate, due to limited data. The matching of the ages of the husband and wife is based on probabilities calculated from the age distribution between the husband and wife from the commissioned table from ONS that are described in data section in Chapter 3. Based on this matrix of the single year ages of husbands and wives, probability of a man at a certain age to marry a woman at a certain age can be calculated. Five year age bands were used for practical reasons. Then the marriage candidates were sorted by age and sex into two lists and one candidate was matched to the closest aged candidate of the three from the search in another list. If a spouse cannot be found, the candidate remains unmarried and enters the marriage market again in next year (Figure 4.12).

In both “marrying” and “after marriage” processes, various HRP (Household Reference Person) scenarios need to be considered for each couple: one is the HRP, both are HRPs and neither is an HRP. The first case is straightforward and the no one’s HRP status will be affected. In the second scenario, the wife’s status will be changed to spouse and the husband will be appointed as the HRP (Figure 4.13). In the model, children always follow their current HRP. Therefore there are 3 scenarios for children upon marriage: no changes to children of a non-HRP, they stay with the current HRP and only the non-HRP moves out; no changes to children of an HRP, while new household member(s) join in and adopt(s) the current household statuses; in the case that children of the current HRP follow he/she to join a new household (when two HRPs get married), all members of this household update their household statuses according to the new household information.



**Figure 4.12 Matching candidates in marriage process**

\*Notes: the pre-superscripts in diagrams represent the simulation step numbers



**Figure 4.13 HRP status updating**

Notes: the pre-super-scripts in diagrams represent the simulation step numbers

### *4.6.3 Household formation probability generation*

As described before, the household formation is a multiple staged process and two sets of probabilities are required. First, the household formation probabilities need to be calculated. Within each small area, the total number of people by age at marriage, sex and previous marital status and the total population by age and sex are used to calculate the probability of marriage by age and sex. Such information comes from ONS data at sub-national level, as there is no suitable data available at finer geographical scales. Such probabilities are then applied by age and sex to populations aged over 16 who are single. This generates a list of marriage candidates.

Then the second stage is to match the marriage candidates together. The distribution of the age difference was analysed, using a commissioned table from ONS (Data details are discussed in Chapter 3). Based on the analysis, the probabilities of age difference between husband and wife was developed. Although the ages are single ages in the source data, it is felt that the single year probabilities will require a lot of computing resource, especially when the household formation process requires a lot of pre and

post processing before and after the event. Age difference probabilities in much cruder bands are decided to be sufficient for this model and calculated sex specific probabilities of husband and wife age difference in 1 year, 2, 3, 4 and 5+ years' bands only.

## 4.7 Migration

The Migration process is much more complex than the others. There are various reasons for people to move. As people are moving from one place to another in this dynamic spatial MSM, the spatial framework is especially important in this process. Migration often has a dependency on other processes. For instance, many household formation processes consequently lead to migration. It is difficult to model also because of the knowledge gaps in both understanding and data availability. It is also felt that modelling the heterogeneous behaviours exhibited in the migration process in a pure MSM is challenging. Therefore different techniques and probabilities are used in different types and stages to model the migration process. This section will only discuss the simplified general migration process within Leeds. The further development of the migration module will be described in Chapter 5 and 6.

### 4.7.1 *Factors*

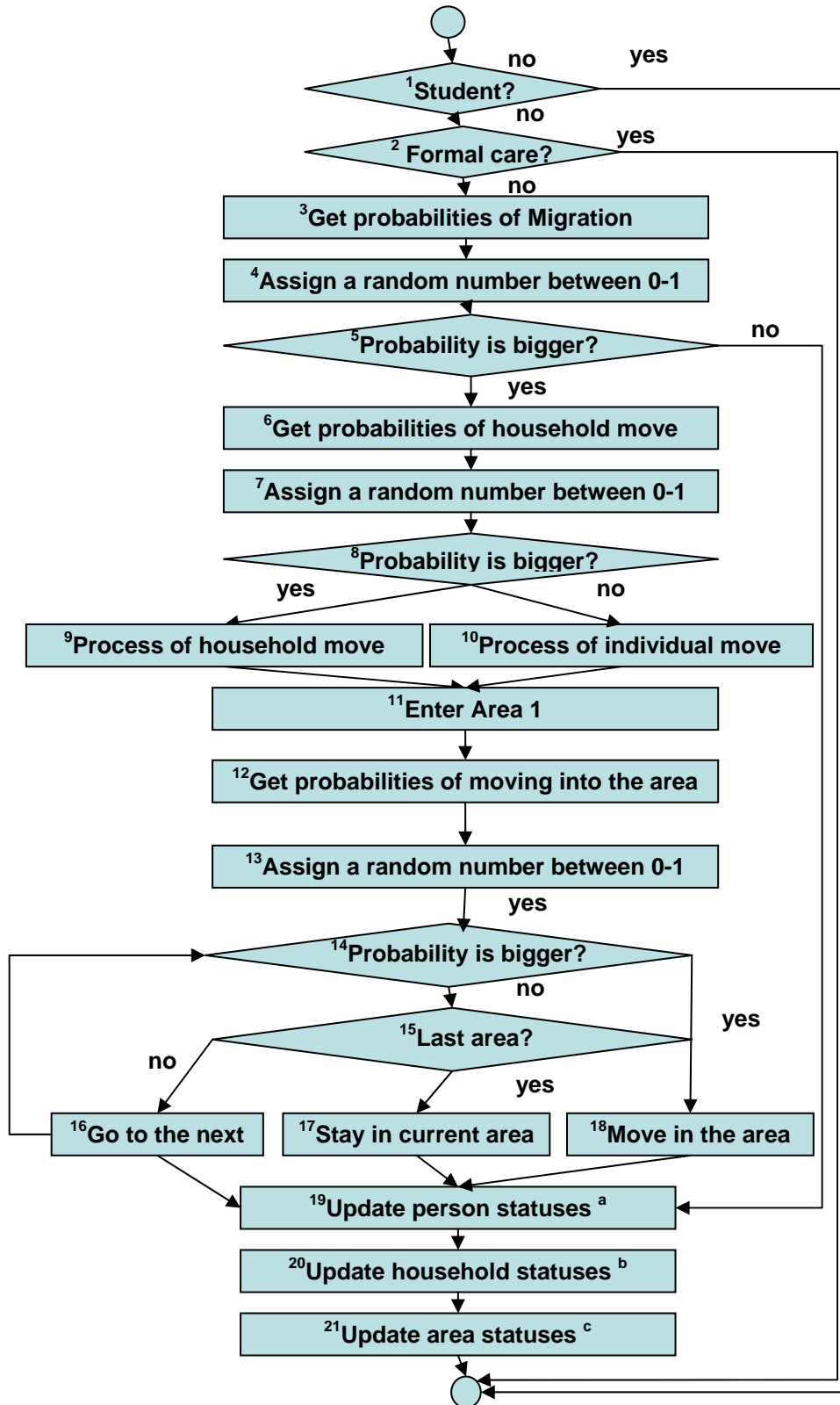
As discussed above, migration is a very complex process and it requires multiple stage modelling. Various factors have been considered in terms of impact on household move and individual move, including: age, dwelling type, household size, tenure, occupation of the HRP, marital status of the HRP, health of the HRP, ethnicity and age of HRP. Age, dwelling type and household size are found to have important impact on household moves, while age, household size and marital status are important for individual moves. However, due to the data unavailability, we have to use information available in ISAR to calculate the probabilities. The migration probability calculation process is described in details in the section 4.7.3.

### 4.7.2 *Migration process in simulation*

The multiple staged migration process requires the modelling of who is to move, how they move (i.e. as a household move or as an individual move) and where they will move to. Student migrants are modelled in a separate agent based model and those already in formal care are not considered in the migration process. The simplified migration process is illustrated in Figure 4.14.

### 4.7.3 *Migration probability generation*

As described in Section 3.4.3, migration probabilities are derived from event counts and counts of population at risk. As described in section 4.7.1, HSAR and BHPS data were explored to select important factors that determine whether a whole household moves or just an individual moves. However, due to the difficulty of calculating such probabilities and handle the computing efficiently, migration rates (including in-flow, out-flow and immigration information) from the SMS (Special Migration Statistics) data were used to calculate the ward-based migration probabilities. The average national emigration rates were calculated using IPS (International Passenger Survey) data due to the unavailability of microdata. The household migration indicator (wholly moving and partly moving) information from the HSAR data is used to calculate the household move rates and the rates are applied to individuals and their household members to generate a list of household migrants. Then probabilities for the individual migrations are adjusted accordingly and applied to the rest of the individual migrants to avoid double counting.



**Figure 4.14 Migration process**

Notes: Student migration is modelled separately. This will be discussed in the next chapter.

The pre-super-script in diagrams represent simulation steps and the post-super-scripts are:

- a. HRP status (see Figure 4.17), current and previous locations etc.
- b. household size, household member type and count, family roles etc.
- c. total population in area, total households in area etc.

To determine the destinations of the migrants, the SMS data are used to calculate the probabilities to move from a specific location to another. As there is no emigration information in SMS, the IPS information on country of visit/residence and intended length of stay is used to calculate the probabilities of moving out of UK. Migrants enter small areas according to the location based probabilities. If the search reaches the last area and is unsuccessful, the migrant remains a migration candidate and start the same process again in the next year.

## 4.8 Initial result analyses

The population is simulated through various demographic processes each year from 2001 to 2031. In the analysis of the results we are free to select any of those years for report and use different years as appropriate. This demonstrates the flexibility of annual interval projections over five year interval projections. The main findings from the initial analyses are described in this section. The strengths and limitations of the model will be discussed at the aggregate level of LA and the small area level of wards. Results are also analysed by demographic changes to provide an insight into the components of change. Possible causes for the limitations are explored and further developments to address such issues are proposed at the end of the section.

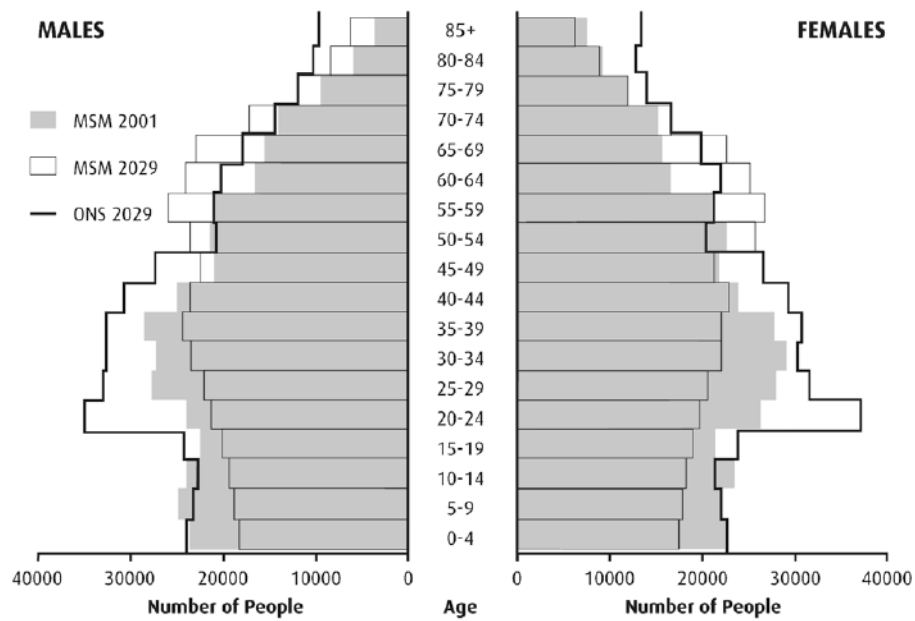
### *4.8.1 Initial result analysis at aggregate level of LA*

Using the dynamic microsimulation, the model can provide projections of an individual based local population with dynamics that truly reflect the demographic changes in a way that they would happen in the real world. The individual changes come from the dynamic updating and interaction of individual attributes from each individual at each simulation step, rather than a statistical procedure of applying general weights statically. Therefore the dynamics of the studied population each year are better represented than in a static MSM.



As this model is also a spatial MSM, in which the populations of a set of small areas (wards) are simulated, we can study the spatial variance in small area populations and population changes within one LA. However, the simulation results can be aggregated to a higher level to provide indications of various population trends to facilitate strategic decisions or to explore different scenarios. For instance, population changes over time such as the ageing pattern can be clearly observed from population pyramids generated by the MSM (microsimulation model) to facilitate high level decision making. At the Local Authority (LA) level of Leeds, population pyramids have been produced to illustrate the age-sex distribution of Leeds population in 2001 and 2029.

The migration process within Leeds is modelled using the 2001 Census Special Migration Statistics (Level 2) data, which provides us the migration flows from one ward to another. Having incorporated the inter-ward migration impact in the MSM, the population changes in Leeds have been analysed. As illustrated in Figure 4.15, we can clearly see the trend of ageing in Leeds population. However, it seems that the current MSM results exaggerate the ageing trend in Leeds population and the whole population in Leeds is shrinking, compared with the ONS (2007b) projection for 2029. This is mainly caused by modelling the area as a closed system, but notice that some studies have pointed out that ONS has radically over-estimated immigration to Yorkshire and Humberside and to Leeds in particular (Boden and Rees, 2010).



**Figure 4.15 The age–sex structure of the Leeds population in 2001 and 2029 generated by the MSM compared with the ONS subnational projections in 2029**

Source: Wu *et al.* (2008)

There are three important sets of differences between the projections. The current MSM assumes that age-specific mortality rates will stay constant over the next three decades, whereas ONS expects a steady improvement in life expectancy. Hence the elderly populations are somewhat larger in the ONS projection. This MSM also assumes constant fertility rates, whereas ONS does not. Finally, the migration process currently only concerns the closed system within Leeds (inter-ward migration flows in Leeds only as described above). While in reality, the inward migration, especially student migration to Leeds, mostly in the 16–29 age range, is substantially greater. This has been reflected in the ONS projections and the projections allow for acceleration in student numbers following the 2001 Census. Such issues in the current model obviously need to be addressed. In Chapters 5 and 6, the details of further development of the model will be discussed in the attempt to address various issues.

However, even at the aggregated level, we can see the potential of the aggregated results from the model. They can clearly demonstrate the trend of the population changes year by year. If we look into the individual details, we can even trace various demographic changes of certain cohorts,

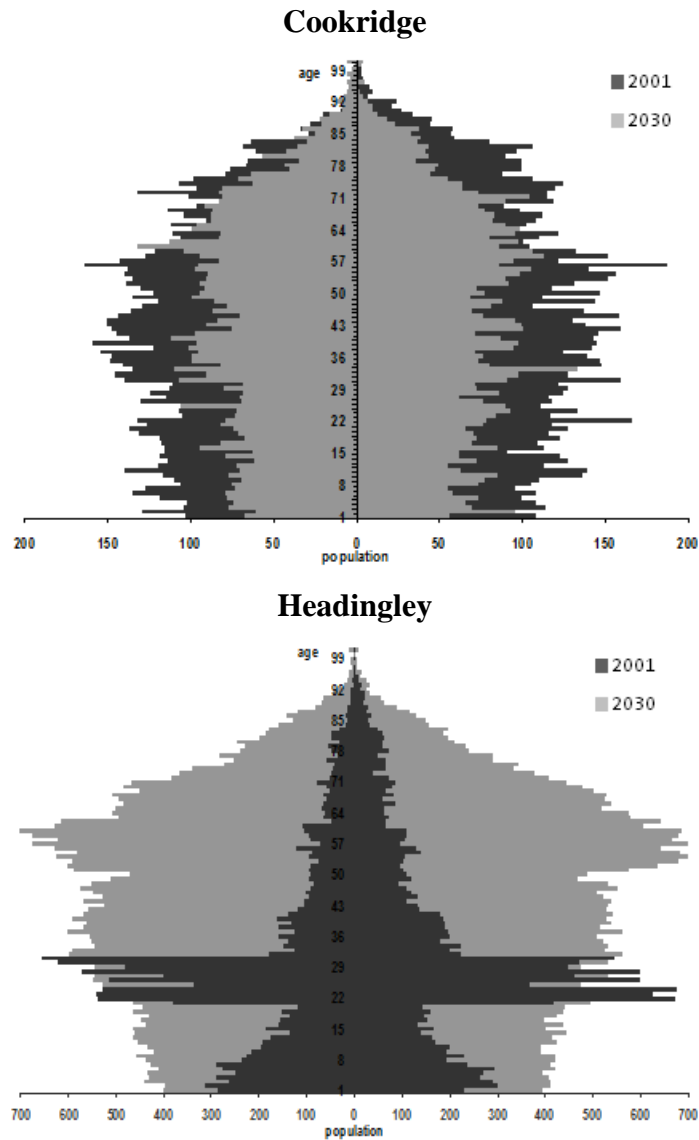
groups or even individuals over a period, for instance, changes of the general health statuses. As all changes are dynamically simulated each year and driven by multi criteria based probabilities, including local area factors, the overall results should present a much more robust representation of the studied population, compared to static models that move population forward through general reweighting procedures. Therefore, this model is potentially more useful for assisting strategic decision making at the aggregate spatial level.

#### *4.8.2 Initial result analysis at small area level of wards*

The aggregated results of the MSM are useful in terms of facilitating the strategic planning or policy making. However, at the tactical level, local context plays an indispensable role in area based intervention measures. Therefore, this study assesses the initial results by small areas to explore the spatial differences.

##### *4.8.2.1 Population pyramids for small areas*

The simulation results over 30 years for all 33 wards have been analysed and a substantial difference is revealed between the small area projections of Cookridge and Headingley. In Figure 4.16, the population changes over time in wards have been presented in population pyramids: Pyramids in darker shade use the simulation results in year 2001 and pyramids in lighter shade use the simulation results in year 2030. Results from any year between 2001 and 2031 can be used in this analysis. Here results in 2030 are used to demonstrate the changes close to the end of the simulation.



**Figure 4.16 The age–sex structure of ward population: Cookridge and Headingley in 2001 and 2030**

Source: Author’s computation using results from this version of model

From the small area population pyramids illustrated in Figure 4.16, we can clearly see that characteristics of the local population evolve differently in small areas. In the projections in 2030, we can see the ageing trend of the local populations. The majority of the population in the ward is aged over 40 and there is a significant increase of the ages over 90. However, Cookridge seems to have a more serious problem than Headingley, where the population has stopped growing and there is a much smaller population in Cookridge in 2030 than in 2001.

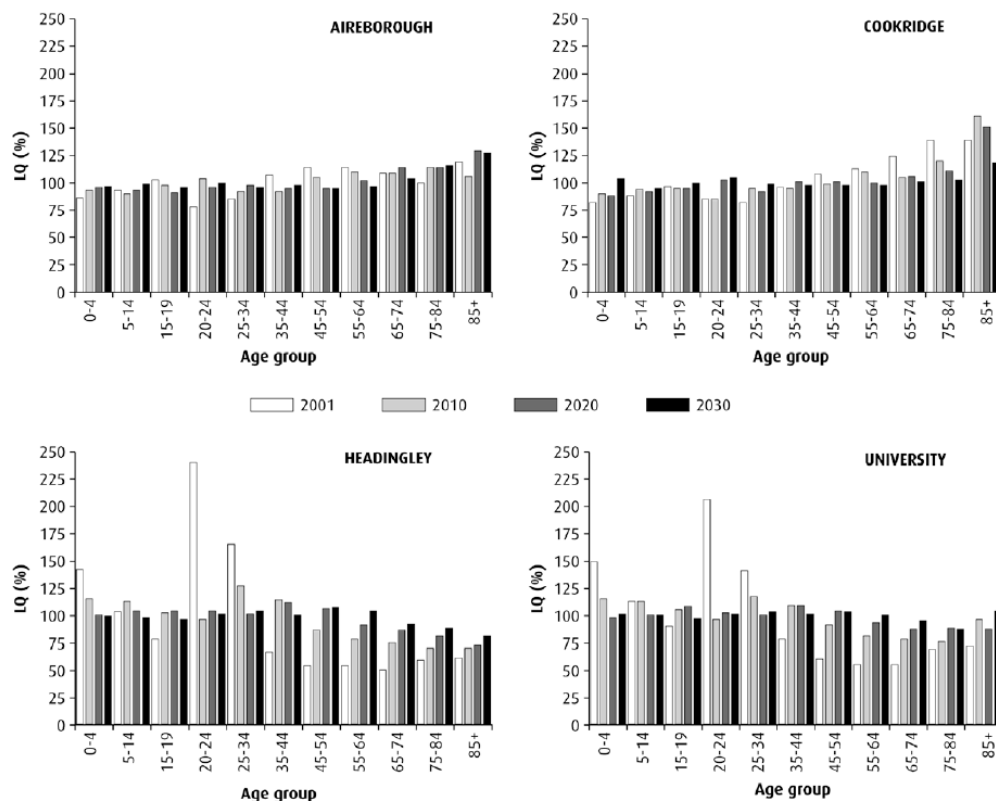
The population in Headingley, on the other hand, keeps growing into a

much larger population. The main reason behind this may be that Headingley has a much larger number of younger population aged 20-30 and under 10 in 2001 (Figure 4.16). Such composition allows the Headingley population keep growing. However, it needs to be noted that this version of MSM models Leeds population within a closed system and there are no housing stock constrains for the areas. Therefore the MSM allows university students to stay and grow old in area after higher education. This has exaggerated the Headingley population growth and the effect of ageing in both wards. Omitting migration to and from the Rest of the UK outside Leeds and to and from the Rest of the World outside the UK is the main limitation, but other reasons are also possible. Such reasons and various issues with the initial model will be discussed and addressed in Chapter 5, 6 and 7.

#### 4.8.2.2 *Location Quotient Analysis in small areas*

To explore more of these spatial differences and in a wider context, a Location Quotients (LQ) analysis has been used to assess the spatial variance in the local populations. The LQ technique is one of the most commonly utilised Economic Base Analysis methods (Wikipedia, 2010). The LQ technique is frequently used in locational analysis, economic geography, and population geography, but it has been applied to a much wider range of studies. The LQ technique compares a certain local characteristic to a reference characteristic, in the process attempting to identify specialisations in the local area. The location quotient technique is based upon a calculated ratio between the local and the reference unit. Generally speaking, the LQ is an index for comparing an area's share of a particular characteristic, e.g. industry concentration, demographic features, with the area's share of some basic or aggregate phenomenon. It therefore provides a way of quantifying how concentrated a particular industry, cluster, occupation, or demographic group is in a region as compared to the nation. It can reveal what makes a particular region “unique” in comparison to the national average and is useful for calculating and mapping relative distributions (Desai *et al.*, 2009; Wikipedia, 2010).

The initial results from a selection of area groups that have similar characteristics have been assessed using LQs as measures of the concentration of a particular group within each geographical area at a point of time. The population structure is analysed in five year age bands. Where the location quotient is greater than 100, this indicates an over-representation of the population group in area, and vice versa. Some examples are shown in Figure 4.17.



**Figure 4.17 Location quotient analyses of ward level projections**  
Source: Wu *et al.* (2008)

The pattern of changes has been found to be consistent in two types of small areas. Here results from four wards are presented: Aireborough and Cookridge represent the suburban areas and Headingley and University to represent city centre student accommodation areas. Aireborough and Cookridge are both established suburban areas in the northwest of the city. In these areas, the changing concentration of demographic groups over time looks reasonable. For example, in Cookridge there is initially an expansion in the very elderly population, but later the quotient for the elderly falls relative to the rest of Leeds as other areas begin to experience a similar

growth in the older age groups. However in wards where student migration has a great impact like University and Headingley, the MSM failed to reproduce the student population renewal. The peaks of young people aged 20–25 disappears after 10, 20 and 30 years simulation. The distinctive population structure in such areas has been lost in the small area projections.

This indicates that the subtlety of the local migration patterns has not been captured successfully in our MSM (microsimulation model). The pure spatial MSM used here cannot differentiate students from other migrants in the migration process. The MSM is probability driven. Migration probabilities at small area scale have been generated using the 2001 Census Special Migration Statistics (Level 2) data. The data provides us the ward-based migration flows from one ward to another for migrants, but we cannot determine the number of the university students within each flow. Therefore student migrants are modelled exactly the same as the rest of the migrants in the pure MSM using the generalised probabilities.

Another factor affecting the results could be that the small area migration probabilities are not disaggregated by age. However, even if we refine the model with disaggregated ages, it still cannot capture the distinctive migration pattern of student migrants. In reality students only move around the area close to the universities where they study, not in the suburban areas. More importantly, most of them will leave the city once they finish their study, instead of settling down and growing old in the area. Due to the replenishment of the student population each year, the population of the wards in which university student stay tends to remain younger than that in other wards. Another issue with student migrants is that they often confuse their answers to term-time and home address in the census. Chapter 6 will discuss such issues in detail.

Given the above reasons, a hybrid modelling approach is adopted to better model the local migration patterns. In chapter 6, a hybrid approach combining MSM and ABM (agent-based model) techniques will be explored.

### 4.8.3 *Initial result analysis by demographic process*

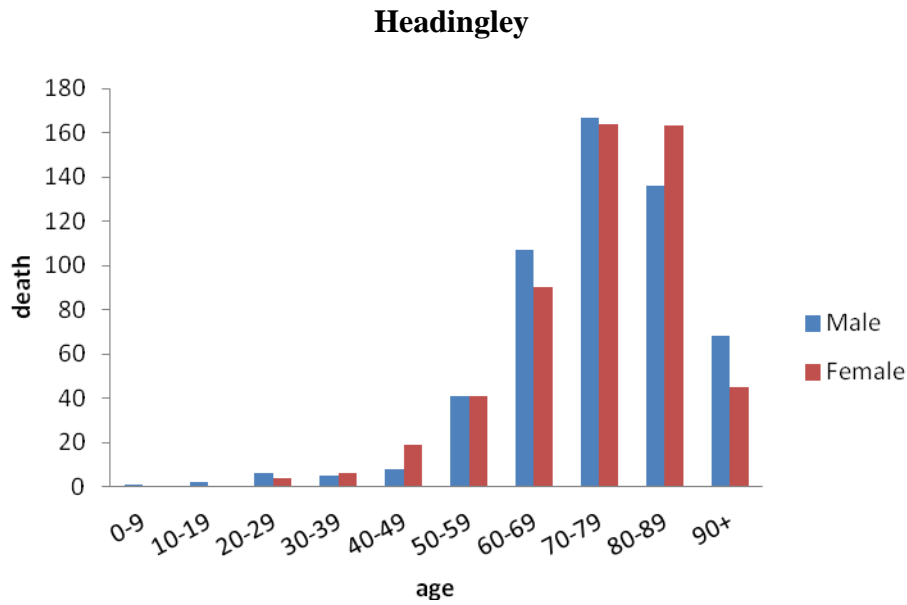
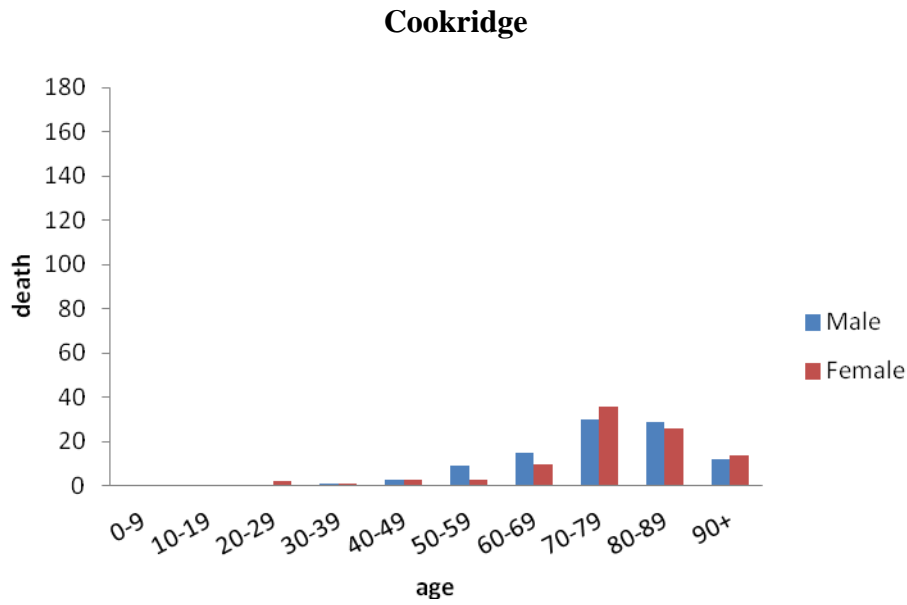
Spatial variances in small areas are also found in the different demographic processes. Two wards, Headingley and Cookridge, have been selected to demonstrate such variances in the small areas in Leeds. Headingley is a ward in the city centre that is popular for the university student accommodation, while Cookridge is a more established suburban area. So the two local populations differ considerably. The two wards will be used as examples throughout the demographic process analyses.

#### 4.8.3.1 *Initial Mortality result analysis*

The simulation results from the Mortality module have been illustrated in Figure 4.18. The current model uses one set of survival probabilities by age, sex and location, therefore the mortality rates do not change over time. This limitation is addressed in further developments of the model, which are described in chapter 5.

The initial results for year 2030 suggest that mortality changes systematically with the age. Both small area populations demonstrate a smaller number of deaths in younger ages. In fact, there is no mortality for ages under 20 in Cookridge and there are only a small number of deaths in Headingley. Mortality starts to increase gradually with age from age 40 onwards and reached the peak around the ages 70-79 and then gradually reduces. However, we have to take into consideration that there are very small numbers of population in the very old ages such as ages 90+. Females in most ages experience fewer deaths than males, except those in ages 70-79 and 90+ in Cookridge and those in ages 40-49 and 80-89 in Headingley.



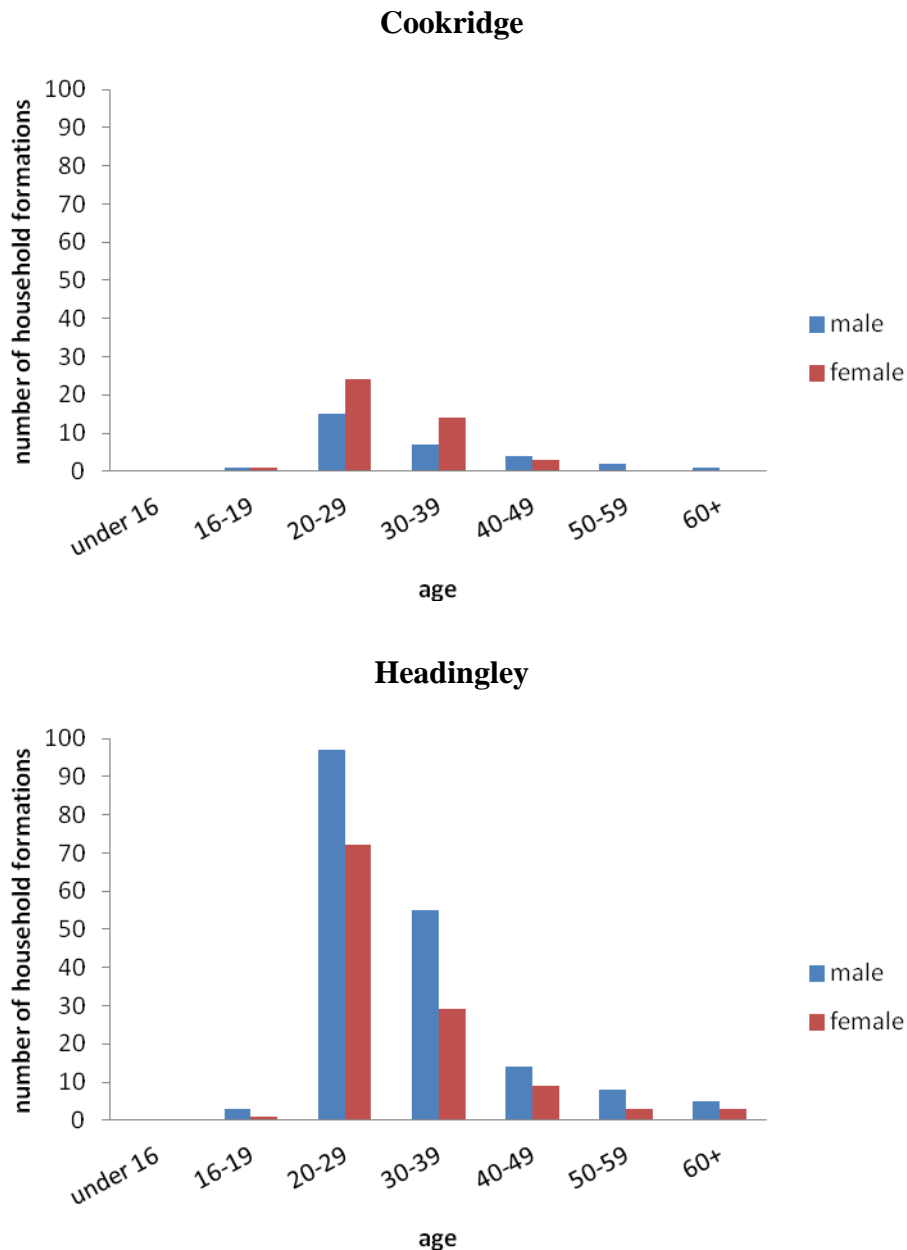


**Figure 4.18 Ward level projections of mortality in 2030**

Overall, Headingley has a larger number of deaths in almost all age bands than Cookridge. This is not surprising, as Headingley also has a bigger base population than Cookridge, as analysed in previous sections. The number of deaths for ages 50 and over in Headingley is substantially higher than that in Cookridge (Figure 4.18).

#### 4.8.3.2 Initial Household Formation result analysis

In the Household Formation module, the household formation probabilities are assumed to be constant again from year 2001. Figure 4.19 shows the household formation projections for Headingley and Cookridge.



**Figure 4.19 Ward level projections of household formation in 2030**

The simulation results in year 2030 show that there are more household formations in Headingley than in Cookridge, due to the larger population in Headingley. However, the initial result analysis indicates that people in both

Headingley and Cookridge are most likely to form a household between the ages of 20 to 30. From the age of 16, a small number of people start to form their households. Such numbers reached a peak at ages 20-29. Then the numbers gradually decrease in both areas and there are very few people that form their households beyond the age 60.

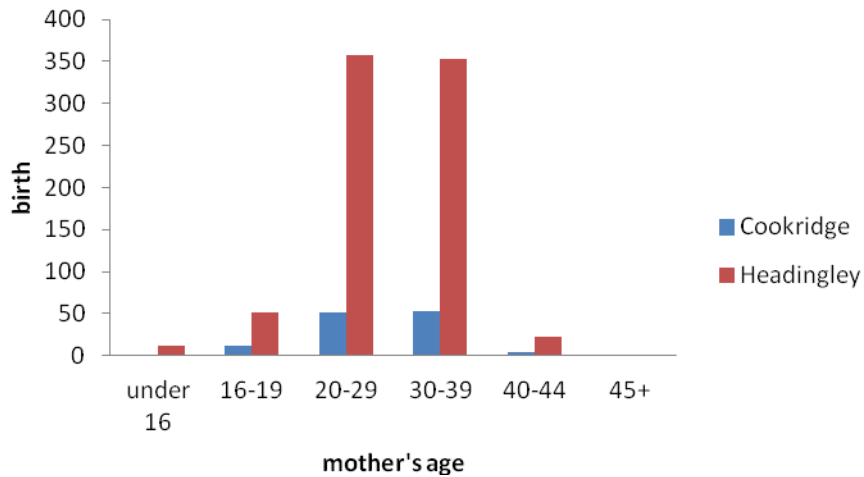
From the ages over 60, the numbers of household formations in both wards are smaller. Compared to Headingley, there are fewer males than females in ages 16-39 that form households in Cookridge; while the contrary relationship holds in Headingley. From ages 40 onwards, there are more males forming households than females in both wards. Although there is a substantial drop in the household formation from the ages 40-49 for both sexes, the reduction in the number of females forming households over ages 40 are significantly smaller than males. In Cookridge, there are hardly any females forming households from the ages 50 onwards.

#### *4.8.3.3 Initial Fertility result analysis*

In the Fertility module, the fertility probabilities generated on the basis of information from year 2001 are applied over time to the women at risk. The simulated fertility results of year 2030 in small areas have been analysed. The analysis of results from Cookridge and Headingley in Figure 4.20 indicates differences in births in the two small areas.

In Figure 4.20, fertility projections in year 2030 show that few females in Headingley give birth to babies at ages under 16 (ages 12-15 in this model); while there are no births in Cookridge until the age 16. Then the fertility increases with age and the numbers of births in both Cookridge and Headingley are found in the age groups of mothers from 20 to 39. Then the fertility drops substantially for females in both areas from the age 40 onwards. There are no births for females aged over 45 in both areas. Such patterns are consistent with the national projection. Overall, Headingley has more births than Cookridge, especially to women aged 20 to 39. This is partly caused by the large projected population in Headingley, as there are many young women ages 16-29 in this area. However, most university

students are unlikely to give birth to babies in reality. This indicates the limitation in the initial model results and this will be addressed in model refinement, which will be discussed in Chapters 5 and 6.



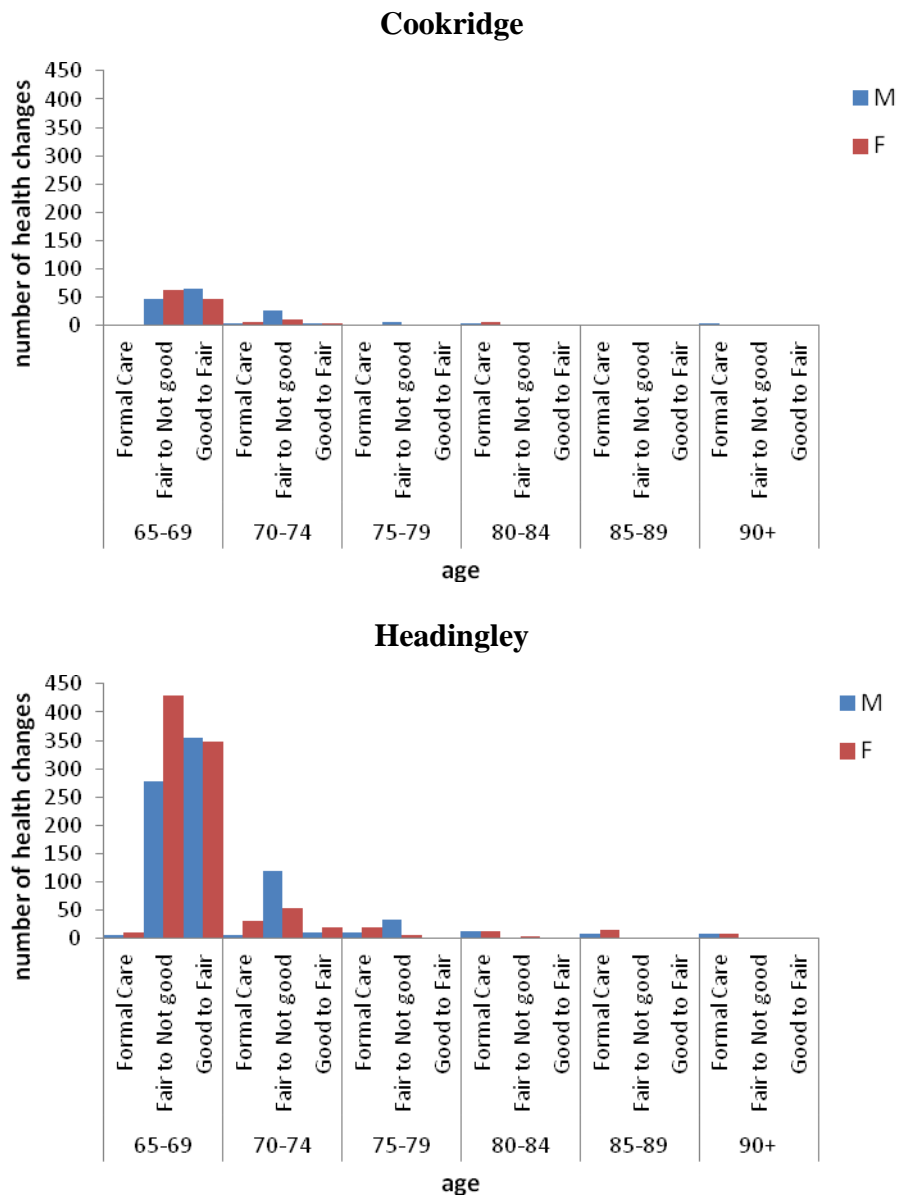
**Figure 4.20 Ward level projections of fertility in 2030**

#### 4.8.3.4 *Initial Health Change result analysis*

In the Health Change module, the changes in general health for population aged 65 or over in small areas are modelled to facilitate the public health planning within an ageing population. The health changes are simulated on the basis of changes in general health status of the population aged 65 onwards, from “Good” to “Fairly Good”, “Fairly Good” to “Not Good” and “Not Good” to “Formal Care”, indicating an individual’s general health has deteriorated to a status where informal care at home is no longer suitable and the person must be transferred into a formal care facility. Projected results in Cookridge and Headingley in year 2030 have been analysed by sex and age groups, as presented in Figure 4.21.

Overall, there are more counts of health changes in Headingley than in Cookridge, due to the larger population in Headingley. Most health changes occur at ages 65-69. Then the number of changes gradually reduces with age. There are very few health changes from age 80 onwards. In both wards, more females experience general health status change from “Fairly Good” health to “Not Good” than males at ages 65-69. On the other hand, this has changed at ages 70-74, where more males experience general health status

changing from “Fairly Good” health to “Not Good” than females and this pattern continues into the age groups 75-79.



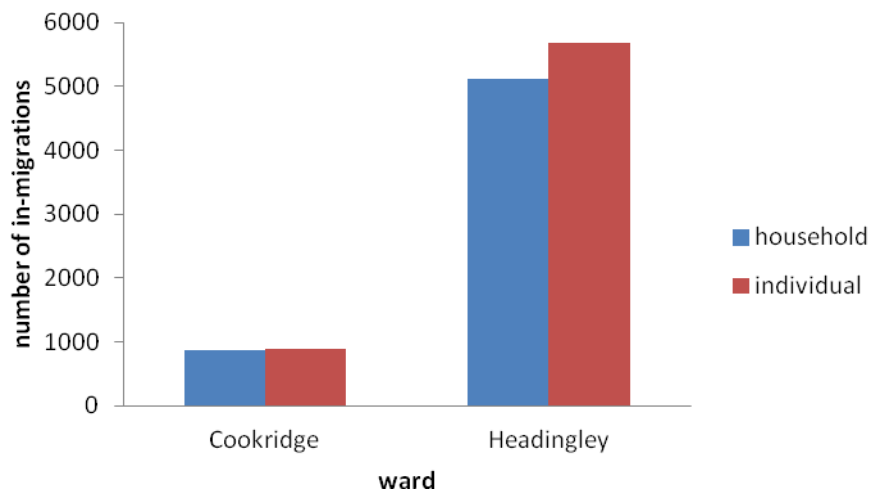
**Figure 4.21 Ward level projections of health changes in 2030 by sex and age**

In terms of the “Formal Care” demand, there is a consistently higher demand from the females than males in most of the older age groups (ages 65 and over). This may indicate a faster deterioration in females’ general health from “Not Good” to “Formal Care” than in male’s for those aged 65 or over. This may indicate that more females require formal care in part because male partner dies first, leaving them living alone without available in-house support. Such patterns are also consistent with the recent

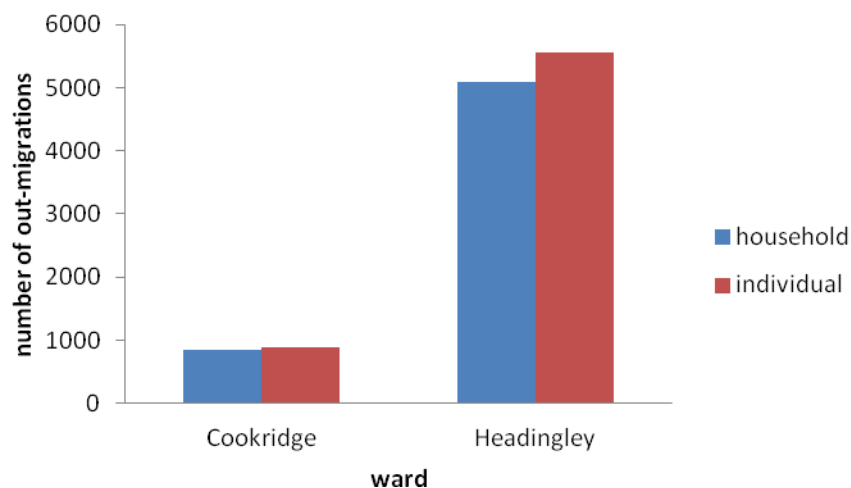
population trends reported by ONS that women live longer in poorer health than men do in UK (ONS, 2008c). Men have shorter lives.

#### 4.8.3.5 *Initial Migration result analysis*

The migration simulation results of Cookridge and Headingley in 2030 are analysed and the analysis indicates that there are significant variances in the migration patterns in small areas. Overall, there is a much higher degree of migration in Headingley than in Cookridge. However, there is very little difference between the in-migration and the out-migration movements in both wards, as indicated by Figures 4.23 and 4.24. Such patterns are consistent in both household and individual migrations.

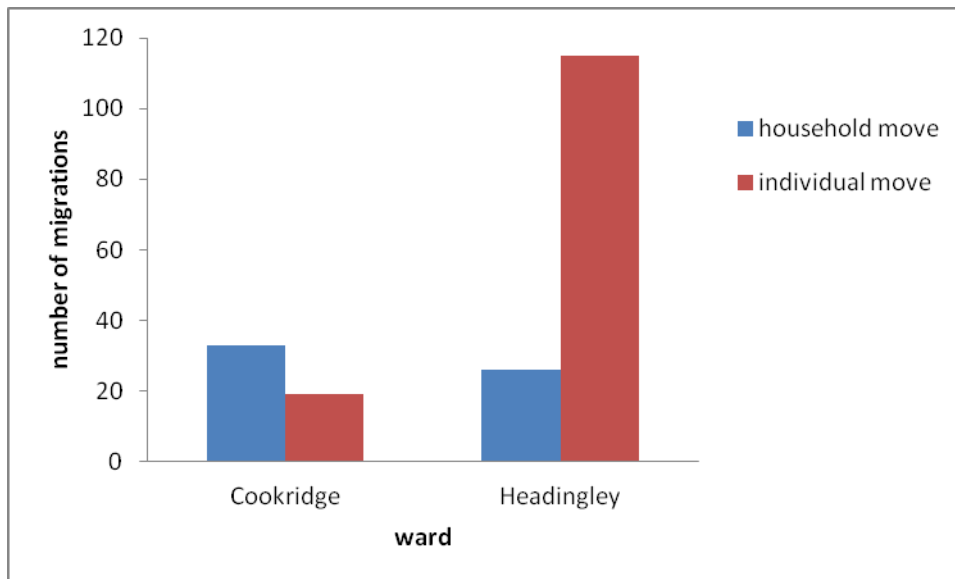


**Figure 4.22 Ward level projections of in-migration in 2030**



**Figure 4.23 Ward level projections of out-migration in 2030**

The net-migration analysis is illustrated in Figure 4.24. Projections in Cookridge indicate that there are more household migrations than individual migrations in 2030, but the difference between two types of migration is not big. Net-migration projection results in Headingley, however, indicate that there are substantially more individual migrations than the household migrations in the area. In contrast to net-migration in Cookridge, net-migration of individual migration is about four times of that of household migration in Headingley. This may reflect the local contexts of the two wards. Cookridge is a more established suburban area, while Headingley is an area popular with student accommodations in the city centre. Such small area characteristics also help explain the difference between the household and individual migrations in the two areas. The much higher individual migration level may be due to the impact of university student migration in this area, while most residents in Cookridge are families so they are more likely to move as a whole household. As most students are full-time student in ages 16-29, who leave their homes and come to Leeds to complete their higher education. Although they are highly mobile during their study duration, they are unlikely to move with households.



**Figure 4.24 Ward level projections of net-migration in 2030**

## 4.9 Conclusions

In this chapter, the general modelling approach used in the dynamic spatial MSM has been described. The reason for this approach is to try to provide a better representation of the Leeds population through the more sophisticated updating/ageing of the individual attributes within a local context. Therefore the population is modelled as individuals within households that have a rich set of important attributes. Such households locate in wards that are the basic spatial units in the area model. Wherever possible, transition probabilities that drive the population changes have also been calculated at the ward level to reflect the local characteristics.

Six demographic processes are being modelled in this dynamic spatial MSM due to their importance in population evolution and planning reference functions. Each process has been developed into a separate module, but they can interact with each other. Various factors that drive the demographic changes have been considered accordingly in each process. Among them, age, sex and location have been selected as the foundational factors that need to be captured in the processes of all demographic changes. Extra factors are introduced according to the requirement of the individual demographic process. Detailed discussions have been provided on each



factor's importance for the relevant processes. Such impact of the factors is captured in the probability calculation process. The probabilities are carefully calculated to reflect the chance of demographic changes for various groups of people with different characteristics. As described in previous sections, all six processes are probability driven and the simulation is based on the application of the Monte Carlo method.

The dynamic spatial MSM developed in this study provides the characteristics of studied population at the level of individuals through a truly dynamic ageing process. As described in Section 4.2.1, at each simulation interval of a year, the attributes of each individual change according to probabilities that reflect their demographic characteristics and local area characteristics. Such characteristics are captured in probabilities that are calculated using relevant demographic and spatial information. The Monte Carlo method is used to determine the probabilities. This is the general method shared by all six demographic process simulations, namely: Ageing, Mortality, Fertility, Health Change, House formation and Migration.

At the aggregate level of Leeds, population age-sex structure diagrams have been generated using the initial results generated by the MSM. They demonstrate the patterns of the population changes year by year. As all changes are dynamically simulated each year and driven by multi criteria based probabilities, including local area factors, the overall results should present a much more robust representation of the studied population, compared to static models that move population forward through general reweighting procedures.

At the small area level of wards, it is found that characteristics of the local population changes differently in small areas. Local population in Cookridge clearly demonstrates a more serious problem of ageing than Headingley. Having used the LQ (Location Quotient) method to analyse the populations of a set of wards, it is confirmed that population evolution does vary geographically. However, it is also found that although the results from established suburban wards such as Cookridge seem to be more

plausible, the results produced in the current MSM are not satisfactory for wards that are largely impacted by student migration such as Headingley. Possible reasons have been discussed and issues found will be addressed in Chapters 5 and 6.

The results from individual processes have also been analysed. The spatial variation is found in small areas even in the same demographic process. Cookridge and Headingley have been chosen to represent two very different wards in Leeds and the simulation results from year 2030 are used in the analysis. In all 6 demographic processes, the two small areas demonstrated geographical variances as described in section 4.8. Due to the page limitation of the thesis, comparisons with the 2001 outputs are not discussed in the initial result analysis. Instead comparisons between the model results at the beginning of the simulation (year 2001) and the end of the simulation (year 2030) are conducted using the revised model results in the following chapters 5, 6 and 7.

In short, the dynamic MSM models the individual changes using dynamic ageing technique. This means that each all simulated attributes are dynamically updated through the evolution of demographic changes and interactions of the demographic processes. At the aggregated level, the trend of the population changes can be traced over time. This can be useful for strategic decision making. At the disaggregated level, spatial variances have been found across different wards. The distinction also persists in all modelled demographic processes. Thus, the spatial MSM can provide valuable information for tactical decisions and location based studies and policies.

Migration is a more complex process because there are various reasons for migration. As there is interdependency between various demographic processes such as household formation, this process is more difficult to model. In the initial version of the dynamic spatial MSM, however, this process is only concerned with the flows within Leeds. Therefore the representation of the migration process has been flawed in this initial version of the dynamic spatial MSM. The lack of migrants that are moving

into Leeds, especially the younger population, may have caused the exaggeration of the ageing trend of Leeds. ONS projected that there are more migrants moving into Leeds than out. Also as the demographic processes interact with each other, a younger baseline population will have important impact on other processes. Consequently, the limitation in the migration module may also affect the population, births, deaths, health changes and household formations. Therefore, the issues introduced by the limitations in modelling of the migration process must be addressed. In Chapters 5 and 6, the further development of this process, attempting to address various issues, will be discussed in detail.

## Chapter 5

# Further development of the MSM

## 5.1 Introduction

As the three most fundamental components of change of any population, mortality, fertility and migration play a vital role in the evolution of the population. Not only do these processes drive the fundamental changes of a population, but the interactions between each of them can also cause critical changes in a studied population. For instance, if a large number of young people migrate into an area and stay, this will have an important impact on the fertility rate. This in turn will change the population evolution pattern and result in a different population structure in the years to come. It is therefore very important to capture the characteristics of the mortality, fertility and migration processes of a population in this population model.

As discussed in Chapter 4, compared to the official projections for Leeds, the initial results produced by MicroSimulation Model (MSM) have

exaggerated the ageing trend in the Leeds population (Section 4.8, Chapter 4). To understand the issue, we have then gone back to its root of the three fundamental components of change and three issues were found, corresponding to each fundamental component.

Due to its central location in England, people from both the north and the south can move to Leeds within relatively shorter distances. The migration flows between Leeds and the rest of the country or the rest of the world have a very important impact on the local migration patterns, which in turn contribute to population structure changes. To model the population evolution within Leeds, it is necessary to capture such impacts in the MSM. In Chapter 4, it has been shown that such migration flows impact not only on the population structure at the aggregated level of LA, but also at the disaggregated small area level of wards. At the LA level, the model results exaggerated the ageing pattern of the population because they did not include the migration from and to other regions inside UK or migration from and to other countries outside UK; on the disaggregated level of wards, the incomplete migration model cannot capture the subtlety of the local migration patterns and failed to capture the impact of distinctive migration patterns of some sub-populations such as the university student population in Leeds.

For the component of fertility, registered births 2002-2006 by age of mother are used to calculate ASFRs (age-specific fertility rate) in the ONS SNPPs. Those are then compared and adjusted according to the national fertility rates, which uses a leading indicator the Total Fertility Rate (TFR) that indicates the average number of live births that a woman will have during her reproductive life time, rather than the General Fertility Rate (GFR) representing the number of annual live births per thousand women aged 15–44 (ONS, 2008b&2008c). In the MSM, the ASFRs in 2001 are used to generate the probabilities that a woman of a given age would give birth in the next year. However, they are not updated in the first version of the MSM to reflect the recent trends of fertility, while ONS NPPs used higher fertility rates in their model for the period 2001 to 2011, and then a high TFR from

2011 onwards. Therefore it is necessary to update the fertility rates over time, using the recently published statistics and long term trends revealed in such statistics such as the TFR (ONS, 2008b&2008c). In this model, the average annual changes of TFRs over period have been applied to update the fertility probabilities.

The ONS also uses improved mortality rates in their projections. In ONS' 2006 based national projections, the age-specific mortality rates are assumed to reduce by approximately 2% annually (ONS, 2008a), although the mortality forecast used in ONS national projection, based on work by the government Actuary's Department, is much more sophisticated. However, this means that modelled population in our system can live longer and our MSM needs to reflect such a trend of mortality improvement in order to make appropriate projections of the mortality process of the population in Leeds.

Based on the analyses in Chapter 4 and the above discussions, a comprehensive migration framework is designed. The revised framework includes an expansion of the migration flows considered from two to six, adoption of the forecasts of TFR to drive the projections of fertility rates and updated mortality rates declining at a constant rate over time. In the rest of this chapter, we will introduce the re-designed migration framework and the further development that has been carried out with the mortality and fertility modelling. Then we will analyse the simulation results for the re-constructed MSM and discuss the results in comparison with the previous simulation results, using the ONS projections as a benchmark.

## 5.2 The extension of migration process

### *5.2.1 New migration framework*

A new framework for migration has been designed on the basis of the previous discussion. Within the new migration framework (Table 5.1), six migration flows will be modelled.

The migration flows from and to Leeds wards within Leeds (flows 1 and 2 in Table 5.1). Such inter-ward flows can cancel each other out in effect on the aggregate level of the LA, but they make a crucial impact on population changes in small areas. According to spatial interaction theories, distance is considered as a disincentive factor in people's movements. The further the distance, the less likely people will move (Wilson and Bennett, 1985). In fact such inter-ward movements make up the majority of the inter area migrations in Leeds and change the composition of the small area population structures each year.

The internal migration flows to and from other areas within UK (flows 3 and 4 in Table 5.1). The migration flows that move in and out of the Leeds wards to the rest of the UK have an important impact for local migration patterns. For example, the student migration patterns are a critical cause of changes in small area population structures. Each year the new groups of freshmen (women) come from other regions of UK and stay in certain Leeds wards for their higher education, while those who have completed their studies leave Leeds for other regions in UK. The internal migrations between Leeds wards and other regions of UK make up the second largest portion of Leeds migration.

The international migration flows from and to Leeds wards (flows 5 and 6 in Table 5.1). Immigration and emigration are the last two flows to complete the migration framework. The international migrations make up the smallest portion of the Leeds migration.

The main datasets used for the further development of the migration module are: the UK Census data in 2001, the Mid-Year Estimation (MYE) of the population 2002-2005 and the ONS Sub-National Population Projections (SNPP) 2006-2031 (ONS, 2008d). As there is no information on the emigration out of UK to the rest of the world in Census 2001, the information from International Passenger Survey (IPS) is used to provide the vital information on this migration flow (ONS, 2001a).

**Table 5.1 The extended migration framework**

To From	Destination			
Origin	Leeds Ward	Other wards in Leeds	Rest of England and Rest of UK	Rest of the World
Leeds Ward	<b>A</b>	<b>1</b>	<b>3</b>	<b>5</b>
Other wards in Leeds	<b>2</b>			
Rest of England and Rest of UK	<b>4</b>			
Rest of the World	<b>6</b>			

Notes:

1= outflow from a Leeds ward to other Leeds wards

2= inflow from other Leeds wards to a Leeds ward

3= outflow from a Leeds ward to the rest of the country

4=inflow to the rest of the country to a Leeds ward

5=outflow from a Leeds ward to the rest of the world (emigration)

6=inflow from the rest of the world to a Leeds ward (immigration)

A=migration flows within a Leeds ward, not modelled

### *5.2.2 Further development under the new migration framework*

When modelling migration within a closed system, all the information needed comes within the system. The populations, locations and the corresponding functions already exist in the system. Modelling migration in an open system, however, is another different matter. Individuals still need to go through the three-stage migration process as described in Chapter 4 (Section 4.7). However, many components in the extended migration framework do not exist in the current system and need to be developed.

In the following section, the further development work under the new migration framework will be described.



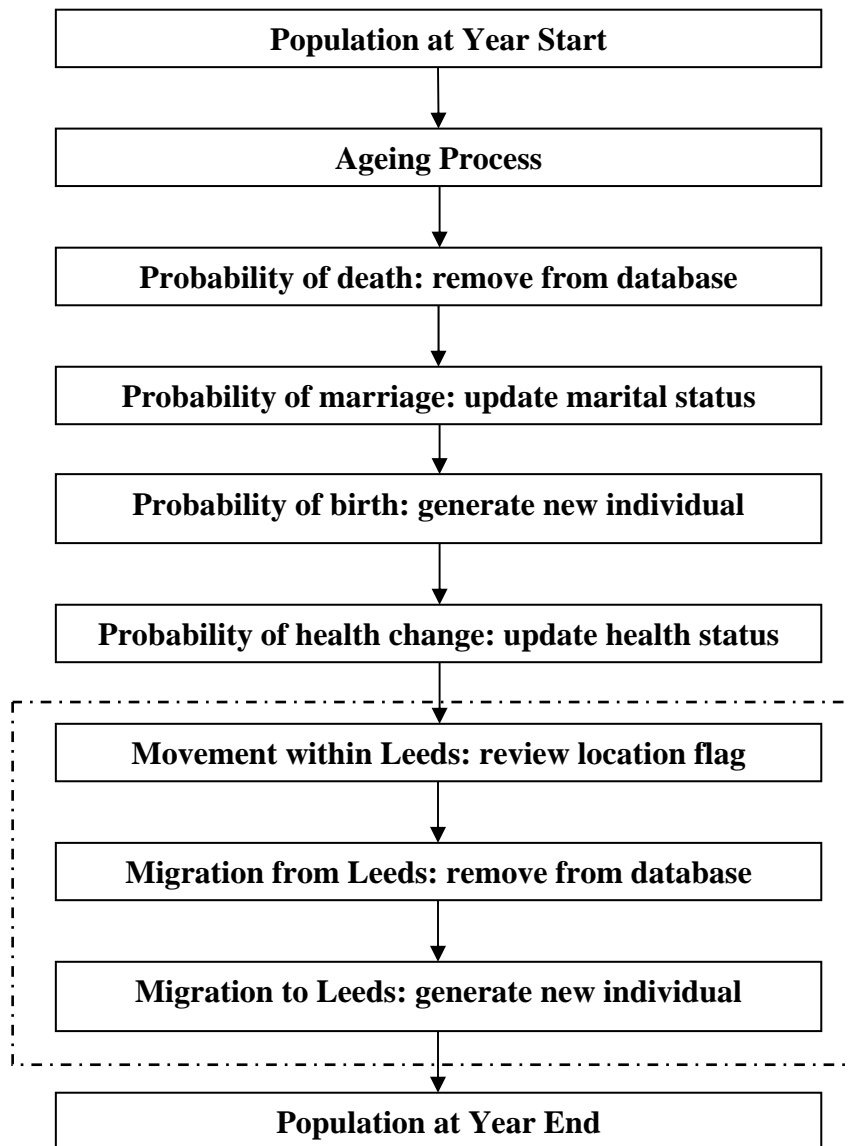
### 5.2.2.1 *Restructuring the system*

There are some structural components of the open system not existing in the current model. To support the modelling of interactions between the closed system and the outside world, structural changes need to be made to make the existing structure compatible with the new requirements and extra development work needed to be carried out to provide the capability to support the modelling of extra activities of extra population.

Within the extended new migration framework (Table 5.1), there are six migration flows. Among them, the inter-ward migration flows within Leeds have already been modelled in the existing MSM. Therefore, only four extra flows need to be modelled: the in and out flows of the internal migration within UK and the in and out flows of the international migration.

To enable the modelling of the UK domestic and international migration, an in-migrant pool and an out-migrant pool have been created to accommodate the migrants that are decided on the basis of probabilities to move in or out of Leeds and placed in the relevant migrant pool (in or out). Then the four flows are created by either selecting and removing the appropriate individuals/households from the out-migrant pool; or selecting the appropriate individuals/households from the in-migrant pool and placing them into the appropriate small areas. To ensure the efficient handling of all the extra data flows, various list processing functions and probability determination methods have also been developed to support the new structure of the extended migration framework.

Such structural changes also affect the modelled process. Within the extended migration framework, the new migration process has become effectively three parts to deal with the migration within Leeds, from/to Leeds to/from the rest of the UK and from/to Leeds to/from the rest of the world. Although they are still all implemented within the same module of Migration, separate sub-modules have been developed for each process that handles different migration flows. A simplified illustration of the main processes under the new framework can be viewed in Figure 5.1.



**Figure 5.1 The main processes under the new framework**

### 5.2.2.2 *Creating the migrants*

When modelling migration within a closed system, the most important work is to calculate the appropriate ward-based migration probabilities. Then they can be applied to each individual through the Monte Carlo simulation. Dependent on the individual characteristics, whether or not they are moving and whether they are moving in households or individually, can then be decided on the basis of corresponding probabilities. Finally the migrants are moved from the current wards to their destination wards according to the

destination probabilities. All the changes are within Leeds and the Leeds population, which exist within the system already.

To enable the modelling of Leeds migration in an open system, many components need to be developed. When modelling the migrants that are moving out of Leeds to other regions of UK or even other countries of the world, it is relatively easy. We follow the 3-staged process and at the last step, remove such out-migrants to some storage or remove them completely from the system. However, when modelling the in-migrants, it is more complicated. The first issue we are facing is to invent the “in-migrants” themselves. There are two options to create such migrants:

1. create them outside the system, i.e., create the populations from scratch with the right combination of characteristics using the original sample records from the census;
2. create them inside the current system, i.e., create the populations using the sample records that are already in the current system.

To create the migrants from the census data, we have to understand what migrant characteristics of each individual are appropriate for a certain small area. However it is not a simple question. Previous research has pointed out that the co-variants are hard to be effectively represented when trying to recreate populations with appropriate characteristics from samples, using a couple of key variables as the overall controls and a couple of extra variables as the constraints. For example, ethnic status is not as strongly spatially clustered within the model as in reality. This probably reflects the fact that none of the controls or optimising attributes currently selected is closely related to ethnic status. So it is not easy to find out what controls and optimising attributes might be selected in order to give the best overall profile of a city and its constituent neighbourhoods (Birkin *et al.*, 2006, Voas and Williamson, 2001). Besides, even if we can work out the answer in theory, it will not be easy to compute it in practice, as such characteristics can be a combination of many factors. Previous experience has revealed, it is not practical to include too many constraints when attempting to recreate a population from a smaller sample. This puts a high demand on both

computing resources and skills and it is not within the priority and time limit of this research.

On the other hand, we know that the local context has an impact on population in the small areas (Chapter 3). In other words, the current local population reflects the local characteristics already. Assuming the distribution of population in the existing system is reliable, a random selection of such individuals will be representative for the local population. Therefore we decided to select randomly the existing population records in the small area and clone them. Then the cloned records of individuals are changed into the new “in-migrants”. The limitation of this approach is that the clones only capture the characteristics of the existing population within the area and cannot provide potentially different characteristics of new migrants from outside the area. In future, this area of the model will need to be strengthened. More details of migration such as probability calculations are described in section 4.7.3 and section 3.4.3.

### 5.2.2.3 *Revising the migration process*

Having restructured the system, we then start to regenerate the full migration process with all six migration flows included. In this section, we will describe how the migration is initialised and updated for both the internal and international migration.

#### **5.2.2.3.1 Migration initialisation**

As the simulation starts from year 2001, we used the data from Census 2001 to initialise the baseline migrants. This process is the similar as the inter-ward flows that have already been modelled in the existing MSM. Therefore the initialisation is mainly in two parts: the internal migration within UK - in and out migration between Leeds and other regions and the international migration outside UK - in and out migration between Leeds and other nations in the rest of the world. As described in Section 5.2.2.2, for the out-migrants that left Leeds, their records were simply removed from the model and for the in-migrants that stayed in Leeds, randomly selected local

population records are used to recreate the synthetic population with local characteristics.

For the internal migration within UK, the Interaction Data from the ONS is used to provide the information on flows of individuals in the UK between origins and destinations. The Interaction Data includes both migration and commuting data. The specific data set used is the Special Migration Statistics (SMS) 2001 level 2 data (ward level) to generate the ward-based probabilities for both the out migration from Leeds ward to the rest regions of UK and the in migration from the other regions of UK to Leeds (ONS, 2001b).

The international migration is more complicated. We used the SMS data again to generate the immigration probabilities, but there is no emigration information in this data. We had to use the International Passenger Survey (IPS) data as the basis to calculate the emigration probabilities, where information on people's self-reported intentions about for how long they plan to leave the UK has been used to estimate the emigration (ONS, 2001a). How the migration probabilities are calculated is described in section 4.7.

#### **5.2.2.3.2 Migration updating**

After estimating the full six migration flows in and out of Leeds small areas using the SMS data and the IPS data, we then need to update the migration probabilities each year to reflect the annual changes in migration patterns. We decided to use the ONS Sub-National Population Projection (SNPP) data (ONS, 2008a), as it uses a finer spatial disaggregation than the national projections and captures more Local Authority (LA) characteristics. However, the most up-to-date SNPP data at the time of the revising the MSM was the projections based on year 2006, therefore we had to use the ONS Mid Year Estimation (MYE) data (ONS, 2008a) for the migration updating in period 2001-2007 and use the SNPP data for the period 2007-2031. To best reflect the local characteristics, we use the LA based indicators from both SNPP and MYE data.

One important thing that needs to be pointed out is that, although these data can be found at the level of LA, we do not use the exact migration rates in such datasets. Similar to what we did in the other demographic processes such as mortality and fertility, we follow the assumptions/trends that are revealed in those datasets, but we use the spatially disaggregated migration probabilities by small area and apply them to potential migration candidates according to their relevant migration/demographic characteristics to determine whether they move, how they move and where they move to.

### 5.3 Fertility updating

To update the Fertility probabilities, we use the trend in the Total Fertility Rate (TFR), a measure which indicates the average number of live children that a group of women would have if they experienced the Age-Specific Fertility Rates (ASFRs) of the calendar year in question throughout their child-bearing lifespan. The TFR is used instead of the General Fertility Rate (GFR), which represents the number of live births per thousand women aged 15–44. This use of the TFR is consistent with the ONS approach, where ASFRs were adjusted in SNPP according to the trends revealed in NPP, where TFRs were used.

The TFR is a synthetic rate, not based on the fertility of any real group of women, since this would involve waiting till they had completed childbearing. It is not based on counting up the total number of children actually born over their lifetime. Instead it is based on the age-specific fertility rates of women in their “child-bearing years,” which in conventional international statistical usage is ages 15–44. As a “period” measure, the TFR is influenced by changes in the timing of childbearing within women’s lives as well as any changes in completed family size. However, as it is independent of the age structure of the population, it provides a better index of fertility in the longer term than the crude birth rates that are calculated from the annual number of births per thousand women (ONS, 2010).

As described above, the two approaches make a significant impact in modelling the fertility process and will bring some variance in the modelling results, especially when simulating the population changes for a longer term (30 years). Based on the features of the two approaches, we decided to use a combination of the two rates. From year 2001-2008, we use the currently available GFRs for the fertility simulation. As the TFR is based on a “period” measure and independent of the annual age structure of the women, we decided to adopt this approach and use the trend revealed in the TFR to update the fertility probabilities from 2008 onwards.

Therefore, we used the data in the most up-to-date data from the population trend at the time to generate the annual fertility rates in years 2001-2008 (ONS, 2008c). Then we updated the fertility rates by mothers’ age, marital status and small areas in Leeds on the basis of those. From year 2008 onwards, we then used the trends that are revealed in TFRs in ONS (2008d) to update the fertility rates for the rest of the simulation period. In 2001, the UK TFR was 1.62 children; it has continued upwards to nearly 2.0 in 2009-10. We then kept the TFR constant as the process that drove up the TFR-women in their 30s catching up on births postponed in their 20s-has run its course.

## 5.4 Mortality updating

As discussed in Section 5.1, ONS project that there will be continuous improvement in mortality of UK population. The ONS projections generally assume higher rates of improvement for the future than experienced over corresponding periods in the past and that is more than 2% of reduction annually for both sexes (Table 5.2).

Due to the advances that have been made in standards of living, disease prevention and overall health care, the expectation of life at birth for UK population, is expected to rise from 77.2 years in 2006 to 82.7 years in 2031 for men, and from 81.5 years in 2006 to 86.2 years in 2031 for women, based on the mortality rates for the year in question (ONS, 2008a).

**Table 5.2 Actual and assumed overall average annual rates of mortality improvement**

England & Wales	Per cent			
	Males		Females	
	Past (actual)	Future (assumed)	Past (actual)	Future (assumed)
Last/next 24 years	2.13	2.12	1.47	2.15
Last/next 44 years	1.54	1.62	1.33	1.64
Last/next 74 years	1.23	1.37	1.27	1.38

Note: Historic estimates are based on comparison of latest (2004–06) interim life tables with English Life Tables for 1930–32, 1960–62 and 1980–82.

Source: (ONS, 2008a, Table 7.3)

In the same document, the ONS also compared the differences between the projected populations in the variant mortality projections and the principal projection. Assuming higher or lower mortality rates than in the principal projection can make significant differences in the projection results, especially over a longer period of projection time. A summary of such differences have been illustrated in Table 5.3.

**Table 5.3 Population differences between variant mortality projections and principal projection by age, 2011-31**

Year	United Kingdom									
	Difference between high life expectancy variant and principal projection					Difference between low life expectancy variant and principal projection				
	All ages	Under 60	60–74	75–84	85 & over	All ages	Under 60	60–74	75–84	85 & over
2011	45	6	10	13	16	-45	-6	-10	-13	-16
2021	209	26	45	56	82	-211	-27	-47	-57	-81
2031	585	57	114	152	262	-605	-63	-124	-163	-256

Source: (ONS, 2008a, Table 9.4)

To reflect trends in mortality, we have to update the mortality probabilities used in our MSM according to assumptions about mortality improvement. As in Fertility module, the mortality updating is also divided into two stages. This is mainly because the most up-to-date mortality data available at the time of the further development of the MSM were based on 2006



projections. Therefore, from 2001-2006, the mortality data in ONS (2008c) were used for the annual update of disaggregated mortality probabilities by sex, age and small areas. While from 2006-2031, the percentage reduction in death rates in ONS (2008a) was assumed to help update the mortality probabilities to reflect the continuous mortality improvements. As in the Fertility module, we decide to apply the difference between the rates to update the base probability instead of applying the ONS rates directly, as the fertility and mortality probabilities used in our model are single year of age, sex and location specific and can therefore provide more individual characteristics.

## 5.5 Result analyses of the revised model

As described in previous chapter and Section 5.1, we found some issues within the initial version of the MSM. To address such issues, we have updated the mortality and fertility probabilities using the most up-to-date data at the time. More importantly, we re-designed the migration framework to include the six migration flows into and out of Leeds small areas.

With the further development of the MSM, we then analysed the results produced by the revised MSM and compared them to the previous results (details of results of the initial model can be found in Chapter 4), using the ONS projections as a guideline. The analyses of results of the revised model have been carried out the same fashion as in Chapter 4. However, the population sex-age structure analyses are discussed in the following sections, while results produced by the revised model have also been analysed separately by demographic processes of mortality, fertility and migration against the original model results to demonstrate the difference made by the updating work carried out on the three processes, using the more up-to-date ONS projection data.

### 5.5.1 Population sex-age structure analysis in Leeds

Having carried out the further development work on the migration, fertility and mortality modules, we then generated the population sex-age structures using the simulated results of year 2001 and 2030. The MSM results are then compared to ONS projection (2006 based). The population sex-age structures are showing in Figure 5.2.

From Figure 5.2, we can see that there has been significant improvement in the projected population compared with previous MSM results. Compared to the previous model results (Figure 4.15, Chapter 4), the new results suggest that Leeds population continues growing instead of shrinking as in the previous MSM. Although it suggests an ageing trend in the population, there is also some increase in the younger population aged 0-24 and the middle aged, compared to the baseline population in 2001. Such trends are consistent with the official projections from ONS.

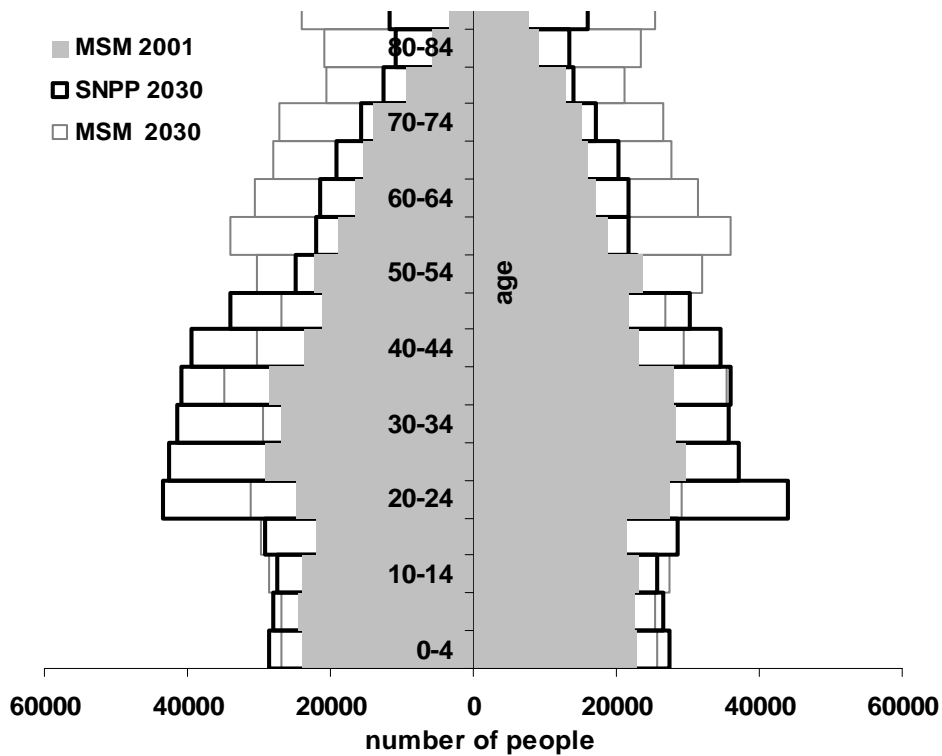


Figure 5.2 Population sex-age structures of Leeds population in 2001 and 2030 generated by the revised MSM compared with the ONS estimates in 2030

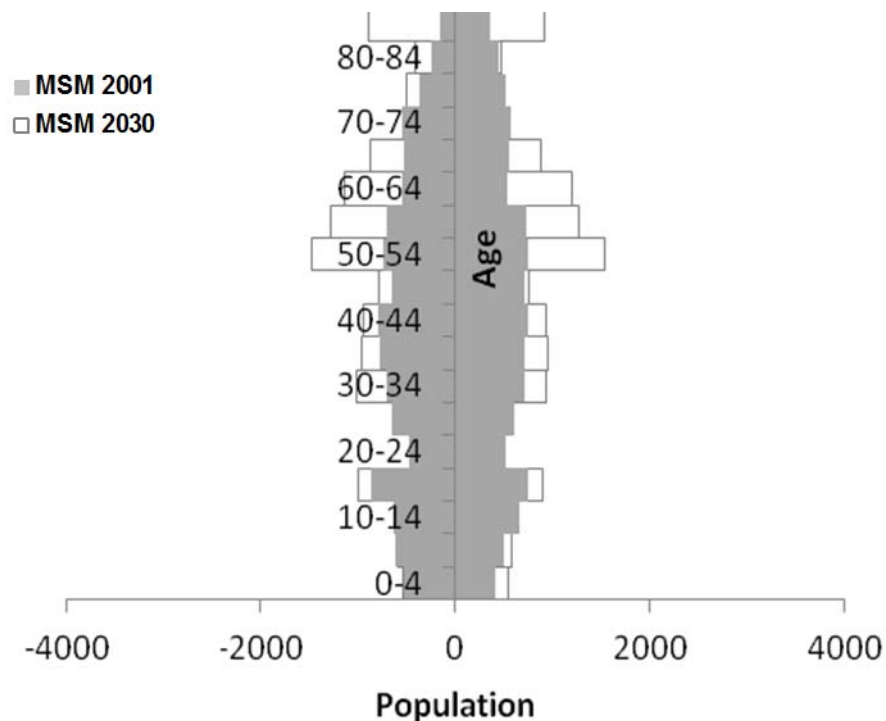
However, when compared to the ONS projections in 2030 in more detail, the MSM is found to still exaggerate the ageing trend of the Leeds population. One possible reason may be that ONS projects a much larger number of immigrants than the MSM. Part of this reflects a real rise in immigration from 2001 onwards. Another part of this reflects the model used by ONS in the 2006 SNPP to estimate the immigration flow to Leeds. This flow was substantially over-estimated, as pointed out by Boden and Rees (2010), where alternative estimates of sub-national immigration have been produced using data from administrative registered and maintained by the NHS, DWP and HESA. They suggest that using proxy variables that covers most of the population is more reliable than using survey local estimates from the IPS and LFS/APS. Another reason is that the MSM applies the same ONS rates to all ages to update the migration probabilities. However, the migration rates change by age. At younger ages, Leeds gains migrants, but at older ages it loses migrants to the surrounding districts and to retirement areas. Due to the two reasons, the MSM retains too many older people in Leeds (Figure 5.2). The analyses and comparisons continue in more detail in the following sections.

### *5.5.2 Population sex-age structure analysis in small areas*

To examine the spatial organisation of the population structure, we also produced the population sex-age structures from the ward based results. As there are no official projections at the ward level, the analysis is only based on the comparison between the simulated results between year 2001 and 2030 (Any other year during the simulation can be used, but year 2030 is selected to represent the projection towards the end of simulation). The two wards, Cookridge and Headingley, have been again selected for the analyses for two reasons. The first reason is because that they are representatives of two distinctively different small area populations and the second is that the same small areas are selected in order to compare with the results from the

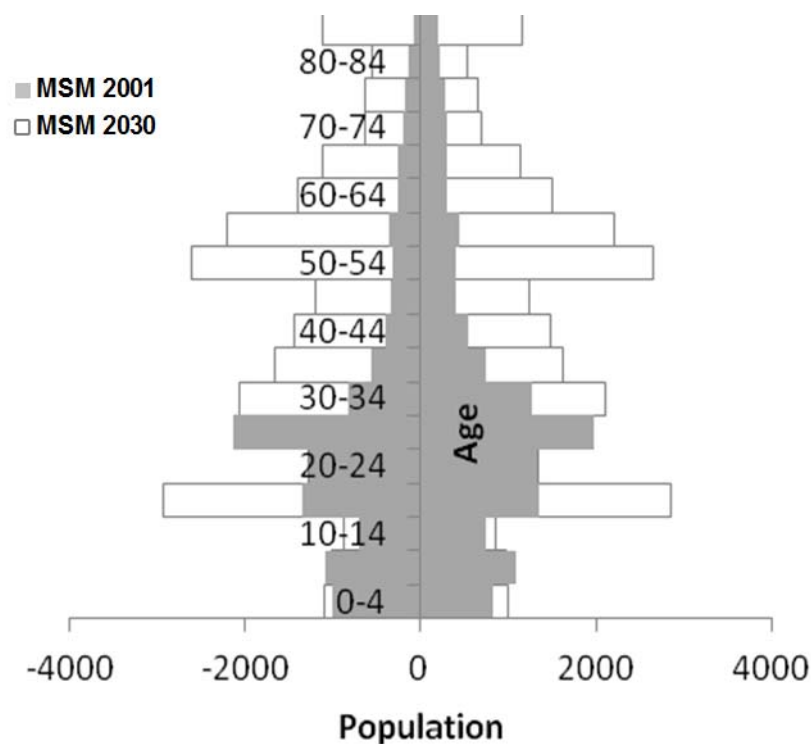
previous version of MSM. Figure 5.3 illustrates the population changes in Cookridge.

Although the new projected population for Cookridge is not shrinking any more in comparison with the earlier model results (Figure 4.16), the older population has been increasing fast in the area. On the other hand there has not been a lot of increase in the younger population. In fact the new results suggest that the largest population group is the ages 50-64, as well as a considerably large population of ages 85 and older. As Cookridge is an established suburban area, the ageing trend should be evident in the projected population in this small area, although the model results indicate some exaggeration (Figure 5.3). The increases in ages 50-69 are probably the result of in-migration from other Leeds wards and from the rest of the UK. They could not be the result of ageing-in-place (younger cohorts living in Cookridge becoming older) nor the result of higher fertility assumptions. The possible causes to such exaggeration to the ageing population have been discussed in Section 5.5.1.



**Figure 5.3 Population sex-age structures of Cookridge population in 2001 and 2030 generated by the revised MSM**

In Headingley, similar patterns have been found in the small area population structure. The population has been increasing and the end population is younger than in previous projections (Figure 4.16). In Figure 5.4, the biggest increase has been found in the population aged 50-69, although there are also significant increases in age groups 85+ and 15-19. The projected results of Headingley population in 2030 by the revised MSM almost present an opposite to the population structure as in 2001. The population in this small area has changed from a population that is mainly young into a population that demonstrates an ageing trend.



**Figure 5.4 Population sex-age structures of Headingley population in 2001 and 2030 generated by the revised MSM**

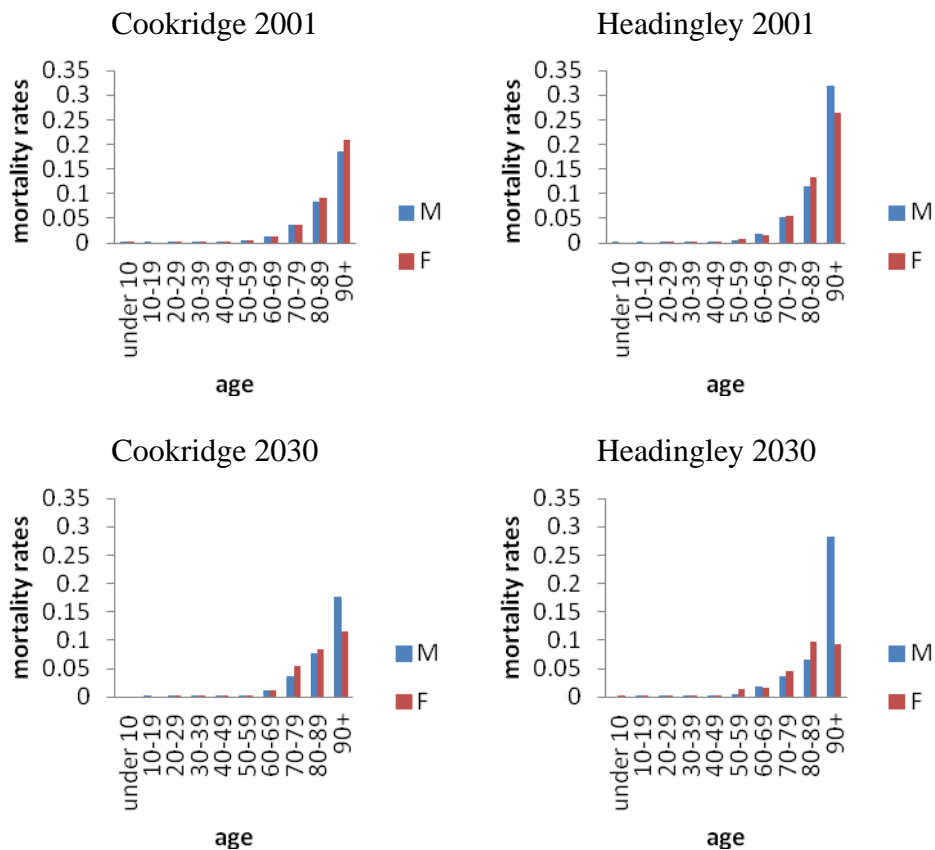
However, this level of population increase is unlikely because of housing constraints. Also university student migration has a great impact on the population structure of Headingley. Each year new university students come into this area to look for privately rent accommodations, while the previous student tenants moving out. Due to the replenishment of the student population each year, the population in this small area stays younger than in others. The projection using the revised MSM therefore has failed to capture the small area subtlety of local migration patterns, even with an extended

migration framework (Figure 5.1). In fact, the age-sex structure analyses of the two small area populations reveal that using the MSM, the two different small area populations have started to change towards similar patterns at the end of the simulation. To retain the local characteristics, this issue will be discussed and addressed in Chapter 6.

### *5.5.3 Result analysis by demographic processes*

Analyses by separate demographic processes have also been conducted. The analyses of Mortality, Fertility and Migration have been discussed here to demonstrate the differences between the original and revised model, due to the updating work carried out on these three modules in the model.

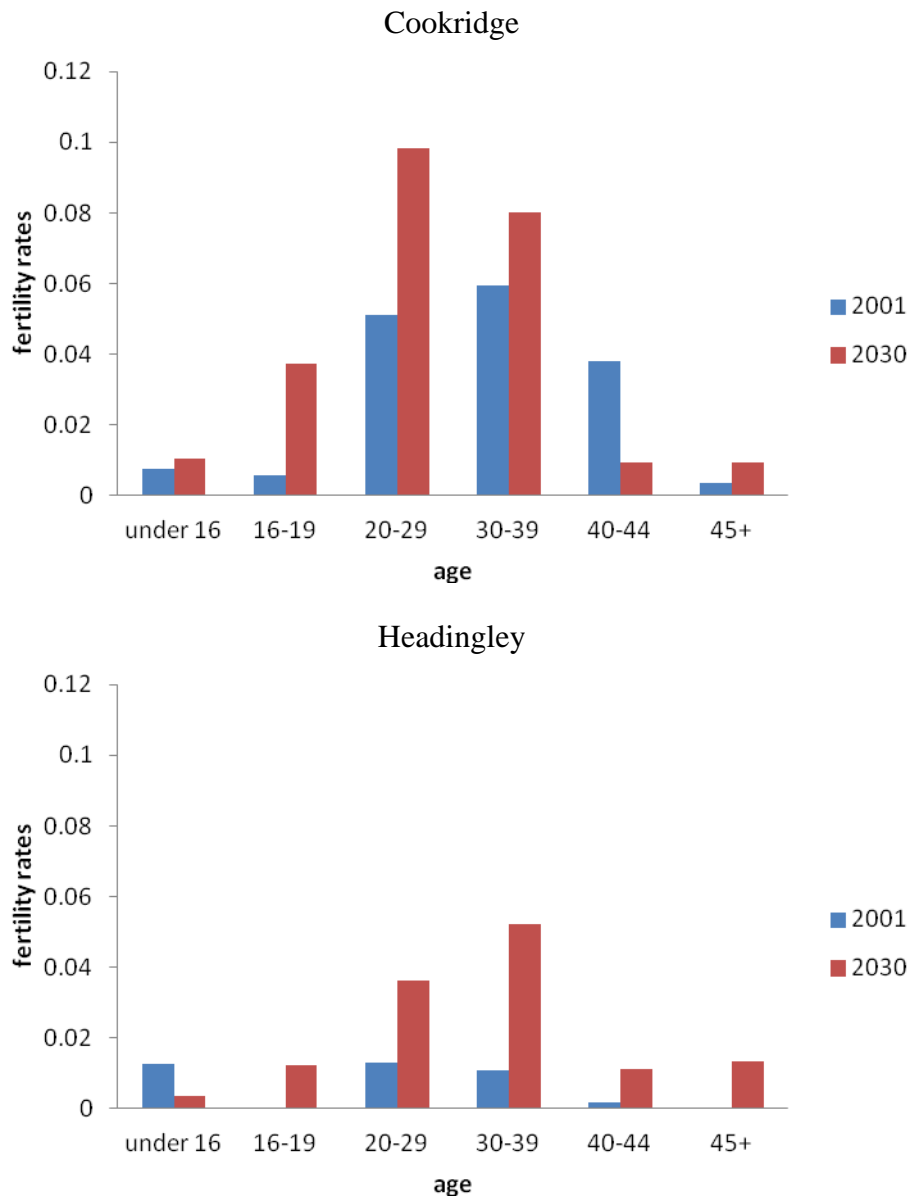
As illustrated in Figure 5.5, the mortality results generated by the revised MSM indicate similar mortality trends as in the original version: very low mortality in the young ages, but the mortality gradually increases with age and the increase speeds up from age 60. The very old ages have high mortalities. Compared with the mortality rates in 2001 to those in 2030, the model results indicates that there is a reduction of mortality in most ages in both small areas, with more significant reductions for older ages. However the mortality rates for men ages over 90 are still high in both wards. In contrast, a substantial reduction for women of the same ages are found in both wards in 2030. Overall, Headingley has a higher mortality than Cookridge in 2001. However, this has changed after 30 years simulation. With the improved mortality rates, mortality in Headingley has greatly reduced and has a slightly lower mortality rate than Cookridge, except the very old ages 80+. The reduced mortality rates and the annual replenishment of young student population in Headingley may have both caused this change. Compared with the initial results in 2030, such changes in mortality in the two small area populations in the revised model results in 2030 demonstrate the impact of the updating of mortality probabilities using the same assumptions as the ONS mortality projections.



**Figure 5.5 Mortality result analyses of the revised MSM: Cookridge and Headingley in 2001 and 2030**

Figure 5.6 shows the difference made by using the trend revealed by TFR (Total Fertility Rate) in the longer term fertility predication. Overall, the fertility rates in the 2030 are higher than the original model results, which uses constant fertility probabilities throughout the simulation. In the revised model results of year 2030, women aged 20-39 still have the highest fertility rates. However, in Cookridge, the highest fertility is found in ages 20-29, while in Headingley, it is the ages 30-39. There are also significant increases of fertility for females of ages 40 onwards in Headingley. This may indicate the trend of postponed child-bearing ages in this area, as suggested in recent ONS population trends. Compared to the results in 2001, there is an increase of fertility for ages under 19 in Cookridge in 2030. On the other hand, there is only an increase of fertility for ages 16-19 in Headingley, while there is a reduction of fertility for ages under 16. Overall

there is a higher fertility level in Cookridge than in Headingley in 2030, especially in ages 20-39. This is consistent with the results in 2001, too.



**Figure 5.6 Fertility result analyses of the revised MSM: Cookridge and Headingley in 2001 and 2030**

Such results may suggest the impact of local population characteristics. There is a large student population in Headingley. Although there is a large number of young females aged 20-39, many of them are likely to be in full-time higher education and are unlikely to give birth to babies.



The revised MSM uses a comprehensive migration framework when modelling the process of Migration. It brings in all migration flows in and out of Leeds. As a result, the revised model can provide a complete picture of the migration for Leeds. Analyses have been carried out by age, sex, flow (flows 1 and 3 are combined to give the total of internal in-migration and flows 2 and 4 are combined to give the total of internal out-migration) and small areas and the results have been presented below. Figure 5.7 indicates that younger ages have the highest migration rates (percentages of migrants against total population of the same group in the area), while older ages are more likely to move less (the high rates for the ages of 70 and older may be caused by the small numbers of population of such ages). The migration results also reveal a much higher rates of internal migrations within UK than that of international migration. There are more internal migration out of Leeds than into it, although there are more immigration than emigration. Such patterns are consistent with the ONS projections for Leeds.

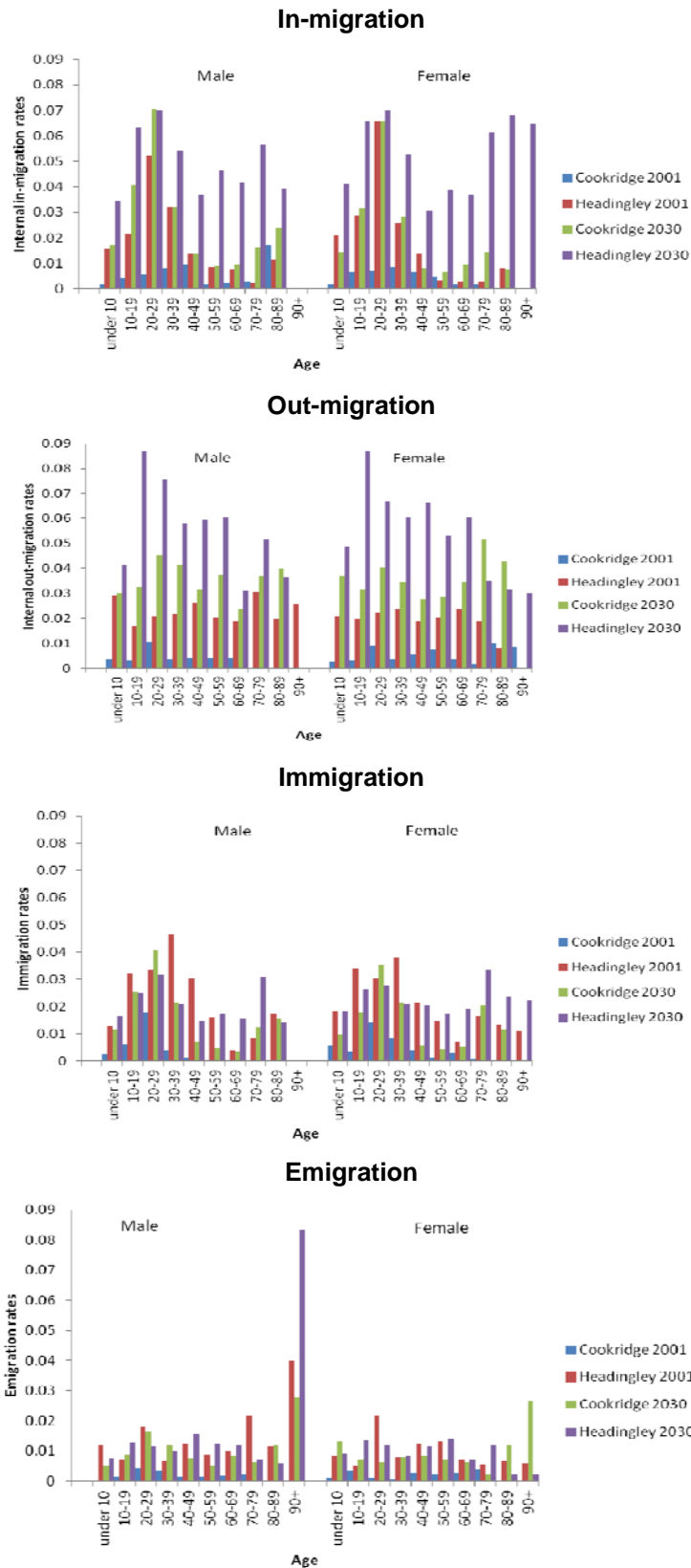
Overall, migration in Headingley has increased considerably over time, especially the internal migration within UK. In 2001, The internal migration rates in Headingley are already higher than that in Cookridge. In 2030, the internal migration rates (both into and out of Leeds) have increased substantially and are much higher than that in Cookridge for both sexes. However, the results in 2030 indicate a trend that there is more increase of out-migrants aged 10-29, but there is more increase of in-migrants aged 40 and older. This does not reflect the observed migration patterns in these wards. The reason of this may be that the ONS probabilities used to update the migration over time are not age specific and such flaws may be further exaggerated by the spatial disaggregation used in this model. In the international migration results, however, indicate that there is an increase of immigration in all ages and both sexes in Cookridge, although there are more increase of immigration in middle aged males and older females in Headingley. There is also an increase in male emigrants in both wards and female emigrants in Cookridge in ages 30 to 69. However, this is not significant for female emigrants in Headingley. There is a high emigration rate in ages 90 and older, especially for males. However, we need to

consider the small number issue for this age band. Both internal and international migration results from the revised model indicate that there is a trend of increasing migration into and out of both Cookridge and Headingley over the years.

## 5.6 Conclusions

As analyses in Chapter 4 revealed, modelling Leeds population in a closed system is not appropriate. As one of the three fundamental components of change in a population, migration has an important impact on the evolution patterns of Leeds population. Due to its central location in England, migrants from both the south and the north can move to Leeds with relative ease and vice versa for the out migration from Leeds to the rest of UK. International migration to and from Leeds also plays a role in the changes in the local population structure. Therefore we have designed a new migration framework that includes all six flows that completes the migration process in the small areas of Leeds. As the bi-direction inter-ward migration flows are already modelled in the previous version of the model, only four more migration flows needs to be modelled under the new framework: the UK internal migration flows into and out of Leeds and the international immigration and emigration flows.

The previous MSM assume the current fertility and mortality trends continue throughout the 30 years' simulation time. To better capture the changes in fertility and mortality, the other two fundamental components of population change, we studied the most up-to-date information used in the official population projections from ONS at the time this refinement work was carried out. We found there are two major differences between our projections and ONS' projection. When modelling fertility, ONS uses a different fertility projection approach. The TFR (Total Fertility Rate) approach is a period based approach, independent of the annual changes of the age structures of the women at risk. Therefore we decided to use the trend revealed by fertility rates calculated by this approach to update the fertility changes over time.



**Figure 5.7 Migration result analyses of the revised MSM: Cookridge and Headingley in 2001 and 2030**

The ONS also uses improved mortality rates in their mortality projections. As this captures a significant improvement annually, this assumption was adopted in the model. Based on the extended framework, the further development work has been carried out, using appropriate data. The projection results for Leeds both at the level of LA (Local Authority) area and small areas of wards have then been analysed. We found that at the LA level, with the extension of the model, the revised model produces better results than the previous version.

In the analyses of revised model results by demographic processes, all three components of change indicated improvements in results. The trends of reductions in mortality, higher fertility rates in women aged over 45, continued increasing migrations both internally (UK) and internationally in the longer term are all consistent with the ONS projections. However, the rich details provide by the revised MSM presents a range of ways to interrogate the projection results. As an example in Section 5.5.3, analyses of the migration projection results by age, sex, flow and small areas are used to demonstrate the capacity to allow comprehensive investigations to gain a fuller picture than the aggregate projections provided by ONS model.

The disaggregated results can be re-aggregated into any higher aggregations to provide an overview of population structures. In this chapter, the projections have been re-aggregated into the LA level to enable the comparison with ONS projections. The revised model results indicate that population of Leeds continues to grow till 2030, instead of shrinking as in the previous model projections. The revised model projections indicate an ageing trend in Leeds population. Such patterns are consistent with the ONS projections. However, the MSM result indicates a more serious ageing trend than in the ONS projection as discussed in Section 5.5.

To further explore the issue and find out the spatial impact, we also analysed the population changes in small areas. Cookridge and Headingley have been selected as the examples again in order to compare the patterns in different types of small areas, as well as to compare to the results of the two wards, generated by previous MSM. In general, the findings are similar to

what we have found at the level of the whole Leeds. Although the populations keep growing till 2030, the ageing trends of the population in small areas have been exaggerated. In more established ward such as Cookridge, such trend may be plausible. However, in wards such as Headingley, it is an issue. As we know that university student migration has a great impact on local population structure in Headingley. Graduates move out and freshmen students move into the student accommodations in Headingley each year. Due to such replenishment, the local population in this small area of ward stays younger than in others. Therefore the results analyses indicate that the revised MSM still fails to capture the subtlety of the local migration patterns in small areas. This issue is also found in the migration analysis. In Chapter 6, more in-depth analyses of this issue will be discussed and a hybrid modelling approach is adopted in the attempt to capture the subtlety of the local migration patterns through the modelling of different behaviours of sub-populations.

## Chapter 6

# Experiments with ABM

## 6.1 Introduction

As described in previous chapters, all demographic processes in this dynamic microsimulation are modelled separately. However, different processes interact with each other throughout the simulation. As a result, one individual process can have important impact on another. Sometimes there can even be interdependency between individual demographic processes. As discussed in Chapter 3, migration is a much complex demographic process than ageing and mortality. On one hand, marriages (to join an existing household or start a new household) and serious health changes can often lead to migrations (to live with other family members or for access to the medical services); on the other hand, large volume of migrants moving in and out of small areas can have critical impact on the demographic compositions. As a consequence of this interaction, there can be significant changes in the outcomes in other demographic processes, especially at the level of small areas. For instance, in the scenario of a large volume of young women at child-bearing age moving into a small area can have a critical impact on local birth rates. Therefore changes in the

magnitude or composition of migration can also lead to changes in outcomes from the fertility modelling process. Moreover, the migration process is much more complicated than other demographic changes, where patterns of the movements, interactions and behaviours of the migrants vary in different sub-populations. For instance, compared with university students have a distinctive migration pattern in small areas than the rest of population. From the initial analyses reported in Chapter 4, it is found that the migration process has a profound impact on other demographic processes and captures the subtlety of local migration patterns that play a vital role in projecting population changes in small areas.

Individual based models such as MSMs (MicroSimulation Models) provide realistic disaggregated information that is often vital for modern policy problems. However, MSM has been criticised for being less strong in modelling interactions between individuals and individual behaviours. MSM also requires realistic micro-data and struggles in situations where such data are not available. In a population MSM, the statistical simulation processes also make it less flexible when modelling interactions and behaviours of various population groups. However, as it has been pointed out in Chapter 4, to appropriately model migration at the level of small areas, it is important to capture different migration behaviours of sub-populations such as university students, which have a significant impact on the local population composition. However, as described in Chapter 5, despite the efforts of further development and refinement of the MSM, the subtlety of the local migration patterns remains an issue for the revised model.

ABM (Agent Based Models) can strengthen our understanding of the interactions between people and their behaviour by modelling demographic processes through interactions with other agents and/or the environment that they live in and modelling actions according to their unique built-in rules of behaviour. The flexibility of being able to construct heterogeneous agents and their rules makes it particularly helpful when there is a knowledge gap or data are unavailable. The combination of MSM and ABM provides a new

approach to enhance complex social modelling. In our study, we attempt to provide better groundwork to facilitate the policy and decision making for the UK population through a hybrid model that combines the strength of the two complementary techniques.

Although another refinement of the MSM could be to include further disaggregation of the model parameters by age (to separate university students from the rest), this can soon make the models unwieldy as explained in Chapter 3. More importantly, due to the statistical nature of the MSM, it limits the capability of the modelling distinctive migration behaviours of the university students. As explained in Chapter 4, the issue will remain (Section 4.8.2). ABM uses an alternative approach that can model individual behaviours through multiple agents. In an ABM, each agent follows their built-in rules and makes decisions and takes actions on the basis of the rules and the knowledge gained through interactions with each other and the environment they live in. Through such interactions, simple and predictable local interactions can generate familiar but unpredictable global patterns. Therefore ABM provides theoretical leverage where the global patterns of interest are more than the aggregation of individual attributes.

As described in Chapters 2 and 3, MSM and ABM have been closely allied, but they also have significant differences in terms of:

1. Programming approach: MSM uses a more traditional programming approach that uses a central control to manipulate data flow, i.e. a central manipulation mechanism in the system handles the input data, processes them and outputs the results. ABM adopts object oriented programming (OOP) approach that models using objects with their own attributes and methods. All objects process the input data and produce results according to their built-in methods (Russell and Norvig, 1995).
2. Computing method: MSM is a more statistical/probability driven simulation where each event occurrence is based on relevant probabilities, whilst ABM is more of a rule-based (built-in intelligence) simulation



process. These rules can, but do not always have to incorporate probabilities or equations (Billari *et al.*, 2003).

3. Interaction and behaviour modelling: Due to its modelling approach, MSM lacks useful means for interaction and behaviour modelling. Although it is possible to incorporate spatial interactions using the probabilities generated from a spatial interaction model, it is inflexible to model the interaction changes between one individual with another and the environment that they live in. Whilst each agent in ABM makes their decisions and takes actions on the basis of the agent interactions with one other and their environment, according to their individual built-in rules. This provides a very flexible way to model the heterogeneous behaviours of different agents (Conte and Gilbert, 1995).

Another approach needs mentioning is CA (cellular automata). CA is similar to ABM in the sense that they are both object-oriented and use agents. However ABM provides agents with more flexibility in terms of mobility, interactions and behaviours, as CA agents are confined within cells and the simulation is driven by the changes of the states of each cell through a set of unified rules. Therefore it is difficult to model the migration process using a CA, where individual movements, interactions and behaviours are of vital importance. ABM on the hand provides great flexibility in modelling such individual behaviours by allowing changing individual agents' rules easily and without disturbing the simulation of other individuals or sub-populations.

Although there are limited examples of using the ABM approach in demographic models, Billari *et al.* (2003) consider ABM as a promising approach to help improve our understanding of demographic behaviours. ABM can provide us an alternative means to study demographic processes as the outcome of interacting agents. It is also useful in explaining individual behaviours by taking into account both micro and macro factors. For instance, agents in an artificial society can follow their built-in rules while subject to the macro behaviour rules (e.g. conventions, institutions). The explanation of behaviour is based on simple propositions about

individual behaviour, but can produce complex situations and feedback at the macro level. By focusing on dynamics of the population instead of the equilibrium, ABM is better suited for modelling demographic processes. The question whether those dynamic changes are due to compositional changes or changing behaviour rules of the individual agents can also be studied within the framework of ABM.

Migration is a complex demographic process where interactions and behaviours play an important role. Using ABM, individual activities and the diversity of migration decisions leading to the observed complex migration patterns can be simulated in detail. Some attempts have produced fruitful outcomes. Espindola (2006) analysed the rural–urban migration using ABM, where the migration of workers is modelled as a process of social learning by imitation. Social systems often demonstrate an emergent property or “emergence”. The emergent property of a system is a structure, a pattern or a property that cannot be seen in the individual units of the system at the microscopic level, instead it emerges from the interaction of the units and is visible at the macroscopic level (Casti, 1997). This view describes the essence of the emergence and is generally shared (Damper, 2010). One simple example of the emergency may be the locomotion in an animal. As emergent properties of the model, transitional dynamics are observed with continuous growth of the urban fraction of overall population towards equilibrium. While Loibl and Toetzer (2003) studied urban sprawl patterns through modelling suburban migration and residential area occupation, distinctive migration behaviours of households with varying socio-economic status have been simulated in an ABM. Makowsky *et al.* (2006) built an ABM to simulate crisis-driven migration of agents within a multi-ethnic population. This study reveals that cultural networks temper an agent’s security calculus, with strong social ties dampening the human security dilemma.

The above studies demonstrate that ABM is very useful in understanding the functioning of complex models and in modelling particular phenomena that are not necessarily mathematically tractable, which is often the case in the

modelling of large complex social systems. ABM to a degree provides a link to the theory and knowledge gap in demographic modelling. ABM is also a much more flexible way to model various behaviours and dynamics during the migration process, as different type of agents can have their own rules. Based on such rules and the information they gather from the environment they live in, agents can make different decisions and act on it accordingly. On the other hand, MSM is a widely applied and tested approach in demographic modelling as discussed previously. While ABM is useful in modelling features in the model where knowledge and theory is lacking, MSM can provide important statistical mechanisms that ensure the similarity between what is predicted and what is actually observed in the gathered data. For instance, the probabilities used in the MSM that are calculated using the empirical data can provide a guideline for the population evolution patterns. This study is one of the first attempts to bring the strength of the two approaches together to provide a better demographic model.

In this chapter, two experiments have been carried out, using the ABM techniques: (i) modelling university student migration, and (ii) in the module of mortality of the MSM. ABM was not applied for the rest of model, mainly because ABM is best suited modelling heterogeneous interactions and behaviours. However, it is less strong in theoretical basis in terms of applications and decision support. On the other hand, MSM is widely used in policy applications. The list-processing power of MSM also provides the strength in computing efficiency to simulate a large complex system. The details of the experiments are explained in the following sections and the results generated by the hybrid model are also compared with the results from the pure MSM.

## 6.2 Student migration experiment

### 6.2.1 *Pure MSM*

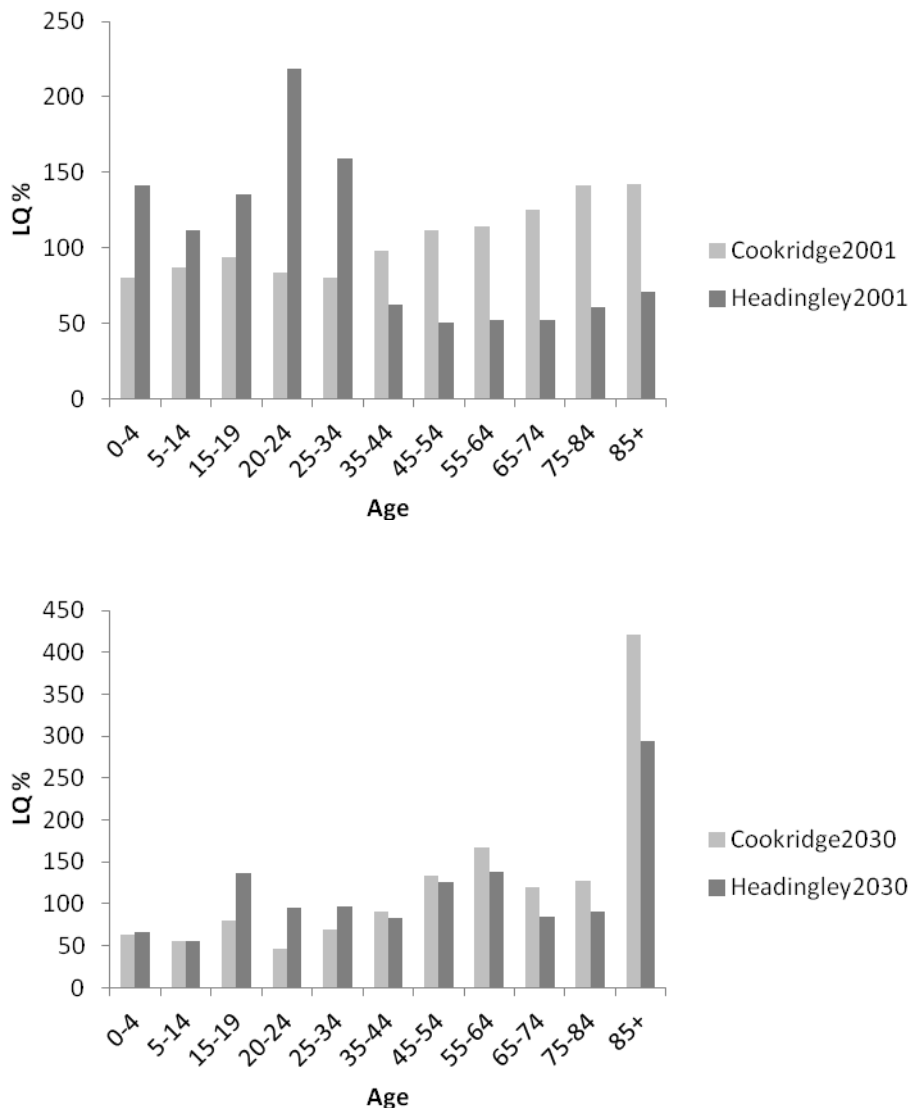
In Chapter 4, the way in which the different demographic processes are

modelled has been explained and the initial results have been analysed. A few issues have been found. Of particular importance, it is found that the MSM failed to capture the characteristics of the student migration. Student migration, however, has an important impact for population changes in Leeds. Each year, due to its strategic location in England, students from both the north and south have been attracted to Leeds for their higher education. Such large numbers of student migrants have a great impact on local population structures, especially in those parts of the city that are popular residential areas. These distinctive patterns need to be captured in the model in order to reflect the local population changes in Leeds. Although ward based migration probabilities are developed, however, they are not student status specified. Even they are, MSM cannot provide the flexibility to model the distinctive student migration patterns, due to its statistical nature: students are highly mobile during their stay in Leeds, but mostly in selected areas; they only stay in Leeds for a certain period of time and most of them leave upon completion of higher educations.

Further work of development and refinement on the model has been carried out and improvement has been discussed in Chapter 5. However, although the projected population stopped shrinking, the ageing trend seems to be exaggerated at both the levels of the LA (Local Authority) and the ward (Sections 5.5.1 and 5.5.2). A set of LQ (Location Quotient) analyses are used in Chapter 4 to measure the concentration of age groups within each geographical area at a point in time to reveal the population characteristics in small areas with the reference to the aggregated population characteristics at the LA level (for more details of the LQ, see Chapter 4). Here the LQ using the projection results from the revised model within two Leeds wards, Cookridge and Headingley, are analysed to provide consistency in comparisons with the previous analyses of local characteristics (Figure 6.1).

As explained previously, Cookridge is a more established and peripheral suburban area whilst Headingley is an inner city area close to the major universities with accommodation that is popular with university students.

Two features of these distributions found in Figure 6.1 are most worthy of further comment.



**Figure 6.1 Location quotient analyses of ward level MSM projections**

Firstly, we have experienced persistent problems in simulating the behaviour of student populations in the university areas where using the MSM. In this case, a typical tendency is found for the model to have difficulty in reproducing the strong peak introduced by student migration into the Headingley area, while student migrants are also retained in greater numbers as they get older. As a result, the LQ analysis of Headingley projections reveals a strong peak in the young population in year 2001, which disappears in projection in 2030. Comparing results of each ward in 2001 to 2030, the projections suggest that the two ward populations have

become similar over time, in terms of their age profiles. As observed in reality, the annual replenishment of young migrants formed mainly by student populations largely happens in a handful wards within Leeds, providing easy access to the universities and suitable student accommodation. The analyses indicated that the MSM simulations may have caused this, where student migrants are not differentiated from the rest. As a result, migrants have been distributed all over Leeds and allowed to settle in the area over a 30 year period. As the older populations are less likely to migrate, a large number of older people have been retained in the area. The behaviour of student populations in the model therefore is a key concern for which we seek a remedy by using a different modeling approach.

Secondly, if we look at overall population in Leeds, then we can again see the importance of the migration of students and young people into the area. The biggest population growth is in the age groups from 16 to 49 in the ONS projection, reflecting students and young people migrating into Leeds (Figure 5.2, Chapter 5). The other effect worth noticing in the ONS projection is significant growth in the elderly population over 75 in Leeds population (Figure 5.2, Chapter 5). The LQ analyses in Figure 6.1 also demonstrate the ageing trend, although exaggerated. The phenomenon of an ageing society has received much comment in both the academic literature and popular media. It is noted that the treatment of the linked questions of ageing, life expectancy and mortality is another vitally important question in projecting the population of a city and its spatial distribution. Mortality provides the focus for another experiment that is chosen to demonstrate the advantage by linking MSM to ABM in demographic models at a fine spatial scale. Of particular importance, it will provide the model with the capacity of tracking personal history during the simulation steps and the difference caused by this is demonstrated in the experiment. The following sections will describe the two experiments separately.

This chapter will focus on the limitations of MSM on the small area level. As discussed in Chapter 5 and here, the limitations of the model found in the

first version of model remain despite the further developments described in Chapter 5. In suburban areas such as Cookridge, the ageing over time may be more plausible. However, in wards where student migration has a great impact on local population structure such as Headingley, the MSM failed to reproduce the student population renewal. The distinctive pattern of the large volume of student migration into and out of the “student” wards each year has disappeared. As a result, the ageing trend of population in these areas has been greatly exaggerated. This indicates that the subtlety of the local migration behaviours has not been captured successfully by MSM (Figure 4.16). The following section will describe how attempts are made to address this issue through the design of a hybrid model combining MSM and ABM features.

The pure MSM models individuals and households within Leeds, which draws on rich attributes of both individuals and households to model and simulate the social systems within an urban environment. MSM provides a means to predict the future characteristics of the modelled micro units and is built with validation ambitions, but it does not model individual interactions and is not based on an explicit behavioural model. When the MSM is run with the 2001 Census migration probabilities, students are found to continue to live in their student neighbourhood. Some may do so in real observation. Most, however, move elsewhere to start their future or jobs.

MSM relies on appropriate microdata to produce reasonable simulation results. However, the empirical information about the migration behaviour of university student is inaccurate. There are two main reasons for this: the problem with base population estimation of students (student status not constrained on) and the dubious nature of 2001 Census data on student locations. As explained previously it is impractical to use too many constraints during population synthesis. More importantly, the Census fails to record comprehensively the sequence of usual residence of students before, during and after their programmes of study. The question of usual residence is only asked at the point of time when the census took place, but the actual address can vary at one of the above three the points of time.

Therefore there is a flaw in the currently available microdata. In this study, the MSM uses probabilities of out-migration to other wards and to other LAs outside Leeds that are captured from the 2001 Census migration tables. In the 2001 Census, the migration question was as follows (Figure 6.2).

**14 What was your usual address one year ago?**

- ◆ If you were a child at boarding school or a student one year ago, give the address at which you were living during the school/college/university term.
- ◆ For a child born after 29 April 2000, ✓ 'No usual address one year ago'.

The address shown on the front of the form

No usual address one year ago

Elsewhere, *please write in below*

--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

**Figure 6.2: Migration question (Q14) in UK census form, 2001**

Source: Adapted from

[https://www.census.ac.uk/Documents/CensusForms/2001\\_England\\_Household.pdf](https://www.census.ac.uk/Documents/CensusForms/2001_England_Household.pdf)

Adding to the answers to this question age and gender information about individuals, we obtain correct tabulations of migration status against age and gender. The migrant counts are divided by the 2001 Census population shifted back 1 year in age to obtain an estimate of the probability of migrating at an annual interval. For most ages their probabilities provide an accurate measure of the population who migrate, but for ages 20-24, the ages which many students leave their higher education, they do not. This is partly due to the conceptual confusion between the student term-time residence and the parental residence. Firstly, there is some conceptual confusion in the minds of student respondents to the Census between their term-time residence and their parental residence. The instruction on the 2001 Census form (Figure 6.2) was to record their term-time residence as their usual residence, but it is probable that many student respondents failed to follow this advice either for their residence at the time of the census or their residence one year earlier. Unfortunately, there was not a Census Quality Survey following the 2001 Census which might have revealed how big this problem was. Lacking appropriate microdata of the student migration in small areas, a hybrid modelling approach was adopted to bring



in the ABM features in modelling heterogeneous behaviours. In the ABM, a set of rules was derived from the way the education programmes works so as to obtain a better view of the student migration process.

The dynamic model simulates continuously from 2001 to 2031 to capture the evolution of the population patterns over time. Here the results for 2031 are presented as normally there are more variations over a longer term in projection results than currently observed. Observed data in 2006 is chosen for the comparison because it is the most up-to-date data available after the simulation starting point of 2001. There are three reasons that we chose to model the population changes from year 2001-2031. One is because there is a lot of need for long-term projections; two is because 2001 Census was the most up-to-date census available at the time of study; finally the alternative to project for a period when actual data are available, the micro populations will need to be reconstructed for 1991 and then roll the model forward to 2001. The student migration data from the 1991 census is a lot worse than that from the 2001 Census, where students were asked to record the usual residence and term-home residence (they could coincide). The migration tables were based on usual residence, so the migration data were even worse than 2001 Census. There is a matrix of usual residence against term residence for local authorities but this only helps with the in-migration to HE, not with out-migration from HE. Due to such difficulties, we have decided not to make a projection for a period for which there is already actual data available.

The problem of modelling student migration is quite well-known in small area demographic modelling, due to its distinctive migration patterns and lack of information on student migrants (Baryla and Dotterweich, 2001; Fotheringham *et al.*, 2004; Koser and Salt, 1997; Rees, 1994). Frequent movement in small areas during their period of university study and the confusion of the definition of term-time and home address all make the production of useful student migration micro-data difficult. Generally speaking, university students tend to only reside in certain areas, mostly around the universities they attend, during the period of their study. Most of

them then leave while other new students move in, instead of growing old with the rest of population in the area. Due to the replenishment of the student population each year, the population in such wards stays younger than that in other wards.

As an ABM is very flexible in terms of constructing heterogeneous agents with different built in rules, it is ideal to experiment on sub-populations that are behaving differently from the rest of population. Section 6.2.2 describes the experiment with an ABM for the student migration process using hypothetical rules in a similar way as in the residential segregation study by Schelling (1971).

## 6.2.2 *Hybrid model combines MSM and ABM*

### 6.2.2.1. *Combination of the two methods*

ABM involves modelling individual interactions and behaviours due to the great flexibility of constructing agents that follow different rules. ABM simulation often begins with values taken from a uniform random distribution. However, the choice of initial conditions can affect the output of the model and a uniform random distribution can be a poor choice. On the other hand, MSM has been known for its capacity to work with real data to simulate the real world phenomena. It seems that neither MSM nor ABM alone is sufficiently comprehensive to model complex social systems and provide a good basis for forecasting. Therefore, recent studies propose a hybrid approach to overcome shortcomings of both ABM and MSM. The hybrid approach will enable us to move towards a more comprehensive way of study of micro unit interactions and behaviours on both microscopic and macroscopic level, along with an empirical validation ambition (Mahdavi *et al.*, 2007; Murphy, 2001; Hassan *et al.*, 2008).

Whilst such proposals and the design of such systems have been developed (Molnar and Sinka, 2007), few models have actually combined the two. ILUTE simulates the activities of individual objects (agents) as they evolve over time. These include persons, transportation networks, the built

environment, firms, the economy and the job market. The simulator evolves the state of the urban system from a specified base month to a specified target month. At any time, the simulation can be branched to test various policy alternatives (Salvini and Miller, 2003). *LABORsim* is a dynamic aging, discrete-event, probabilistic agent-based MSM of labour supply, where agents are used to represent individuals, institutions, etc., each with their own variables and methods to increase the modularity of the MSM and the transparency of the code (Leombruni and Richiardi, 2006).

ABM can model individual interactions with each other and their local environment easily. It does not necessarily require high quality micro data or well formulated equations. This is particularly useful when appropriate data on such behaviours are not available. Agents can also store their personal history and retrieve such information in future easily. This model tried to experiment with three useful features of ABM to enhance the capability of MSM in modelling individual behaviours to provide better groundwork for demographic forecasting. Simple rules have been developed for individual agents and through their interactions at the microscopic level, demographic patterns can be observed at the macroscopic level to facilitate demographic forecasting.

This model exploits the well-known strength of MSM for list processing for the scale issue of the model. Ageing, fertility and mortality can all be simulated easily by this means. However, once we start to introduce concepts of movement and interaction, the standard MSM process begins to struggle. Forecasting the behaviour of the student migrants at the small area level proves to be difficult using a pure MSM approach, due to the distinctive migration pattern of this sub-population (Wu *et al.*, 2008). Despite its many strengths and advantages, a MSM of a spatially distributed population depends on good data about the important demographic transitions which are experienced by individuals. Because of this dependency, MSM is less strong in modelling individual behaviours where realistic micro-data are unavailable. The structure of MSM is also found less

flexible to accommodate various rules for different individuals. An ABM can be introduced into the MSM to strengthen such aspects of the model.

It is generally recognised that agents are an effective way to represent individual entities that move around and interact with one another and with their environment based on built-in rules (Billari *et al.*, 2003). Such rules can be very simple and flexible at the individual level, but simulation through a large number of agents can reproduce complex social phenomena at the macroscopic level, with leverage for unexpected patterns that are not necessarily intentionally described in the rules. For instance, in Schelling's segregation model, individual agents only follow the rules under the assumption that people prefer living with "similar types" to themselves. However on an aggregated spatial level, residential segregation patterns are revealed (Schelling, 1971). Such a property makes it possible for an ABM to reflect the characteristics of a society more realistically and makes it a technique complementary to MSM in modelling a complex social system. Mobile agents in an ABM interacting with an environment and other agents look like an ideal partner to static, self-contained individuals in an MSM, especially in more complex demographic processes such as migration, where interactions and behaviours play an important role. In his famous study of residential segregation, Schelling's (1971) analysis posits a relationship between individual decisions to move and the composition of neighbourhoods. This process can easily be represented within an agent based migration model. Therefore an agent-based mechanism can be used to in this way to help maintain geodemographic discrimination within urban population projections.

#### *6.2.2.2 Construction of the ABM*

In this chapter an investigation of the usefulness of a hybrid modelling approach has been carried out in a series of experiments, where the MSM is combined with ABM to allow the flexibility to explore evolutions of demographic structures in various scenarios.

A sub-population of university students has been formed by selecting relevant records by small areas on the basis of two criteria: age and full-time student status. Empirical data from the Higher Education Statistics Agency (HESA): admission data 2004–2005 has been used to make the assumptions of age for this sub-population. As most university students start at age 17, 18 or 19 for their 1st degree study, most people will finish their PhD program before 30. The transitional probability to a higher degree of study is generated by comparing counts of students in studies for different degrees in the HESA admission data for the current and the following year, e.g. number of first degree graduates during the period of 2004–2005 and number of masters students 2005–2006. Therefore, individuals that are recorded as ages between 17 and 29 and “full-time students” have been selected to form the baseline population of students for the initialisation.

Four types of student migrants have been identified in this study due to their distinctive behaviours during the process of migration. For instance, the first year students tend to stay in university accommodations. From the second year on, as students are more familiar with the area, they move out and find privately rented accommodation that are close to the university where they study, often with their fellow students in the same area. Hence the four groups of student migrant groups by migration patterns are:

- first year undergraduates;
- other undergraduates (years 2 and 3);
- masters students and
- doctoral students.

The number of housing vacancies and their spatial distribution comes from three sources:

- Accommodation at University of Leeds:  
<http://www.leeds.ac.uk/accommodation/overview.html>
- Accommodation at Leeds Metropolitan University:  
<http://www.leedsmet.ac.uk/visiting/accomm/index.htm>

- Privately rented accommodations from the Unipol database at: <http://www.unipol.org.uk/leeds/>.

Unipol is a charity working nationally in student housing and Unipol Leeds is the largest student housing provider in Leeds.

One reason for using student accommodation information is that we have seen the effect of exaggerating of population growth in student areas such as Headingley using MSM without a housing stock constraint in Figure 5.4. Another reason is that the base population is not constrained by student status. The dubious nature of 2001 Census data on student locations means that we cannot provide accurate information on student locations. Using the student accommodation data, we can adjust the distribution of students in the baseline population. Then students are assigned to appropriate locations according to the probabilities of a student living in that area, which are calculated using the counts of current vacancy and total student accommodations in area.

The ABM of student migration simulation is based on the following assumptions/hypotheses:

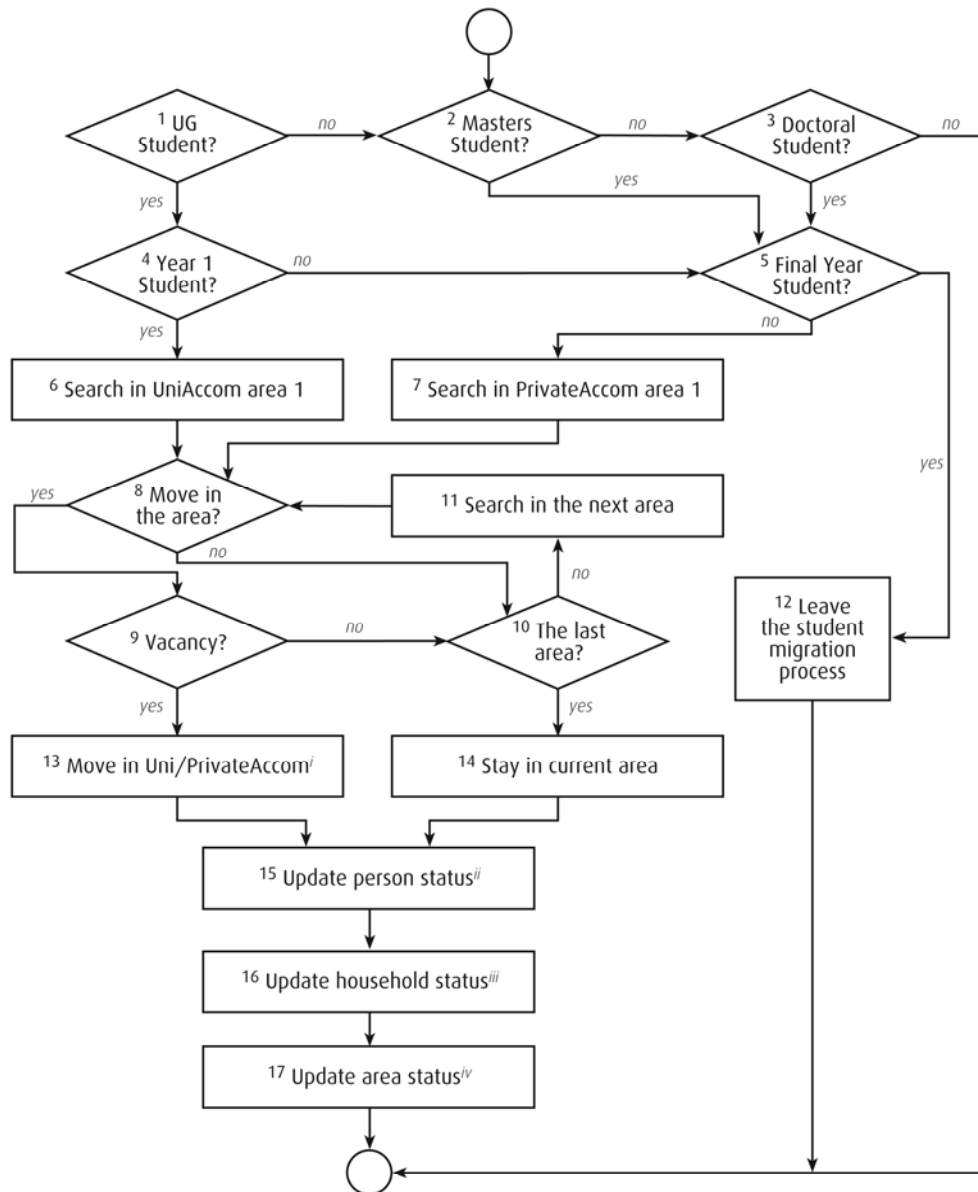
- students prefer to stay with their fellow students – choose to stay in the area where student population resides and
- students prefer to stay close to the university, subject to the availability of accommodation.

Due to the simplicity of the model at this stage, price and other attributes such as housing quality are not modelled. Also the changing supply of vacancies is not modelled, i.e. new stock other than places available from the previous year. The above assumptions are based on the behaviour in housing choice that has been recognised in previous studies. Many studies including Schelling's (1971) famous model of housing segregation have suggested that people have a tendency to live where "similar" types of residents are. The criteria can be based on characteristics such as ethnic group and socioeconomic class. Here we just focus on the different renting

behaviour between the four types of university students. Student areas are preferred, but are not the only areas students can live in. There is also a limit on vacancies for each area. If students cannot find a vacancy in preferred area, they will stay where they are. However, using “term-time address” to define students means students are those taking degree courses away from their parental homes, although confusions when answering this question resulted in a small number of students living in far suburban locations (see section 6.2.1). It is difficult to identify university students and the definition used here is not exhaustive. However, it is not the purpose of this model to track accurately mature students, distance-learning and part-time students. Based on the above assumptions, we therefore apply the following general rules to the “student agents”:

- each group is allowed a set number of years to stay in an area (based on their study period at the universities);
- students stay close to their university of study, subject to housing availability and
- they are excluded from the processes of marriage and fertility during their study (Such simplification may result in lower marriage and fertility rates in an area where there is a large proportion of students. In a future model, this can be improved by re-adjusting fertility probabilities for the rest of the population at risk in area).

The typical student migration process that a student agent experiences each year during their stay in Leeds has been illustrated in Figure 6.3. Depending on the type of the students, their rules vary slightly. For example, a year two undergraduate student can stay in the area for two years. He/she then can have the chance to continue studying towards a master’s degree for one more year or leave. A masters student can stay in the area for one year and then continue with a doctoral study for three more years or leave. The typical interaction between the agents in this model would be finding the fellow students in order to move to the area they reside in and the interaction with the environment would be checking if there is a vacancy in that area. If the searching result is negative, the agent/student moves on to



- i.* Year 1 students in University accommodation and other students in private accommodation;
- ii.* type of students, year(s) stayed in Leeds and household ID etc.
- iii.* household size, household reference person, number of various household members etc.
- iv.* number of people in the area, number of household in the area, vacancies in area etc.

**Figure 6.3 Student migration process**  
Source: Wu *et al.* (2008)



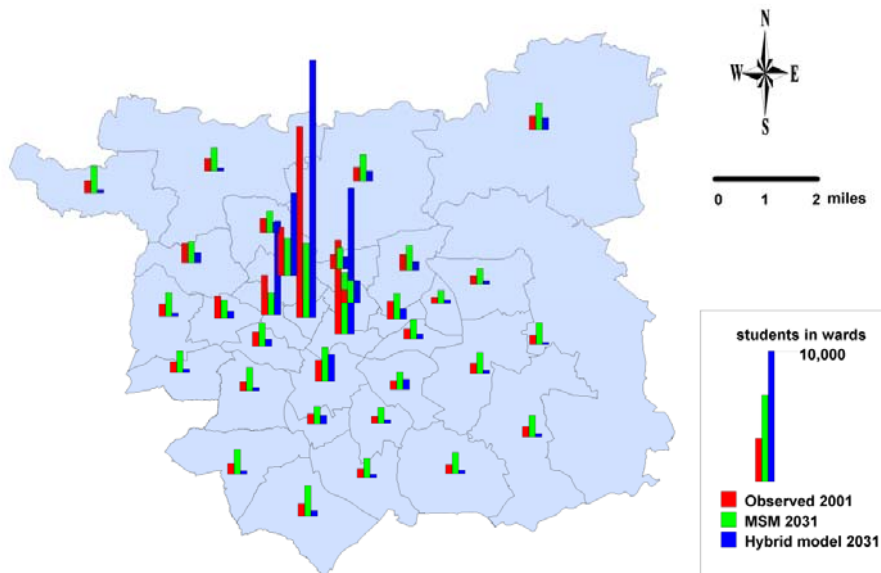
the next area. At the end of the search, the agent without success then stays in the area for one more year. The student agents will carry out the similar process (depending on the updates of the student status) if they are going to stay in Leeds for at least one more year. At the end of their study, they will be marked as final year students and leave the student migration process and join the general migration process (Figure 6.3). The current ABM does not model the changing supply of vacancies from such as new-builds and demolishes and this can be improved in the future versions of the model.

### *6.2.3 Comparison of the simulation results using pure MSM and the hybrid model*

Using only the pure MSM, the model does not differentiate the student migrants from other migrants and they join the general migration process as described in previous sections. By applying different rules to individual “student agents”, the hybrid model presents a better reflection of the observed student population in wards of Leeds. The hybrid model constructed the agents to represent the student migrants within the MSM. Each student migrants then follow their own built-in rules in the migration processes in the microsimulation, while the rest of the population still go through the same processes as in the pure MSM. The advantage of the hybrid model is that adding the ABM features to the sub-population of student migrants does not need to substantially change the system structure. Other simulations of demographic processes remain the same as in the pure MSM. All individuals’ simulated attributes get updated and records of students who finish their higher education and leave Leeds are removed at the end of each year. Then records of the new students are added in the system to create the baseline population for the next year’s simulation.

After 30 years’ simulation, the simulation results of the student population using the pure MSM and using the hybrid model have been compared to the observed distribution of the student population. From Figure 6.4, it can be

seen that the hybrid model provides a much better reflection of the observed student population concentration around the city centre close to the universities, instead of students almost evenly scattering around the whole city in the MSM. Most students will leave upon completion of their study in the hybrid model as new students come into the area each year. In other words, students are no longer growing old together with the rest of the population in the suburban areas as simulated in the pure MSM. The distribution patterns of the student population within Leeds city can be observed clearly in the map (Figure 6.4). Comparing the results in the map, it suggests that the hybrid model results provide a much better reflection of the observed population in reality than the pure MSM simulation results.



**Figure 6.4 Migration experiment: students in small areas 2001 and 2031**

Source: Wu *et al.* (2010)

A formal comparison of model results from the MSM and ABM is shown in Table 6.1. The level of similarity between four groups: the total Leeds population and the student population, as observed in the 2001 Census data, the distribution of students from the MSM and the ABM. The comparison is presented as an index of dissimilarity (IoD).

The IoD is a measure of the evenness with which two groups are distributed across the component geographic areas that make up a larger area. IoD has

been used as a means for comparing the distribution of different social and ethnic groups in cities (Rees and Birkin, 1983) and reflects the average variation in concentration between areas. However, the IoD can be used to compare the amount of spatial segregation or spatial dissimilarity between two population (or ethnic/racial/immigrant) groups. The index ranges from 0 to 100 (when reported as a percentage), with 0 meaning no segregation or spatial disparity, and 100 being complete segregation between the two groups with no spatial intermingling. If all the students were in Headingley ward and the rest of the population evenly distributed elsewhere then the IoD would be 100. If the students are spread uniformly across the city then the IoD will be zero. The mathematical expression of this relationship is:

$$IoD = 100 \sum_i \left[ \frac{X_i}{\sum_i X_i} - \frac{Y_i}{\sum_i Y_i} \right] / 2 \quad (\text{Equation 6.1})$$

where  $X_i$  is a count for the small area population of ward  $i$ , and  $Y_i$  is the count for students. Other distributions for  $X$  and  $Y$  are substituted as appropriate.

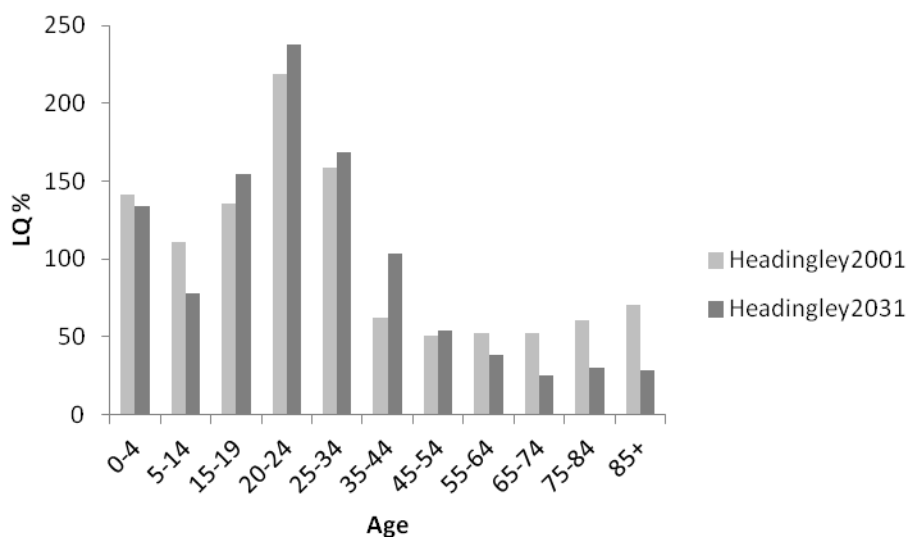
**Table 6.1 Index of dissimilarity for student populations**

	<b>Leeds population, all ages</b>	<b>Students, aged 17-29 Census 2001</b>	<b>Projected students, aged 17-19, MSM, 2031</b>	<b>Projected students, aged 17-19, ABM, 2031</b>
<b>Leeds population, all ages</b>	--	<b>61.6</b>	<b>12.6</b>	<b>62.3</b>
<b>Students, aged 17-29 Census 2001</b>	--	--	<b>49.7</b>	<b>11.7</b>
<b>Projected students, aged 17-19, MSM, 2031</b>	--	--	--	<b>50.7</b>

Source: Wu *et al.* (2008)

The results in Table 6.1 show that the student population is indeed highly concentrated against the general population of Leeds (IoD = 62). This is reflected well in the ABM (IoD = 62) and much less well in the MSM (IoD = 13). The same pattern is reflected if we compare the student population with the models: in this case we find a high level of dissimilarity to the MSM model (IoD = 50) and a much lower level of dissimilarity to the ABM (IoD = 12). Last, it is found that the ABM distributions are rather different to the MSM (IoD = 51) (Wu *et al.*, 2008).

The hybrid approach in modelling the student migration also results in improvement in projections of the local population profiles. Here the LQ analysis is used again to provide the same basis for comparison with the previous results (Details of LQ technique can be found in Chapter 4). The analysis results are presented in Figure 6.5. Here we can clearly see that the hybrid model results in year 2031 retained the similar local population characteristics as that of Headingley in 2001. Compared to the MSM results illustrated in Figure 6.1, the hybrid model provided a local population profile that is much closer to the real observation.



**Figure 6.5 LQ analyses of Headingley projections by hybrid model: 2001 and 2031**

## 6.3 Mortality experiment

Encouraged by the findings from the student migration experiment, another experiment has been carried out using ABM. This time the most straightforward demographic process in the MSM, mortality, has been chosen as a contrast with the student migration experiment, but also to further assess the impact of migration. Details of the comparisons of the two approaches in modelling mortality processes are discussed below.

### 6.3.1 *Pure MSM*

Modelling the mortality process in the dynamic microsimulation model (MSM) is pretty straightforward. All individuals are tested to see whether they survive each year. This is achieved through a Monte-Carlo simulation on the basis of individual probabilities that are calculated to correspond to single year of age and location (the ward where the individual currently resides) for each sex of the simulated individuals. Based on the outcomes of the determinations, the non-survivors are then marked “dead” and their records are removed from the system and the rest of the population forms the baseline population for the simulation of next year.

One of the advantages with simulating geographically identified populations is that the influence of the local context on the individual characteristics can be taken into account to a certain extent. However, sometimes it is not only the current small area that influences behaviour but the places individuals originally came from or used to live in often also play an important role. Such history can have greater impact in some demographic processes than others. In terms of the mortality rates, it is known that immigrants from other countries often show significant differences from their descendants in the destination countries. Another simple example of the internal migration may be: if a person has worked as a miner all his/her life, his/her mortality rates should not suddenly change a great deal just because he/she retired to a pleasant residential area. Norman *et al.* (2004) found a trend of healthy migrants moving away from the most deprived wards. Consequently the

accumulation of survivors occurred through migration to the least deprived areas with a net loss of migrant survivors from the most deprived locations. In order to model such impacts, there is a need to assess each individual's behaviour using information from their personal histories.

However, it is not easy to manage this in MSM. Due to its statistical nature, it is difficult to trace a single individual's history without affecting the whole simulation process. The information of all individuals is managed collectively by the central control of the system. This means the simulation cannot or will be difficult to start without the completion of checking one single individuals' history. Due to the volume of the data, it is impossible to carry all life history information such as migration origins and destinations in machine memory. Instead they are overwritten and output annually, which means to check the information in MSM, the simulation has to be stopped while tracking the information from relevant output files. If the system has to come to a halt as soon as it has to check some individuals' history and update relevant probabilities, this will lead to very inefficient processing. Obviously it will be much easier and more efficient if individual history can be accessed and updated locally and individually in the system.

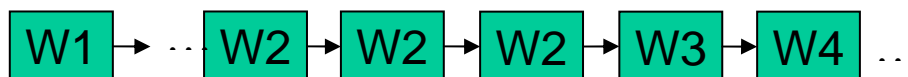
### 6.3.2 *Hybrid model combines MSM and ABM*

An ABM can meet such requirements with much more ease than an MSM, as the object-oriented agents can simply have a function to retrieve specified information from their own history that they carried along (see Chapter 3, section 3.3.2 and section 3.3.4). Based on the hypothesis that mortality/morbidity probabilities depend not only on the current personal and environmental conditions of the individual, but also depend on their personal histories, three scenarios have been explored of mortality projections in this hybrid model: (1) mortality projected on the basis of current residence location; (2) mortality dependent on first residence location in the system/birth places and (3) mortality dependent on personal migration histories. In the first scenario, all individuals are simulated using a MSM. Their survivals are determined against an age, sex and location

specific mortality probability generated on the basis of local information about the current location of residence. A hybrid approach is used in the second and third scenarios, where agents carry their own histories along with them and have the ability to check on such histories. In the second scenario, the survivals are determined on the basis of the mortality rates of individual first residence location/birth places. In the third scenario, we tried to model the impact of personal migration history on mortality. Mortality is projected on the basis of the mortality rates of the area where the individual stays the longest.

Take a simple example of an individual migrant in Leeds, who we can call migrant A and his migration history is illustrated in Figure 6.6. If his origin is ward 1 and in the last five years his migration destinations are: ward 2 in year 1, ward 2 in year 2, ward 2 in year 3, ward 3 in year 4, ward 4 in year 5. Then in the first scenario, migrant A will check his current location, which is ward 4. Then he decides that his mortality probability in year 6 will be determined against the age-sex specific probability in ward 4 that corresponding to his age-sex group; while in the second scenario, his mortality probability will be based on his age-sex specific probability in his origin ward: ward 1; in the third scenario, his mortality probability will be based on his age-sex specific probability in the area that he stayed the longest, which is ward 2 (Figure 6.6). Adopting these different rules for the mortality regime, a migrant is exposed to should improve the projections of deaths by location in the system. Such variances in the individual migration processes will also result significant changes in the local population structure.

Year 1 ... Year X-4 Year X-3 Year X-2 Year X-1 Year X ...



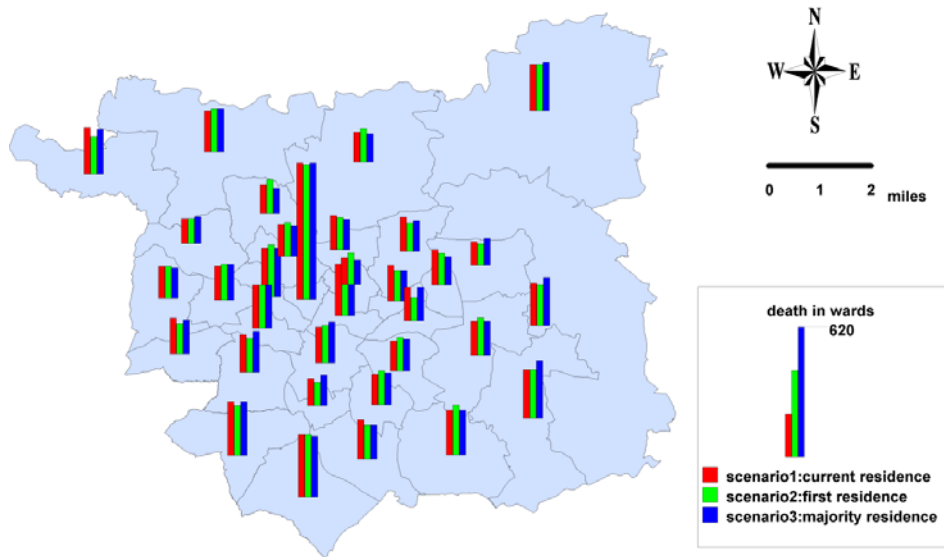
**Figure 6.6 Migration history of migrant A**

### 6.3.3 *Comparison of the simulation results using pure MSM and the hybrid model*

In the experiments, the whole population of Leeds has been simulated under the three assumptions for 30 years and the results of the year 2031 are then compared spatially to assess the difference in the mortality distribution within the city. Lacking information on migration histories of individuals from their births to ages at 2001, the focus here is the variations in results in 2031. Although the distribution pattern of mortality is similar on the whole, the experiments reveal some interesting variations in small areas.

Interestingly we can see from the map (Figure 6.7) that there tends to be a higher mortality in the more established suburban wards in the north of the city such as Otley and Wharferdale, Cookridge and Weetwood in scenario 2 than in the current residence based projection. In comparison, scenario 3 indicates a reduction of mortality compared with the current residence based projections in the northern suburban areas. It also indicates an increase of mortality in the southwestern belt around the city centre such as CityandHolbeck, Beeston and Wortley, as well as in the eastern strip from Wetherby to Garforth and Swillington. The mortality projection based on individual migration history in scenario 3 seems to differ from scenario 2 by indicating more deaths in traditionally immigrant/more deprived areas in the city. Such findings demonstrate that personal history could have an important impact on mortality in the small areas (Figure 6.7).





**Figure 6.7 Mortality experiment: deaths in small areas 2001 and 2031**  
 Source: Wu *et al.* (2010)

Empirical research on the relationship between limiting long-term illness and migration (Norman *et al.* 2004) established that the illness status of migrants is mid-way between that of their origin and destination locations. If this finding also applies to mortality, then a combination of all three scenarios may be needed to represent the mortality chances of migrant properly. The mortality projections will continue to be improved in the light of such evidence. Although the mortality experiments discussed here are purely based on hypothesis, it demonstrates that there are many more aspects of the population MSM which can be strengthened through the use of ABM techniques. Important elements of the model such as marriage behaviour, fertility patterns and change in health status might all benefit in a similar way. Such explorations are not only just interesting experiments, but potentially can play a vital role in facilitating the decision making where the impact of personal history is required to be taken into account.

## 6.4 Conclusions

In this chapter, two ABM experiments have been introduced to address the issues that are found of the spatial dynamic MSM. The MSM has faced the challenges of modelling a large complex social system at a fine geographical scale, but it is in need of attempts to capture the distinctive

patterns of movement, interaction and behaviours of various sub-populations more effectively.

As described in Chapter 5 and previous sections of this chapter, the subtle behaviours of some sub-populations within small areas still proved to be an issue, despite the efforts made on improving the MSM. In order to improve the model results at the finer level of spatial scales, an ambitious hybrid model has been built, which combines MSM and ABM features. Experiments on modelling the most and least complex demographic processes using both pure MSM and the hybrid modelling approach have suggested that the hybrid modelling approach is a useful extension of the MSM and is helpful in modelling both complex and simple demographic processes. Adopting the hybrid approach, the goal of both the effective handling of large scale individual based system and extra flexibility of modelling various behaviours of sub-populations in different scenarios has been achieved. The hybrid modelling approach does not only improve the behaviour modelling of the student migration, by capturing of the subtlety of the student migration patterns, the hybrid model can also produce improved projections in terms of population profiles at the disaggregate spatial level of small areas. In the examples described above, the most straightforward processes are deliberately chosen and ABM techniques are applied with considerable simplification. However, given more time and resource, the modelling of interaction, behaviour and personal history impact can be significantly more complex with intelligent agents. For example, people can have much more complex interactions with others and their environment to make a decision or take an action based on their personal history or requirement. The examples discussed in this chapter are simply used to demonstrate the potential of the hybrid model.

This model aims to provide support for both strategic and tactical decisions. Strategic planning is supported by long-term projections running to the year 2031. Tactical decisions can be supported by “what if?” scenarios concerning influences such as changes in the provision of housing or transportation. More detailed results including the alignment of model

estimates at the metropolitan area level to national forecasts are described in Birkin *et al.* (2009). The unique advantage of this model over the national projections is an ability to disaggregate both socially and spatially. This model can provide more details, for example, about household composition and health status of individuals within the households, but it can also provide very fine spatial detail about their spatial patterns. The MSM features ensure such benefit of this model. At the same time, the addition of the ABM features to this hybrid model has provided a framework to enable the effective modelling of heterogeneous individual decision making units on a large scale, as well as adding the flexibility to introduce different modelling techniques to strengthen various aspects of the model.

The model itself provides a useful tool in assisting decision making, exploring various “what if” situations and testing different hypotheses. As a demographic planning tool, this model can monitor the evolution of population structures and various demographic change patterns on a fine geographical scale. This provides vital information for demographic planning/policy making. This model can also benefit other public policy making or public service planning. For instance, the ageing trends in certain suburban areas may promote changes in health service and public transport service provision in order to enable easy access to such services for the old and frail in the area. The rich attributes captured in the system are also very useful in various policy analyses or research purposes. As discussed above, the hybrid modelling approach demonstrates great potential in demographic modelling and it is believed that continuous efforts to improve various aspect of this model using this approach can provide a better groundwork for decision making and predicting the future.

## Chapter 7

# Model validation and alignments

## 7.1 Introduction

Zaidi and Rake (2001) suggest that model validation is a broad term that incorporates various techniques to ensure the internal consistency of models, as well as their external credibility among the broader scientific community. Given this broad aspiration, it is not surprising that some disagreement exists about the techniques best suited for validating models. However, they suggest that one obvious way to address the broadest questions of validity is to compare the model results with external data. Despite its many advantages described previously (Chapter 3), the MicroSimulation Model (MSM) is more difficult to build and validate than the macroscopic model, due to the disaggregate nature (Harding, 2007). To assess the results of a dynamic spatial MSM, it is often necessary to re-aggregate the disaggregate simulation results to a more aggregated level. It is a great challenge to ensure such re-aggregated microsimulation outputs are in line with the externally produced macro projections. However, the developments of all current major MSMs make an

effort to compare the model results with external data to demonstrate the validity of their models (Harding, 2007; Rephann, 2001).

In comparison with the static MSM, the dynamic MSM can model the endogenous changes of the characteristics of the micro units, as well as project them over time to include demographic processes and social economic transitions, such as ageing, mortality, fertility or social and geographical mobility (O'Donoghue, 2001). Spatial MSM, as described in previous chapters (Chapter 3), takes the small area variances into account and allows us to model the individuals within a local context (Ballas *et al.*, 2005b). However, typically there is not any single set of microdata comprehensive enough to contain all the disaggregated individual and small area information embodied in projections of a dynamic spatial MSM. Therefore this indicates that a dynamic spatial MSM often has to use multiple datasets with different assumptions in various processes/modules. On the other hand, the small area characteristics modelled within a dynamic spatial MSM can also lead to disaggregated spatial variances that may not be consistent with the aggregate projections. The complexity of a dynamic spatial MSM can therefore lead to many possible causes of the issues in the process of model validation.

As dynamic MSMs normally project over a long period of time to assess the longer term impact of population changes, there are often limited suitable microdata that can be used for comparisons. It is generally agreed that the internal validation through exercises such as debugging and sensitivity tests is not sufficient to validate the results from dynamic spatial MSMs. The summed results of dynamic spatial MSMs often need to be aligned to external aggregate data (Caldwell and Morrison, 2000; Zaidi and Rake, 2001). One attractive way is to run a retrospective MSM and analyse the results against historical data. However, it is expensive in terms of computing and time cost and the historical data such as Census 1991 are often of poorer quality. As future trends can be quite different from the past, Zaidi and Rake (2001) points out that it is an incomplete way of establishing the validity of a dynamic MSM. Another alternative way is to use data from different

empirical studies for validations of the various results produced by the model. However, such independent studies are often conducted with different underlying assumptions and cannot provide a consistent view. Therefore a consensus has emerged during the past decade about the need for alignment of the micro projections to future macro economic and demographic aggregates. Today, most dynamic MSMs are aligned to external projections of aggregate results when used for policy analysis (Anderson, 2001).

In the past years, there has been substantial methodological work undertaken in the field of alignment and various alignment exercises have been adopted in current major MSMs (Rephann, 2001; Caldwell and Morrison, 2000; Harding, 2007). Based on the discussion of alignment practice and various previous alignment exercises, a framework of the model alignment has been designed for the current model. In the rest of the chapter, the details of the alignments will be discussed and the model results will be analysed against the official projections from the government agency, Office for National Statistics (ONS).

## 7.2 The need of alignments for MSMs

### *7.2.1 Issues in validating dynamic spatial MSMs*

There are different reasons why the summed results of dynamic spatial MSMs may have to align to external aggregate data. From the users' point of view, many governments have already provided their own official projections for the country. As a result, policy makers tend not to take seriously the model results that drift away too much from these official numbers. Also there is a need for the summed results of some of the dynamic models to match the official actuarial projections of their agencies, in order to provide a consistency between the macro actuarial projections and the summed micro results. Finally, the decision makers and some researchers often have been used to the usage of the official datasets/models in the processes of decision making or a study. Such users are often

reluctant to accept results/models drifting too far away from the official estimates.

More importantly, given the length of projections and the level of detail simulated by dynamic spatial MSMs, policy makers and scientific researchers are likely to raise questions about the validity of the dynamic spatial MSM. At the same time, the capability of a dynamic spatial MSM to model a complex social system in great details can lead to many challenges in building and validating such models. The main causes can come from the following areas (Harding, 2007; O'Donoghue, 2001; Zaidi and Rake and Rake, 2001):

- Data;
- Specifications and
- multiple sources of error (and at multiple levels for models with submodules).

Dynamic spatial MSMs face numerous and daunting data challenges. One of the first crucial challenges comes from the baseline data for the model. Often the microdata will be subject to sampling errors and typically there are no microdata readily available to provide all the information required by a dynamic spatial MSM. Often dynamic spatial MSMs have to recreate synthetic microdata (Birkin *et al.*, 2006). Given that one of the objectives for creating synthetic microdata is to create data that does not currently exist for required small geographic areas, it is not hard to see that it is going to be difficult to validate the baseline data. However, even after dealing with the associated challenges with the selected base data, various transitional probabilities still need to be generated. A dynamic spatial MSM uses such probabilities to move the simulated individuals from one state to another as they progress through time and various important demographic transitions. The projection of the future population characteristics depends on consistent estimates of all transitions probabilities and state specific expectations. However, as with the creation of the baseline data, finding suitable data to update the parameters and/or probabilities over time for a dynamic spatial MSM is difficult and often such data are not available at the required disaggregated spatial scale. Typically a selection of empirical data is used to

estimate the various transition probabilities. The need to use multiple datasets for the probability calculations can introduce inconsistency to the model and cause issues in the process of the model validation. The probability based simulation itself will introduce a degree of randomness to the model results as well.

Short term projections in static models are often not able to capture critical dynamics which have been occurring in the longer term, e.g. the trend of delayed childbearing age for women in UK during the past decades (Tromans *et al.*, 2008). Sometimes a short term projection only captures behaviours when they are about to change for reasons not self-evident at the time. Consequently the model results can become misleading. Dynamic MSM attempts to capture critical changes in the population dynamics over time and such information has become increasingly important in modern policy development or strategic planning (Harding, 2007). However, projections over time at the micro-level are susceptible to misspecification error due to the details involved and insufficient knowledge about micro-behaviour. The parameters used in such projections are often not (sometimes cannot be) accurately specified, for the precise information on future trends is unattainable. Sometimes the underlying assumptions of the specification can change in light of more up-to-date information that is not yet available at the time of specification (O'Donoghue, 2001; Zaidi and Rake, 2001).

Due to the complexity of a dynamic spatial MSM, there are multiple sources of errors that can be introduced to different aspects of the model. As discussed above, the mistakes in the choice and manipulation of multiple data, the inaccuracy of the system specification, the mis-specification of individual processes or interactions between them can all result in errors in the model results. However, the errors can be introduced to the system at the multiple levels as well. Due to the nature of demographic processes, some are more complicated than others and may require sub-modules that specifically designed for modelling such process, e.g. the migration module needs sub-modules to handle different aspect of modelling of this process. The migration process is composed of multiple flows at multiple spatial



scales (e.g. domestic and international migration) and proceeds in multiple stages (e.g. who is moving, how they move, where to move to?). There is also an interdependent relationship between the migration and other demographic processes (e.g. household formation often leads to migration). There is a detailed discussion in Chapter 3. Modularised system provides great flexibility in both structural and functional capabilities of the model. However, the sub-modules can give the opportunity for multiple sources of error to impact at multiple levels of the model, if not handled carefully. Subsequently, errors or mis-specification of individual processes or interactions within the sub-modules can have an impact on the overall output from the model.

Given the length of the projection period and the level of detail simulated by dynamic spatial MSMs, it is, therefore, not surprising that the predictions of unaligned dynamic MSMs can drift away from benchmark aggregates such as the official population projections. Despite the challenges posed by dynamic spatial MSMs, it is believed that considerable efforts are likely to continue to be made to develop and improve such models, because there is a need for investment in the requisite modelling infrastructure in order to tackle the forthcoming fiscal challenges caused by population ageing (Harding, 2007).

### *7.2.2 The alignment approach*

In the recent past few years, the ability to align the micro output to benchmark macro estimates has emerged as a crucial component of many MSMs, as alignment can help capture the macroscopic impact in microsimulation aggregate results and provide an indicator of the aggregate performance of the model (Rephann, 2001). As Anderson (2001) states, without alignment, projections of future years may begin from a base that already is subject to errors and such errors can continue to accumulate over the entire simulation period of time. Aligning the micro values produced by dynamic MSMs with known or projected macro aggregates can help to enable integration into the model of the more recently available trends and

macro effects. Previous researches also believe that the alignment approach can make up for some of the unexplained system effects resulting from the disaggregate differences. It can also be justified as a way of compensating the imperfection of data and estimation techniques, as well as rectifying sub-module defects such as mis-specified behavioural equations and data limitations/errors (O'Donoghue, 2000). Another advantage of adopting the alignment approach is that it allows the aligned MSM retain the disaggregate characteristics. Although the alignment practice usually involves some modification of the model estimates, it generally only changes the aggregate outputs of the model, not the distributions (Anderson, 2001).

Today it is generally agreed that one alternative way to validate the dynamic spatial MSM is to compare aggregate or mean model output over time with official data derived from administrative records and/or survey results (Rephann, 2001; Caldwell and Morrison, 2000). Most current MSMs, such as MOSART, DYNAMOD2, CORSIM, and NEDYMAS, all make use of the technique to some extent (Rephann, 2001). Zaidi and Rake (2001) believes any alignment can operate to constrain the outputs of the model in order to bring them in line with the external data, although the alignment exercise can vary in actual operations. The alignment practice allows the MSM adjust individual simulation results to achieve aggregate results which correspond to historical or forecasted results. Typically this is achieved through a series of adjusting exercises of the results in new individual outcomes, totals and flows (O'Donoghue, 2000; Anderson, 2001). Various alignment exercises can be conducted, ranging from alignments to assumptions and rates to alignments on multi-levels of individual processes or modules (Caldwell and Morrison, 2000; Rephann, 2001; Zaidi and Rake, 2000). The alignment to the macro outputs can be any important macro or aggregate outcomes, such as macro-demographic and macro-economic outcomes. For example, DYNACAN aligns to target rates for mortality, fertility, migration, marriage and divorce propensities and SESIM is constructed to easily align to different exogenous demographic and macroeconomic assumptions (Harding, 2007).

Based on the above discussions, an alignment approach is adopted in this study to demonstrate the validity of the model through a series of alignment exercises. The aggregated results of the official projections produced by ONS have been used for the comparisons.

## 7.3 The alignment framework

On the basis of the discussion of previous alignment exercises, it is decided that the results of this dynamic spatial MSM need to be aligned to the official projection results produced by ONS models. By aligning the aggregated results to ONS projections, it also enables the incorporation of the recent and future population trends revealed in ONS assumptions that have significant impact on simulation results. The assumptions underlying the ONS projections based on the 2006 population estimates (the most up-to-date at the time the model was built) were employed to modify the probabilities used in various demographic transitions and to align the model to the ONS projections.

A framework of the validation has been developed on the basis of the alignment exercises, composed of 2 assessments between 3 models. The framework of the validation is illustrated in Figure 7.1. Model A is the ONS aggregate projection, Model B is a naïvely disaggregated hybrid MSM applying the same aggregate probabilities for all wards and Model C adopts a full disaggregation. The first assessment is through the alignments to the ONS projections by applying the naïvely disaggregated/averaged estimations that are used in ONS projection to all populations in Leeds wards. Then the results from the modified simulation are re-aggregated and compared against ONS' aggregated results to test the consistency of the model. In the second assessment, the ONS estimations are fully disaggregated to the ward level and the results are re-aggregated and compared to test the robustness of the model.

This validation framework uses a combination of two sets of model results by ONS. As the most up-to-date LA level projections by ONS were the

SNPP (Sub-National Population Projection) based on 2006 populations at the time when the study was conducted, results from it for the period 2007-2031 are used as the basis data for the alignment exercises. For the period prior to the SNPP 2006, year 2001 to 2006, the MYE (Mid-Year Estimation) data are used instead. The MYE data are used instead of SNPP 2004, mainly because it is considered to be more accurate, while important changes in population trends have been found missing in the 2004 population based projections (ONS, 2008e). While assumptions revealed in SNPP2006 (ONS, 2008e) are used to update the migration probabilities from 2007 onwards.

There are three windows identified during the simulation period. The period from year 2001 to 2006 is called the Estimation Window in our validation framework, because the alignments are based on the MYE data. As the SNPP 2007 projection is directly projected on the basis of the 2006 MYE, the period 2007-2010 provides a Validation Window to assess results of the hybrid MSM against the ONS results. The SNPP projections are selected as the baseline for comparison as MYE data were only used for 2001-2005; therefore, the simulated results are compared to SNPP results instead of MYE, as the most up-to-date LA level projections by ONS that cover the simulation period of our model were the SNPP 2006 based projections at the time when the study was conducted. MYE data were used for the period prior 2006. Given the time limit of this PhD programme, we cannot compare the results against the 2007-2010 MYE data available now. However, this can be carried out in future work. The period from year 2011 to 2031 is called the Projection Window in the framework, as the SNPP 2006 projections with the long term assumptions are used for comparison with the aligned model results.

It should be noticed that in these alignment exercises this study only tries to adopt the assumptions/trends found in ONS models and these exercises are not trying to recreate the results of ONS in this study. The ONS SNPP model projection do not consider the characteristics of the small area population by further spatial disaggregation transitions. However, the

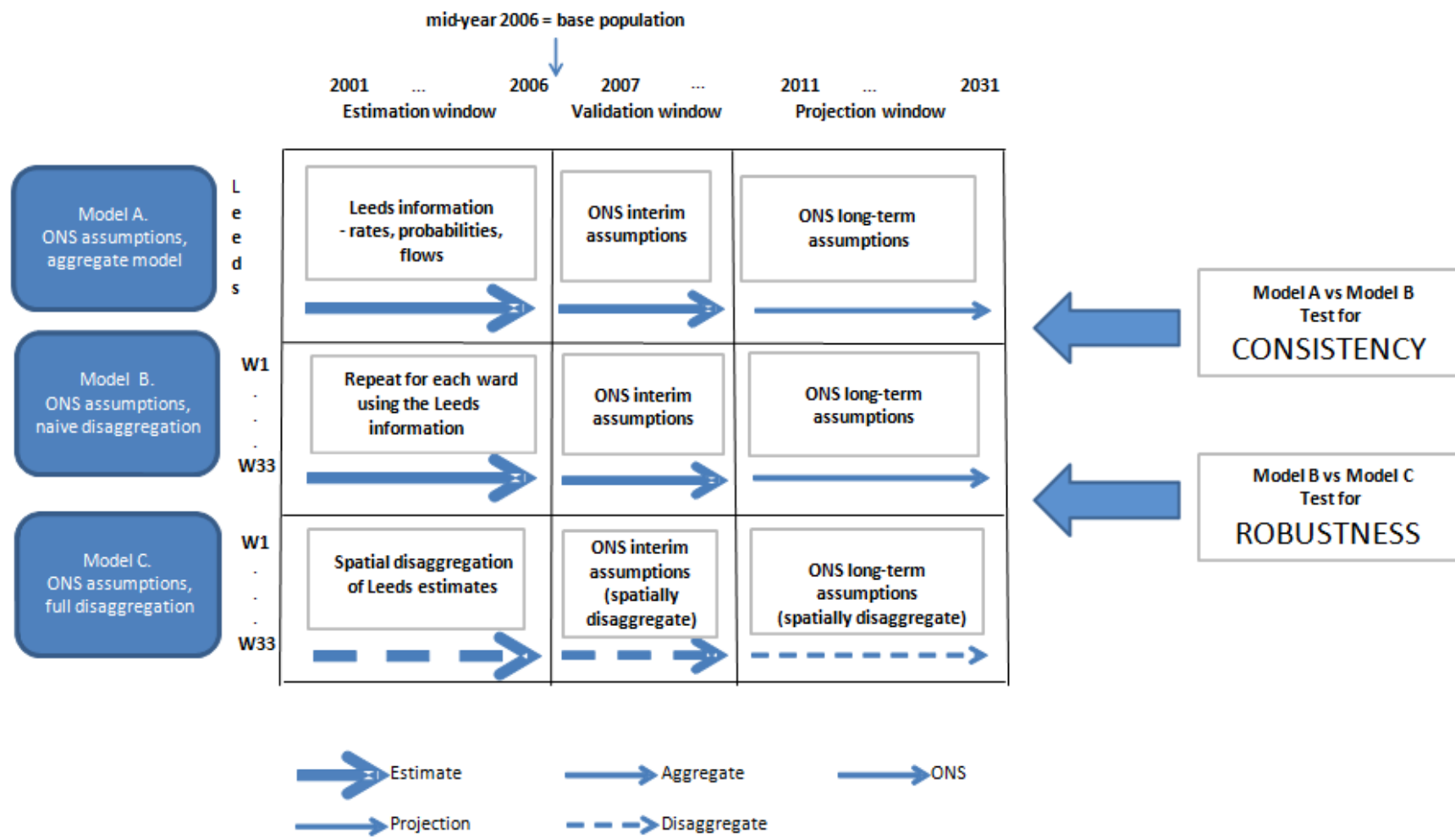
individual based model projects the population by age, sex, location and various attributes that are important for specific demographic transitions. Therefore, recreating exactly the same results is not the aim of this study. Instead, the alignment exercises are an attempt to assess the consistency and robustness of the individual based hybrid model.

## 7.4 Model validation and alignment

As discussed above, the unaligned dynamic spatial MSM can drift away from the official projections. Due to both the need for system validation and the user requirements, alignment of projection has become an increasingly popular practice for MSMs. In the following section, two alignment exercises will be described and results from the two models, Model B and C, will be compared to the official projections from Model A by ONS (Office of National Statistics). The following sections will describe the models in detail.

### *7.4.1 Model A*

Model A is an multi-regional, cohort-component aggregate model that uses the same set of ONS interim assumptions for the whole England and Wales for the period of 2002 to 2010 and long term assumptions for the period of 2011 to 2031. The most up-to-date MYE and SNPP data at the time of the exercises were carried out have been used for Model A results. As described previously, for the period of 2002 to 2006, estimations at the LA (Local Authority) level from the ONS MYE of the populations are used for accuracy. However, for the period from 2007 to 2031, 2006 based SNPP projection results at the LA level of Leeds are used, as it was the most up-to-date LA level projections for the period of 2007-2031. The Model A results are the aggregate results at the level of the whole Leeds, which do not provide details on age, sex and small area characteristics. Model A



**Figure 7.1 Result alignment framework**

provides the basis for the series of alignment exercises and analyses against Model B and C as illustrated in Figure 7.1.

### *7.4.2 Model B*

Model B is a disaggregate model that naïvely distributes into each small area populations of Leeds the aggregated probabilities calculated using the Model A results, i.e., the Leeds level probabilities from Model A are applied for all wards in Model B. Model B is aligned to the ONS model with the relative population trends. The main purpose of this alignment exercise is to find out if the individual based hybrid model can produce population evolution patterns that are consistent with the ONS projections, given the same assumptions and trend information. Model B uses the information available at the aggregate level of Leeds LA in ONS projections and tries to align the assumptions and trends in individual processes or sub-modules.

As mortality, fertility and migration all plays crucial roles in population change, alignment exercises have been conducted on all the three components of changes. In the mortality and fertility processes, Model B naïvely disaggregates specific probabilities using the available Leeds information from the ONS aggregate projections at the LA level and applies them to every ward of Leeds.

The migration alignment requires more work. There are five different migration flows in our migration framework: inter-ward migration (within Leeds), internal in migration and internal out migration (within UK), immigration and emigration (outside UK), but there are only four migration flows used in the ONS projection outputs: internal migration in, internal migration out, international and cross-border migration in and international and cross border migration out. Therefore, the results of the inter-ward migration are not aligned and compared with the ONS projections, due to the lack of information. Only results from the four flows are re-aggregated to the level of LA for comparison. However, this is not a problem for the aggregate analyses, as the inter-ward migration in cancel out the effect of the inter-ward migration out at the level of the LA.

### *7.4.3 Model C*

Compared to Model B, Model C is a fully disaggregated model at the level of the small areas. All probabilities used in modelling the components of change within the population have been disaggregated at the fine spatial scale of wards, as described in Chapters 3 and 4. This provides a disaggregated baseline population of year 2001, in terms of both spatial and demographic characteristics. From the year 2002 onwards, the information on the trends of population changes revealed in the ONS projections are used to update the probabilities for simulations of 2002-2031.

In the mortality updating, Model C adopts the ONS assumptions that the mortality will continue to improve during the period of 2001-2031. As described in chapter 4, the base survival probabilities of year 2001 are disaggregated by single year of age, sex and ward. From 2002 to 2006, the annual mortality differences within the ONS MYE data for Leeds populations have been used to update the survival probabilities by single year of age, sex and ward for the annual simulations during this period (ONS, 2008e). From 2006 to 2031, the mortality improvements are calculated on the basis of SNPP information from year 2006 onwards (ONS, 2007c).

Similarly in the fertility updating, from year 2002 onwards, spatially disaggregated ONS probabilities have been applied to the base probabilities of year 2001 that are age, marital statuses and ward specific. During the period of 2002-2006, the ONS MYE data for Leeds populations have been used to calculate the aligned updating probabilities Model C. Then the longer term trends revealed in 2006-based SNPP results (Bray, 2008) are used to align the fertility probabilities during the simulation period of 2007-2031, by applying the difference between fertility rates of two consecutive years in the ONS data to Model C probabilities. Changes over 5 years are averaged to obtain the annual change.

In the migration updating, the base probabilities in year 2001 are spatially disaggregated to the level of wards by age and sex, using the SMS 2001



(Special Migration Statistics) data to directly calculate the internal migration in, internal migration out and immigration probabilities (ONS, 2011). Due to the small IPS (International Passengers Survey) samples, the emigration probabilities are then calculated on the basis of the internal out probabilities and the MYE outputs, assuming the ratio between the internal-out migration and emigration rates is consistent. On the basis of the 2001 probabilities, the migration probabilities 2002-2031 for all four flows of migration are then disaggregated to the level of ward and used to update the simulation probabilities annually. Again, during the period of 2001-2006, the migrations in Model C are simulated using probabilities aligned to the ONS probabilities in MYE data (ONS, 2007c) and SNPP 2006 (ONS, 2008e) are used to update the migration probabilities from 2007 onwards.

## 7.5 Analysis of alignment results

As described previously, this model aligns with the assumptions made in the ONS projections, but is not aligned with the parameters, input and output in the ONS model. Analyses have been conducted on model results from three models: Model A - ONS aggregate model, Model B - naïve disaggregate model and Model C - fully disaggregate model. Results from Model B and C are re-aggregated to the level of Leeds LA to enable the comparisons with Model A (ONS projections). As age and sex are the most important characteristics of any population, the age-sex structures of the Leeds population from Model B and C at the beginning and the end of the simulation period have been analysed to assess the consistency and the robustness of the hybrid individual based model against the official results of Model A. Due to the fundamental impact on population changes, the three components of changes of Mortality, Fertility and Migration from the three models have also been analysed to reveal the details of the population changes. The component of change of three years, 2002, 2007, and 2031 in Model A, B and C are selected to represent the three important windows in the validation framework: the Estimate Window, the Validation Window and the Projection Window specifically.

To explore the consistency, robustness and plausibility of the individual based model, age-sex structures and components of change of small area populations have also been analysed and discussed. The following sections will describe the analyses separately.

### *7.5.1 Analysis of age-sex distribution of Leeds population*

#### *7.5.1.1 Model B*

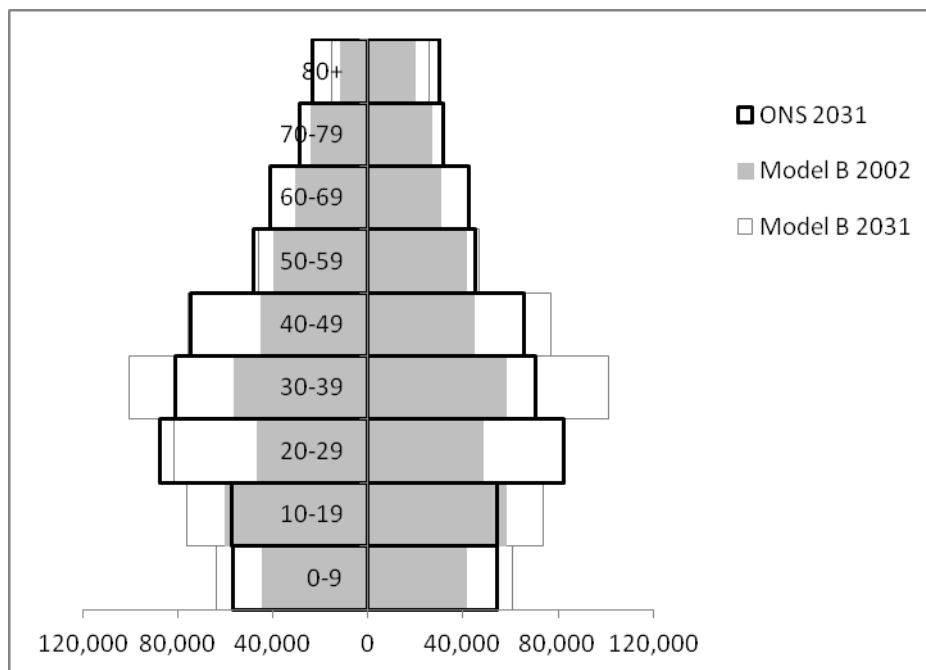
Model B naïvely distributes the same set of probabilities from the ONS aggregate model, Model A, to update the simulations of population changes in all Leeds wards without considering the local spatial and demographic characteristics within the small areas. The simulated results from Model B are then re-aggregated to the level of the whole Leeds for comparison with projections from Model A. The results from Model B demonstrate a good consistency with the Model A results.

As can be seen from Figure 7.2, Model B projects that Leeds population will grow steadily during the period of 2001-2031. The age-sex distribution patterns of the projected Leeds population in 2031 remain similar to those in 2002 results. Such patterns are also parallel to those revealed in the ONS projections by Model A in 2031. However, the results from Model B suggest a bigger population (1,026,889) than Model A (974,300). There are also some slight differences in some age groups. Model B results suggest a slightly smaller group of ages 80+, but bigger groups of ages 0-9, 10-19 and 30-39 than Model A (Figure 7.2). Possible sources of the difference include:

1. the impact of disaggregate characteristics within the base population, although updated with naïvely disaggregated probabilities;
2. the ABM simulation of the student migration process in the hybrid model does not use the same migration assumptions for this sub-population and it may result in more younger people migrating into Leeds and affecting the age-sex distributions of the small area populations (Chapter 6) and

- Model B starts with a slightly larger baseline population of 723,272, while Model A starts with 715,600, where some variations of the population characteristics in the two datasets may exist (Birkin, *et al.*, 2006).

With such sources of variance, the naïvely spatially disaggregated Model B produces quite similar patterns in the age-sex distributions after 30 years' simulation. To assess the impact of further disaggregation on the age-sex distributions of Leeds population, the next section will discuss the analysis carried out on the results from Model C in comparison with both Model A and B.



**Figure 7.2 Population sex-age structures of Leeds population in 2002 and 2031 generated by Model B compared with the ONS projections (Model A) in 2031**

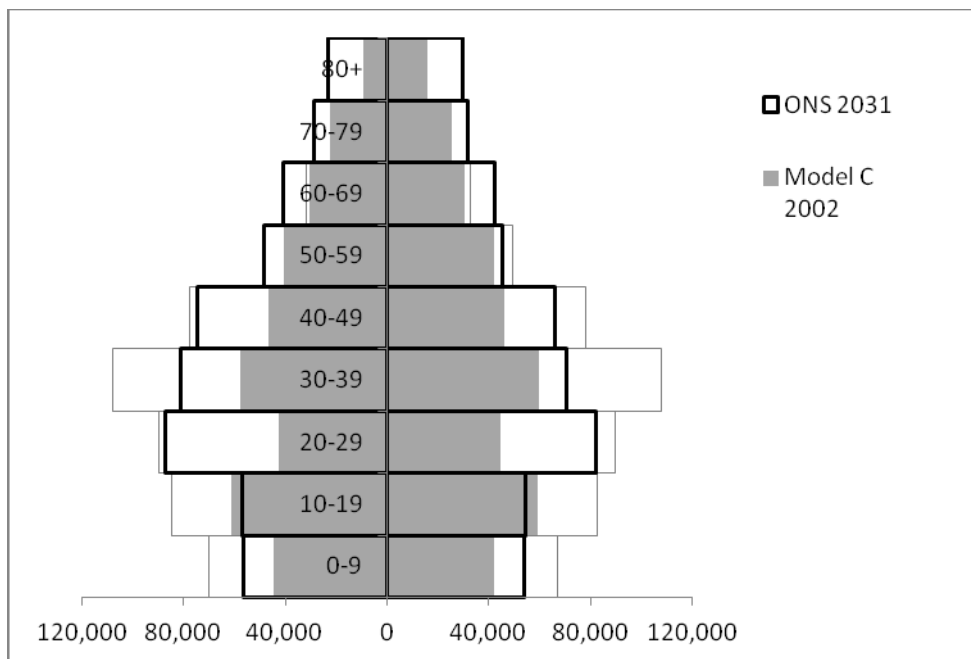
Source: ONS (2008) and author's own computation

Note: Male population is on the left, female population is on the right of the axis.

### 7.5.1.2 Model C

Unlike Model B, Model C fully disaggregated the probabilities to the level of wards to capture the small area variance. Although still demonstrate most similar patterns to the projections from Model A, the simulated results by

Model C reveal further differences than the results produced by Model B. In Model C, the population is projected to be even bigger (1,070,246) than in Model B. As Figure 7.3 illustrates, there is also some further variation in the age-sex distributions. There are bigger numbers of population for the ages 0-49 than in the ONS projections, while much smaller number for the ages 80+. In fact, it is even smaller than what is projected in 2002 by Model C. The age-sex structure analysis of Model C results indicate that there is a less obvious trend of population ageing than projected by the ONS projections in Model A. Considering the full disaggregation that Model C has adopted, the age-sex structure analysis still reveal patterns that are consistent with those produced by Model B. Apart from the degree of the disaggregation used in the updating process, Model C is completely consistent with Model B, including the hybrid feature of ABM. The largely similar model results produced by this alignment exercise demonstrate the robustness of the hybrid model in reproducing the evolution of the age-sex structure over the period of 30 years.



**Figure 7.3 Population sex-age structures of Leeds population in 2002 and 2031 generated by Model C compared with the ONS projections (Model A) in 2031**

Source: ONS (2008) and author's own computation

Note: Male population is on the left, female population is on the right of the axis.

The big difference between Model A and models B and C is in the 30-39 age groups, followed by the 0-9 and 10-19 ages, where results from Model B and C are larger than those from Model A. This probably suggests difficulties in estimating the migration flows into and out of Leeds, as well as the impact of spatial variances.

### *7.5.2 Analysis of components of change of Leeds population*

As the Mortality, Fertility and Migration components all play crucial roles in population change, alignments have been conducted on all the three components of changes and results from Models A, B and C are analysed by each component of change, as well as the total populations. Results in selected years of 2002, 2007 and 2031 from each model represent the results within Estimation Window, Validation Window and the Projection Window in the validation framework. In Migration process, individual migration flows are analysed separately as well. All components of change of Leeds population are shown in Table 7.1 and the findings are discussed in the following separate sections (six decimal places are used to present the demographic rates so that they can be compared between models).

**Table 7.1 Components of change analyses of Leeds population**

Model	Model A			Model B			Model C		
Year	2002	2007	2031	2002	2007	2031	2002	2007	2031
<b>Total population</b>	720,000	759,400	974,300	731,047	778,774	1,026,889	721,660	756,857	1,070,246
<b>Fertility</b>	0.048888	0.053162	0.054042	0.041786	0.045935	0.049709	0.041630	0.045663	0.050982
<b>Mortality</b>	0.009722	0.008428	0.006364	0.009687	0.008646	0.007842	0.011063	0.009831	0.006775
<b>Internal In</b>	0.038889	0.040163	0.031305	0.038575	0.039525	0.031019	0.041820	0.046425	0.044613
<b>Internal Out</b>	0.041250	0.040953	0.037463	0.041293	0.040674	0.037503	0.044360	0.047334	0.053329
<b>Immigration</b>	0.013611	0.017777	0.016935	0.013500	0.017493	0.016783	0.014903	0.021542	0.027467
<b>Emigration</b>	0.006389	0.006321	0.008314	0.006377	0.006259	0.008278	0.006874	0.007308	0.011935

Notes:

Fertility probabilities: all women give birth ages 15-45+/all women ages 15-45+

Mortality probabilities: all death/all Leeds population

Internal in-migration probabilities: all migration from the rest of UK to Leeds/all Leeds population

Internal out-migration probabilities: all migration from Leeds to the rest of UK/all Leeds population

Immigration probabilities: all migration from the rest of world to Leeds/all Leeds population

Emigration probabilities: all migration from Leeds to the rest of world/all Leeds population

### 7.5.2.1 Model A

Model A is the ONS sub-national projection for Leeds LA based on year 2006 population, which provides the most up-to-date information of official projections for the period of 2001-2031 at the time of this study was conducted. The analyses use the most detailed results, which is at the level of the LA (the whole of Leeds) from the aggregate model for the purpose of comparisons with the results produced by the disaggregate models.

As the results in Table 7.1 indicate, Model A projects that there will be a steady increase in Leeds population during the period of 2001-2031. Mortality is projected to be reduced over the projection period, under the assumption that there is a continuous improvement in life expectancy, due to the progress made in health and well-being in our society. Fertility, on the other hand, is predicted to increase. The increases in fertility occurred

between 2001 and 2011 as the TFR rose nationally from 1.6 to 2.0 approximately. Although later the TFR fell back from 2.0 to 1.85 approximately, it is still higher than 1.6 in 2002. This explains the rapid increase of fertility in 2007, then slower increase from 2007 to 2031. ONS projects a rapid increase of immigration into Leeds during the period of 2002-2007, but a slight reduction in the emigration. After 2007, the immigration results reduce at a slower pace while the emigration increases. The immigration trends are consistent with the ONS projections that predicate a reduction in immigration from the high levels of 2004-2011 in the period of 2011-2036. The emigration reflects the SNPP projections of the increase of emigration from 0.006321 in 2007 to 0.009752 in 2011 approximately. Although the emigration falls after 2011 gradually, the emigration level is still higher in 2031 than that in 2007. The internal migration in results have seen steady increases before 2007 and then reduces afterwards, while the internal migration out results seem to reduce gradually throughout the whole simulation period. At the end of the simulation in year 2031, Model A projected the total population to have increased from 720,000 in 2002 to 974,300.

#### *7.5.2.2 Model B*

Model B fully adopted the Model A assumptions and distributes the aggregated probabilities to all ward populations of Leeds, assuming there are linear changes each year between the five year gaps within the ONS results. As indicated by the figures in Table 7.1, the spatial variation does have an important impact on the Leeds population change patterns. Although Model B has produced close results as those from Model A, the disaggregated characteristics has produced some significant differences in the model results. Model B results suggest a consistently larger population than Model A results throughout the simulation. Looking into the specific components, we can see that there is more variation in the natural changes than in migration. Model B results suggest that the mortality reduces at a slower pace than Model A results, however, the fertility increases faster than Model A. On the other hand, the migration changes in Model B are quite

close to what projected in Model A. In all four migration flows, the trends of the migration changes in Model B are consistent with Model A. The naïve disaggregation only produced slight variances in the aggregated migration results. Despite the very similar population change patterns identified in Model A and B, the naïve disaggregation still made an impact on the population changes. Compared to Model A with no spatial composition effect, changes in the spatial characteristics of the population across the wards in Model B will affect the aggregate totals. For example, assuming the high mortality wards increase their share of the Leeds population, the average mortality probabilities will shift upwards as a result.

Model B starts with a larger population of 731,047 and projects a larger total population of 1,026,889 after 30 years' simulation (Table 7.1), approximately 50 thousands more than that projected by Model A. There are a couple of reasons for fertility in Model B to be lower than the SNPP projections. Firstly, both Models B and C start with the same baseline population. Although the aggregate ONS probabilities for all population from SNPP component projections are used in updating, only the difference between years were applied to base probabilities as we were only aligning to ONS assumptions, not inputs (see chapter 3 and previous section). Secondly, ONS uses TFRs to model the fertility, but this model uses ASFRs. Different assumptions of the age distributions of women at risk are adopted: the baseline population was not constrained by numbers of women at fertility ages, as too many constraints reduce the effectiveness of the population synthesis. Thirdly, the exclusion of university students from the birth process can also introduce further difference into the fertility results. The fertility probability probabilities need to be adjusted to allow the exclusion of student population in future versions of the model.

### *7.5.2.3 Model C*

In comparison with Model A and B, the fully disaggregated Model C updates the mortality, fertility and migration probabilities that are both disaggregated by spatial and demographic characteristics. The results from Model C suggest a slightly different population change pattern during the



period of 2001-2031. When we look at the total population results before year 2007, Model C results suggest a slower growth in total population than that of the ONS projections. However, the growth speeds up after 2007 and Model C has produces a much larger a population than Model A in year 2031 (Table 7.1). Looking into the details of the components of changes, a changing pattern closer to Model B rather than Model A has been found. Changing trends in fertility, mortality and all four flows of migration in Model C have been consistent with Model B, although all with higher rates. Compared with the results of Model A up to the year of 2007, Model C results suggest lower fertility rates than Model A and they are lower than those in Model B, too. However, the fertility increases at a faster pace after 2007. As a result, the fertility of Model C at the end of the simulation is much higher than that of Model A. Model C projections suggest consistently higher mortality in Leeds population throughout the simulation. The migration results suggest a consistent pattern with both Model A and B, although all the rates are higher. At the end of the simulation in Model C, the total population is projected to be 1,070,246, nearly 100 thousands more than that in Model A. As Model C uses a full disaggregation, local area and demographic characteristics are captured within the simulation. The ONS probabilities are not applied directly in each process; instead, the difference from one year to the next is applied to the original disaggregated probabilities. In mortality process, such differences are used to update the survival probabilities by location, sex and single year of age. In fertility process, they are used to update the fertility probabilities by age groups, marital statuses and location of women at risk. Combined with the existing difference in the baseline population, the two processes reveal more impact of the full disaggregation. As the migration probabilities are only disaggregated by location, but not by age and sex, due to the theoretical and practical reasons described in Chapter 4, the migration results demonstrate a closer pattern to the ONS results. Compared with Model B, Model C results demonstrate the same the spatial composition effect as discussed in Model B results. However, the full disaggregation of demographic characteristics within simulations each year has also produced an impact that is equally important.

#### *7.5.2.4 Population change trends*

As described in previous sections, the different degrees of disaggregation in Model B and Model C have produced some variances when compared to the results produced by Model A. However, when we look at the population evolution patterns using all annual results over the whole simulation period, there has been a good consistency within the trends that all three models reveal. Not surprisingly, Model A and B results demonstrate the same trends in all components of changes:

1. fertility increases before the long term assumptions are adopted in 2011, then fertility decreases gradually afterwards;
2. mortality improves throughout the whole simulation;
3. internal migration in increases until 2007, then starts to decrease gradually afterwards;
4. internal migration out decreases throughout the whole simulation;
5. immigration increases until 2007, then starts to decrease gradually afterwards and
6. emigration decreases until 2007, then starts to increase gradually afterwards.

Model C results also suggest mostly the same trends, except:

1. internal migration out keeps growing throughout the simulation and
2. similarly the immigration also keeps growing after 2007.

As the main difference between Model B and C is the different degree of disaggregation, further analyses using the disaggregate results at the level of small areas have been conducted for explorations of the differences in the population change patterns.

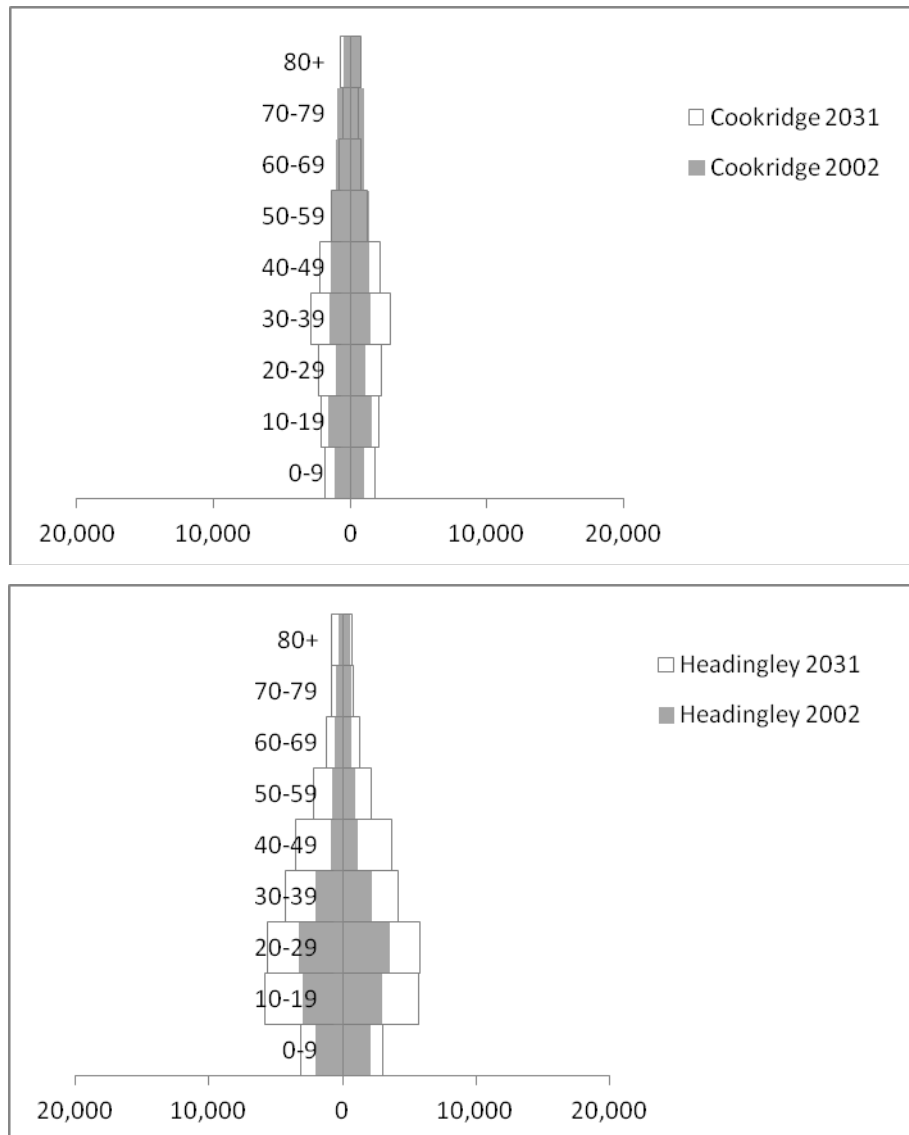
### *7.5.3 Analysis of age-sex distribution of small area populations*

To explore the impact of the small area characteristics on the population change, the aligned model results for the same two wards, Cookridge and Headingley, have also been analysed. The two wards are selected to provide the consistent basis for comparisons, as the unaligned results from the same wards have been used in the small area analyses in previous chapters. The age-sex structures of the projected small area populations of the two wards at the beginning (2002) and end (2031) of the simulation have been analysed, using the results produced by Model B and C specifically. The following sections will discuss the findings separately.

#### *7.5.3.1 Model B*

As illustrated in Figure 7.4, although the same average probabilities have been applied all individuals in all areas of Leeds, the small area characteristics still made an impact on the age-sex structure changes. Populations within both Cookridge and Headingley will grow steadily over time. There will be significant growth from the age 0 to 49 in Cookridge, while a slight reduction will see in ages 60 to 79. On the other hand, all age groups in Headingley will see considerable growth. Noticeably a sizeable growth will be found in ages 40 to 59, although the ages 0 to 39 will grow considerably. The age-sex structures analyses of the Model B results in the two small areas suggest that populations in both areas have grown over time and the structures in 2031 remain mostly similar to what has been found in 2001. However, there is some significant variation between the two areas, too. There will be more increase in all age groups in Headingley than in Cookridge. Also, populations in ages over 50 in Cookridge stop growing and start to decrease and there is even a smaller population in ages 70-79 in 2031 than in 2002. On the other hand, the ages over 50 keep growing in Headingley and produce a bigger population of ages over 50 in 2031 than in

2002 (Figure 7.4). As Model B applies the same aggregate probabilities from ONS projections to all wards, the growth rates of all small area populations are assumed to be the same. However, with the impact of different structures of the two small area populations, such naïve disaggregation produces a slightly larger growth in Headingley than in Cookridge.

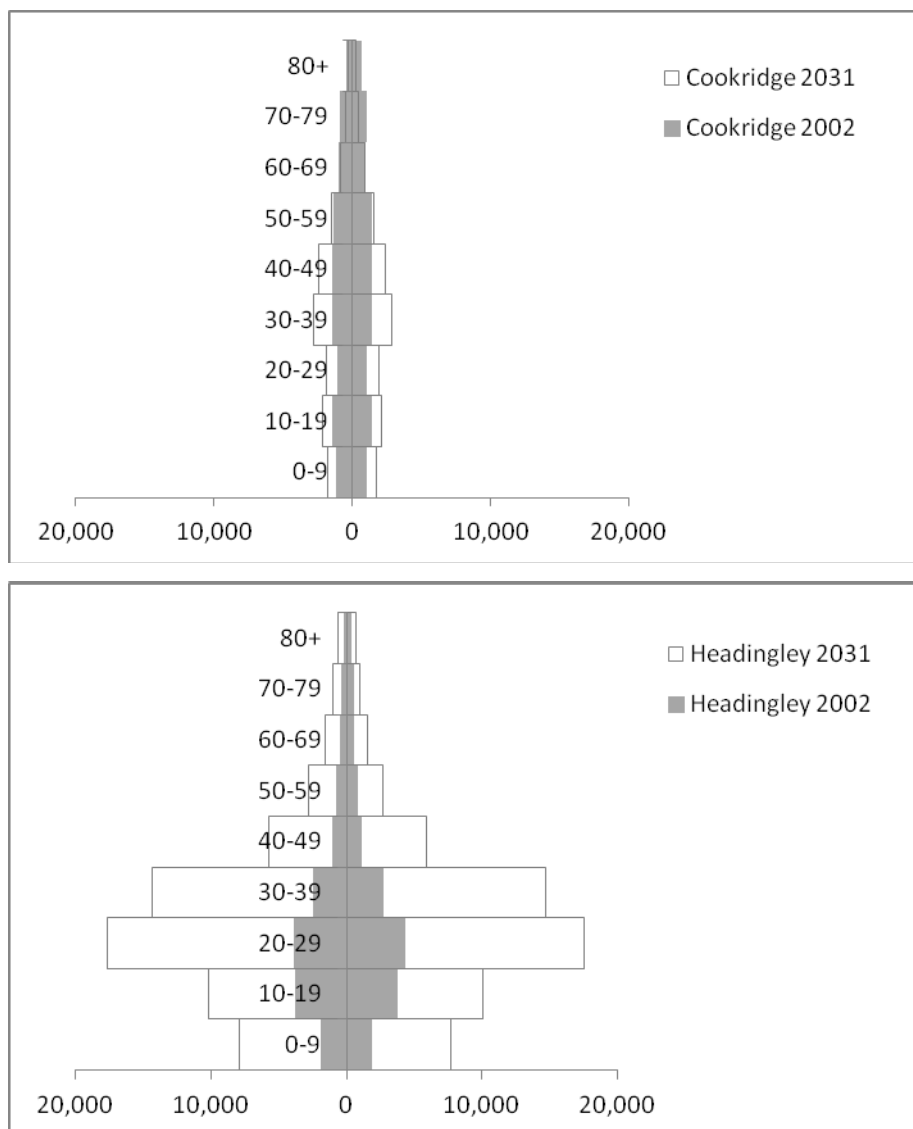


**Figure 7.4 Age-sex structures in small area populations projected by Model B**

Note: Male population is on the left, female population is on the right of the axis.

### 7.5.3.2 Model C

Analyses of age-sex distribution of populations in Cookridge and Headingley have also been carried out using the results produced by Model C. Compared to the analyses using the results from Model B, Model C results suggest even more variations between the two small area populations. As can be seen in Figure 7.5, the age-sex structure of Cookridge population remains similar to what has been projected by Model B, while the age-sex structure analysis of Headingley population reveals an even bigger and younger population in 2031 than projected in Model B.



**Figure 7.5 Age-sex structures in small area populations projected by Model C**

Note: Male population is on the left, female population is on the right of the axis.

In fact, the population in 2031 more than doubles that in year 2001. Both areas have substantial growth in younger ages 0-49. On the other hand, from ages 60 onwards, population starts to decrease in Cookridge. Although population aged over 60 still increases, it is increasing with a much smaller proportion, compared to the increase in ages 0-49. As a result, populations in both Cookridge and Headingley have grown over time in Model C projections. The population of Headingley grow faster than that of Cookridge, especially in ages 0-39. In Model C projections in 2031, Headingley has a much larger and younger population than projected in Model B (Figure 7.5). Clearly Model C has projected a large number of migrants into the ward of Headingley and this is an exaggeration of the real migration in this area, as there is not housing capacity in Headingley to cope with such increases. This MSM alignment analysis indicates that a future MSM should be integrated with the housing sector. Without the ABM, where agents always check the availability of housing vacancy before moving in, the future MSM will need to include a microsimulation of the housing stock, a house rent/purchase market and apply capacity constraints.

#### *7.5.4 Analysis of components of change of small area populations*

##### *7.5.4.1 Model B*

As discussed in Section 7.5.3, the spatial variations have been found in the analyses of population changes between small areas. To further explore the reasons behind such differences, the components of change of small areas are analysed and results of Cookridge and Headingley are used as an example for discussion again. Table 7.2 provides details of the components of changes of the populations in the two wards in years 2002, 2007 and 2031.

Compared to the patterns of components of change that have been revealed in aggregate Model A results, Cookridge seems to have a closer pattern than Headingley in terms of the changes. However, as indicated by the age-sex structure analyses above, Headingley is projected to have a younger

population than the Leeds average. It was probably the reason for the lower mortality, but considerably higher migration rates in Headingley than that of Cookridge and the average of Leeds (Table 7.2). As previously described, Headingley has a large university student population. Such findings are consistent with the local characteristics. The impact of the university student population may also help to explain the lower fertility rates, as the students are more unlikely to give birth during their higher education than other people of the same age. On the other hand, Cookridge is a settled suburban area. Therefore it is understandable that it has a higher mortality, but much lower migration rates in this area.

**Table 7.2 Components of change of small area populations projected by Model B**

Model	Model A			Cookridge			Headingley		
	2002	2007	2031	2002	2007	2031	2002	2007	2031
Total population	720,000	759,400	974,300	21529	22778	29618	28158	38330	53425
Fertility	0.048888	0.053162	0.054042	0.038286	0.044156	0.049481	0.020399	0.026284	0.034059
Mortality	0.009722	0.008428	0.006364	0.009615	0.008868	0.008036	0.007778	0.005896	0.006364
Internal In	0.038889	0.040163	0.031305	0.039110	0.040653	0.030488	0.085766	0.071484	0.056584
Internal Out	0.041250	0.040953	0.037463	0.041943	0.040697	0.036768	0.051744	0.067258	0.059785
Immigration	0.013611	0.017777	0.016935	0.013424	0.017824	0.014923	0.010477	0.013045	0.014413
Emigration	0.006389	0.006321	0.008314	0.006178	0.006322	0.007934	0.004723	0.004774	0.007674

#### 7.5.4.2 Model C

The age-sex structure analysis has found that Model C suggests more disaggregated variation from the aggregate model than Model B. Looking into the projected components of changes in the two small areas by Model C, the patterns of the Model C results in Cookridge and Headingley confirm again that spatial variation does have an important impact on the population

changes. In fact, the Model C projections suggest an even bigger difference between the two wards.

It can be seen from Table 7.3 that Model C projects much lower fertility and mortality in Headingley than in Cookridge and the Leeds average. On the other hand, all four migration flows are considerably higher than that in Cookridge or Leeds average. In fact, nearly all flows in Headingley are tripled the average migration rates elsewhere in Leeds. From the components of change analysis of the small area populations, it seems that the full disaggregation has caused even more spatial variation (Table 7.3).

**Table 7.3 Components of change of small area populations projected by Model C**

Model	Model A			Cookridge			Headingley		
	2002	2007	2031	2002	2007	2031	2002	2007	2031
Total population	720,000	759,400	974,300	20756	20783	28426	31132	52240	121239
Fertility	0.048888	0.053162	0.054042	0.043478	0.046078	0.047684	0.008478	0.015554	0.031500
Mortality	0.009722	0.008428	0.006364	0.012093	0.012077	0.005945	0.004690	0.003235	0.004116
Internal In	0.038889	0.040163	0.031305	0.028474	0.035991	0.033279	0.173262	0.119334	0.073244
Internal Out	0.041250	0.040953	0.037463	0.030015	0.034596	0.039049	0.142040	0.115907	0.085022
Immigration	0.013611	0.017777	0.016935	0.008817	0.013809	0.019032	0.046255	0.044047	0.037529
Emigration	0.006389	0.006321	0.008314	0.004722	0.005581	0.008197	0.019947	0.014835	0.016612

Looking back at the discussion on the aggregated results produced by the three models in Section 7.5.2.4, the Model C results seems to indicate variation in the migration trends, especially in internal migration out flows and immigration flows. In light of the analyses of the small area populations above, an experiment is provided to illustrate the seemingly unaligned internal migration out is actually still following the same trends in Table 7.4. On the aggregated level, the simulation results indicate that the internal migration out flows in Model C does not follow the trend of gradually decrease over



time. However, as can be seen in Table 7.4, this merely demonstrates the impact of disaggregation. The disaggregated out-migration probabilities that still give an average (0.0037461) are applied to populations in small areas in simulation year 2031. This average is close to the aggregate probability and is lower than the average of out-migration probabilities in 2030 (0.0037469). However, when the number of the projected migrants are re-aggregated and divided by the number of total population, the results of 0.056081 turns out to be larger than 0.055985 in 2030 (Table 7.4).

The example in Table 7.4 proves that the impact of the disaggregated characteristics is such even the application of lower probabilities can end up with a higher outcome. Examples like this demonstrate the strength of the disaggregated model over the aggregate model, where individual and local characteristics are retained in the simulation.

**Table 7.4 Disaggregation impact on internal migration out in Model C**

	Ward 1	Ward 2	...	Ward 32	Ward 33	Total	Average probability
<b>2030 start population</b>	33009	34125	...	10518	22983	1043328	N/A
<b>out migrants</b>	813.138	883.129	...	101.424	304.243	58411	0.055985
<b>out migration probability</b>	0.02463	0.02588	...	0.00964	0.01324	N/A	0.037469
<b>2031 start population</b>	33364	34598	...	10565	23146	1056791	N/A
<b>out migrants</b>	821.749	895.224	...	101.847	305.527	59266	0.056081
<b>out migration probability</b>	0.02463	0.02588	...	0.00964	0.0132	N/A	0.037461

On the other hand, previous studies have pointed out the issues with the ONS migration projections. Boden and Rees (2010), during their investigation of designing a new migration framework for UK, have found differences of migration data varied from complementary administrative sources such as National Insurance and GP registrations at a national,

regional and local level. The fact that substantial population change varies considerably across the local authorities of the UK can result in profound change in the size and composition of the local populations. However, in the ONS SNPP projections, the international migration is calculated by distributing the national projection assumptions of migration and like internal migration, future assumptions on the distribution of the international migration is based on observed patterns in the five years prior to the projections base year (ONS, 2008e). In one of ONS' accuracy report, the 2006-based SNPP for Yorkshire & Humberside and East Midlands have been found too high in comparison with the 2007 MYE (ONSCD, 2008).

Therefore, future work should look into other empirical studies of Leeds population, align the model to assumptions of the locally based empirical model and explore the alternative projection patterns with the ONS aligned results to further assess the disaggregation impact on Leeds population changes.

## 7.6 Conclusions

The dynamic spatial MSM is known to be as hard to validate. The unaligned MSM can drift away from the official projections. Due to both the need for system validation and the user requirements, alignment has become an increasingly popular practice for MSMs. This chapter introduces two exercises of model alignments to the official model. Two models B and C have been designed to align and test the consistency and robustness against the ONS aggregate model, Model A. The projections for Leeds by Model A for the period of 2007-2031 are used in the alignment exercises. For the period prior 2006, the ONS mid-year estimations for the Leeds population changes have been used instead. Model B is a naïvely disaggregated model that distributes the aggregated probabilities from Model A into all small areas, while Model C is a fully disaggregated model that captures the local population characteristics by sex, age and small area. Results of the two aligned models are re-aggregated to the level of Leeds, analysed and compared to results of Model A.

At the aggregate level of Leeds, Model B reveals a good consistency with the Model A. The age-sex structure analysis indicates similar patterns to that of Model A. However, Model B produces a larger and slightly younger population than Model A. Three reasons may contribute the differences. The disaggregated characteristics captured in the baseline population may cause some variations, despite the application of averaged probabilities for updating; the ABM of student migration without the same migration assumptions, may introduce more younger population into certain small areas than others and bring changes to the age-sex structure and finally Model B starts with a slightly larger population that contains different population characteristics than Model A. There are some variances in Model C. It produces an even larger and younger population than Model B. In the components of change analyses, Model C results suggest slightly different migration patterns than Model A and the migration levels are higher than that of Model A. Generally speaking, results from both aligned models demonstrate patterns that are close to the aggregate Model A and a consistency have been found throughout the model results.

As disaggregation is the main difference between the three models, a series of analyses have been conducted using the projection results from two small areas, Cookridge and Headingley, which have been used in analyses in previous chapters. On the disaggregate level, Cookridge seems to produce patterns that are closer to what Model A projects in terms of population changes, while results of Headingley suggest slight variances within the components of changes. However, given the distinctive local population characteristics and migration patterns in Headingley, such differences are not surprising. As indicated in the findings in Section 7.5.4, despite some variances in the small areas, both Model B and C still produce patterns that are consistent with Model A. Although with the full disaggregation, Model C still produces the population changing trends that are mostly consistent with Model B and A.

Such findings suggest that the hybrid MSM is robust enough to be able to provide the consistency needed through different disaggregation. The

variations within the Model C and B also demonstrate the importance of capturing the local characteristics and individual changes, which are overlooked in the ONS aggregate model.

As pointed out by other studies, there are some issues with the ONS SNPP model. Therefore, the next step of the future work may be looking into other empirical studies of Leeds population, such as UPTAP (Understanding Population Trends and Processes), project undertaken by Rees *et al.* (2010), which captures the local characteristics. Then we would try to align the hybrid MSM to assumptions of the locally based empirical model and explore the alternative projection patterns with the ONS aligned results to further assess the disaggregation impact on Leeds population changes. Investigations of more empirical studies and datasets will be carried out in future work. The findings will be used to help refine this model so that the model provides a better representation of the Leeds population.

## Chapter 8

# Health application: obesity scenarios

## 8.1 Introduction

As described in previous chapters, the hybrid modelling approach brings together the strength of two closely related individual based modelling approaches to accommodate both the theoretical and practical demands of this model. The dynamic spatial MSM has provided the theoretical basis for modelling a large complex social system at a fine geographical scale. It enables the projection of the studied population with individual characteristics that change annually through important demographic transitions within a local context. Through the added ABM features, the goals of both the effective handling of a large scale system and the extra flexibility to model various behaviours of sub-populations in different scenarios have been achieved.

As a demographic planning tool, this model can monitor the evolution of population structures and various demographic change patterns at a fine geographical scale. As people are at the centre of the society, the model can

also provide a useful tool in assisting decision making in public planning or policy formulation, as well as exploring various “what if” scenarios and testing different hypotheses. There are many possible ways to apply the model.

Public health planning is one potential application area that interacts intensively with the population changes. As obesity has become one focus for global epidemic studies, as well as a focus of UK public policies (WHO, 2004a; NHS, 2010), an application model of obesity projection has been developed to demonstrate one of the potential applications of this hybrid model, aiming to facilitate public health planning.

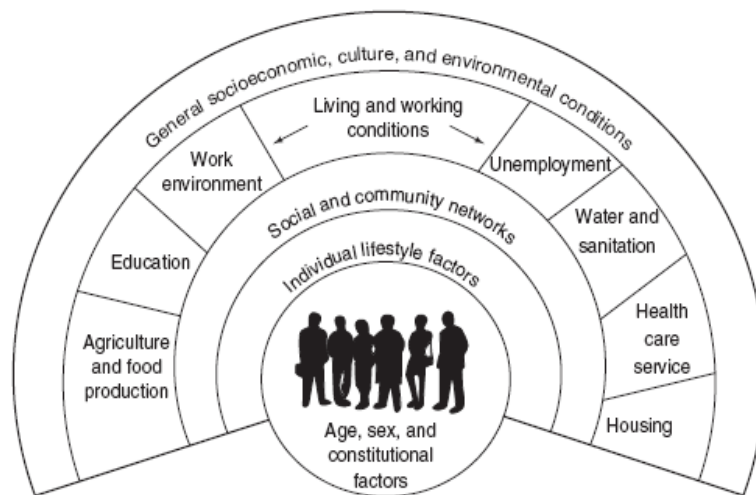
## 8.2 Public health planning with a focus on obesity

### *8.2.1 Public health planning*

Public health literally means the health of the public. Public health planning is the implementation of measures needed to maintain and improve health levels in the population. Public health officials fulfil two important functions: they maintain health levels of the population in collaboration with statistical agencies and they organise campaigns of illness prevention and increasingly wellness development for conditions where there are collective benefits. One example is the programme of childhood immunisation against infectious diseases. Each child who is immunised reduces the pool of the potentially infected and the incidence of disease. Medical science continues to identify infecting agents and to develop preventive exercises. Another example is the campaign to reduce the level of smoking by people through restricting the availability of cigarettes, banning advertising of cigarettes, reducing the opportunity to smoke in public places where one person’s pleasure motivated exhaled smoke is other people’s unwanted inhalation.

Public health reflects complex interactions between population characteristics and the environment. The understanding of small area effects

on health helps explain the cause, facilitate the prevention and/or treatments and enable the design and implementation of targeted interventions and policies. Traditionally public health policy has laid emphasis on implementing biomedical solutions to public health problems, e.g. the implementation of cancer screening campaigns based on careful tests. Though the long-term effectiveness of the biomedical model cannot be denied, there have been some critical perspectives. Among those critics, studies argue that there was a need to recognise the social–theoretic context of health (Moon, 2009). Dahlgren and Whitehead (1991) have pointed out that people’s health is the result of the complex interactions between the individual demographic factors and their socioeconomic, culture and environmental conditions, ranging from the agriculture and food production, education, work environment, (un)employment, water and sanitation, health care service to housing conditions (Figure 8.1).



**Figure 8.1 The Dahlgren-Whitehead health model**

Source: Dahlgren and Whitehead (1991)

The public’s health reflects complex interactions between population characteristics and the environment. Although genetics can predispose people towards certain health conditions, economic status, access to health care and life styles also have important impact on the health of individuals. In fact, Wilkinson and Marmot (2003) pointed out that environmental

factors (both physical and social) are the most common underlying causes of the ill health. Changes in environmental factors have a more immediate impact than the genetic changes, as they reflect the changes in the way people live. The environmental (especially social) factors also play an important role in health inequalities. After all, the relief of deprivation is not just simply the provision of materials and services if who gets these is socially determined. Public policies therefore play an important role in shaping the social environment to promote public health (Marmot, 2005). The Marmot Review (2010) on health inequalities in England proposes a new way to reduce health inequalities in England post-2010. The review suggests that the conditions in which people are born, grow, live, work and age and can lead to health inequalities. The report argues that, to improve health for all of us and to reduce health inequalities, action is needed to reduce the disadvantages faced by the poorest groups in society.

There are many reasons why it is difficult for the lower socioeconomic groups to adopt beneficial health behaviours. Firstly, information about how to behave healthily may not reach some of such groups; secondly, they may lack the material resources and the environments to live healthily and thirdly, for people living difficult lives, changing health behaviour is unlikely to be a major priority. Therefore, promoting people's well being through studies about who they are, where they live, and how they live within a spatial framework provides a more holistic approach to public health than the traditional medical geography. In fact, some health geographers view a place as an experiential construct that is negotiated, flexible, contingent, and socially constructed, instead of simply an activity container (Moon, 2009). Place is clearly important when considering public health and social inequalities. For instance, the access to public health services has a direct and profound impact on local population's well being (Kearns 1993). On the other hand, the understanding of place effects on individual health could help to explain the cause or the diffusion pattern of diseases as well as providing additional information to facilitate preventions and/or treatments. A national guarantee of free health care at the point of delivery contributes to everybody's wellbeing. However, local interventions



to persuade people to consult their doctors early in the course of a disease are also important. An increase in our understanding of health patterns within a geographical context can enable the design and implementation of more informed and targeted interventions and health policies for the promotion of public well being. Such a targeted method is more effective than a broad-brush approach (Lang and Rayner, 2007). This application model projects the individual obesity changes within wards to capture both the individual and small area characteristics.

### *8.2.2 Obesity focus in public health planning*

The definition of obesity is the excess of body fat sufficient to be detrimental to health (WHO, 2006). Obesity is often measured by BMI (Body Mass Index) and details of BMI will be described in Section 8.3.1. As health risks, obesity related diseases account for a substantial proportion of costs of health care resources worldwide. Previous research has pointed out that the treatment of obesity is not only expensive, but also at best is difficult and time consuming (WHO, 2004a). Obesity is an important risk factor for premature death (Allison *et al.*, 1999; Fontaine *et al.*, 2003; Flegal *et al.*, 2005) and health problems like diabetes, gallbladder disease, coronary heart disease, high cholesterol, hypertension and asthma (Must *et al.*, 1999; Mokdad *et al.*, 2001; McTigue *et al.*, 2006). More recently research has also linked obesity with the incidence of several cancers (White, 2003, Stamatakis, 2006). Excess weight reduces the quality of life, raises medical expenditures, places stress on the health care system and results in productivity losses due to disability, illness and premature mortality (Quesenberry *et al.*, 1998; Finkelstein *et al.*, 2003; Andreyeva *et al.*, 2004).

The risk of a global epidemic of obesity has been identified in various documents of WHO (2004a, 2010). A number of UK studies (NAO, 2001; Foresight Programme, 2009) have produced estimates of the differences in life expectancy of obese persons compared with normal weight persons of between 1.5-9 years. However, the emerging consensus is that the additional morbidity associated with obesity is the more important public

health problem than premature mortality. Obesity has become a fast growing problem in the UK in recent years (British Nutrition Foundation, 1999; House of Commons Health Committee, 2004; NAO, 2001). The Foresight Programme (2009) in UK found that one in four adults is now obese. The obesity has nearly doubled for men from 1993 – 2004 (13% to 24%) and obesity prevalence for the period 1995–2004 has increased from 14% to 24% for boys and from 15% to 26% for girls (UK Growth Chart definitions). In UK, if current trends continue obesity is expected to overtake smoking as the primary cause of premature deaths in the next few years (House of Commons Health Committee, 2004). The Government now is paying particular attention to the rising levels of obesity (Moon *et al.*, 2007).

The Government's health White Paper (DoH, 2004) recognises that encouraging healthy choices and associated behavioural change is a complex process and it requires more than just increasing public awareness of health issues. Therefore policy making needs to move away from considering disease groupings in isolation, towards a population approach that considers the determinants of health problems. WHO (2004a) states that major social and environmental changes are required to make healthier choices more accessible and preferable to prevent obesity. In the Parma Declaration (WHO, 2010), WHO has identified the key challenges of environment and health and made it their mission to reduce the health inequalities posed on the vulnerable populations by both poor physical and human environmental conditions. The environmental approach allows significant benefits be achieved from small changes when the improved environment benefits a large number of people (Swinburn and Egger, 2002). Indeed, reducing obesity and health inequalities are now at the centre of the UK Government's health policy. In this light, WHO also encourages the development of knowledge and tools for policy-making and implementation that help to tackle such environment and health challenges (WHO, 2010).

### 8.2.3 *Previous obesity research findings*

Previous researches have suggested that the prevalence of obesity is not uniform across the whole of a studied geography. Instead, obesity and overweight are subject to area effects at scales, varying from the very local (Edwards and Clarke, 2009) to the regional (Rami *et al.*, 2004, Moon *et al.*, 2007) and national (Chaix and Chauvin, 2003; Dollman and Pilgrim, 2005; Kim *et al.*, 2006). However, individual factors alone cannot however explain geographical variations in obesity rates (Ellaway *et al.*, 1997; Kahn *et al.*, 1998). Edwards (2010) pointed out that obesity is a complex interaction between biophysical, social, environmental and psychological factors. The spatial patterns of obesogenic environments allow us to identify the socio-economic, demographic, and environmental behaviours at specific times and locations.

Therefore, various researchers have attempted to model obesity from different angles: the aggregate models tend to focus on the environment while the qualitative models focus on the individual characteristics. The Multi Level Models (MLMs) attempt to capture environmental impact in modelling the health outcomes/behaviour in areas (Moon *et al.* 2007; Pearce and Witten, 2010; Twigg and Moon, 2002). Such MLMs normally consider the proportion of cases explained by various individual factors (e.g. age and sex), then introduce area characteristics (e.g. deprivation) to explain the remaining cases. Based on this, the model's parameter estimates can be applied to the individuals and areas to create new prevalence estimates (Twigg and Moon, 2002). However, such estimates remain as the zonal average that cannot truly reflect the individual characteristics. More seriously, if different zoning systems are used for individual and area characteristics, it can produce different spatial patterns than intended. Edwards (2010) suggests that aggregated models are too far removed from the actuality of the real life and the individual-based qualitative data on individuals cannot provide a system wide understanding of the environment

and its potential relationship with obesity. Small area level models may offer the best compromise between two methodologies and allow compact problem areas of high risk to be identified and analysed whilst capturing the important local characteristics (Edwards, 2010).

Previous researchers also point out that the treatment of obesity is not only expensive, but also at best difficult and time consuming. Therefore the prevention of obesity is just as important as the treatment (WHO, 2004a). The micro spatial analysis concept enables governments and health professionals to respond to local differences in health behaviours, and to develop and implement more targeted interventions and health policies for prevention. It also allows conceptualising and devising population-level and place-level interventions and health policies to help prevent obesity, transforming the interventions that focus only on individual-level activities.

As part of the advisory group to help Government think systematically about the future, Foresight Programme (2009) was commissioned by the Government to produce a long-term vision of how to deliver a sustainable response to obesity in the UK over the next 40 years. The Tackling Obesities programme projects the changes of obesity rates through to 2050 and predicts the consequences for health, health costs and life expectancy. Four scenarios have been developed in this project to assess the impact of individual and public decisions on the obesity changes. This project represents the first major development and utilisation of a dynamic MSM of chronic disease in the UK. The report authors believe that this work demonstrates that a public-health-orientated, dynamic MSM, although with limitations, offers the best tool for estimating future levels of avoidable chronic disease and anticipating the scale of interventions required to have a significant impact.

Our application model of obesity attempts to project the obesity changes within the small areas in Leeds, using the hybrid dynamic spatial MSM. Three scenarios adopt the assumptions used in the Foresight Programme, which have been developed to explore some “what if” situations. Section 8.3 describes the model in detail.

## 8.3 Obesity application model

As discussed in Section 8.2, it is important to include the impact of a person's environment when studying and planning people's health. Understanding the characteristics of the obesity environment and moving the study of obesity towards a population approach rather than isolated groups of individuals has considerable potential for developing the theoretical understanding of obesity, as well as delivering successful policy interventions (DoH 2004). Using the dynamic spatial MSM, changes in both the characteristics of the locality and the characteristics of individuals can be simulated simultaneously in small area populations. In the following sections, the application model will be described and the results will be analysed to demonstrate the potential use of the model.

### *8.3.1 Modelling method*

#### *8.3.1.1 BMI approach*

The definition of obesity is the excess of body fat sufficient to be detrimental to health (WHO, 2006). In this application model of obesity projections, obesity is defined on the basis of Body Mass Index (BMI). BMI is a weight-for-height index that is commonly used to classify overweight and obesity. BMI is a useful proxy measurement of body fat at a population perspective. The BMI has been validated by the IOTF (International Obesity Task Force) and its use to measure obesity is generally accepted (Dietz and Bellizzi, 1999). The BMI of a person is calculated as their weight in kilograms divided by the square of their height in metres (Equation 8.1).

$$BMI = \frac{weight(kg)}{height^2(metres)} \quad (\text{Equation 8.1})$$

The definitions of different categories of BMI are listed in Table 8.1.

**Table 8.1 Definitions of BMI**

<b>Definition</b>	<b>BMI range (kg/m<sup>2</sup>)</b>
Underweight	Under 18.5
Normal	18.5 to less than 25
Overweight	25 to less than 30
Obese	30 to less than 40
Obese I	30 to less than 35
Obese II	35 to less than 40
Morbidly obese	40 and over
Overweight including obese	25 and over
Obese including morbidly obese	30 and over

Source: NHS, 2010

An adult, either male or female, is considered to be obese if their BMI score is or over 30 (Table 8.1). Child obesity varies with the stage of the growth. During childhood BMI shows significant variations, which is the reason that age and gender specific reference standards must be used. This study adopts the BMI approach and used single year and sex specific probabilities to simulate the obesity changes in children aged 2-15. Obesity of population aged over 16 is modelled in age groups of 10 years until ages 75 or over.

BMI provides the most useful population-level measure of the overweight condition and obesity, as it is the same for both sexes and for all ages of adults. However, BMI may not correspond to the same degree of fatness in different individuals and populations due to different body proportions. The health risks associated with increasing BMI are continuous and the interpretation of BMI scores in relation to risk may differ for different populations (WHO, 2004b). For example, children's and adolescents' BMI varies with age and sex, due to different stages of growth. The standards for Asians also vary slightly. The WHO (2004b) Expert Consultation concluded that the proportion of Asian people with a high risk of type 2 diabetes and cardiovascular disease is substantial at BMI's lower than the existing WHO cut-off point for overweight. However, the consultation recommended that the current WHO BMI cut-off points (Table 8.1) should be retained as the international classification. Despite the limitations, BMI measurement has

been widely used in obesity studies, due to its simplicity and practicality (It is relatively easy to obtain weight and height information of individuals).

#### *8.3.1.2 Data*

As in the other parts of the dynamic spatial MSM, the base population is recreated from Samples of Anonymous Records from the Census 2001 data (ONS, 2001m). The process of the population recreation has been conducted by other colleagues and described elsewhere (Birkin, 2006). The base population data provides the demographic and socio-economic characteristics of individuals, as well as information on their geographical locations. All individuals are simulated within households that are located in small areas (wards). Each year the obesity probabilities are applied to the individuals according to their demographic, socio-economic and spatial characteristics to simulate the changes in their obesity statuses.

The obesity information is derived using the obesity trends that have been published in the Health Survey of England (HSE) trend tables and revealed through the analyses using the HSE raw data. The obesity probabilities were computed using HSE data for 2008, which was the most recent information at the time that the model was developed.

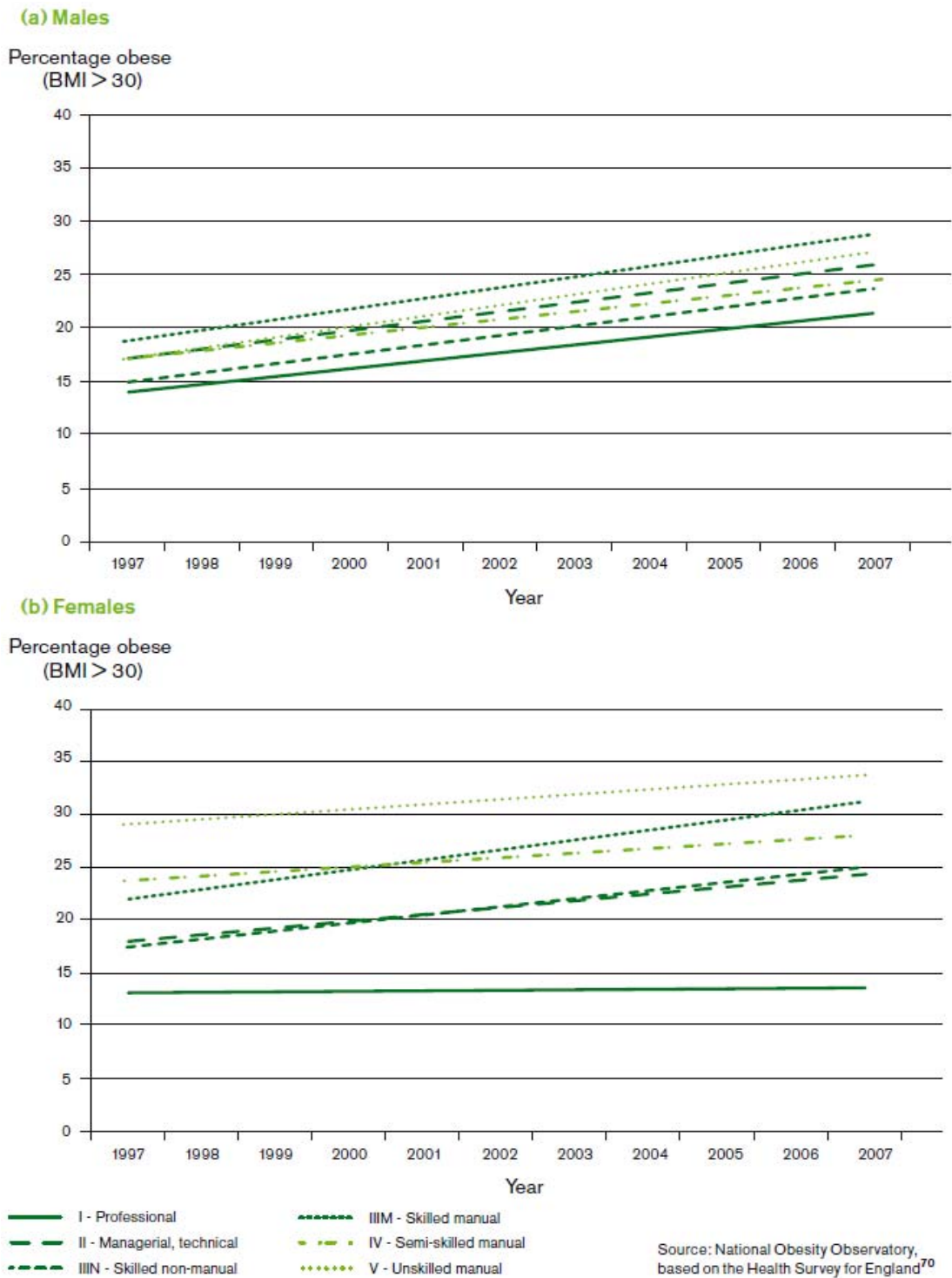
#### *8.3.1.3 Risk factors*

The aetiology of obesity is complex and multi factorial, so it is impossible to study all obesogenic factors simultaneously. In Foresight's study, 29 drivers of change under the categories of Health, Science and technology, Leadership, Education and information, Shape of society, Values and behaviour, Resource constraints, Economy, Food production and retailing and Living environment have been considered to be important for modelling obesity. However, due to the limited resources, in the development of this application model, it is necessary to focus on several key obesogenic factors identified as having a significant relationship with BMI, as it is not the aim of this study to identify and explore the full range of risk factors of obesity. Also due to the limitation of the data availability, it is important for this

study to identify not only the obesogenic variables that correlate with BMI, but also to ensure that the same variables that are available in both the population data (census ward) and obesity data (survey ward in HSE).

There is a strong argument that we must include the factors of age and sex, because obesity, like most demographic changes, varies systematically across the age range and different sexes experience the obesity in different patterns (Foresight programme, 2009). Lifestyle variables (for instance, physical exercises, smoking, fruit and vegetable intake) have been estimated for small areas using aggregate statistical model (Twigg *et al.*, 2005) or static microsimulations (Edwards and Clarke, 2009), but integration into the dynamic MSM would be challenging. Unlike the statistical or static models, where a weight can be used to update the obesity changes, the dynamic MSM requires a good understanding of what factors and how they drive changes in those lifestyle variables in order to calculate the transition probabilities for the dynamic ageing process each year. It is also difficult and computationally demanding to update each of these factor changes at each step of simulation for each individual. Due to such reasons, when modelling previous demographic changes, location has been used as a useful proxy variable for the simultaneous operation of socio-economic, ethnic, life style and environmental variables. However, it is hard to obtain reliable small area obesity prevalence data. To compensate the lack of the small area obesity information, a third attribute, the socio-economic status (SES) has been added. The impact of SES on people's health is well-recognised in previous literatures (Marmot, 2005; Marmot, 2010; Foresight programme, 2009; McLaren, 2007; Stunkard, 2007). Figure 8.2 demonstrates the impact of social class on obesity prevalence.





**Figure 8.2 Obesity prevalence of ages 16 and over by social class, 1997–2007**

Source: The Marmot Review Commission (2010)

As Figure 8.2 illustrates, socioeconomic status does have an important impact on obesity and the lower socioeconomic groups seem to experience more obesity. In our model, the occupationally based classification of

NSSEC (National Statistics Socio-economic Classification) of individuals is used to assess the socioeconomic impact on obesity, as both the population data and HSE (Health Survey for England) have this attribute. Here the coarse three categories of NSSEC are used, namely:

- NSSEC1: Managerial/Professional;
- NSSEC2: Intermediate and
- NSSEC3: Manual/Routine workers.

This classification is used to accommodate the different NSSEC categories between the census data and health survey data.

### 8.3.2 *Description of the model and scenarios*

The obesity application model projects the obesity changes within small areas at the individual level using Model C. Using the dynamic spatial MSM, it assumes that the obesity risk is a function of age, sex and socio-economic status. An individual becomes obese if

$$Ran(0,1) \leq P(o|a,s,S) \quad (\text{Equation 8.2})$$

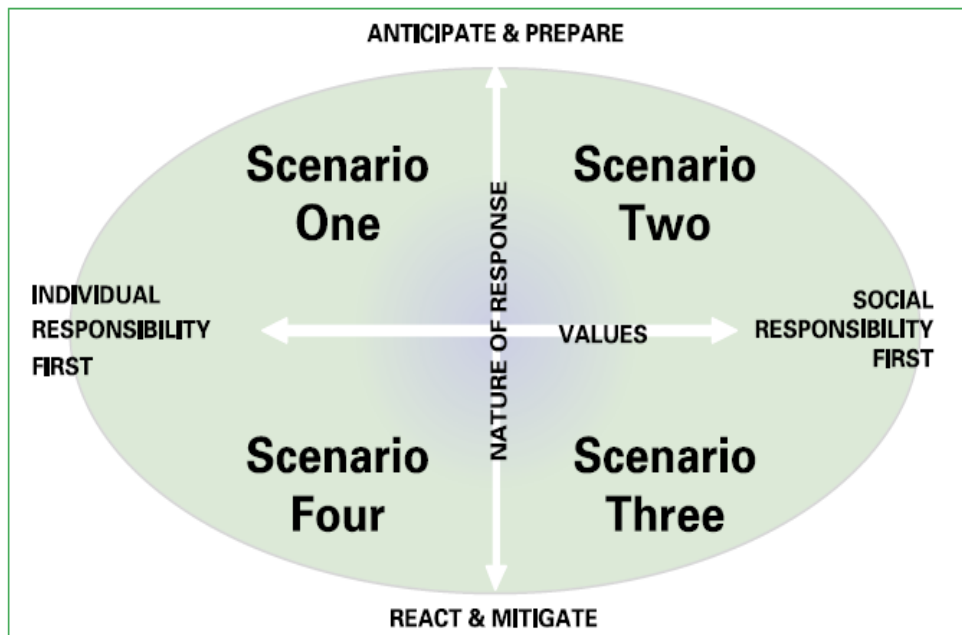
where  $Ran(0,1)$  is a random number between 0 and 1,  $a$  is the age,  $s$  is the sex and  $S$  is the socio-economic status (SES) of the individual.  $P(o|a,s,S)$  is the probability of the state of obesity  $o$  occurring to an individual of age  $a$ , sex  $s$  with a NSSEC category  $S$ . This model does not capture the process of obese population becoming normal weight, based on the assumption that obesity is hard to cure and weight reduction is at best time-consuming (section 8.2.3). Therefore it is hard to reflect this in the yearly changes. Although SES is used in the initial probability calculation, the evolution of SES is not simulated in this model. This and other limitations are discussed in section 8.5 and the need for more refinement and validation in future work is also recognised.

The projected individuals are simulated with obesity probabilities that are calculated on the basis of obesity data from the Health Survey for England (HSE) 2008 (NHS, 2009), which is the most up-to-date data available on obesity at the time of model development. The HSE estimates the obesity rates for the relevant year to the aggregate population level of England. The bases for these survey estimates are provided in the HSE 2008 trend tables (NHS, 2009).

Then these obesity probabilities are applied to individual population records recreated from the census data according to age, sex and SES attributes to produce the annual projection of prevalence of obesity in small areas. As described before, children's BMI is more complicated due to the variations during growth stages. Therefore only the obesity of children in ages 2-15 is projected in this model. The single year of age, sex and SES specific obesity probabilities are used in child obesity modelling. For adults, age is in a cruder band of 10 years until ages 75 and over. The MSM projected the annual changes in the obese population in Leeds wards for 30 years and the model results have been analysed on both LA level and ward level to understand the obesity changes in small area populations.

To demonstrate the strength of MSM in implementing "what if" explorations, three scenarios have been developed to explore the impact on the obesity patterns of individual and public decisions. The scenarios use the assumptions in some of the Foresight scenarios (2009) as illustrated in Figure 8.3.

As Figure 8.3 indicates, Scenario 1 and 4 emphasise individual responsibility, while Scenario 2 and 3 emphasise social responsibility. Our application model adopts some of the assumptions found in Foresight's report in Scenario 1 and 4 to assess the impact of individual decisions on the obesity patterns. Assumptions in Scenario 2 are also used to create a scenario that demonstrates the impact on the obesity patterns when public decision making is involved.



**Figure 8.3 Foresight obesity scenarios**  
 Source: Foresight Programme (2009)

The three scenarios are described in more details below:

- Scenario 1- Individual Anticipation (Foresight scenario 1):

This scenario is constructed with the assumptions that individuals are responsible for their own obesity and have a long-term focus of anticipation on obesity. However, such an effort is insufficient on its own to overcome the impact of an obesogenic environment, especially the built environment, to create a population-level effect. As a consequence, the obesity level will increase from the current level. In their analysis, the Foresight report (2009) uses a qualitative point of +1 to represent the increase in adult obesity (qualitative scale of 0-3), as the HSE data indicate that obesity probabilities for ages 21-60 from year 2008 to year 2030 would increase by 30% approximately (Foresight report, 2009), our model uses an annual increase of 0.001 to update adult obesity probabilities in 2001. Under the assumptions that a long-term focus on child welfare, generational issues and ill-health prevention start to pay off, the child obesity level has been predicted to be 1.5 points lower than current (Chipperfield *et al.*, 2007) and an annual reduction of 0.0015 is used to update the child obesity probabilities in the model.

- Scenario 2 - Social Anticipation (Foresight scenario 2):

The assumption of this scenario is that within a socially responsible, community-based society with support for long-term systemic change, the key features of society and environment will indirectly influence obesity levels and improve the adult obesity by a point of -2 at the population level. For child obesity, the assumption is that long-term focus, co-ordinated systemic change and interest in the impacts on subsequent generations to improve the child obesity by a point of -3 (Chipperfield *et al.*, 2007). In this model, similar to the practice in simulating Scenario 1, annual reductions of 0.002 and 0.003 are used to update the adult and child obesity probabilities respectively.

- Scenario 3 - Individual Reaction (Foresight scenario 4):

In this scenario, the assumption is that the rise in obesity levels increases dramatically due to the passive reaction attitude towards obesity instead of proactive anticipation. The focus is on treatment and its normalisation instead of prevention and result in a point of +3 for increased adult obesity. Under the assumption that people remain concerned about children, so action is still taken on their behalf, as knowledge and awareness increases, the younger generation does better than the adults, but still scores a point of +1 (Chipperfield *et al.*, 2007). In this model, similar to the practice in simulating Scenario 1, annual increases of 0.003 and 0.001 are used to update the adult and child obesity probabilities respectively.

The Foresight scenarios 1 and 4 are developed to demonstrate the impact without social approach and scenario 2 is chosen instead of scenario 3 to demonstrate the contrast between social and individual approaches towards tackling obesity, under the assumption that Social Anticipation approach (Foresight scenario 2) can make more impact on the population level than the Social Reaction approach (Foresight scenario 3). To understand the variances in different areas and sub-populations to facilitate the purpose of the healthy planning, the results from the 3 scenarios have also been analysed with a focus on children to provide information on child obesity.

The discussions on the results under the three different “what if” scenarios is described in Section 8.4.

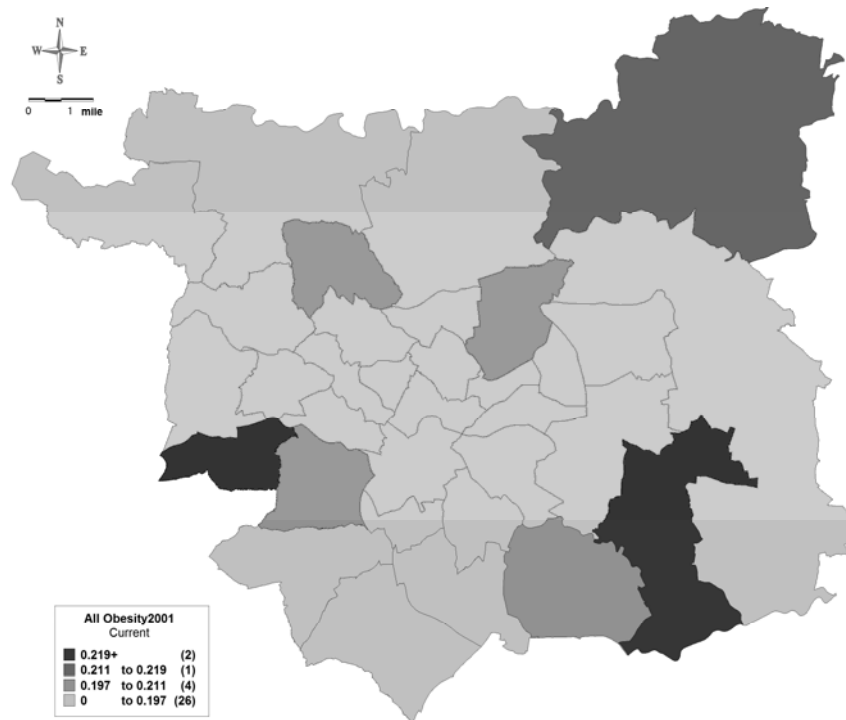
## 8.4 Analysis of results

The obesity changes have been projected for populations in Leeds wards. Three sets of analyses have been carried out to demonstrate the use of the model results: we explore the projected obesity patterns for Leeds as a whole, for individual wards of Leeds and for sub-populations by demographic and socio-economic characteristics. As preventing child obesity is particularly important in the obesity prevention and control, this sub-population has been selected for analysis. The same two wards, Cookridge and Headingley, have been used here again here to demonstrate the small area variance.

### *8.4.1 Obesity patterns at LA level*

The model results of the three scenarios have been analysed to assess the obesity changes in Leeds population in year 2001 and 2030. All three scenarios start with the same obesity characteristics in year 2001 (Figure 8.4) and the same ranges are used in Figures 8.4 and 8.5 for the purpose of comparison: under 0.197, 0.197-under 0.211, 0.211-under 0.219 and 0.219+, represented by the shades from light to dark.

At the beginning of the simulation in 2001, the obesity map suggests that the obesity is more serious in the east side of Leeds and only two wards in the west side of Leeds are of higher obesity level and one ward is of the highest obesity level. Overall, the obesity map of 2001 indicates that obesity rates of most of the city are in the lowest range. Only four wards are found with slightly higher rates of 0.197-0.211, 1 ward at the top north east corner with 0.211-0.219 and 2 wards in the south side of the city with obesity rates over 0.219.



**Figure 8.4 Obesity in Leeds 2001: all population**

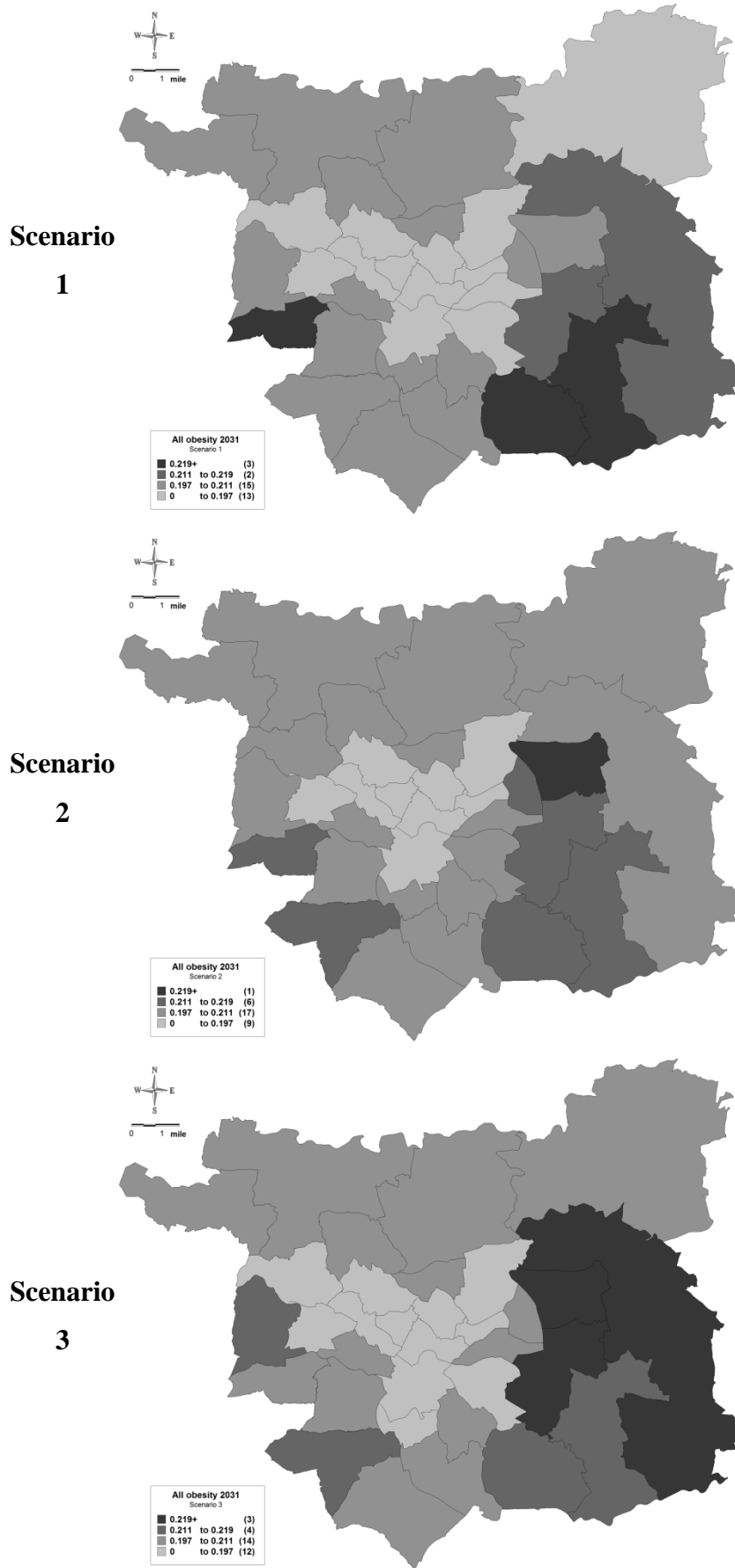
After 30 years' simulation, such patterns have changed. In Figure 8.5, all three sets of scenario results in 2031 suggest that there has been an increase in the level of obesity in Leeds. The zone of lowest obesity level has shrunk to only the areas around the centre of the city. However, the simulation results of all three scenarios in 2031 still suggest some patterns that are consistent with the patterns in 2001: there is a higher degree of obesity in the east of the city than in the west and the lowest obesity levels in area are found around the city centre. The low obesity level around the city centre may suggest the impact of migration. As young people, including university students, enjoy the city life, they tend to move into the city centre area. As described in previous chapters (5 and 6), such annual replenishment of the younger population may help to explain the lower obesity in the city centre than the rest of the city.

However, there are some variations in the results of different scenarios (Figure 8.5). Scenario 1 results indicate a much higher obesity level in the southeast corner of the city than elsewhere. Nearly all top two ranges of high obesity levels are found in this area, except one ward in the east of the city. All wards in the north of the city are in the two lower obesity ranges

and the top northeast ward has even improved to move down to the lowest level of obesity. Compared to the results in 2001, the majority of the Leeds wards has seen an increase of obesity. As Scenario 1 relies on individual anticipation to tackle the obesity, it is hard to make a difference to change the trend of increasing obesity at the population level. The south is significantly worse off in this scenario than the north of the city. This may indicate, when individuals demonstrate individual anticipation towards the obesity epidemic, the more established suburban areas such as those in the north do better than those less affluent areas. Scenario 2 results indicate a different pattern, with a more even distribution of obesity levels throughout the city. In comparison with the other scenarios, there is a significant improvement in the southeast area of the city, where the obesity level is consistently higher in both year 2001 and in the other two scenarios in 2031. Also there is only one ward with the highest level of obesity in Scenario 2, less than in the results of 2001 and in the other two scenarios. As a result, Scenario 2 presents a less contrasting pattern in obesity increase in small areas. This may indicate that the social anticipation approach adopted in Scenario 2 has made an impact on the population level. The social response to the obesity has reduced the difference in obesity levels in the city.

In Scenario 3, on the other hand, a significant increase has been found in most wards throughout the city in the 2031 projection results. This perhaps is not surprising, as Scenario 3 adopts an individual reaction approach, where individuals do not actively seek obesity prevention and lacks of the social response. The results of Scenario 3 in the north of the city are worse than those in Scenario 1 and the south of the city are worse than those in both Scenarios 1 and 2.





**Figure 8.5 Obesity in Leeds 2031: Scenarios 1, 2 and 3**

The three sets of scenario results in Figure 8.5 indicate the impact of different individual and social attitudes towards the obesity epidemic. When social responsibility is emphasised (Scenario 2), it can make a difference on the population level and the results is observable at the city level. The differences in the obesity levels between the small areas are significantly reduced. However, when individual responsibility is emphasised (Scenario 1 and 3), it cannot make an impact on the population level and the gaps between west and east of the city persist. Of the two individual responsible approaches, the individual anticipation approach (Scenario 1) indicates a better outcome than the individual reaction approach (Scenario 3). This indicates that a proactive attitude is more useful in tackling the obesity epidemic, even without relying too much on the support from the society.

#### *8.4.2 Obesity patterns at ward level*

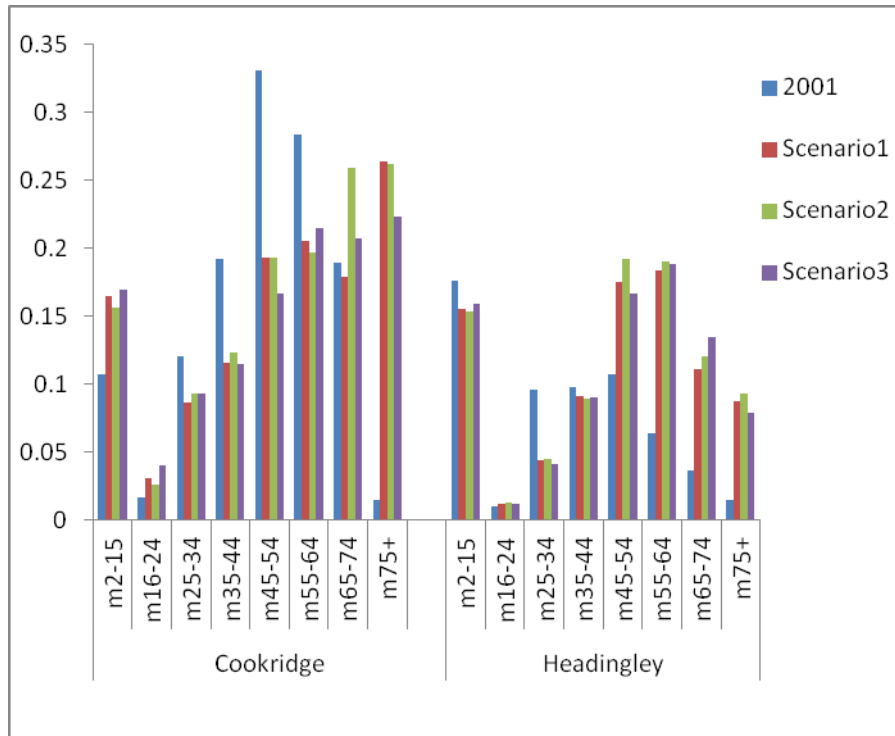
The projection results from wards Cookridge and Headingley are used to explore the spatial variances of obesity changes in small areas. The obesity rates by age and sex within the two small areas at the beginning of the simulation are compared to the projection results at the end of the simulation in Scenarios 1 (Figure 8.6). Child obesity is banded into ages 2-15 instead of single year of ages and adult age groups are 16-24, 25-34, 35-44, 45-54, 55-64, 65-74 and 75+.

As Figure 8.6 indicates, spatial variances reflecting demographic, social economic and other characteristics of the local population in the small areas are captured in the model. In year 2001 Cookridge has a higher obesity level in the adult population (ages 16+) than Headingley, especially in ages 45-54. However, the analysis indicates that child obesity is higher in Headingley than in Cookridge. In fact, ages 2-15 score the highest of all age groups of the Headingley population in 2001. In both wards, the obesity rates increase with the age from the ages 16-24 until ages 45-54. Then they gradually decrease. The lowest obesity is found in ages 75+ for both sexes in both areas in 2001. In terms of the sex differences, in ages 16-24, females have a higher obesity than the males in both wards. From the age 35

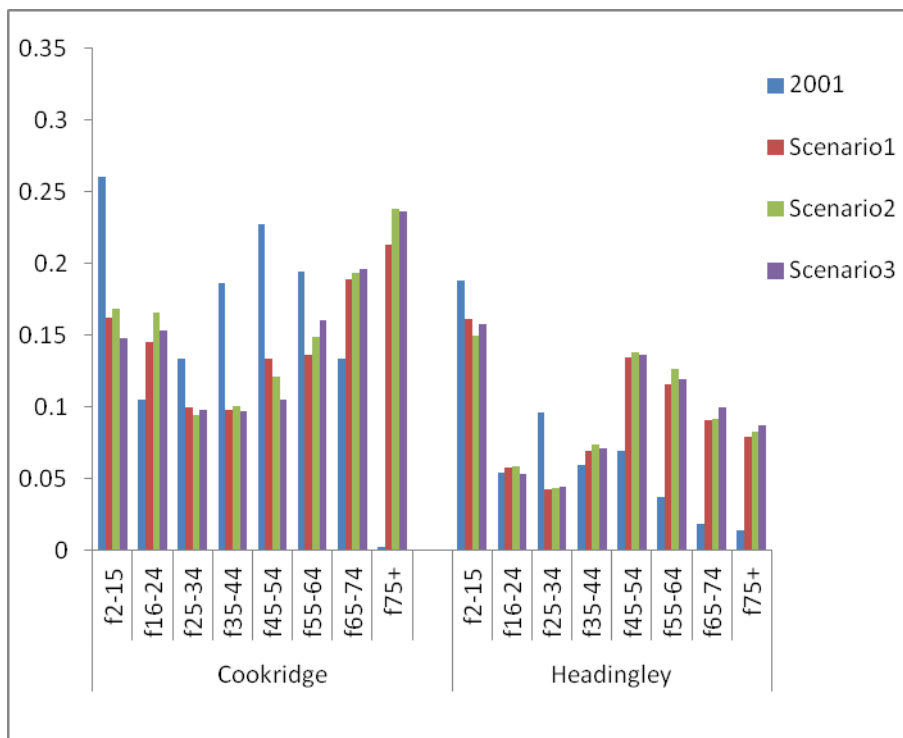
onwards, male obesity is found higher than females in both wards, although the difference is more significant in Cookridge than in Headingley.

However, such obesity patterns have changed in the projections in year 2031. All three Scenario results indicate an increase of obesity in older ages in both wards. Such a shift to obesity in older ages perhaps is not surprising, given the high obesity level in ages 45-74 in 2001. The reduction in younger ages may be the result of the education of child obesity prevention. As in all three scenarios, the assumptions are that adults will try to protect their children from obesity. This is reflected in a reduction of obesity in ages 2-15 for both sexes in Headingley in all scenarios. In contrast, the child obesity in Cookridge keeps growing. There is an obesity reduction in the middle aged population (ages 45-64) in Cookridge, while there is an increase in the same age groups in Headingley. Overall, there has been a much faster growth of adult obesity in Headingley than in Cookridge. However, Headingley still has a lower obesity level than Cookridge, due to the low obesity level at the beginning of the simulation. This perhaps reflects the age structure of the small area populations. Headingley has a much younger population than Cookridge, due to the impact of migration as explained in previous chapters (Chapters 5 and 6). All scenarios indicate that there has been a much faster growth of female obesity than male obesity in both wards. Especially in Headingley, except ages 2-15 and 25-34, obesity levels of the other ages of females in Headingley have increased in all three scenarios in 2031 (Figure 8.6).

### Male



### Female



**Figure 8.6 Obesity rates by age and sex in Cookridge and Headingley: year 2001 and 2031 (scenarios 1, 2 and 3)**

Of all three scenarios, Scenario 2 seems to be more effective in child obesity reduction in both areas. There is a bigger reduction in both areas for ages 2-15 than in other scenarios, except for females in Cookridge. However, it is found less effective in adult obesity reduction. The impact of the different scenarios is not as significant in the two small areas as on the more aggregated spatial level of the whole city. Overall, there has been more growth of obesity for both sexes in Headingley than in Cookridge, where there is a considerable reduction of obesity in ages 35-64. However, the overall obesity level in Cookridge is still higher than that in Headingley. Analyses indicate that older people aged 65+ in Cookridge are more likely to have a higher risk of obesity than those in Headingley. Such differences can be explained by the higher level of obesity in ages 35-54 in Cookridge in 2001 than those in Headingley. The lowest level of obesity is found with males aged 16-24 in both areas and females of the same ages in Headingley, but not with females of the same ages in Cookridge. However, the child obesity level remains very high for both sexes in Headingley. The analysis reveals the spatial variation in the small area populations of age and sex groups.

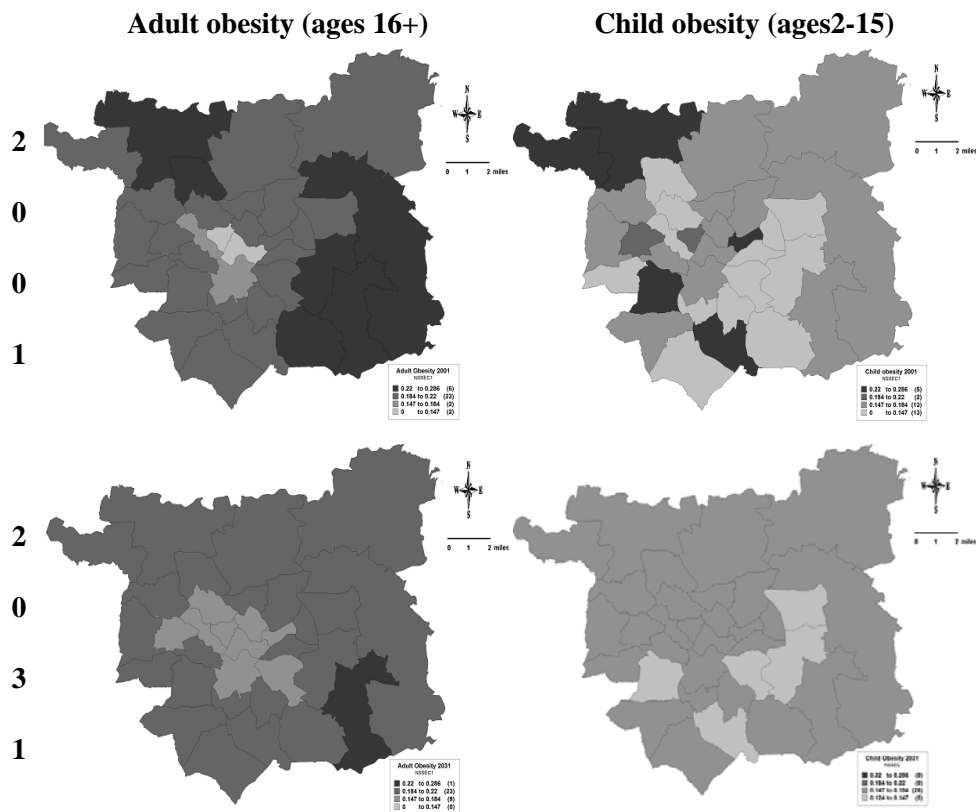
### *8.4.3 Obesity patterns by sub-population*

As discussed in Section 8.4.1 and 8.4.2, the dynamic spatial MSM enables us to analyse the changes of obesity patterns at different spatial scales so as to obtain both the global view, as well as the small area variations. As the MSM is individual based, we can also assess the obesity patterns by different individual characteristics. Three analyses of different populations by different demographic and socioeconomic characteristics will be discussed in this section. Only results from Scenario 1 are used here for illustration purpose.

Age has an important impact on obesity. Obesity changes in different patterns at the different stage of human life. An analysis of the patterns of the child obesity has been carried out against the adult obesity patterns, using the population of the same socioeconomic status (NSSEC1).

Projection results in year 2001 and 2031 are presented in Figure 8.7 to illustrate the obesity patterns at the beginning and the end of the simulation. The shades from light to dark indicate the ranges of: under 0.147, 0.147- under 0.184, 0.184- under 0.22, 0.22 and over in the maps.

From Figure 8.7, we can see that child obesity in Leeds has a different pattern from that of the adult obesity.



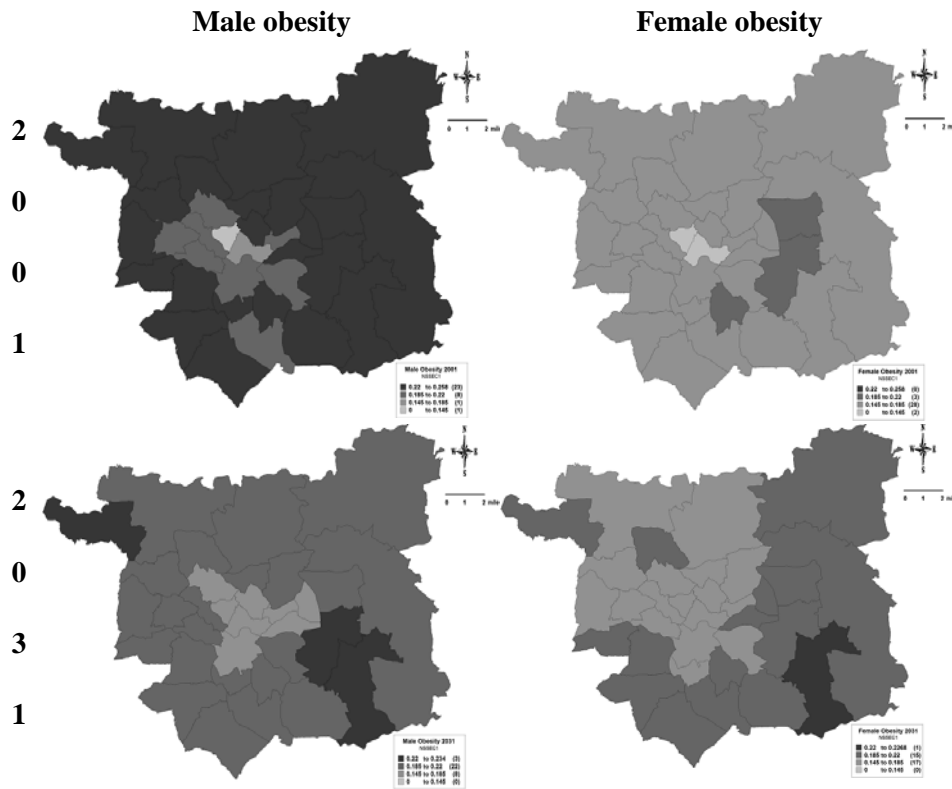
**Figure 8.7 Obesity by age (adult and child) in 2001 and 2031 (NSSEC 1)**

In the projections of 2001, the city centre seems to have a very low level of obesity for adults with a socioeconomic status of NSSEC1 and the higher prevalence of adult obesity have been found in suburban areas, mostly in the southeast and northwest areas. In contrast, the higher child obesity prevalence has been found scattered around the city, although a higher level in the west of Leeds than in the east.

After 30 years' simulation, the level of obesity is less different between the small areas and the results indicate an improvement for both child and adult populations. There is a reduction in the lowest range of the obesity levels of

under 0.147, but an increase in the range of 0.147-udner 0.184, while the same level of obesity remains in most areas. In 2031 projections, there is still significant difference between the adult and child obesity spatial distribution patterns. For adult obesity, only one area in the south (the darkest) has been found with a level of obesity in the range of over 0.22, most areas are in the range of 0.184-under 0.22, while the lower level of adult obesity (0.147-under 0.184) remains in the city centre. On the other hand, child obesity improves even more and the whole city is in the lower ranges of 0- under 0.147 and 0.147-under 0.184. Compared to the adult obesity, the projection results indicate that children in the south are slightly better off than those in the north of the city, to the contrary of the adult obesity distribution patterns. Similarly children in the east are slightly better off than those in the west, which is also to the contrary of the patterns of adult obesity (Figure 8.7).The analyses indicate that in the Individual Anticipation scenario, both adults and children in NSSEC1 group do better in future. This is consistent with the assumption that this sub-population group is likely to act proactively towards obesity prevention, given that they are likely to have more awareness and resource to reduce obesity.

Sex is another factor that has a strong influence on the obesity patterns. In Figure 8.8, we can see the sex impact on the obesity patterns of the population with a socio-economic status of NSSEC1 in Scenario 1 projections. The shades from light to dark indicate the ranges of: under 0.145, 0.145-0.185, 0.185-0.22, 0.22+ in the maps. The maps indicate that there is a much higher level of obesity for males in Leeds than for females, especially in 2001. Although there has been an increase in the projected female obesity in 2031 and a reduction in the male obesity, the obesity level of the males is still higher than that of the females. In projections for both years, the results indicate that there is a higher level of obesity in the east of the city for both sexes. The analyses indicate that in Scenario 1, males in this group improve more than females in future (Figure 8.8). This may indicate that males in this group are more likely to take individual responsibility towards obesity prevention than females.

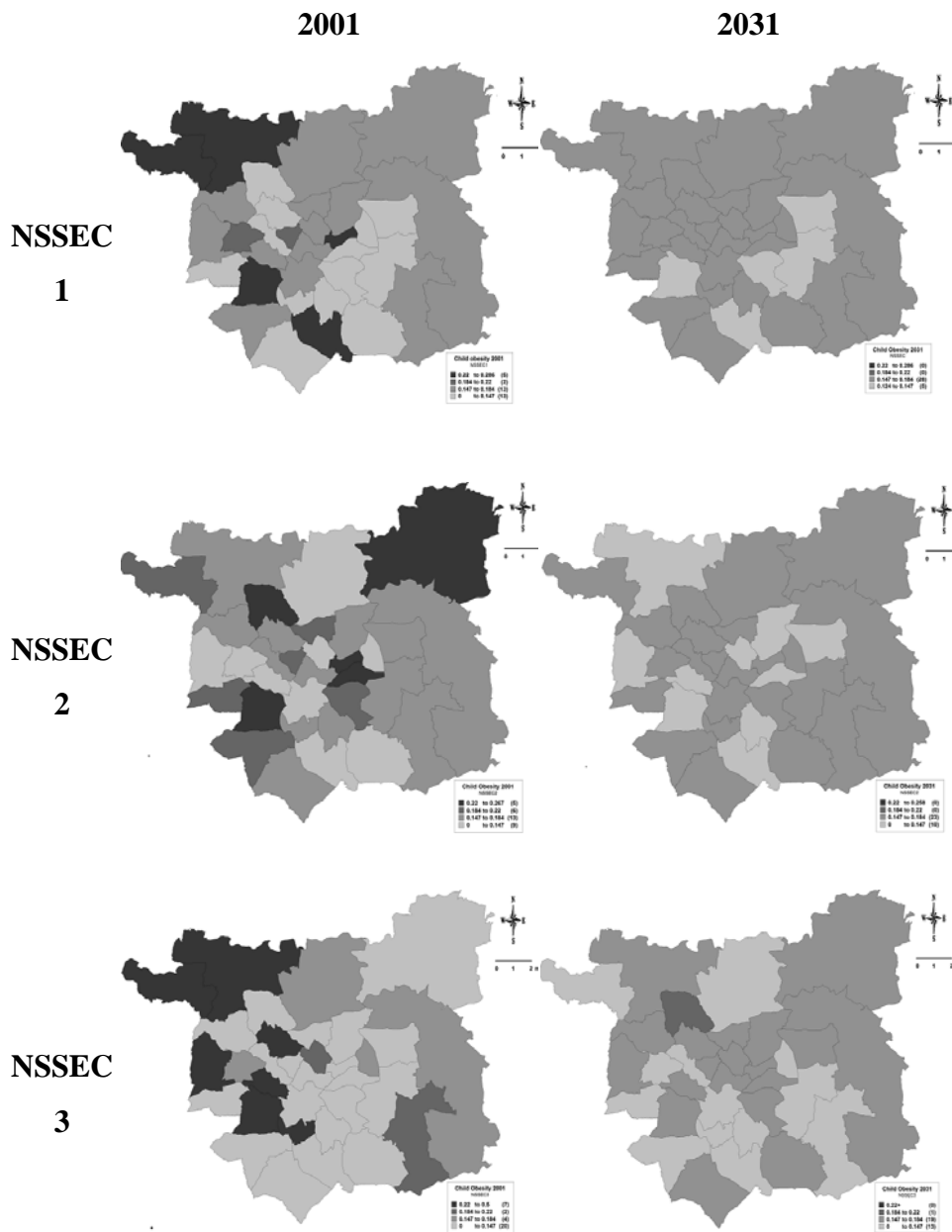


**Figure 8.8 Obesity by sex in 2001 and 2031 (NSSEC 1)**

Finally the impact of socioeconomic status on obesity is analysed. Child obesity has an important impact on the future patterns of the obesity changes in area and it has a different pattern than the adult obesity. Due to its vital role in the obesity prevention and control, child obesity projections are analysed by NSSEC groups 1, 2 and 3 under the assumptions in Scenario 1. As it can be seen from Figure 8.9, the obesity patterns of children from the three groups are quite different. The shades from light to dark indicate the ranges of obesity rates in an area: under 0.147, 0.147 to under 0.184, 0.184 to under 0.22 and 0.22+.

In 2001 projections, the higher level of obesity are mainly found in the north and the east of the city for children from NSSEC group 1 with some areas scattered around city centre area; higher level of obesity for children from NSSEC group 2 are found mostly around the city centre area and northeast of the city; while higher level of obesity for children from NSSEC group 3 are found in the northwest of the city and city centre areas.





**Figure 8.9 Child obesity by socio-economic status in 2001 and 2031**

After 30 years' simulation, child obesity in all NSSEC groups has improved in 2031 and the most significant change is that the east side of the city has replaced the west with higher child obesity levels. Compared to the obesity levels of the other two NSSEC groups, children from NSSEC1 improved the least (Figure 8.9). The improvement is due to the assumption in Scenario 1 that parents will protect their children from future obesity. However, analyses show that the individual anticipation approach work better for sub-populations in NSSEC groups 2 and 3 than for NSSEC1. This may be due to

the easy access to unhealthy food and limited physical activities for children from NSSEC1 group. There is a consistent high level of child obesity around the city centre area for all NSSEC groups, in both year 2001 and all scenario projections in 2031.

## 8.5 Conclusion

Using the dynamic spatial MicroSimulation Model (MSM), both the individual and local characteristics can be captured and impacts from previous simulation steps can be built into the next step. The holistic modelling approach produces a better reflection of the studied populations and allows us to understand the population changes. Therefore it provides a better basis for various strategic planning and policy making. As obesity has become a serious and fast growing issue in public health in UK, an application of the MSM to model obesity changes in Leeds has been developed to demonstrate one potential application of the MSM. The projection of the obesity changes in small area populations within three scenarios demonstrates that the MSM can provide a new approach to promote the well-being of people, through studies about who they are, where they live, and how they live within a spatial framework.

Based on the BMI approach, a function of three risk factors (age, sex and socioeconomic status) has been used to simulate individual transitions of obesity. There are three main reasons to use socioeconomic status instead of the variable of location as has been used in other important demographic transitions. The first is due to the lack of small area obesity prevalence data, the second is because of the important impact of the socioeconomic status on obesity and finally socioeconomic information is available in both population data and health data. Three scenarios have been developed under different assumptions where the tackling the obesity problem relies on Individual Anticipation (Scenario 1), Social Anticipation (Scenario 2) and Individual Reaction (Scenario 3). A series of analysis have been conducted the model results have been discussed at the aggregate spatial level of Leeds and at disaggregate spatial level of the small areas. Various patterns of

obesity changes of sub-populations by different demographic characteristics such as age, sex and socioeconomic status have also been analysed and discussed.

The socioeconomic changes are not simulated. Firstly this is because it is a huge task. It will require measuring transitions by education, occupation and income. Then these transitions need to be related to the transitions to an economic model of the Leeds/UK economy, which are not always reliable. Secondly the current socioeconomic statuses are dependent on the occupation information, which excludes those who do not have an occupation and the children. Thirdly the possible impact of keeping socioeconomic statuses the same would be: if the population gets poorer, obesity may rise because of the association with poor diets, unless extreme poverty reduces food expenditure so much that people lose weight. However, the spatial structure of the deprivation is rather stable over time. It is unlikely that spatial structure of obesity will change because of socioeconomic changes.

The analyses at the LA level of the whole Leeds indicate that at the beginning of the simulation, there are some hotspots of obesity in the east side of the city in 2001. After 30 years' simulation, there has been significant increase in obesity level in the city and the east remains the worse-off side of the city. Compared to the north, the south of the city seems to have higher obesity levels. The city centre maintains the lowest level of obesity consistently in all scenarios. At the small area level, projection results in Cookridge and Headingley have been selected and analysed to explore the spatial variance in wards. After 30 years' simulation, the obesity patterns in both wards have changed and the projections indicate a trend of higher obesity levels in older population in both wards. However, the results suggest that the obesity trend changes have a more significant impact on the population in Headingley than in Cookridge, especially for the female population. The projection results of sub-populations by age allow comparison of the child obesity against the adult obesity patterns. Distinctive patterns have been found among the two age groups. Similarly

the analyses of sub-populations by sex and socioeconomic status have also revealed the impact of such factors on the patterns of obesity changes. Such disaggregate analyses provide insights into the understanding of the obesity patterns on a more aggregate level, as well as providing useful information to assist other further explorations and health planning.

Although overall at both the aggregate and disaggregate spatial scales there is a consistent indication that obesity in Leeds is like to increase in future, spatial variations have been found in small area projections. Such small area information is important to understand the population trends found at a more aggregate level. The micro approach is novel in obesity studies as it allows different disaggregation and provides rich details of the studied populations. By also considering the relationships of the risk factors at the micro-level using small area analyses, the model results can pin point to micro-areas that stand out from the normal obesity patterns. Therefore in such areas, different interventions or health policies may be more appropriate to have the most effect on improving obesity trends in local populations. Thus it is important that these geographical differences are understood in order that they can be acted upon. This provides a more informative position for a public health planner to work from.

In the three obesity scenarios, the impact of different social and individual approaches to tackle the obesity problems can be seen clearly. Despite the trend of the increasing prevalence of obesity in future, the Social Anticipation approach (Scenario 2) has made an impact on the population level and produces a more even distribution of the obesity levels in the city of Leeds than the Individual Anticipation (Scenario 1) and Individual Reaction (Scenario 3) approaches. Such scenarios demonstrate the potential application of this model in exploring the alternative options and providing useful information to assist public health planning in future under the “what if” situations.

When compared to results from other research, it is encouraging to note that some findings are similar. In our child obesity projections, future hotspots have been consistently recognised in all scenarios in city centre area and

isolated cases in the north of the city (Figure 8.7 and 8.9). Such child obesity projections are consistent with the results from Edwards and Clarke (2009). Identifying obesogenic environment in Leeds, the authors have pointed out that there are higher rates in the central areas of Leeds, with some isolated cases in the north of Leeds. The projected child obesity hotspots in the north are also consistent with their identification of areas with high levels of poor diet and low activity amongst children in more affluent and rural parts of Leeds. Although the authors use a determinist model with a focus on the child obesity and use specialised obesity datasets in their model, such consistent patterns have been found in the projections produced by this dynamic spatial MSM.

As the dynamic spatial MSM is a population based model, it can be used for various strategic planning where the population changes are important. Public health is one of the potential areas. There are three main themes in the medical geography: disease ecology, health care delivery and environment and health. The obesity application model of the dynamic spatial MSM demonstrates the potential applications in all three areas: it can be used to study how a disease distributed spatially over time, identify the hotspots and design the health care provision accordingly. It can also be used explore the relationship between the people's health and where they live. It is possible to look into the individual details, using the annual output of individual records from the MSM. For exploration or planning purposes, it is also possible to trace various demographic changes of certain cohorts, groups or even individuals over the period using the unique ID of individuals in the model, although it is not demonstrated here due to the computing and time constraints. Finally, as the MSM projects the studied populations into the future from 2001 to 2031, the projection results are particularly useful in providing the groundwork for various explorations or facilitating medium and longer term strategic planning.

For the obesity application model alone, like the Foresight report (2009) has pointed out, there are many other possibilities of the further use of the model, including the application decision making areas such as obesity

related social inclusion and reduction of health inequalities. Further explorations using the model can also provide useful information to increase obesity related workforce productivity, as well as to support the promotion of general public well-being and healthy ageing. There is also potential for this model to contribute to the reform of child related policies or even wider health policies.

## Chapter 9

# Conclusions

## 9.1 Introduction

This study aims to facilitate the understanding of a complex social system through the development of a population based model. To achieve the aim and objectives of the study, the work described within this thesis adopts a novel modelling approach that combines the strength of two individual based modelling approaches: dynamic spatial Micro-Simulation Model (MSM) and Agent Based Model (ABM). To our knowledge there are no other published examples of similar models being applied to the modelling of the dynamic evolution of populations through heterogeneous demographic changes, as well as individual movements, interactions and behaviours at ward scale.

Using the hybrid approach, this work enables the modelling of a complex social system that is both theoretically and practically challenging. Essentially the model needs to be able to efficiently simulate the discrete changes of a large number of individuals within small areas that

demonstrate heterogeneous characters and behaviours of individuals. Such behaviours can arise not only from the individual demographic characteristics, but also from the interactions with each other and their local environment. To capture a fuller picture of such population evolution would be a challenge to existing models.

This chapter concludes the thesis. Section 9.2 explains how the overall research aim has been achieved by summarising the findings of the research with reference to the objectives outlined in Chapter 1 and highlights the main discoveries made. Following the discussion of the limitations of the research in Section 9.3, possible future work is proposed in Section 9.4. The chapter ends by drawing attention to areas where the applications of this hybrid model may prove beneficial potentially in the future and by providing concluding statements regarding the key findings.

## 9.2 Summary of main findings

### *9.2.1 Research objectives*

This research aims to study a complex social system to facilitate understanding through the development of a population based hybrid model that combines the strength of both MSM and ABM techniques. The thesis has sought to demonstrate the importance of individual based modelling and simulation tools within the scope of demographic planning, as well as a variety of substantive research and planning environments. To achieve this, more specific objectives of this study have been established in Chapter 1:

1. to review and discuss the microscopic approach in social modelling;
2. to develop an innovative modelling framework that enables the study of population evolutions through individual changes over the period of 2001-2031;
3. to provide a complete representation of the studied population at a fine spatial scale;



4. to produce rich, detailed and robust projections of the future population and

5. to investigate scenarios for demographic related public planning.

In terms of the demographic planning, this project aims to use the power of computational simulation to get beyond traditional macroscopic approaches to demographic analysis and forecasting in terms of:

6. Spatial disaggregation of population projections at a small area scale;

7. Rich representation of heterogeneous characteristics in individuals and their behaviours and

8. Scenario-based analysis of population changes.

Those objectives are achieved in the thesis. Table 9.1 summarises the chapters where the specific objectives (O1-8) are achieved.

**Table 9.1 Objectives achieved in chapters**

	Objective 1	O2	O3	O4	O5	O6	O7	O8
Chapter2	✓							
Chapter3		✓						
Chapter4			✓	✓		✓	✓	✓
Chapter5			✓	✓		✓		
Chapter6		✓	✓			✓	✓	✓
Chapter7			✓	✓				
Chapter8					✓			✓

### *9.2.2 Main findings against objectives by chapters*

This section will describe how this work achieved the objectives are achieved through the findings from specific thesis chapters in detail.

### 9.2.2.1 Chapter 2

To achieve objective 1, a review on the available literature and models has been conducted in Chapter 2. The literature review suggests that as an alternative to traditional macroscopic models, microscopic models provide powerful ways to model social systems by providing rich information at the individual level. MSM is now widely used, especially in relation to public policy research. A spatial MSM takes local contexts into account and therefore has many advantages in exploring spatial relationships and analysing the spatial implications. Compared to a static MSM, a dynamic MSM can produce a better representation of the modelled population. The dynamic ageing technique enables the updating of the demographic structure by applying transition probabilities to individuals on the basis of their individual and spatial characteristics. With a dynamic MSM, it is possible to capture the impact of structures from the previous year for the next, as well as assessing the impact of population changes over a longer term.

On the other hand, to understand the panorama of population changes, changes in individual movements, interactions and behaviours need to be captured within the model. However, to effectively model such changes associated with a large number of individuals with a rich set of attributes is both challenging in theory and practice using a pure MSM. ABM can complement it by providing the capacity for modelling individual behaviours, especially when such behaviours are not necessarily mathematically tractable, or where we have insufficient knowledge/data about the behaviours. However, ABM as a new modelling approach is less strong in its empirical foundations in social modelling, especially in terms of the validation/calibration of the model results. Another practical issue is that the system performance of the rule-driven simulation is less impressive. Without the ABM, a system run from 2001-2031 completes in 8 hours. However, with the ABM applied to student population in migration process only, it will take about 15hours. So the ABM model results in about 90% more running time (on a pc with 3GHz CPU and 8GB RAM).

A hybrid modelling approach that incorporates both MSM and ABM techniques can achieve better results than using only one of the approaches. The widely applied and tested MSM approach complements the ABM by providing the efficient list-processing power and empirical basis of using real data with the ambition to facilitate decision making. ABM focuses on the dynamic processes and the heterogeneous movements, interactions and behaviours at the individual level. Chapter 2 reviews and discusses the microscopic approach in social modelling and provides the groundwork to achieve the aim and objectives of this research.

### 9.2.2.2 *Chapter 3*

To achieve objective 2, Chapter 3 described the methodology of this study, including system design, system development and validation methods. The modulated system provides the flexibility for adding or subtracting functions and modules as required. The object oriented development method complements the individual based model, as individuals can be modeled naturally in the form of objects. The hybrid modelling approach has also been described in terms how the system design corresponds to the system development. The individual based model of MSM within a spatial framework enables the easy integration of an ABM, as agents in an ABM originated as autonomous individuals on a lattice of cells. The flexibility of rule implementation in an ABM enables agents to observe both external and internal rules. As explained in researches on artificial societies, agents obey social conventions, as well as their individual rules. Therefore, when individual agents switch from general external rules to individual internal rules, any general system setting or other individuals are not affected. Thus the hybrid model can capture both the temporal and spatial aspects of a complex system whilst charting the effects of an individual's behaviour at different scales of analysis.

MSMs are known to be difficult to validate, due to the rich details captured in the model. A framework is designed to attempt the validation of the hybrid model in the form of a series of alignment exercises, as they are now generally practised to benchmark MSM results against the aggregates such

as the official population projections. The application of such methods enables the exploration of the population evolutions and provides a useful platform to build up further application models or develop various scenarios with different public planning focuses, therefore allows the model to achieve the objectives of the representation, projection, exploration and planning of the studied populations.

#### 9.2.2.3 Chapter 4

Chapter 4 outlines the Model Development through detailed descriptions and illustrations in various diagrams and charts. As pointed out previously, this model is built on the base of a dynamic spatial MSM, with added features of ABM. The MSM is used for modelling the population and its dynamics, with the model structure paralleling the macro multistate cohort-component projection model. To achieve the objective 3 “to provide a complete representation of the Leeds population at a fine spatial scale”, individuals of the studied population are projected at the ward level to enable the study of the population evolution within a local context. Each individual record in the base population (provided by colleagues, see Birkin *et al.*, 2006) comes with more than 60 various attributes, including demographic, social-economic and spatial information and thus enables the simulation of the individual changes with great details.

In order “to produce rich, detailed and robust forecasts of the future population of Leeds” (objective 4), six demographic processes, Ageing, Mortality, Marriage, Fertility, Health Change and Migration, are modelled in this dynamic spatial MSM. Each of the processes has been developed as a separate module to provide structural flexibility, but they can interact with each other as observed in reality. Details of the modelling of such processes are illustrated in flow charts. Using the Monte-Carlo simulation method, transitional probabilities based on Age, Sex and Location are applied to individuals dynamically simulate the discrete demographic transitions. Additional variables are introduced according to the specific requirements of each demographic process. Thus the dynamic spatial MSM captures the changes in individual attributes at each step of the simulation for each

individual. Running the MSM for several decades enables us to assess the impact of population changes.

In terms of demographic planning, the population within Leeds is modelled at the individual level within a set of small areas called wards. Thus individual characteristics are simulated at the ward level to enable the achievement of the “spatial disaggregation of projections” (objective 6) at various spatial scales. Also such a structure of simulating individuals within areas allows the ABM features to be added on with ease. The ABM features have been developed under the MSM framework where agents are constructed to represent individuals and they act autonomously according to their built-in rules. Agents can interact with each other and their environment (the features of the ward that they live in). Like individuals who follow the social conventions within an artificial society, agents in the hybrid model follow the general simulation procedures in MSM as the common external rules, unless the MSM procedures conflict with their internal individual rules. When this happens, the prioritised rules (in this model, their internal rules) overwrite the external rules and the agents follow their individual rules. As agents can react to the changes of their fellow agents or of the environment by following rules in the form of “If A happens, then do B”. Great flexibility in modelling the complex process is enabled without running into problems of mathematical tractability. Thus the hybrid model overcomes various practical issues raised when using a pure MSM associated with poorly measured transitions such as student migration and provides “rich representation of heterogeneous characteristics in individuals and their behaviours” (objective 7). The ease of exploring different scenarios by updating relevant probabilities used in the dynamic ageing procedures and the usage of different individual rules for the agents both enable the achievement of objective 4: “To investigate scenarios in demographic related public planning”. The hybrid model framework also provides the flexibility for developing various “scenario-based analyses of population changes” through explorations of changes in sub-populations or different demographic processes by applying different rules or probabilities to groups of individuals with ease (objective 8).

#### 9.2.2.4 Chapter 5 and 6

The initial results produced by the MSM have been analysed in Chapter 4. The analyses have indicated some limitations of the pure MSM. The main problem comes from the most complicated process of migration. To provide “rich representation of heterogeneous characteristics in individuals and their behaviours” (objective 6) and to understand the population evolution patterns through individual heterogeneous changes within rich, detailed projections (objective 3 and 4), the model needs a series of refinements.

Chapter 5 describes the further developments of this process by extending the migration framework to include migration not only within Leeds, but also from and to the rest of UK and the rest of the world. Chapter 5 also describes the updating of fertility and mortality probabilities to reflect recent demographic trends. However, the student migration issue is poorly modelled by the MSM.

Chapter 6 introduces two ABM simulations to address this shortcoming of the MSM in order to capture the distinctive patterns of movement, interaction and behaviours of various sub-populations to achieve objectives 3, 4 and 6. The student migration simulation tries to model and understand the distinctive pattern of student migration in Leeds, which is understudied and lacks suitable and accurate data. There is an annual replenishment of the university student population in wards that offer popular student accommodation. To capture this process, students are modelled as agents that follow their built-in goal of moving to the areas where their fellow students stay, subject to the availability of housing vacancies. Depending on their higher education program, students stay in Leeds for a certain period of time before they leave the city. The typical interaction of a “student agent” with another would be “finding my fellow-students in the area they stay” and the typical interaction with the “environment” will be “checking if the targeted area has a vacancy to move in”. Thus the heterogeneous migration behaviours of student migrants can be captured within the projections and impact of annual replenishment of young population in small area population projections is captured, whereas the pure MSM failed to

reproduce this behaviour. The hybrid model results also suggest clear concentrations of the student populations around the city centre that are close to the university campuses in Leeds. In contrast, the MSM results suggest an even distribution of the student population throughout the whole city, even in the suburban areas.

Another simulation using ABM demonstrates the impact of personal migration history on demographic changes. The mortality process is selected for a simple demonstration using the hypothesis that the current environment has a great impact on the mortality of local populations, but the mortality risk of an individual should not suddenly change just because he or she moved into a different location. ABM enables the assessment of individuals' behaviours using information from their personal histories, as agents can store and retrieve their personal history easily. Three scenarios of mortality projection using the hybrid model have been developed: mortality based on current destination location, original residence location and personal migration history. The resulting analyses indicate the impact of different personal histories on the mortality risks faced by the individuals. This allows us to assess the impact of personal history under different scenarios of personal migration history circumstances and explore the relationship between the mortality rates and different environments. Such changes in mortality risks for both migrants and non-migrants will result in significant changes in local population structure and can be interesting for various applications. This also helps to achieve objective 8, the "scenario-based analysis of population changes" and objective 7, the "rich representation of heterogeneous characteristics in individuals and their behaviours".

In the examples described above, the most straightforward processes are deliberately chosen and ABM techniques are applied with considerable simplification. However, with some modifications, this model can be applied to any other demographic process or sub-population to achieve particular goals.

#### 9.2.2.5 *Chapter 7*

As dynamic MSMs normally project with rich individual details over a long period of time to assess the longer term impact, there are often limited suitable microdata that can be used for comparisons. It is generally agreed that the summed results of dynamic spatial MSMs often need to be aligned to external aggregate data. Based on the discussions of the need of the alignment and various previous alignment exercises, a framework for the result alignment has been developed “to produce rich, detailed and robust projections of the future population” (objective 3) and to provide the groundwork “to investigate scenarios of demographic related public planning” (objective 4). Chapter 7 describes the details of an alignment exercise where the model results are compared with the official projections from the Government (ONS). The official aggregate model is called Model A and its projections are used to compare with two disaggregate models, Models B and C.

Under the framework, alignment exercises have been conducted. The result analyses indicate that the results from Model B are closer to Model A, as the same set of probabilities are naively applied to every ward. Results from Model C, though, demonstrate the spatial variations and resulted in a larger population in the projections for year 2031. There are also some variations in individual components of change. However, the main population trends and patterns are consistent in Models B and C. This demonstrates the robustness of the model results. The main reason behind the differences is that Model C uses a full disaggregation and different probabilities are applied to each ward to reflect the spatial variations in small areas. Also the dynamic simulation using the Monte Carlo method and the ABM can introduce a certain degree of randomness as well. In future, further development can be carried out to reduce such randomness.

#### 9.2.2.6 *Chapter 8*

Chapter 8 demonstrates the potential of the hybrid model through an application. Obesity has become a serious fast growing issue in public



health planning. “To investigate scenarios of demographic related public planning” (objective 5), obesity changes in small areas have been projected in three scenarios that use the Foresight (2009) scenario assumptions of adopting Individual Anticipation (Scenario 1), Social Anticipation (Scenario 2) and Individual Reaction (Scenario 3). Spatial variations have been found in small areas in the projections. Such information is important to understand the spatial variation of population trends at an aggregate level and to facilitate more targeted public planning. The analysis of scenario results indicates that there will be an increase of obesity levels in future under all scenarios. However, Scenario 2 made an impact at the population level and changed the obesity distribution patterns in small areas. The gaps between the obesity levels in small areas are very small and the inequity in terms of obesity risks was reduced for small areas. As both Scenarios 1 and 3 rely on individual responsibility, they did not make an impact at the population level and the differences between wards in year 2001 still persist in 2031. This application model and analysis in Chapter 8 therefore also achieves objective 8, the “scenario-based analysis of population changes”.

As the dynamic spatial MSM is a population based model, it can be used for various strategic planning where the population changes are important. Public health is one of the potential areas, as the MSM can provide a new approach to promote public well being, through studies about who they are, where they live, and how they live within a spatial framework. The obesity application model demonstrates the potential of this model in such areas as disease ecology, health care delivery and environment and health: it can be used to study how a disease is distributed spatially over time; it can identify the hotspots; and it can help in the design of the health care provision. It can also be used explore the relationship between the people’s health and where they live. Finally as the MSM projects the studied populations into the future from 2001 to 2031, the projection results are particularly useful in providing the groundwork for various explorations or facilitating medium and longer-term strategic planning.

## 9.3 Limitations

This work has developed a hybrid IBM. The model attempts to track the evolution patterns of the local population through changes to each and every individual, which is both theoretically and practically challenging. The hybrid approach that brings together the strength of both MSM and ABM has proved to be a successful modelling strategy. As explained in Section 9.2, this research has achieved its main aim and specific objectives. However, it is important to recognise the limitations of the research. This section produces a self-critique of the research and section 9.4 suggests areas for potential future work.

One of the main limitations of the current model is its confinement to Leeds alone. This limits its usage. To extend its application, it would be better to apply the model at a bigger scale, i.e to the regional, sub-national or even national level. This would enable the model to demonstrate more obvious spatial differences, for instance, at the regional level. Such a model can also provide better groundwork to facilitate research to address a wider range of issues, as well as assess the impacts on bigger areas. As the current model structure is generalised for any ward based population, it can be extended to a bigger scale easily in future, given more computing resource and time.

The model results indicate Leeds immigration will continue to increase throughout the simulation period and as a result, it will have a population of just short of one million in year 2031. However, other research (Boden and Rees, 2010) has pointed out that the 2008 based ONS projections used sub-national immigration estimates which were flawed. One of the major issues is that the ONS projections have overestimated the international migration into Leeds. As the migration changes the dynamics of the population, the ONS projected a much bigger population in 2031 than in research led by experts using alternative estimates of subnational migrations (Rees *et al.*, 2010). In this study, the model results have been aligned to the official projections, but the next step may be to align our model with other academic models that capture local characteristics better. The comparison of the

results to both ONS and academic models will help to improve the robustness of this model.

Another limitation may be the degree of sophistication of the ABM within the hybrid model. The current ABM is simplified due to two reasons: limited understanding of the individual behaviours and limited availability of suitable microdata. Although the hybrid model provides the example to demonstrate the usefulness of the ABM to model the impact of individual behaviours and their personal histories, the examples have been limited to rather straightforward processes where individual agents operate with limited intelligence. However, as many studies of artificial societies indicate, ABM is capable of much more sophistication. If the knowledge and data gaps can be bridged, then sophisticated decision making rules for the agents can be developed to assist their choices and actions to explore complex research questions with more sophistication.

The application of the MSM-ABM to the projection of obesity levels is also very simple. Lacking the small area prevalence data, the probability of obesity has to be calculated on the basis of age, sex and socio-economic status only. As a result the application only reflects the demographic and socio-economic characteristics of the small area populations and the impact from the environments of small areas themselves is currently not modelled. Given suitable data, this model could be strengthened to capture more spatial variation in obesity prevalence. If such data were available, it would benefit the hybrid model to model the impact individual behaviours or/and personal history through development of different individual rules using ABM. There has been a lot of research on the “obesogenic environments” that reveals the environment that foster the obesity problem (Edwards 2009, 2010). In the Foresight Programme (2009), 29 drivers of change under the categories of Health, Science and technology, Leadership, Education and information, Shape of society, Values and behaviour, Resource constraints, Economy, Food production and retailing and Living environment have been considered to be important for modelling obesity. Due to the time and resource constraints, it was necessary to focus on the key obesogenic

factors. Given more time and research, this application model could be improved into a better version that reflect the “obesogenic environments” approach very well, as ABM provides the capacity to capture interactions between individuals and the environment that they live in.

Another data related limitation is that the model is simulated at the ward level. This may be criticised as one potential weakness for not providing disaggregation at the OA level. However, due to the Small Cell Adjustment Method (SCAM) used in UK Census 2001 data, population microdata at the OA level can be obscured. Also the OA level results may not be as effective as the ward level results in terms of facilitating policy making, due to the sizes of areas at this is scale. Switching to Middle layer Super Output Area (MSOA) seems to be a better option and in future work, the model can be restructured to be based at the level of MSOA.

Other specific limitations include:

1. No housing stock constraint in migration process
2. Modelling of flows into Communal establishments treats people living alone or with others the same
3. Unresolved modelling issues such as the exclusion of HE students from marriage and fertility
4. ABM arguably confines students too closely to existing ‘student areas’, and treats all students as studying away from home rather than allowing for students who ‘study at home’
5. Obesity made conditional upon SES, but SES status is not modelled
6. MSM model is Monte Carlo based, impact of this run-to-run random variability need to be addressed
7. System performance needs improvement
8. Emigrants and immigrants assumed to look like existing ward residents

These limitations will need to be addressed in future work on this model.

## 9.4 Future work

Section 9.3 has discussed the main limitations of the model. Given more computing resource and time, the model can be extended to the scale of the whole UK. This will better demonstrate the strength of the dynamic spatial MSM. The spatial variation between the regions or sub-nations will become more obvious and different dynamics within local populations will lead to different patterns in the population evolutions. A model of the whole UK population would be more useful in providing rich information about population changes at more different levels of aggregation. It therefore could provide an even better groundwork to facilitate various researches or decision making on different scales and would be especially useful for research questions such as the impact of recent European immigration at the regional or the national level. The current model is completely generalised for any area of UK. However, the computing power required for both the main simulation and the probability matrix process will increase exponentially. For instance, in the most straightforward demographic process of Mortality, the current Leeds model already needs to process a matrix that contains 6,666 ( $2 \times 101 \times 33$ ) mortality probabilities (by sex, single year of age and ward: Standard Table wards). Extending the model to the UK scale will need to process 2,132,716 ( $2 \times 101 \times 10558$ ) mortality probabilities each time when each of just under 59 million individuals is simulated in the mortality process. Not to mention that the base population is increasing each year and the mortality is the simplest demographic process. More complicated demographic process requires multiple stage simulation - hence multiple probability matrices and demands for more computing power to process the large matrices. However, given sufficient resource and time, it is possible to achieve the ambition to build such a large scale model with comprehensive individual details.

Another possible improvement may be to build more sophistication into the ABM. The current ABM is deliberately simplified due to two reasons: lack of understanding of the individual behaviours and availability of suitable

data. Given the opportunity to solve the two issues, the ABM can enrich the behaviour modelling of the individuals with much more sophistication. For example, in the marriage process, the ABM can potentially benefit the modelling of this complex process in many areas. The preference of a marriage candidate for an individual can be very different according to their individual characteristics such as age, sex-orientation, socio-economic status, education or cultural/religion. Given sufficient information on such preferences and complex rules can be developed in ABM to guide individuals to select their candidates. Similarly, using ABM to model the process from how an individual starts to search for a marriage candidate to finally forms a household could be of interest across a wide range of social science research.

The challenge of applying the understanding of obesity related behaviours of the individuals and appropriate microdata also limit the obesity application considerably. The probabilities of obesity have to be calculated on the basis of age, sex and socio-economic status, instead of age, sex and location due to the unavailability of prevalence data at the ward level. If there are spatially disaggregated data available, it will strengthen the modelling of spatial variation substantially. Another issue with the current application is due to the author's lack of knowledge and data on individual obesity-causing behaviours. Given more time and research, it would enable the hybrid model to model the impact of individual behaviours and personal history on obesity (smoking/diet/obesity: especially important when their parents were obese). The analysis could possibly be extended to consider the impact of social networks and broader social context on obesity such as discussed in Christakis and Fowler (2007) and Wilkinson and Pickett (2009). The performance of the model is, of course, limited by the information available at the time the model was built. It is also constrained by uncertainties about future changes in lifestyle and behaviour, as well as various assumptions about migration and demographic change. One way to handle this is to run the microsimulations a large number of times and to describe the probability distribution of outcomes through a median projection. Another way is to adapt the probabilistic methods used in

macromodels for use in microsimulations. Continuous effort into further refinements to the simulation outputs could potentially be achieved by an ensemble modelling approach using methods already introduced in the context of climate simulations (Murphy *et al.*, 2004).

Finally, specific limitations identified in last section will need to be addressed in future versions of the model.

## 9.5 Conclusion

This study has developed a hybrid model of the Leeds population, whose method is grounded in the form of a dynamic spatial MSM, but with an ABM providing insight to model individual movements, interactions and behaviours. It has demonstrated the importance of a hybrid spatial modelling and simulation tool within demographic related planning environments, as well as its potential in a variety of substantive research contexts in social science. As explained in Section 9.2, this research has achieved the main aim and specific objectives. The added ABM features enhance the capacity of the model in behaviour modelling and enable personal histories to be taken into consideration. This is particularly useful when there is a knowledge gap or when appropriate microdata are not available. The hybrid model can therefore present a comprehensive representation of the studied population and provide projections with rich information of individual changes at a fine spatial scale.

Such strategy seems to be successful as described in previous sections. The individual based approach enables the modelling and analyses of populations at different spatial aggregations. The main advantage of such analysis is that it enables us to focus on key problem areas, rather than relying on averages for the whole region or even nation. In contrast to the generalised results, the hybrid model highlights the spatial variations within small areas and thus enables more focused interventions to be designed and implemented.

The rich individual details provided by the hybrid model can facilitate the understanding of variations in the small areas and individual behaviours at the micro level. It also enables the consideration of the relationship between the populations themselves and between the population and the environment that they are living in. By focusing on the dynamics of the population, the hybrid model is better suited for capturing all the rich details within discrete demographic changes than traditional macroscopic demographic models. The question whether those dynamic changes are due to compositional changes or changing behaviour rules of the individual agents can also be studied within the framework of the hybrid model.

The MSM and ABM are used to complement each other to strengthen different aspects of the model. ABM is used to provide flexibility in modelling movements, interactions and behaviours where knowledge and data are limited. The MSM roots of using real data and use for real application can provide valuable guidance for the projections and the statistical mechanisms in MSM can ensure the similarity between what it predicts and what is actually observed in the gathered data. The model also benefits from the power of list processing from MSM to enable a large scale model. Ageing, fertility and mortality can all be simulated easily by this means. However once the movement and interaction are introduced, the flexibility of the MSM is restricted by its statistical nature, when lacking appropriate data. All MSMs rely on appropriate microdata to produce robust projections and they will struggle if such data are not available, as described in the student migration application. To address such issues and enable flexibility of the modelling of movement, interaction and behaviour, a hybrid modelling approach has been adopted to bring the ABM insight to strengthen such areas of the model. As previously discussed e.g. Billari *et al.*, 2002 and by adopting the hybrid approach, we have achieved the goal of both the effective handling of large scale individual based system, as well as providing extra flexibility to model various movements, interactions and behaviours of sub-populations in different scenarios, especially when there is a limitation on the microdata.



As a demographic model, this model can monitor the evolution of population structures and various demographic change patterns on a fine geographical scale. This provides vital information for demographic planning/policy making at different aggregations. The rich representation of heterogeneous characteristics and behaviours by the hybrid model provides a good basis for demographic planning. The capability of incorporating “what if” scenarios enables the investigations of population changes under different circumstances, as well as explorations of various hypotheses. Similarly, this model can also benefit other public policy making or public service planning. For instance, the ageing trends in certain suburban areas may promote changes in health service and public transport service provision in order to enable easy access to such services for the ageing population in the area. The rich attributes captured in the system are also very useful in various policy analyses or research purposes.

On the other hand, this hybrid model has provided a framework to enable the effective modelling of individual decision making units on a large scale, as well as adding the flexibility to introduce different modelling techniques to strengthen various aspects of the model. Such an innovative model is applicable to more general social science studies, including facilitating the study and understanding of a complex social system, the testing of various behaviour scenarios and other research interests. As discussed in Chapter 2, the hybrid modelling framework provides a linkage between the macro and micro approaches, as agents within the ABM can observe both internal and external rules, which can include both macro and micro factors. Due to its flexibility of modelling, the hybrid approach can also provide a bridge between different disciplines and a useful multidisciplinary tool when the mathematics is intractable (Axelrod, 2005; Conte *et al.*, 1998).

Although the main aim and objectives have been achieved, some limitations of the model have also been discussed. Based on the discussion, some future work has been proposed, including the extension of model to the whole UK scale and development of more sophisticated ABMs, as well as continuous effort in refining the model with more available information. It is believed

that continuous efforts to improve various aspects of this model using the hybrid approach can provide a better framework to study the population evolution through individual changes. This innovative model can thus provide a better groundwork for various research purposes, as well as facilitating population related decision making.

# References

Aaberge, R., Holmoy, E. and Colombino, U. (2003) An integrated micro-macro general equilibrium framework for analysing tax reforms. Proceeding of the International Microsimulation Conference on Population, Ageing and Health, Canberra, Australia.

Allison, D., Fontaine, K., Manson, J., Stevens, J. and VanItallie, T. (1999) Annual deaths attributable to obesity in the United States, *Journal of American Medical Association*, 282, 16, 1530-1538.

Anderson, J. (2001) Models for Retirement Policy Analysis, Report to the Society of Actuaries, available from: <http://www.soa.org/research/pension/research-models-for-retirement-policyanalysis.aspx>, accessed 12/03/2011.

Andreyeva, T., Sturm, R. and Ringel, J. (2004) Moderate and severe obesity have large differences in health care costs, *Obesity Research*, 12,12, 1936-1943.

Armitage, B. (1997) Variation in fertility between different types of local area, *Population Trends*, 87, 20-8.

Axelrod, R. (2005) Agent-based modelling as a bridge between disciplines, in Judd, K. and Tesfatsion, L. (eds.) *Handbook of Computational Economics*, 2, Agent-based computational economics, Handbooks in Economics Series, North-Holland, Amsterdam.

Ballas, D., Clarke, G. and Wiemers, E. (2005a) Building a dynamic spatial microsimulation model for Ireland. *Population, Space and Place*, 11, 157-172.

Ballas, D., Rossiter, D., Thomas, B., Clarke, G. and Dorling, D. (2005b) *Geography matters: simulating the local impacts of national social policies*, the Joseph Rowntree Foundation, UK, ISBN 1859352669.

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

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

Barceló J., Casas J., Ferrer J. and García D. (1999) Modelling advanced transport telematic applications with microscopic simulators: the case of AIMSUN2, in Brilon, W., Huber, F., Schreckenberg, M. and Wallentowitz,

H. (eds.) *Traffic and Mobility, Simulation, Economics, Environment*, Springer.

Baryla, E. A. and Dotterweich, D. (2001) Student migration: do significant factors vary by region? *Education Economics*, 9, 3, 269–280.

Batty, M. and Densham, P. (1996) Decision support, GIS, and urban planning. *Systema Terra*, 1, 72-76.

Beebout, H. and Bonina, P. (1973) TRIM: a microsimulation model for evaluating transfer income policy, Working Paper No. 971-4, The Urban Institute, Washington, DC.

Bækgaard, H. (2002) Micro-Macro Linkage and the Alignment of Transition Processes: Some Issues, Techniques and Examples, National Centre for Social and Economic Modelling (NATSEM), University of Canberra, available from: [https://guard.canberra.edu.au/natsem/index.php?mode=downloadandfile\\_id=567](https://guard.canberra.edu.au/natsem/index.php?mode=downloadandfile_id=567), accessed 11/01/2012.

Billari, F., Ongaro, F., and Prskawetz, A. (2002) Introduction: agent-based computational demography. in F. Billari and A. Prskawetz (eds.) *Agent-based computational demography: using simulation to improve our understanding of demographic behaviour*, 1–18, London/Heidelberg: Springer/Physica, ISBN:3790815500.

Birkin, M., Wu, B. and Rees, P. (2009) A dynamic spatial microsimulation model of local populations, 2<sup>nd</sup> General Conference of the International Microsimulation Association, June 8<sup>th</sup> to 10<sup>th</sup>, 2009, Ottawa, Canada, [http://www.microsimulation.org/IMA/Ottawa\\_2009.htm](http://www.microsimulation.org/IMA/Ottawa_2009.htm), accessed 18/06/2012 .

Birkin, M. and Wu, B. (2008) Dynamic social simulation models enabled by eresearch. In Proceedings of the Fourth International Conference on E-Social Science. Manchester: National Centre for E-Social Science, available from: <http://www.ncess.ac.uk/conference/programme>, accessed 06/09/2010.

Birkin, M., Turner, A. and Wu, B. (2006) A Synthetic Demographic Model of the UK Population: Methods, Progress and Problems, proceeding of The Second International Conference on e-Social Science, Manchester, UK, June, 2006, available from: <http://www.geog.leeds.ac.uk/people/a.turner/projects/MoSeS/documentation/articles/BirkinTurnerWu2006.pdf>, accessed 20/08/2012.

Birkin, M., Clarke, G. and Clarke, M. (1996), Urban and regional modelling at the microscale, in G. P. Clarke (ed.) *Microsimulation for urban and regional policy analysis*, 10-27, Pion, London.

Birkin, M. and Clarke, G. (1995) Using microsimulation methods to synthesize census data, in S. Openshaw (ed.) *Census users' handbook*, 363–

87. GeoInformation International, London. Boden, P. and Rees, P. (2010) New Migrant Databank: concept and development, Chapter 5 in Stillwell, J., Duke-Williams, O. and Dennett, A. (eds.) *Technologies for migration and commuting analysis*. IGI Global, Hersey, PA.

Boman, M. and Holm, E. (2004) Multi-agent systems, time geography, and microsimulations, in Olsson, M. and Sjöstedt, G. (eds.) *Systems approaches and their application*, 95-118, Kluwer Academic.

Boyle, P., Norman, P. and Rees, P. (2004) Changing places: do changes in the relative deprivation of areas influence limiting long-term illness and mortality among non-migrant people living in non-deprived households? *Social Science & Medicine*, 58, 2459-2471.

Bratman, M. (1987) *Intentions, plans, and practical reason*. Harvard University Press, Cambridge, MA.

Bray, H. (2008) 2006-based National Population Projections, *Population Trends*, 131, 8-18, available from: <http://www.ons.gov.uk/ons/rel/population-trends-rd/population-trends/no--131--spring-2008/2006-based-national-population-projections-for-the-uk-and-constituent-countries.pdf>, accessed 20/06/2012.

British Nutrition Foundation (1999) *Obesity: The report of the British Nutrition Foundation Task Force*. Oxford: Blackwell Science.

Britton, M. (1989) Mortality and geography, *Population Trends*, 56, 16-23.

Brooks, R. (1991) Intelligence without reason. *Proceedings of the twelfth International Joint Conference on Artificial Intelligence (IJCAI-91)*, 569-595, Sydney, Australia.

Brown, L. and Harding, A. (2002) Social modelling and public policy: application of microsimulation modelling in Australia, *Journal of Artificial Societies and Social Simulation* 5, 4, available from: <http://jasss.soc.surrey.ac.uk/5/4/6.html>, accessed 06/09/2010.

Caldwell, S. and Morrison, R. (2000) Validation of Longitudinal Microsimulation Models: Experience with CORSIM and DYNACAN, in Mitton *et al.* (eds.) *Microsimulation in the New Millennium*, Cambridge University Press: Cambridge, 200-225.

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

Caldwell, S. (1993) Corsim 2.0 model documentation, Cornell University, Ithaca, NY, mimeo.

- Caldwell, S. (1990) *Static, Dynamic and Mixed Microsimulation*, Department of Sociology, Cornell University, Ithaca, New York.
- Casti, J. (1997) *Would-be worlds: how simulation is changing the frontiers of science*, Wiley, New York, ISBN 0-471-12308-0.
- Centre for Census and Survey Research (2005) *2001 Household SAR (Licensed) Codebook*, Version 2.4. Manchester: CCSR, University of Manchester, available from: <http://www.ccsr.ac.uk/sars/2001/indiv/resources/ukind-v2codebook20060426.pdf>, accessed 08/05/2010.
- Centre for Interaction Data Estimation and Research (CIDER) (2001) *2001 Census Special Migration Statistics*, available from: <http://cider.census.ac.uk/>, accessed 11/01/2012.
- Centre for Census and Survey Research (2005) *2001 Household SAR (Licensed) Codebook*, CCSR, University of Manchester.
- Chambers Dictionary* (2013) available from: <http://www.chambers.co.uk/search.php>, accessed 12/02/2013.
- Champion T., Fotheringham S., Rees P., Bramley G. and others (2002) *Development of a Migration Model*. Office of the Deputy Prime Minister and the Stationery Office, London, ISBN/ISSN: 1851125833.
- Chaix, B. and Chauvin, P. (2003) Tobacco and alcohol consumption, sedentary lifestyle and overweightness in France: A multilevel analysis of individual and area-level determinants, *European Journal of Epidemiology*, 18, 531–538.
- Chipperfield, T., O'Brien, R., Bolderson, T., Eidinow, E., Shafner, I., Parry, V. (2007) Visualising the Future: Scenarios to 2050, in *Foresight Tackling Obesities: Future Choices* available from: <http://www.foresight.gov.uk>, accessed 26/06/12.
- Christakis, N. and Fowler, J. (2007) The Spread of Obesity in a Large Social Network Over 32 Years, *New England Journal of Medicine*, 357,4, 370-379.
- Citro, C. and Hanushek, E. (1991) Review and recommendations, in Citro, C. and Hanushek, E. (eds.) *Improving information for social policy decisions: The uses of microsimulation modelling*, 1, Washington, DC, National Academy Press.
- Clarke, G.P. (1996) Microsimulation: an introduction. G.P. Clarke (ed.) *Microsimulation for Urban and Regional Policy Analysis*, 1-9. Pion, London.

Clements, P., Papaioannou, T. and Edwards, J. (1997) Aglets: Enabling the Virtual Enterprise. Proceeding of the international conference of *Managing Enterprises - Stakeholders, Engineering, Logistics and Achievement* (ME-SELA '97).

Colella, V., Klopfer, E., & Resnick, M. (2001) *Adventures in Modelling: Exploring Complex, Dynamic Systems with StarLogo*. Teachers College Press, New York.

Collier, N., Howe, T., and North; M. (2003) Onward and upward: the transition to Repast 2.0. Proceedings of the First Annual North American Association for Computational Social and Organizational Science Conference, Pittsburgh, PA, USA.

Cohen, P. and Levesque, H. (1990) Intention is choice with commitment, *Artificial Intelligence* 42, 3, 213-261.

Conte, R., Gilbert, N. and Sichman, J. (1998) MAS and Social Simulation: A Suitable Commitment, in Sichman, Conte and Gilbert (eds.) *Multi-Agent Systems and Agent-Based Simulation*, Lecture Notes in Artificial Intelligence, Springer –Verlag, 1534, 1-9.

Conte, R. and Gilbert, N. (1995) Introduction, in Gilbert, N. and Conte, R. (eds.) *Artificial Societies: The Computer Simulation of Social Life*, Taylor & Francis, London. ISBN 1-85728-305-8, available from: <http://www.scribd.com/doc/16416522/GilbertConte-Artificial-Societies-the-Computer-Simulation-of-Social-Life>, accessed 26/06/12.

Curry, C. (1996) PENSIM: A dynamic simulation model of pensioners' income. *Government Economic Service Working Paper No. 129*. London: Analytical Services Division, Department of Social Security.

Dahlgren, G. and Whitehead, M. (1991) *Policies and strategies to promote social equity in health*, Institute of Futures Studies, available from: <http://www.framtidsstudier.se/wp-content/uploads/2011/01/20080109110739filmZ8UVQv2wQFShMRF6cuT.pdf>, accessed 26 June 2012.

Damper, R. (2000) Editorial for the Special Issue on Emergent Properties of Complex Systems: Emergence and levels of abstraction, *International Journal of Systems Science*, 31, 7, 811-818. DOI: 10.1080/002077200406543.

David, N., Marietto, M., Sichman, J. and Coelho, H. (2004) The structure and logic of interdisciplinary research in agent-based social simulation. *Online Journal of Artificial Societies and Social Simulation*, available from: <http://jasss.soc.surrey.ac.uk/7/3/4.html>, accessed 23/11/2011.

Davidsson, P. (2000) Multi agent based simulation: beyond social simulation. Moss, S. and Davidsson, P. (eds.) *Multi agent based simulation*, Lecture Notes of Computer Science, 1979, Springer Verlag, Berlin.

De Man, W. (1988) Establishing a geographic information system in relation to its use: a process of strategic choices, *International Journal of Geographical Information Systems*, 2, 3, 245-261.

Department of Health (DoH) (2004) *Choosing health: making healthy choices easier*. Department of Health White Paper, London.

Desai, A; Greenbaum, R and Kim, Y (2009) Incorporating Policy Criteria in Spatial Analysis, *Journal of the American Review of Public Administration*, 39 (1), 23-42, DOI: 10.1177/0275074007311387.

Dietz, W. and Bellizzi, M. (1999) Introduction: the use of body mass index to assess obesity in children, *The American Journal of Clinical Nutrition*, 70, 123-125.

Dijkstra, J., Jessurun, A. and Timmermans, H. (2000) A multi- agent cellular automata system for visualising simulated pedestrian activity. Bandini *et al.* (ed.) *Theoretical and practical issues on cellular automata*, proceedings of the fourth International Conference on Cellular Automata for Research and Industry, 29-36, Springer-Verlag, Berlin.

Dollman, J. and Pilgrim, A. (2005) Changes in body composition between 1997 and 2002 among South Australian children: Influences of socio-economic status and location of residence. *Australian and New Zealand Journal of Public Health*, 29, 166–170.

Dubuc, S. (2012) Immigration to the UK from high-fertility countries: intergenerational adaptation and fertility convergence, *Population and development review*, 38,2, 353–368.

Dubuc, S. and Haskey, J. (2010) Ethnicity and fertility in the UK, in J. Stilwell and M. van Ham (eds.), *Understanding Population Trends and Processes, Volume 3: Ethnicity and Integration*, 63–82.

Duley, C.; Rees, P. and Martin, C. (1988) A microsimulation model for updating households in small areas between censuses, *School of Geography Working Paper* 486, iv, University of Leeds, available from: <http://lib.leeds.ac.uk/search~S4?Xmicrosimulation&searchscope=4&SORT=D/Xmicrosimulation&searchscope=4&SORT=D&SUBKEY=microsimulation/1%2C33%2C33%2CB/frameset&FF=Xmicrosimulation&searchscope=4&SORT=D&12%2C12%2C>.

Edina (2011) Description of UKBORDERS, available from: <http://edina.ac.uk/ukborders/description/>, accessed 12/12/2011.



Edwards, K. (2010) Defining and mapping obesogenic environments for children in Lake, in Townshend, T. and Alvanides, S. (eds.) *Obesogenic environments: complexities, perceptions and objective measures*, Blackwell Publishing Ltd.

Edwards, K. and Clarke, G. (2009) The design and validation of a spatial microsimulation model of obesogenic environments for children in Leeds, UK: SimObesity, *Social Science and Medicine*, 69, 1127–1134.

Ellaway, A., Anderson, A. and Macintyre, S. (1997). Does area of residence affect body size and shape? *International Journal of Obesity*, 21, 304–308.

Espindola, A., Silveira, J., and Penna, T. (2006) A Harris–Todaro agent-based model to rural–urban migration. *Brazilian Journal of Physics*, 36, 3A, 603– 609.

Ettliger, M. and O’Hare, J. (1996) Revenue and incidence analysis of state and local tax systems: tools, technique and tradition, the National Tax Association's 89<sup>th</sup> Annual Conference on Taxation, USA.

Falkingham, J. and Hills, J. (1995) The effects of the welfare state over the life cycle. Falkingham, J. and Hills, J. (eds.) *The dynamic of welfare: the welfare state and the life cycle*, 83-107, Prentice Hall/Harvester Wheatsheaf, New York.

Favreault, M. and Smith, K. (2004) A primer on the dynamic simulation of income model (DYNASIM3), an Urban Institute Issue Focus, available from: [www.urban.org/uploadedpdf/410961\\_dynasim3primer.pdf](http://www.urban.org/uploadedpdf/410961_dynasim3primer.pdf), accessed 23/11/2011.

Fellendorf, M. and Vortisch, P. (2010) Microscopic Traffic Flow Simulator VISSIM, in Barceló, J. (ed.), *Fundamentals of Traffic Simulation*, International Series in Operations Research & Management Science 145, 63-93, DOI 10.1007/978-1-4419-6142-6\_2.

Finkelstein, E., Fiebelkorn, I. and Wang, G. (2003) National medical spending attributable to overweight and obesity: how much, and who’s paying, *Health Affairs*, w3.219-w3.226, DOI: 10.1377/hlthaff.w3.219.

Flegal, K., Graubard, B., Williamson, D. and Gail, M. (2005) Excess deaths associated with underweight, overweight, and obesity, *Journal of American Medical Association*, 293, 15, 1861-1867.

Fontaine, K., Redden, D., Wang, C., Westfall, A. and Allison, D. (2003) Years of life lost due to obesity, *Journal of American Medical Association*, 289, 2, 187-193.

Foresight Programme (2009) *Tackling Obesities: Future Choices – Modelling Future Trends in Obesity and the Impact on Health*, 2nd Edition,

Modelling Report for Government Office for Science, available from: [www.foresight.gov.uk](http://www.foresight.gov.uk), accessed 20/07/2011.

Fotheringham, A., Rees, P., Champion, T., Kalogirou, S., and Tremayne, A. (2004) The development of a migration model for England and Wales: Overview and modelling out-migration, *Environment and Planning A*, 36, 9, 1633–1672.

Galler, H. and Wagner, G. (1986) The microsimulations model of the Sfb3 for the analysis of economic and social policies. Orcutt *et al.* (eds.) *Microanalytic simulation models to support social and financial policy*, 227-247, Amsterdam, Holland.

Gardner, M. (1970) Mathematical Games - The fantastic combinations of John Conway's new solitaire game "life". *Scientific American*, 223, 120–123, available from: [http://web.archive.org/web/20090603015231/http://ddi.cs.uni-potsdam.de/HyFISCH/Produzieren/lis\\_projekt/proj\\_gamelife/ConwayScientificAmerican.htm](http://web.archive.org/web/20090603015231/http://ddi.cs.uni-potsdam.de/HyFISCH/Produzieren/lis_projekt/proj_gamelife/ConwayScientificAmerican.htm), accessed 11/22/2011.

Georgeff, M. and Lansky, A. (1987) Reactive reasoning and planning. *Proceedings of the Sixth National Conference on Artificial Intelligence (AAAI-87)*, 677-682, Seattle, WA.

Gilbert, N. and Troitzsch, K. (2005) *Simulation for the Social Scientist*. Open University Press, Berkshire, England, ISBN 0-335-19744-2.

Gilbert, N. (2002) Varieties of Emergence. Proceeding of the Workshop on Agent 2002 Social Agents at the Ecology, Exchange, and Evolution Conference 2002, available from: [http://www.soc.surrey.ac.uk/staff/ngilbert/ngpub/paper148\\_NG.pdf](http://www.soc.surrey.ac.uk/staff/ngilbert/ngpub/paper148_NG.pdf), accessed 23/11/11.

Gonçalves, A., Rodrigues, A. and Correia, L. (2004) Multi-agent simulation within geographic information systems. Proceedings of the Fifth workshop on Agent-based Simulation, Lisbon.

Grayer, M. (2011) Analysis of variation in small area life expectancy within London, 2000-2008. PhD Dissertation, Queen Mary London.

Griffiths, C. and Kirby, L. (2000) Geographic Variations in Conceptions to Women aged under 18 in Great Britain during the 1990s, *Population Trends*, 102, 13-23.

Grimm, V (1999) Ten years of individual-based modelling in ecology: what have we learned and what could we learn in the future? *Ecological Modelling*, 115, 129-148.

Hägerstrand, T. (1985) Time-Geography: focus on the corporeality of man, society, and environment, in *The science and praxis of complexity*, 193-216, The United Nations University.

Hägerstrand, T. (1957) Migration and area, in Hannerberg, D., Hagerstrand, T. and Odeving, B. (eds.) *Migration in Sweden*, 27-159. Lund, Sweden, C.W.K Gleerup.

Hägerstrand, T. (1953) *Innovation diffusion as a spatial process*, the University of Chicago Press, Chicago and London.

Harding, A. (2007) Challenges and opportunities of dynamic microsimulation modeling, in plenary paper presented to the 1st General Conference of the International Microsimulation Association, Vienna, August 2007, [www.euro.centre.org/ima2007/programme/papers/Challenges\\_Harding\\_Paper.pdf](http://www.euro.centre.org/ima2007/programme/papers/Challenges_Harding_Paper.pdf), accessed November 2011.

Harding, A. (2002) Using microsimulation models in the policy process: DYNAMOD, paper presented at Modelling Policy in an Ageing Europe Conference. London, U.K., available from: [www.natsem.canberra.edu.au/storage/cp2002\\_001.pdf](http://www.natsem.canberra.edu.au/storage/cp2002_001.pdf), accessed 20/08/2012.

Harold, E.R. (1996) *Java developer's resource: a tutorial and on-line supplement* (resource series), Prentice Hall, ISBN-10: 0135707897.

Hassan, S., Pavon, J. and Gilbert, N. (2008) Injecting Data into Simulation: Can Agent-Based Modelling Learn from Microsimulation? Working paper from the *Centre for Research in Social Simulation*, available from: <http://epubs.surrey.ac.uk/cress/4>, accessed 28/01/2010.

Helbing, D., Farkas, I. and Vicsek, T. (2000) Simulating dynamical features of escape panic. *Nature*, 407, 487-490.

Holm, E., Lindgren, U., Lundevaller, E. and Strömgren, M. (2006) The SVERIGE spatial microsimulation model, paper presented at the 8<sup>th</sup> Nordic Seminar on Microsimulation Models, Oslo, 8-9 June 2006, available from: [http://www.ssb.no/misi/papers/holm\\_einar\\_svesim.doc](http://www.ssb.no/misi/papers/holm_einar_svesim.doc), accessed 11/01/2012.

Holm, E., Lingren, U., Makila, K. and Malmberg, G. (1996) Simulating an entire nation. Clark, G. (ed.) *Microsimulation for urban and regional policy analysis*, Pion, London.

House of Commons Health Committee (2004) *Obesity*, The Stationery Office, London.

Jennings, N. (2000) On agent-based software engineering. *Artificial Intelligence*, 117, 277-296.

- Jennings, N. R. (1992) On being responsible. Werner, E. and Demazeau, Y. (eds.) *Decentralized AI 3. Proceedings of the Third European workshop on modelling autonomous agents and multi-agent worlds (MAAMAW-91)*, 93-102. Elsevier, Amsterdam, the Netherlands.
- Johnston, R. (2000) Microsimulation, in Johnston *et al.* (eds.) *The dictionary of human geography*, Blackwell, Oxford, ISBN 0-631-10721-5.
- Kahn, H. S., Tatham, L. M., Pamuk, E. R., and Heath, C. W. (1998) Are geographic region with high income inequality associated with risk of abdominal weight gain? *Social Science & Medicine*, 47, 1–6.
- Kearns, R. (1993) Place and health: towards a reformed medical geography? *Professional Geographer*, 45, 139–47.
- Kennell, D.L. and Sheils, J.F. (1990) PRISM: dynamic simulation of pension and retirement income, in Lewis G. H. and R.C. Michel (eds.) *Microsimulation techniques for tax and transfer analysis*, the Urban Institute Press, Washington D.C.
- Kim, D., Subramanian, S., Gortmaker, S., & Kawachi, I. (2006) US state and county level social capital in relation to obesity and physical inactivity: A multilevel multivariable analysis. *Social Science and Medicine*, 63, 1045–1059.
- Kliemt, H. (1996) Simulation and rational practice. Hegselmann *et al.* (eds.) *Modelling and simulation in the social sciences from a philosophy of science point of view*, 13-28, Kluwer, Dordrecht.
- Konolige, K. (1986) *A deduction model of belief*. San Mateo, CA, Pitman Publishing, London.
- Koplan, J. (2001) The continuing epidemic of obesity and diabetes in the United States, *Journal of American Medical Association*, 286,10, 1195-1200.
- Koser, K., and Salt, J. (1997) The geography of highly skilled migration. *International Journal of Population Geography*, 3, 4, 285–303.
- Kripke, S. (1963) Semantical analysis of modal logic. *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik*, 9, 67-96.
- Krupp, H. (1986) Potential and limitations of microsimulation models. Orcutt *et al.* (eds.) *Microanalytic simulation models to support social and financial policy*, North-Holland, Amsterdam.
- Kwan, M. (1998) Space-time and integral measures of individual accessibility: A comparative analysis using a point-based framework, *Geographical Analysis*, 30, 191-216.

Laird, J., Druitt, S., and Fraser, D. (1999) Edinburgh city centre: a micro-simulation case study, *Traffic Engineering and Control*, 40, 2, 72-76.

Lambert, S., Percival, R., Schofield, D. and Paul, S. (1994) An introduction to STINMOD: a static microsimulation Model, NATSEM Technical Paper No 1, University of Canberra, Australia, available from: [www.natsem.canberra.edu.au/storage/stp1.pdf](http://www.natsem.canberra.edu.au/storage/stp1.pdf), accessed 20/08/2012.

Lang, T. and Rayner, G. (2007) Overcoming policy cacophony on obesity: an ecological public health framework for policymakers, *Obesity reviews*, 8, 1, 165-181.

Langford, I. and Bentham, G. (1996) Regional Variations in Mortality Rates in England and Wales: An Analysis Using Multi-Level Modelling, *Social Science & Medicine*, 42, 6, 897-908.

Langton, C. (1991) Life at the edge of chaos, in Langton C. (ed.) *Artificial Life II*, 41-91, Addison-Wesley.

Law, A. (2007) *Simulation Modeling & Analysis*, Fourth edition, McGraw-Hill, Boston, ISBN: 9780073294414.

Leck, I. (1989) The north-south divide in England: implications for healthcare resource allocation. *Community Medicine*, 11, 102-107.

Leombruni R. and Richiardi M. (2006), LABORsim: An Agent-Based Microsimulation of Labour Supply, *Computational Economics*, 27, 1, 63-88.

Liu, R., Van Vliet, D. and Watling, D. (1995) DRACULA: dynamic route assignment combining user learning and microsimulation. Proceeding of PTRC summer conference, E, 143-152, available from: <http://www.its.leeds.ac.uk/software/dracula/downloads/Paper17-Paper%20presented%20at%20PTRC.pdf>, accessed 20/08/2012.

Loibl, W., and Toetzer, T. (2003) Modeling growth and densification processes in suburban regions-simulation of landscape transition with spatial agents. *Environmental Modelling and Software*, 18, 6, 553-563.

Lopez, A. (1983) The sex Mortality Differential in Developed Countries, in Lopez and Rukicka (eds.), *Sex differentials in Mortality: Trends, Determinants and Consequences*, 53-120, Australian National University, Canberra.

Luck, M., McBurney, P. and Preist, C. (2003) Agent technology: enabling next generation computing (A Roadmap for Agent Based Computing), ISBN 0854 327886.

Lutz, W., Sanderson, W. and Scherbov, S. (eds.) (2004) *The End of World Population Growth in the 21st Century*, Earthscan, London.

Magne, L., Rabut, S. and Gabard, J. (2000) Towards an hybrid macro-micro traffic flow simulation model, paper presented at INFORMS spring conference, Salt Lake City.

Mahdavi, B., O'Sullivan, D. and Davis, P. (2007) An Agent-based Microsimulation Framework For Investigating Residential Segregation Using Census Data, in Oxley, L. and Kulasiri, D. (eds.) *MODSIM 2007 International Congress on Modelling and Simulation*, 365-371. ISBN: 978-0-9758400-4-7.

Makowsky, M., Tavares, J., Makany, T., and Meier, P. (2006) An agent-based model of crisis-driven migration, in proceedings of the complex systems summer school 2006, Santa Fe, NM: Santa Fe Institute, available from:  
[http://192.12.12.16/events/workshops/images/6/67/Sf\\_csss06\\_makowsky\\_et\\_al.pdf](http://192.12.12.16/events/workshops/images/6/67/Sf_csss06_makowsky_et_al.pdf).

Marmot, M. (2005) Social Determinants of Health Inequalities, *The Lancet*, 365, 1099-104.

McDonald, P., Kippen, R. and Temple, J. (2006) Net transition probabilities: an approach to subnational level projections of households and housing demand based on census data, *Population, Space and Place*, 12, 479-495.

McLaren, L. (2007) Socioeconomic Status and Obesity, *Epidemiologic Reviews*, 29, 1, 29-48. DOI: 10.1093/epirev/mxm001.

McTigue K., Larson, J., Valoski, A., Burke, G., Kotchen, J., Lewis, C., Stefanick, M., Van Horn, L. and Kuller L. (2006) Mortality outcomes and cardiac and vascular outcomes in extremely obese women, *Journal of American Medical Association*, 296, 1, 79-86.

Meiss, J. (2007) Dynamical systems. *Scholarpedia*, 2, 2, 1629.

Miller, H. (1991) Modelling accessibility using space-time prism concepts within geographical information systems, *International Journal of Geographical Systems*, 5, 287-301.

Minar, N., Burkhart, R., Langton, C. and Askenazi, M. (1996) The Swarm Simulation System: A Toolkit for Building Multi-agent Simulations, available at: <http://www.santafe.edu/research/working-papers/abstract/70e0c84b2cff6e8e87f2940fa60965e9/>, accessed 23/11/2011.

Möhring, M. (1996) Social science multilevel simulation with MIMOSE. Troitzsch *et al.* (eds.) *Social Science Microsimulation*, 123-137, Springer Verlag, Berlin, ISBN 3-340-61572-5.

Mokdad, A., Bowman, B., Ford, E., Vinicor, F., Marks, J. and Must, A., Spadano, J., Coakley, E., Field, A., Colditz, G. and Dietz, W. (1999) The disease burden associated with overweight and obesity, *Journal of American Medical Association*, 282,16, 1523-1529.

Molnar, I. and Sinka, I. (2007) Toward Agent-Based Microsimulation – Another Approach, in proceedings of the First Asia International Conference On Modelling and Simulation (AMS'07), 403-408, DOI: 10.1109/AMS.2007.104.

Moon, Graham (2009) Health geography, in Kitchin, R. and Thrift, N. (eds.) *International Encyclopedia of Human Geography*, 35-55, Elsevier, Oxford, UK. DOI:10.1016/B978-008044910-4.00338-2.

Moon, G., Quarendon, G., Barnard, S., Twigg, L. and Blyth, B. (2007) Fat nation: deciphering the distinctive geographies of obesity in England, *Social science and medicine*, 65, 25-31.

Moon, G., Altankhuyag, G., Barnard, S. and Twigg, L. (2006) Smoking Prevalence in Scotland: 2003/4 sub-national estimates (A Report for NHS Health Scotland), available at: <http://www.scotpho.org.uk/nmsruntime/saveasdialog.asp?IID=3901&sID=3256>, last access 28/04/10.

Moore, R. C. (1990) A formal theory of knowledge and action, in Allen, J., Hendler, J and Tate, A (eds.) *Readings in Planning*, 480-519, Morgan Kaufmann Publishers, San Mateo, CA, ISBN: 1558601309.

Moraitis, P. and Spanoudakis, N. (2004) Combining Gaia and JADE for multi-agent systems development, proceedings of The 17th European Meeting on Cybernetics and Systems Research (EMCSR 2004), Vienna, Austria, April 13 - 16, 2004.

Morrison, R.J. (2003) Making pensions out of nothing at all, *The International microsimulation Conference on Population, Ageing and Health: Modelling our Future*, available from: [https://guard.canberra.edu.au/natsem/conference2003/papers/pdf/morrison\\_ichard1.pdf](https://guard.canberra.edu.au/natsem/conference2003/papers/pdf/morrison_ichard1.pdf), accessed 01/11/2011.

Moss, S. (2000) Messy Systems - The Target for Multi Agent Based Simulation. Moss, S. and Davidsson, P. (eds.) *Multi Agent Based Simulation*, Lecture Notes of Computer Science, 1979, Springer Verlag, Berlin.

Murphy, J., Sexton, D., Barnett, D, Jones, G, Webb, M, Collins, M. and Stainforth, D. (2004) Quantification of modelling uncertainties in a large ensemble of climate change simulations, *Nature*, 430, 768-772.

Murphy, M. (2001) Bringing behaviour back into micro-simulation: Feedback mechanisms in demographic models, in proceedings of Workshop

on Agent-based Computational Demography, Max Planck Institute for Demographic Research, February 21-23, available from: [www.demogr.mpg.de/Papers/workshops/010221\\_paper10.pdf](http://www.demogr.mpg.de/Papers/workshops/010221_paper10.pdf), accessed 18/06/12.

Murphy, L. (1995) Geographic information systems: are they decision support systems? *Proceedings of 28th Hawaii International Conference on System Sciences*, 4, 131-140, available from: [http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=375736&url=http%3A%2F%2Fieeexplore.ieee.org%2Fxppls%2Fabs\\_all.jsp%3Farnumber%3D375736](http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=375736&url=http%3A%2F%2Fieeexplore.ieee.org%2Fxppls%2Fabs_all.jsp%3Farnumber%3D375736), accessed 20/06/2011.

NAO (National Audit Office) (2001) *Tackling obesity in England*. The Stationery Office, London.

NHS (National Health Service) (2010) *Statistics on obesity, physical activity and diet: England*, The Information Centre for Health and Social Care document, available from: <http://www.ic.nhs.uk/statistics-and-data-collections/health-and-lifestyles/obesity/>, accessed 26/06/12.

NHS (2009) *Health Survey for England – 2008 Trend Tables*, available from [www.ic.nhs.uk/pubs/hse08trends](http://www.ic.nhs.uk/pubs/hse08trends), accessed 12/12/2010.

Norman, P., Boyle, P. and Rees, P. (2004) Selective migration, health and deprivation: a longitudinal analysis. *Social Science and Medicine*, 60, 12, 2755-2771.

O'Donoghue, C. (2001) Dynamic microsimulation: a methodological survey. *Brazilian Electronic Journal of Economics*, 4, 2, available from: [http://www.microsimulation.org/IMA/BEJE/BEJE\\_4\\_2\\_2.pdf](http://www.microsimulation.org/IMA/BEJE/BEJE_4_2_2.pdf), accessed 23/11/2011.

ONS (Office for National Statistics) (2011a) *A Beginner's Guide to Geography*, available from: <http://www.ons.gov.uk/ons/guide-method/geography/ons-geography/index.html>, last access 19/12/2011.

ONS (Office of National Statistics) (2011b) *Frequently asked questions: births & fertility*, available from: <http://www.ons.gov.uk/ons/rel/vsob1/parents--country-of-birth--england-and-wales/2010/faq-births-and-fertility.pdf>, accessed 10/02/2011.

ONS (Office for National Statistics) (2009) *UK Interim Life Tables, 1980-82 to 2006-08*, available at: <http://www.statistics.gov.uk/StatBase/Product.asp?vlnk=14459&Pos=&ColRank=1&Rank=422>, accessed 27/04/10.

ONS (2008a) *Subnational Population Projections Accuracy Report*, available from: <http://www.ons.gov.uk/ons/rel/snpp/sub-national-population-projections/2006-based-projections/subnational-population-projections--accuracy-report.pdf>, accessed 10/02/2011.



ONS (2008b) *National Population, Projections, 2006-based, Series PP2 No 26*, available from: <http://www.ons.gov.uk/ons/rel/npp/national-population-projections/2006-based-annual-reference-volume/2006-based-reference-volume-no--26.pdf>, accessed 10/02/2011.

ONS (2008c) *Population Trends No. 134*, available from: <http://www.ons.gov.uk/ons/rel/population-trends-rd/population-trends/no--134--winter-2008/population-trends.pdf>, accessed 10/02/2011.

ONS (2008d) *Population Trends No. 131*, available from: <http://www.ons.gov.uk/ons/rel/population-trends-rd/population-trends/no--131--spring-2008/population-trends.pdf>, accessed 10/02/2011.

ONS (2008e) *National Population projections: 2006-based*, available from: <http://www.ons.gov.uk/ons/rel/npp/national-population-projections/2006-based-projections/index.html>, accessed 12/03/2012.

ONS (2007a) Web supplement, Annual Update: Births in England and Wales 2006, *Population Trends 130*, available at [http://www.statistics.gov.uk/downloads/theme\\_population/Births\\_update\\_web\\_supplement.pdf](http://www.statistics.gov.uk/downloads/theme_population/Births_update_web_supplement.pdf), last access 27/04/10.

Office for National Statistics (2007b) *2004-based sub-national population projections for England (revised)*, National Statistics Centre for Demography. <http://www.statistics.gov.uk/statbase/product.asp?vlnk=997>, accessed 17.04.2008.

ONS (2007c) *2006 Mid-year estimates*, available from: <http://www.ons.gov.uk/ons/rel/pop-estimate/population-estimates-for-uk--england-and-wales--scotland-and-northern-ireland/2006/index.html>, accessed 12/03/2012.

ONS (Office for National Statistics) (2006a) *2001 United Kingdom Sample of Anonymised Records, Household Special Licensed File*. Distributed by the UK Data Archive, University of Essex.

ONS (2006b) *2001 United Kingdom Sample of Anonymised Records, Individual Licensed File*. Distributed by the Cathie Marsh Centre for Census and Survey Research, University of Manchester.

ONS (2006c) *Population Trends - No. 123, Spring 2006*, <http://www.ons.gov.uk/ons/rel/population-trends-rd/population-trends/no--123--spring-2006/index.html>, accessed 12/12/2011.

ONS (2006d) *2001 Individual Licensed SAR Codebook*, available from: <http://www.ccsr.ac.uk/sars>, accessed 11/01/2012.

ONS (2002) Birth statistics, England and Wales (Series FM1) - No. 30, available from: <http://www.ons.gov.uk/ons/rel/vsob1/birth-statistics-->

england-and-wales--series-fm1-/no--30--2001/birth-statistics-series-fm1.pdf, accessed 11/01/2012.

ONS (2001a) *Guide to 2001 Census data*, available from: <http://www.ons.gov.uk/ons/guide-method/census/2011/get-data/guide-to-census-data/index.html>, accessed 12/12/2011.

ONS (2001b) Table VS4: vital statistics for wards 2001 - 2001 boundaries, Produced by Program Annual\_ICD10\_VS4\_2001, *Vital Statistics for England and Wales, 2001*. Distributed by UK Data Archive, Colchester, Essex, SN: 4608, <http://dx.DOI.org/10.5255/UKDA-SN-4608-1>.

ONS (2001c) Table 3.2: Maternities (total), live births and stillbirths (totals and female): age of mother and occurrence within/outside marriage, 2001, in Series FM1 no 31, *Birth Statistics*, <http://www.ons.gov.uk/ons/rel/vsob1/birth-statistics--england-and-wales--series-fm1-/no--31--2002/birth-statistics-series-fm1.pdf>, accessed 12/12/2011.

ONS (2001d) Table PBH63A: Births: Maternities with multiple births: rates per 1,000 maternities, age of mother, 1938-2004, a. all maternities, *Historic Births on Statbase*.

ONS (2001e) Census 2001 Standard Table: ST 016: Age and sex by general health and limiting long-term illness), available from: <http://casweb.mimas.ac.uk/>, last access 11/01/2012.

ONS, (2001f) Commissioned Table C0729 - Age of Married Couples England: contains age of husband and wife's age from 16 to 100+. For access to commissioned tables, see <http://www.ons.gov.uk/ons/about-ons/who-we-are/services/unpublished-data/census-data/census-commissioned-tables/index.html>.

ONS (2001g) 2001 Census Standard Table: Table ST01: Age by sex and resident type, available from: <http://casweb.mimas.ac.uk/>, last access 11/01/2012.

ONS (2001h) VITAL STATISTICS FOR WARDS 2001 - 2001 BOUNDARIES, Table VS4: produced by Program Annual\_ICD10\_VS4\_2001. The LA level data is provided by ONS Vital Statistics and Paul Norman distributed into wards.

ONS (2001i) 2001 Census Standard Table 016: General Health by Age and Gender, available from: <http://casweb.mimas.ac.uk/>, accessed 11/01/2012.

ONS (2001j) *Historic Births on Statbase*, Table PBH63A: Births: Maternities with multiple births: rates per 1,000 maternities, age of mother, 1938-2004, a. all maternities.

ONS (2001k) International Passenger Survey (IPS) 2001, data available through:

[http://www.statistics.gov.uk/ssd/surveys/international\\_passenger\\_survey.asp](http://www.statistics.gov.uk/ssd/surveys/international_passenger_survey.asp), accessed 10/02/2011.

ONS (2001l) 2001 Census Special Migration Statistics (Levels 2), Available through the Centre for Interaction Data Estimation and Research (CIDER) at: <http://cider.census.ac.uk/cider/wicid/query.php>, accessed 10/02/2011.

ONS (2001m) Census 2001 Disclosure Protection Measures, <http://www.statistics.gov.uk/census2001/discloseprotect.asp>, accessed 22/02/2011.

ONSCD (Office for National Statistics Centre for Demography) (2008) Subnational Population Projections Accuracy Report, available from: <http://www.ons.gov.uk/ons/rel/snpp/sub-national-population-projections/2006-based-projections/subnational-population-projections--accuracy-report.pdf>, accessed 12/03/2012.

Orcutt, G., Merz J. and Quinke, H. (eds.) (1986) *Microanalytic Simulation Models to Support Social and Financial Policy*. Elsevier, Amsterdam, North Holland, ISBN: 9780444878762.

Orcutt, G. (1957) A new type of socio-economic system. *Review of Economics and Statistics*, 58, 773-797.

Parker, M. (2001) What is Ascape and why should you care? *Journal of Artificial Societies and Social Simulation* 4(1), available at: <http://jasss.soc.surrey.ac.uk/4/1/5.html>, accessed 23/11/2011.

Pearce, J. and Witten, K. (2010) (eds.) *Geographies of obesity: environmental understandings of the obesity epidemic*, Ashgate, Aldershot, ISBN: 9780754676195.

Power, D. (2002) *Decision support systems: concepts and resources for managers*. Westport, Quorum Books, ISBN: 156720497X.

Pred, A. (1977) The choreography of existence: Comments on Hagerstrand's time-geography and its usefulness, *Economic Geography*, 53, 207-21.

Propper, C. (1995) For richer, for poorer, in sickness and in health: the lifetime distribution of NHS health care, in: Falkingham J. and Hills J. (eds.) *The dynamics of welfare*, Prentice Hall Wheatsheaf, London.

PTV AG (2011) VISSIM - Multi-Modal Traffic Flow Modeling, available from: <http://www.vissim.de/>, accessed 23/11/2011.

Pudney, S. (1994) How reliable are microsimulation results: an analysis of the role of sampling error in a U.K. tax-benefit model, *Journal of Public Economics*, 53, 327-365.

Quesenberry, C., Caan, B. and Jacobson, A. (1998) Obesity, health services use, and health care costs among members of a health maintenance organization, *Archives of Internal Medicine*, 158, 5, 466–472.

Rami, B., Schober, E., Kirchengast, S., Waldhor, T., & Seifanek, R. (2004) Prevalence of overweight and obesity in male adolescents in Austria between 1985 and 2000: A population based study. *Journal of Paediatric Endocrinology and Metabolism*, 17, 1, 67–72.

Raney, B. and Nagel, K. (2003) An agent-based simulation model of swiss travel: first results, 3rd Swiss transport research conference, available from: [www.strc.ch/conferences/2003/raney.pdf](http://www.strc.ch/conferences/2003/raney.pdf), accessed 20/08/2012.

Rao, A. and Georgeff, M. (1991) Modeling rational agents within a BDI-architecture, in Fikes, R. and Sandewall, E. (eds.) *Proceedings of Knowledge Representation and Reasoning (KR&R-91)*, 473-484, San Mateo, CA, Morgan Kaufmann Publishers.

Redmond, G., Sutherland, H. and Moira Wilson, M. (1998) *The arithmetic of tax and social security reform: a user's guide to microsimulation methods and analysis*, Cambridge University Press, Cambridge.

Rees, P., Norman, P., Wohland, P. and Boden, P. (2010) What happens when international migrants settle? : ethnic group population trends and projections for United Kingdom local areas, British Society for Population Studies annual conference 2010, available from: [http://www.geog.leeds.ac.uk/fileadmin/downloads/school/research/projects/migrants/Ethnic\\_group\\_projections\\_for\\_UK\\_local\\_areas.pdf](http://www.geog.leeds.ac.uk/fileadmin/downloads/school/research/projects/migrants/Ethnic_group_projections_for_UK_local_areas.pdf), accessed 20/08/2012.

Rees, P. (2009) Demography, in Kitchin, R. and Thrift, N. (eds.) *International Encyclopedia of Human Geography*, 3, 75-90. Elsevier, Amsterdam, ISBN 978-0-08-044911-1.

Rees P, Wohland P and Norman P (2009) The estimation of mortality for ethnic groups at local scale within the United Kingdom, *Social Science and Medicine*, 69, 1592-1607, DOI:10.1016/j.socscimed.2009.08.015.

Rees, P. (2007) GEOG3031 Life tables lecture, available from: <http://vle.leeds.ac.uk/site/nbodington/geography/geoglev3/geog3031/week3/>, accessed 28/04/10.

Rees, P., Norman, P. and Brown, D. (2004a) A framework for progressively improving small area population estimates. *Journal of the Royal Statistical Society, A, Statistics in Society*, 167, 1, 5-38.

Rees, P., Kupiszewski, M., Eyre, H., Wilson, T. and Durham, H. (2001) *The evaluation of regional population projections for the European Union*. Eurostat Working Paper, Population and Social Conditions 3/2001/E/no9. European Commission: Luxembourg. Catalogue No. KA-AP-01-041-EN-I.

Rees, P. (1994) Estimating and projecting the populations of urban communities, *Environment and Planning A*, 26, 11, 1671–1697.

Rees, P. (1984) Spatial population analysis using movement data and accounting methods: theory, models, the “MOVE” program and examples, in Working Paper404/Computer Manual 23, School of Geography, University of Leeds, Leeds, UK.

Rees, P. and Birkin, M. (1983) Census-based information systems for ethnic groups: A study of Leeds and Bradford. *Environment and Planning A*, 16, 1551–1571.

Rees, P. and Wilson, A. (1977) *Spatial population analysis*, Edward Arnold, London.

Rephann, T. (2001) Economic-demographic effects of immigration: Results from a dynamic, spatial microsimulation model, the 2001 Annual Meeting of the Mid-Atlantic Division of the Association of American Geographers, Maryland, available from: <http://www.equotient.net/papers/immic.pdf>, accessed 12/03/2012.

Rephann, T. (1999) The *Education Module for SVERIGE: Documentation V 1.0*, available from: <http://www.equotient.net/papers/educate.pdf>, accessed 23/11/2011.

Reynolds, C. (1999) Individual-Based Models, available from: <http://www.red3d.com/cwr/ibm.html>, accessed 23/11/2011.

Robertson, D. (2003) Agent-based models of a banking network as an example of a turbulent environment: the deliberate vs. emergent strategy debate revisited, *Emergence*, 5, 2, 56-71.

Rogers, A. (1995) *Multiregional demography: principles, methods and extensions*, Wiley, Chichester.

Rogers, A. (1975) *Introduction to multiregional mathematical demography*, Wiley, New York.

Rouchier, J. (2001) Review of Ferber, J. (1999) Multi-Agent System: an introduction to distributed artificial intelligence, in *journal of artificial societies and social simulation* 4 (2), available from: <http://jasss.soc.surrey.ac.uk/4/2/reviews/rouchier.html>, accessed 23/11/2011.

Rowland, D. (2003) *Demographic methods and concepts*, Oxford University Press Inc, New York, USA, ISBN: 019875263.

Rubinstein, R. and Kroese, D. (2008) *Simulation and the Monte Carlo method*, Wiley, Hoboken, NJ, ISBN: 9780470177945.

Russell, S., and Norvig, P. (1995) *Artificial intelligence: A modern approach*. Englewood Cliffs, NJ: Prentice Hall, ISBN 0136042597.

Salvini, P. and Miller, E. (2005) ILUTE: An Operational Prototype of a Comprehensive Microsimulation Model of Urban Systems. *Networks and Spatial Economics*, 5, 2, 217-234. ISSN: 1566113X, DOI: 10.1007/s1106700526305.

Salvini, P. and Miller, E. (2003) ILUTE: An Operational Prototype of a Comprehensive Microsimulation Model of Urban Systems, proceedings of the 10<sup>th</sup> International Conference on Travel Behaviour Research, Lucerne, August 2003, available from: <http://www.ivt.baug.ethz.ch/allgemein/pdf/salvini.pdf>, accessed 18/06/12.

Sawyer, K. (2003) Artificial societies - Multiagent systems and the Micro-Macro link in sociological theory. *Sociological Methods & Research*, 31, 3, 325-363, DOI: 10.1177/0049124102239079.

Schelhorn, T., O'Sullivan, D., Haklay, M. and Thurstain-Goodwin, M. (1999) STREETS: an agent-based pedestrian model. Presented at Computers in Urban Planning and Urban Management, Venice, 8-11 September.

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

Seeman, T., Crimmins, E., Huang, M., Singer, B., Bucur, A., Gruenewald, T., Berkman, L. and Reuben, D. (2004) Cumulative biological risk and socio-economic differences in mortality: MacArthur Studies of Successful Aging, *Social Science & Medicine*, 58, 10, 1985-1997.

Shaw, M., Davey Smith, G., Thomas, B., and Dorling, D. (2008) *The Grim Reaper's road map: an atlas of mortality in Britain*, Policy Press, Bristol, ISBN: 9781861348234.

Smith, D., Edwards, K., Clarke, G. and Harland, K. (2010) Measuring obesogenic environments-representing place in studies of obesity. In Pearce, J. and Witten, K. (eds.) *Geographies of obesity: environmental understandings of the obesity epidemic*, Aldershot, Ashgate.

Smith, S. and Tayman, J. (2003) An evaluation of population projections by age, *Demography*, 40, 741-757.

Smith, S. and Shahidullah, M. (1995) An evaluation of population projection errors for census tracts. *Journal of the American Statistical Association*, 90, 64-71.

Stamatakis, E. (2006) Chapter 5: Obesity, eating and physical activity, Focus on Health 2006,

[http://www.statistics.gov.uk/downloads/theme\\_compendia/foh2005/05\\_Obesity\\_Eating\\_PhysicalActivity.pdf](http://www.statistics.gov.uk/downloads/theme_compendia/foh2005/05_Obesity_Eating_PhysicalActivity.pdf) , accessed 12/12/2010.

Stunkard, A. J. (2007) Socioeconomic Status and Obesity, in D. J. Chadwick and G. Cardew (eds.) *Ciba Foundation Symposium 201 - The Origins and Consequences of Obesity*, John Wiley & Sons, Chichester, UK. DOI: 10.1002/9780470514962.ch11.

Sutherland, H. (2007) EUROMOD – the tax-benefit microsimulation model for the European Union. Chapter 10, 477-482 in Gupta, A. and Harding, A. (eds.) *Modelling Our Future: Population Ageing, Health and Aged Care*, Amsterdam: Elsevier Science, ISBN: 9780444530059.

Sutherland, H. (2001) EUROMOD: an integrated European benefit-tax model, Final report, EUROMOD Working Paper No EM9/01, Department of Applied Economics, University of Cambridge, Cambridge.

Swinburn, B. and Egger, G. (2002) Preventive strategies against weight gain and obesity, *Obesity Reviews*, 3, 289-30.

Sykes, P., De Jong, F., Bradley, R., Jennings, G. and McDonnell, G. (2010) Planning urban car park provision using microsimulation, *Traffic Engineering and Control*, 51,3, 103-107.

Taylor, M., (ed.) with Brice, J., Buck, N. and Prentice-Lane, E. (2010) British Household Panel Survey User Manual Volume A: Introduction, Technical Report and Appendices. Colchester: University of Essex, available from: [https://www.iser.essex.ac.uk/bhps/documentation/pdf\\_versions/survey\\_docs/index.html](https://www.iser.essex.ac.uk/bhps/documentation/pdf_versions/survey_docs/index.html), accessed 12/12/2011.

The Marmot Review Commission (2010) *Fair society healthy lives: A Strategic Review of Health Inequalities in England Post-2010* (The Marmot Review), available from: <http://www.marmotreview.org/>, accessed 26/06/12.

Toffoli, T. and Margolus, N. (1987) *Cellular Automata machines*, the MIT Press, Cambridge, Massachusetts.

Torrens, P. and Benenson, I. (2005) Geographic Automata Systems. *International Journal of Geographical Information Science*, 19, 4, 385-412.

Transims Travelogue (1996) Transims Travelogue (LAUR-96-1430), available at: [http://transims.tsasa.lanl.gov/PDF\\_Files/Travel4.pdf](http://transims.tsasa.lanl.gov/PDF_Files/Travel4.pdf), last viewed 10/12/05.

Treasury Board of Canada Secretariat (2000) The itthink® and PERSIM forecasting and succession planning models, available from: [http://www.hrma-agrh.gc.ca/hr-rh/psds-dfps/dafps\\_f\\_model1\\_e.asp](http://www.hrma-agrh.gc.ca/hr-rh/psds-dfps/dafps_f_model1_e.asp), last viewed 10/12/05.

Tromans, N., Natamba, E., Jefferies, J. and Norman, P. (2008) Have national trends in fertility between 1986 and 2006 occurred evenly across England and Wales? *Population Trends*, 133, 7-19, available from: <http://www.ons.gov.uk/ons/rel/population-trends-rd/population-trends/>, accessed 19/06/2012.

Twigg, L. and Moon, G. (2002) Predicting small-area health-related behaviour: a comparison of multilevel synthetic estimation and local survey data. *Social Science and Medicine*, 50, 1109-20.

Vallin, J., Caselli, G. and Surault, P. (2005) Life Styles and Socio-Cultural Factors of Mortality, in G.Caselli, J. Vallin and G. Wunsch (eds.) *Demography: Analysis and Synthesis*, 2(1), Academic Press. ISBN: 978-0-12-765660-1.

Valkonen, T. (2005) Social inequalities in mortality, in G.Caselli, J. Vallin and G. Wunsch (eds.) *Demography: Analysis and Synthesis*, 2,1, Academic Press. ISBN: 978-0-12-765660-1.

Van der Gaag, N., de Beer, J. and Willekens, F. (2005) MicMac, combining micro and macro approaches in demographic forecasting. The Joint Eurostat-ECE Work Session on Demographic Projections, Vienna, available from: [http://circa.europa.eu/irc/dsis/jointestatunece/info/data/paper\\_vandergaag.pdf](http://circa.europa.eu/irc/dsis/jointestatunece/info/data/paper_vandergaag.pdf), accessed 23/11/11.

Van Imhoff, E. and Post, W. (1998) Microsimulation methods for population projection. *Population: An English Selection*, 10, 97–138.

Voas, D. and Williamson, P. (2001) Evaluating goodness-of-fit measures for synthetic microdata, *Geographical and Environmental Modelling*, 5, 2, 177-200.

Von Neumann, J. and Burks, A. (1966) (ed.) *Theory of Self-Reproduction Automata*, University of Illinois Press.

Walker, A., Percival, R. and Fischer, S. (1998) Microsimulation model of Australia's pharmaceutical benefits scheme. Technical Paper No. 15, National Centre for Economic Modelling, University of Canberra.

WHO (1948) The Constitution of the World Health Organization, available from: [http://policy.who.int/cgi-bin/om\\_isapi.dll?infobase=Basicdoc&softpage=Browse\\_Frame\\_Pg42](http://policy.who.int/cgi-bin/om_isapi.dll?infobase=Basicdoc&softpage=Browse_Frame_Pg42). Accessed 11/01/2012.

Wikipedia (2010) Economic base analysis, available from: [http://en.wikipedia.org/wiki/Economic\\_base\\_analysis](http://en.wikipedia.org/wiki/Economic_base_analysis), accessed 11/08/2010.

Wilkinson, R. and Marmot, M. (2003) *The Solid Facts*, WHO document, available from:



[http://www.euro.who.int/\\_\\_data/assets/pdf\\_file/0005/98438/e81384.pdf](http://www.euro.who.int/__data/assets/pdf_file/0005/98438/e81384.pdf), accessed 22/02/2011.

Wilkinson, R. and Marmot, M. (2003) (eds.) *Social Determinants of Health: the Solid Facts*, WHO Regional Office for Europe, ISBN: 9789289013710.

Willekens, F. (2005) Biographic forecasting: bridging the micro-macro gap in population forecasting, *New Zealand Population Review*, 31, 1, 77-124.

Wilson, B. and Stuchbury, R (2010) Do partnerships last? Comparing marriage and cohabitation using longitudinal census data, *Population Trends* 139, 37-63.

WHO (World Health Organisation) (2012) BMI classification, available from: [http://apps.who.int/bmi/index.jsp?introPage=intro\\_3.html](http://apps.who.int/bmi/index.jsp?introPage=intro_3.html), accessed 22/02/2011.

WHO (2010) Parma Declaration on Environment and Health, Parma, Italy, available from: [http://www.euro.who.int/\\_\\_data/assets/pdf\\_file/0011/78608/E93618.pdf](http://www.euro.who.int/__data/assets/pdf_file/0011/78608/E93618.pdf), accessed 22/02/2011.

WHO (2006) Fact Sheet No. 311: Obesity and overweight, available from: <http://www.who.int/mediacentre/factsheets/fs311/en/>, accessed 20/07/2011.

WHO (2004a) Report of a WHO Consultation on Obesity: preventing and managing the global epidemic. WHO Technical Report Series; 894. Geneva, available from: <http://www.who.int/nutrition/publications/obesity/en/index.html>, accessed 12/12/2010.

WHO (2004b) Appropriate body-mass index for Asian populations and its implications for policy and intervention strategies, *The Lancet*, 157-163.

WHO (2000) Obesity: preventing and managing the global epidemic: Report of a WHO Consultation, WHO Technical Report Series 894, Geneva.

White, C. (2003) Cancer rates in Europe are linked to overweight, experts warn. *British Medical Journal*, 327-700.

Wilkinson, R. and Pickett, K. (2009) *The Spirit Level: Why More Equal Societies Almost Always Do Better*, London, Allen Lane, ISBN 978-1-84614-039-6.

Williamson, P. (1999) Microsimulation: An idea whose time has come? Paper presented at The 39th European Congress of The European Regional Science Association, August 23-27, Dublin, Ireland.

Wilson, A.G. and Bennett, R.J. (1985) *Mathematical methods in human geography and planning*, John Wiley & Sons, ISBN 047110521.

Wilson, A. (1967) A statistical theory of spatial trip distribution models, *Transportation Research*, 1, 253-269.

Wolfram, S. (1984) Universality and complexity in cellular automata. *Physica D*, 10, 1-35.

Wolfson, M. and Rowe, G. (1998) LifePaths – Toward an integrated microanalytic framework for socio-economic statistics, paper presented to the 26th General Conference of the International Association for Research in Income and Wealth, Cambridge, UK.

Wooldridge, M. and Jennings, N. (1995) Intelligent agents: theory and practice, *Knowledge Engineering Review*, 10, 2, 115-152, available from: <http://www.csc.liv.ac.uk/~mjw/pubs/ker95/ker95-html.html>, accessed 23/11/2011.

Woods, R. (2000) *The demography of Victorian England and Wales*. Cambridge University Press, Cambridge. ISBN: 0-521-78254-6.

Wu, B. M. and Birkin, M. H. (2012) Moses: A dynamic spatial microsimulation model for demographic planning, in Tanton, R. and Edwards, K. (eds.) *Spatial Microsimulation: A Reference Guide for Users*, Understanding Population Trends and Processes 6, DOI 10.1007/978-94-007-4623-7\_11.

Wu, B. M., Birkin, M. H. and Rees, P. H. (2010) A dynamic microsimulation with agent elements for spatial demographic forecasting, *Journal of Social Science Computing Review*, 29 (1), 145-160, DOI: 10.1177/0894439310370113.

Wu, B. M., Birkin, M. H. and Rees, P. H. (2008) A spatial microsimulation model with student agents, *Journal of Computers, Environment and Urban Systems*, 32, 440-453, DOI: 10.1016/j.compenvurbsys.2008.09.013.

Wu, B.M. and Hine, J.P. (2003) A PTAL approach to measuring changes in bus service accessibility, *Journal of Transport Policy*, 10, 4.

Zaidi, A. (2004) Estimation of earnings in the SAGE dynamic microsimulation model, *Technical note 10*, ESRC SAGE Research Group, London School of Economics, London.

Zaidi, A. and Rake, K. (2001) Dynamic Microsimulation Models: A Review and Some Lessons for SAGE, SAGE Discussion Paper no. 2, available from:

<http://citeseerx.ist.psu.edu/viewdoc/download?DOI=10.1.1.96.1328&rep=rep1&type=pdf>, accessed 12/03/2011.

# Appendix A

## Abbreviations

ABM: Agent Based Model

CA: Cellular Automata

DSMSM: Dynamic Spatial Micro-Simulation Model

HSAR: Household Samples of Anonymised Records

IBM: Individual Based Model

ISAR: Individual Samples of Anonymised Records

LA: Local Authority

MAS: Multiple Agent System

MSM: Micro-Simulation Model

ONS: *Office for National Statistics*

OOP: Object Oriented Programming