Using a Hybrid Model for Investigating Residential Segregation: An Empirical and Simulation-based Study

Babak Mahdavi Ardestani

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

The University of Auckland

2013

Abstract

The impact of residential segregation on the dynamics of spatial, social and economic spheres of society is a topic of great interest in geography, sociology and economics. Residential segregation and its spatial separation effects have been acknowledged as having significant impacts on education, healthcare, business as well as social network and social/ urban structures. The complexity and multidimensional nature of the residential segregation phenomenon and the centrality of the individual based decision making process made this topic an ideal case for investigation using a micro/ individual based simulation modelling approach. An example of such an approach is agent-based modelling (ABM) as theorised by Thomas Schelling. However, most Schelling-type models are often too simple and small. This lack of sophistication and 'small-village' syndrome remain though among the major weak points of existing models generally. More importantly, the lack of empirical support for informing and verification has long impeded the widespread acceptance of most simulation modelling approaches of this kind.

Various individual based simulation modelling approaches to investigate residential segregation are reviewed. In particular, two simulation modelling approaches – agent-based modelling (ABM) and microsimulation (MSM) – are compared with the aim of embracing a combined design approach that will also include key features of geosimulation models. For this reason a series of model prototypes are built initially to examine different aspects of a combined design approach, and with consideration of available census data (in aggregate format), the *HAAMoS* model is ultimately presented. It can simulate the entire population of the Auckland metropolitan area whilst dealing with up to four major ethnic groups each of which exhibit heterogeneous behaviours and have multi-level preferences. It can also measure various dimensions of segregation, including local and spatially sensitive ones at different geographical scales. The implementation of these features is described.

A descriptive statistical analysis of the modified data describes past and present patterns of residential location by ethnicity in the Auckland region area, which in turn are used as 'benchmarks' against the model's outputs. Using specific scenarios, it is demonstrated that this relatively simple Schelling-type model – informed by empirical data – has the potential to replicate plausible residential distribution patterns, even though the detailed representation of decision making behaviours are not available/ used. This demonstration confirms that the development of a potential 'test-bed' consisting of an agent-based model using census data for future modelling-based research which can address residential and socio-spatial segregation questions and similar theoretical issues in urban geography, sociology or economics would be feasible.

The methodological approaches built and used in this research are among its important achievements. Further possible extensions of this research under different topics are also discussed.

__World__ is a lunatic sphere,

Don't always agree it's __real__,

Even with my feet upon it, […],

My __address__ is somewhere else.

– (Hafez, Persian poet)

In memory of Suri Anisi

For Penney & Aryana

with love

Acknowledgments

As with any project with this magnitude, debts of gratitude are accumulated along the way. Many people deserve a vote of thanks for their inspiration, support and encouragement over the duration of this project. Special mention must go to my supervisors, Professors David O'Sullivan and Peter Davis, for their invaluable advice, insights, guidance, support, endurance and patience. Each of them contributed a unique perspective that shaped the direction of this research work, whilst also proving willing to give me a free hand to build a unique modelling framework and to bring the thesis to a successful conclusion.

Numerous collaborators, colleagues and friends have also given me much needed support at each stage of the process of pursuing this degree. I am therefore indebted to each of them. Alan Kwok Lun Cheung was instrumental in helping me with GIS related issues and questions. Roy Lay-Yee and Lyndon Walker were also helpful in many different ways throughout this research study. Mark Whelden and Martin von Randow helped me with the SNZ datalab issues and related SAS inquiries.

Many people provided me with instructive feedbacks, encouragement or simulating thoughts and discussions during or after seminars and meetings. Thanks are due to Alastair Scott, Alan Lee, Gordon Winder, Ward Friesen, and George Perry.

I would also like to thank previous and current COMPASS Research Centre associates, collaborators, and colleagues. These include Daniel Patrick, Katharina Bauer, Gerry Cotterell, Mervyl McPherson, Janet Pearson, and Rocio Gomez.

Special thanks goes to my postgraduate and undergraduate teachers, advisors and mentors, Professors Monty Newborn, Nelly Khouzam, and Dimitri Vouliouris who always believed in me and did not hesitate to support me. The self-imposed desire to live up to the trust, hopes, and expectations of others is indeed a powerful incentive.

I also extend a special word of thanks to Steve MacDonell and Andy Connor for their insights and generosity.

I must also acknowledge the Royal Society of New Zealand's Marsden Fund for their research funding of this project.

Finally, I would like to dedicate this thesis to the soul of my mother and to my wonderful wife Penney and my sweet and brilliant daughter Aryana, for their support, love, sacrifice, and encouragement which has enabled me to achieve this academic accomplishment.

Contents

List of Figures

List of Tables

THE UNIVERSITY OF AUCKLAND
SCHOOL OF GRADUATE STUDIES

Co-Authorship Form

Graduate Centre
ClockTower – East Wing
22 Princes Street, Auckland
Phone: +64 9 373 7599 ext 81321
Fax: +64 9 373 7610
Email: postgraduate@auckland.ac.nz
www.postgrad.auckland.ac.nz

This form is to accompany the submission of any PhD that contains research reported in published or unpublished co-authored work. **Please include one copy of this form for each co-authored work.** Completed forms should be included in all copies of your thesis submitted for examination and library deposit (including digital deposit), following your thesis Abstract.

Please indicate the chapter/section/pages of this thesis that are extracted from a co-authored work and give the title and publication details or details of submission of the co-authored work.

Table 4.1 in section 4.2 is extracted from the following joint publication:

Mahdavi, B., O'Sullivan, D., & Davis, P. (2007). An Agent-based Microsimulation Framework for Investigating Residential Segregation Using Census Data. In L. Oxley & D. Kulasiri (Eds.), MODSIM07 International Congress on Modelling and Simulation (pp. 365-371).

Nature of contribution by PhD candidate	Lead author
Extent of contribution by PhD candidate (%)	100

CO-AUTHORS

Name	Nature of Contribution
David O'Sullivan	Advisory
Peter Davis	Advisory

Certification by Co-Authors

The undersigned hereby certify that:

- the above statement correctly reflects the nature and extent of the PhD candidate's contribution to this work, and the nature of the contribution of each of the co-authors; and
- in cases where the PhD candidate was the lead author of the work that the candidate wrote the text.

Name	Signature	Date
David O'Sullivan	David O'Sullivan	2/12/11
Peter Davis	peter davis	6/12/2011

Chapter 1

1. Introduction

"Theories should be as simple as possible but not simpler" – (Albert Einstein, physicist theoretician)

1.1 Overview

Residential choice influences many aspects of our lives. It affects our access to centres of education, healthcare and business. It determines the composition of our urban neighbourhoods, thereby impacting our social network and social structures. It determines the spatial separation and concentration of groups within the broader population, thereby impacting social cohesion. This can manifest in the form of spatial divisions: when the observed housing patterns are divided along ethnic, cultural, religious, or class lines in urban areas; that is, spatial separation/ segregation or simply residential segregation.

Residential choice is influenced by many factors. Some of them are at a 'micro' level, for instance – personal preferences – as implied by the word 'choice'. Some factors are considered 'macro', for instance, institutional forces such as governmental policies in regards to immigration quotas and anti-discrimination laws. The word choice may even be considered deceptive, as it does not always imply a total 'freedom' of options. For instance, with an explicit systematic discrimination in force, any choice implies 'coercion'.

From a historical point of view, residential segregation is not a new phenomenon. However, the present period of globalization has increased demand for skilled and professional workers, and international student migration and freedom of movement have endowed it with a new dimension. For these reasons today's urban centres in many countries around the world have undergone significant changes in the ethnic, racial and cultural makeup of their respective populations. These rapid demographic changes present new challenges to political and social institutions. The quality of urban life is to a significant degree determined by how successful these institutions are in integrating racial and cultural minorities into the political, economic, and social mainstream. The extent of spatial integration of ethnic minorities is an important indicator of the level of social, political, and economic incorporation that these groups can achieve in society.

Increased awareness of the problems of residential segregation combined with demand for knowledge about its causes and effects and for solution to problems associated with it has led scholars from different disciplines, encouraged by policy makers, to investigate different facets of this phenomenon. As a result of this relatively new systematic research, a reasonable body of knowledge on residential segregation has emerged over the years. Different theories have been proposed to explain the phenomenon of residential segregation and to suggest ways to alleviate its manifestations. Because of the inherent multidimensional and multidisciplinary nature of the phenomenon, multifaceted approaches and explanations are needed.

The study of segregation using dynamic models is a relatively more recent endeavour and promises new perspectives on this issue. Models are excellent methods to represent, investigate and understand a complex system. Models are simpler versions of the real world and range across a spectrum of levels of abstraction. They are useful tools in helping us to learn and better understand a complex system ('exploration'), and when possible, can provide qualitative ('behaviour') and quantitative ('future state') perspective on it ('prediction'). They are also sometimes used for testing, refining and developing theories ('prescription') – a practical instrument for discovery and formalization. In such a way, they can be instruments to theorize about some of the most complex and multidisciplinary systems in a simpler way. In fact, "the term 'model' is sometimes used [loosely] as synonym for 'theory' " (Kaplan, 1998, p. 264).

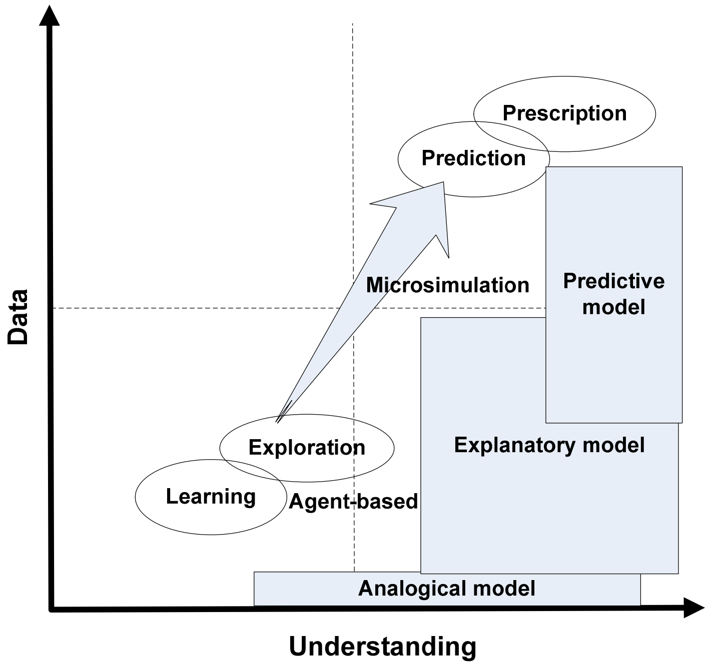


Figure 1.1 Modelling Aims, Scope and Deployment[1]

Within this broad range of prospects and expectations, an attempt is made in figure 1.1 to illustrate these notions and to show how (simulation) models may be used in analogical, explanatory and predictive investigations.

As our aims in using models may vary from 'general' to more 'precise', and from 'complex' to 'simple' and more 'accurate', one of the main challenges in modelling is to keep the balance between 'generality' and 'precision', and between 'complexity', and both 'accuracy' and 'simplicity'.

This work seeks to investigate a number of theories and hypotheses on residential segregation by building a dynamic model which fashions a simplified version of a real urban area, but which is sophisticated enough to serve, in turn – as a medium – not

[1] This figure in its basic format has been first seen in Dr. George Perry's slides in the 'GIS and Modelling' course at the University of Auckland, which has been since modified and elaborated in its current arrangement.

only for improving our understanding, but also as a vehicle for developing new ideas and theories.

1.1 Research Components and Their Relationships

An overview of research components and their interconnected relationships is depicted in figure 1.2. These components can be found in different sections and chapters throughout this thesis. Theory can be derived from observation of real world phenomenon (for example, by using 'exploratory spatial data analysis' (ESDA) (Batty & Xie, 1994b, p. 452)) and aims to describe it. Theory can be expressed in the model; we can inform the model (as an 'incarnation' of theory) by using empirical data (collected from the observation of real world phenomenon, here residential segregation). Research objectives and research questions (constructed to address research objectives) are also derived from theory and observations of the real world phenomenon and prepare the ground for insightful testing of 'specific hypotheses' designed to tackle specified research questions.

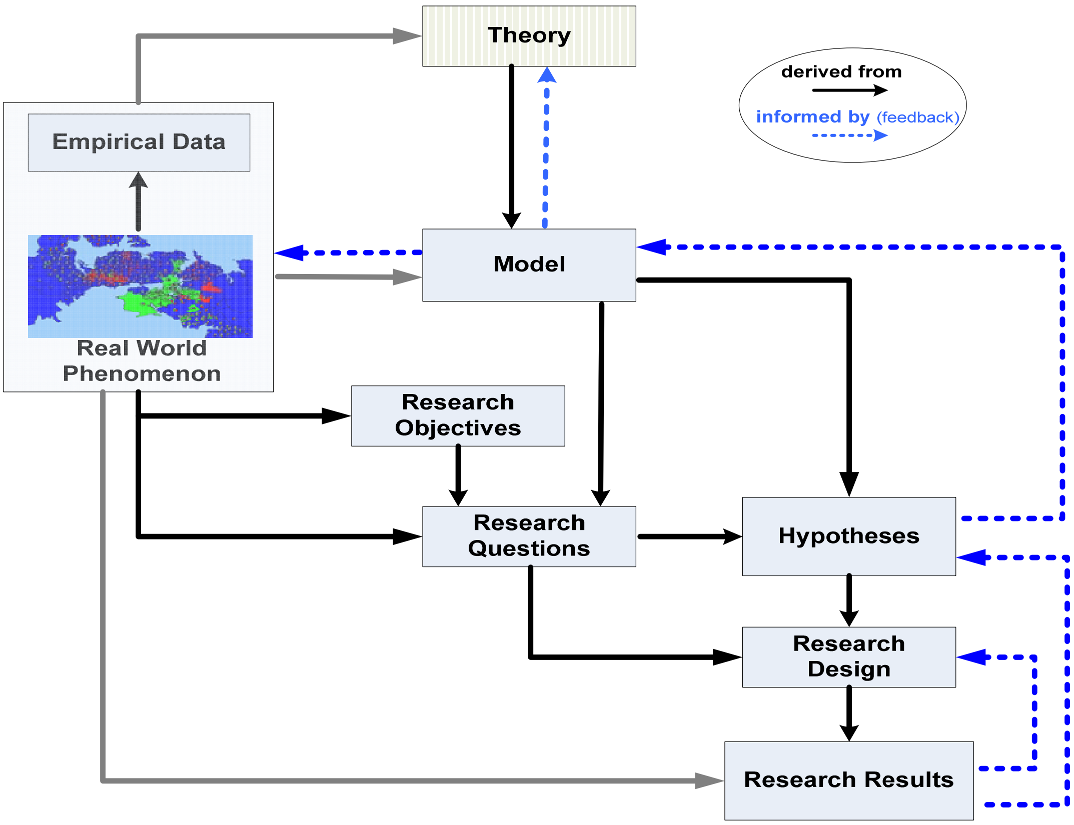


Figure 1.2 Overall Research Components and Their Relationships

The results obtained after careful analysis of model outputs can in turn inform and reshape the research design, hypotheses and eventually the model itself. This is how theories are operationalized, informed, tested/ validated and (new ones) potentially built. It is also how the knowledge (about the real world phenomenon under study) is built using modelling: through families of simulated experiments. These research components are presented in the remainder of the thesis.

1.2 Motivations

While "our knowledge of the formation of ethnically segregated neighbourhoods is still incomplete" (Aydinonat, 2007), racial and cultural minorities have become increasingly visible in large urban areas around the globe, and Auckland in New Zealand is no exception. Auckland is the dominant port of entry for immigrants coming to settle in New Zealand. For quite a long time cosmopolitan Auckland has been considered the largest Polynesian city in the world (Friesen, 2000). In 2006 about 37% of the population of the Auckland region had been born overseas. (i.e. over one-third of the Auckland region population). This overseas born population constitutes more than half of the overall overseas born population of New Zealand.

Recent changes in the demographic landscape of Auckland have been more dramatic. For example, the 2006 Census data by Statistics New Zealand (2006a, 2006b, 2006c, 2007a) show that Pacific peoples had the second-largest increase (14.7%). While they represent 6.9% of the total New Zealand population 67% live in the Auckland region. Moreover, between 2001 and 2006, the Maori ethnic group increased by 7.4%, representing 14.6% of the entire New Zealand population. Of these about 24% live in the Auckland region. Most dramatically, the Asian ethnic group had the fastest growth between 2001 and 2006 in New Zealand's population: an increase of about 50% in five years. In 2006, they constitute 9.2% of the entire population of New Zealand. Of these two-thirds live in the Auckland region.

This synopsis gives some idea of the rapidity of demographic change in New Zealand and in particular the Auckland region and offers a motivation for investigating residential pattern change (and the social structure in broader terms) in

New Zealand's largest city (metropolis). The motivation is particularly reinforced as insight into patterns of segregation in New Zealand has received relatively modest attention (Johnston et al., 2004). Although some research into patterns of segregation in the country has been conducted in recent years none used modelling or simulation.

An important aspect in the analysis of urban demographic change and residential segregation is the ability to relate that change to other aspects of complex urban systems. This can be achieved through a dynamic model. For this reason many studies on residential dynamics have been carried out using modelling simulation. Many of them are based on Schelling's original work. However, the majority of these models are vulnerable because of lack of empirical grounding (i.e. they use synthetic data) and/ or realistic sequencing of spatial units that influence the choice of relocation to a proximate neighbourhood.

More importantly, these models are considered overly simplistic. In particular, their cell-based environments populated by binary ethnicity choice of household agents can barely represent a village. In reality, this 'small-city' syndrome makes these models exceedingly vulnerable to criticism (see, for example, Goering, 2006).

1.3 Research Hypotheses

A research study based on modelling that seeks to provide more insightful explanations of residential segregation should take a more sophisticated methodological approach than those previously taken and should also be empirically validated. Taking advantage of this opportunity for a distinctive research study of residential segregation this research applies the following two general hypotheses in its objectives and modelling:

- Using a micro/ individual-based modelling approach (e.g. ABM, MSM or hybrid) to build an urban geosimulation model of a realistic scaled city would be more apposite for investigation into residential segregation phenomenon. An eventual evolution of such a model towards a complete 'artificial city/ society' version (whilst using sensible computational resources) can be envisaged.

- Using empirical data (particularly those that are publicly available, such as census data), such a model can be informed, calibrated and validated. This would be essential for a model that aspires to be used as a decision support and policy-informing tool.

1.4 Research Aims and Objectives

As discussed in the previous section, the relative lack of research on residential segregation in New Zealand, and the need for and absence of an innovative simulation modelling approach in particular provide an opportunity for a novel investigation. This research was part of *Modelling Social Change* (MoSC) project funded by the Marsden Fund of the Royal Society of New Zealand. The central theme of the MoSC research proposal was to investigate the stratification of the social structure of New Zealand. This stratification was that reflected in the distribution of matching socio economic and ethnic choices of co-habitation partner across households by application of simulation techniques and by using the census as a faithful data source for 'societal coverage'. The residential dimension was added later (in 2006) to the MoSC project in addition to the co-habitation aspect, and similarly it aimed to quantify, in part, the degree of social integration (or segregation) expressed in the location choices (or the cohabitation choices, in the case of the cohabitation facet). The five-yearly data collections of the census provided a strong time-series element that could be used as the basis for modelling social change in New Zealand.

As the residential part of the MoSC project, the overall aim of this study is to develop and validate a dynamic model to explore relationships of interest in the dynamics of urban neighbourhood change in New Zealand. More specifically, this thesis aims to achieve the following objectives:

i. The establishment of census datasets as potential 'test-beds'[2] for future modelling-based research which can address theoretical issues in urban geography and sociology.

ii. The examination of changes in New Zealand's social structure from a residential perspective. This includes the measurement and description/ explanation of past, present, and possible future patterns of residential location by ethnicity, by focusing on change in Auckland urban settings.

iii. The operationalisation of micro and macro dynamics

- the development of a simulation-based model as an approach to understanding the dynamics of urban neighbourhood change
- the development of an approach to the transparent verification of simulation models of residential segregation using aggregate census data
- examination of the influences of some of the underlying drivers of residential location choice in the Auckland region through the period 1991 to 2006
- investigation of linkages between the social micro (household level) and macro levels (social structure expressed through demographic patterns in geographical space), and to operationalise these linkages in the simulation model.

1.5 Research Approach

The study of relationships in residential segregation has a long history in sociology, economics and geography. This study, on the other hand – reflecting the bias of an

[2] This refers to the potential usage of the census data as a 'platform' that can virtually map a society (and its smaller sub-components) as completely and as faithfully as any other data sources into the model. In other word, this project can be seen as a demonstration for possibilities and advantages of developing and deploying modelling-based simulation frameworks that are empirically informed by the census data.

urban geographer – is concerned with modelling and investigating how some of the multidimensional forces of residential segregation are related to adjustments by Auckland population in how they occupy and relocate within their urban environment. This study, therefore, is concerned with the spatial manifestation of a social/ behavioural process. The research will particularly focus on describing the changes in residential patterns (e.g. residential segregation or integration) that took place in the period 1991-2006, whilst describing and demonstrating the social and behavioural processes (through individual preferences) that led to these changes.

As stated in the previous section, this research has two central objectives. The first is to examine changes in New Zealand's social structure from a residential perspective (based on ethnicity). The second is to operationalise the link and dynamic between micro level (e.g. individual or household's[3] preferences), macro level elements and forces (e.g. number of groups, their proportions, sizes) and higher level structures generated by micro level elements' behaviour. The latter feeds back to the lower level, influencing the micro elements to behave differently (i.e. a bottom-up – top-down synergy). Finally, the overall research objective is to demonstrate that by using census data, an empirically informed model can be used as a 'test-bed' (framework/ platform), where future modelling-based research can address theoretical questions and issues in urban geography and sociology. Therefore, the key product of this research (i.e. the proposed framework along with its methodological approach) should be able to demonstrate its potential. This has possible implications in social and urban theory, and therefore, will be of interest in the fields of geography, sociology, urban planning, demography, economics, as well as public policy.

The first part of the central objective will be addressed by descriptive statistical data analysis. By describing and examining the changes in the (census) data inferences can be made about whether the residential patterns in New Zealand have become more highly segregated (stratified), or integrated, or remained unchanged over the period 1991 to 2006. To tackle the second objective, a simulation-based framework will be

[3] The terms 'individual' and 'household' (even 'agent') are used interchangeably throughout this thesis. If individual is mentioned alone, it can equally be read as 'individual household' or 'individual agent'.

used to model the changing residential patterns of the Greater Auckland urban area. This part of research approach also comprises calibration of the customized scenarios, and an analysis of the outcomes including prediction of future residential segregation change in Auckland region. With a special focus on operationalising the linkages between the micro-level and macro-level, a direct connection can for instance be drawn on how changes in the behaviour of the agents (micro-level), can impact urban structure as a whole (emergent macro-level outcomes). The overall objective will be demonstrated by simulating specific scenarios where the generated results is similar to the residential distributions (patterns) observed in empirical data (census), either quantitatively (aggregate statistics/ measures), qualitatively (spatial/ visual distribution patterns), or both.

1.6 Specific Research Questions

Within the research aims, objectives and approach presented and explained in the previous sections, this study will address the following primary research questions:

- Can census data be used to inform and validate the model (in particular, a geographically referenced model) and therefore be used as a test-bed for similar future modelling-based research?

- What are the most noticeable/ remarkable changes of patterns of residential location by ethnicity in the Auckland urban area during the past four censuses (1991-2006)?

- Is the model capable of reproducing some of these observed changes and validate them without detailed knowledge and presence of all elements in the residential choice decision-making process?

Furthermore, secondary research questions that will be explored in this thesis include the followings:

- What are the advantages and limitations of using census data for the purpose of informing and validation of a residential segregation simulation model and even broader 'social simulation' models? To what extent they can census data be used in a 'geographically referenced social simulation' modelling context?

- What are the distinctive methodological characteristics and strengths of a residential segregation model capable of simulating a realistic large scale urban area? Can these be materialized with small or medium computational resources?
- What is the potential of the current model (or its future extensions) to be used as a decision support and policy-informing tool?

1.7 Organisation of the Thesis

The dissertation begins in Chapter 2 with a definition of residential segregation and neighbourhood. The importance of residential segregation and its dynamics as a complex system is explained. The potential factors and forces involved in the formation of residential segregation are discussed, along with those which will be used in the simulation modelling based investigation in this thesis. This chapter also reviews dynamic modelling of residential segregation and in particular the Schelling based models.

The research design is presented in Chapter 3 which begins with an explanation of the hybrid modelling approach vision. The unit of analysis as well as the multidimensional capability of the model to explore geographical space at different scales are then clarified. This is then followed by theorizing a conceptual framework on which the final model is built and explored. The chapter will also expand on a detailed schematized version of the research model. The final section explains how the time dynamic of the system will be operationalized and how it compares with the time scales of other models.

Chapter 4 unpacks the design of the final version of the simulation model framework by explaining its key design concepts, functionalities, and magnitudes, along with an overview of the scheduled processes. The model is also compared with other Schelling-type models.

Chapter 5 starts with a detailed description of the measures implemented in the model for estimating different dimensions of segregation. The rest of this chapter focuses on the demonstration of the model's conformity with two of the most well-known related models, namely the Schelling and Fossett models. However, no-exhaustive exploration of the parameter space is undertaken using these theoretical configurations and experiments.

Chapter 6 begins with the examination of empirical data. In part, it explains the level of error ('noise') in the data and the measures taken (when possible) to attenuate it. The delimitation of data will be also discussed. The analysis of data and segregation measures will reveal the changes in the ethnic landscape of the Auckland region through the census periods of 1991-2006. Some of the measures are used as the basis for a series of quantitative (aggregate statistics/ measures) and qualitative (spatial patterns) 'benchmarks' to which the outcomes of scenario-based simulations (in chapter 7) will be 'qualitatively' compared for validation purpose.

Chapter 7 demonstrates the capability of the model to generate interesting and plausible patterns of changes in residential distribution over time. These patterns should reasonably correspond to the empirical census-based benchmark patterns. To do this, a pattern-oriented calibration (POC) approach is proposed and used with specific attention to the parameterization of two important minority groups in the Auckland region (namely Asian and Pacific peoples). Using a series of scenarios, including policy-oriented types, the potentials of the model as a decision-informing tool is demonstrated.

Chapter 8 concludes by summarising the materials presented in previous chapters. The summary provides the rationale for the modelling approach along with its unique elements and the implications of its theoretical, methodological and public policy viewpoints. The chapter closes with recommendations for future research.

Chapter 2

2. Background

2.1 Introduction

In order to build a model that can capture as accurately as possible the fundamental elements of the real world of residential segregation dynamics, an understanding of residential segregation and its modelling is required. Using the literature, this chapter begins (section 2.2) with a description of residential segregation, its importance, complexity and the causes of it. Also explained are the factors used in the model developed in this thesis to explore the dynamics of residential segregation. Section 2.3 provides a background into the modelling of residential segregation dynamics, starting with an explanation of Schelling's groundwork. Most agent-based extensions of Schelling model are still vulnerable to the severe limitations of artificial (grid-based) environments and lack of empirical data. By explaining microsimulation and geosimulation, this chapter introduces the intention of this thesis to employ state-of-the-art modelling approaches along with a hybrid mechanism to overcome some of the limitations of Schelling based models.

2.2 Residential Segregation

The International Encyclopaedia of Social Sciences (Sills, 1997, p. 144) describes segregation as "an institutionalized form of social distance expressed in physical separation". This definition makes no reference to coercion (compulsory) or choice (voluntary) factors. In practice, different researchers advocate different definitions of segregation (Massey & Denton, 1988, p. 282). However, segregation can be conceived of as an aspect of social organisation, as it "defines the boundaries between groups, locates the groups in the hierarchy and regulates their interactions" (Sills, 1997, p. 144). The International Encyclopaedia of Social Sciences definition also does not expose the complexity of segregation phenomenon in relation the dynamics of individual choice (Schelling, 1987, p. 140). After all, the ultimate consequences of the process of segregation although aggregate, "the decisions are exceedingly individual" (Schelling, 1987, p. 145). The aggregate result of these individual decision-makers eventually manifests in the form of change in the composition and spatial distribution of the population among neighbourhoods.

Neighbourhood

There is also a lack of consensus among social scientists on how to define a neighbourhood (Logan & Zhang, 2004, p. 113). In fact, different terms such as 'quarter', 'district', 'enclave' or even 'ghetto' are often used interchangeably to refer to the same notion. For a lay person, the differences and variations among neighbourhoods are often characterized by terms for example such as 'poor', 'rich', 'safe', 'unsafe', 'rough', 'white', 'black', 'Asian', 'Hispanic'.

In practice, people often define their neighbourhood by reference to their own location. But Schelling (1987, p. 155) observed a common definition of the neighbourhood and its boundaries: "A person is either inside it or outside".

Table 2.1 attempts to capture types of neighbourhood classified by their most prominent defining factor(s). These defining characteristics are usually ethnicity (race), religion, profession, sexual orientation, age, income, wealth/ class, business/ industry type and distance from the heart of a city.

This thesis focuses exclusively on the ethnicity factor with reference to the residential type of segregation (i.e. residential segregation). Residential segregation is seen as "the degree to which two or more groups live separately from one another, in different parts of the urban environment" (Massey & Denton, 1988, p. 282). These different parts of the urban environment can be seen as different neighbourhoods. In the realistic environment envisaged in this thesis, the unit of study (which corresponds to administrative boundaries of the city) is called 'area unit' (AU). A neighbourhood can also be seen as an area unit and all its adjacent units. The administrative boundaries of the city are explained in chapter 6.

Table 2.1 Neighbourhoods Types Examples

Grouping name	Defined mostly by	Names and References Examples
Ethnic sectoring district	Ethnicity	- Immigrant/ethnic enclaves (Logan et al., 2002) - Immigrant neighbourhoods (Davies & Fagan, 2002)
Muslim, Christian, Jewish, Hindu, … area	Religion	- Religious district (Wedam, 2003)
'Latin quarter'	Profession	- Artist[ic] neighbourhood/enclave/district (Hudson, 1988) - Student Ghettos (Smith, 2005, p. 73)
Gay district	Sexual orientation	- Gay (/Lesbian) districts (Rosenblum, 1995) - Gay neighbourhood (Knopp, 1997)
Retired people neighbourhood	Age	- Gentrification areas (e.g. young professionals) (Kotze & Van Der Merwe, 2000, p. 40)
Socio-economic zone	Income	- Socio-economic (SES) zoning (Laurence, 1994, p. 18) - Working-class neighbourhood (Tomes, 1978, p. 330) - Middle-class enclave (Hudson, 1988) - Low-income ghettos (Gibson, 1968, p. 1188) - Low/high income neighbourhoods (Gross, 2005)
Prestige neighbourhood/ Gated block	Class/ Wealth	- High-prestige neighbourhood (Fleming, 1989, p. 480) - Old-money neighbourhood (Mickelson, 2006, p. 170) - New elite neighbourhood (Nathans, 2003, p. 7) - New-money neighbourhood (Milano, 2008, p. 43)
Commercial district	Business	- 'Old' industrial district (Herrigel, 2000, p. 296) - Central business district (Burgess, 2002)
Suburbia	Location/ Distance from city	- New housing in suburban areas; Residential communities lying immediately outside a city (Hayden, 2003) - Inner-ring suburbs (Lee & Leigh, 2005)

2.2.1 The Importance of Residential Segregation

The impact of (residential) segregation on health, income, education and other social and economical issues has been studied and reported. Residential segregation often causes or reinforces inequalities in these important spheres of society. Therefore, it is not surprising to see a policy related interest in this topic. By understanding the role and impact of different factors implicated in the residential segregation phenomenon, there is a prospect of reducing or eliminating inequalities in society. Some of policy-oriented scenarios in chapter 7 will demonstrate the potential of this model to inform the decisions that can help in attenuating the impact of segregation in urban areas, for example, by planning (budgeting) the construction of new public service centres long before the anticipated new urban realities emerge.

2.2.2 Complexity and Determinants of Residential Segregation

The causes of residential segregation are not well understood (Bruch & Mare, 2006, p. 672). Overall, the urban spatial structure is inherently multidimensional and residential segregation does not stem from a single process, but from a complex interplay of many different social and economic processes (Massey & Denton, 1988). In other words, residential segregation is the result of many factors.

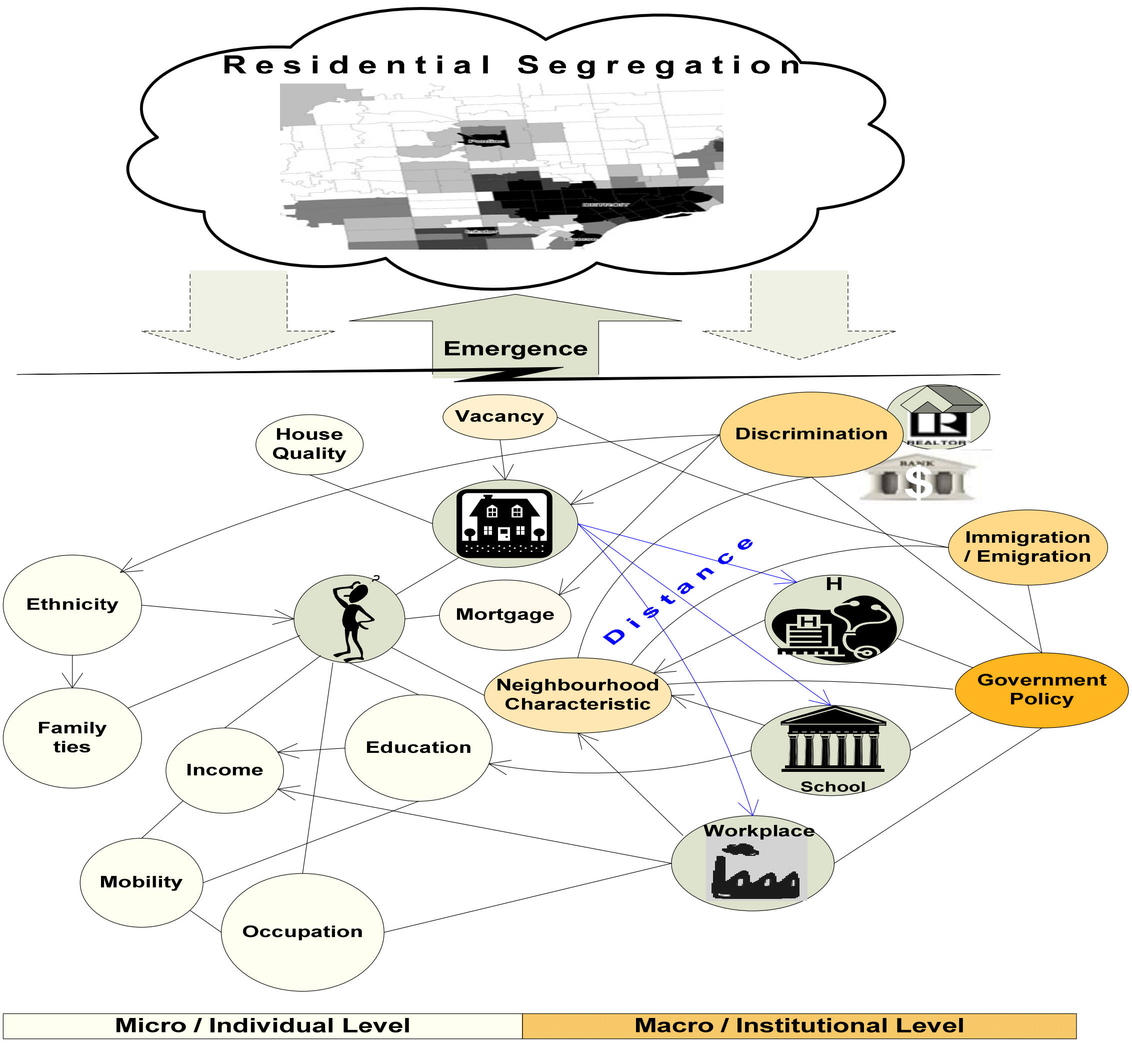


Figure 2.1 Complex System Dynamics of Residential Choice

Figure 2.1 illustrates the complexity of the residential choice dynamics (Charles, 2003, see) and its overall manifestation as spatial separation (i.e. segregation). This 'network' exhibits the characteristics of a 'complex system'; that is, a product of many parts that interact, and the aggregate phenomenon (behaviour, pattern, or structure) that emerges as a result, cannot easily be explained in terms of interactions between these individual constituent elements. Moreover, this system "cannot be simplified in the conventional way by reduction or aggregation" (Batty & Torrens, 2005, p. 747). In this sense, the scientific study of the residential segregation phenomenon falls under the construct of 'complexity science'. Interactions between parts and elements of this dynamic system are the source of both complex dynamics and emergence.

2.2.3 Emergence and Self-organisation of Residential Segregation

The notion of emergence has several meanings (see Dessalles & Phan, 2005, for example). It is also possible to discern different kinds of emergence (e.g. weak or strong). In this thesis, emergence is seen as a coherent developing behaviour, pattern, or structure that manifests at the macro-level as a result of interaction between the parts at the micro-level. The properties of these macro-level behaviours, patterns, and structures are not necessary contained in the property of the parts.

In this sense, although knowledge of individual attributes and interaction rules is not usually sufficient to predict the emerging behaviour of the whole system, a dynamic model will permit a better understanding of the properties of interaction structures, and therefore will make it easier to gain a better understanding of the emergence process. From this viewpoint, designating the residential segregation manifestation as an emergent phenomenon does not mean that it would be impossible to explain or model it (Dessalles & Phan, 2005).

Self-Organisation

In choosing residential location, individuals (households) do not have a global notion of segregationist structures. In other words, they do not choose to live or not to live in a segregated urban area (this is sometime regarded as 'weak emergence' condition). They are however simply conditioned by different optional (e.g. distance to work, proximity to family member) and no-optional (e.g. race/ ethnicity) factors in selecting their residential location. The dynamic of these local interactions; that is, these individual micro choices, each shaped by different (and often insubstantial) conditions and criteria for choosing a neighbourhood, will eventually induce a 'chain reaction'

(Schelling, 1987, p. 150) that engenders the formation of spatial homogeneous patterns. In other words, this 'unravelling process' occurs by micro/ individual behaviours which are organised autonomously (or at least semi-autonomously).

The notion of self-organisation is naturally present in a system where individuals make decision and act autonomously. That is presumably how 'self-forming neighbourhoods' (Schelling, 1987, p. 147) occur. In a purely lateral meaning of the term, self-organisation can be described as a system that can organise itself without any external control or manipulation. It is logical to think that a sophisticated residential segregation model should normally allow decision-maker entities (e.g. agents) to be totally autonomous, learn, and adapt themselves to their environment (which is also constantly changing). In this case, it would be more appropriate to describe such a model as a 'complex adaptive system' (CAS).

However, like many other concepts (e.g. emergence, complexity science, and the like), there is a degree of confusion (or at least lack of consensus) about the meaning of self-organisation. For many, the above lateral description is unsatisfactory. For instance, Prigogine's theory of self-organisation emphasises the process of dissipation (and self-organisation of dissipative structures) whilst Portugali (2000, p. 51) emphasises the 'creative' nature of such a system, as well as its interconnection between parts and components in a 'nonlinear' fashion.

This thesis takes a computer science view of self-organisation by considering the primary areas of its application, which are 'learning' and 'adaptation' (Wolf & Holvoet, 1998). In the final version of the model presented in this thesis, agents do not have an aptitude for learning and adaptability (in its pure meaning). In this sense, they are considered 'proto-agents' (Howe et al., 2006). Therefore, although in the loose and lateral meaning of the terminology, the behaviours of agents (in the model presented in the thesis) can be described as autonomous (i.e. self-organized). But, because of lack of learning and adaptive capability, this thesis does not claim that this model (in its current version) should be recognized as a system that exhibits (complete) self-organisational behaviour.

2.2.4 Explanatory Factors

Figure 2.1 also shows examples of the factors that potentially play a role in the dynamics of residential location choice and hence in the occurrence of segregation. Some of these factors are at an individual (micro) level, including demographic or socioeconomic factors such ethnicity, education, and income. Ethnicity, for example, is a factor that cannot normally be changed, unlike the family-ties factor which is an individual preference. On the other hand, ethnic (and cultural) background can potentially play an influential role in deciding how important it would be for an individual to live or not live near his or her relatives. In Li's (2006, p. 11) view, these may be seen as "internal pull factors", which can also include 'ethnic solidarity' and 'mutual interest'.

On the other side of the spectrum, macro and communal forces such as government (or municipal) policies or different forms of institutional discrimination play an important role in residential segregation formation. From Li's (2006, p. 11) viewpoint, these may be seen as "external push factors by the host society" (e.g. discrimination). In fact, the labour market (e.g. type and level of skills demanded) and access to social benefit systems (e.g. housing subsidies or social housing programs) , both of which are influenced by government policies, impact significantly on the position of people in the social ladder (i.e. their socio-economic position) and hence on where they can afford to live. Choices may also be influenced by attributes (such as distance) related to schools or places of work.

Although it will not be practical for the model presented in this thesis to include all these factors, a useful model with a policy-informing ambition should preferably allow a dynamic investigation of many of the key factors. Ideally investigation of residential segregation dynamics that lie outside the limitations of the scope and space imposed on a thesis should gradually be included. In any case, the flexibility of the model and its capacity to be extended to allow inclusion of additional factors does not mean that this should necessarily be done in a simultaneous fashion. Better understanding of residential segregation requires numerous questions and hypotheses to be tested and analysed. A systematic approach should be used and the selection and value range of factors should be determined by research questions.

Although it may not be necessary in contemporary society for the model to include discriminatory government policies based on racism and ideology (such as the one that existed in apartheid South Africa), other form of subtle institutional discrimination still exist. For instance, Sugrue (1996) and Ferguson (2008, pp. 249-250) assert that "segregation [in the city of Detroit] … was not accidental, but a direct consequence of government policy".

Furthermore, other institutionalized forms of discrimination also exist. One is 'redlining' (Pol et al., 1982; Shlay, 1988; Sugrue, 1996; Taggart & Smith, 1981), a practice by a financial institution to systematically give negative credit-rating and refuse to grant credit/ loans to specific urban neighbourhoods (areas). Redlining on the basis of racial discrimination has not ceased completely (see Lefebvre, 2004, for example).

There is also discrimination by means of 'racial steering' (Pearce, 1979). This is a practice in which real estate agents guide prospective client towards or away from certain neighbourhoods based on their socioeconomic characteristics, and especially their race (ethnicity). In fact, the design of the model was influenced by an initial intention to include ethnic steering and 'managerialism'[4] behaviours (in addition to micro/ individual 'behaviourism') such as those by realtor agents. It was later decided not to operationalize these features in order to set the level of complexity of the model to a manageable level. However, the realtor object is still present in the design of the final model presented in this thesis. In this context, the realtor object should be seen as a model mechanism and not as a representation of a real world real estate professional, as was initially intended. Notwithstanding the fact that the scenario has not been explored in this thesis, the realtor object in the model's current form (as a part of the model's 'placement' mechanism) can be still used as a single agency playing a central discriminatory role against individual residence searchers. On the other hand, the

[4] *Managerialism* seeks to explain the unequal distribution of life chances by reference to those agencies whose responsibility and power is to undertake that distribution (Cater & Jones, 1989, p. 51).

presence of the blue-print (place-holder) realtor object in the design should both emphasise and facilitate the inclusion of this feature in the future extension of this model.

2.2.5 Operationalizing Neighbourhood Characteristics

Numerous characteristics can be attributed to a neighbourhood. However, this thesis focuses mainly on the characteristic of its ethnic composition. This is operationalized through an individual/ household agent's preference for living in a specific neighbourhood (spatial unit) whose co-ethnic composition percentage is equal or above its threshold. That is, after an individual (household) agent in new house-searching mode verifies the composition percentage of an alike ethnic group in a candidate location against her own threshold preference. If the co-ethnic composition percentage is lower than her threshold, she will not settle and continue her search to find another suitable location to relocate.

Often terms such as 'in-group', 'co-ethnic', 'co-racial', or 'own-race' are used to refer to this preference (see, for example, Bobo & Zubrinsky, 1996; Clark, 1992; Fossett & Waren, 2005; Wilson & Hammer, 2001). In this thesis the first two terms are applied (i.e. in-group and co-ethnic).

Moreover, like Schelling (1987) and Clark (1991, p. 4), this thesis also uses the terms 'preference' and 'tolerance' (even 'choice') interchangeably when referring to neighbourhood composition preference, or "preferences for different racial and ethnic neighbourhood composition" (Clark, 1991, p. 1).

2.2.6 Thesis Focus on Selective Factors

Before concluding this section, it is important to mention that researchers have studied and measured the role and influence of different factors (including some of those shown in figure 2.1), using traditional techniques, methods and tools (i.e. empirical and static[5] approaches). The majority of these studies deal with hypotheses related to socioeconomic factors and inequalities. However, the inconclusive and contradictory nature of research results suggests that socioeconomic factors account for only a small proportion of segregation levels.

Notwithstanding that choice or preference based on neighbourhood characteristics factor (e.g. proportions of people with the same ethnicity) alone also fall short in explaining residential segregation, the findings of several studies demonstrate how changing preference criteria can (often radically) alter the outcome. There is consequently little doubt about its central role in causing or reinforcing residential segregation (see, for example, Bruch & Mare, 2009; Clark, 1991, 2006; Fossett, 2006a).

In debates about the relative role of these forces, the consensus is that patterns of separation have a multifaceted explanation; no one factor explains those patterns (Clark, 1986, 1987, 1991; Galster, 1988). Urban problems, including those relating to race and housing are best understood as the result of a complex inter-relationship between many processes (Sarre et al., 1989).

The scenarios presented in this thesis will focus particularly on the role of the factors ethnicity, freedom of movement (which can also be seen as a proxy for income), mobility, immigration/ growth, vacancy and neighbourhood characteristics (through in-group contact preference).

2.3 Modelling Residential Segregation

Throughout this thesis, the terms simulation and modelling are often used interchangeably (sometimes even together). Simulation is considered to be a particular type of modelling (in many cases simulation is actually the implementation of a model), which is especially suitable for investigating social processes. Like other forms of modelling, simulations are considered as a third symbol system (beside natural and mathematics ones) (Gilbert & Troitzsch, 1999). Simulations are used for exploration,

[5] In contrast to a dynamic approach which involves simulation modeling, a static approach refers to those techniques and tools that do not employ dynamic simulation, or are formula-driven static 'simulation' modeling that are usually performed using spreadsheet programs.

prediction (of specific outcomes) and possibly prescription (the testing of assumptions, propositions and theories could result in refining or refuting them or could lead to the development of other propositions and theories).

Researchers have studied and measured the role and influence of different factors on segregation using traditional static techniques, methods and tools. These include empirical studies and statistical analyses of census data. However, there are some limitations to the static approach. For instance, in examining residential preference hypothesis, such studies cannot explain how residential preferences at the individual level are translated to the overall residential pattern at the aggregate level. Moreover, because of the multidimensional nature of the residential segregation phenomenon, many components of this complex system are interconnected in a nonlinear fashion (by a complex network of feedback loops). Only a dynamic model allows the creation of a more sophisticated and realistic representation of the real world. These include the establishment of feedback loops and interconnections, the simultaneous execution of certain factors and forces, and the ability to examine the possible role of preference at the micro individual level, as well as their emerging macro effects.

2.3.1 Schelling Groundwork

Boman and Holm (2004) reported an early model of dynamic spatial simulation at the micro level made by Hagerstrand in 1953 (this had to be executed using 'manual' calculations), and provided a long list of dynamic approaches that have been adopted since the 1960s for the investigation of demography, population and socio-economic issues (see Merz, 1991, for example). But it was the pioneering work of Thomas Schelling (1969, 1971a, 1971b, 1972, 1987) that provided the foundation for one of the most influential dynamic models for investigating residential segregation dynamics[6].

Schelling's work presents influential examples of an abstract study of the interactive dynamics of (discriminatory) individual choices showing that segregation can result from individual preferences. However, he was not the first one who pointed out that

[6] Almost at the same time Sakoda (1971) also proposed a very similar model based on the same principles.

residential segregation may result from individual choices. For instance, Becker (1957) identified potential links between preferences and residential segregation, and other earlier discussions about this hypothesis are also recorded (see Muth, 1969, for example).

Nevertheless, Schelling 'spatial proximity' (also known as 'checkerboard') and 'bounded neighbourhood' (also known as 'neighbourhood tipping') models are considered seminal examples of agent-based models (ABMs)[7] in social science.

Schelling presented his ideas about how "macrobehaviour" effects in a society might be disproportional or different from "micromotives" of its individual residents. Using his models he demonstrated how integrated cities could unravel into segregation due to non-economic forces. This is notwithstanding relatively mild co-ethnic contact preferences of individual agents.

His 'spatial proximity' or 'checkerboard model' consisted of fixed (and equal) numbers of black and white agents who were randomly distributed over squared cells, leaving some of the squares empty (unoccupied). The agents had (similar) preferences in regard to the racial composition of their immediate neighbourhoods and a discontented agent (dissatisfied with its current location) could move in random order to another neighbourhood that satisfied its preference. An agent could become discontented with its location as other agents move into its neighbourhood. This moving process could continue until every agent is satisfied with its current residential location. Schelling found that even a mild individual preference for like-colour agents could result in complete segregation. This simple checkerboard model has since become seminal groundwork for future agent-based simulation for studying residential segregation.

---

[7] This thesis prefers not make any distinction between ABM and cellular automata (CA), particularly when most agent-based Schelling models have been built on cellular grid environments where "cells are agents" (Batty, 2005, p. 1377). Cellular automata often consists of a grid, cell states and a transition rule, determining changes in cell state at each time step. The behavior (state) of each entity (cell) depends on the state of neighboring cells as well as its own current state (O'Sullivan, 2001, p. 687).

Through his other 'bounded-neighbourhood' (Schelling, 1972) (or 'neighbourhood tipping') model, he explained the phenomenon of 'neighbourhood tipping' which occurs when a minority group enters a neighbourhood in numbers sufficient for long established majority residents to decide to evacuate at an accelerating rate. The checkerboard model could be seen as a miniature city with many neighbourhoods where an agent (local) neighbourhood is defined by reference to its own location (like surrounding squares defined in Moore). On the other hand, in the 'neighbourhood tipping' model, Schelling used a single bounded neighbourhood to explain his ideas, where an agent was considered either inside or outside of it. In this environment, the agent is concerned with the composition ratio of the whole neighbourhood, and is not interested in the configuration or composition ratio within the immediate (local) neighbourhood.

The simulation model developed and presented in this thesis consists of multiple bounded neighbourhoods which appear like a checkerboard in the artificial environment. It is compared with the Schelling model characteristics in chapter 4 (section 4.2.5).

2.3.2 Agent-based Model

In an agent-based model, a system (such as residential segregation) can be modelled as a collection of autonomous decision makers and interacting entities called agents (Bonabeau, 2002). However, it seems that there is no consensus on what exactly constitutes an agent (see, for example, Wooldridge & Jennings, 1995, p. 116). Many consider autonomy (the capability to independently make decisions, process information or interact with others) and heterogeneity (being different from others in the same group) as defining characteristics. But other characteristics include mobility (being able to move around the space), pro-activity (being able to pursue a goal) and reactivity (being able to perceive others and the environment and react accordingly).

Many agent-based models designed to investigate residential segregation are simply variations of the Schelling's original checkerboard model, either by modification or extension of its initial settings and conditions. Examples include works by Bruch and

Mare (2006), Laurie and Jaggi (2003), Fossett and Waren (2005), Sander et al. (2000), and Zhang (2004a, 2004b, 2009). Clark (1991) also used survey data to evaluate the underpinning of the Schelling model.

Schelling was careful to recognize the limits of his model and contributions (Fossett & Waren, 2005, p. 1894). For instance, his behavioural assumptions were limited. Preferences for ethnic composition were the only operative preferences (Skvoretz, 2006, p. 182). But subsequent efforts to extend these limits were rather modest too. The Fossett model (2006a), for example, considered by some scholars as one of the "most sophisticated version[s] to date of Schelling's basic ideas", relaxed many of the constraints of the original model" (Skvoretz, 2006, p. 182). But it also suffers from some weaknesses. These were pointed out by some critics, notably by Goering (2006). "Segregation has a plethora of overlapping causes"(Goering, 2006, p. 302), and Fossett's overemphasis of the role of preference (particularly minority preferences[8]) as "single factor explanation[s] [is] likely simplistic" (Goering, p.302, quoting Fossett, 2006a). The problem may not be only in relation to "extravagant or premature claims about what can be learned from" (p. 307) his model, or "one-sided modelling of preferences and social distance" (p. 313), but also on the "limits of the methodology" (p. 307). Fossett's model consisted of about 15,000 people and 5,000 family homes in 112 neighbourhoods (around 50 single-family homes in each neighbourhood) but it is not a "recognizable metropolitan" area. Furthermore, his "small town" has a very "tight housing market" (p. 308) (i.e. few vacancies). In Fossett's "preference-ville" (p. 308) model, it is assumed that quality of houses (e.g. better and more expensive houses) improves with distance from the city centre. This may not be a compelling assumption in many contemporary metropolitan areas (including Auckland). Another notable weak area in the Fossett model is the use of a single and limited measure of segregation (i.e. dissimilarity index). In particular, this measure indicates a higher score for the smaller scale of a geospatial unit.

[8] Later Fossett in his 'defending' paper (2006b) clarifies that he acknowledges the existence of discrimination as an important force in residential systems (p. 292), and states that "research on multi-factor explanations of segregation is desirable" (p. 296).

One of the notable characteristics of Fossett's models is the setting of the racial composition as minority and majority groups, whilst assigning different preferences to each group (although homogenously the same among the same group). Another notable characteristic is his emphasis (based on his simulation model outcomes and findings) on the role of minority group preference in formation of segregation, to the point that it appears he virtually blames the minority group for causing segregation (Fossett, 2006a, p. 200).

In retrospect, the subsequent agent-based modelling of Schelling's original checkerboard model have been so simple that they prompt questions about how well they portray the neighbourhood dynamics of real cities (O'Sullivan, 2009, p. 507). Some of the shortcomings of these models will be addressed in this thesis. Notably, by the use of a more realistic configuration (using aggregate data), representing a population that (roughly) corresponds to the real size of the population who live in the Auckland metropolitan area units, and includes exogenous factors (such as newcomers who immigrate from outside world into the city, realistic vacancy rates (based on district area), and the possibility of relocation to immediate or city-wide neighbourhoods. Moreover, the central role of how choices are made and the question of whether it is the individual choice or preference or other conditions (e.g. immigration, vacancy) which drives the system level change can be explored in this model.

2.3.3 Microsimulation

Microsimulation (MSM) is another way to approach the modelling of individual and household dynamics, and has been particularly used for policy analysis. Merz (1991) categorizes such models into static and dynamic model types and then gives an exhaustive list of major microsimulation models in each category in the US and Germany up to 1990.

Ballas et al. (2005) also provide an account of those studies not explicitly labelled microsimulation, but according to them, "even calculating simple social statistics such as life expectancy is a method that is similar to microsimulation". They quote Birkin et al. (1996) that "microsimulation models developed so far do not take spatial scale into account". They then go on to give an account of more recent endeavours. These include a spatial microsimulation approach to the analysis of the spatial impacts of social policies which identifies which households and geographical areas would be affected the most under different social policies (Ballas & Clarke, 2001). Their 'SimBritain' model is another spatial dynamic microsimulation endeavour which simulates the population of Britain under different scenarios. Other similar microsimulation models which attempted to simulate a population dynamically (and sometimes spatially) are CORSIM (Caldwell et al., 1998), and SVERIGE (Rephann & Holm, 2004; Vencatasawmy et al., 1999). CORSIM (Cornell Microsimulation Model) was used as a modelling tool for basic social science research and policy analysis. It has also been used as a template for a Canadian microsimulation model called DYNACAN (Dynamic Microsimulation Model for Canada), as well as a starting point for a new Swedish dynamic microsimulation model, SVERIGE.

SVERIGE (System for Visualizing Economic and Regional Influences in Governing the Environment) microsimulation model was inspired by the success of CORSIM and in part focused on exploring person-environment interactions and internal migration in Sweden.

In practice, the distinction between agent-based simulation and dynamic microsimulation is not clear-cut (Birkin, 2008). Microsimulations are generally used as tools for predicting or estimating future behaviour (hence their usual applications for policy analysis). In other word, by having a more aggregate focus with projection goals, they are more suitable for 'real systems' and for representing entire national or urban populations. On the other hand, agent-based models have a more exploratory focus, where behaviour and the emergence of patterns/ structures are considered important dimensions of the investigation. They often use synthetic data with no validation ambition.

In part, because most studies of Schelling-based models have been rather 'theoretical' with no ambition for empirical validation (i.e. no considerable amount of real data used), agent-based models have traditionally been used to study the effects of residential segregation. Based on the assumption that these two modelling methodologies can fit together, as explained in more details in section 3.2, chapter 4 will present a distinctive methodological framework that embraces this combined approach. Underlying this methodological approach is the belief that an agent-based model would include microsimulation too (as a sort of superset) and therefore ABM can and should ideally use empirical data as well (micro, aggregate or semi-aggregate).

2.3.4 Geosimulation

Geosimulation is a term used to describe a recent and innovative 'wave' of approaches and research in the application of spatial modelling, particularly in the context of urban modelling. This innovative approach – which rests on the property of spatial phenomena features (Moore, 2011) – is characterised by special management and characterisation of spatial entities (with emphasis on individual/ micro scale objects/ entities with autonomy of behaviour such as householders), representation of wider spatial relationship and interaction, more realistic and flexible handling of space (often using GIS[9]) and time (Benenson & Torrens, 2004; Editorial, 2004).

A good example of a geosimulation approach applied to residential dynamic context is that done by Benenson et al. (2002). However, it is not clear whether the GIS presentations (e.g. colour pie-charts of ethnic group population in houses) in their paper were prepared statically or whether the model is actually able to dynamically display and update them as simulations take place. It should also be noted that their model benefited greatly from access to 'high resolution' data. This type of investigation may not be possible in other countries (as in the case of New Zealand) where 'fine-grained' (micro-level) census data is not publicly and legally available.

In residential segregation modelling, a grid-based environment has been traditionally more predominant (see, for example, Crooks et al., 2008). This is in part because residential segregation studies using simulation models (e.g. those in Sociology) are typically approached through theory-based considerations inspired by

[9] GIS (Geographic/ Geographical Information System[s]) is usually referred to a software system which enables users to efficiently produce, store, retrieve, organize, manipulate, analyze and output (display) spatial and geographical data. In particular, it is an ideal tool for mapping demographic data to geographic entities.

the Schelling approach which uses grid-based environment (see Fossett, 2006a, for example). Moreover, building more sophisticated models requires more advanced programming skills, not available to many researchers.

The reference to geography (space) in these environments is restricted to the grid topology which is not quite similar to the topology of a real city. The geographical/ spatial units of a city are often defined by governmental authorities (e.g. the statistics agency office in a country) based on which (census) data are collected. At best, empirical/ census-based validation of grid-based model would be less attractive, whilst a simulation model which utilizes a more realistic representation of the real city, based on which empirical data are collected would be more apposite for validation and prediction of the future states of a real city.

The final version of the model presented in this thesis will include the micro dynamics of individual residential segregation at a realistic urban scale and will employ similar approaches used in a geosimulation approach. It is therefore proper to classify this model as a geosimulation work/ category too[10].

2.4 Conclusion

Residential segregation emerges as a result of decision-making individuals in regard to their residential choices. These decisions are influenced by many factors and forces, some of them at institutional level. Charles (2003, p. 199) has quoted Logan and colleagues who "lamented" that "we are near the limit of what can be accomplished through the analysis of publically available census data" and argues that the "major limitation to research of this sort … is its inability to capture the dynamic nature of residential segregation".

Simulation modelling seems to be the only possible way to engage actively with this "complexity", and the individual-based Schelling model has been a seminal

[10] Although GIS functionalities are 'embedded' within the model, the modules are implemented outside the GIS. Moreover, the grid-based environment could technically have been developed without the GIS. For these reasons, this model-based GIS framework can be still considered "loosely-coupled" (Batty & Xie, 1994a).

groundwork towards the realisation of this approach. However, as even more sophisticated extensions of Schelling models still suffer from lack of realism, 'small-village' syndrome as well as limited measures (to mention a few), more innovative approach can be used which would take advantage of the large amounts of empirical data (as in MSM) as well as realistic (GIS-based) spatial treatment (as in geosimulations).

Chapter 3

3. Research Design

3.1 Introduction

In their article, propounding a 'neo-positivist' position, entitled "towards good Social Science", Moss & Edmonds (2005) recognize 'agent-based' simulation as an approach to facilitate the move towards more 'formal' modelling and description.

This chapter is dedicated to the description of various aspects of the research design. It starts with the demystification of the hybrid modelling approach in section 3.2. This is followed by explication of the unit of analysis used in the thesis investigation in section 3.3. The conceptual framework which provides a structure within which various control variables can be used for investigation of urban population dynamic and residential segregation and its intensity using different measures is explained in section 3.4. A figure in section 3.5 illustrates the entire research model framework, while the notion and choice of time scale is covered in section 3.6.

3.2 Hybrid Approach

Since the beginning of this project, it has been the intention of this thesis to use innovative methodological approaches for a fresh investigation into the residential segregation dynamic. For example, one of the first developed prototypes had a topology that embraced both bounded and continuous neighbourhoods (for more information on preceding prototypes, see appendix C), whilst practically all of the grid-based models use only continuous neighbourhoods (for more information on bounded and continuous neighbourhoods, see appendix C.1).

The hybrid approach used in the conceptualisation, design and implementation of the final model presented in this thesis embraces some of the most important characteristics of agent-based modelling and microsimulation models.

In their originally conceived form, agent-based modelling emphasises the interaction between individuals (agents) according to the established governing behavioural rules which may evolve stochastically over time as a result of these interactions (e.g. failure or success of interactions with other individuals or environment). On the other hand, agent-based models do not usually and easily support validation. With behaviour as a central dimension, ABM is deployed in exploratory investigations where emergence of new patterns/ structures is an important characteristic. The nature of this type of research approach is more qualitative and inductive in nature.

Microsimulation on the hand neither attempts to model interaction between micro units (agents), nor tries to represent the motivation or intentions of the units (Gilbert & Troitzsch, 1999). Unlike agent-based models which are rule-driven, events in microsimulations are driven by probabilities (they can be partly written and executed by transitional probability matrices). This in part makes them lack an evolutionary dimension – something that normally occurs as a result of the interaction of independent decision-making agents with each other and the environments in agent-based models. Microsimulation in general uses large amount of data and explores their attributes and relations (e.g. characterization of agents by gender, ethnicity, age, marital status, and so on). Hence, there is tendency to use MSM in data-driven environments with validation in mind. The characteristics of investigations and research approach using MSM models are often quantitative and deductive in nature with a more aggregated focus, whilst the projection of simulation outputs are often among the main objectives. When data is in aggregated format, it is usually disaggregated to individual-levels, although aggregate statistics are calculated and used as estimates of the future characteristics of population. A comparison between microsimulation and agent-based simulation is shown in table 3.1 (Mahdavi et al., 2007).

Table 3.1 Dynamic Microsimulation vs. Agent-based Simulation

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

As microsimulation models add more spatial and behavioural interaction and agent-based models add both space and other socioeconomics and demographic attributes to their agents, these simulation modelling approaches have moved towards more common ground in the recent years (see, for example, Wu et al., 2008).

The present research aims at an integration of these modelling approaches in a hybrid environment. Similar to microsimulations, 'micro' units (individuals/ households) will be disaggregated from data identified by one of its important characteristics (here ethnicity). Movements are also generated by a probability based process which is applied to each unit of ethnic group population in area units. On the other hand, as in agent-based models, moving agents use rule-based mechanisms to choose and settle in their preferred locations.

Another hybrid facet not previously exploited by other residential segregation models is provided due to the realistic handling of spatial environment in the model – a feature often found in geosimulation models. This hybrid feature provides an opportunity to explore both settings within the same framework. That is: a typical artificial environment where exploratory experiments can be performed with ease and without much concern about data, and to link (or compare) the results with realistic environments inhabited by merely representative populations (Mahdavi et al., 2009). The potential to develop new insights in such an environment for examining the parallels and differences between these two worlds exists, in particular to see how the outcomes and conclusions made in artificial settings compare with those in a more realistic environment. This can reinforce the credibility of grid-based experiments or potentially shed some doubts on (some of) their conclusions.

Finally, another hybrid feature of this model is its ability to deal with both exogenous and endogenous activities simultaneously. In Schelling single bounded-neighbourhood model (called tipping), events triggering the tipping process are assumed to be exogenous and not generated within the model, whilst in his checkerboard model, agents only move inside the model and their moves generate residential segregation endogenously (Zhang, 2009). The latter endogenous element is obviously present in the model presented in this thesis, as in most typical Schelling-based models. However, the exogenous element can be also operationalized simultaneously in the model with multiple bounded-neighbourhoods by enabling the immigration process to take place – another feature not apparently exploited by other related models.

3.3 Units of Analysis

In this thesis the 'focal unit' consists of an individual population. This can be seen as a micro scale object or entity with autonomy and heterogeneity of behaviour used in geosimulation models. Focal units based on which generalizations can usually be made represent units of analysis or observation (sometimes in aggregate form).

Also in the realistic environment, the micro-geographical scale (based on which neighbourhoods are defined) is chosen by one of the administrative and geographical unit structures defined by SNZ called 'census area unit' (CAU) or simply 'area unit' (AU). There are 316 of these micro units in the urban/ metropolitan area (UA/ MA), as is shown in figure 3.1. However the segregation measures and visual representations/ inspections can be also performed at both meso (territorial authority [TA], 5 of them in total) and macro (one single metropolitan area) levels. Information and statistics about these micro area units, territorial authorities, and metropolitan area are provided in chapter 6.

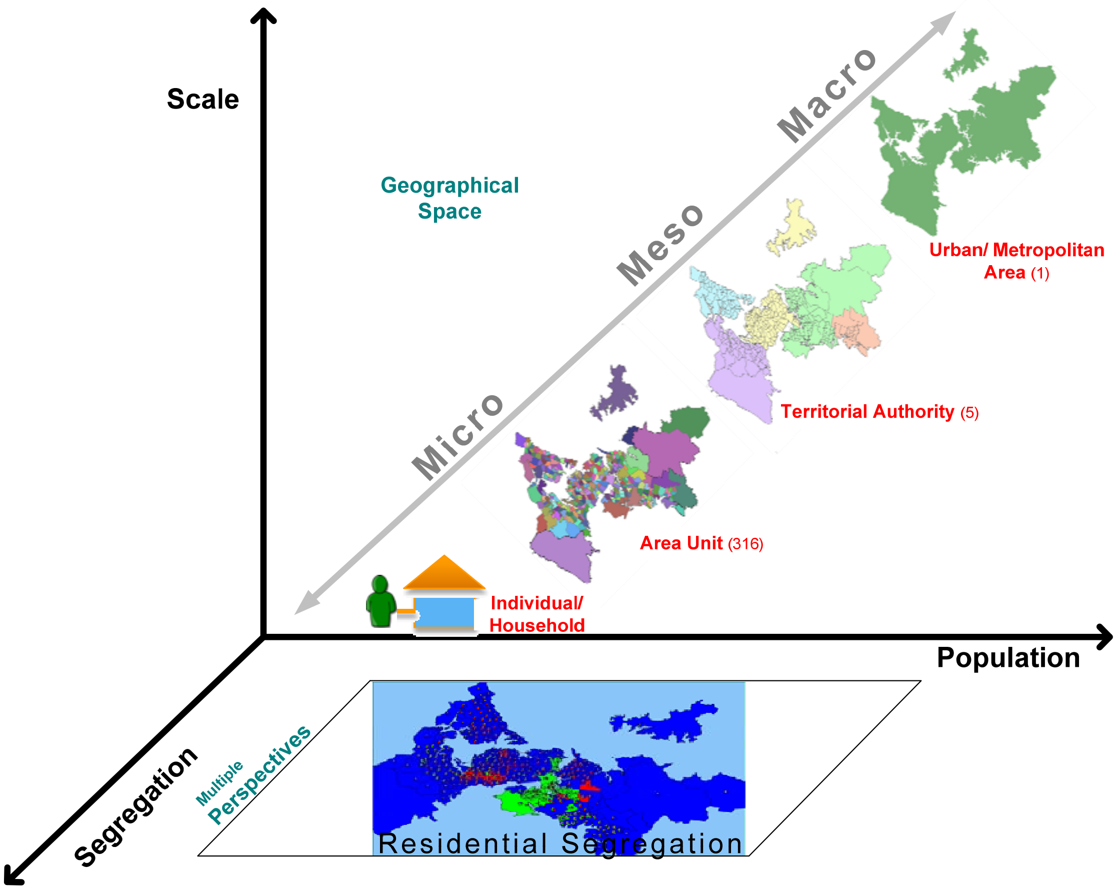


Figure 3.1 Multidimensional and Multi-scaled Perspectives and Capability

As illustrated in figure 3.1, different perspective on the investigation of residential segregation would be possible by focusing on different geographical scales and data level. The methodological approach undertaken by this thesis will allow the study of residential segregation in the model in both inter and intra-urban at three different levels:

- **Micro-geographical level**: allowing individuals/ households described by their qualitative characteristics (such as ethnicity), as well as possible quantitative characteristics to live in one of 316 micro spatial units throughout the large urban area in the model where they may be forced to make choice about a new residential place by searching for a suitable location among many and relocate there.

- **Meso-geographical level**: with few agent-based models developed at the 'mesoscale' (Batty, 2005, p. 1393), this feature enables an intra-urban perspective where several territorial authorities (TAs) can be seen as smaller regions/ districts (although potentially, a spectrum of meso-geographical scales can be defined and used in this model). Each of these mid-size spatial constituent entities can then be studied separately and their semi-aggregate emerging behaviours can be quantitatively measured by various indices, and 'qualitatively' examined by visual comparison with their initial or previous states.

- **Macro-geographical level**: the ultimate goal of the model is to provide a holistic inter-urban perspective of the entire urban system/ area which will provide a global picture/ view derived from the micro and meso geographical entities; that is, a "macroscopic level of description [that] emerges from the interplay of the elements at lower level" (Sanders, 2007, p. 149). This macro spatial entity can also be quantitatively measured by various indices and 'qualitatively' examined by comparison of its visual maps at different time scales.

It is important to mention the likelihood of 'scale-sensitivity' of measures that rely on neighbourhood boundaries because they depend on neighbourhoods of different shapes and sizes (Reardon et al., 2009; Reardon et al., 2008; Wong, 1997). From a broader perspective, this is related to the scale and aggregation problems defined under the 'Modifiable Area Unit Problem' (MAUP) (Openshaw, 1983; Openshaw & Taylor, 1979; Openshaw & Taylor, 1981). This can be seen as the source of statistical bias since variations in the spatial units used for aggregation cause variations in statistical results. It arises in the measurement of residential segregation (Wong, 1997) because residential population data are collected and aggregated for spatial units that have no necessary correspondence with meaningful social and spatial division (Reardon, 2006, p. 174).

The model presented in this thesis is not immune to the issues associated with the MAUP and therefore different choice of micro-geographical 'focal unit' would probably lead to different aggregated results at meso and macro scales. On the other hand, with some technical adjustments, the model would have potential to also be used as a vehicle for investigation of the effects on using different (micro) geographical units; that is, the effects of MAUP on residential segregation results.

3.3.1 Sociological Micro-Macro Link and Geographical Micro-Meso-Macro Link

In the model, agents do not have cognitive capacities. In fact, these 'proto-agents' do not even initiate their own departure when they are dissatisfied with their current locations. A proxy mechanism enforces the departure of a fraction of their population at specified ratios for each ethnic group. After all, people move out for variety of good reasons, for example, employment, departure of children, aging, to mention a few (see, for example, Gilbert, 2006).

On the other hand, despite abstracting and 'proxying' the micro motivations regarding the origins of moves, the essential Schelling micro motivation is very much in place. It is generally assumed that the micro-macro mapping of the Schelling models depends greatly on the householders' behaviours at the micro level. In this model, in addition to threshold for (in-)tolerance, there is also a provision for global-local mobility preference (agents who can move globally can relocate to any part of the 'city' whereas the local movers can only move within the immediate neighbourhood, see also section 4.2.2). This provision adds another important micro level behaviour into the model's dynamic whose effects are linked to the macro structure.

Evidently, "relocations in a segregation model are a form of positive feedback, and when the conditions are right such a feedback can cause major cascades to wash across the systems" (Miller & Page, 2007). It is the combination of these multiple micro motives and micro behaviours, relocations and feedback loops in this model which enables the micro-macro link at the macro social level. Additionally, this model provides another unique geographical layer for linking geographical levels of micro (individual households), meso (district /TA level) and macro (urban/ metropolitan area) to enable and investigate the links between them. Working with a larger population (as is possible with this model) can affect localised variations (Courgeau, 2003) and the possibility of multi-scale analysis in this model will allow simultaneous study at different aggregate levels (allowing better appreciation of the extent of variations).

3.4 Conceptual Framework

Figure 3.2 depicts the conceptual framework for this thesis. Confronting the challenge of making conclusions that attempt to explain the effects of collective phenomena through aggregate (macro), reduced/ semi-aggregate (meso), individual (micro) or combined attributes, it is practical to adopt a framework which does not set a rigid distinction between these elements.

At the micro/ individual level, the most interesting data characteristic in this thesis is ethnicity, although other attributes such as age, gender, education, income or other socio-economic factors could be relatively easily added. The behavioural mechanism at this level is operationalized by micro-individual preferences, either in form of in-group contact or location (local or global). The turnover probability parameter set globally will influence the number of relocations at micro level.

At macro level, more aggregate characteristics such as neighbourhood number or proportion of groups can be positioned. Like turnover probability, the vacancy is set at higher level, with ramifications to be propagated at micro level. Most economics factors are at aggregate level, although they are not currently part of this thesis study (identified by greyed colours). Similarly, discriminatory behaviours by institutional forces or realtors are not currently explored in this thesis. However, two types of behaviours at two different levels have been initially provisioned in the framework, but only discriminatory behaviours by more micro-entities (households) through their preferences will be investigated. Once the realtor mechanism is fully implemented (as explained in chapter 2) in a future version of the model, it will be possible to examine other scale of behaviours, if necessary simultaneously, from two different angles: the micro-individual level with reference to the exercise of choice by the dose of tolerance/ intolerance towards other ethnic groups ('behaviourism' approach), and from the macro level with reference to the exercise of power through realtor discriminatory practice ('managerialism' approach).

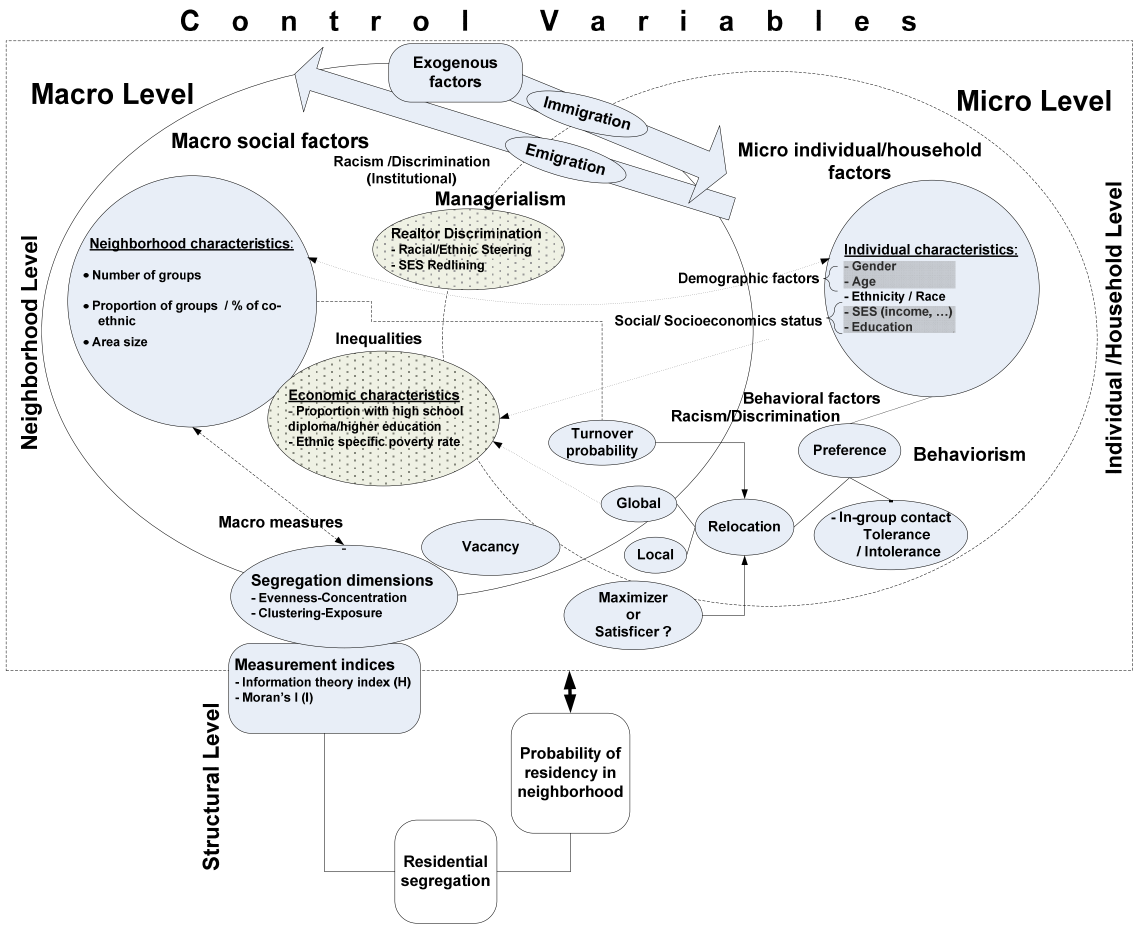


Figure 3.2 Conceptual Framework (greyed areas are not currently operationalized)

As depicted in the figure, the exogenous effect of immigration (i.e. individuals who enter the simulated world (often in mass numbers) as macro force (which is also influenced and controlled by government policies) will eventually affect the dynamism at the micro level.

All the measures are calculated in aggregate format at macro (or its semi-aggregate/ meso) level to gauge changes in segregation at the structural level.

It is hoped that this conceptual framework provides a foundation as well as some boundaries for the upcoming series of hypotheses that will investigate the relationship between residential segregation and selected variables at different (multi) levels. It also presents the base upon which the thesis objective and research questions can be operationalized and upon which the outcomes can be evaluated and analyzed.

3.5 Research Model

The research model for this study is depicted in figure 3.3. As can been seen in this figure, data can used based on censuses, or they can be purely 'synthetic'. The execution of each scenario will take place via a preset but flexible environment of selected key variables (some of them from the census data, such as ethnic composition), while examination will be carried out by systematically varying parameters (time T and T+1 in the figure), measuring specific output variables, comparing input variations with output variations, and comparison of the results using advanced analysis for discerning interesting behaviours and patterns, as well as comparing model results with empirical or demographic data. This experimentation involving systematic variation of the applied parameters and variation of input can also test the model's sensitivity (Krink, 2004). Alternatively, the pattern-oriented calibration (POC) – which will be explained in chapter 7 – can be used to calibrate the model in order to attune it for specific investigations. Aggregate effects can be obtained by combining individual simulations – that is, synthesizing from relatively simple parts. Each finding will constitute knowledge about a theory, gained from a scenario-based investigation.

Usually the scenarios are simulated by a repeated number of runs using identical initial conditions and parameters, but using different random number seeds. This helps to determine whether particular observed patterns are idiosyncratic or typical (Castle & Crooks, 2006, p. 38). The process is also necessary due to the existence of stochastic processes in the model, so that "many runs are necessary to obtain a complete picture of all possible implications …" (Werker & Brenner, 2004, p. 10).

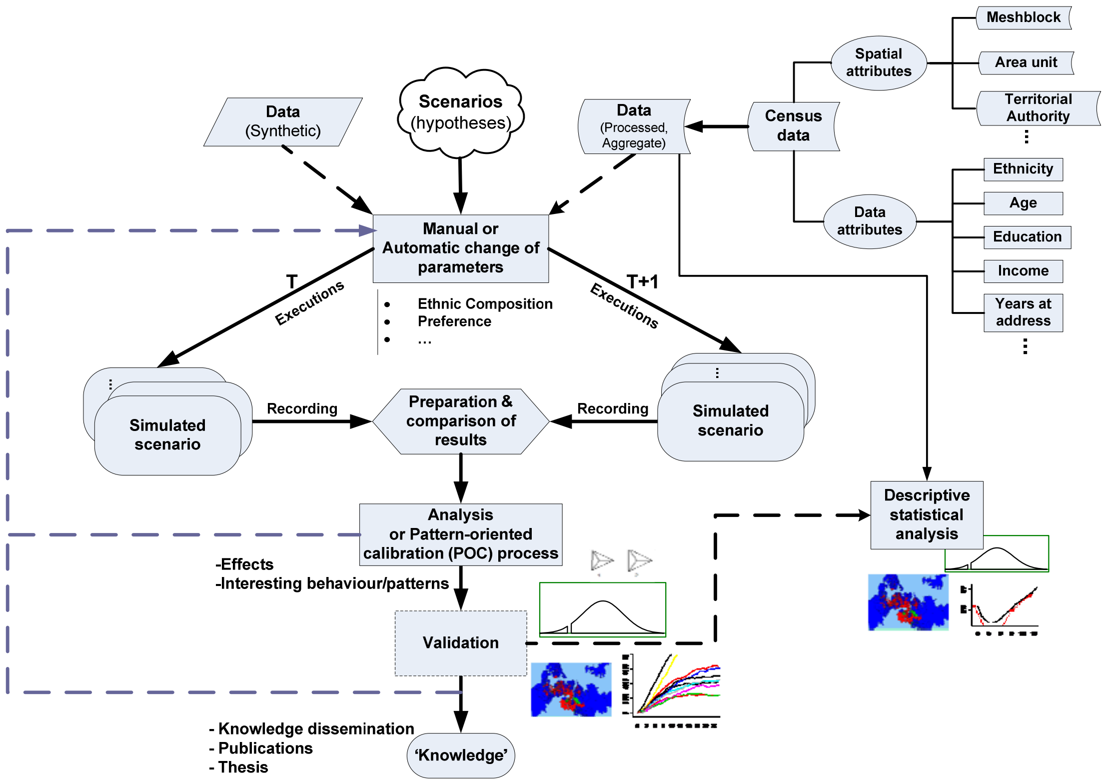


Figure 3.3 Research Model

At the micro/ individual level, the most interesting data characteristics in this thesis is ethnicity, although other attributes such as age, gender, education, income and other socio-economic factors could be added with some additional (but minor) work. The behaviourism mechanism at this level is operationalized by micro-individual preferences, either locational (local or global) or in the form of in-group contact. Set globally the turnover probability parameter will influence the number of relocations at micro level.

Likewise, the results of the model will be analyzed to discern interesting behaviours and patterns. Uncertainty and sensitivity analysis can be optionally used to assess the range of model outcomes given uncertainty in parameters, model error or exogenous factors (in case of uncertainty analysis) or measure the change in model output associated with the change/ perturbation in model output (in case of sensitivity analysis). The measured values (also dynamically plotted and available during the simulation lifetime) are also captured in a variety of formats, for further advanced analyses and generation of descriptive and explanatory graphs.

Data is an important part of this research and greatly shapes its nature and direction and the design of model. Data is explained and analyzed in chapter 6.

3.6 Time, Equilibrium and Convergence

The decision about the time scale (i.e. how many cycles or time ticks scenarios will be simulated) is based on a combination of pragmatic and logical reasons. The model is able to measure different segregation dimension indices at different scales. The calculation of these measures are computationally demanding and other factors also play a role in the duration of simulations (for instance, even an asymmetric threshold can cause segregation to take much longer).

In the majority of Schelling models, a specific variable exists which counts the number of unsatisfied (unhappy) or satisfied (happy) agents (see, for example, Daude & Langlois, 2007). In these models, the convergence occurs (or is studied) when the number of unsatisfied agent approaches zero (or everyone become happy).

Table 3.2 shows several models for investigating residential segregation which use different time scales (ticks), from "millions of iterations" in Bruch & Mare's model to 30 cycles (where each cycle roughly corresponds to time periods of 6-12 months) in Fossett's model. Although Bruch & Mare believed that their large iteration was necessary "to reveal the essential behavior of the models, albeit not at equilibrium"(Bruch & Mare, 2006, p. 680), their statement that their model could be potentially calibrated to real time suggests that the choice might have been influenced by such an intention.

Table 3.2 Comparison of Selective Residential Segregation Models

Parameter/ Feature	Bruch & Mare (2006)	Fossett (2006a)	Benenson et al. (2002)	Sander et al. (2000)
Time tick	"1 million iterations"	"30 cycles" (times)	"100 times"	"Thousand times"
Vacancy	15%	6%	'Realistic'	0% to 20% (e.g. 5%)
Neighbourhood type	Moore radius 2	Moore radius 1	Contiguity based	Moore radius 1

Itzhak Benenson (1999) agrees that "the problem of which time scale fits the given approach is more complicated" and that the model solutions would describe the chosen time dynamics of the system. He furthermore adds that "the proper selection of the processes in regards to the time scale depends greatly on the system's specificity, and no other confirmation of certain preferences can be obtained by any means other than the operational, Do it!"(p. 168).

The scenarios simulated in this thesis do not aim to achieve equilibrium. In other words, the dynamics of the urban system and aggregate emergence via observed measures have not been defined in term of convergence. However, the decision to limit the number of iterations to 30 (each iteration equivalent to one year) was not solely based on speed constraints. The model was not originally designed to be used in "real time", and the objective of its verification with quinquennial empirical census data necessitated an appropriate usage of a time-scale. Similar to Fossett's simulations, it was believed that a cycle representing a year would be appropriate for this study.

3.7 Discussion and Conclusion

The design of this research and the model that will follow in the next chapter incorporate several unique features, including an ABM-MSM hybrid geosimulation modelling approach where the simulated environment comprises multiple bounded-neighbourhoods, capable of simulating the 'world' in both artificial and realistic modes. Since "the outcomes of residential segregation models may strongly depend on the way that neighborhoods are conceptualized and represented" (O'Sullivan, 2009, p. 508), it would be interesting to investigate the residential segregation phenomenon in an environment and conditions different from typical Schelling-based models.

The idea of exploring the dynamics of residential segregation and the linkage between macro and micro elements using a hybrid approach is illustrated in figure 3.4. The micro individuals constitute parts of the system (city) and their interactions with each other along with the environment will determine meso/ macroscopic manifestations (e.g. segregation, mosaic structure) at different scales.

Figure 3.4 Agent-based Microsimulation Design

It is expected that the entire framework (conceptual, research model and simulation model) along with other analysis methods will reveal certain facets of a complex residential segregation system, and will allow engagement in an analysis that involves using census data to investigate the patterns of residential segregation in Auckland. Through the proposed framework, the evaluation of the intensity and character of divisions as well as influence of various selective factors and forces involved in residential segregation dynamic can be examined.

Underlying this methodological approach is the belief that improved theoretical insights and more effective guidelines for future inquiries and policy making can be delivered by directing research towards the study of existing population structures using census data and evolving them dynamically through time using simulation-based study in a theoretically driven research framework.

Chapter 4

4. Design of the Simulation Model

4.1 Introduction

This chapter will unpack the design of the final version of the simulation model. Software engineering principles as well as object-oriented design (OOD) concepts were applied in the development of this model. Nonetheless, like art works, most design endeavours (including software developments) are creative products. This is reflected through the combined (hybrid) design approach. However, the role of existing data (necessary for the validation objective covered in chapter 6) in the shaping of the model should be particularly emphasized.

Section 4.2.1 will explore some of the model's functionalities through its graphical user interface (GUI). This is followed by the model 'magnitude', its key design concepts and with an overview of its scheduled processes (in particular the *placement* algorithm) in the subsequent sections. The comparison of the final simulation model and other Schelling type models is presented in section 4.2.5.

4.2 *HAAMoS*

The development of *HAAMoS* (Hybrid Agent-based Microsimulation of Segregation) proceeded after the development and trial of several prototypes (for more information about these prototypes see appendix C). The purpose of building prototypes was to conduct feasibility studies and explore potential avenues.

Any software system that undergoes continuous changes will grow in complexity. As 'software's entropy' increased during the development of the model presented in this thesis, it was important to apply software engineering principles as well as object-oriented design concepts early on to manage the complexity. The applied principles include 'separation of concern', 'abstraction', 'generality', 'design for change'/ 'anticipation of change', 'encapsulation', 'modularity', 'maintainability', 'reusability', 'incrementality'/ 'extensibility', and to some extent 'robustness' (related to proper handling of invalid inputs; since this model was not developed for a third party, less attention was paid to this latter principle as I was the sole developer and user at the same time). Further principles included using 'design patterns' and making sure to 'get it right before making it faster'.

The initial modelling approach employed was closer to KISS (Keep It Simple Stupid) paradigm which started with the simplest possible model and gradually added more complexity whenever it was necessary and justifiable. Nonetheless the final model could have been possibly characterized better as a descriptive model, in part because it allowed for more of the available evidence to be applied, as it is defined for the KIDS (Keep It Descriptive Stupid) approach (Edmonds & Moss, 2004).

Some aspects of the model presentation style used in this chapter are recommended by ODD (Overview, Design concepts, and Details) protocol (Grimm et al., 2006; 2010). Albeit being promoted as a 'standard', Grimm et al. (2010) recognized some difficulty of employing the full protocol in all cases. This may be related to the 'ecological outlook' of ODD, as well as its evolving maturity (for example, there have been significant changes between the initial 2006 proposal and the 2010 version of protocol). In fact, the latest version of ODD would be unlikely to be the final version. For these reasons, this thesis did not thoroughly follow the original protocol when it was not deemed to be completely suitable.

4.2.1 Model GUI and Functionality

This section reveals the key functionalities of *HAAMoS*, mostly through its graphical user interface (GUI). *HAAMoS* is implemented in Java using *Repast J 3.1* [11] libraries. As a hybrid model, *HAAMoS* provides two modes of simulation environments. Figure 4.1 shows the realistic environment of the model in execution.

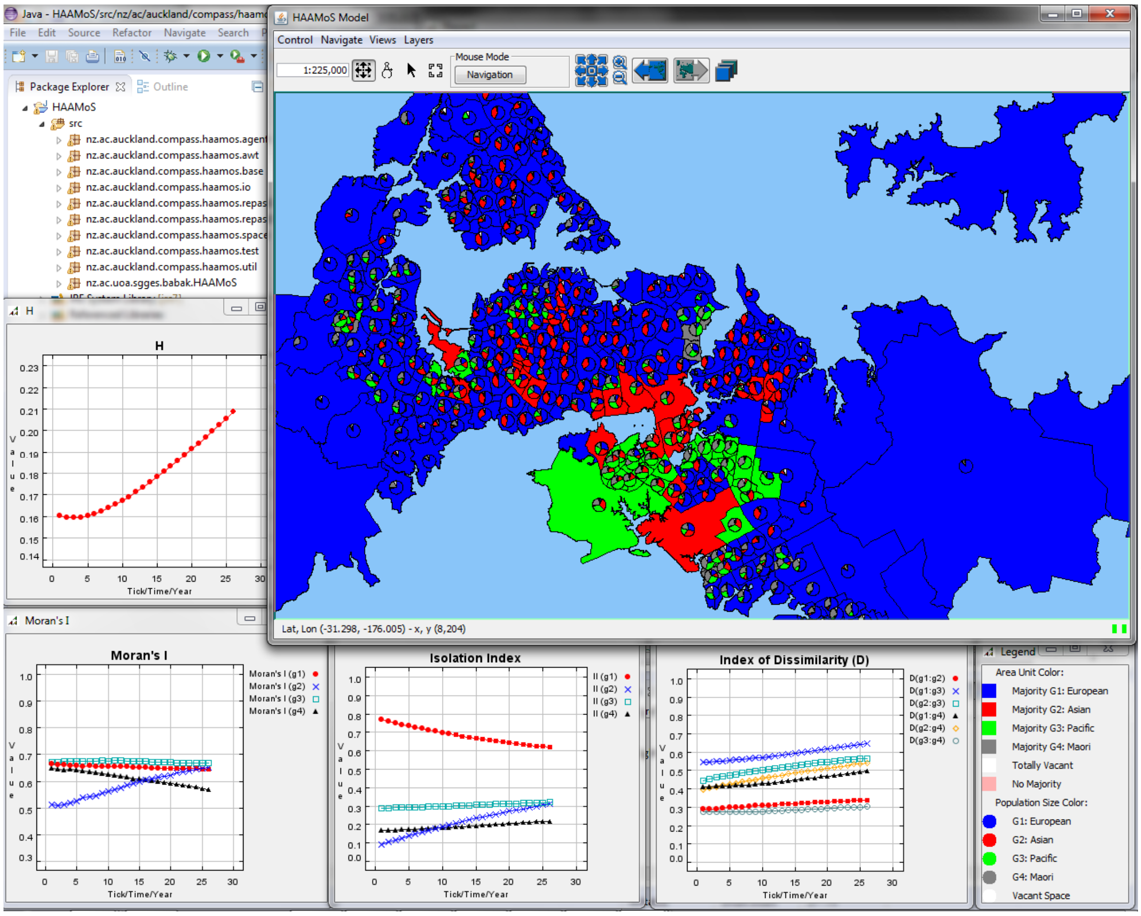


Figure 4.1. Snapshot of *HAAMoS* – Realistic Mode

The model uses GIS map of Auckland region which represents the real map of Auckland region area shown in figure 4.2.

[11] Repast (Recursive Porous Agent Simulation Toolkit), http://repast.sourceforge.net

Figure 4.2. Real Map of Auckland Region

Up to four ethnic groups can be simulated in the current model, namely European (group 1), Asian (group 2), Pacific people (group 3) and Maori (group 4). As indicated by the model legend (see figure 4.3), the population size of these groups are respectively represented in Blue, Red, Green and Grey. Each area unit also displays the colour of the group in plurality (relative majority) position in that area unit (update at each time cycle/ year). When an area unit is totally vacant, this is shown in White. In rare situations when it would be impossible to determine a plurality group (e.g. two predominant groups are of exactly the same size), the area unit is rendered in Pink.

The population sizes of each area unit are dynamically represented by small pie-charts. These are shown in a zoomed version of the simulated urban area in figure 4.4.

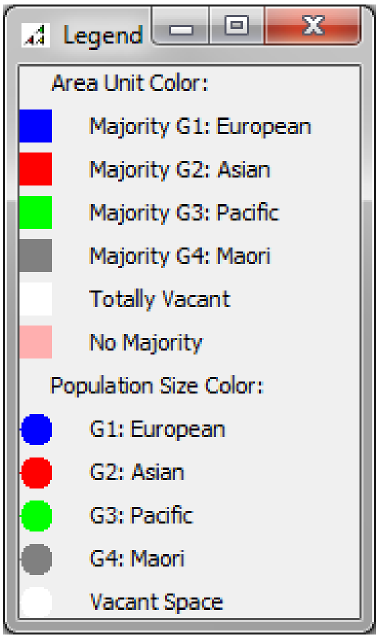


Figure 4.3 Colour Legend

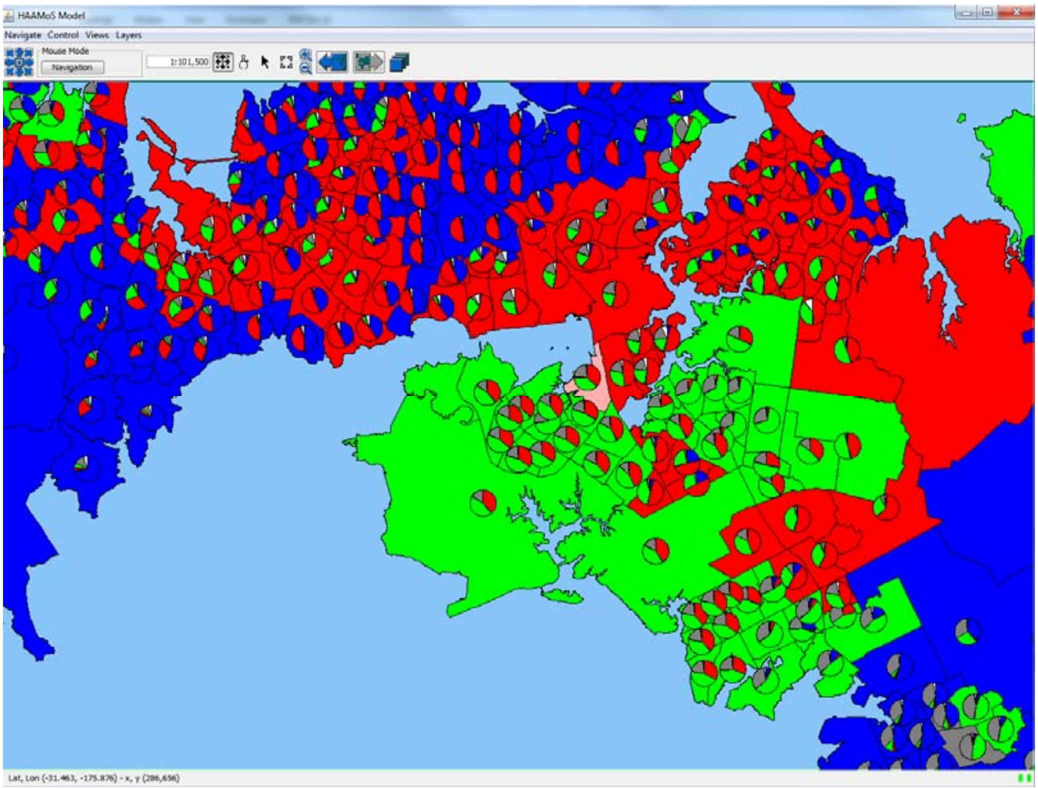


Figure 4.4 Population Proportion in each AU as Pie Chart – Realistic Mode

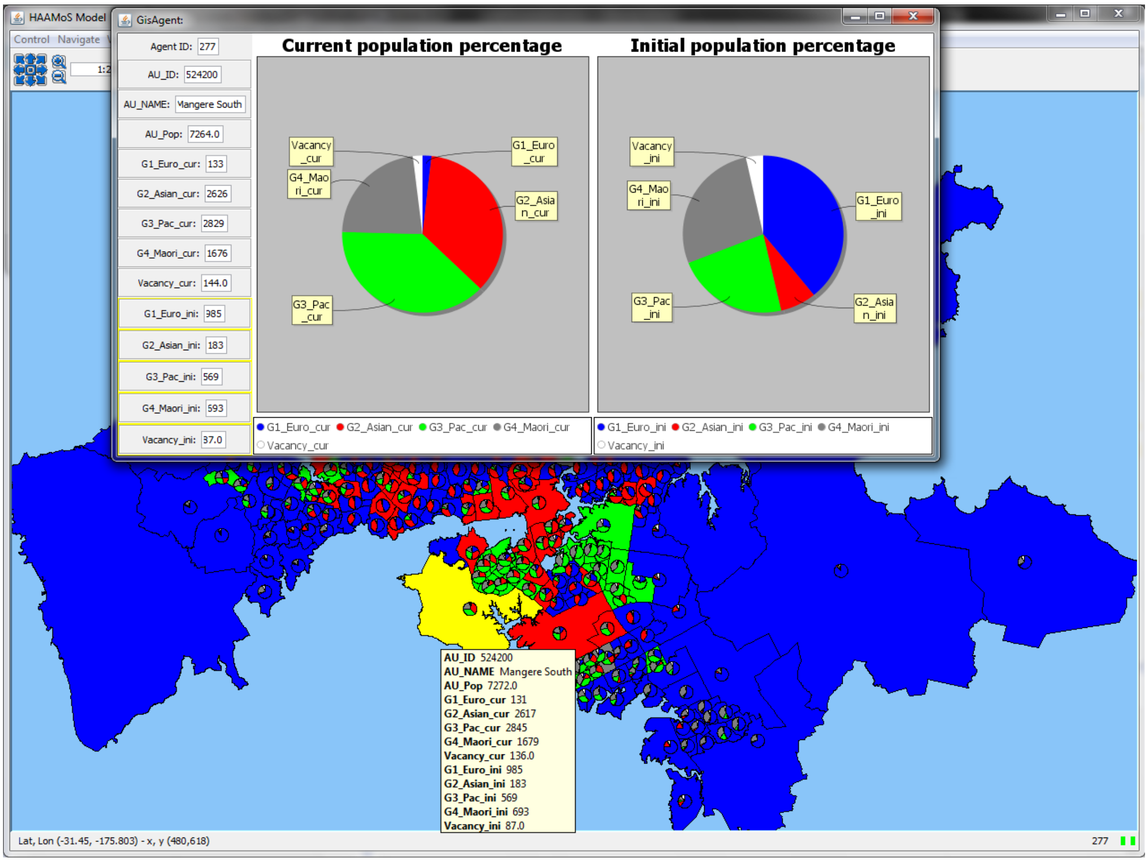


Figure 4.5 Detailed Properties of Individual AU – Realistic Mode Example

Area units can be inspected in more detail by selecting the 'Properties' item from the pop-up menu. This will invoke a large window similar to the one shown in figure 4.5. The left side pie-chart with 'Current population percentage' title is similar to the one drawn dynamically on top of the area unit, but is larger and more legible. The other pie-chart on the right (entitled 'Initial population percentage') shows the composition of the area unit at the initial state of simulation. By comparing these two pie-charts, the extent of change from the start of the simulation until the time of inspection becomes possible (a series of 'textfields' also display values for both current and initial states).

Batch Execution

The multi-runs function of the model allows a batch execution mode by loading a prewritten 'parameter file' (i.e. batch file). This mechanism permits multiple executions of predefined scenarios in automated fashion, whilst elected measures and outputs are recorded. The parameter file has a specific syntax and format for defining the parameter space and describing how the model should explore that space. An example is provided in appendix E.

4.2.2 Model 'Magnitude'

Following on the main goals schematised in the conceptual framework (see section 3.4), key controlled variables in the current version of this model through which various scenarios are built are the following:

- Number of ethnic groups
  - Up to four different ethnic groups can be currently simulated in the model.
- Proportion of groups
  - For empirical experiments, the proportion are predefined at the micro-geographical level used by this model (i.e. area unit) using census data, although it would be possible select the starting year among one of three census periods (default is 1991).
- Turnover probability percentage
  - A percentage defined as probability which determines the fraction of a given ethnic population who 'decide' to relocate (for unspecified reasons) at each cycle. This can be separately set for each ethnic group.
- Global location preference probability percentage
  - A locality percentage defined as a probability which determines the fraction of the moving ethnic population (i.e. a subset of the moving population determined by the turnover process) who can relocate globally (i.e. they are not restricted to local moves within their

immediate neighbourhoods only) at each cycle. This can be separately set for each ethnic group.

- In-group preference (tolerance/ intolerance)
    - A value expressed in percentage term which indicates the threshold below which the agent prefers not to settle in a given location (referred to as tolerance). This can be separately set for each ethnic group.
- Net inflow (immigration)
    - For theoretical experiments, a total net number of newcomers' inflow into the urban area at each cycle along with the probability percentage share of each ethnic group can be specified. For empirical experiments, the number of newcomers (including endogenous and exogenous immigration, and growth by birth) are calculated based on the available statistics (and projections) for each territorial authority.
- Net outflow (emigration)
    - For theoretical experiments, a total net number of outflows from an urban area at each cycle along with the probability percentage share of each ethnic group can be defined. For empirical experiments, the specified number of outflows that are removed from the 'world' at each cycle is calculated based on the available statistics for each territorial authority.
- Vacancy percentage
    - For theoretical experiments, a specified percentage is used to introduce the number of vacant spaces as a specified fraction of total population in each smallest spatial entity (area unit) in the urban area. For empirical experiments, a vacancy function is used to

    - maintain the vacancy rate (proportion of vacant spaces to population of the area unit) within the empirically-based rates for each TA.

- Persistent probability (for searching in an area unit for a suitable location before moving to another)
    - A percentage-like value specified to represent the probability that an agent will shift her efforts from searching current area unit to another one.
- Maximum number of visits
    - A mean value to randomly determine and assign (Poisson distributed) the maximum number of visits for each individual agent after which it will settle in the best found location – if no place would exactly satisfy its preference (once this is set, it will be kept constant for the rest of simulations presented in this thesis).
- Precision of the estimated (guessed) co-ethnic proportion
    - A standard deviation value which will be used to determine the precision (or deviation) of randomly obtained estimates of a given location co-ethnic percentage relative to its real value by an agent (this will be kept constant in the simulations presented in this thesis).

4.2.3 Design Concepts

This section gives an overview of the model characteristics using the descriptive items of the ODD's 'design concepts' promoted by Grimm et al. (2010).

Basic Principles

The model is built based on theories about residential population dynamics/ movements and the householders' decision making process for selecting a new place to live. Similar to Schelling model concepts (and most Schelling-like models), household preference plays an important role in the dynamics of the *HAAMoS* model. In this way, micro behaviours at individual level and macro effects at aggregate (population) are operationalized.

However, there are many other differences. In addition to typical tolerance (a personal threshold indicating a desired co-ethnic percentage perimeter in a potential neighbourhood above which a householder considers it suitable for relocation), preferences in this model are extended to include both local (neighbourhood) and global (metropolitan-wide) indicators for relocating to a new location. Each individual household also exhibits a different behaviour by setting an individual (different) number of visits limit to potential places before being forced to settle in a best found, if no other matching location was found. Furthermore, each ethnic group has a different level of persistence for searching in a given area unit for a suitable location before moving to another. Although agents' behaviours are not directly informed by census/ survey, they are represented by 'empirical rules', in the sense that their rates are usually described as 'realistic' probabilities. (A table containing the definitions of variables, their range as well as their empirically-informed levels is provided in appendix B).

The motivation for moves does not stem from individually-aware agents, meaning that household agents do not decide to move because they have become dissatisfied with their current location (e.g. because of the change in the neighbourhood ethnic composition). It occurs as a result of a turnover process where a percentage of population 'decide' to move for various valid but unspecified reasons.

The model enables the simulation of an entire metropolitan area at aggregate level using empirical data (census-based). Unlike most Schelling-style models, this model allows simulations to be conducted in both 'realistic' (census tract-shape) and 'artificial' (cell-shape based) environments. The model will be able to shed some light on the role and intensity of various agents' behaviours (e.g. different preferences, rate of movements), as well as scarcity or abundance of vacant spaces in the formation, persistence and degree of observed segregation.

Although in the current version no economic variable is included, the capability of some agents to move locally or globally (anywhere in the wide urban area) can be seen as a proxy for personal economic freedom (wealth). With all these factors the model allows the testing of theories and assumptions about the influence and scale of individual or group decisions in the dynamic of residential segregation. A more detailed comparison of the model with the characteristics of Schelling-based models can be found in section 4.2.5.

Emergence

As a result of population dynamics (i.e. movements of individual-level household agents (of different ethnic groups) combined with their heterogeneous behaviours, endogenous interactions with the urban environment, as well as exogenous factors such as immigration, the distribution and proportion of ethnic groups throughout meso/ macro geographical space change and segregation gradually emerges. That is, the main emergent phenomenon in this model manifests in forms of spatial distribution of an ethnic population (aggregate households) emanating from the behaviour of micro actors (individual household). The level of segregation is calculated by several measures for different dimensions of segregation. These measures are described in chapter 5.

Objectives

The aim of household agents who (for different reasons) decide to relocate is to find a suitable place that corresponds exactly or as close as possible to their preferences. In the pursuit of this goal, as soon as agents visit a location that matches their prerequisites (i.e. co-ethnic proportion of the neighbourhood is equal or higher than their preferred threshold) they move to it. If however during this search period and after the exhaustion of their search visit limit they cannot find an ideal (perfect match) place, they will compromise by moving to the best found (i.e. with nearest proportion ratio to the preferred value) among those already inspected. The decisions on relocation are governed by 'encapsulated' behaviours of agents which are controlled via variables and parameters which are set prior to the execution of scenarios. More detailed description of the relocation process can be found in the next section.

Prediction

Individual agents cannot estimate the future consequences of their decisions and cannot predict the future condition of their environment. Like in the original Schelling model, agents do not have any knowledge of the 'global' emerging phenomenon (i.e. segregation) as a result of their individual local interactions. This is sometimes referred to as 'weak emergence' as opposed to 'strong emergence' (see, for example, Dessalles & Phan, 2005) which is sometime defined as an environment where agents can sense the emerging phenomenon which consequently affects their decision (and even worse, they work actively for the emergence or intensification of segregation). In other words, in weak emergence, the observer of the emerging 'global structure' (segregation) is external to the system, but in strong emergence, the agents (observers) are able to perceive it and get explicitly involved in its emergence. As external observer to this system, we can use this model and learn more about individual behaviours and their effects on the segregation level. Subsequently, we can use what we learn about their behaviour to predict the emerging aggregate. But the agents in this model cannot do this. They cannot see how their micromotives and local decisions can eventually lead to a global macro effect/ structure called residential segregation.

Sensing

Individual agents are assumed to know about their own preferences. However, they can only 'sense' (i.e. have an 'estimated' information of) of a neighbourhood's characteristics (e.g. ethnic composition). This 'sense' may more or less correspond to the real/ exact situation on the ground. Acquiring this imprecise knowledge through stochastic approximating mechanisms in the model is empirically realistic, as few people in the real world have (or will bother to obtain) a perfect census-based statistics about a neighbourhoods' characteristics or ethnic composition during the decision-making process relating to their residential moves.

Interaction

Individual agents do not directly interact with each other, but rather through the environment. They are competing for the same available resources (i.e. limited vacancies in the neighbourhoods). In their pursuit of this objective, they have to adapt to an environment that consists of other agents also interacting with the same environment. To some extent, their choices depend on (or are influenced by) the behaviour or choices of other agents. This 'system of interaction' in the model consists of the interaction between individuals and environment (micro or macro), and the individual and the collectivity (aggregate population).

Stochasticity

Many processes in this framework are modelled in a stochastic manner. This is often because the underlying detailed knowledge about some of the processes are unknown (or not fully understood), or there is not enough information/ data to implement them or they are simply unimportant (how a certain process functions exactly in the real world has no or little effect on the results and analysis). So the variability needed in the model is randomly generated using appropriate probability distributions. Some of the processes in this model can optionally work stochastically.

The 'turnover' process which causes a percentage of population (for each ethnic group) to 'move' at each simulation cycle can be specified stochastically, so that at each cycle a closely estimated number of agents are randomly forced to move. Also, the percentage of these moving populations (i.e. a subset of the previous number) predestined to move globally is also similarly specified. The exact number of agents with this type of behaviour are determined randomly (albeit within a relatively close range) at each cycle. When the inflow (immigration) and outflow (emigration) are not based on census (e.g. for theoretical experiments) the model allows a total number of agents to enter (or leave) the city at each specified cycle. However, the precise number of agents entering (or leaving) is determined stochastically by the model within an adjustable portability range (more or less close to a total specified number, but not quite exactly).

A maximum number of visits limit is also allocated randomly (Poisson distributed) to each individual agent who can do global moves (for agents moving locally, this value is set to the size of immediate neighbourhood) after which the agent will settle in the best visited location. The persistence of each ethnic group in the continuous search for a desirable place within a given spatial unit before eventually moving on to another is specified by percentage. This is tested during the placement process in a probabilistic fashion to see whether the persistence has been exhausted or still stands (search within the same area during the placement process or move to another area).

A random mechanism is also used in the model to ensure that agents would not be able to obtain completely accurate (exact) information about the neighbourhoods they consider moving to. This is to mimic the reality where few people will/ can obtain such faultless information which will be crucial in the final decision about relocation. An option exists in the model to apply an empirically informed vacancy rate within some probability limits for each territorial authority (and constituent area units), although it is also possible to strictly apply the exact census-based percentages which are available at TA level. Albeit with the presence of many stochastic elements in the model it would be possible to reproduce the same results using the same random seed.

Collectives

Individual households belonging to the same ethnic group exhibit some of the same behaviours and characteristics (for example, they share the same level of tolerance towards other ethnic groups, although they may differ in their preference for relocating locally or globally, or in how many places they visit before deciding to settle). The population of a given spatial area/ unit (at different micro, meso or macro geographical scales and for each different ethnic group) is a collective representation of individual households at a given time, although these latter actors resuscitate only when they are transmuted to moving agents by the turnover process. The aggregate information is encapsulated within their respective geographical unit. In this way, they can be seen as separate entities with their own variables (measurement are done using these), although they share many traits with the individual actors. While the population as a collective entity has an important role in this model, the aggregated macro effects and behaviour outcomes (e.g. segregation) caused by the micro motives and behaviours at household level are not collectively sought.

Observation

At the end of each simulation cycle, segregation in the forms of inter-group indices of D and H, (which measure ethnic groups' evenness and concentration dimensions), Moran' I index, (which measures spatial dependence and neighbourhood clustering), and I.I. index, (which measures both isolation and interaction of an agent to others or same group people) are calculated (see section 5.2 for description of these measures) at different spatial area scale (meso and macro, as explained in section 3.3). The counts of AUs with a given ethnic group in plurality (relative majority) position is also calculated at both macro and meso geographical scales. Moreover, when emigration is used in the model, it would be possible to measure the absolute vacancy number as well as the vacancy rate (proportion of vacant space to population size) at both territorial authority and metropolitan area levels. Although these are not used for the simulation scenarios presented in this thesis the model can also calculate local Moran's I (LMI) as well as location quotient (LQ) values for each area unit. These measures are both displayed in various charts as simulations are carried out and recorded for further analysis. Each of these measures can be controlled by explicitly specifying whether to include or exclude it (and at which scale) from either the chart display or recording process or both (with some level of effects on the speed of simulations).

4.2.4 Process Overview and Scheduling

The development of *HAAMoS* (along with preceding prototypes) has been part of the research process itself. Nevertheless, a comprehensive detailed description of its implementation is beyond the scope of this thesis. Moreover, the source code of the *HAAMoS* model is included in the attached CD-ROM. Sufficient information about the process and scheduling of the model necessary for the replication of the model are provided. The UML diagrams describing Java packages, model use-cases, static class and sequence diagrams can be found in appendix D.

Initialization

The initial state of the model world (at time t = 0) for a given simulation depends on the selected or inputted values for variables presented in the 'Settings' panel of the model, or coded in the 'parameter file'. Default values are initially offered for all these variables and if they are not modified, they will be used to initialize the model. In this way, initialization of the model can vary among simulation experiments. The initialisation of simulated scenarios presented in this thesis will be explained at the time they are presented.

Scheduling

The time in this model can be seen as discrete, meaning that it assumes scheduled actions or events occur at regular time steps (also called 'ticks' or cycles). A tick is a single iteration over all the scheduled actions, and these actions are scheduled relative to each other. A scheduled action occurs using the results of a previous action as its basis. For example, actions scheduled for tick 2 will complete execution before actions scheduled for tick 6 begin execution. However, if nothing is scheduled for the ticks between 2 and 6, the actions scheduled for 6 will begin execution immediately after those at tick 2 complete (as if ticks 3 -5 do not exist).

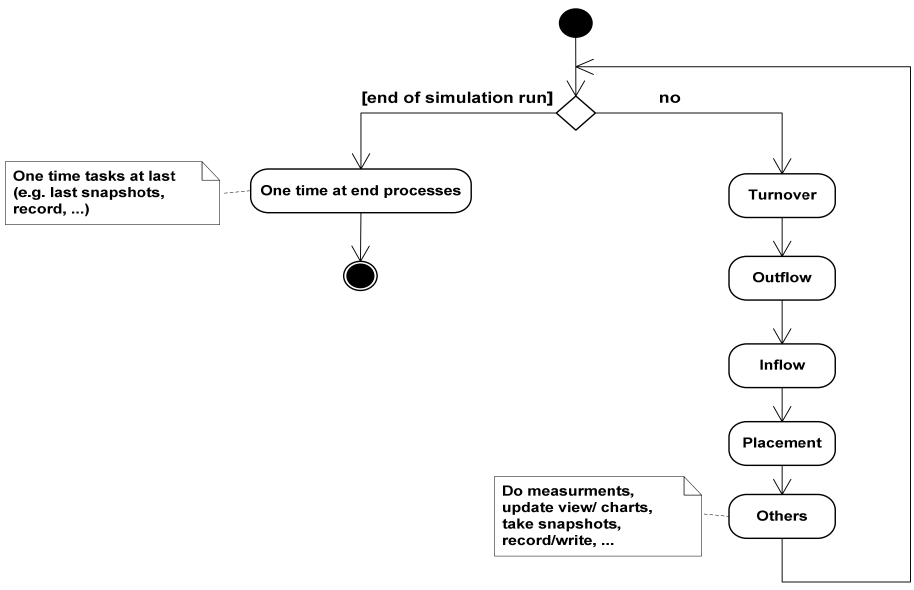


Figure 4.6 Scheduled Processes and Their Orders

At each time step, the following processes are executed in the order shown in figure 4.6:

- Turnover:
    - For each area unit (AU) in the urban area generates [stochastically] moving household agents as specified fractions (proportional to the size) of each ethnic population residing there. A smaller fraction of these moving agents will be designated global movers [stochastically] according to other specified percentages for each ethnic group.
- Outflow:
    - *Theoretical experiments*: For total number of outflow (emigrants) [if not zero], reduce stochastically the size of ethnic population proportional to the specified probabilities for each ethnic group.
    - *Empirical experiments*: For the simulated period (e.g. year) where the empirical (census) data exists, calculate territorial authority (TA)-based on quinquennial population change. If there is a population decline, reduce the size of that population randomly from AUs in the TA where the decline has taken place.
- Inflow:
    - *Theoretical experiments*: For total number of inflow (immigrants) [if not zero], add stochastically new household agents to the urban area entry 'pool' proportional to the specified probabilities for each ethnic group.
    - *Empirical experiments:* For the simulated period (e.g. year) where the empirical (census) data exists, calculate TA-based quinquennial population change (e.g. immigration/ growth after emigration/ population decline) and introduce them randomly into area units [or if stated, in area units under their respective TAs] and in equal

portions according to the specified time elapse for census periods in the model. A specific vacancy function takes care of maintaining empirical TA-based vacancy rates specified [exactly or stochastically] in area units.

- Placement:
    - *See below*.
- Others:
    - Update model's visual display (and population pie-charts).
    - Calculate specified segregation measures (e.g. D, H, I.I, GMI, PI …) for specified spatial scales (meso/ macro) and [record] and/ or [update the charts].
    - Take snapshots of specified graphs and specified intervals.
    - Do one time tasks at a specified time (e.g. beginning or end of simulation).

Placement Process

An important scheduled process in the execution of simulated scenarios is the placement phase which is separately discussed in this section.

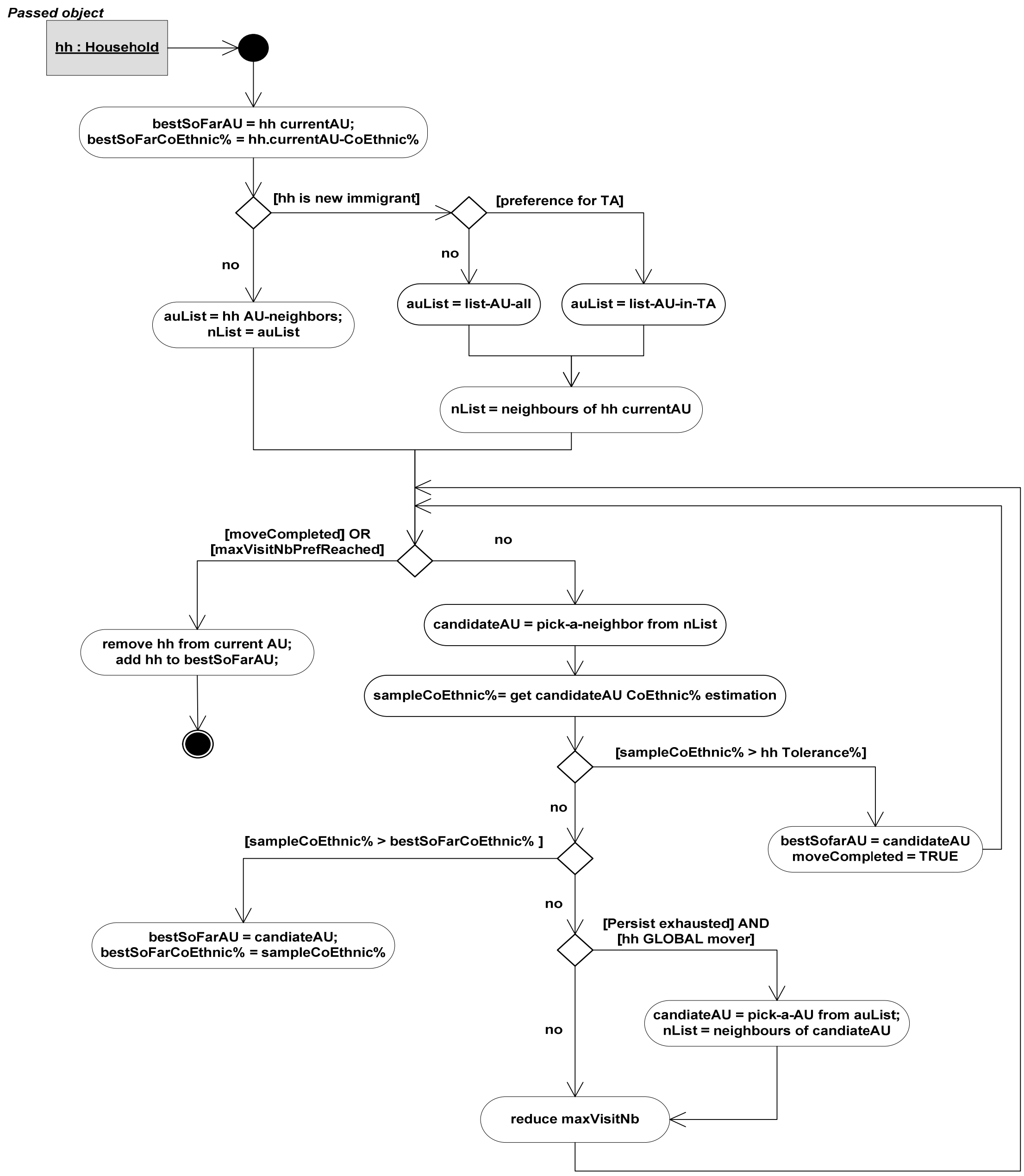


Figure 4.7 Activity Diagram of Placement Process

Figure 4.7 illustrates how the placement of a single household agent in a single simulation cycle is implemented.

Each household agent in the moving pool is randomly selected and passed to the entity that manages its placement. The behaviour of household agents conditioned by the reality and constraints of the housing market environment can be summarized as followings:

- The search for a new place to settle is limited to the
    - Immediate neighbourhoods for those agents with local move preference.
    - Specific territorial authority (TA) for newcomer agents (calculated based on the growth of population) when this preference is enabled in the model.
    - Entire urban area (UA) for those agents with global move capability or newcomers – if preference for TA settlement is not specified.
- Except for newcomers (immigrants) who have no previous presence history (reference) in area units (in this case, an area unit is chosen randomly in UA, or TA when specified), the search for suitable location starts locally (i.e. in the neighbourhood). If such a place is not found within an agent's 'persistence' limit (*p-persit*) and its search 'patience' (number of visits) is not exhausted, the agent broadens its search to include the entire UA (or TA for newcomers – if preference for TA settlement is selected).
- During the search, the best co-ethnic proportion so far among those visited places is retained by the agent. The co-ethnic proportion of a potential (candidate) location is obtained by an imperfect 'sampling' process (which is proxied by a stochastic mechanism in the model). This reflects the fact that complete, transparent and accurate information about the ethnic proportion of a neighbourhood may not be accessible to householders who are looking for an ideal location to settle. If the discriminatory behaviour of an agent is satisfied; that is, if the co-ethnic percentage of a candidate (visited) location is more than the threshold tolerance of the agent, the placement process completes by moving the agent to this place. Otherwise, the agent compromises by moving to the best location (i.e. the one with highest co-ethnic preference) found so far during the placement search. In this sense, agents exhibit 'satisficers' behaviour (after certain visits they settle for a

‘good enough’ option, and are not worried about possibility that there might be something better).

4.2.5 *HAAMoS* and Characteristics of Schelling Model

Despite fundamental differences between the final model presented in this thesis and the original Schelling model and its extended versions, the *HAAMoS* model at its core is a Schelling type model. Therefore, it would be important to compare features of the original Schelling model or its descendant versions (i.e. later generations of Schelling-type models) with the *HAAMoS* model to see how and to what extent the *HAAMoS* model differs or compares with them. This is done in table 4.1.

Table 4.1 Schelling Models vs. *HAAMoS*

Schelling models’ feature	*HAAMoS* feature
“Schelling’s results are derived from an extremely small population” …“13 x 16 grid populated with 138 individuals” (Bruch & Mare, 2006, p. 670).	The population of selected area units and territorial authorities of the Auckland region for this study correspond to their respective census-based population sizes. However, population in each area unit is processed at aggregate level. Individual household agents are only created when they are forced to relocate (due to turnover process) or when they enter the urban area for the first time as new settlers (immigrants).
The Schelling “model is limited to only two race-ethnic groups” (Bruch & Mare, 2006, p. 670).	Up to four race-ethnic groups can be simulated in the model.
In the Schelling model, vacancy is applied uniformly as a fixed number or percentage (of vacant cells).	Vacancy rates can be established either uniformly as a fixed percentage; or dynamically (at each cycle) as realistic/ empirical rates applied either stochastically/ approximately or more precisely to area units within different territorial authorities.
The Schelling checkerboard model is a grid with many neighbourhoods in which each cell square could accommodate an individual agent and a neighbourhood of an agent located in a square is defined as the immediate surrounding squares (Moore). In other words, it is a continuous neighbourhood (see, for example, Bruch & Mare, 2006; Zhang, 2009).	*HAAMoS* consists of bounded neighbourhoods. Individual agents are supposed to live within area units (surrounded by other area units), but their exact locations are not defined and not considered a matter of concern.
The Schelling neighbourhood tipping model is a single bounded neighbourhood where an individual agent is considered either inside or outside of it, and not concerned with its immediate local surrounding/ neighbourhood ethnic composition (see, for example, Zhang, 2009).	*HAAMoS* urban environment is composed of bounded neighbourhoods. It is therefore similar to Schelling tipping model, but with multiple neighbourhoods.

"Schelling assumes that players move to the nearest satisfactory position" (Pancs & Vriend, 2007, p. 8).	Household agents – including first time settlers – can move to any location (which may be the furthest position) [or any location within a given TA, if this option has been enabled] within an urban area, although many others can only make local relocations (i.e. within immediate neighbourhood).
"In Schelling all unsatisfied agents simultaneously put their name on a list, which is then processed sequentially in some arbitrary order" (Pancs & Vriend, 2007, p. 8).	Agents simultaneously put their name on the list (pool), not because they are unsatisfied, but as a result of the turnover process (motivations for moves, although justifiable, are not a matter of concern and considered unimportant). Similarly, the list (pool) is processed in a random order.
Schelling "assumes inertia. That is, satisfied agents will always stay put, whereas it is not clear what happens with non-satisfied agents who cannot find a satisfactory position" (Pancs & Vriend, 2007, p. 8).	Agents do not stay or move because they are satisfied or unsatisfied with their current places. They are initially forced to move (for unspecified reasons, although reasons can include peculiarities of the location, financial problems, and job/ school related matters). They may be able to relocate to locations that fully correspond to their preferences, or they have to compromise and settle in the best visited locations. Although inconsequential, the rest of the (aggregate) population can be thought of satisfied people, yet it is similarly fine to think of a fraction of them as unsatisfied who cannot afford to move.
Schelling "distinguishes preferences expressed in absolute terms (number of like or unlike agents within a neighborhood [sic]) or relative terms (ratio of like to unlike neighbors [sic] within a neighborhood [sic])" (Pancs & Vriend, 2007, p. 7).	Ethnic composition preferences are expressed in relative terms, i.e. ratio of like-neighbours within the neighbourhood.
Schelling does not specify "preferences over completely empty neighborhoods [sic]" (Pancs & Vriend, 2007, p. 7).	A completely empty neighbourhood is not treated as the least preferred one, but as a satisfactory position – if it is visited and the agent estimation of a neighbourhood's ethnic composition happens to match its threshold or make it as the 'best' found choice.
Schelling used a "flat"/ "threshold" utility function. For instance, an ethnic group is satisfied as long as the adjacent neighbours belonging to other ethnic group remain below a flat threshold. (see, for example, Bruch & Mare, 2006; Pancs & Vriend, 2007). It is flat and straightforward in comparison to some 'compound' utility functions such as Benenson's "dissonance" function for example (Benenson et al., 2002).	In-group contact preference (tolerance) can arguably be seen as a flat threshold percentage (which can be different for each ethnic group) based on which a household agent decides to settle immediately in an area unit (if the neighbourhood like-group ratio is greater than her threshold). However, the agent eventually has to settle anyway (in the best visited location), even if her threshold condition has not been met. In this sense, it is only a dynamic indicator and not a purely 'flat threshold', *per se*.
In the Schelling model, ethnic preferences are homogenous within (if not also across) ethnic groups. (i.e. member of the same group have the same level of preference threshold for living in an own-group neighbourhood).	Agents who are members of the same ethnic group have the same level of preference (tolerance) for living in the neighbourhoods which correspond to their own-group ratio level. However, in addition to ethnic proximity preference, agents in *HAAMoS* have other inclinations too. These include the desire to move locally or globally, meso- geographical preference if they are first time

	settlers, how many places to visit before they eventually settle, and how much they persist before broadening their search area – if they are open to global moves. As these can be done differently across each ethnic group, the relatively large choice combination has heterogeneous effects.
The Schelling (model configuration) starts with either - “a random initial distribution of agents”, or - “reshuffling a perfectly integrated board ” (Pancs & Vriend, 2007, p. 6).	Experiments can start with both integrated and random distribution of agents. Many scenarios start with the realistic configuration based on census data.
In Schelling agents move until equilibrium is reached (e.g. all agents are happy/ satisfied, or no further more move is possible, e.g. “no pair of agents can increase their utility by trading residential locations” (Zhang, 2009, p. 18)). If the configuration at the initial stage is already at equilibrium, the final equilibrium state will represent a residential pattern different from the original equilibrium state.	Like most recent schelling-like models, *HAAMoS* simulations are limited to a fixed number of iterations (cycles), by which the model may not (often) be at equilibrium. On this aspect, it is similar to Fossett model which simulates 30 cycles as “they roughly correspond to time periods of 6–12 months” (Fossett, 2006a, p. 233).
In the Schelling checkerboard model, “agents only move inside the model and their moves generate residential segregation endogenously” (Zhang, 2009, p. 10). The Schelling tipping/ bounded neighbourhood model deals with a single neighbourhood and “although agents […] are assumed to move into and out of the neighborhood [sic], the model does not specify where they come from nor where they go” (p. 9).	In the bounded multi-neighbourhoods environment of *HAAMoS*, agents move inside the model and segregation can be generated endogenously. At the same time, other agents can move into and out of urban areas. In such way, it would be possible to influence the residential patterns and potential segregation exogenously. This is done similar to Benenson et al.’s model (2002) which uses a fix flow number for in and out migration.

4.3 Discussion and Conclusion

This chapter has briefly introduced the *HAAMoS* model (and its design concepts) which is the main focus of the remainder of this thesis. Some of the hybrid elements of the model (explained in previous chapters) are expressed via the key processes examined in this chapter. For example, the *turnover* process is driven in a similar way as in microsimulations (MSM) by probabilities applied to ‘units’ of populations, while the *placement* process defines the rules based on which individual decisions are made (as in ABMs). The update of layers and visual display process operationalize the geosimulation feature of the model, whilst the *inflow* process enables exogenous activities which along with typical endogenous actions in the model take place in a bounded multi-neighbourhood environment (as in a mix of Schelling tipping and checkerboard models).

Although a Schelling-type model in its core fabric, the 'magnitude' of the *HAAMoS* model (even in its current version) combined with possible geographical types (artificial and realistic) and scales along with empirical validation objectives present both exciting opportunities and formidable challenges for detailed examination and analysis.

The model can measure different dimensions of segregation, and with validation based on empirical data will provide additional perspectives for the investigation of relationships between behavioural processes and structural patterns that emerge. Despite the methodological and procedural sophistication employed by this model, the decision-making process of agents is still far less sophisticated (if not still primitive) than that in a real world situation. Thus generating patterns similar to the empirical measure with different data perspectives (i.e. different empirical-based measures and for different geographical scales) all at the same time would perhaps be unattainable.

Bearing these challenges in mind, chapter 7 will demonstrate that the model is able to generate plausible patterns fairly similar to several empirical benchmarks at different geographical scales at the same, although understandably not all of them will be conforming.

In the next chapter, the descriptions of the segregation measures used by this model will be presented followed by presentation of some theoretically-based simulations.

Chapter 5

5. Theoretical Experiments

5.1 Introduction

Chapter 4 presented the design of the simulation framework, particularly in the light of Schelling-based models. The aim of this chapter on the other hand is to test *HAAMoS* model behaviour and to demonstrate the extent to which it would conform/ exhibit the behaviours of Schelling-type models. This is done through a comparison of the characteristics of two prominent models – the original Schelling model and the Fossett model (as one of the recent and most sophisticated successors) – whilst exploring the effects of some of the key elements of the parameter space. As in Schelling and Fossett models, these experiments will be conducted in an artificial (grid-based) environment (notwithstanding that agents are in bounded neighbourhoods where the relationship between grid-cells and individual agents' geographical references [residences] is not necessarily one-to-one) without the involvement of exogenous elements (e.g. immigration) in order to test and explore some theoretical hypotheses. Before doing so, we need first to see the details of the selected measures which are implemented and used by the model to determine different facets of segregation.

5.2 Measures and Dimensions of Residential Segregation

Figure 5.1 shows a range of spatial pattern occurrences (from 'integrated' to 'segregated' urban area) which can usually be observed in cell-based simulation models with two different groups (here identified by blue and yellow colours). The question that arises is would it be possible to measure the variability in observed patterns as truthfully as possible.

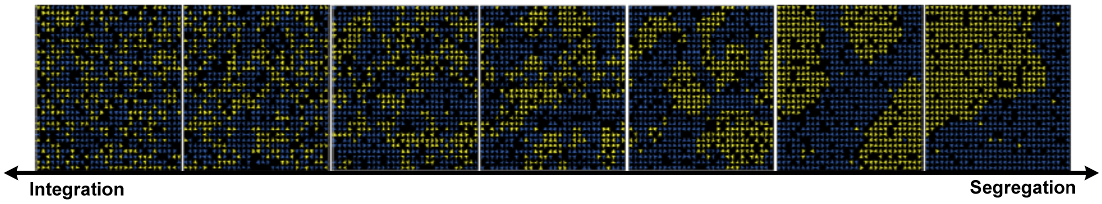


Figure 5.1 Various Manifestations of Spatial Patterns

Measurement of residential segregation is a complex matter. Finding or developing techniques that conceptually, methodologically and empirically are suitable is challenging. At the same time, reliable and meaningful measurement of residential segregation is essential to the study of the causes and patterns of racial and socioeconomic segregation (Reardon & O'Sullivan, 2004). Over a number of years a large number of indices have been proposed to quantify segregation. The literature on segregation measurement offers a bewildering array of proposed indices and an equally extensive literature examining or criticizing these indices (Reardon, 2006).

Table 5.1 Description of Segregation Dimensions

Dimension	Description	Relationship with Segregation
Evenness	It measures differential distribution of a specific ethnic group over a specific or entire geographical area	Evenness is minimized and segregation maximized when none of the two groups' members share a common spatial area.
Exposure	It measures the degree of potential inter-ethnic contact (likelihood of interaction or isolation) of a group with others within a neighbourhood (i.e. extent to which different ethnic group members come into contact with one another and share a common residential area)	Low levels of exposure mean high levels of segregation
Concentration / Isolation	It measures the physical space that is occupied by an ethnic group in a geographical area (i.e. to what extent the group is concentrated in relatively small residential areas)	If the physical space occupied by an ethnic group (e.g. minority) is relatively small, segregation would be high
Clustering	It measures the extent to which the residential areas inhabited by minority members adjoin (crowd) one another in space	A high degree of clustering implies a residential structure where ethnic group areas are contiguous and closely packed, creating a single large ethnic enclave (high level of segregation)
Centralization	It measures the proximity of a residential distribution to the urban core, (i.e. degree to which members of an ethnic group spatially live near to the centre of an urban area)	levels of centralization denote high levels of segregation

In the midst of these divergent opinions and continued waves of new or alternative measures, Massey and Denton (1988) conducted a seminal study that recognized different indices used to measure different dimensions of segregation. Their conceptualization highlighted five key dimensions of segregation, which are described in detail in table 5.1.

Of these five dimensions, centralization which measures the proximity of residential distribution to the centre of an urban area has attracted less interest , as these days large sprawl metropolitan areas (such as the Auckland region) are increasingly polycentric/ multi-nodal. Moreover, Reardon and O'Sullivan (2004) argue that "centralisation and concentration dimensions can be seen as specific subcategories of spatial unevenness". They propose two conceptual dimensions of spatial residential segregation: exposure/ isolation and evenness/ clustering.

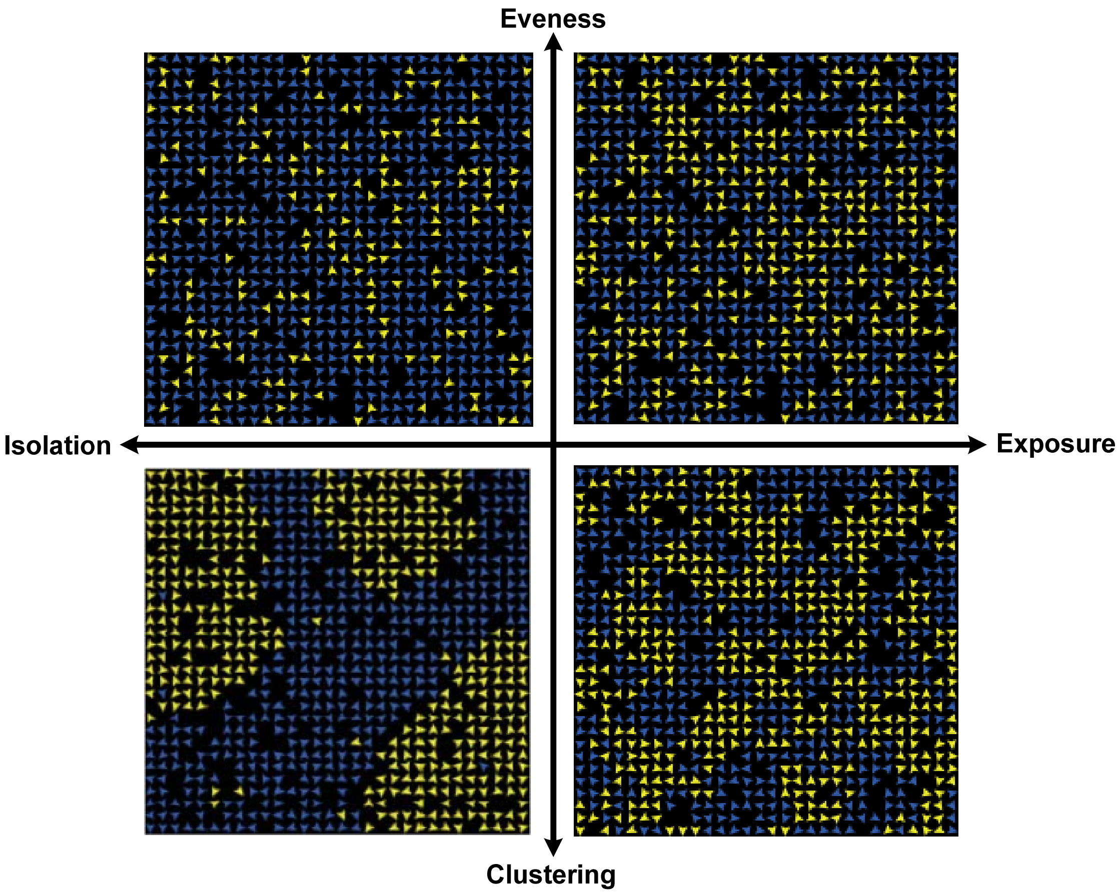


Figure 5.2 Four Dimensions of Segregation

Figure 5.2 attempts to capture the conceptualisation of Reardon and O'Sullivan's (2004) spatial dimensions. They reason that evenness and clustering are spatially related and both are independent of population composition and refer to the extent to which groups are similarly distributed in a geographical area. Ethnic movement of agents from one area to another can create underrepresentation in one area and overrepresentation in another (hence damaging evenness), whilst at the same time create or increase the clustering in the overrepresented area. On other hand, exposure/ isolation depends on the overall ethnic composition of the residents in the spatial unit under study, as it refers to the probability that members of the same group share (/interact) in the space with a similar or another ethnic group.

Reardon and O'Sullivan's association (or rather simplification) of the segregation dimension implies that fewer fine measures than initial numbers of conceptual segregation dimensions would be sufficient to get a good indication of segregation intensity.

In most studies, including those using simulation approaches, only few measures of segregation have been used overall, and the dissimilarity index (D) which measures the evenness dimension remains the most utilized among them (Acevedo-Garcia et al., 2003, p. 217). However, the model in this thesis employs a combination of indices that can measure at least two major axes of the segregation dimension simultaneously. In fact, this simple feature does not seem to have been previously employed in similar studies. Details of these measurements are explained in the following subsection.

5.2.1 Selected Measures

Table 5.2 summarizes the list of measures that are used by the model to operationalise different facets of the segregation dimension and related aspects. Chapter 7 explains that one of the most challenging part of the validation and calibration of any model is related to the 'multiple realizability' of its outputs. This thesis argues that using multiple measures which gauge different (or even comparable) dimensions of segregation in a multi-level modelling environment (such as this modelling simulation framework) would be beneficial as it will result in better calibration of the model –

even if the model outputs will not match the best the empirical benchmarks. This is because using multiple measures would eventually lead to the reduction of realizable space by allowing detection and elimination of those 'unfruitful' configurations (i.e. 'bogus fits'). For example, using binary feature of D index (along with the multi-ethnics measure H) proved to be helpful in better calibration of the model by comparing D-values across four ethnic groups. For these reasons, this thesis uses and advocates employing multiple measures.

Complete notations of variables, parameters and measures can also be found in appendix B. In the formulas presented in this section, larger (meso/ macro) geographical areas for which a measure is calculated are implicit in the formulas (i.e. they do not appear in the signature of the formula as a parameter, but can be thought of as if the UA/ TA reference is passed when such a function is called).

Table 5.2 Operationalization of Measurements

Measurement category	Operationalization
Segregation	
Inter-group measures	
Evenness	Index of dissimilarity (**D**); Info. theory (**H**)
Exposure	Index of interaction (**I.I.**$_{xy}$)
Isolation	Index of isolation (**I.I.**$_{xx}$)
Geographical measures	
Global clustering	Global Moran's I (**GMI**)
Local clusters	Local Moran's I (**LMI**)
Concentration	Location quotient (**LQ**)
Other geographical indicators	
Spatial plurality metric	Plurality index (**PI**)

Index of Dissimilarity (D)

Allowing only for two groups, this index compares the spatial distribution of an ethnic group (e.g. majority, g_x) versus another (e.g. minority, g_y). It measures the evenness with which two ethnic groups in the model are distributed across the metropolitan area or other meso-geographical entities (i.e. TAs) defined in the model. It is calculated by the following formula (see, for example, White, 1983):

$$D = \frac{1}{2}\sum_{u=1}^{U}\left|\frac{\varphi_y(u,t)}{\Phi_y(t)} - \frac{\varphi_x(u,t)}{\Phi_x(t)}\right| \qquad (5.1)$$

where:

- **U** is the total number of area units in the larger geographical area for which D is being calculated
- φ_x(u,t) is the population of group **x** (majority) in the **u**th area unit at time **t**
- φ_y(u, t) is the population of group **y** (minority) in the **u**th area unit at time **t**
- Φ_x(t) is the total population of group **x** (majority) in the large geographical area at time **t** for which D is being calculated
- Φ_y(t) is the total population of group **y** (minority) in the large geographical area at time **t** for which D is being calculated

The index of dissimilarity can vary from 0 (even distribution) and 1.0 (maximum segregation/ uneven distribution).

Information Theory Index (H)

The information theory index[12] (H) measures the variation in diversity across area units, where the ethnic diversity of a population is defined as the (extent of) 'entropy' of the population (Reardon, 2006). With two groups entropy reaches a maximum when both groups are of equal size. The entropy of a large geographical area (here: entire metropolitan area or a territorial authority) is the extent of its ethnic diversity which can be defined by the following formula (see, for example, Reardon & Firebaugh, 2002):

$$E = \sum_{g=1}^{G} \Pi_g(t) \ln \frac{1}{\Pi_g(t)} \qquad (5.2)$$

[12] Information theory index (H) also known as the information index, the entropy index, the Theil index or Theil's H, after its originator who originally proposed it as a measure of school segregation. Its multigroup version is sometimes explicitly referred to as the multigroup entropy index, the multigroup version of Theil's H, or the multigroup information theory index.

where:

- **G** is the total number of simulated groups
- **$\Pi_g(t)$** is the proportion of the group **g** in the large geographical area at time **t** for which **E** is being calculated

A unit's entropy is analogously calculated by (see, for example, Massey & Denton, 1988):

$$E_u = \sum_{g=1}^{G} \pi_g(u,t) \ln \frac{1}{\pi_g(u,t)} \qquad (5.3)$$

where:

- **$\pi_g(u, t)$** is the proportion of the group **g** in the area unit **u** at time **t**

The information theory index can be seen as a measure of unevenness, measuring departure from evenness by assessing each area unit's departure from ethnic entropy of the larger spatial entity. The index can be defined as the weighted average deviation of each area unit's entropy from the larger 'container' geographical area entropy, expressed as a fraction of the larger geographical entity's total entropy (Massey & Denton, 1988, p. 285):

$$H = \sum_{u=1}^{U} \frac{\varphi(u,t)\,(E-E_u)}{\Phi(t) * E} \qquad (5.4)$$

where:

- **U** is the total number of area units within the larger geographical entity for which H is being calculated
- $\varphi(u,t)$ is the population of the **u**th area unit at time **t**
- $\Phi(t)$ is the total population of the large geographical area at time **t** for which H is being calculated

The information theory index can vary from 0 (when all areas have the same composition) to 1.0 (maximum segregation/ when all areas comprise only one group).

Interaction/ Isolation Index (I.I)

This index measures the exposure of two groups as a dimension of segregation. There are two basic measures of residential exposure. First, is the degree of exposure of a group of household agents to another ethnic group; that is, the probability that members of group **x** share a spatial unit with members of group **y**. This is often called the index of interaction and can be defined – based on Bell's (1954) and Lieberson (1980, 1981) – as follows:

$$\mathrm{I.I_{xy}} = \sum_{\mathrm{u}=1}^{\mathrm{U}} \left(\frac{\varphi_x(\mathrm{u,t})}{\Phi_x(\mathrm{t})}\right)\left(\frac{\varphi_y(\mathrm{u,t})}{\varphi(\mathrm{u,t})}\right) \qquad (5.5)$$

where:

- **U** is the total number of area units within the large geographical entity for which I.I is being calculated
- φ_x(u,t) is the population of group **x** in the **u**th area unit at time **t**
- φ_y(u,t) is the population of group **y** in the **u**th area unit at time **t**
- φ(u,t) is the total population (all groups) in the **u**th area unit at time **t**
- Φ_x(t) is the total population of group **x** in the large geographical area at time **t** for which I.I is being calculated

Second, is the extent to which a group of household agents are exposed to (only) one other (rather than others); that is, the probability that members of group **x** share a spatial unit with alike people, which is often called index of isolation. It can be defined as a weighted average of each unit's group proportion:

$$\mathrm{I.I_{xx}} = \sum_{\mathrm{u}=1}^{\mathrm{U}} \left(\frac{\varphi_x(\mathrm{u,t})}{\Phi_x(\mathrm{t})}\right)\left(\frac{\varphi_x(\mathrm{u,t})}{\varphi(\mathrm{u,t})}\right) \qquad (5.6)$$

Both of these indices vary between 0 and 1.0. Since the interaction index depends on the composition of populations – it is asymmetric – which means that in general $I.I_{xy}$ does not equal $I.I_{yx}$. When two groups have the same proportion of the population, the indices would be equal. In general, the sum of all interaction indices plus the isolation index will equal 1.0. For example, when there are only two groups $I.I_{xx} + I.I_{xy} = 1.0$.

Moran's I

Moran's I is a measure of spatial autocorrelation, measuring spatial dependence and correlation among nearby locations in geographical space.

Moran's I can be separated into two indices, one for measuring the clusters at local level (local spatial autocorrelation or local Moran's I) and the other for measuring the global clustering (global autocorrelation or global Moran's I).

Local Moran's I can be classified and used as a 'Local Indicator of Spatial Associations' (LISA) measure, i.e. a statistic that "gives an indication of the extent of significant clustering of similar values around [an] observation" (Anselin, 1995, p. 94), whilst the sum of these statistics for all observations would be proportional the global indictor of spatial association (here global Moran's I). It can be used as an "indicator of local packets of inconsistencies" (hot spots) as well as a tool in assessing the influence of individual local locations on global Moran's I statistics for identification of 'outliers' (Ibid).

Local Moran's I (LMI) allows for the decomposition of global Moran's I into the contribution of each observation, where each observation **i** is measured as follows:

$$\mathrm{LMI} = \mathrm{I_i} = \frac{\mathrm{Z}_i \sum_{j=1}^{U}(w_{ij}\,\mathrm{Z}_j)}{\mathrm{m}_2} \qquad (5.7)$$

$$\mathrm{m}_2 = \frac{\sum_{i=1}^{U} Z_i^2}{U} \qquad (5.8)$$

where:

- **U** is the total number of area units within the large geographical entity for which $\mathbf{I_i}$ is being calculated

- **Z** is the deviation of the variable of interest (x) with respect to the mean ($\bar{x}$), i.e. $Z_i = (x_i - \bar{x}_i)$; where $\bar{x}_i = \frac{\sum_i x_i}{U}$, and the variable of interest for Moran's I calculation in this model can be set to either population of group **g** in area unit **i** at time **t**; i.e. $\varphi_g(u,t)$, or the proportion of group **g** in area unit **i** at time **t**; i.e. $\boldsymbol{\pi}_g(u,t)$

- w_{ij} (where $w_{ii} = 0$) is an element of the spatial weight matrix that indicates the spatial connectivity between area units of $\boldsymbol{i}$ and $\boldsymbol{j}$. If the observation $\boldsymbol{i}$ is neighbour of (/ adjacent to) observation $\boldsymbol{j}$, then $w_{ij} = 1$, otherwise $w_{ij} = 0$

The global Moran's I (GMI), as a global spatial statistics about the entire geographical area under the study (observation) can be defined as:

$$\text{GMI} = \text{I} = \frac{\sum_{u=1}^{U} \text{I}_i}{\text{S}_o} \qquad (5.9)$$

$$\text{S}_0 = \sum_i^U \sum_j^U w_{ij} \qquad (5.10)$$

GMI-values can vary between -1.0 to 1.0. A negative autocorrelation indicates the juxtaposition of high-values next to low-values (dispersion), whilst a positive autocorrelation (denoted by positive values) indicates the existence of both high-value clustering and low-value clustering (correlation) (Anselin et al., 2000; Haining, 1990).

Although the model is capable of calculating both local and global Moran's I values, LMI-values are not used and analyzed in the simulated scenarios presented in this thesis. There are more than 300 local area units in the selected territorial authorities of the Auckland region, which would make it impractical to report the major changes in this study. However, section 6.3.3 will present a method that in part uses LMI-values to identify the 'hottest' and 'coldest' area units for the Asian ethnic group during the last three censuses.

Location Quotient (LQ)

The location quotient can be used as a way of measuring and comparing the relative concentration of an ethnic group in a small (micro/ meso) geographical area to the relative concentration of the same group in a larger (meso/ macro) area. For example, it would be possible to relate the proportion of the various ethnic groups at metropolitan area (or territorial authority) level to their concentration or underrepresentation in area units (see, for example, Agyei-Mensah & Owusu, 2010). In other words, it indicates the degree to which the area unit departs from the overall proportion of a given group in the larger urban area.

Location quotient (LQ) at location **u** can be defined as:

$$\mathrm{LQ}_u = \frac{\varphi_g(\mathrm{u,t}) \;/\; \Phi_g(\mathrm{t})}{\varphi(\mathrm{u,t}) \;/\; \Phi(\mathrm{t})} \qquad (5.11)$$

where:

- φ_g(u,t) is the population of group **g** in the **u**th area unit at time **t**
- φ(u,t) is the total population of the **u**th area unit at time **t**
- Φ_g(t) is the total population of group **g** in the large geographical area at time **t** for which QL is being calculated
- Φ(t) is the total population of the large geographical area at time **t** for which QL is being calculated

The quotient is a ratio of the ethnic group's population percentage in the area unit relative to its population percentage in the territorial authority or the metropolitan area. A QL = 1 indicates that the area unit has exactly the same relative representation of an ethnic group as can be found in the larger geographical area (e.g. TA or MA). A value lower than 1 indicates under-representation of the ethnic group (low concentration) in the area unit, whilst a value greater than 1 signifies the group is overrepresented (higher concentration).

Although the model is capable of calculating LQ-values for each area unit in the metropolitan area, the LQ (as in the case of the LMI) is uniquely used in this thesis to identify 'hottest' and 'coldest' area units for the Asian ethnic group during the last three censuses presented in section 6.3.3.

Plurality Index (PI)

The plurality index is used to calculate the number of spatial units (enclosed within the larger geographical area for which the index is being calculated) where a given ethnic group is in plurality (relative majority) position. If, for example, in a GIS map each area unit is entirely painted with a specific colour designated to that ethnic group when the group is in plurality, this index will represent the number of area units in the map having the same colour.

Plurality index (PI) for group **x** can be defined as:

$$\mathrm{PI} = \sum_{u=1}^{\mathrm{U}} \mathrm{p_x}(u, \mathrm{t}) \qquad (5.12)$$

where:

- p_x(u,t) is the plurality status of group **x** in the **u**th area unit at time **t**. If the group x is in plurality in the area unit **u** at time **t**, then p_x(u,t) = 1, otherwise p_x(u,t) = 0

The plurality index ranges from 0 to **U**, where **U** is the maximum number of spatial (area) units contained in the macro geographical area for which PI is being calculated. Higher index values are normally indicative of higher level of segregation.

5.3 Testing the Model's Conformance

In the following sections, a series of experiments is used to demonstrate that when the model is configured to typical Schelling and Fossett model settings – it will exhibit similar behaviours – in spite of the essential differences that exist between *HAAMoS* and the Schelling and Fossett models. This establishes a point of departure for discussion and analyses of the emerging (segregation) outcomes by placing the model in a known paradigm that encompasses Schelling-like models.

In this section we refer to the Schelling and Fossett models in a 'narrow way', and distinguish between them in a subtle way according to their initial settings. First, we only engage two ethnic groups (g_1 and g_2). In the Schelling setting, we consider that the results in the model do not depend on the proportion of the population. So, as in the Zhang model (2009), the number of people in two groups can be chosen exactly (or roughly) the same (i.e. 50 – 50). In other words, there is no minority or majority group.

On the other hand, in the Fossett configuration, the minority group (and its preference) is the key feature (Goering, 2006) and we will examine the model's behaviour by distinctively creating two groups, namely the minority and majority and examining whether "minority preference regarding neighbourhood ethnic mix are unambiguously segregation-promoting" (Fossett, 2006a, p. 200). It is important to mention that among the many models built on the basis of Schelling's model – either by extending his original work or by loosening some of the initial constraints and assumptions – Fossett's model is actually one of the most sophisticated. This is because he included preferences for housing quality and neighbourhood status and examined how these preferences are implicated in the production of patterns of residential segregation (Skvoretz, 2006). It is true that from a different perspective and for some other people Fossett's model can still be considered "simplistic" and relatively "one-sided" (Goering, 2006), particularly because of his emphasis on the role of minority preference in creating (or sustaining) segregation. This is exactly the feature that has been captured in a 'caricature' way for the Fossett-like experiments.

Table 5.3 clearly shows the values of key variables which set apart Schelling and Fossett in these demonstrative experiments. The initial ethnic composition population setting for the Schelling experiment in each area unit will be 50-50 for each group (i.e. equal size groups), whilst it will be set at 80-20 for the Fossett experiment; that is, a main majority group (g_1, blue) and a small size minority group (g_2, red). Also, in the

Schelling setting each moving agent from either ethnic group wants to resettle in a neighbourhood where at least half of the residents are alike (i.e. $T_1 = T_2 = 0.50$). For the Fossett setting, we will vary the in-group contact preference of the minority ethnic group to show its effects, whilst the majority ethnic group are often totally neutral (i.e. $T_1 = 0$: no preference regarding the ethnic composition of neighbourhood). We will, however also show what would be the effects with higher value ($T_1 = 0.50$).

Table 5.3 Setting Differences in Schelling and Fossett Experiments

Model Setting Type	Variable		g_1	g_2
Schelling	Initial proportion	**Π (t=0)**	50	50
	Tolerance	**T**	0.50	0.50
Fossett	Initial proportion	**Π (t=0)**	80	20
	Tolerance	**T**	0 or 0.50	0.10 or more

As in typical Schelling-based models, the environment in which these experiments will be conducted is grid-based, although the grid-cells are not usually linked to a single individual residency, but can be seen as aggregates of many agents' geographical references. The most important common variables are shown in table 5.4. The setting values for the main experiment are purposely chosen to keep these experiments and the environment simple and more importantly similar to their original configurations.

Table 5.4 Common Settings for Schelling & Fossett Experiments

Variable/ Parameter		Value
G	Number of groups : (g_1 and g_2)	2
$P(M)_g$	Turnover probability percentage, for each ethnic group, **$P(M)_1 = P(M)_2$**	0.2
$P(GL)_g$	Global location move preference probability, **$P(GL)_1 = P(GL)_2$**	1.0
MX_F_i	Inflow (immigration)	0
MX_F_o	Outflow (emigration)	0
$P\text{-}PERSIT_g$	Probability of persistence in same area before moving to another (g_1=g_2)	0.50
λ_{VISIT}	Mean for maximum number of visits during search	12
σ	SD for accuracy of co-ethnic% stochastic estimation	0.07
V_{pc} (t)	Initial vacancy percentage applied at t=0 only	10%
MINv	Minimum number of vacancies in area unit at t=0	100
t_s	Stop tick time	30
RS	Random seed	1 - 10
U	Number of area units (using Artificial Env./ Grid-based)	289 (17 × 17)
φ_g (u,t); @ t=0	Total population in each area unit (cell) at t=0	1,000
Φ (t)	Total population in the world at all the time ($0 \le t \le t_s$)	289,000

For instance, as indicated in the table, there is no inflow (immigration) or outflow (emigration) of population into or out of the simulated world. The agent population of each ethnic group has 10% turnovers at each cycle (if not specified otherwise), and the probability of global locational move preference P(GL) [for simplicity is noted as L) is set to 1.0. This is in order to allow all the agents to be able to move anywhere in the city (similar to Schelling and Fossett models).

At t = 0, each area unit contains a population of 1000 (both group combined), although depending on the experiment the proportions of ethnic groups are different. Also at t = 0, with initial vacancy set at 10%, each area unit has 100 vacant spaces. With 289 cells in the grid (of 17 × 17), the total population in the grid city will be 289,000 households (along with a total of 28,900 vacant spaces). For each of these experiments, 10 separated simulations will be conducted, each time using a different random seed. As the variation among simulated outcomes are relatively small (as will be shown in the next section), a larger number of separated simulations is not required.

The probability of persistence for each member of these two groups search in the same location before looking elsewhere is set equally to 0.50.

Unless a matching location is found, each agent would visit an average of 12 locations (Poisson distributed) before it would finally settle in the best (fitting one) among them. The estimation of co-ethnic percentage of a given location would be obtained stochastically by agents by an approximate value close to the real percentage with a standard deviation of 0.07. The values of these two latter parameters (λ_{VISIT} and σ) will be kept constant throughout the simulations presented in this thesis (for a complete list of notations and descriptions of variables and parameters, see appendix B).

5.3.1 Schelling-like Exhibit

The starting state of the Schelling model is typically highly integrated (Pancs & Vriend, 2007). At the time of initialisation (t = 0), there are equal numbers of agents belonging to two ethnic groups in each area unit. This can be seen in figure 5.3, where pie-charts (on each grid's cell) have the same shapes (that is, the same proportion for each ethnic group and the same ratio for vacant spaces). Moreover, area units at t = 0 are neither painted blue (colour of group 1) nor red (colour of group 2), but pink to indicate that no group is in a relative majority (plurality) position.

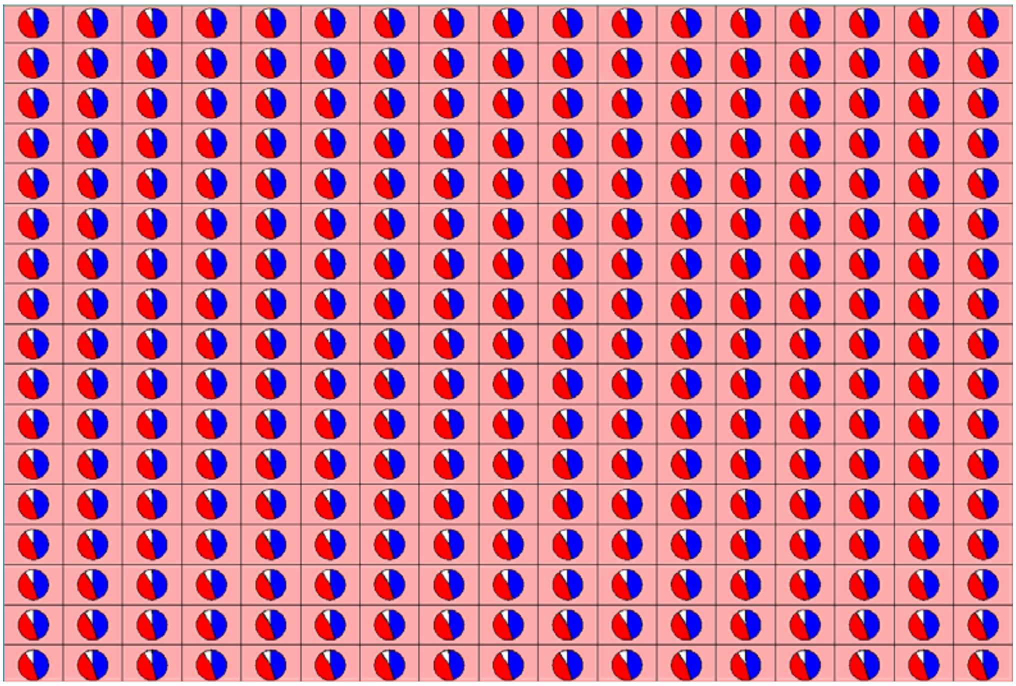


Figure 5.3 Schelling Test: Initialisation State (t=0)

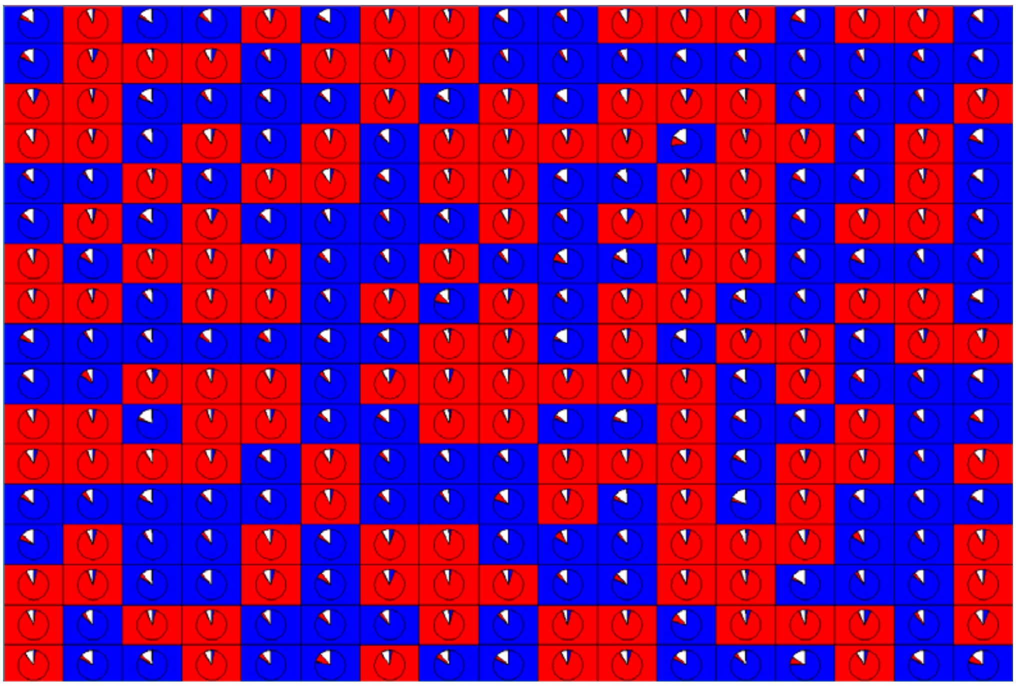


Figure 5.4 Schelling Test: End of One Simulation (t=30)

The simulations are repeated ten times using different seed numbers (same seed numbers are used for Fossett-like experiments), each until t = 30. Figure 5.4 shows the 'world' at the end of one of ten simulations (i.e. at t = 30). All the cells (area units) have now turned either blue or red depending on the group that forms the relative majority (plurality).

The pie-charts show the proportion of the population and can indicate the presence of another ethnic group, although in this case at the end of simulation, the presence of another group (minority) in the areas claimed by the majority group is small.

These snapshots (along with the measures which will follow) demonstrate how an unrealistic 'world' in equilibrium (perfect integration) between two existing groups who wish to live alongside each other (50:50 co-ethnic preference) can unravel to a state of intense segregation, as initially demonstrated by Thomas Schelling.

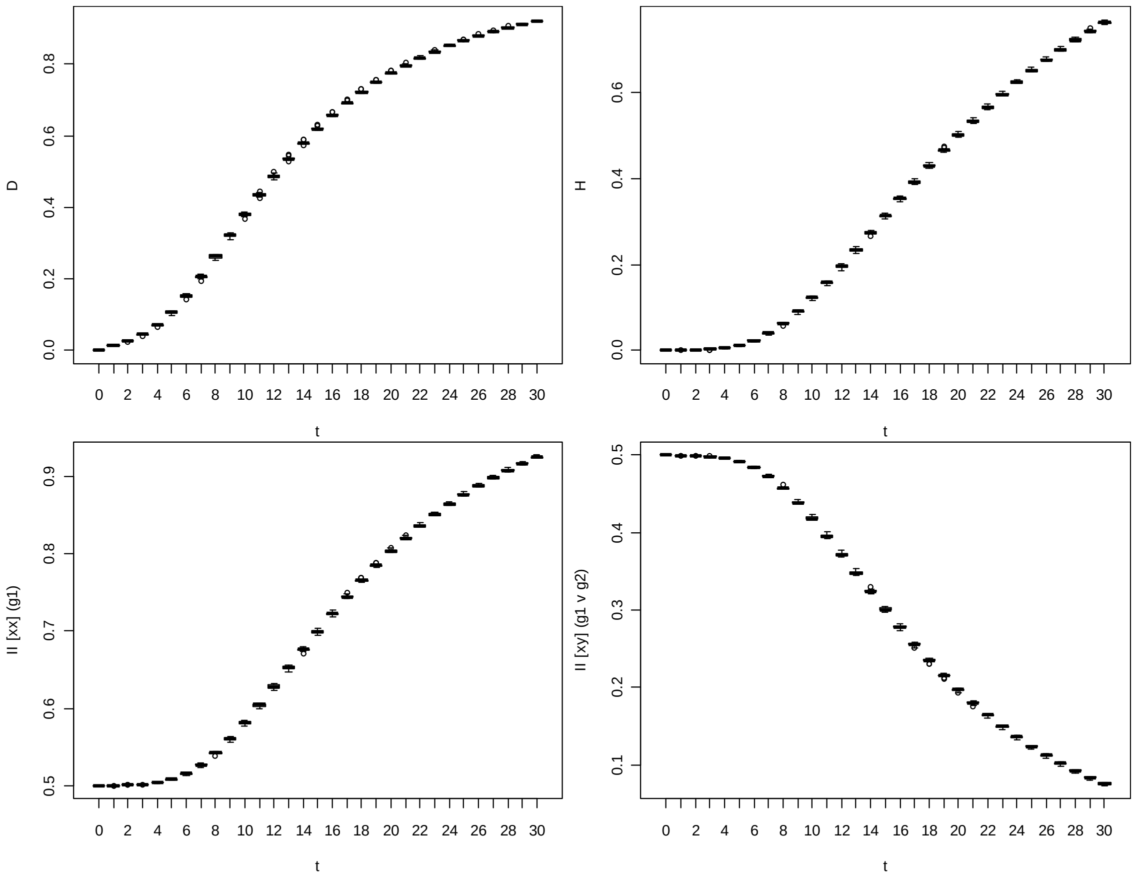


Figure 5.5 Schelling Test: D, H and I.I-values (10 runs)

Figures 5.5 shows that both the index of dissimilarity (D) [top left] and information theory index (H) [top right] have increased gradually from initial integration status (D

= H = 0) to a high level of unevenness (about 0.90 in case of D and less than 0.80 in case of H).

The box plots show the variation of values in 10 different runs (using different random seeds). The whiskers extend to the lowest and highest values within one-and-a-half interquartile ranges (IQR), while any other value outside these ranges are considered outliers and marked individually with small points (the fences mark the end values for the whiskers). The median line (bare) on the boxes can reveal whether the values are right or left skewed (these are more noticeable in figure 5.6).

The variation in the choices of agents' locations which may be dissimilar at different runs can affect the distribution of groups and hence the values of the measures. Overall, the level of variations in these simulations appears to be within an acceptable range (exhibiting somewhat steady patterns) so that additional number of simulations (with different random seeds) does not seem to be necessary.

Figure 5.5 (down left) also shows that the isolation index has increased from initial 50% to about 90%. This is not surprising from the situation captured in the final snapshot of the simulation (figure 5.4) which indicates that higher concentrations of the same group's agents in specific areas of the grid environment has dramatically increased the chance of own-group contacts over time.

On the other hand, the interaction index (see figure 5.5, down right); that is, the chance of meeting other group members, has dropped from 50% to about 7% over the same period. Similarly, the rather balanced desire to live in a totally mix (50-50) neighbourhoods has not stopped the gradual but rapid build-up of segregation between these two groups so that their level of interaction has deteriorated extensively over the simulation period. As noted by Schelling, the underlying motivations of these agents are still far less extreme than the observable patterns of separation.

The variation for the global Moran's I values appears to be wider as it is shown within a narrower oscillation range (see figure 5.6).

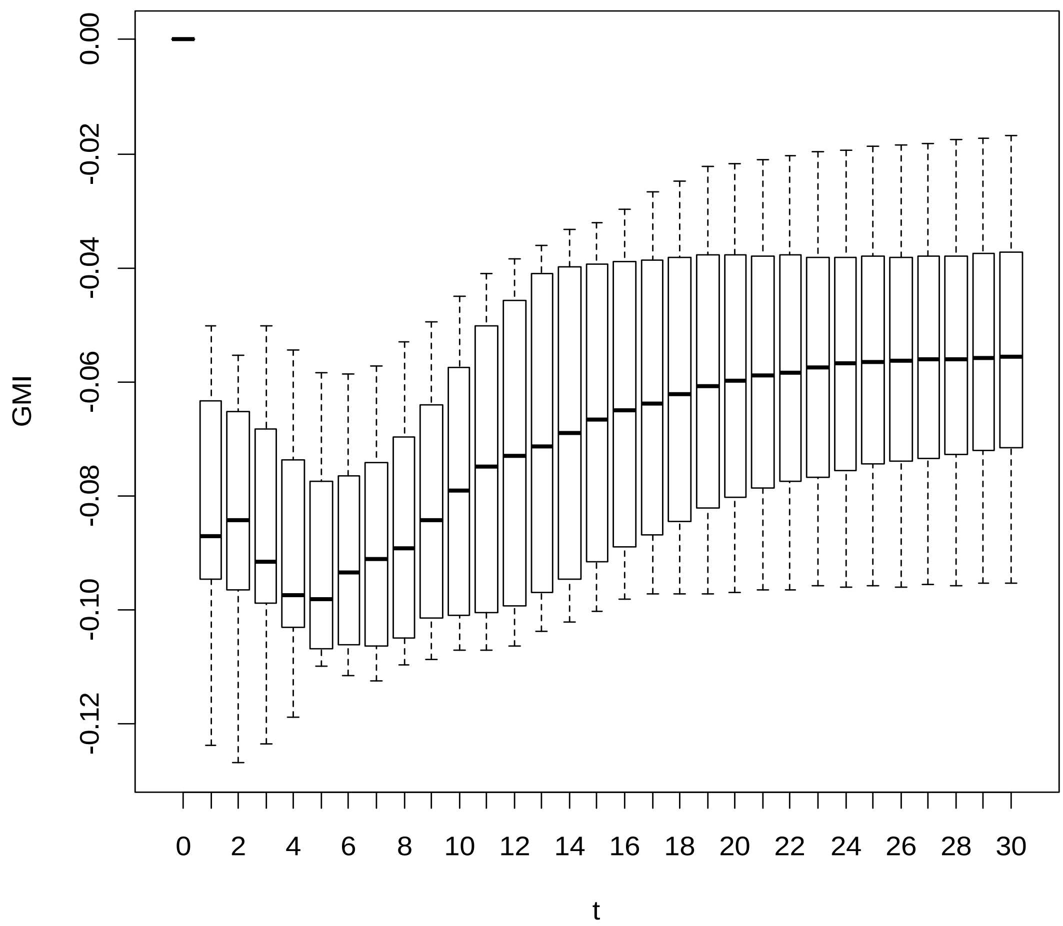


Figure 5.6 Schelling Test: GMI-values

With only two groups and the proportions used as variable of interest, the GMI-values are the same for both groups. Initially at zero, GMI drops to negative territory (as low as about -0.12), indicating that some level of dispersion has been caused by the occurrences of high-values of concentration (proportions) next to low-ones at the macro level (i.e. entire grid structure). Despite the variation, the median values indicate modest and gradual reduction of dispersion before it becomes more stabilised at about the last third period of the simulation during which the agents' movements in the already segregated 'world' hardly lead to substantial change measured by the GMI. At the beginning – with a perfect equilibrium – the next initial steps will see the disparity aggravated and momentum built up so that segregation unravels rapidly. Although the segregation is high as measured by H, D and I.I values, the (negative) autocorrelation remains somewhat within a moderate range.

With complete equilibrium at the beginning of the simulation, more than 90% of agents move to locations that match their co-ethnic percentage preferences (while about another 8% find them within their current locations and only a small fraction by moving to the best compromise locations) although more than 80% of them inaccurately estimate the co-ethnic compositions of the new locations. As the unevenness becomes more and more pronounced throughout the cell-based world, the percentage of estimation error gets gradually smaller (similarly in the real world the perception of a neighbourhood type, such as the ones described in table 2.1, is formed by the 'size' of the group for which the neighbourhood is known), so that by time t = 15, the percentage of mistaken estimations becomes negligible. In the same period, the percentage of agents who found matching locations fell about 5% in favour of those who remained in their current locations (i.e. seen as finding a suitable house in the same cell area).

But more importantly, the value of *P-Persist* parameter plays an important role in the dynamic of segregation as measured by GMI. This is explained in the next section.

Partial Exploration of Parameter Space (with Schelling model)

The parameter space of this modelling framework is relatively large and its exhaustive exploration would be beyond the main focus, objectives and limits of this thesis. Nevertheless, several instructive cases will be presented here to evaluate and demonstrate the effects on the simulation outcomes of changes to some of the key variables. These changes are implemented by keeping all the other variables fixed whilst changing (partially/ selectively) the key variable under evaluation.

Many of these (partial) parameter space explorations will be done using the Fossett configuration for it is a somewhat more realistic environment with asymmetrical population sizes. However, variation of *P_Persist* value is more apposite within the Schelling configuration as the change of value and outcomes will be applicable to both groups (having the same size).

Effects of *P-Persist*

In this experiment, the initial Schelling configuration (presented in the previous section, with *P_Persist* = 0.50) is compared to two farthest cases of 0.10 and 0.90. Unlike the probability for global move capability (to be examined in the next section) which can also be seen as a proxy for an economic factor (e.g. ability to purchase a home and move to anywhere in the world/ city), *P_Persist* has not been firmly defined in this thesis. However, it can be regarded and used as a parameter that accommodates the persistence (perseverance) that householders will put into finding a house within a particular neighbourhood even though this might be at odd with their financial ability. Such persistence may be because of cultural reasons (e.g. ties and proximity to other family members), lifestyle and preferences not related to racial/ ethnic composition of neighbourhoods (e.g. types of house, reputation of neighbourhood, reputation of a public school nearby in the catchment zone, proximity to work), or simply because they (or their realtors) are not as active as others in broadening their search horizons to other areas/ neighbourhoods.

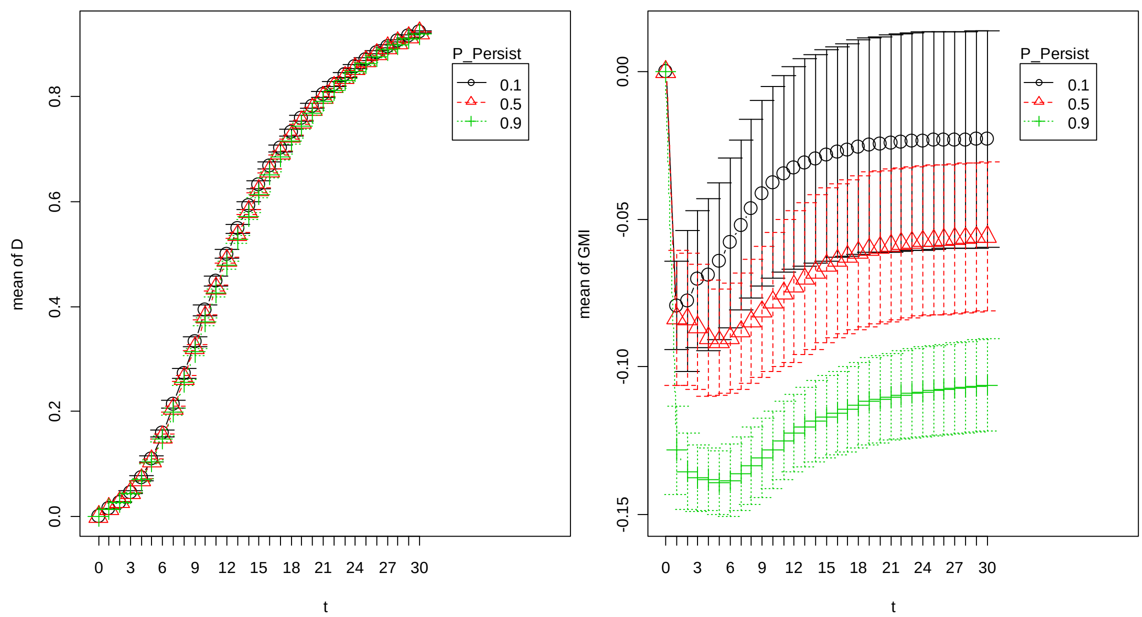


Figure 5.7 Schelling *P-Persist* Test: Mean of D & GMI-values (10 runs)

As can be seen in figure 5.7 (left side), the change of *P-Persist* has hardly any significant effect on the segregation dimensions measured by H (as well as D and I.I, not shown here). The values are plotted as means (of 10 different run values) with small ‘error’ bars giving plus or minus one standard deviation (i.e. variations) around them.

With careful attention it can be observed that the level of segregation measured by these indices is slightly (though subtly) higher with a lower *P-Persist* value (although it is less clear in this experiment). In other words, it seems that the segregation can have a ‘modest’ inverse relationship with these parameters, so that they may be lower when *P-Persist* is more elevated.

On the other hand, the effect of *P-Persist* value change is more noticeable on the GMI-values (see figure 5.7, right side). Initially starting by looking at neighbouring area units for suitable locations (see *Placement Process* section in chapter 4), with lower *P-Persist* values agents continue their search for desirable locations at other neighbourhoods much sooner than when they have higher level of persistence.

In other words, they have a better chance of finding a matching place by broadening their search at earlier stages before their personal maximum visit number (the ‘self-imposed’ limit that they establish for visiting different locations if no suitable place is still found) is exhausted. On the other hand, trying to find a desirable location in a specific area with high level of persistence can result in a higher possibility of not finding such a place within the search/ visit limit and a higher prospect of remaining at the current location would persist (viewed in this thesis as finding an unspecified location within the current location). In fact, by the end of simulation at t = 30, the overall (accumulative) percentage of minority group agents (g_2) who ‘remained’ in their initial locations was about 11%, 12% and 18% (for the last iteration alone: 13%, 15%, and 23%) respectively for *P-Persist* of 10%, 50% and 90%. More importantly though, the overall (and for the last iteration too) percentages for those who moved to the matching locations were higher when *P-Persist* values were lower. The co-ethnic proportion guessing errors were also historically lower when *P-Persist* was 10% (i.e. the overall error rate when neighbourhood was changed more often when searching a desirable place was lower than when sticking with a neighbourhood for a long time in searching for a suitable place).

This dynamic can help explain why the autocorrelation variations are somewhat 'acuter' when the persistence values are lower and how at the same time additional moving to matching locations contributed to reversing some early decline of autocorrelation values.

5.3.2 Fossett-like Exhibit

The Fossett model is yet another extension of Schelling's original model. As previously explained, the role of the minority group and its preference are central to the Fossett model.

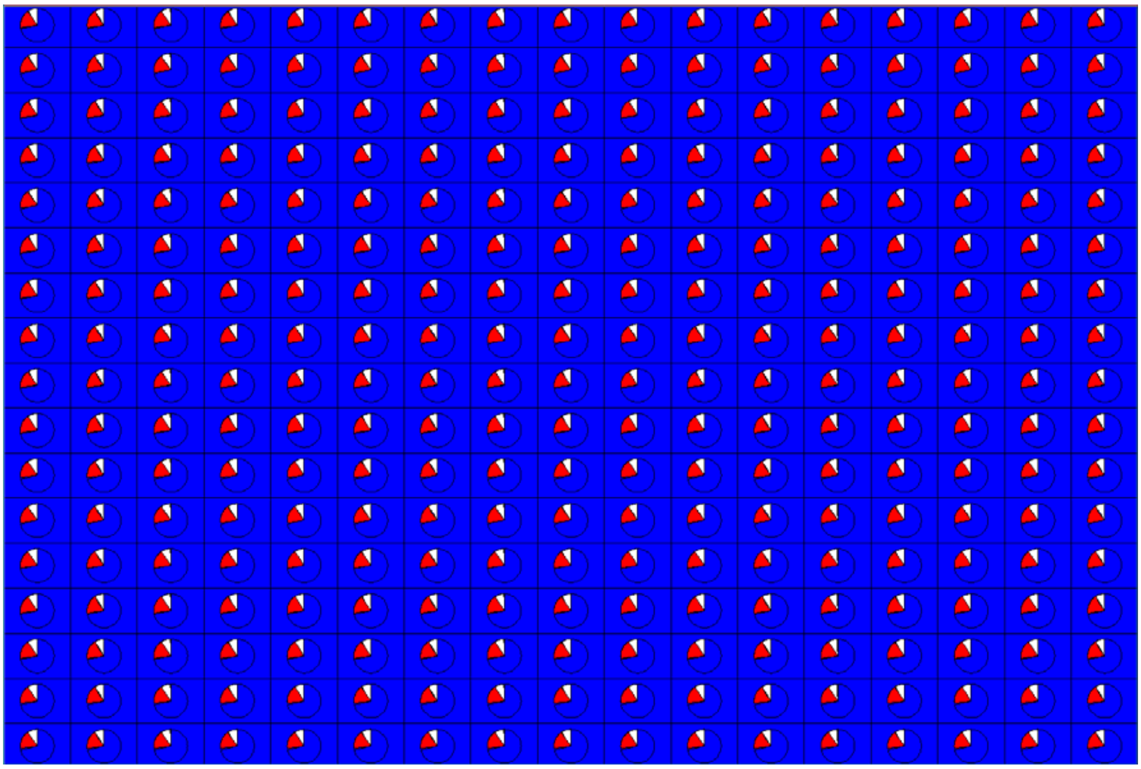


Figure 5.8 Fossett Test: Initialisation State (t=0)

It is therefore important to form two distinct groups of minority and majority and to test the effects of minority preferences on the segregation dimensions. Unlike the Schelling 50:50 population setting, the Fossett configuration conforms more to real situations where populations are of different sizes. The proportion ratio of minority-majority (g2-g1, red-blue) is chosen as 20:80 (excluding vacant spaces) in the following experiments. The starting state of the city is shown in figure 5.8. The pie charts inside each area unit all have the same shapes. Each cell unit's population consists of 800 blues and 200 reds along with 100 vacant spaces.

While the majority is completely neutral in their choice regarding neighbourhood composition (i.e. $T_1 = 0$) [although we later show how changing this would affect the outcomes], the level of tolerance for the minority group is varied (from 10% to 40%, including proximate values before, after and at the critical proportional weight of the minority group at 20%). The level of segregation measured by D, H and I.I are all positively correlated with the increase of the minority group tolerance (see figure 5.9), although the intensity of effects diminishes as the tolerance percentages approaches twice that of the overall minority proportional size. At $T_2 = 0.10$ the minority group agents (group 2) seem to always find first visited locations (or at worst within their search perimeter) suitable, so that moving randomly to these new satisfactory locations would not generate noteworthy segregation. The level of segregation however increases more significantly as the [in-]tolerance of the agents (of the minority group) approaches and passes their proportional size, as seen in the following figures.

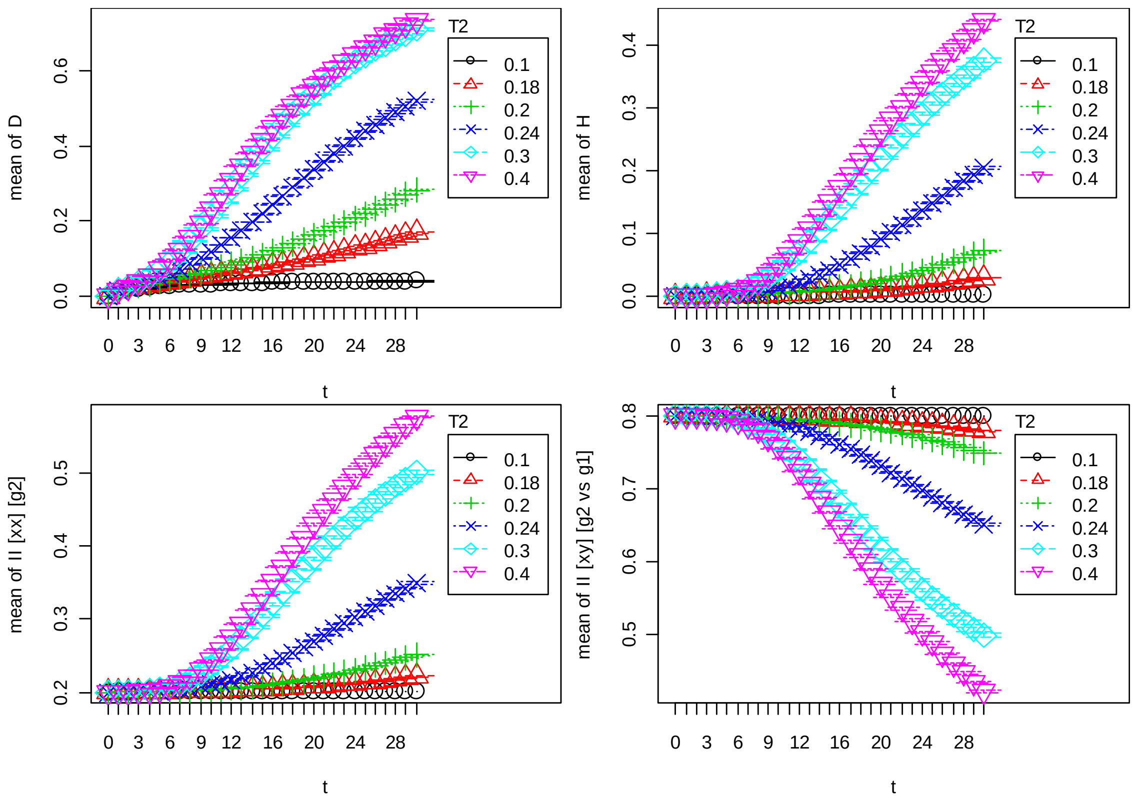


Figure 5.9 Fossett Test: Mean of D,H and I.I-values (10 runs)

The differences observed between the dynamic states when the value of T_2 was high (i.e. 30% and 40%) and when it was lower (i.e. 0%, 18%, 20% and to large extent for 24% too) could be helpful in making sense of the GMI-values (in figure 5.10). The initial phase of simulation when T_2-value was high is marked by high ratios of best (found so far) moves (in particular higher when T_2 was 40%) as well as high ratios of error when agents estimated co-ethnic proportion of candidate (visited) locations; while those who remained in the same locations were about 10% of the total moving agents. However, after more than a third into the simulation time (about $t \geq 12$) things started to gradually and continually change with smaller ratios for 'best' and 'error estimation', whilst at about the same time, the ratio of those who remained in the same locations continued to grow.

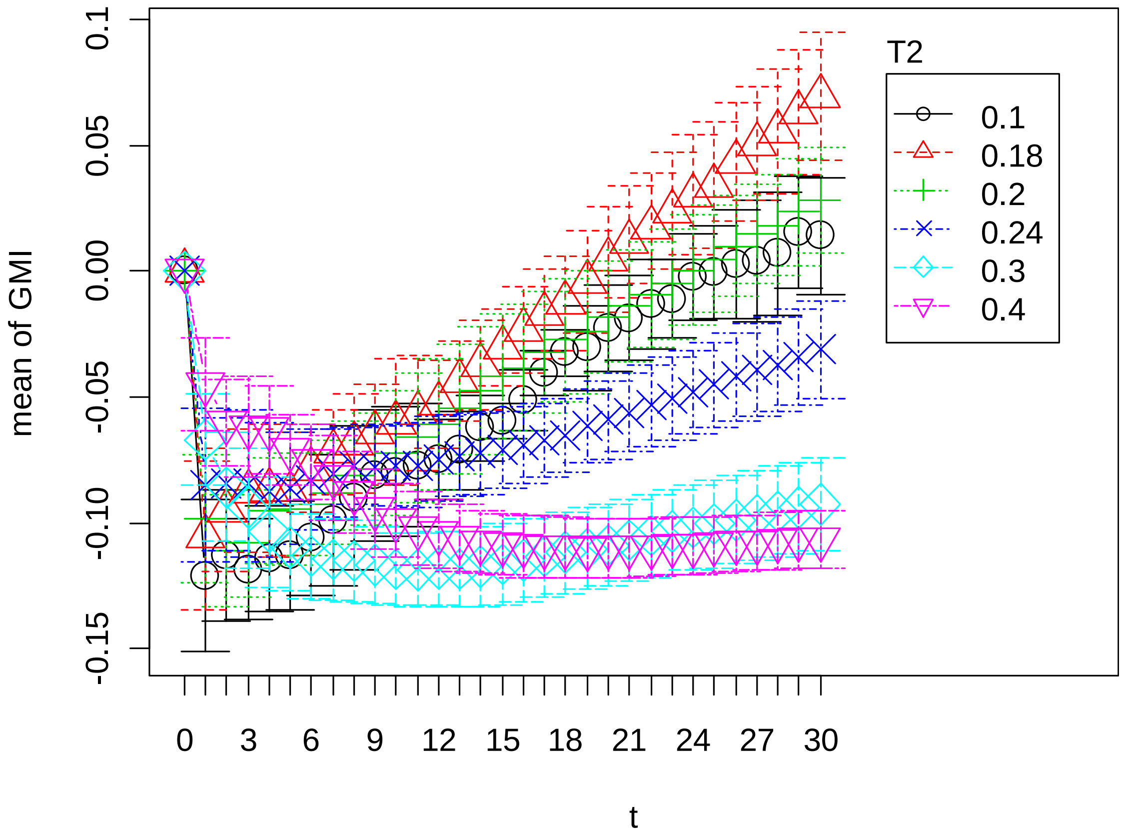


Figure 5.10 Fossett Test: GMI-values (10 runs)

By the end of simulation, T_2 = 0.3 & 0.4 were the only two cases with low but relatively significant ratios of those agents who settled in the 'best' available locations, but at the same time had considerably larger ratio of those agents who remained in the same locations. Although the estimation errors were no-occurring at the end of the simulation, the ratios of those who moved to the 'matched' locations were also lower for scenarios with higher T_2-values (particularly when it was 40%).

The 'world' was initially in a perfect equilibrium at the beginning of the simulation with exactly 20% of the minority group (g_2) population in each cell. When T_2 was high, the agents could not initially find any location that could satisfy their lofty expectations in a perfectly integrated world. That is why they compromised at the end of their visits' limit by settling in the 'best' locations they had found during their visits. Moreover, with little or no ethnic minority sector/ district where the minority group is locally a majority force, most of the guesses (regarding the estimation of the prospects' co-ethnic proportions, see chapter 4) were actually incorrect (there was nonetheless a considerable difference between 30% and 40% regarding the ratio of those agents who moved to the matching places, with 40% being the much smaller ratio).

When areas with higher percentages of a minority group started to emerge, the ratio of erroneous estimations declined, so agents could find 'matching' locations (those where the co-ethnic percentage were truly higher than their sought T thresholds) more easily. So, gradually the ratio of 'moved to best locations' fell at the expense of 'moved to matched locations'. Although there was over-concentration in these minority 'districts' the margin of estimation error became smaller, the minority agents at one point could not easily find vacant spots in these areas, nor could they find other locations with a satisfactory percentage of a co-ethnic population. For this reason their current locations started to appear more satisfactory than any other potential location. This was the time the ratio for moving to the best found locations dropped and the ratio for those who remained in their current locations (viewed as finding a suitable place in the same location) grew.

In contrast, the low T_2-level cases can be summarized by high ratios for agent moves to matching locations, and relatively steady ratios for those who remained in their original locations (about 12%) along with an altering estimation error percentage rate.

For $T_2 = 0.18$ for example, in the first iteration no significant difference could be observed in comparison with $T_2 = 0.10$. But as the population distribution changed (starting from the first iteration) and unevenness kicked in, a small fraction of co-ethnic percentage estimation error occurred at the second iteration (when T_2 = 0.18). This indicated that there were locations with a less than 18% minority presence but some household agents moved there because they made inaccurate assumptions (i.e. they assumed wrongly that proportions were comprised of at least 18% co-ethnic people). The estimation errors became more and more frequent over time amid more unevenness throughout the 'world': there was seemingly not enough of a clear line of separation (i.e. centres of minority concentration) for the agents to make lesser mistakes in evaluating the co-ethnic percentages of prospects' candidate locations. In fact, as can be observed from the measures and the visual maps (see figure 5.11) this level of intolerance (T_2 = 0.18) was not sufficient (at least within this simulation time period) for the emergence of a crystallised segregation type.

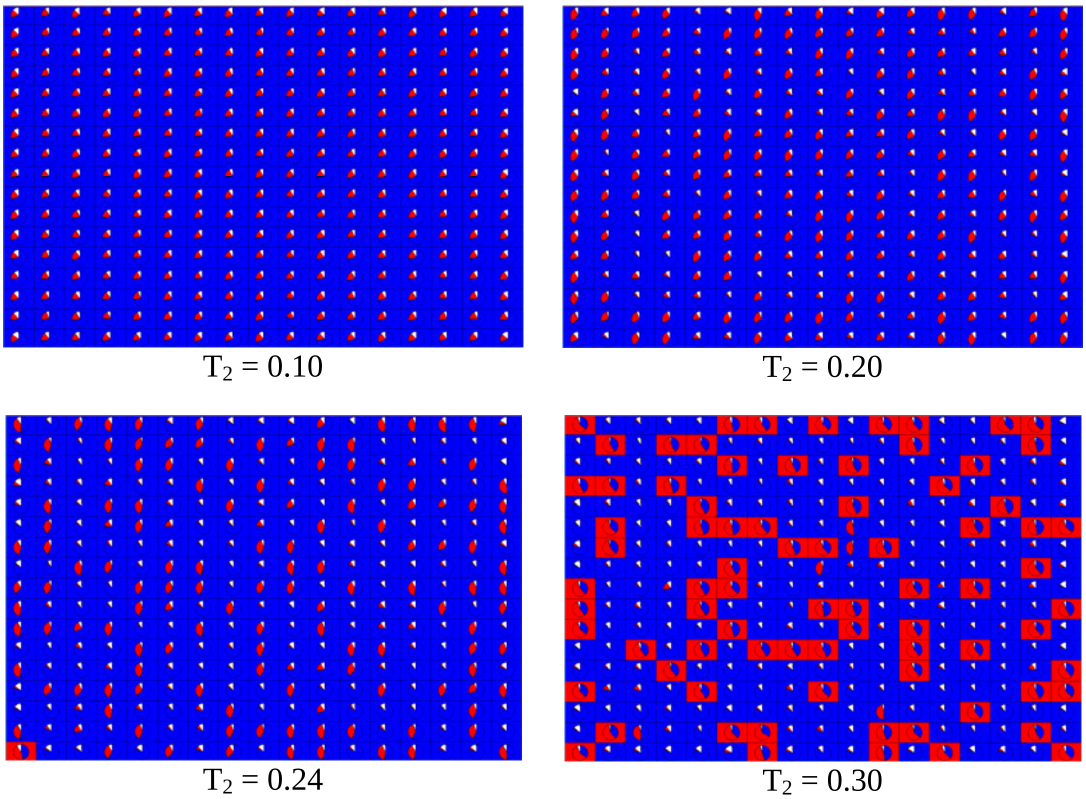


Figure 5.11 Fossett Test: World Snapshots as T_2 is Varied

At the end, despite the growing percentage of estimation errors for low T_2 level cases (such as 0.18), the real numbers of those agents who settled in the matching location (i.e. correct estimations minus those who did not) were still higher than the cases when T_2 was high (e.g. 0.3 and 0.4). This caused a more elevated realization of the spatial correlation which can be observed in figure 5.10.

As a spatial function relationship, higher correlation between the proportions of a minority group in one cell in relation to other neighbouring cells can be seen in the visual map (see for example, case of $T_2 = 0.20$ in figure 5.11, top right) by paying close attention to the more balanced red proportions in the pie-charts in one cell and its neighbouring cells (and overall in the grid world). In case of $T_2 = 0.30$ (or 0.40, not shown here), although the manifestation of local clusters can be seen in several cells where the minority group had a majority status, the global statistic of Moran's I indicates a lower (negative) autocorrelation (at least within this simulation time period) which at first instance may appear counterintuitive (figure 5.11, bottom right).

Partial Exploration of Parameter Space (with Fossett model)

<u>Effects of Higher Turnover [$P(M)_g$]</u>

With a 'world' in symmetry at t = 0, and with the movements of agents initially starting locally, the unevenness occurs slowly and gradually. In other words, it is the movements of agents that cause the 'world' to progressively degenerate from equilibrium towards segregation. It is therefore reasonable to expect that a higher level of turnover would result in a significantly quicker unravelling of this process. Figures 5.12 and 5.13 confirm this. They only show D and GMI values respectively when T_2 = 0.10 whilst comparing the outcomes between turnover 0.10 versus 0.20. As can be seen, a higher number of movements led to a faster and relatively more intense segregation from the initial time when the 'world' had an unrealistic symmetrical structure.

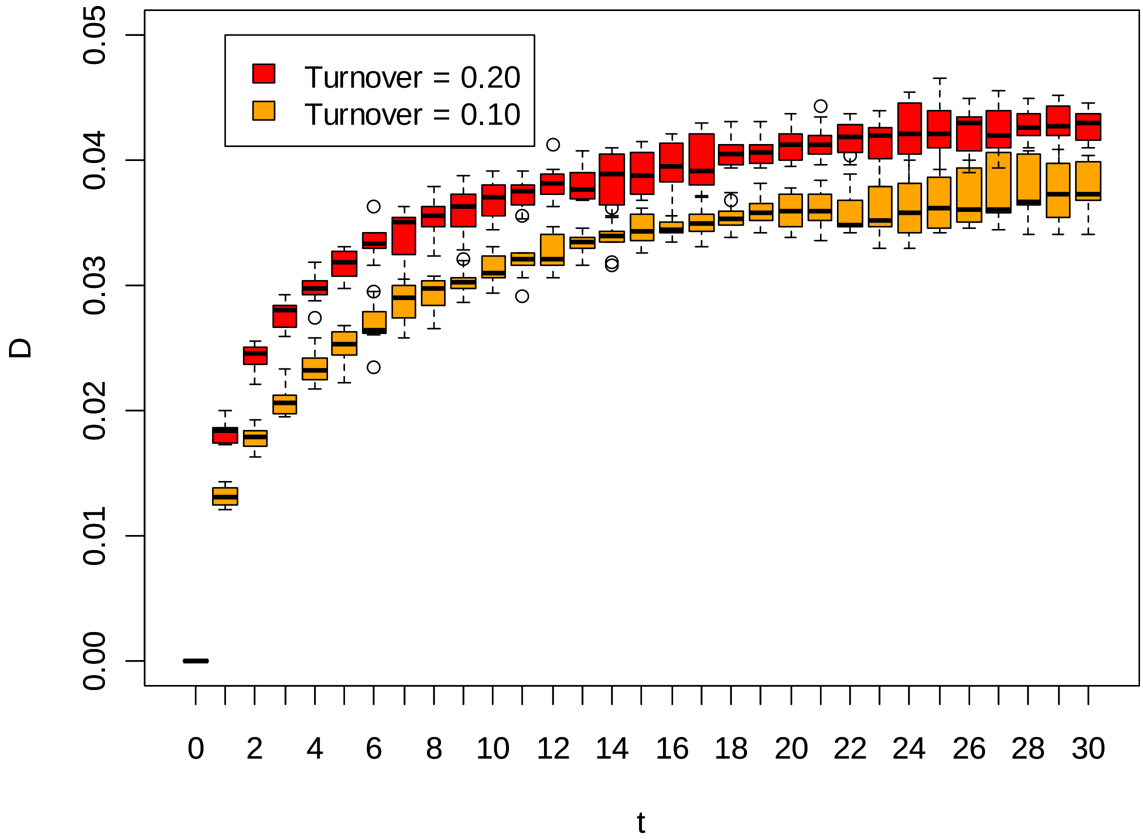


Figure 5.12 Fossett Test: D-values with Different Turnovers

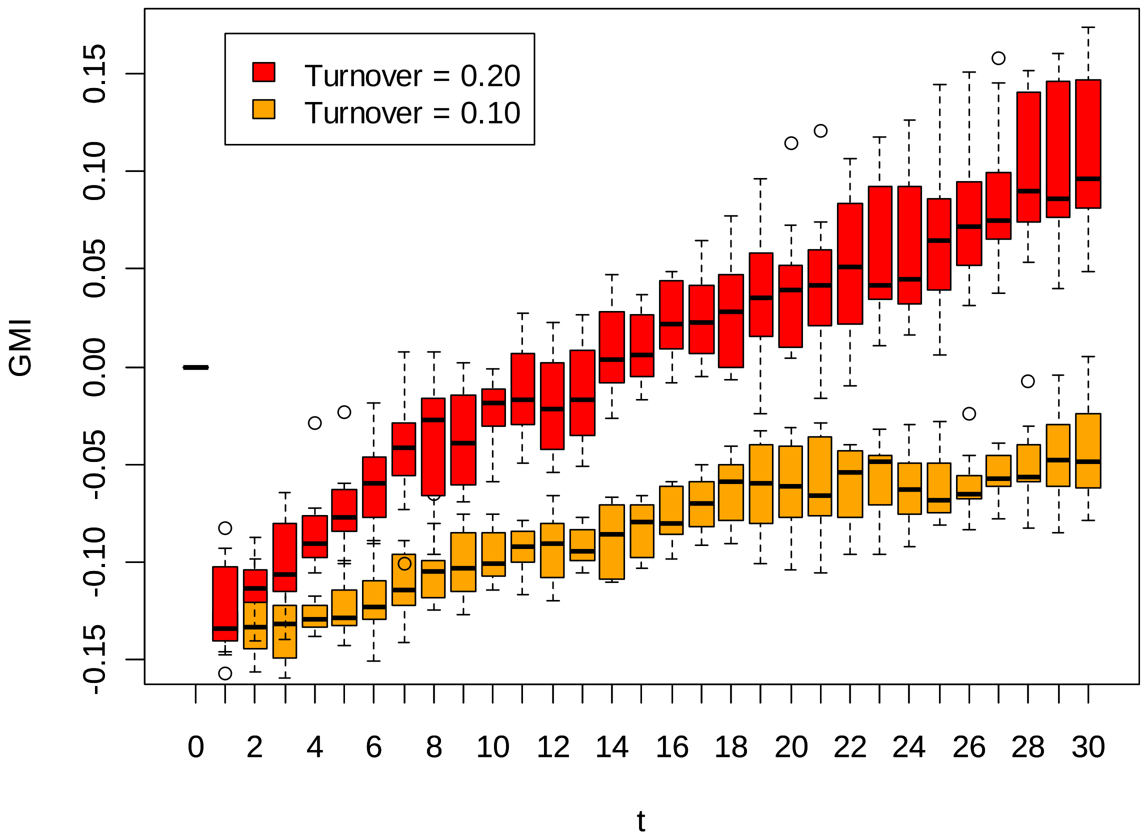


Figure 5.13 Fossett Test: GMI-values with Different Turnovers

Effects of Discriminatory Behaviour by the Majority Group ($T_1 \neq 0$)

So far it has been assumed that the larger group (majority) which comprised 80% of the entire population was neutral regarding the composition of neighbourhoods. This way, the overall observed level of segregation was relatively small. Figure 5.14 shows what happens to H and GMI values whilst figure 5.15 displays a snapshot of the world at the end of the simulations when the majority group starts to exhibit some discriminatory behaviour. The co-ethnic tolerance of the minority group (T_2) is kept constant at 0.20 and the turnovers for both groups are set to 20% for these experiments. Given that the group constitutes a high proportion of the population (80%) and the fact that agents decide about their relocations on the percentage of the co-ethnic population for the prospective neighbourhood (Miller & Page, 2007, p. 145); that is, they are sensitive to the overall proportion of a given population, a significant change in the results (related to T_1 / majority factor, and not T_2) appears only at T_1 = 0.70. More intense change occurs at 0.80+, notwithstanding the fact that increasing inflexibility in finding matching locations starts taking a toll in terms of autocorrelation values.

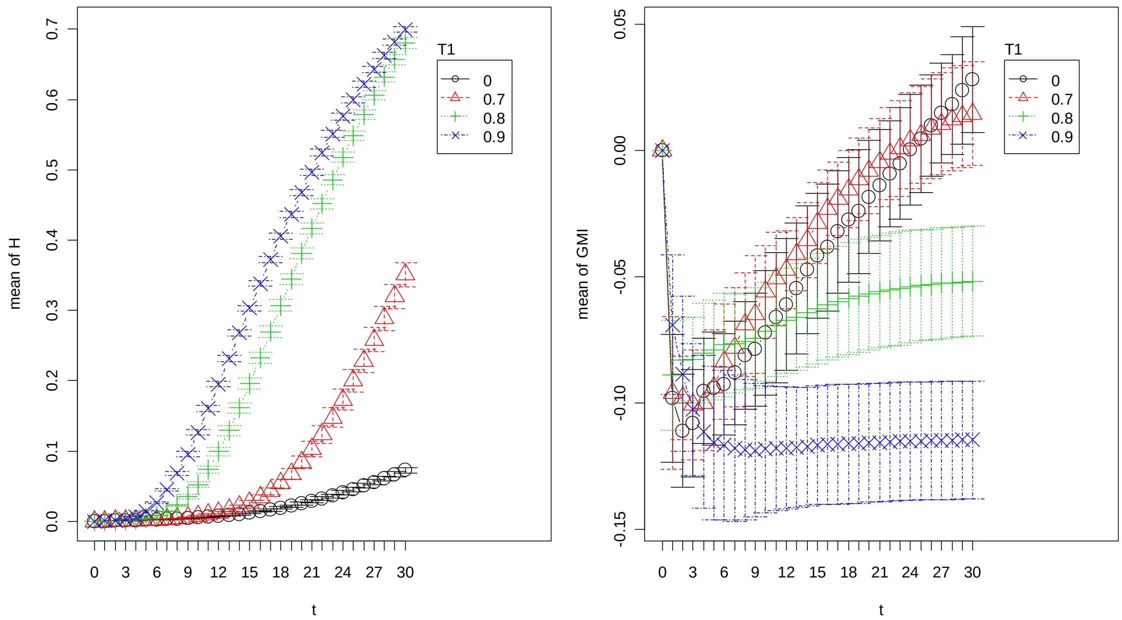


Figure 5.14 Fossett Test: H-values as T_1 Varies

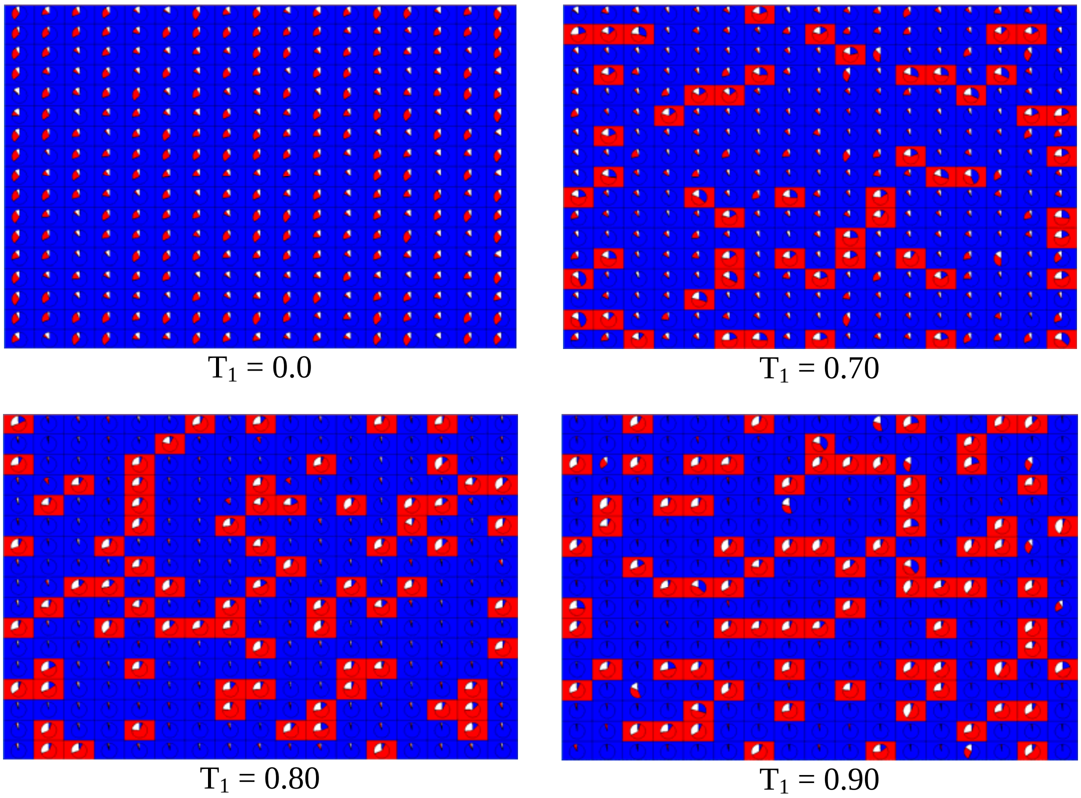


Figure 5.15 Fossett World Snapshots as T_1 is Varied

Although the dissimilarity between T_1 values of 0.80 vs. 0.90 is visually less evident in figure 5.15, subtle differences in quantitative measures still exist (see figure 5.14).

<u>Effects of Locational Preference Change (L < 100%)</u>

The movements of agents in the original Schelling and Fossett models are not restricted. They can relocate randomly to any part of the 'world' when they search for suitable locations and move there.

However, with the *HAAMoS* model, it is possible to arrange for agents to have a preference for relocating to local neighbourhoods only. Theoretically it is expected that agents who can relocate to anywhere in the 'world' find their preferred spots more easily. This should therefore lead to a higher level of segregation. On the other hand, agents who only prefer to move to the immediate neighbourhoods have a higher chance of ending up in locations below their original expected thresholds.

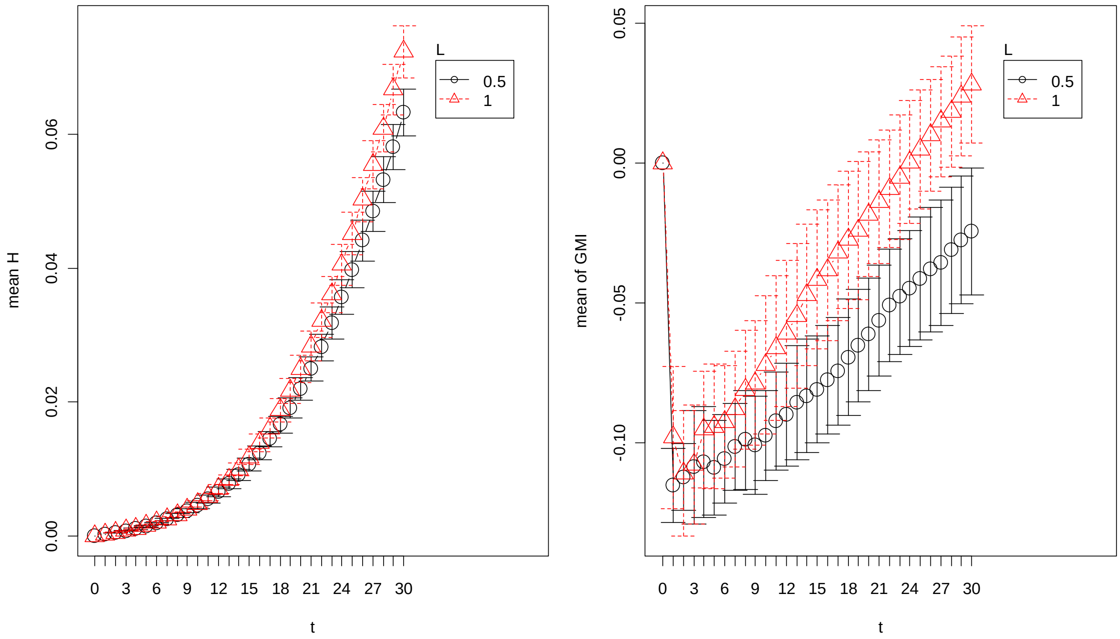


Figure 5.16 Fossett Test: H and GMI-values as L is Varied

However, in practice this may be influenced by other factors such as the ability (and extent) of agents to make accurate guess-estimations (regarding composition of neighbourhoods), as well as their persistence (in finding a desirable spots in specific areas) and the number of visits they are willing to make before final settlement.

The results of a simple experiment (with 20% turnover) presented in figure 5.16 confirms the expected outcome. With the majority group neutral and T_2 = 0.20 the segregation measured by H and GMI is somewhat more elevated when the probabilistic percentage of global movers (or global locational probability move preference, P(GL), which for simplicity will be referred to as L only) is higher (L = 1). , By the end of the simulation period in this particular example there were about 2% (overall)/ 4% (in the last iteration) more agents who moved to matching locations when L-level was higher, whilst at the same time the overall numbers of accurate guesses were also higher.

5.4 Summary and Discussion

This chapter began with a description of the measures used in this thesis to evaluate and monitor various aspects of segregation dimensions. There are arguably some levels of overlap between some of these measures. In particular D and H measures appear to be redundant when only two groups are involved. However, the utility of H manifests with multiple ethnic groups, as will be seen in chapter 7. The I.I measures the odds of isolation or interaction of a group member with others (in/ out group members) and is sensitive to the size and proportion of populations. Unlike the experimental examples shown in this chapter where the population sizes were constant throughout the simulations, the advantage of I.I will be better demonstrated in chapter 7 where the scenarios involve the exogenous inflow (immigration) of population as well as other growth.

Using the original Schelling model setting along with the minority-majority centric configuration of the Fossett model, it was demonstrated how the *HAAMoS* model exhibits similar behaviours when basic conditions are present.

A comprehensive and meticulous exploration of the parameter space would be beyond the focus of this thesis. Nevertheless several experiments were used to demonstrate how the changes in agents' behaviour could result in different emerging outcomes. Generally, it is expected that a higher level of intolerance (T) will cause a higher level of segregation measured by D, H and I.I., whilst at the same time this would be sensitive to the percentage of T relative to the population percentage in the entire 'world' (and smaller spatial areas) under investigation. Simultaneous intolerances (by more than one group) would also lead to higher (and faster) segregation. Higher levels of turnover seem to accelerate and potentially aggravate the emergence of the phenomenon initiated by the movements of agents. Higher probabilities for having a higher number of agents who move globally (L) are generally more segregation-prone than having them with a preference for uniquely local moves.

When it comes to the segregation aspect measured by GMI, conclusions are not always straightforward. In fact, the impression that can be gained from visual observation of the 'world' may seem counterintuitive and misleading. This is because unlike typical individual cell-based Schelling-generation models, this model uses aggregate based populations and representations. The colours of spatial units indicate which group is in plurality locally and a collection of them adjacent to each other shows with the formation of 'clusters' (detectable by local Moran's I measure and not the global statistics of GMI which is used for measuring 'clustering'). In other words, visual clustering may not necessary translate into a higher level of autocorrelation at the global scale. On the other hand, a 'world' where a single group appears to dominate the other group by being the plurality 'force' in all the spatial area may have a high level of autocorrelation (GMI). This simply depends on the sum of the spatial relations of each individual spatial/ area unit's group proportion in relation to its adjacent units – which can actually be observed (although not easily, but with careful attention to detail) by inspecting the shape of pie-charts over spatial units.

In general, higher numbers of agents who relocate to matching locations (i.e. higher co-ethnic proportion than their T threshold) would lead to higher (positive) autocorrelation level (GMI). But other factors can also be influential. For example, the extent to which the move to matching places is based on accurate estimation of neighbourhood composition (and not by erroneous guessing).

The Schelling and Fossett experiments started with perfect but unrealistic symmetrical/ equilibrium configurations – but in chapter 7 – realistic scenarios will take place in a realistic urban area in disequilibrium. The effects of comparable conditions in this realistic environment with the presence of multiple ethnic groups along with exogenous factors such as immigration will be demonstrated using different scenarios. Before this, we will learn more about this real environment (i.e. the Auckland region urban/ metropolitan area) in chapter 6.

Chapter 6

6. Empirical Case Study: Auckland 1991-2006

6.1 Introduction

This chapter begins with the unpacking of the empirical data in section 6.2. This will include their sources and the delimitation of them. The unpacking will also reveal the extent of their interwoven error ('noise') and explain how the level of error has been attenuated.

The chapter then moves on to section 6.3 by examining the changes that have occurred in the urban Auckland metropolitan area over the census periods of 1991-2006. This is done through a series of descriptive tables of quantitative values as well as by time-series plots with an emphasis on changes in population sizes, proportions, as well as by examining the results of selected measures (introduced in the previous chapter) applied to census data. These include global measures for gauging the exposure, concentration, evenness, and clustering dimensions at the metropolitan area level as well as some local spatial measures at the micro-geographical scale.

Finally, the chapter proposes a method to identify the 'hottest' and 'coldest' local spots (outmost outliers) based on specific variables of interest.

6.2 Empirical Data

6.2.1 Data Sources

The New Zealand Census of Population and Dwellings is conducted every five years by Statistics New Zealand (SNZ)[13]. In this process individual and household information from every household in the country is collected. This study draws on the aggregate and unit level data from the 1991 through to the 2006 Census datasets. This aggregate data had already been processed and conformed with to the SNZ data confidentiality rules such as randomly rounding of counts. This means the counts are not exact, but they nonetheless correspond closely to the situation at the time of census (for more information about these rules, see appendix F).

In general, the data and its format (e.g. aggregate nature) have played a significant role in shaping the final model. It is correct to say that the model with the ultimate intention of using real data for its validation has practically been built around the data.

6.2.2 Data Delimitation

The census data for the Auckland region include a total of 399 area units. Each of these 399 area units is included in one of the following territorial authorities (TA) shown in table 6.1.

Table 6.1 Auckland Region's 7 Territorial Authorities
(Rodney and Franklin districts are excluded in this study)

Regional Council	TA: District/ City or other areas	TA-Code
Auckland Region (code: 02)	1- Rodney District	004
	2- North Shore City	005
	3- Waitakere City	006
	4- Auckland City	007
	5- Manukau City	008
	6- Papakura District	009
	7- Franklin District	010

[13] Published census volumes can be obtained (as pre-packaged release, otherwise by using 'Table Builder') from Statistics New Zealand's website (www.stats.govt.nz).

The Auckland region consists of seven districts and cities. In this thesis, the terms TA or district are used to refer to these geographical census boundaries, but not 'city', as this may insinuate a reference to the entire geographical space (here called metropolitan or urban area).

The *Rodney* and *Franklin* districts are relatively uninhabited compared to the other five districts/ TAs (in other words, they are not really 'urban' in the traditional sense). For this and practical reasons including difficulty regarding the visual representation and analysis of the entire metropolitan area (*Franklin* and *Rodney* in particular have large geographical sizes), only five TAs are used in this study (excluding *Franklin* and *Rodney*). More importantly, the two latter TAs have been excluded from this study as the current model does not include a mechanism (function) that 'ranks' a location's 'attractiveness' based on its distance to the centre of business/ activities (e.g. job opportunities, public services, amenities, and the like). This means agents could move to distant and low populated area units in *Rodney* district, for example, with the same likelihood that they would possibly have chosen an area unit in the *Auckland* district. This could have led to the formation of implausible patterns.

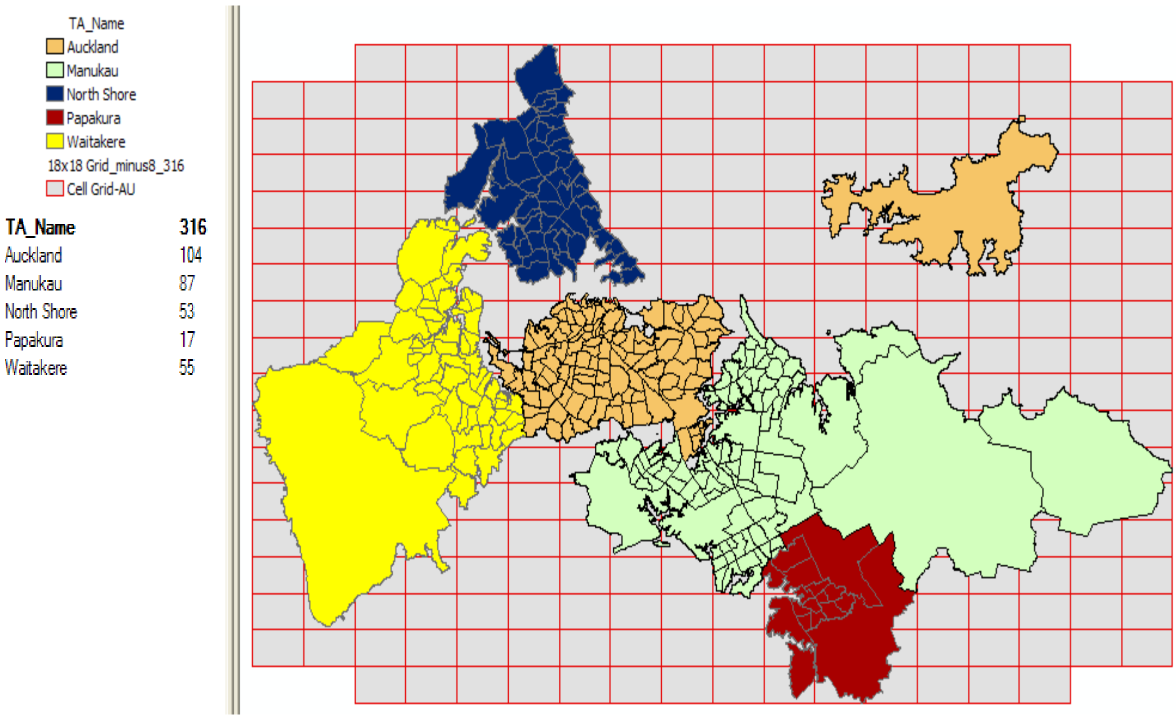


Figure 6.1 Auckland Region HES: 5 TAs, 316 AUs

Figure 6.1 shows the hybrid environment used for the simulation experiments: a realistic setting of the Auckland region with five selected TAs along with its 316 area units; and an artificial grid-based equivalent of size 18×18 minus eight cells (two in each corner of the grid).

The main data variables used in this analysis and in the model come from ethnic counts of individuals. The SNZ standard classification of ethnicity is a hierarchical classification of four levels. The data used for this study corresponds to the category at level one (Statistics New Zealand, 2005), shown in table 6.2. Only the four major ethnic categories (i.e. European, Asian, Pacific and Maori) are used in this study and can be simulated in the model, since the percentage of other ethnicities are relatively marginal (MELAA[14], 'Other Ethnicity' and NEI[15] categories combined comprise of about 6% of the total population counts in the last three censuses, whilst 5% is attributed to the NEI category alone). Nevertheless when necessary, additional groups can be added to the model with relatively minimal programming effort.

Table 6.2 SNZ Standard Classification of Ethnicity

Ethnicity Categories (Level One)	Used in Model
1. European	✓
2. Asian	✓
3. Pacific Peoples	✓
4. Maori	✓
5. MELAA	
6. Other Ethnicity	
7. Not Elsewhere Included (NEI)	

The data (censuses 1991-2006) do not include the counts of households based on ethnicity. Dividing the individual ethnic population counts by the size of national

[14] MELAA refers to Middle Eastern/ Latin American/ African.

[15] Not Elsewhere Included (NEI) encompasses at least one, several or all of the following categories: 'not stated', 'response outside scope', 'response unidentifiable', 'refused to answer' and 'don't know' (based on information found in SNZ datasets, data dictionaries and SNZ website http://www.stats.govt.nz).

average household would just be a matter of scaling. Moreover, this approach does not take into account the composition of households, which consist of one, two, three, four, or more people per household. Methods that classify the entire household in a specific ethnic group based on a selected member of the household (e.g. 'head' of the family/ household) are also far from ideal, since the multi-ethnic nature of households is suppressed. More importantly, some of the empirically-based scenarios in this thesis will use the SNZ individual-based population percentage estimate projections (which are based on districts/ TAs only, see appendix G.3) for the simulated years beyond which census data is not available. Therefore, considering the data limitation and these related issues, it was decided to conduct this study using population counts of individuals. It is believed the fundamentals of the rationale and methodological features of this study, as well as results and conclusions will remain adequate, valid and applicable. This way, it would be desirable to view the agents in this model as one-person households (which already form about 20% of the entire total counts of households in the Auckland urban area). Notwithstanding the use of individual counts, the reference to 'household' (interchangeably or in combination with, for example, agent or individual) throughout this thesis will persist.

6.2.3 Data Noise and its Attenuation

The problems with census data and its 'quality' (accuracy) are not unique to the SNZ census data (see, for example, Alexander et al., 2010), although the nature and causes of problems and inaccuracies are not necessarily the same. Certain factors such as confidentiality rules and intercensal inconsistencies can potentially cause some level of 'noise'/ error to 'interweave' with the data (e.g. during data collection or after processing). Sometimes the noise effects are negligible. However, in some cases this issue can become significant and when possible should be remedied. The following sections explain some aspects related to the quality of data, the nature of the noise, its potential effects and how, when possible, it has been remedied.

Data Discrepancy Related to the Change of Ethnicity Question

The lack of consensus about the precise definition of ethnicity (Banks, 1996; Hutchinson & Smith, 1996), is reflected in 'ambiguous' questions with possible multi-interpretations (as was the case in 1996 New Zealand Census which encouraged people to answer on the basis of ancestry rather than identification (Allan, 2001)). For this reason respondents, when asked for ethnic group identification, report a range of identities from nationality (citizenship), country of birth or residence, ancestry, socio-cultural affiliations (e.g. language, culture, religion) to race. Some may report one ethnic group but identify with another, or report more groups but in fact identify with fewer groups. A number of people may refuse to answer questions on ethnicity or may answer facetiously (Statistics New Zealand, 2004; 2005).

In fact, changes to the ethnicity questions and their encoding for census periods between 1991 and 2006 have resulted in some data inconsistency. These include:

- The effects of the question change in 1996, and particularly the impact of the drop down tick boxes for "Other European" categories (see appendix C.3). Previous work by Statistics New Zealand has shown that these categories led some respondents to think of ethnicity in terms of their ancestry.

- Consequently, the results from the 2001 Census are not in line with results from the 1996 Census. In fact, the 2001 Census ethnic question is much comparable with the 1991 Census results, as the questions asked in these census years (1991 and 2001) are almost identical.

- Moreover, six ethnic responses were captured and coded in 2001, compared with a maximum of 3 responses in 1996. 'Ethnic Group - Up to Three Responses' is a derivation which reduces the number of ethnic responses from a maximum of 6 down to 3. This derivation is used to compare 2001 data with data from the 1991 and 1996 Censuses which only collected up to 3 responses. However, the impact of the change on the 2001 data is relatively small, as only 0.3% of the population gave more than 3 responses in the census of that year.

Unfortunately, there is no possible corrective measure regarding these aspects as available public data had already been processed and aggregated. Although this should not be considered a major concern, awareness of these issues is useful since it can partly explain some of the inconsistencies in the data such as occasional rapid falls or abrupt increases in the population of some ethnic groups from one census to another.

2006 'New Zealanders' Category Effect

One significant disparity in the data was particularly noticeable in the 2006 Census which was caused by an unusual (and unexpected) surge of responses under the 'New Zealander' ethnic group classification. This classification was later placed under the 'Other' category (this category is designated to include ethnicities outside the major ones recognized by SNZ). The sudden surge was attributed to a widely circulated campaign on national television and other media just prior to the 2006 Census to invite people to select this category.

Keeping with SNZ recommendations, it was assumed that 99% of the New Zealander category was 'European' and therefore data have been adjusted accordingly to correct this discrepancy (by adjusting 'European' and 'Other' categories' populations).

Multiple Ethnicity Choice Effect

The census question on ethnicity allows people to provide more than one response. People who reported more than one response will be counted in each group they reported. This means that the total population will be greater than the usual subject population.

The difference between 'sum of ethnic groups' and 'total people stated' in the census data can determine the number of extra people counts (i.e. those who reported more than one ethnicity). An attenuation technique for the multiple ethnicity choice effect was performed by reducing the extra ethnicity reported counts proportionally from the population size of each ethnic group. This noise reduction process has been applied on the already modified data for 2006 'New Zealander' category effect. Although, the final processed data could not be completely free from this noise, it has nonetheless been greatly minimized and thus the quality of data in this regard has been improved significantly.

6.3 Descriptive Statistics Analysis

6.3.1 Overall Changes

Tables 6.3 show a statistical summary of population in 5 territorial authorities and their constituent 316 area units. With the exception of Europeans, the population of all ethnic groups have constantly increased over each census period, albeit at different rates. The Asian population category for example has more than quadrupled, from about 50,000 in 1991 to more than 200,000 in 2006.

Table 6.3 Metro Area Stats Summary on Ethnic Groups

Year	Stats	Euro	Maori	Pacific	Asian
1991	Prop	70.55%	10.42%	12.01%	5.65%
	Total	605,953	89,517	103,146	48,497
	Max	4750	1430	2629	622
	Min	54	0	0	0
	AVG	1917.6	283.3	326.4	153.5
1996	Prop	62.34%	10.48%	11.67%	9.41%
	Total	602,459	101,283	112,774	90,926
	Max	5010	1488	2536	1613
	Min	51	3	0	0
	AVG	1906.5	320.5	356.9	287.7
2001	Prop	57.98%	9.90%	12.95%	13.11%
	Total	605,423	103,419	135,236	136,924
	Max	5552	1593	2871	2167
	Min	23	9	0	3
	AVG	1915.9	327.3	428.0	433.3
2006	Prop	53.05%	9.33%	13.14%	17.92%
	Total	620,711	109,209	153,782	209,695
	Max	6080	1984	3302	4262
	Min	16	20	3	8
	AVG	1964.3	345.6	486.7	663.6

The total population growth of the Pacific ethnic group has also been significant, from about 100,000 in 1991 to more than 150,000 in 2006 (constituting about 13% of the overall population in 2006).

The population growth of Asians and Pacific people is largely the product of immigration (although in case of Pacific people, natural growth also plays a considerable role as this group has the highest fertility rate among the four major ethnic categories). Since the mid 80s, the NZ immigration system has been liberalized by using a point-system which allows candidates to obtain an immigration category visa based on the points awarded for factors such as skills, education/ qualifications, work experience, and age, regardless of their race, ethnicity, religion, or place of birth. This is supplemented by the family (reunification) visa category which allows family members or partners to be sponsored for a residency visa. In the case of Pacific people, there is also an annual quota that grants residency visas based on a ballot system.

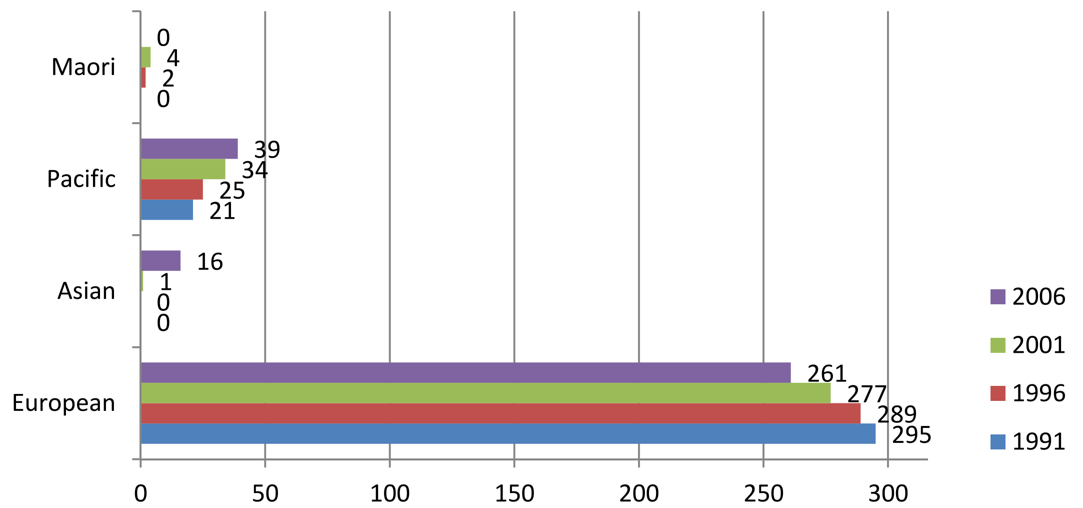
The Maori total population in the Auckland region (5TA) has also grown steadily, albeit more modestly. This is the case even though their overall proportion has slightly declined (about 1%) as a consequence of the rapid growth the of Pacific and Asian categories. Unlike Asians and Pacific people, the Maori population has not primarily been located in the metropolis of New Zealand (i.e. the Auckland region) as they have large population bases in other regions in New Zealand such as *Northland*, *Waikato*, and the *Bay of Plenty*.

In proportional terms, the growth in the proportion of Asian and Pacific people has mostly been at the expense of European people. In fact, the European proportion is the only group that has continuously declined over these census periods, from more than 70% in 1991 to around 53% in 2006. This is perhaps not surprising as the Auckland region has been the primary immigrant destination of non-European ethnic groups. Despite the loss of proportional weight, the overall European population was higher in 2006 than it was in 1991.

Descriptive and Visual Representations of 'Plurality Landscape' Changes

The overall picture obtained from table 6.3 is a growing population in general (as a result of natural growth and immigration) where the Pacific and in particular the Asian populations categories are increasing at a much faster pace than the other two groups.

Another indicator of the ensuing changes in the Auckland region landscape by the Pacific and Asian population categories can be seen in the counts of area units where ethnic groups constitute a plurality force. These changes in landscape can be seen visually in snapshots of the metropolitan area (figures 6.3-6.5) as well as in the graphical statistics shown in figure 6.2. As can be seen in figure 6.2, in 1991 and 1996 there were no area units where Asians constitute a plurality group.



	European	Asian	Pacific	Maori
2006	261	16	39	0
2001	277	1	34	4
1996	289	0	25	2
1991	295	0	21	0

Figure 6.2 Counts of Area Units Where Ethnic Groups Form a Plurality

In 2001, there was only one, and in 2006 this number had dramatically increased to 16. Pacific people on the hand have long concentrated in certain areas of the city and the number of these areas has grown steadily over the last censuses (from 21 in 1991 to 39 to 2006). Despite the dramatic increase in the number of areas with a large concentration of Asians in 2006 they are in comparison with other groups still to great extent more widely distributed. In 2006, they constituted about 18% of the entire population of the 5TAs in the Auckland region with a plurality status in 16 area units. In comparison Pacific people had only about a 13% share of the entire population with a plurality status in 39 area units. Maoris with about 9% of the entire population formed no plurality in any part of the metropolitan area in 2006. So, it is clear that Pacific people for some unidentified reasons (potentially combination of cultural, socio-economical and other factors) 'prefer' to stay closer to each other – in comparison to Maori and Asian groups.

The mosaic-like geographical map of ethnic representation of the metropolitan area (5TAs/ 316AUs) based on 'adjusted' census data can also provide a useful ground for visual inspections, as well as qualitative comparisons in analysing census data and particularly the simulation outputs. For example (as can be noticed in figure 6.3) in 1991 apart from the majority of area units inhabited by Europeans (identified in blue), there were only a handful of area units (relatively clustered in the middle central part of the map) where Pacific people are the plurality group (identified in green). In 2006 (figure 6.4) this ethnic mosaic had changed, with more area units having plurality-status concentrations of Pacific people, along with 16 area units where Asians had become the largest group (identified in red).

As for Maoris, a zoom version of the 2001 Census map (see figure 6.5) is used to show the presence of the four area units where this ethnic group was the largest group (identified in grey, within the white circle indicator). The fact that despite forming more than 9% of the entire population in 2006, the Maoris do not hold any plurality area suggests that their preference for residential co-ethnic contact may be lower than that for Pacific people. (We will see in chapter 7 that the lower residential co-ethnic preference of Maoris than of Pacific people [in both absolute and proportional terms] generated more realistic patterns during calibration).

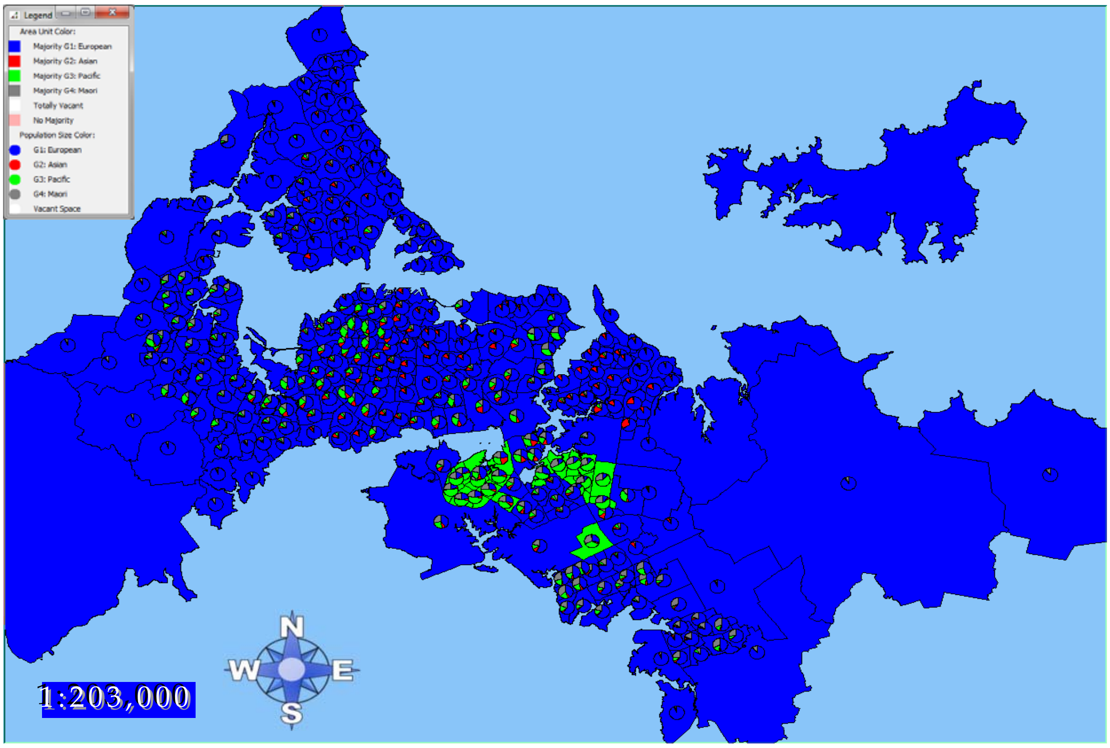


Figure 6.3 Metropolitan Area Map, 1991

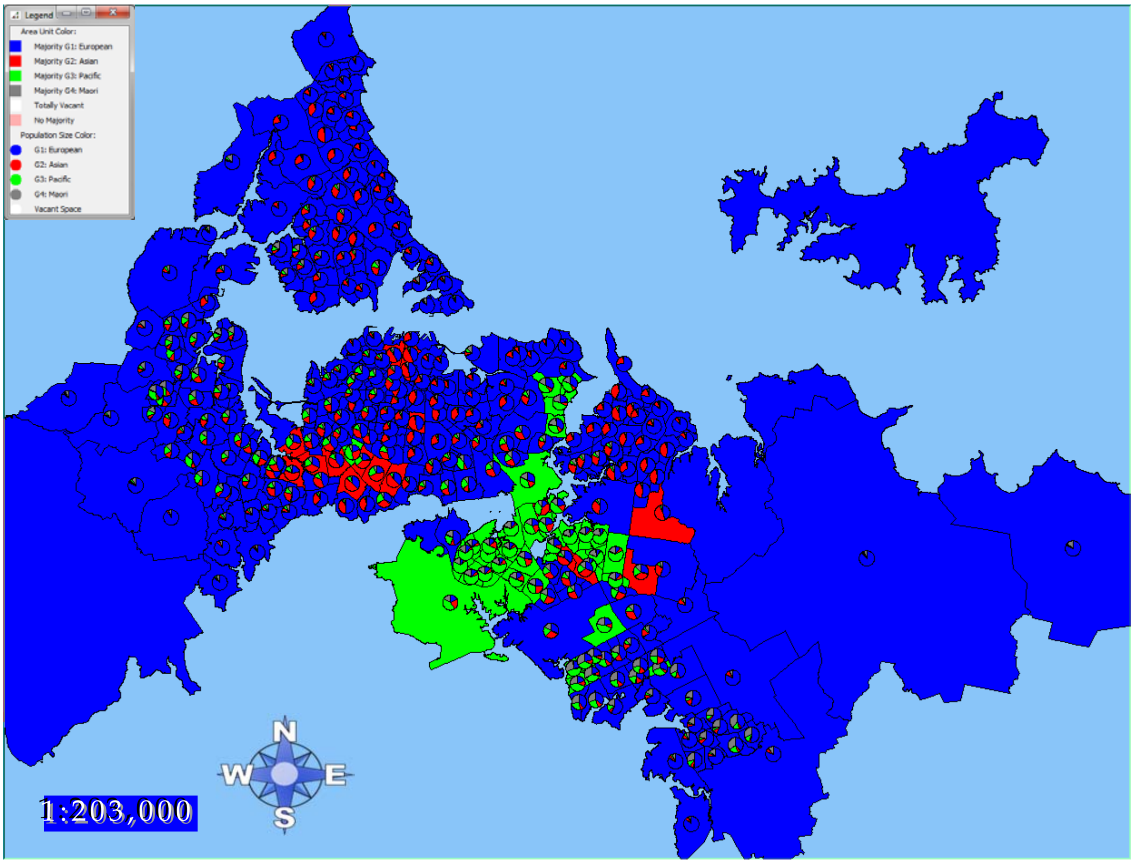


Figure 6.4 Metropolitan Area Map, 2006

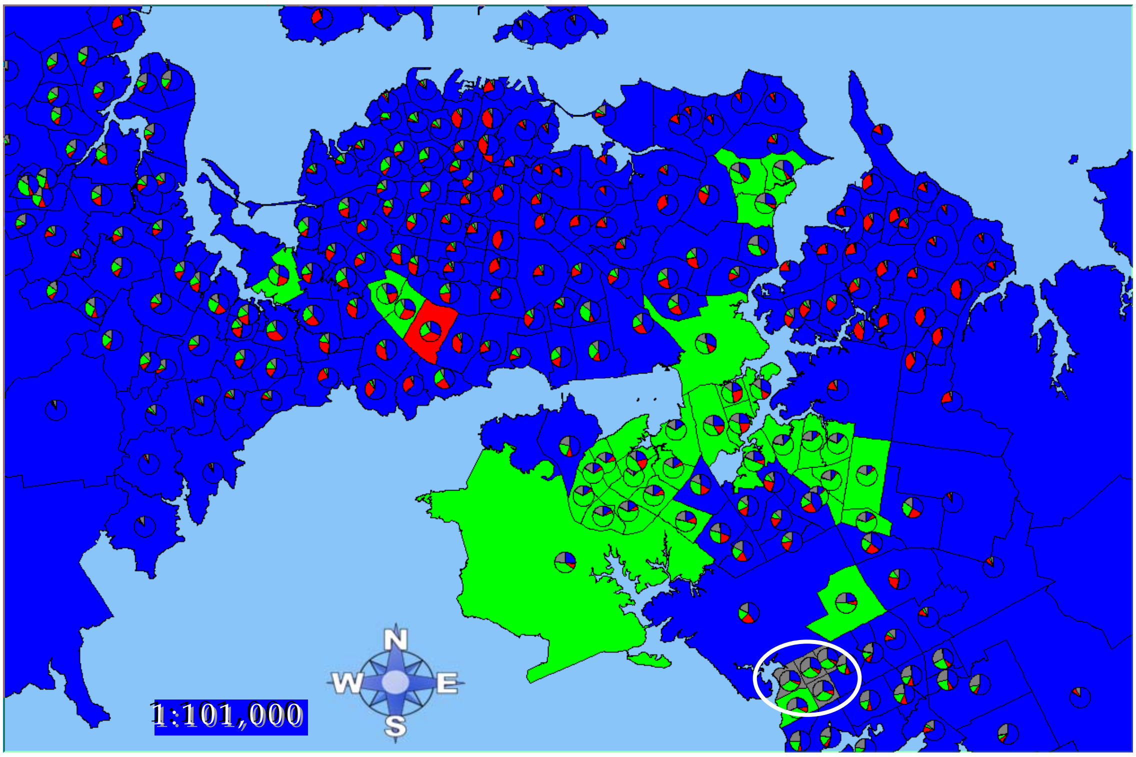


Figure 6.5 Zoomed Map of 2001 with Identified Maori AUs in Plurality

Important features of the geographical maps with census administrative boundaries also manifests through detailed representation of population proportions using pie-charts. For example, in figure 6.5, although many area units are still dominated by the presence of the European population (in blue), the growing Asian population (in red pie slices) visually imposes themselves. In this way, changes are not uniquely noticeable at the aggregate level when transformation of plurality status in spatial units occurs. It is actually possible to monitor the micro geographical level of population change in each area unit as well.

These maps (along with the actual counts of the area units with ethnic plurality) will be used as visual references in order to compare the results produced by the simulations for discerning qualitative changes in the urban area landscape.

The Unbalanced Nature of Changes

In general this thesis deliberately avoids a focus on local changes (e.g. by using LMI and LQ measures). However, it uses pertinent examples related to some of the 316 individual area units to demonstrate the effects of changes locally at the lowest geographical scale in the model (see also section 6.3.3).

The overall changes from the aggregate level and macro geographical perspective (i.e. at urban area scale) were examined in previous sections. However, these changes did not occur simultaneously and with the same intensity in smaller geographical constituents.

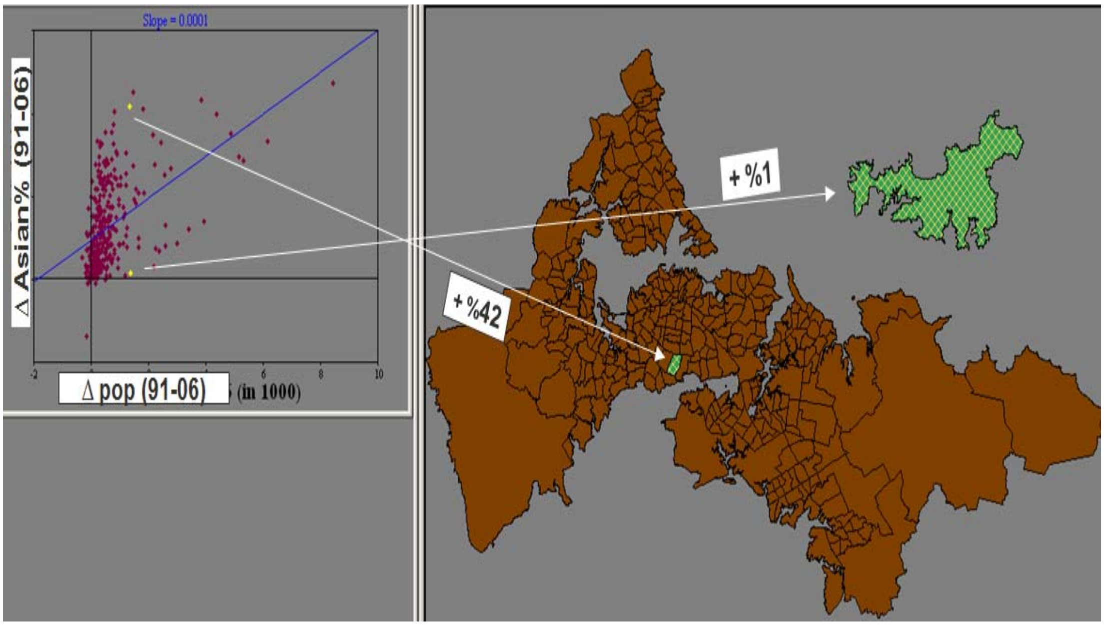


Figure 6.6 Changes in Population vs. Changes in Asian Population (1991-2006)

Figure 6.6, for example, shows the overall population change in the horizontal axis versus the population change of the Asian ethnic group in the vertical axis from 1991 to 2006. Every dot in the graph (left side of figure) represents an area unit. Finding two or more dots (area units) in the same horizontal axes but with a large distance in the vertical axes indicates that although the change in the size of the population has been exactly (or roughly) the same, the increase of the Asian population in those area units has been quite different. As can be noticed in this figure, the area units of *Hillsborough West* and *Waiheke Island* (shown by yellow dots and green polygons) had roughly the same level of population change (growth) over last the 15 years (1828 and 1965 individuals respectively). However, their share of the Asian ethnic group population is remarkably different: *Hillsborough West* with 2931 (a 42% increase), whilst *Waiheke Island* had only 109 (a 1% increase).

When we look at the counts of area units based on ethnic plurality presence in each of the five meso-geographical entities (TA), the intensity of change is mostly concentrated in two of them.

As can be seen in table 6.4, all the area units in the territorial authorities of *North Shore* (NSH) and *Papakura* (PAPK) have constantly been dominated by Europeans (E) the largest group in 'town' from 1991 until the last census in 2006. Except for a single area unit in plurality made by Asians (A) in 2006, the situation is the same in *Waitakere* (WTK) as well. The main changes have occurred in the *Auckland* (AKL) and *Manukau* (MKU) districts where Asian and Pacific people (P) have become new plurality forces in previously mostly European area units.

Table 6.4 Counts of AUs Where Ethnic Group Forms Plurality (TA-based)

	NSH				WTK				AKL				MKU				PAPK			
	E	A	P	M	E	A	P	M	E	A	P	M	E	A	P	M	E	A	P	M
1991	53	0	0	0	55	0	0	0	103	0	1	0	67	0	20	0	17	0	0	0
1996	53	0	0	0	55	0	0	0	100	0	4	0	64	0	21	2	17	0	0	0
2001	53	0	0	0	55	0	0	0	92	1	11	0	60	0	23	4	17	0	0	0
2006	53	0	0	0	54	1	0	0	84	11	9	0	53	4	30	0	17	0	0	0

In chapter 7, we will see that under certain realistic scenarios, changes can also occur in stable districts (TAs, such as *Waitakere*, *Papakura* and *North Shore*) which have historically been dominated by a single European 'majority' group. This will also demonstrate the potential of the model as a useful tool for forecasting these changes in the future under potentially different conditions.

Change Trends in TAs by SNZ Projections

In 2010, the SNZ released its projections regarding the population change (until 2021) for four major ethnic groups in different territorial authorities based on 2006 values.

The projections were given as average annual changes based on three alternative series designated 'high', 'medium' and 'low'. These projections are used and simulated in the model and compared in the section 7.5.1. The details of these projections can be found in tables included in appendix G.3.

6.3.2 Changes in Segregation Dimensions

Indicators of change in segregation dimensions for four major ethnic groups in the Auckland region (5TAs) based on available census periods are presented and discussed in this section.

Table 6.5 Isolation/ Interaction Index for Metropolitan Area

Year		Euro	Maori	Pacific	Asian
1991	**Euro**	0.7768	0.0859	0.0817	0.0556
	Maori	0.5815	0.1651	0.1990	0.0543
	Pacific	0.4798	0.1727	0.2887	0.0588
	Asian	0.6945	0.1003	0.1251	0.0801
1996	**Euro**	0.7208	0.0955	0.0849	0.0987
	Maori	0.5682	0.1598	0.1874	0.0847
	Pacific	0.4535	0.1683	0.2912	0.0870
	Asian	0.6543	0.0943	0.1079	0.1436
2001	**Euro**	0.6886	0.0890	0.0877	0.1347
	Maori	0.5211	0.1565	0.2058	0.1166
	Pacific	0.3927	0.1574	0.3321	0.1178
	Asian	0.5956	0.0881	0.1164	0.1999
2006	**Euro**	0.6498	0.0846	0.0872	0.1784
	Maori	0.4806	0.1500	0.2123	0.1572
	Pacific	0.3522	0.1508	0.3402	0.1569
	Asian	0.5282	0.0819	0.1151	0.2749

The isolation/ interaction index (I.I) values (shown in table 6.5) provide a measure of population composition and their exposure. As previously explained (see section 5.2.1) the index has two forms. The isolation index (values in diagonal cells) can be seen as a probability for members of the same group sharing the space with 'alike' people. The interaction index (values in non-diagonal cells) on the other hand is the probability for members of a given group sharing the space with members of another group. The I.I-values in non-diagonal cells are asymmetric. For example, in 1991 a randomly selected European individual had about a 6% chance of meeting a person of Asian origin in the metropolitan area, whilst the likelihood for an Asian to encounter a European was about 69%. In 2006, these values are respectively about 18% and 53%. These changes show how considerably the exposure of the Asian population to Europeans has increased over the last 15 years, whereas the exposure of Europeans to Asians has decreased during the same period. As was explained in chapter 5, the exposure (isolation/ interaction) dimension is dependant of population size and its composition, and hence, the changes in exposure values can easily be explained by the growth of population and their proportion sizes.

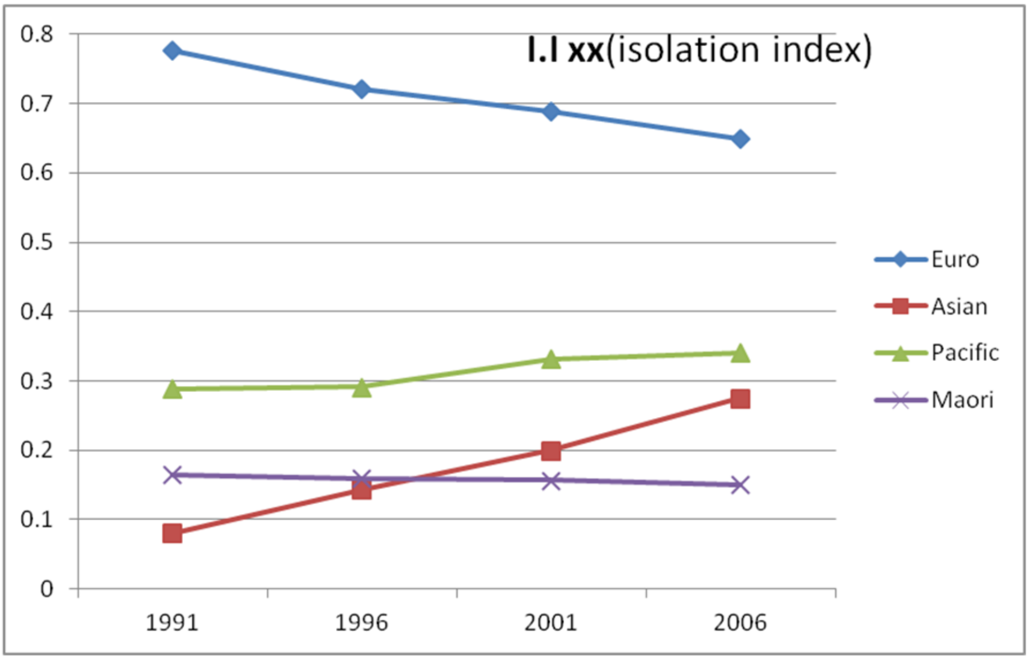


Figure 6.7 Isolation Indices for Metropolitan Area

The isolation values are plotted as time-series in figure 6.7. In 1991, on the average a randomly selected Asian could meet another fellow Asian about 8 times out of 100. This chance increased significantly to about 27 times out of 100 in 2006. The Pacific people on the other hand have seen an increase from about 0.29 in 1991 to 0.34 in 2006. As we saw, the population growth of Pacific people has not been as intense as that of the Asian population in the same period. But more importantly, the aggregate indices also indicate that compared to Pacific people, Asians have a lower likelihood of encountering each other. This is because they are scattered (distributed) in different neighbourhoods throughout the entire large urban area, whereas Pacific people have a high concentration in specific areas. On the other hand, the trend has been reversed for Europeans (and to lesser extent for Maoris) who have lost a significant portion of their proportional weights over the last 15 years. On average, a randomly selected European could meet another European about 72 times out of 100 in 1991, but only about 65 times out of 100 in 2006.

H, D and GMI values are presented in table 6.6. These measures are particularly sensitive to the distribution of ethnic groups across micro spatial entities. The most known and used value, the index of dissimilarity (D), is a measure of the evenness for binary group comparison. Unlike I.I, the dissimilarly index (D) values are symmetrical. For example, the dissimilarity of Asian vs. European (or European vs. Asian) has increased from about 0.29 in 1991 to about 0.35 in 2006. These values can be interpreted as a percentage of one of these groups that needs to move to different area units in order to create a distribution that would match that of the larger urban area. In 2006, a higher percentage of Asians and Europeans needed to move around (comparing to 2001) to create a better balance between members of these two groups.

Table 6.6 H, D and GMI for Metropolitan Area, Diagonal cells contain Global Moran's I, non diagonal cells contain the D index.

Year	H		Euro	Maori	Pacific	Asian
1991	0.1617	**Euro**	***0.6661***	0.4104	0.5439	0.2915
		Maori		***0.6529***	0.2729	0.3895
		Pacific			***0.6698***	0.4339
		Asian				***0.5071***
1996	0.1426	**Euro**	***0.6647***	0.3660	0.5315	0.2842
		Maori		**0.6992**	0.2897	0.4148
		Pacific			***0.6780***	0.4905
		Asian				***0.5739***
2001	0.1596	**Euro**	***0.6791***	0.3906	0.5586	0.3166
		Maori		***0.7402***	0.3007	0.4384
		Pacific			***0.7175***	0.5086
		Asian				***0.5800***
2006	0.1671	**Euro**	***0.6870***	0.3966	0.5722	0.3460
		Maori		***0.7689***	0.2983	0.4476
		Pacific			***0.7428***	0.5268
		Asian				***0.5765***

By looking at the plotted D-values shown in figure 6.8, we can observe that the intermixing imbalance between Europeans & Pacific people, Asians & Pacific people, and Asians & Maoris have deteriorated more than that, for example, of Pacific people & Maoris. In fact, the unevenness has been slightly (although not significantly) improved (or at worst, remained the same) between Europeans & Maoris. The intermixing imbalance has always been high between Europeans & Pacific people (highest among all too), and it got slightly worse in 2006. On the other hand, it is still the deterioration of the Asian groups' relationship in this regard with all the other groups that emerges as the most perceptible trait. As Asians have become increasingly present in more numerous area units throughout the urban area (although not in all), the lack of an ideal balance between their population size and that of other groups manifests in terms of higher D values.

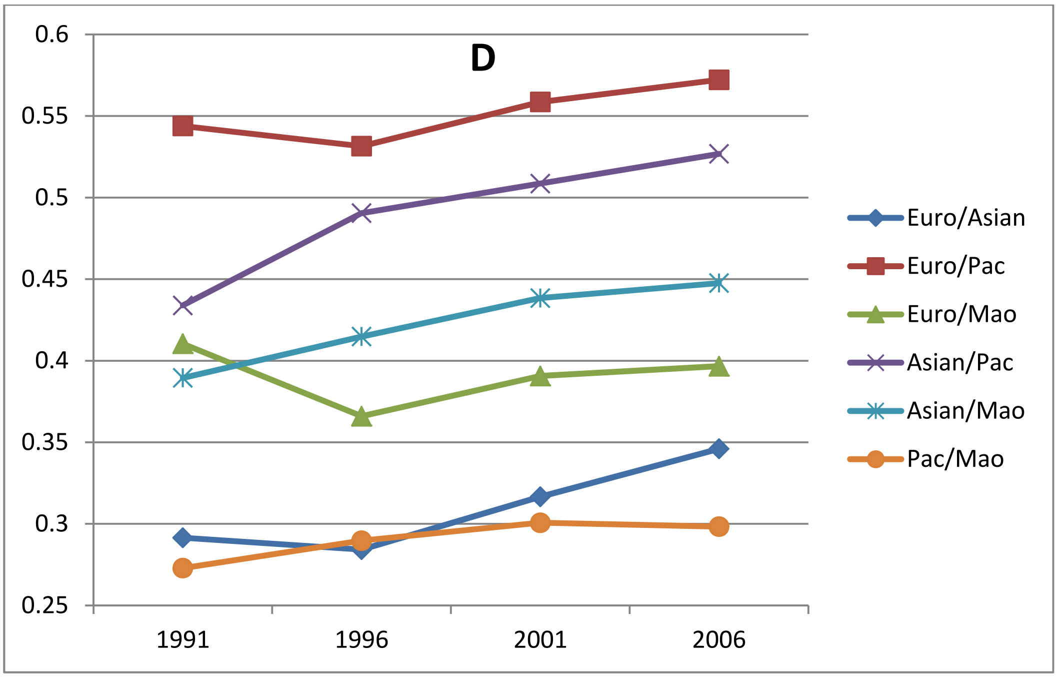


Figure 6.8 D-values for Metropolitan Area

The 'aspatial' aspect of D as a global segregation measure and its lack of spatial sensitivity have been reported and discussed in the literature (see, for example, Brown & Chung, 2006, pp. 127, Figure 1), and it constitutes an important limitation of this measure. This means that although we know the groups are not evenly distributed throughout the urban area, we have no idea how their spatial patterns looks like. It is partly because of this lack of spatial sensitively in D and other measures (including H and I.I) that this thesis (and the model) also uses Moran's I as a complement measure.

However, before examining the MI-values, it would insightful to also look at the multi-group measure of H, which is the weighted average deviation of each area unit's entropy from the entire metropolitan area entropy (plotted also in figure 6.9). Like D, this is another global measure for measuring unevenness by focusing on how diversely ethnic populations are distributed throughout urban area spatial units. Despite the up and down in the plotted H-values in figure 6.9, the intensity of variations differ by second decimal digits. Nonetheless, the result of comparison between H-values in 2006 and 1991 reveals a growing disparity in terms of diversity of ethnic groups throughout the urban area. The drop in the 2006 value may be explained by a better (larger) presence of the three minority ethnic groups (in particular Asians) versus the European majority. Between 1991 and 1996, the overall percentage of population growth for Asians, Maoris and Pacific people was about 87%, 13% and 9% respectively, while the Europeans had actually lost about 1% of their population during the same period. Hence, it is possible that with the more tangible and diverse presence of the minority groups – in particular the Asians and to some extent Maoris – in wider area units throughout the Metropolitan area, the global H-value declined slightly in 1996.

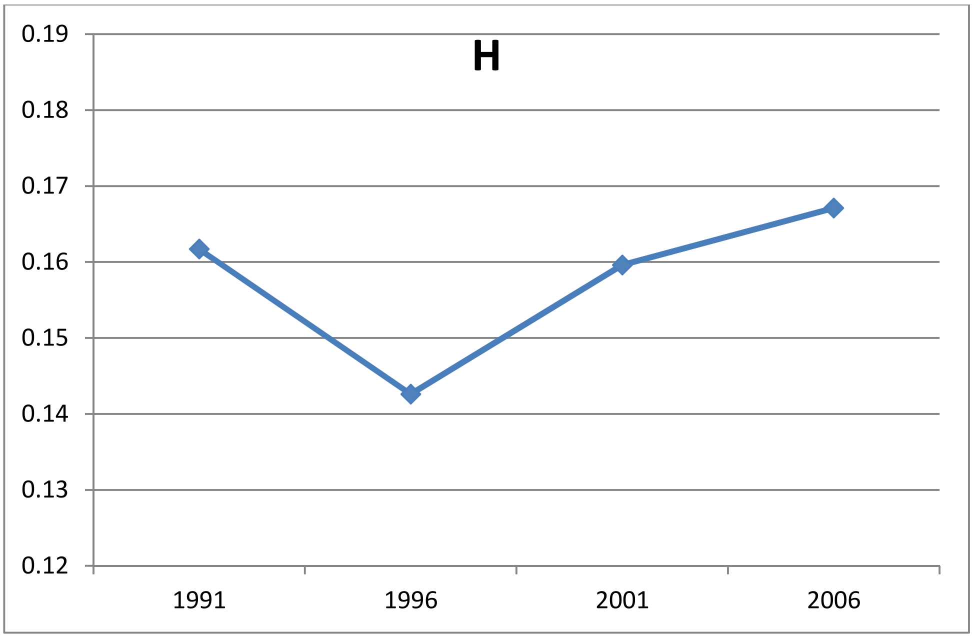


Figure 6.9 H-values for Metropolitan Area

The percentage of growth or decline of population size was given based on the data values presented in table 6.3. Although it is believed that these values to great extent represent the actual number of people counted in the censuses, it cannot be excluded that a small percentage of respondents were actually misled by the census question on ethnicity in 1996, as was explained in section 6.2.3.

The global Moran's I values are shown in figure 6.10. These are the values that are obtained using the proportion of groups (as variable of interest). It is believed that this would provide a more accurate representation of the global clustering tendency in the urban area as will be explained.

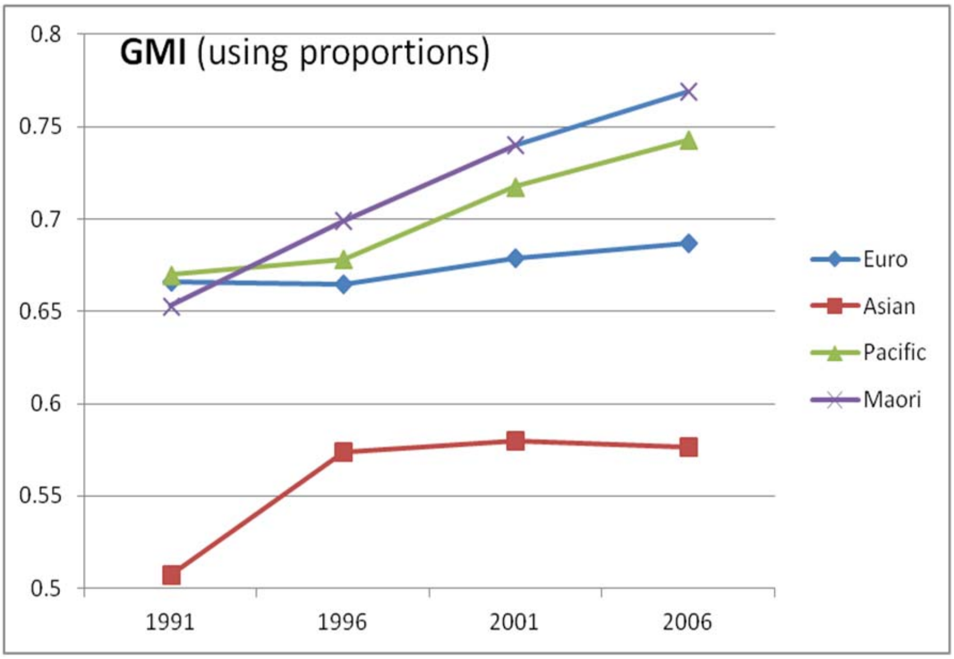


Figure 6.10 GMI-values for the Metropolitan Area (Using Proportions)

GMI-values in this case are the aggregate results of individual autocorrelations calculated for an ethnic group's proportion in an area unit relative to its adjacent neighbourhood, repeated for the entire urban area. The positive values indicate that in aggregate, the area units' ethnic group proportions were positively correlated with their neighbouring area units (i.e. either having high proportions surrounded by high proportions, or low proportions surrounded by low proportions). As such, this spatial dependency index was already elevated for Europeans, Pacific people and Maoris and became even worse over the last 15 years, particularly for Maoris and Pacific people. On the other hand, although the GMI-value increased for Asians in 1996, it has been relatively stable since then, indicating there is a lower and more stable clustering level for this ethnic group.

It should be noted that a high level spatial autocorrelation for a group does not always translate into a visual and apparent clustering on the geographical map of the city (for example, see Figures 5.10 and 5.11, where T_2 = 0.18 had some degree of positive autocorrelation at the end the simulation but had a geographical map with no clear/ discernible sign of clustering). This is because the area unit turns to a group's colour only if the group is in plurality (i.e. has relative majority status). On the other hand, a group can have a very high spatial correlation between its proportion percentages in one unit relative to other neighbouring areas without being a plurality 'force'. This is indeed the case of Maoris which have the highest level of GMI without being much visible in the geographical map of the city. However, a careful (or zoomed mode) of the map that show the proportion of population sizes in pie-charts can expose these spatial correlations.

Finally, it should be obvious that these global aggregate measures at macro geographical scale would be different in smaller meso divisions of the city. As an example, figure 6.11 shows the H-values in each of the 5 districts (TAs) which together form the metropolitan area of Auckland. *Manukau* has the highest level of segregation measured by H for it is home to the largest Pacific people population with a seemingly lesser representation (diverse distribution) of other ethnic groups (the situation has been more or less the same since 1996).

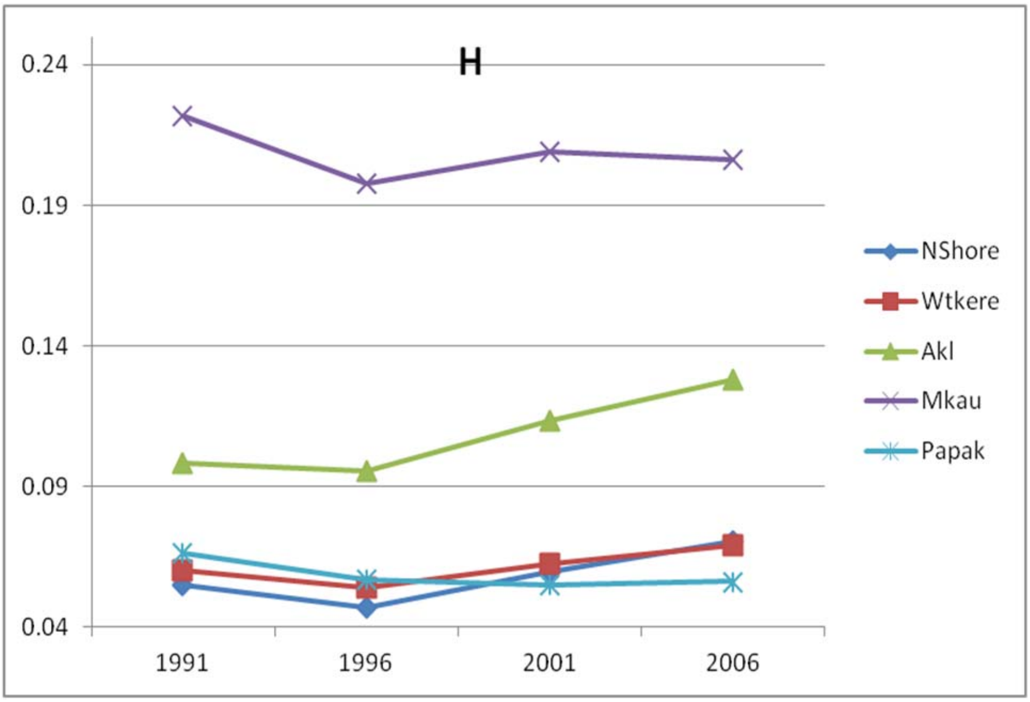


Figure 6.11 H-values for Five TAs

On the other hand, the diversity characteristic of *Auckland* TA (and to lesser extent in *North Shore* and *Waitakere*) has deteriorated since 1996. This is most likely because the large immigration of Asians has made the entropy of each constituent area unit more divergent from its containing district. *Papakura* remains the district where this entropy divergence was the least in 2006 among the 5 TAs (lowest segregation measured by H index).

In general, when describing segregation levels relatively and qualitatively using terms or adjectives such as 'very low', 'low', 'moderate', 'high' and 'very high' and the like, a 'cut-point' is used (often in a rather *ad-hoc* manner) to separate the values based on these categories (see, for example, Massey & Denton, 1989).

6.3.3 Local Changes and Hottest/ Coldest Spots: the Case of Asians

The global measures such as those used and presented in the previous section are not able to capture local variation (Brown & Chung, 2006). On the other hand, unless the results can be presented in an aggregate and yet meaningful format, using measures that enable us to discern changes at the local level may not really be practical if the numbers of local areas are numerous (such as for the environment in this thesis consisting of 316 area units).

Two different spatial measures and one spatial tool were used to investigate the changes at local level and to identify the 'hottest' or 'coldest' spots. This investigation is performed for the measures obtained for the Asian ethnic group in the metropolitan area (with 5TAs). Since there are 316 area units in this environment, the number of 'hot spots' (high values) or 'cold spots' (low values) using recommended 'cut-points' in some literatures would still lead to the identification of high number of cases (which make the tracking process unmanageable or at least less noticeable). For instance, to gauge the significance of location quotient (LQ), Brown and Chung (2006) use a LQ of 1.2 or greater or a LQ of 0.85 or less to indicate respectively the significant concentration or under-representation of ethnic groups. Using these values for the 316 area unit environment will lead to the identification of numerous hot and cold spots of the order of 100 to 200. Similarly, they propose using z-scored local Moran's I (LMI) of z = +0.667 or greater to indicate significant clustering and z = -0.667 or less to indicate significant under-representation of the ethnic group of concern. Similarly, this will lead to the identification of a large number of potentially significant area units.

Table 6.7 LMI & LQ-values, as well as Hottest & Coldest Spots for Asians
The area units identified by bold* indicate the hottest/ coldest spots detected by GeoDa's univariate LISA tool utilizing each group's proportion values.

	1991		1996		2001		2006	
	LMI	**LQ**	**LMI**	**LQ**	**LMI**	**LQ**	**LMI**	**LQ**
Max	36.6055	5.4953	56.1385	4.3661	33.2501	3.1184	33.6828	3.0539
Min	-15.282	0	-3.8331	0	-4.2961	0.0650	-4.1337	0.0468
Mean (μ)	2.5770	1.0317	2.9168	0.9699	2.9480	0.9616	2.9301	0.9445
SD (σ)	4.5566	0.7708	5.7820	0.6826	4.6783	0.6630	4.3818	0.6468
# >= μ+2σ	12	12	12	15	16	15	16	12
# <= μ-2σ	1	0	0	0	0	0	0	0
LMI-LQ Intersection (Hottest)	6 **Grafton E.*** (Akl.), Grafton W. (Akl.), Sandringham E. (Akl.), Otahuhu N. (Akl.), **Millhouse*** (Man.), Dannemora (Man.)		8 Epsom central (Akl.), Meadowbank S (Akl.), **Lynfield N**.* (Akl.), Maungamanungaroa (Man.), Golfland (Man.), **Millhouse*** (Man.), **Dannemora*** (Man.), **Kilkenny*** (Man.)		8 Auckland central E. (Akl.), Epsom central (Akl.), Hillsborough W. (Akl.), **Akarana*** (Akl.), **Lynfield N.*** (Akl.), **Millhouse*** (Man.), **Dannemora*** (Man.), **Kilkenny*** (Man.)		8 Lynnmall (Wait.), Aukland central E. (Akl.), Epsom central (Akl.), Hillsborough W. (Akl.), **Akarana*** (Akl.), **Lynfield N.*** (Akl.), Pigeon Mountain N. (Man.), **Dannemora*** (Man.)	
LMI-LQ Intersection (Coldest)	0		0		0		0	

The term 'hottest' or 'coldest spot' refers to the process presented in this thesis section which employs both local indices of LMI and LQ, along with a spatial tool to identify those in common (i.e. those area units which have highest values of LMI and LQ and are identified by the spatial tool as well). Only a handful of area units have gone through acute change (or lack of it). For LQ and LMI, the significant values should be greater or lower than the mean value (μ) plus two times the standard deviation (SD or σ). This process is shown in table 6.7. For example, the counts of area units where their LMI and LQ values were equal or greater than μ + 2σ in 2006 are respectively 16 and 12. However the intersection between these identified outliers (i.e. numbers of the area units where both of these values are present) will only lead to the selection of 8 hottest spots (outmost outliers). The reason for using both of these local indices (LQ and LMI) is because one is better for spotting concentrations while the other is better for identifying clusters.

As an additional test GeoDa[16]'s univariate LISA tool was used by utilizing ethnic group proportion values. This process is about evaluating spatial dependence; that is, measuring the functional relationship through the correlation between each observed area unit and the rest of the area units to which it is spatially linked.

A GeoDa's 'LISA cluster map' (see figure 6.2) was then constructed which depicted four types of spatial association: similar positive local spatial autocorrelation values were categorized as *high-high* (high values surrounded by neighbours of similar high values) or *low-low* (low values surrounded by neighbours of similar low values), as well as dissimilar negative local spatial autocorrelation values labelled as *high-low* (high values surrounded by neighbours of low values) and *low-high* (low values surrounded by neighbours of high values) (Anselin, 2005).

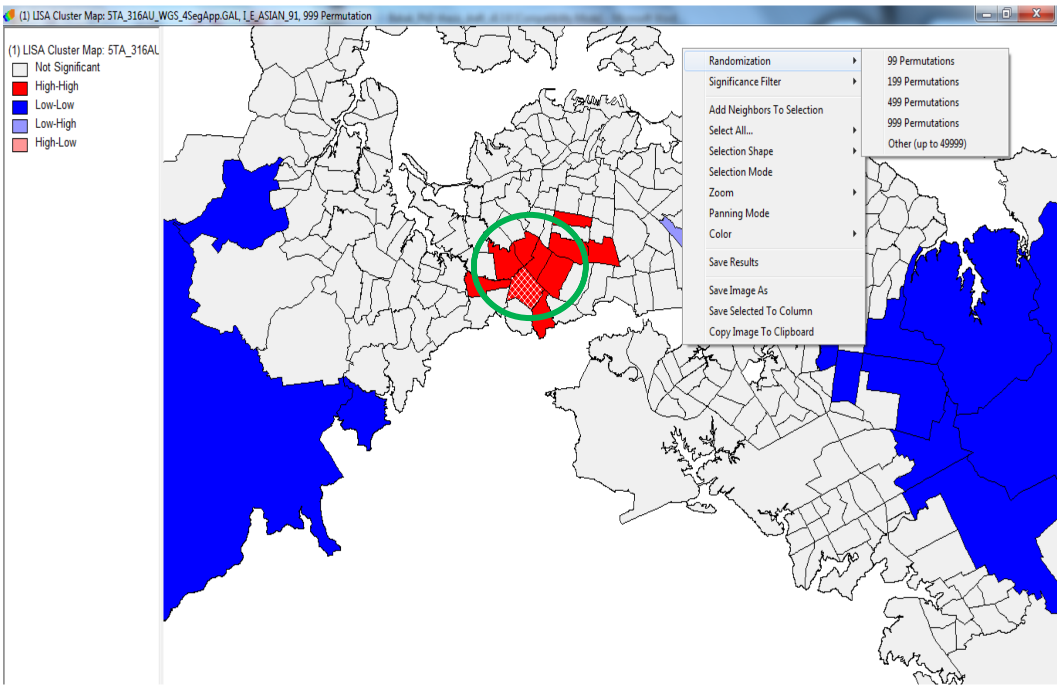


Figure 6.12 GeoDa LISA Cluster Map Example

[16] GeoDa is a software package mostly for spatial data analysis and geovisualization; https://geodacenter.asu.edu

For this analysis, a combined randomization/ permutation and significant tests were performed using 999 permutations at a significance level of $p \leq 0.001$ for *high-high* category values. Limitations of this test are that it is sensitive to the number of selected permutations and it can lead to slightly different results between permutations. However, several permutations (of the same order, i.e. 999) were performed after which exhaustive results was obtained. Those significant area units identified by GeoDa's LISA cluster tool and intersected with the hottest spots identified by the process described earlier are shown in bold and with asterisks (*) in front of their names in table 6.7.

For example, the area unit of *Lynfield North* (highlighted with mesh-like white in the map, figure 6.12) in *Auckland* TA has been identified by both LQ and LMI measures as well as GeoDa LISA cluster (at 0.001 significant level) as one of the hottest spots since the 1996 Census year. In 2006, its percentage of the Asian ethnic group was slightly more than 50%, a dramatic increase from about 14% in 1991. It is also surrounded by neighbours with a relatively high concentration of Asians. For instance, *Akarana*'s Asian proportion has jumped from about 12% in 1991 to about 46% in 2006 (the *Akarana* area unit is actually among hottest spots since 2001). *New Winsdor's* Asian proportion from 1991 to 2006 jumped from 12% to 43%, that of *Waikowhai West* from 12% to 38%, *Blockhouse Bay* from 6% to 29%, *Avondale South* from 8% to 39%, *Walmsley* from 10% to 31%, and *Wesley* from 12% to 40%. Similar situations were also true for other places such as *Dannemora*, although *Millhouse* which has been a hottest spot between 1991 and 2001 has not been 'labelled' as such in 2006.

As can be seen, the hottest spots have been either located in the *Auckland* or *Manukau* TAs, with the exception of one new entry in 2006 (*Lynnmall* area unit) from *Waitakere* TA.

6.4 Summary and Conclusion

Due to practical reasons (e.g. the large scale of two mostly uninhabited districts), the metropolitan area of Auckland region for this study will comprise the five most important and populated districts (territorial authorities). Furthermore, conforming to the available data, individuals are the most micro focal point entities incarnated as agents in the model.

Data had to be adjusted due to the existence of known 'noise' (including errors). However, the data should not be considered totally noise-free. Certain 'inherent' levels of noise are already embedded (e.g. due to change of questions, etc.) in the census data so it would be virtually impossible to completely eradicate.

In general, our initial impression, assumption or certainty about the existing discrepancy and errors in data needs to be challenged, and uncertainty analysis needs to be included. This is particularly the case if the model is to fully evolve for the evaluation of policy-related scenarios.

The analysis of census data – both qualitatively and quantitatively – exposed the extent of change in the ethnic residential landscape of the Auckland metropolitan region over the last 15 years. Two findings were particularly significant and by far the most noticeable. First was the high level of segregation (in particular concentration) of Pacific people in a few area units of the city. The second was that although the growth of Asian population had been remarkable – particularly in the past 10 years – their distribution was relatively more widespread (and their level of segregation lower), particularly in comparison with Pacific people. This is despite the fact that by 2001, Asians had become the second largest ethnic group after Europeans in the Auckland metropolitan region. In the scenarios presented in the next chapter, particular attention will be given to these two ethnic groups (e.g. as key variables of interest during the calibration) in order to replicate these observed spatial characteristics.

Most of these quantitative values (along with the qualitative maps of segregation based on census data) presented and examined in this chapter will form a series of real empirical benchmarks that will be used to inform and validate the scenarios and simulated results in the next chapter.

Chapter 7

7. Simulation of Scenarios

7.1 Introduction

This chapter is dedicated to the simulation of selective realistic scenarios. It begins with explaining the scenario building approach followed by a proposed pattern-oriented method for calibration of the model. The simulation outputs are analysed and compared with the census based historical data to identify similarities and explain the divergences. A series of tables and graphs are used to describe the generated outputs alongside historical census data based measures. A number of policy-oriented scenarios will be presented where the simulations are run for a longer period than for which census based data exist. The projections of scenarios analysis will demonstrate the potential of the model to be used as a decision support and policy informing tool for shedding light onto some *what-if* questions.

7.2 Scenario Building Approach

As explained previously, the parameter space for building various scenarios in this model is large and so would be difficult to explore fully in this thesis. One of the characteristics that emerged from the exploration of the census data was the relative integration (or at least more widespread distribution) of the Asian population in contrast to the high concentration of Pacific people in a few areas within the urban metropolitan area. Some of the scenarios presented in this chapter will demonstrate the ability of the model to replicate (to some extent both quantitatively and qualitatively) the distribution and contrast between Asian and Pacific people based on historical census-based observations without detailed empirical information of the decision making behaviour of householders.

The potential validation of this model using census data has been a key impetus behind its development. In this thesis the validation term is defined as referring to 'external' (or 'operational') 'validity' which is "concerned with the linkage between the simulated and the real [world]" or the "adequacy and accuracy of the computational model in matching real world data" (Carley, 1996).

This thesis avoids entering the debate about whether it is truly possible to establish the validity of a model (e.g. "all models are wrong, but some are useful" (Box, 1976)). For more on this perspective, see, for example, Oreskes et al. (1994), Sterman et al. (1994) and David (2009).

However, it is useful to recognize that there is no such thing as absolute model validity, and the aim is simply to gain sufficient confidence in the model (see, for example, Law, 2009; Robinson, 1997). After all, simulation modelling of complex residential segregation phenomena can always only be an approximation of the actual system. This in part, is because of the complexity of the system, but more importantly due to the principle of model building which stipulates a model should normally be a simpler version of the real world. Therefore, a model has simply "a certain degree of validity" (Castle & Crooks, 2006) and cannot be classified in a binary manner as valid or invalid.

Various approaches have been suggested in the literature for the process of validation. These include subjective evaluation by the development team based on various tests (Sargent, 2008), and experts' opinion as validation evidence (Brade, 2003; Carley, 1996).

However, in line with the thesis objective and because of the nature of data, the validity of this model will mostly be ascertained by comparing model outputs (and behaviour) with comparable data (or system behaviour) from a real-world system (Balci, 1994, p. 121; Castle & Crooks, 2006, p. 37) using graphical comparison (Sargent, 2008). The intended approach in this thesis can also be called "pattern validity"; that is, when the pattern generated by the model matches a real world pattern (Carley, 1996, p. 10). This process is akin to Grimm et al.'s (2005) strategy of Pattern-Oriented Modelling (POM). The idea behind POM is to use multiple patterns observed in real systems that seem to characterize the system and its dynamics and then 'tune' certain variables in the model so that these patterns could emerge in the model (Grimm et al., 2005, p. 987). "Empirical observations are used to develop both theories and the patterns used to test and falsify them" (Grimm & Railsback, 2006, p. 139).

The patterns of interest in this chapter will show the changes observed in the census data – explained in the previous chapter.

Finding the "Medawar Zone" (Grimm et al., 2005) (i.e. finding the optimal level of resolution or areas which are most likely to produce fruitful results) in the POM strategy is similar to calibrating a model; that is, a process of tuning a model to fit detailed real data. This "approach is generally used for showing that it is possible for the model to generate results that match the real data" (Carley, 1996). Carley also addresses the critics of calibration who "argue that any model with sufficient parameters can always be adjusted so that some combination of parameters generates the observed data" and thus "calibration does not establish the validity of a model in a meaningful way" (p. 16). Carley does so by asserting that this criticism is less appropriate for "models where the process is represented not by parameterized equations but by rules, interactive processes" (p. 16).

In fact, the 'indirect calibration', 'Werker-Brenner calibration' and 'history-friendly' approaches discussed by Windrum et al. (2007) and Moss (2008) all aim to restrict (constrain) the parameter space of the model using different but comparable techniques. The history-friendly approach, for example, attempts to first use data to assist "the identification of initial conditions and parameters on key variables likely to generate the observed history" and then to "empirically validate the model by comparing its output (the 'simulated trace history') with the 'actual' history …" (Windrum et al., 2007).

In an interesting paper by Axtell et al. (2002) on modelling the historical society of Kayenta Anasazi and its population – cited also by Grimm et al. (2005) as one example of the "indirect parameterization" approach – the aim is to "generate model realization[s] having a total population closest to the historical data" ("best realization as well as best average set of runs"), so that in each period of the model, the number of simulated households at specific times are compared to the historical record and that optimizing the model with respect to the eight adjustable parameters yields the best configurations (Axtell et al., 2002, p. 7277) . In this paper, Axtell et al. present only a single figure (graph) of the best single run of the model where the pattern of simulated (output) number of (current) households is reasonably similar to the historical household pattern. They do though state that "other best runs based on other norms yield very similar trajectories" and that "the average run, produced by averaging over 15 distinct runs, looks quite similar to this one as well" (p. 7278).

The approach used in this thesis is relatively similar, except that the process does not depend on numerical optimisation methods to restrict (constrain) the parameter space. The visual comparison of differences (in size and patterns) is ultimately used for selecting more promising configurations of the model, while rejecting others (to restrict the parameter space). In part – to reduce uncertainty – simulations are repeated with different random seeds to make sure the conclusions would be typical.

7.3 Pattern-Oriented Calibration

The term Pattern-Oriented Calibration (POC) refers to the process employed in this thesis to tune the model so that it can generate plausible patterns which can best match the best the pattern(s) obtained from the real world or phenomenon under investigation (e.g. by using empirical or census data) for the envisaged scenario(s).

In general, patterns can be obtained in a qualitative or quantitative manner, although quantitative measures are still 'mapped' to graphs and plots for visual comparisons. The idea is to have not only numerical points as close as possible to the empirical data, but also to the shape of the real world pattern as whole. This is because census data are only available every five years, whilst the simulation time step is usually shorter (annual in this case.)

The pattern-oriented calibration can be seen as an iterative, funnel-like method (depicted in Figure 7.1). If the parameter space is not large there will only be one (set of) variable(s) of interest (VOI). This can theoretically be done by a single iteration. However often, even if enough computation power exists, to sweep over a large and yet sensible parameter space with single batch (parameter) file may be impracticable, for it would be necessary to compare numerous large results – both numerically and visually – in order to pick up the best. Therefore, the POC can be useful regardless of computational power, as long as visual verification is also an important characteristic of the process.

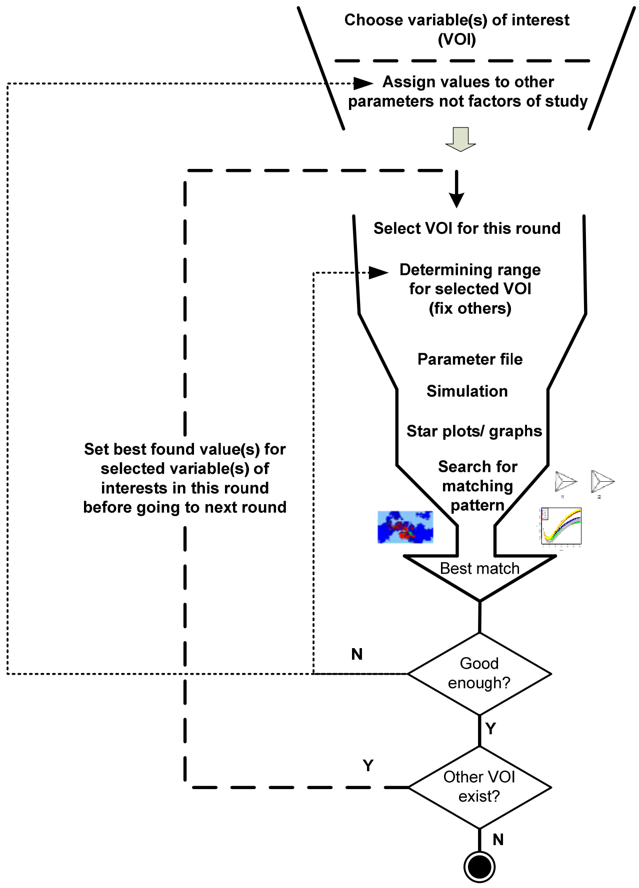


Figure 7.1 The Pattern-Oriented Calibration (POC) Method

The procedure for the pattern-oriented calibration can be described as follows:

1. Choose a variable(s) of interest [e.g. L: local-global | T: tolerance preference]

2. Assign sensible values to other parameters (parameters that are not factors of study [e.g. P(M) for all groups])

3. Select the variable(s) of interest to vary for the current iteration round [e.g. T]

4. Establish a sensible range of values to vary selected variable(s) in step 3, whilst fixing other variable(s) of interest (to be used in the next round) – if there is any

5. Build a parameter file (for batch execution)

6. Perform the simulation (as batch)

7. Generate graphs and star plots

8. Search visually for the best match (using graphs, plots and spatial/ qualitative maps) and mark down the value(s) that generated them

9. If the match is not satisfactory, return to step 4 and try another value range (in some cases, it may be necessary to return to step 2 and change the values of those parameters which are not factors of study); otherwise, when satisfactory (and when still other variables involved in the calibration exist), set (as constant) the value of the calibrated variable(s) to the value(s) marked in step 8, and return to step 3 by choosing the next round of variable(s) of interest for calibration (e.g. L).

Figure 7.2 shows a part of the POC process where the level of turnover for all groups and the tolerance and locational mobility levels for Europeans and Maoris are fixed, and the level of tolerance for Asians and Pacific people are varied within a plausible range with consideration of their proportional average sizes at the beginning and end of simulation (8% to 15% for Asians and 9% to 15% for Pacific people, both at 1% intervals). The selected combinations generated 56 different pattern outputs which were tested against empirical benchmark patterns (in this case, H-values at MA level), using star plots for quick detection.

Star plots are a useful graphical way to display multivariate observations with an arbitrary number of variables (here 3). It consists of a sequence of spokes, where each spoke represents one of the variables (here, the absolute difference between simulated H-value at specific tick year, and its corresponding census-based empirical value for years 1996, 2001, and 2006). By looking at the generated stars, it would be possible to verify their relative values as well as the similarity or dissimilarity among them. It would also be possible to see whether there are clusters of observations. For example, as can be seen in Figure 7.2, T-values of less than 11% (or greater than 13%) for both Asians and Pacific people do not clearly generate close (favourable) matches.

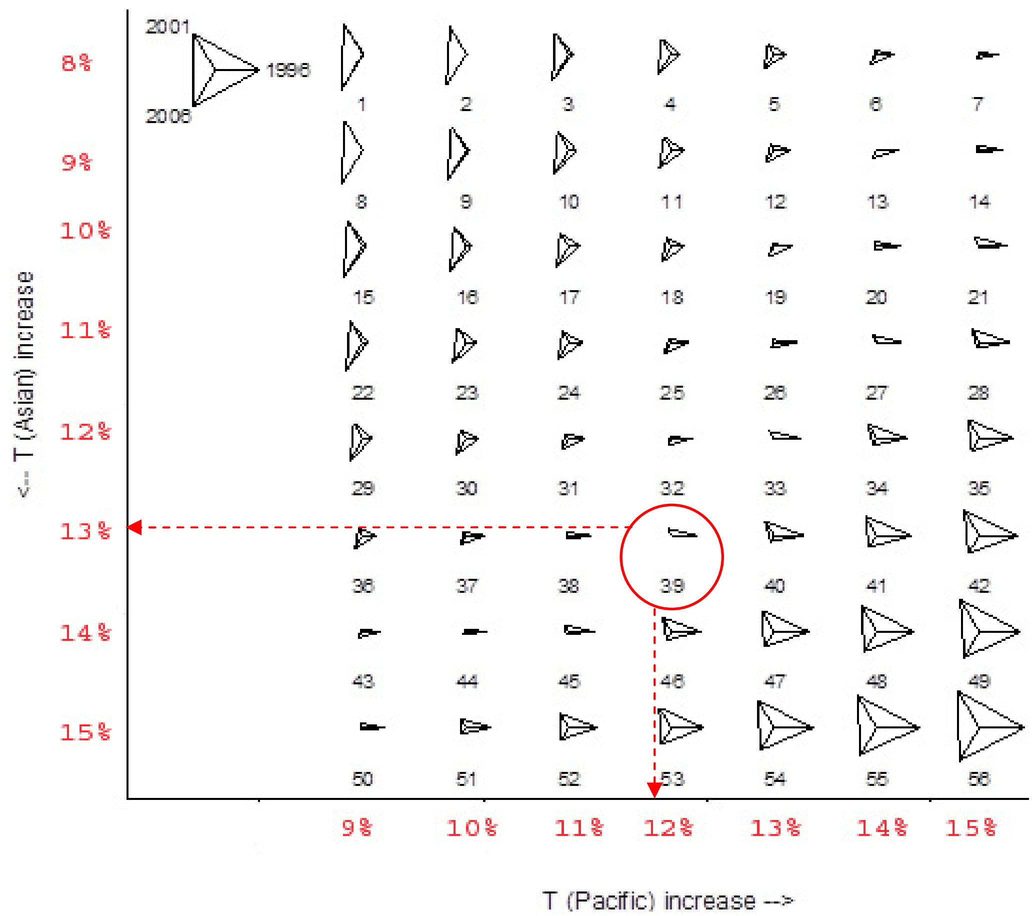


Figure 7.2 Star-plots of Abs. Differences between Simulated and Empirical Values

To ensure the best pattern is chosen, interaction plots can be used. In this way, an overall fit of simulated patterns to empirical samples can be visually verified (e.g. the shape of the curves in the graph compared to a five-year census based sample).

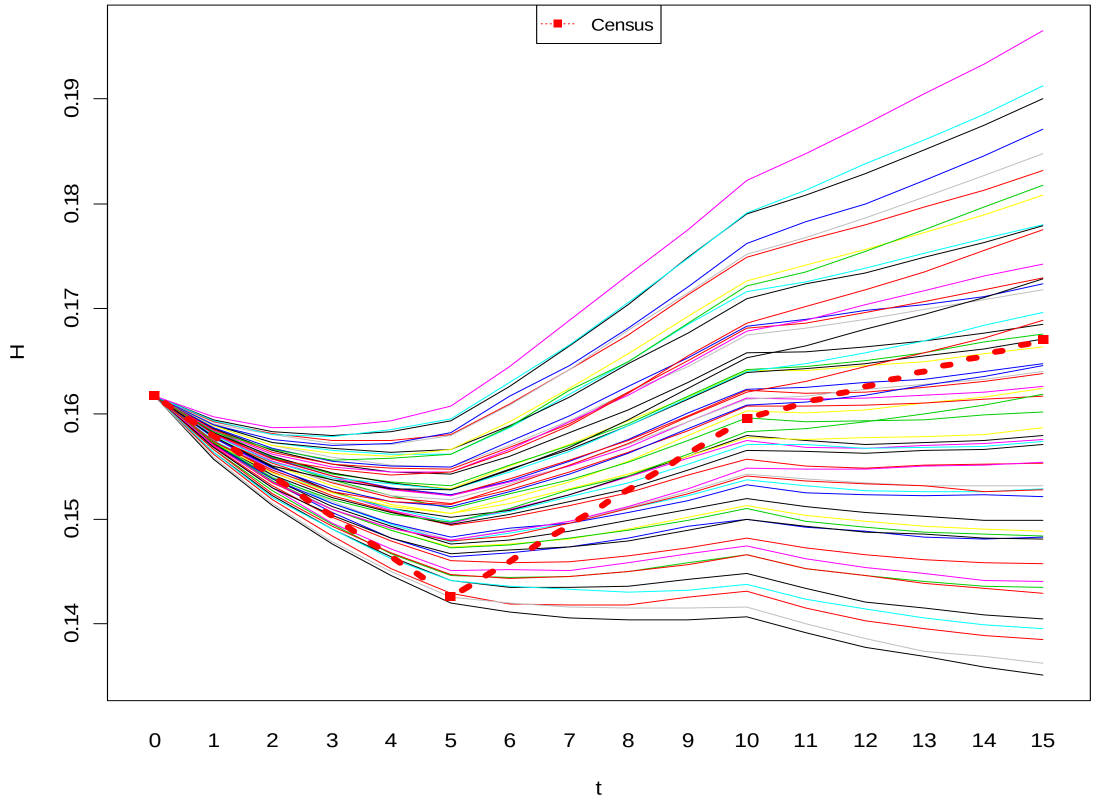


Figure 7.3 Plot of Simulated Outputs and Empirical Benchmark

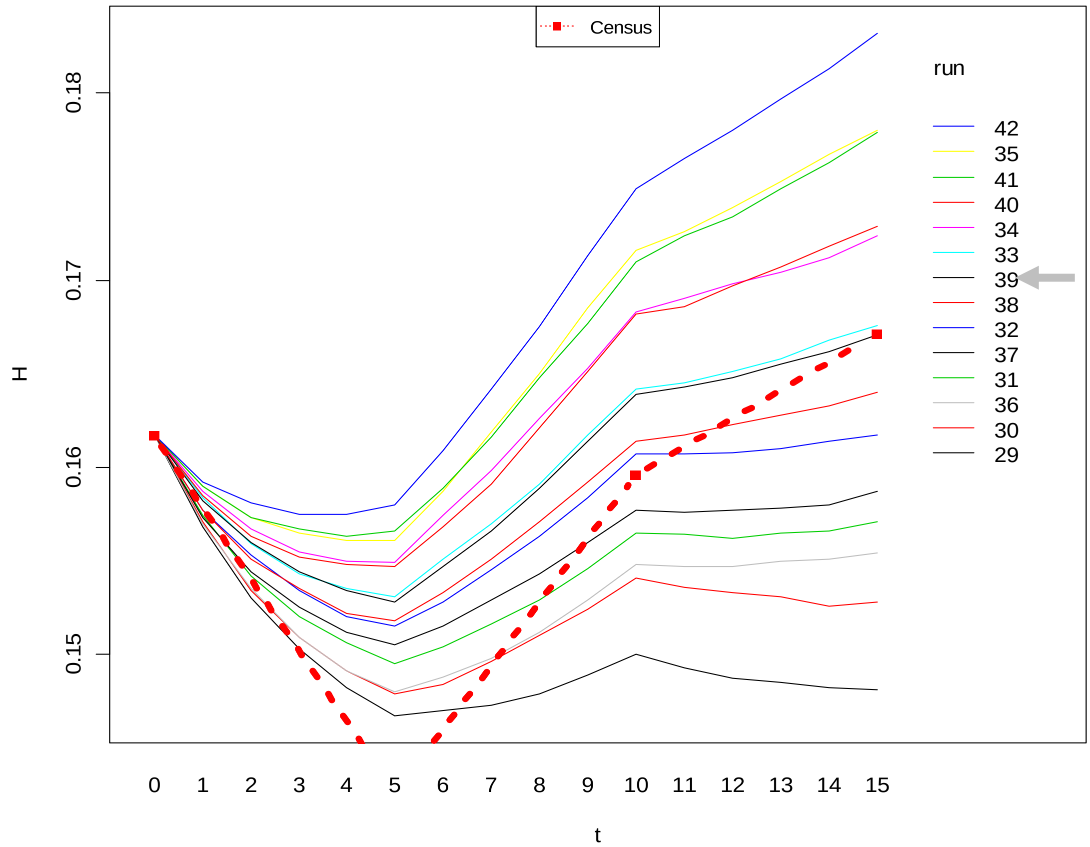


Figure 7.4 Smaller Set of Simulated Outputs and Empirical Benchmark

When the number of simulations is large (as in this example with 56 different runs), graphs with a smaller set of simulations can be created if closer visual inspection and selection is desirable. This can be seen in figure 7.4. In this particular case, the run number 39 was the closest pattern to the empirical benchmark shape and value in 2006. In cases where patterns are still very similar and the selection of the best match is not still obvious enough, other measured outputs and patterns such as snapshots of the city map can be used to help find the best match. In fact, other measures (potentially those at smaller geographical scales such as in TAs) are very useful in reducing the search space and ultimately for a most informed selection.

The POC process is to a large extent visual and in order to accelerate the process graphs and plots can be created automatically using R[17] scripts. Many aspects of this process can also be used for sensitivity analysis. It is possible that the combination of some parameters can potentially result in 'lock-in sensitivity' – a situation in which subsequent changes to some of the parameters' peripheries would not lead to any (significant) change in the behaviour of the model. If necessary, the model can be adjusted, otherwise it would be reasonable to use the smallest values of parameters at the start of the lock-in sensitivity state.

[17] R software environment and programming language for statistical computing and graphics: http://www.r-project.org

7.4 POC of the Model Results

The results of the calibration of the realistic mode of the model, largely based on H-values using the POC method with particular attention to the parameterization of Asians and Pacific people as variables of interests amid certain conditions and assumptions, are presented in this section. The calibrated values are presented in table 7.1.

Table 7.1 Values of Key Variables at the End of POC

Variable/ Parameter		Value
Y_0	Starting year at t=0	1991
G	Number of groups : ($g_{1/E}$, $g_{2/A}$, $g_{3/P}$, $g_{4/M}$)	4
$P(M)_{g\,[1,2,3,4]}$	Turnover probability percentages for all four groups	0.10
T_1	Tolerance/ In-group preference for group 1 (Europeans)	0.38
T_2	Tolerance/ In-group preference for group 2 (Asians)	0.13
T_3	Tolerance/ In-group preference for group 3 (Pacific people)	0.12
T_4	Tolerance/ In-group preference for group 4 (Maoris)	0.08
$P(GL)_1$	Global location move preference probability for g_1 (Euro), L_1	0.70
$P(GL)_2$	Global location move preference probability for g_2 (Asian), L_2	0.60
$P(GL)_3$	Global location move preference probability for g_3 (Pacific), L_3	0.30
$P(GL)_4$	Global location move preference probability for g_4 (Maori), L_4	0.30
MX_F_i	Inflows (immigration); uses empirical values set at each cycle	see Table 7.2
MX_F_o	Outflows (emigration)	see Table 7.2
V_{pc} (t)	Vacancy percentage applied at t ≥ 0 randomly with p=0.80	see Table 7.3
MINv	Minimum number of vacancies in area unit at t=0	50
P-PERSIT$_{1,4}$	Probability of persistence for Europeans & Maoris	0.50
P-PERSIT$_{2,3}$	Probability of persistence for Asians & Pacific people	0.30
λ_{VISIT}	Mean for maximum number of visits during search	12
isGATA	Is a growth agent TA oriented?	TRUE
RS	Random seed	Vary from 1 to 10 (by 1)
t_s	Stop tick time (each tick time is equivalent to 1 year)	30
U	Number of area units (realistic environment)	316
φ_g (u,t); t=0	Total population in each area unit at t=0	Based on 1991 Census data
Φ (t); t=0	Total population of urban area (at t=0)	847,113
Φ (t); t=15	Total population of urban area (at t=15)	~ 1,093,288

The assumptions and conditions include:

- Uniformly applied turnover probability (10%):
    - Only about 10% of the population of each group search for new residential places at each simulation cycle (equivalent to one year). Although there are different reasons for people moving to new locations, the mechanism in this model abstracts out the motivations behind individuals' moves. People simply move for different reasons.

> For example, members of a [Maori] household might want to move out of a white [European] area because they want to be close to others of the same ethnicity, because they were priced out of an expensive white neighbourhood, because they were the victims of abuse and discrimination, or any number of other reasons (Gilbert, 2006, p. 433).

- Empirically informed growth and immigration:
    - Although between 1991 and 1996, the European population had actually shrunk the analysis of census data in the previous chapter showed an overall increase between 1991 and 2006 of the population for all four ethnic groups in the Auckland metropolitan area (see table 6.3). To introduce these population changes exogenously in the simulated urban area, a specific procedure has been written and scheduled in the model so that census-based numbers of household agents for each ethnic group are dynamically calculated in each cycle and removed from or injected into the urban area. In other words, after 15 cycles (tick times equivalent to 15 years), the population of each group will roughly correspond to 2006 Census values. An estimation of these numbers is shown in table 7.2. The growth and immigration is currently treated in the same manner in the model. The 'immigrant' agents are created and usually by default can settle in any part of the urban area. However, for this calibration, these exogenous agents were allowed to limit their search within the TAs for which the inflow numbers had been calculated. There is an empirical underpinning behind this approach as most newcomers to a city have an idea (preference) about which area(s) they would like to settle when they arrive. This might be because of known (or prospective) job location, proximity to relatives or friends, specific school location for children, reputation of district, propinquity to the first temporary accommodation (e.g. hotel) or simply because they did some research prior to arrival.

Table 7.2 Number of Out/ Inflows for Each Group until 2006

	[1991-1995]	[1996-2001]	[2001-2006]	Total in 15 Years
European	- 996	592	3,056	14,745
Asian	8,484	9,197	14,552	161,165
Pacific	1,923	4,489	3,707	50,595
Maori	2,352	426	1,156	19,670

- Empirically-based vacancy rate stochastically applied
  - The TA-based vacancy rates used are realistic according to the SNZ censuses of unoccupied dwellings (as percentages). These are shown in table 7.3. However, to avoid the 'edge effects' observed during experimentations as a result of 'abrupt' changes between census periods, the average of the last two censuses (shown in the last column) is used stochastically as the 'trial number' with an 80% chance of success (accuracy). The model will make sure that after provision of additional space (new housing developments in the city) for new immigrants (growth) according to the population change based on census, the vacancy percentages for each area unit will corresponds to their respective TA values shown in Table 7.3.

Table 7.3 Empirical Vacancy Rates

	[1991-1996]	[1996-2001]	[2001-2006]	Model
North Shore	4.1	5.5	5.2	5.4
Waitakere	4.1	6.1	5.5	5.8
Auckland city	5.9	7.3	8.4	7.8
Manukau	3.6	5.1	4.6	4.9
Papakura	3.0	5.8	5.1	5.4

Fixing initially the tolerances (T) for Europeans and Maoris and taking Asians and Pacific people as variables of interest whilst systematically varying them, several iterations of the POC process (as described in the previous section and illustrated in figure 7.1) was necessary to 'tune' the tolerance preferences for the best H-based matching result. It is important to reiterate that tolerance percentages are sensitive to population proportions. It would be difficult (and at a certain level almost impossible) for a small minority group with a high level of (in-) tolerance to find locations with a matching level of co-ethnic concentration. In most existing models, the population size is fixed, and a suitably constant level of tolerance is used (as shown in Schelling and Fossett tests in chapter 5). However, in the scenarios presented in this chapter, the realistic growth of the population sizes alters the proportional sizes of the groups' populations. For example, the proportional sizes of European and Asian populations (see table 6.3), in particular, are quite different at the beginning (1991), middle and end of the simulation (the end of the simulation for the current calibration is 2006, but subsequent scenarios will project the simulations until 2021). With proportional sizes of about 57%, 18%, 14% and 10% respectively for Europeans, Asians, Pacific people and Maoris in 2006 (see table 6.3), selecting T = 0.38, 0.13, 0.12, and 0.08 would roughly bring about a 66%, 72%, 86% and 80% intolerance impact for each group in that order. This way, the desire to be surrounded by like people is stronger for Pacific people, followed by Maoris, Asians and Europeans. This may not be surprising, as some previous studies also found higher levels of in-group preferences for Pacific people comparing to Asians, (for example see, for example, Grbic et al., 2009; Johnston et al., 2008). The fact that Europeans are the group with the least 'change of residence' may be also interpreted as a high desire (preference) by this group for remaining at their current residences surrounded by long time (alike) neighbours. On the other hand, highly mobile people (such as Maoris) may still choose to search for new residences where the higher concentration of their own group usually live.

In the next sections, we will see what will happen if some of these assumptions and conditions are challenged.

It must be noted that one of the most challenging part of the validation and calibration processes of any model is dealing with the cases in which different configurations (conditions) would lead to similar outcomes (sometimes this issue is discussed under the topics of 'equifinality' and 'multiple realizability'). From the perspective of a single global value (measure) such as H at macro level, there are potentially various configurations of the urban system which can result in a similar pattern (or values). The decision making mechanism used by household agents in their choice of location is still primitive in the current version of the model. But even with a detailed knowledge of the key factors and a sophisticated implementation of how these will influence the relocation choice of the householder agents, it would still be extremely challenging (if not impossible) to regenerate the same movements and compositions in every corner of the urban area as occurs in reality by free and independent decision-making agents. Moreover, the odds of achieving this goal would be even slimmer if more than one perspective is sought in the validation process (for example, if in addition to H, we also want GMI and I.I values to correspond to their empirical counterparts, as sought here).

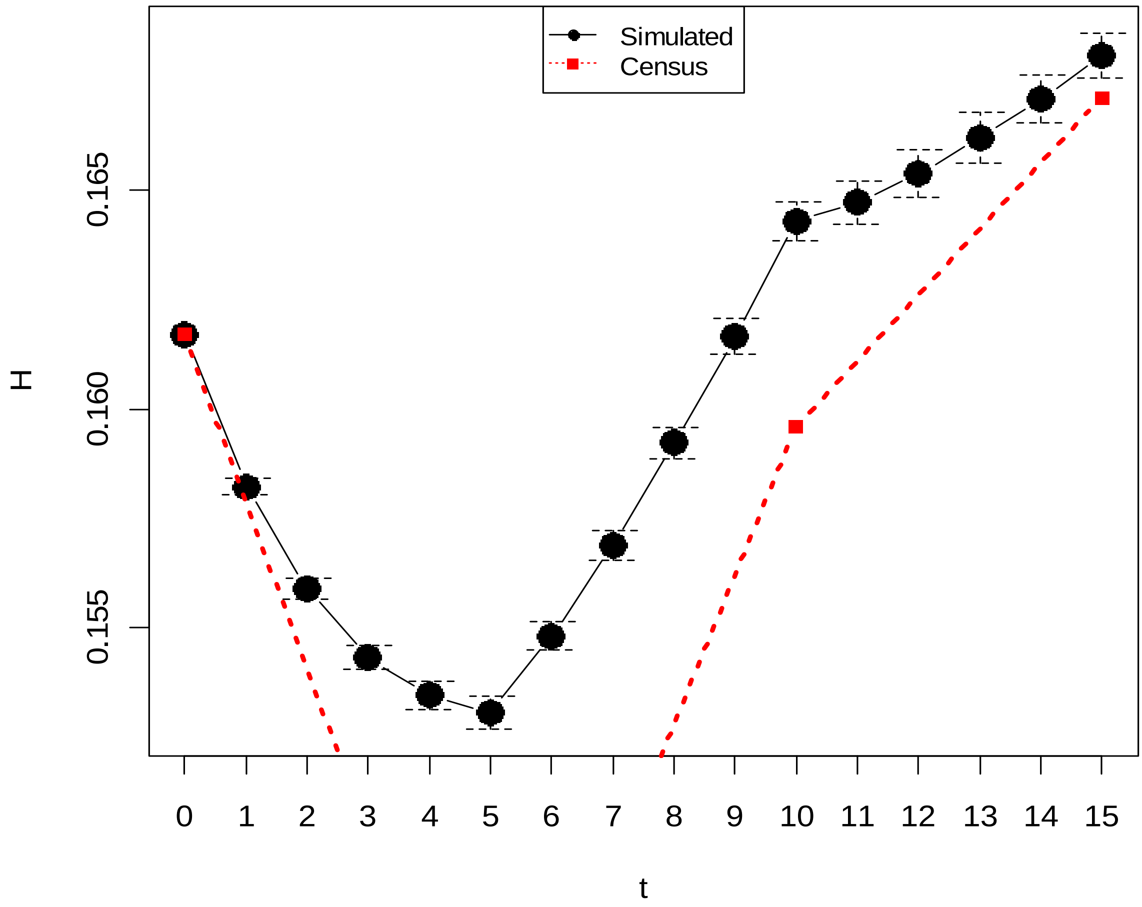


Figure 7.5 Means of Calibrated H vs. Census-based Values (MA level)

Some of these issues (or difficulties) are exposed in several of the subsequent graphs which show the extent of calibration fitness/ precision. Figure 7.5 shows the mean of 10 runs of H-values using the configuration described above along with a census-based empirical benchmark (in red). Although the H (at macro urban area scale) was mostly used as a target in the POC process, attention was given to other indicators as well. For example, although it was possible to obtain a better fit of the H pattern, the results of other indicators (such as D and GMI) including graphical map of urban area would have been less encouraging. Considering our ambitious and multi-perspectives validation aim in this thesis, it would be helpful to view the current calibration (and other subsequent) results presented in this chapter as 'optimal' values.

For example, the H-values at meso-levels were not as good as those at macro-level (see figure 7.6). Nonetheless, these graphs (here and others in general) still carry some significance in qualitative sense. A 'qualitative similarity/ comparison' (in this sense applicable to both 'aggregate' and 'spatial' similarities) is not purely concerned with the 'depth' of absolute 'precision' – but rather with the 'breadth' and degrees of relative 'accuracy' – the approximation at multiple scales and with multiple measurements/ data perspectives.

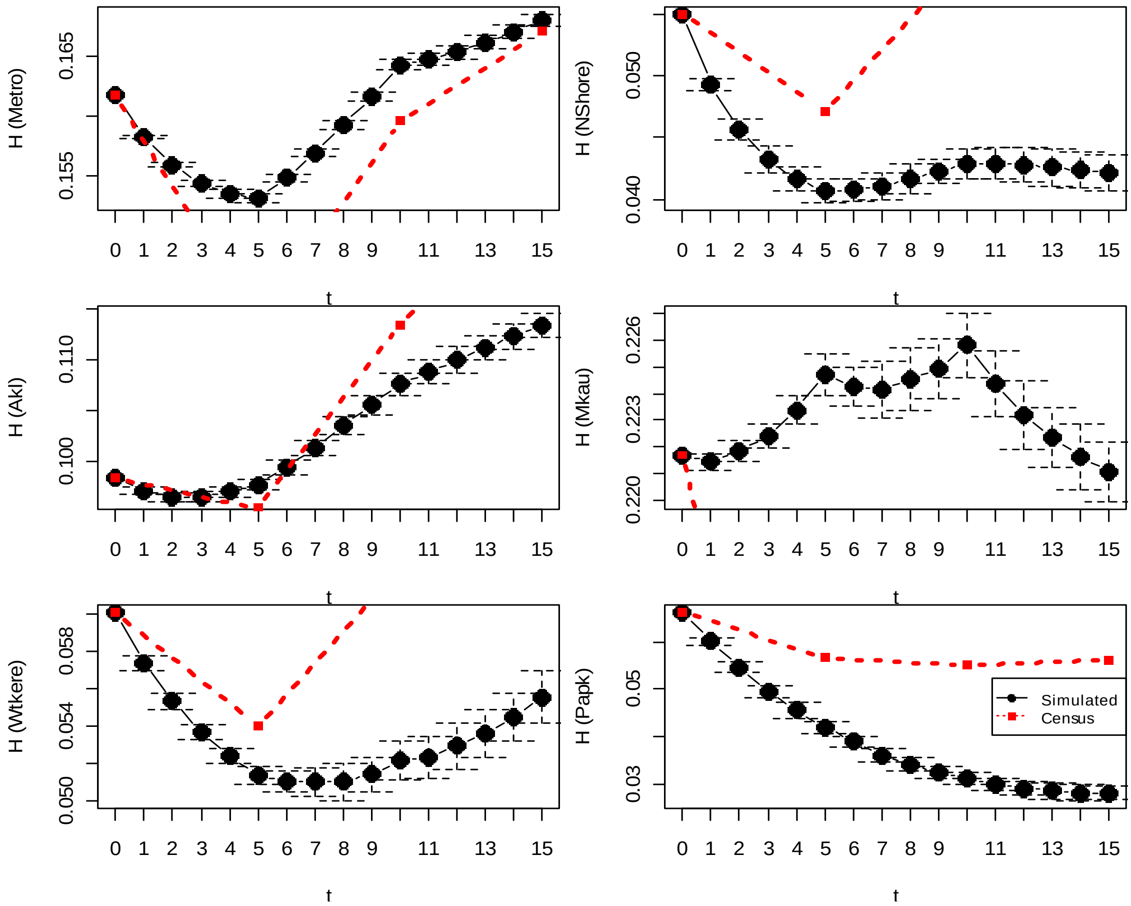


Figure 7.6 Means of H vs. Census-based Benchmarks (MA and 5TAs)

On the other hand, the population growth and sizes in the territorial authorities were relatively similar to the real situation on the ground as seen in figure 7.7 (notwithstanding that the simulation generated lower than real values in *Manukau* and higher in *Auckland* and *Papakura* TAs). However, the population sizes are only one important factor among others. Even at the meso-scale, these are still aggregate values and they do not reveal the actual population sizes of individual area units.

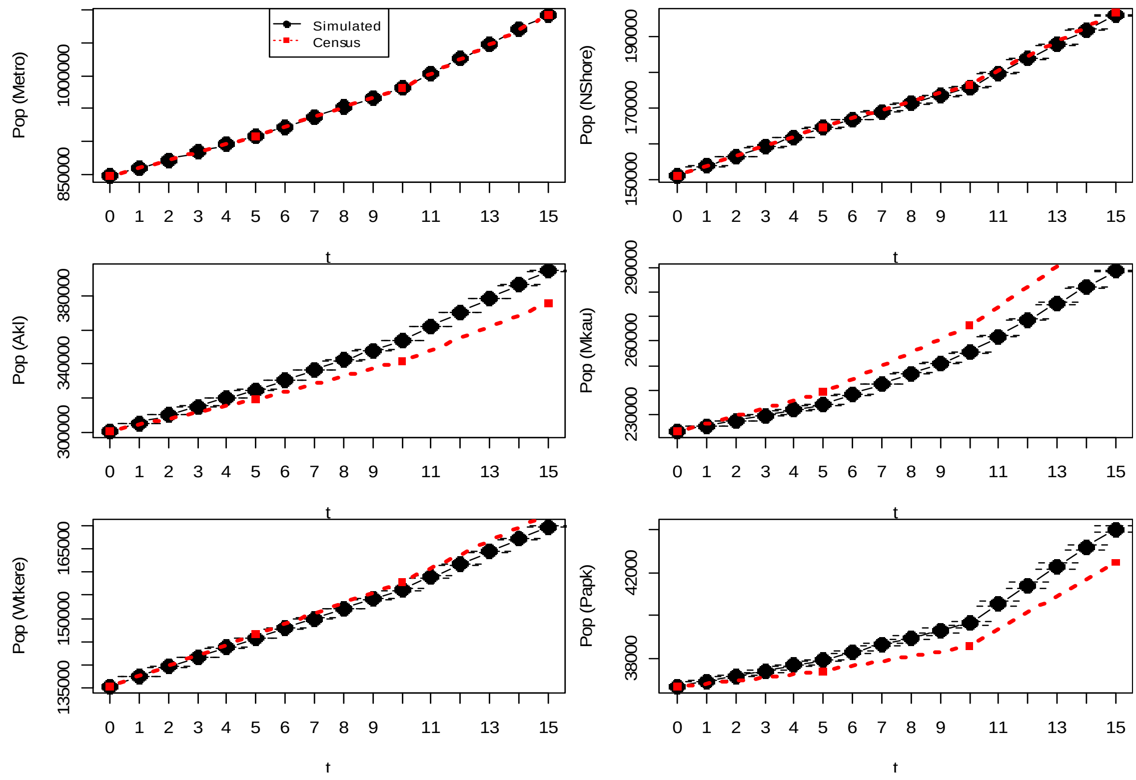


Figure 7.7 Means of Population Sizes vs. Census-based Values (MA and 5TAs)

Moreover, the proportions of these groups in each of the 316 area units are different from the real situation, although the combinations of these micro arrangements have actually produced a fairly favourable realization outcome at the macro level.

On the other hand, the values of GMI and I.I (both isolation and exposures) at metropolitan area (MA) level are more encouraging as can be seen in figures figure 7.8, 7.9, and 7.10 although many of them, like H (at MA level) are far from being perfect matches.

It must be reiterated that the 'error bars' on the graphs give only plus or minus one standard deviation of the mean of 10 different runs values and as such they do not show the maximum and minimum ranges. In such way, for example, the highest GMI among simulated runs for Pacific people in 2006 (figure 7.8, bottom left) was 0.7426 which is really close to the census value of 0.7428, but in general 'close hit' values are not well captured in this graph type. This perhaps is because they are 'outliers'.

Although ideally it would be interesting to have the simulated values matched (as closely as possible) to their empirical counterparts at five years census intervals (i.e. t = 5, 10 and 15), in the final analysis, the importance of the last value at the end of simulation (i.e. t = 15) is paramount. As such, it was more important, for example, to obtain GMI for Asians (figure 7.8, top right) at the end of simulation close enough to the census value and it was less important how closely this 'trajectory' has been traversed in previous census periods (1991, 1996 or 2001).

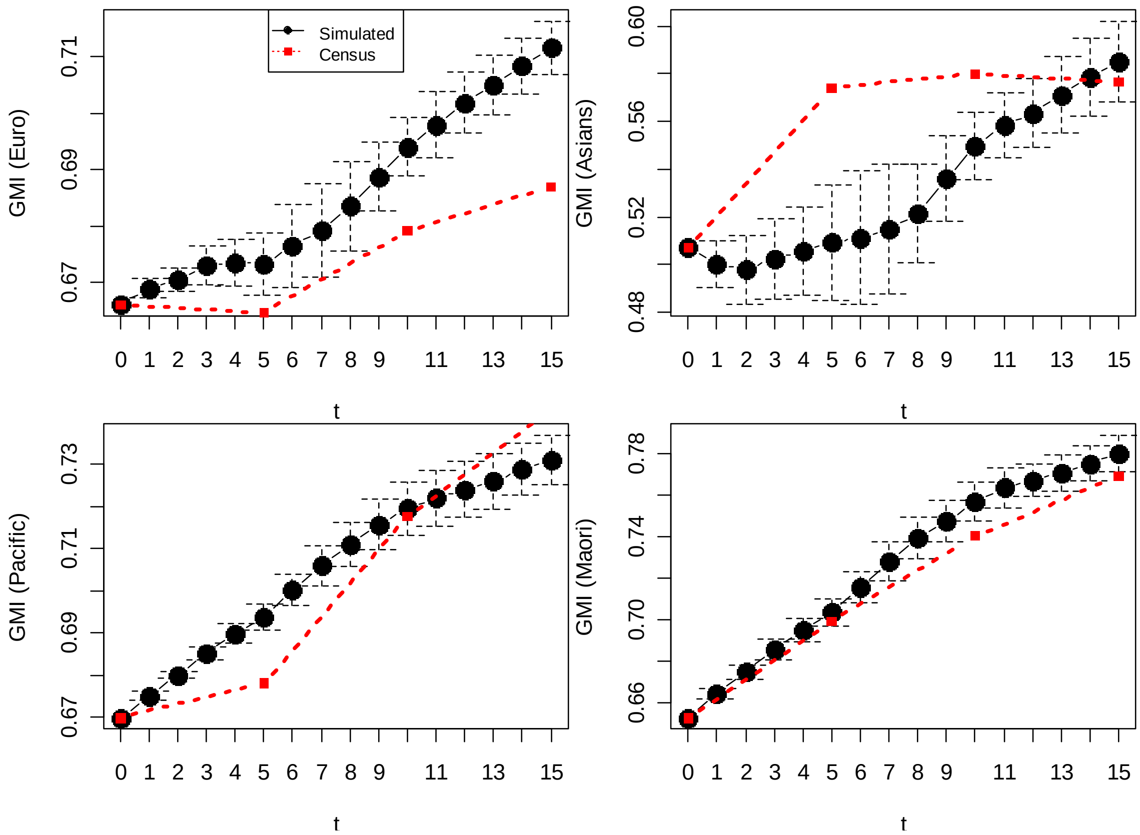


Figure 7.8 Means of GMI vs. Census-based Values by Ethnicity (MA level)

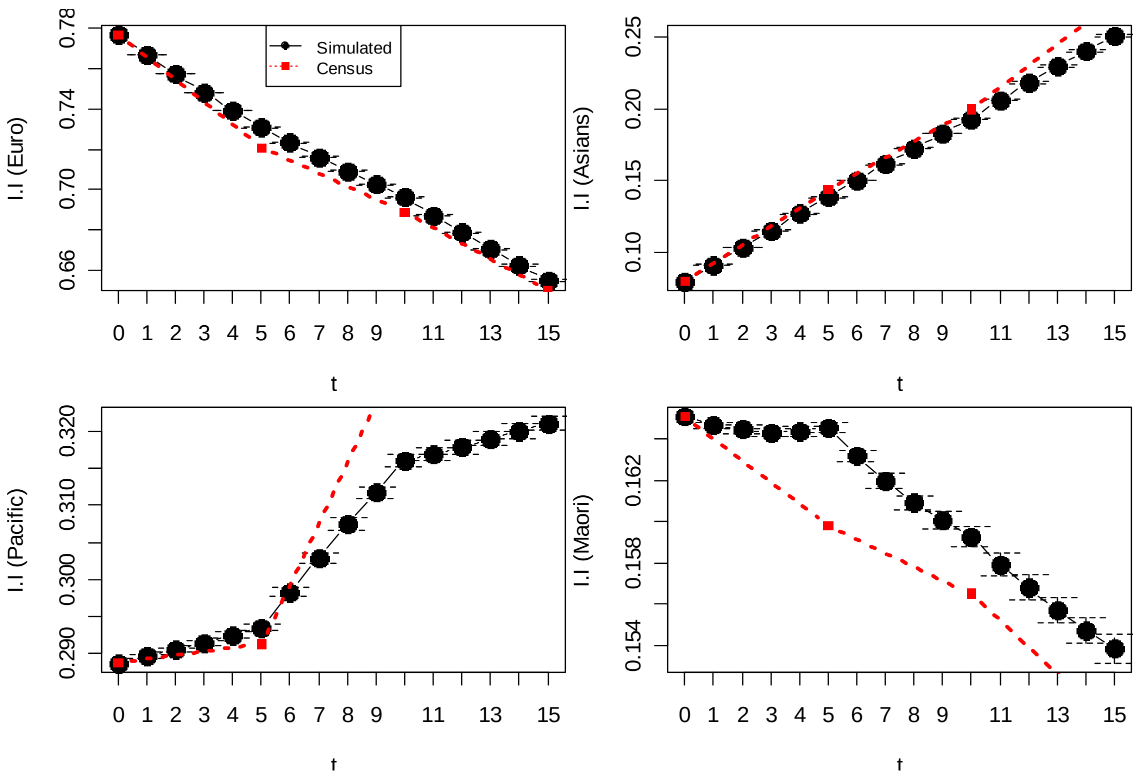


Figure 7.9 Means of I.I. (Isolation) vs. Census-based Values by Ethnicity (MA level)

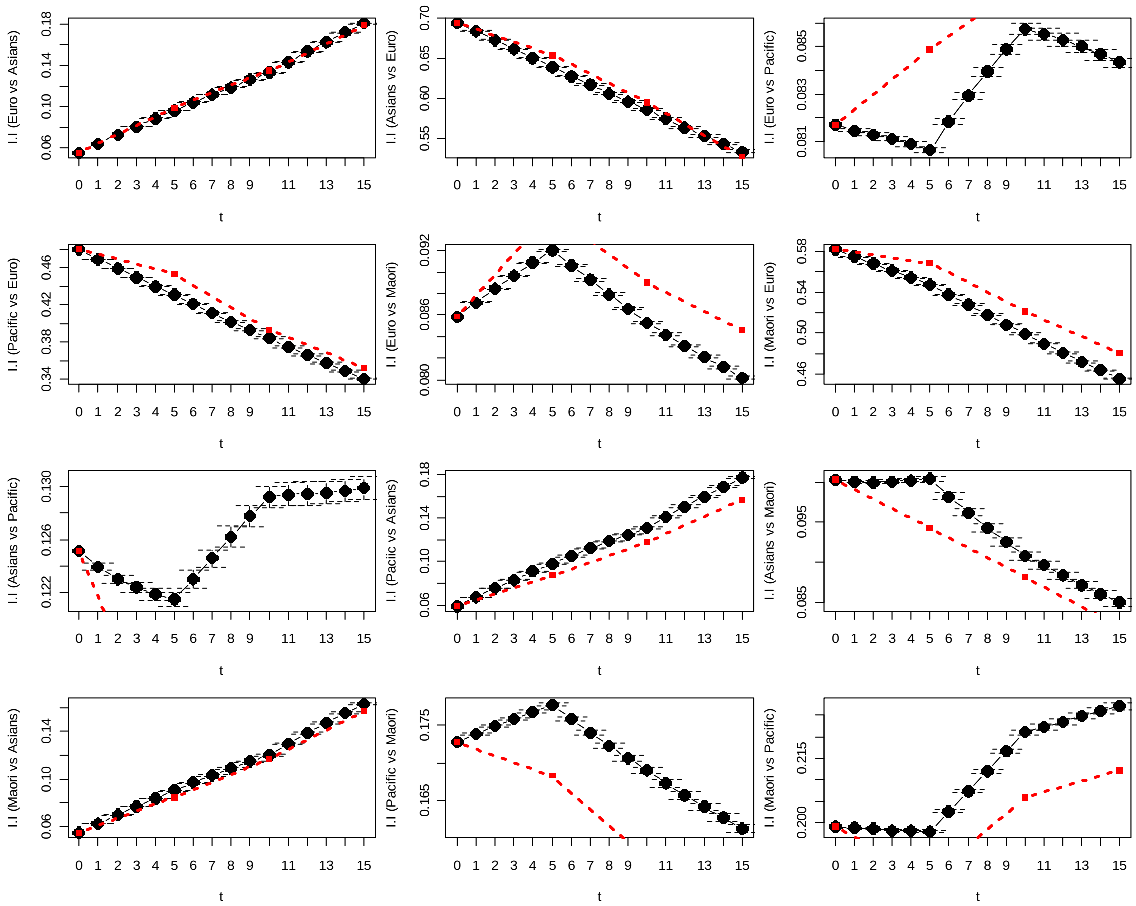


Figure 7.10 Means of I.I (Exposure) vs. Census-based Values (MA level)

The average global isolation index for Maoris (figure 7.9, bottom right) in the simulated version was somewhat higher than the realistic measure value (at t = 15, by about 0.01). For Pacific people the average global isolation index (figure 7.9, bottom left) was lower (at t = 15, by about 0.01, although the highest simulated value was almost the same as the census-based value). On the other hand, the chances in the simulated version for Europeans and Asians (figure 7.9, top) to meet people from their own ethnic group were fairly close to the realistic situation.

The exposure values in the POC version have also been relatively in compliance with the real world situation (figure 7.10). But looking at the individual graphs provides perhaps the biggest disappointment for the Asians vs. Pacific people (first one in left in the third row), end up higher in the simulation. This means that on the whole (at urban area scale) the probability for an Asian to meet a Pacific Islander is higher in the simulated world (i.e. the presence of Pacific people in the same area units where Asians typically live is greater than in reality). Even so, a randomly selected Asian had about 12% (0.115 precisely) chance in 2006 of meeting a Pacific Islander at urban area level in reality (according to census), but had about 13% (0.129 precisely) on average in the simulated world. By all means, this is still a good reconstruction of the real world.

Except for the case of Pacific people vs. Maoris case, even D-values (at MA level) at least point in the same direction as their census-based counterparts (see figure 7.11). Nonetheless, the intermixing and the global unevenness between these two groups has been more favourably simulated comparing to the real situation on the ground (i.e. a more balanced intermixing between these two groups has taken place in the simulated version). The average simulated D-value in 2006 is less than 0.4 lower than the census-base value.

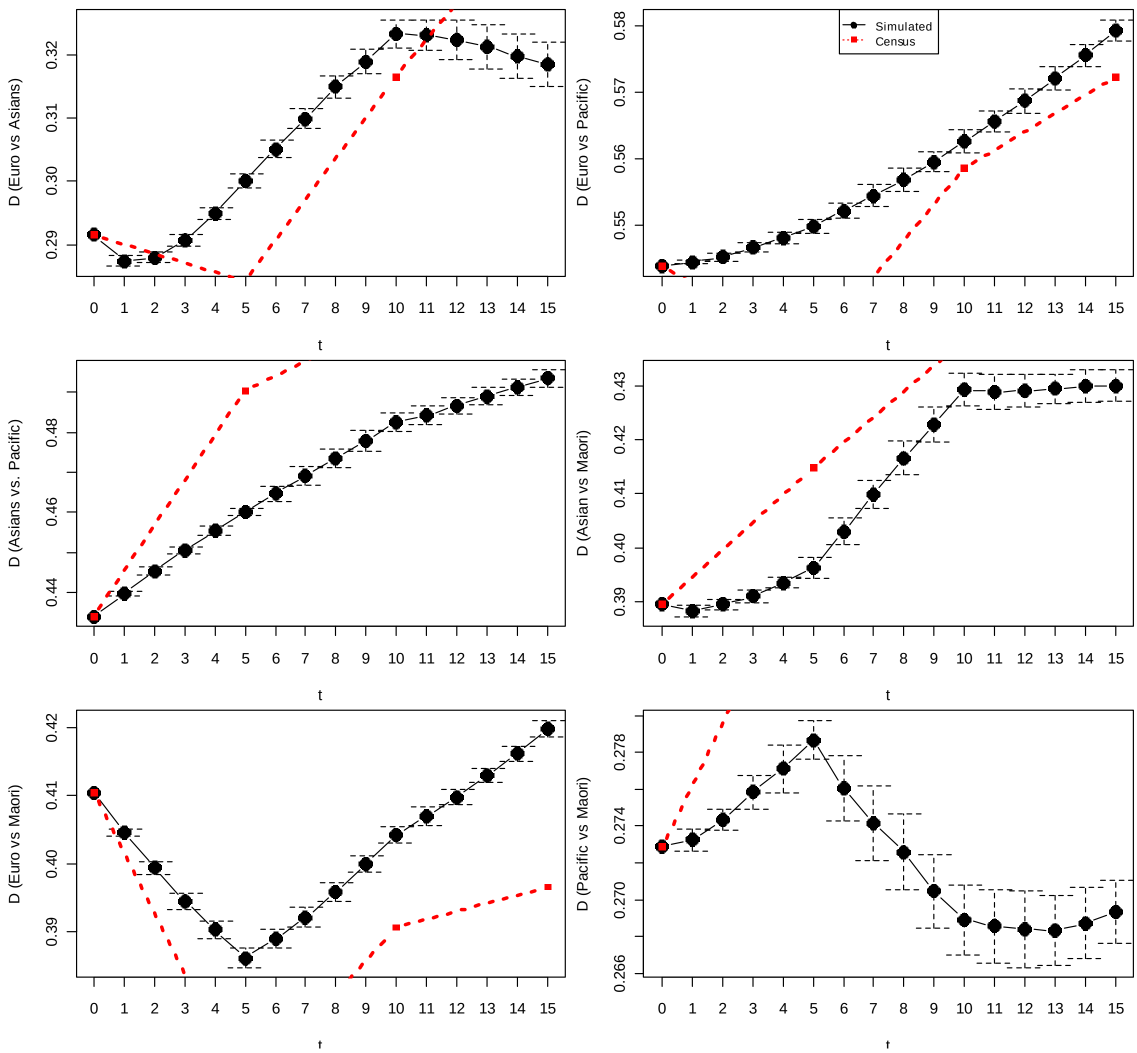


Figure 7.11 Means of D vs. Census-based Values (MA level)

There are also significant spatial similarities between the snapshots of the census-based and simulated metropolitan area (see figure 7.12). However, with 316 area units, many of them with tiny superficies, it would be rather difficult to meaningfully discern and appreciate the differences and similarities (producing many zoomed versions is not a practical solution for this thesis). The representative colours of the large area units often outshine those of smaller AUs, so that changes in smaller area units can easily go unnoticed.

At this scale, it would be even more difficult to discern something meaningful, had we been additionally interested in examining the proportions of populations too (i.e. pie-charts). Nevertheless, such a usage of the visual inspection of the urban area characteristics could have been advantageously performed using the *HAAMoS* model, if the focus of the investigation and comparison was on specific (targeted) areas. This is not the case for the urban-area focus scenarios presented in this chapter.

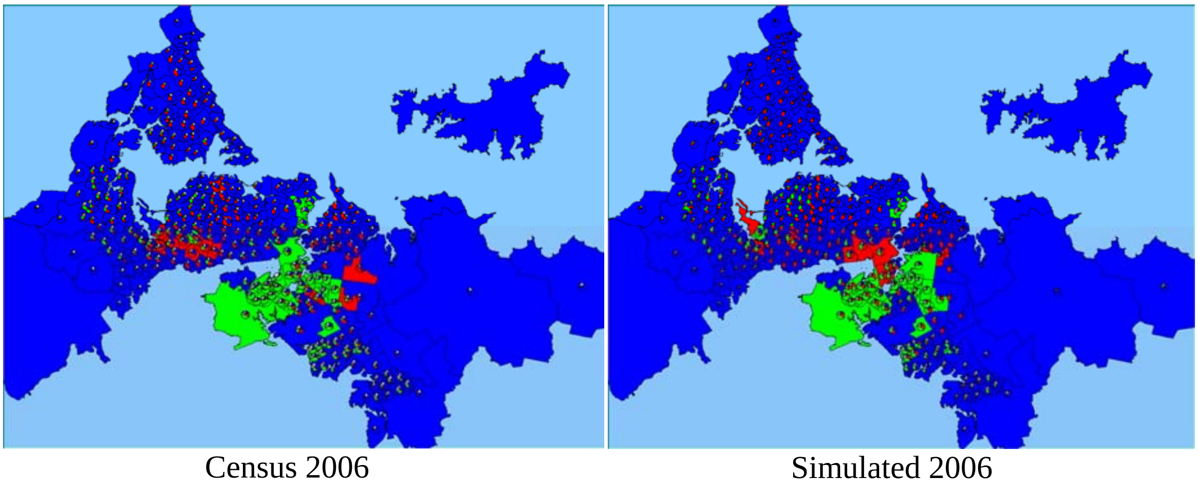


Figure 7.12 MA Snapshots, Census-based vs. Simulated

Moreover, at the urban area level, although the concentrations of groups in specific areas (i.e. in districts/ TAs) are important, it is perhaps less vital (at least at this stage of study) to find out in which specific area units they occurred.

As Brown et al. (2005) found, it seems that simulations generate some "invariant" area units or regions where the concentrations or plurality of a specific group virtually occur all the time (i.e. it is "almost certain") whereas in other area units (or regions) it is "path dependant" (i.e. depends on a particular series of conditions and events). For example, as can be seen in one of the simulated snapshots presented in figure 7.12 (right side), the simulations were successful in replicating the real situations where many specific area units are represented by Pacific people (green) in relative majority (plurality) positions. On the other hand, the simulations did not do so well in generating the (exact location of) Asian plurality cluster (in red) appearing on the census-based map of the urban area (figure 7.12, left side, around southwest region). This is notwithstanding that the total number of area units with an Asian plurality in the simulated version is roughly the same as in reality. After all, "demanding that modelers get the locations right may be asking too much" (Brown et al., 2005, p. 154) and despite the importance of spatial similarity, "model validation should not be [uniquely] based on mapping correspondent spatial patterns" (Wu, 1999, p. 202).

For this reason the count of area units where each ethnic group forms a plurality (PI) is a useful alternative (or complement) to the visual analysis of simulation outcomes. This is shown in figure 7.13. As can be seen, the numbers in the simulated runs were relatively close to the census-based numbers. In 2006, the census-based counts of plurality-based AUs for Europeans, Asians, Pacific people and Maoris were 261, 16, 39, and 0 respectively (see also figure 6.2). The simulations generated minimum and maximum values of 265|269 for Europeans, 12|14 for Asians, 35|37 for Pacific people and 0|0 for Maoris. Although these are not exactly the same as in reality, they are nevertheless an encouraging regeneration of the real word phenomenon, even though these majorities might have been formed (by simulation) in different AUs than those on the real ground.

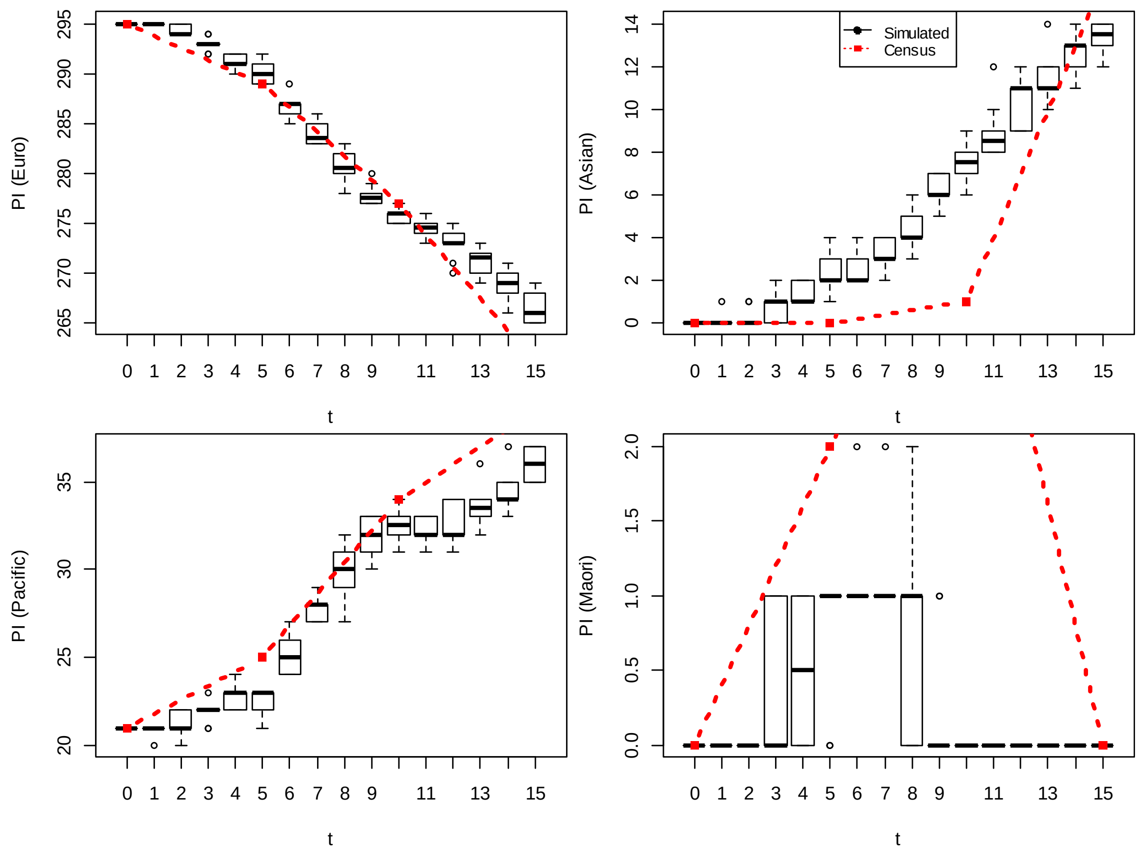


Figure 7.13 PI vs. Census-based Values by Ethnicity (MA level)

The D, GMI, I.I and PI values at meso levels (TAs) are not presented in this section but have been included in the appendix H. Like the MA level results presented in this section, they show the extent of similarities and discrepancies between simulated and census-based values.

Although they are largely due to the way the decision-making individual household agents make their moves across the simulated world at each cycle (year) there are several reasons for the discrepancies between the simulated and real world. The movement of agents are conditioned by multiple factors, among them their level of tolerance, the percentage at which they are endowed to move globally (anywhere in the urban area) and whether there are vacant houses available at the time of their visit (decision-making). Although it would be possible to engage in the explanation of the cause-and-effect relationship in a complex system investigation such as this study, this thesis does not focus on explaining (at least not in every single presented case) what has exactly caused a divergence between the real situation and the simulated world. Chapter 5 (section 5.3) has briefly engaged in explaining the emerging results. These often point to the assumptions made in the logic of the residential relocation process and the way it is implemented in the model (e.g. algorithm).

Nonetheless, specific scenarios in the following sections will show and discuss the effects of changing parameters (micro conditions, behaviours, etc.) on the aggregate (macro) or semi-aggregate (meso) outcomes.

7.5 Policy-oriented Scenarios and Projections

Models are a relatively inexpensive way to acquire relevant information (particularly about the future trends) that can be used to inform policy-related questions. Population growth, relocation, and segregation are important issues that affect our daily lives by influencing the availability of adequate services in health and education and shaping the fabric and structure of society.

The scenarios presented in the following sections will demonstrate the potential of the *HAAMoS* model (even in its current state) as a decision-informing tool by investigating some of the relevant *what-if* questions.

7.5.1 Effects of Population Growth and Immigration Policies

States/ governments exercise certain levels of control over population growth particularly by control of immigration policy. This control is usually enacted, for example, through changes to immigration rules and visa quotas, or by providing free birth control or sterilisation (if the aim is to reduce the population) or by incentivising the natural growth by 'baby-bonus' programs and the like.

Regardless of whether (and to what extent) these controls are effective and without entering into the debate about the moral and potential success or failure of such social engineering, we will here simply simulate and compare the results and effects of three different population growth projection scenarios.

This will be done by using Statistics New Zealand (SNZ) projections (which used 2006 values as the base until 2021) for four major ethnic groups in different territorial authorities. The assumption is that different policies (by the government/ lawmakers) during the period to 2021 can change/ drive current trends.

The projections are given as average annual changes based on three alternative series designated 'high', 'medium' and 'low'. The medium case will later be used as the base for other scenario enactments and comparisons to simulate the population growth/ immigration rate of each ethnic population after 2006 (the year of the last census). The SNZ specifies that projections have been produced for each ethnicity and each TA using different birth rate (fertility), mortality, migration, and other inter-ethnic mobility assumptions. The terms 'low', 'medium', and 'high' do not correspond to probabilities, but merely indicate the combination of these assumptions. The medium projection series uses medium fertility, medium mortality, medium migration, and medium inter-ethnic mobility for each ethnic group in each district area (TA). The low projection series assumes low fertility, high mortality, low migration, and high inter-ethnic mobility for each ethnic group in each TA. The high projection series assumes high fertility, low mortality, high migration, and low inter-ethnic mobility. Moreover, the low and high series are independent of other national or sub-national projections as they represent plausible alternative scenarios for each ethnic group and each area only, rather than for any collective ethnic or geographic level.

Table 7.4 Scenario Table: Comparing Growth/ Immigration Effects

Sc ID	Scenario Name	Operative Time	Pop. growth/ immigration rate
Sc1	POC + SNZ Proj **Med**	t ≥ 16 (2007)	see tables in appendix G3
Sc2	POC + SNZ Proj **Low**	t ≥ 16 (2007)	see tables in appendix G3
Sc3	POC + SNZ Proj **High**	t ≥ 16 (2007)	see tables in appendix G3

The details of the simulated scenarios are presented in table 7.4 (the exact number of the growth rates for each ethnic group and TA are included in appendix G3). It should be noted that these growth rates will only be used during the simulation after 2006 (t = 15). In other words, the lapse of time between t = 15 and 30 are the projections until 2021 for which no empirical data is currently available for validation purposes.

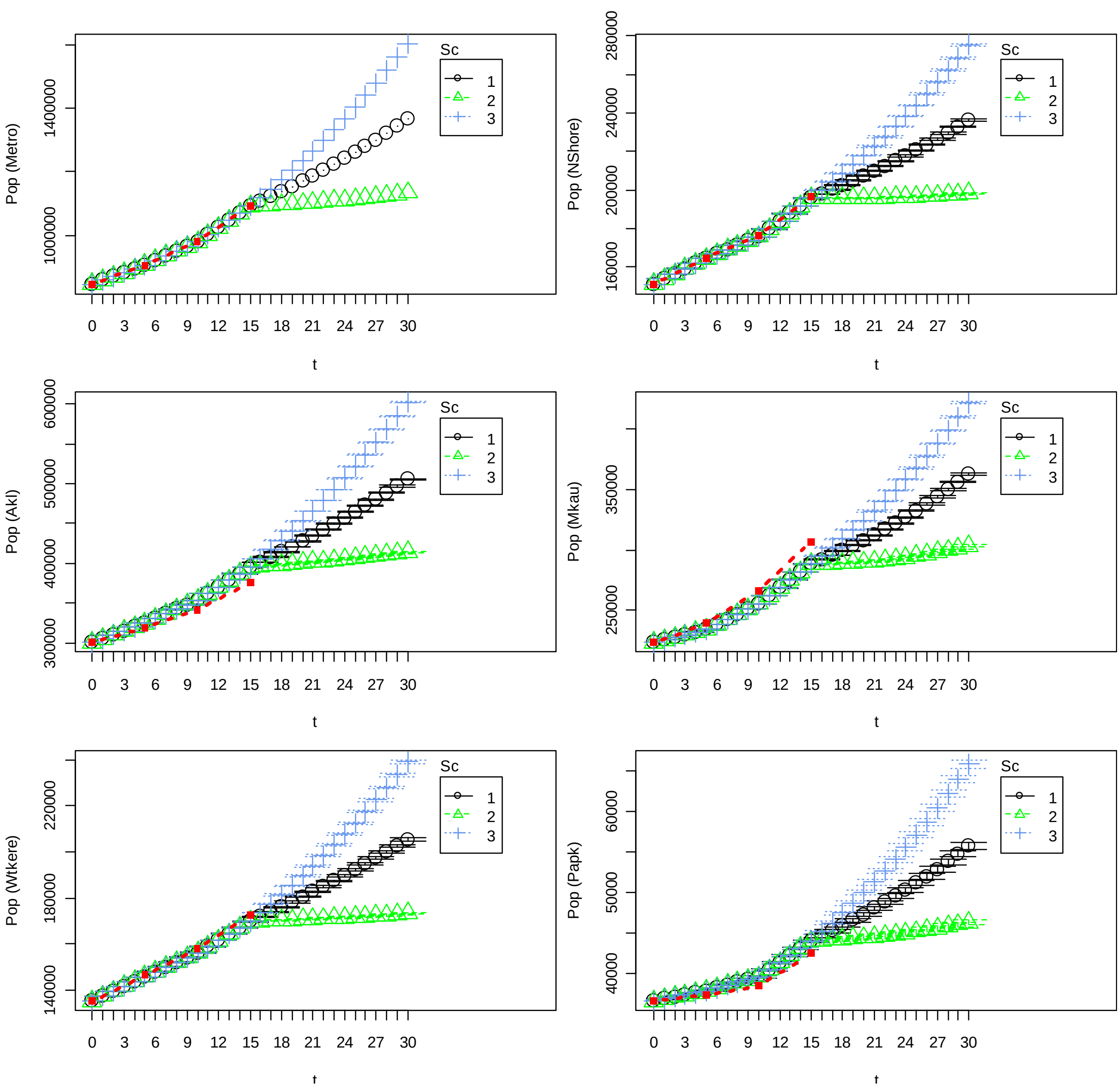


Figure 7.14 Population Sizes Under Three Different Scenarios (MA and 5TAs)

As expected the population sizes correspond to the names and rates of these projections; that is, higher population was generated when 'high' rates were used and lower population when 'low' rates were used and so on (see figure 7.14).

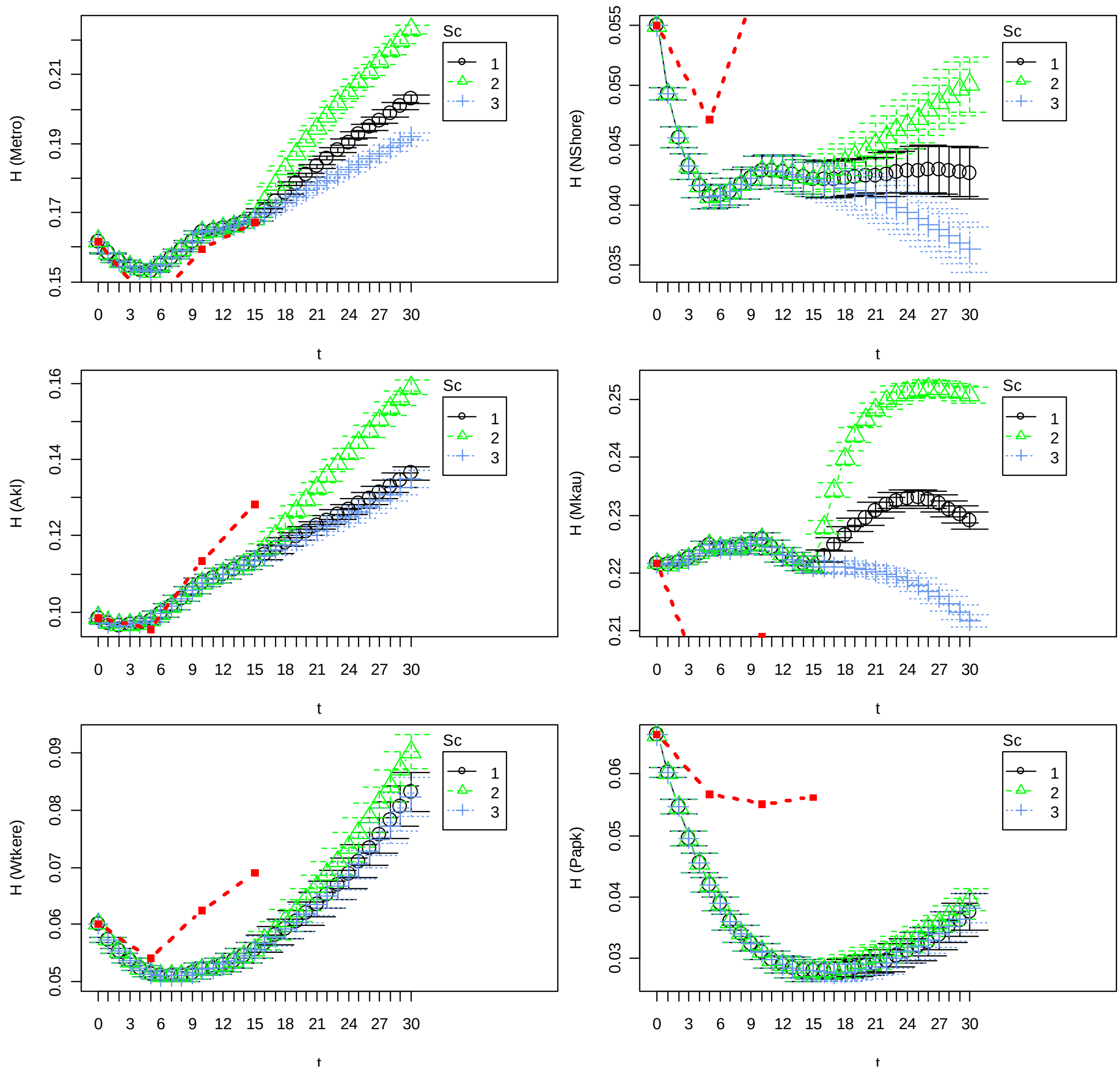


Figure 7.15 H-values Under Three Different Scenarios (MA and 5TAs)

On the other hand, the level of segregation measured by H (and D, not shown here) may appear counterintuitive (see figure 7.15). The 'low' rates (Sc2) have generated higher level of segregation, followed by 'med' (Sc1) and 'high' (Sc3). This is because these measures are sensitive to the distribution patterns (e.g. unevenness) and not the population sizes. Under the SNZ 'low' rate case, several TAs (and their constituent area units) will have their percentage of long time majority group (Europeans) shrunk and a new composition order would naturally take place. For example, under the 'low' case, the population of Europeans will contract in all the TAs (and in *Manukau* under the 'medium' case too) whilst the Pacific and Maori populations will also lose some weight in *Auckland*. These happen simultaneously in each constituent area unit along with different annual growth rates for each group which are applied proportionally to their existing sizes.

The results indicate that although at the metropolitan scale the H measure will be higher than its current value in 2006, the increase would be more dramatic under the SNZ 'low' case scenario (Sc2). Moreover, in *Manukau* and *North Shore*, the H level has actually dropped even under Sc3. In *Manukau*, for instance (see figure 7.16), the loss of the European population (which had not already reached the census-level in the simulated version) occurs in all the cases as a result of combined population change rates and the dynamic effect of 'white flight' (i.e. departure of Europeans from this district which has taken place within the conditions set for these simulations). This is not to suggest that there is direct correlation between global H and the size of the population *per se*. Rather, it is the proportional changes and imbalances/ unevenness throughout the urban area and its smaller geographical constituents that are the most likely culprits.

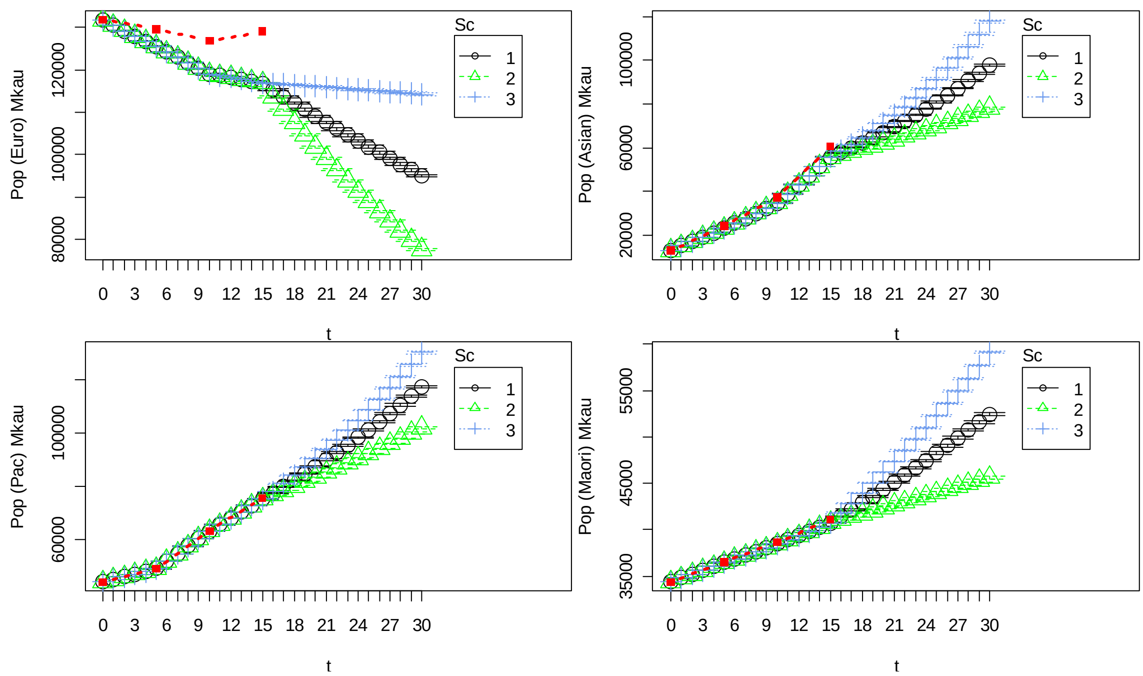


Figure 7.16 Population Sizes by Ethnicity Under Three Different Scenarios (MKAU)

Higher growth or immigration does not automatically exacerbate the level of segregation. The disproportionate level of birth rate and immigration by ethnicity on the other hand can eventually change the level of segregation, even if future adults continue to stay in the same locations and the new arrivals settle in the already known ethnic 'ghettos' (neighbourhoods).

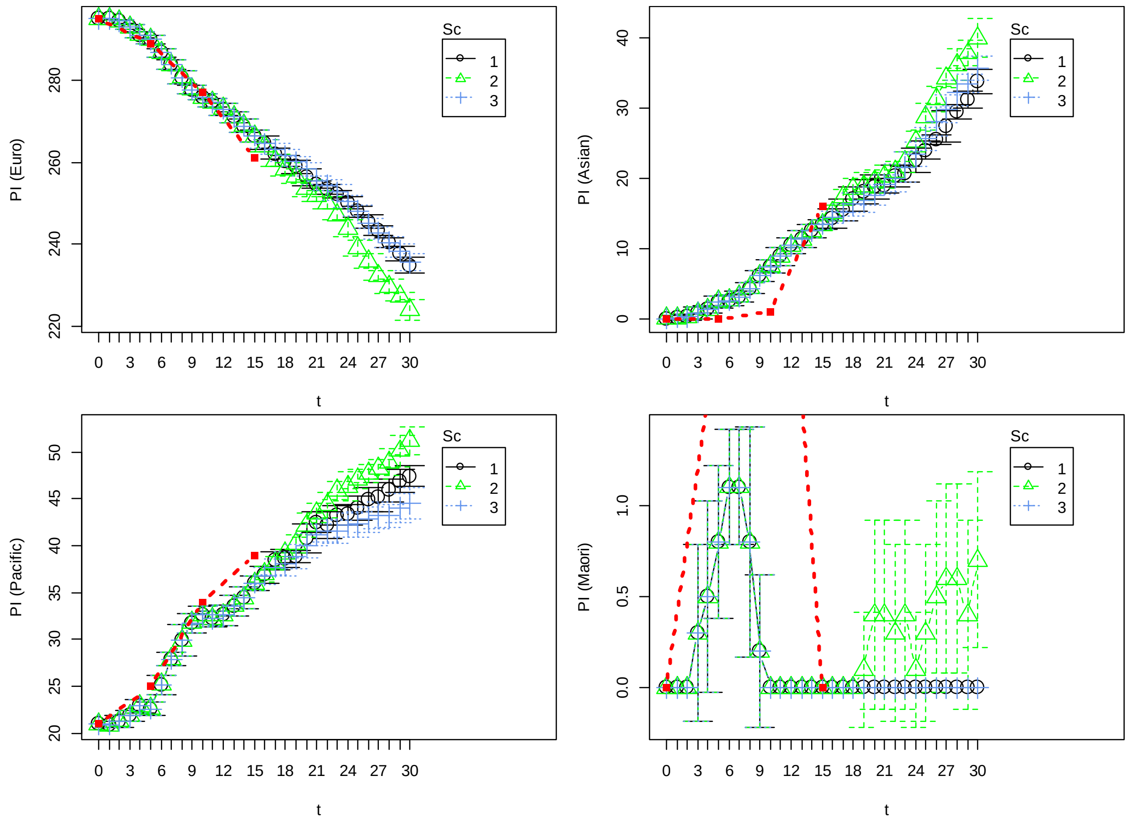
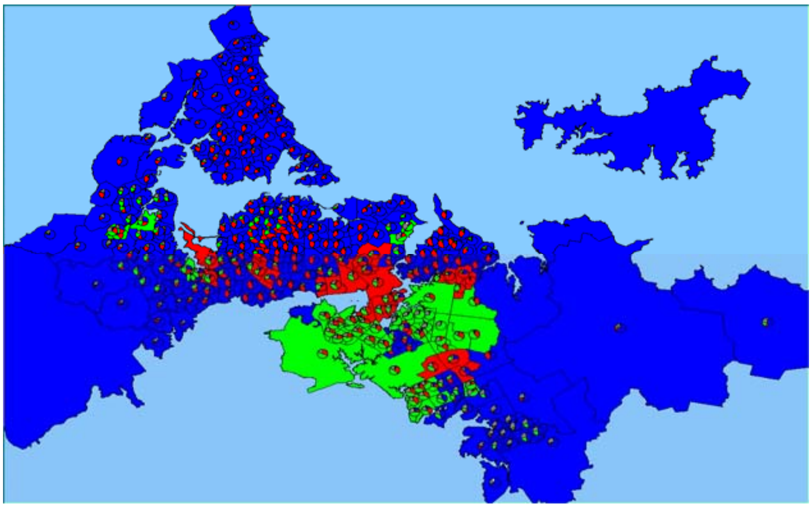
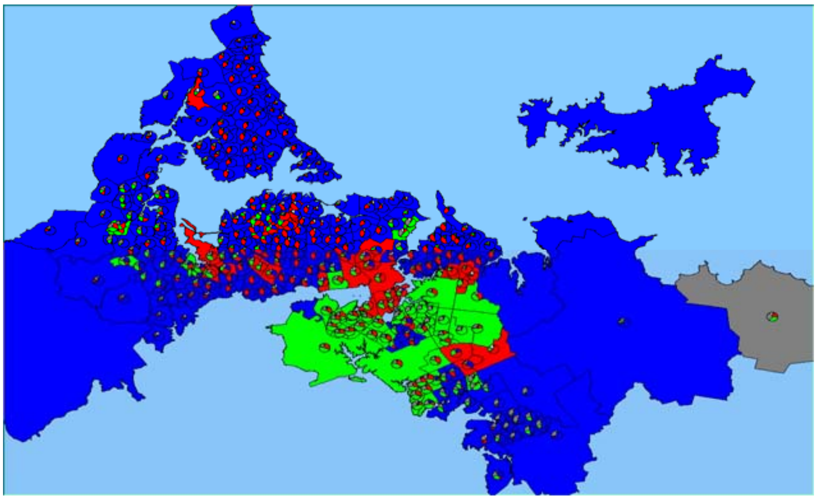


Figure 7.17 PI-values Under Three Different Scenarios (MA and 5TAs)

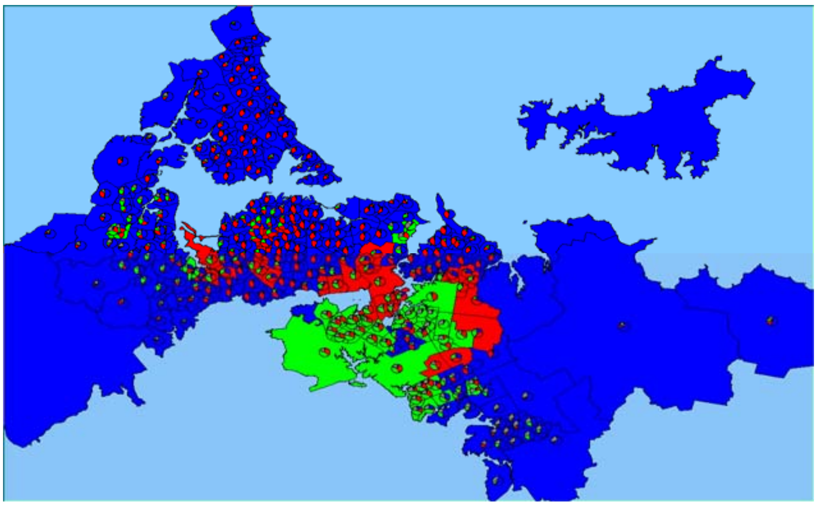
As can be seen in figure 7.17 and 7.18 the three different levels of growth rate projections simulated here have not produced remarkably different numbers of ethnic-occupied pluralities. Nevertheless, the seemingly inevitable decline of the total number of area units where Europeans form the majority group (measured by PI) has somewhat worsened under Sc2 (down to about 220 from 261 in 2006). On the contrary, Asians and Pacific people will have their PI-values boosted by as high as 46 (up from 16 in 2006 for Asians) and 53 (up from 39 in 2006 for Pacific people), whilst Maoris can also become a plurality group in up to 2 AUs (up from nil in 2006). These changes, however, are less noticeable in the visual snapshots of the urban area in figure 7.18.



Sc1



Sc2



Sc3

Figure 7.18 MA 2021 Year Snapshots, Sc1 vs. Sc2 vs. Sc3

Some of the changes in the number of plurality held AUs occur in districts (TAs) that have been stable in the last 15 years as was seen in chapter 6 (see table 6.4). In *Waitakere*, for example, under all these three simulated scenarios (albeit with different numbers) there will, for the first time, be several area units where Pacific people would be in plurality in 2021 (from nil for the past 15 years).

Arguably, in addition to planning of population growth rates and immigration numbers, the government and policy makers can play a role in stemming the deterioration of the spatial ethnic distribution and its landscape balance. They can do so, for example, by creation of job centres/ industrial zones, housing development, health and education services and other amenities in the areas and neighbourhoods which are identified to attract (specific) segments of the population. For this to work in the long term, such schemes should continually be monitored and engineered. More likely though, the model can simply inform policy makers about inevitable changes that will occur in the future, so that adequate planning can take place and a budget for providing and maintaining essential services to the population could be predetermined.

The rest of this chapter continues to use the SNZ 'medium' rate as a base for conducting other investigations using different assumptions and *what-if* questions.

7.5.2 The Effects of Incentivising Population Mobility

There are various reasons behind the dissimilar mobility patterns of people across different ethnic groups. The impetuses are certainly numerous, but employment opportunity (where jobs are located) is the most likely one of them. With the expectation of constituents from their elected lawmakers in government to create/ boost jobs, the effects of population mobility on the urban system or its segregation would become relevant matters.

During the calibration, we assumed an equal turnover rate (stochastically determined as 10% of the AUs' population) among all four major ethnic groups. However, survey studies, such as the Survey of Dynamics for Migration in New

Zealand published by Statistics New Zealand (2007b), suggest that the Maori population, for example, has become much more mobile, whereas Europeans have become less likely to leave their current residence.

In the scenarios presented in this section, we will show the effects of different (rather more realistic) turnover percentages (assuming the incentives to create higher mobility of population for selected ethnic groups had been made available) by comparing the outcomes with the base case (behaviour as usual/ BaU).

Table 7.5 shows old (base) and new turnover probability percentages for each ethnic group for two simulated scenarios. Following the insights from the SNZ mobility survey, we assume that Asians and Maoris have higher mobility than Europeans, whilst keeping the original rate constant for Pacific people.

Table 7.5 Scenario Table: Comparing Different Mobility Effects

Scenario Name	**Sc1 (Med, BaU)**	**Sc2 (Med, new turnovers)**
Operative time	t ≥ 1 (1991)	t ≥ 1 (1991)
Europeans	10%	**8%**
Asians	10%	**12%**
Pacific people	10%	10%
Maoris	10%	**12%**

The simulated H-values indicate slightly higher levels of segregation in the long run when the turnover levels are increased for two major minority groups and decreased for the majority group (Sc2) (see figure 7.19). The rise in segregation is more evident in GMI and D-values where these two increasingly more mobile minority groups are involved (figure 7.20 and 7.21). At the same time, the GMI level for lesser mobile Europeans shows signs of retreat.

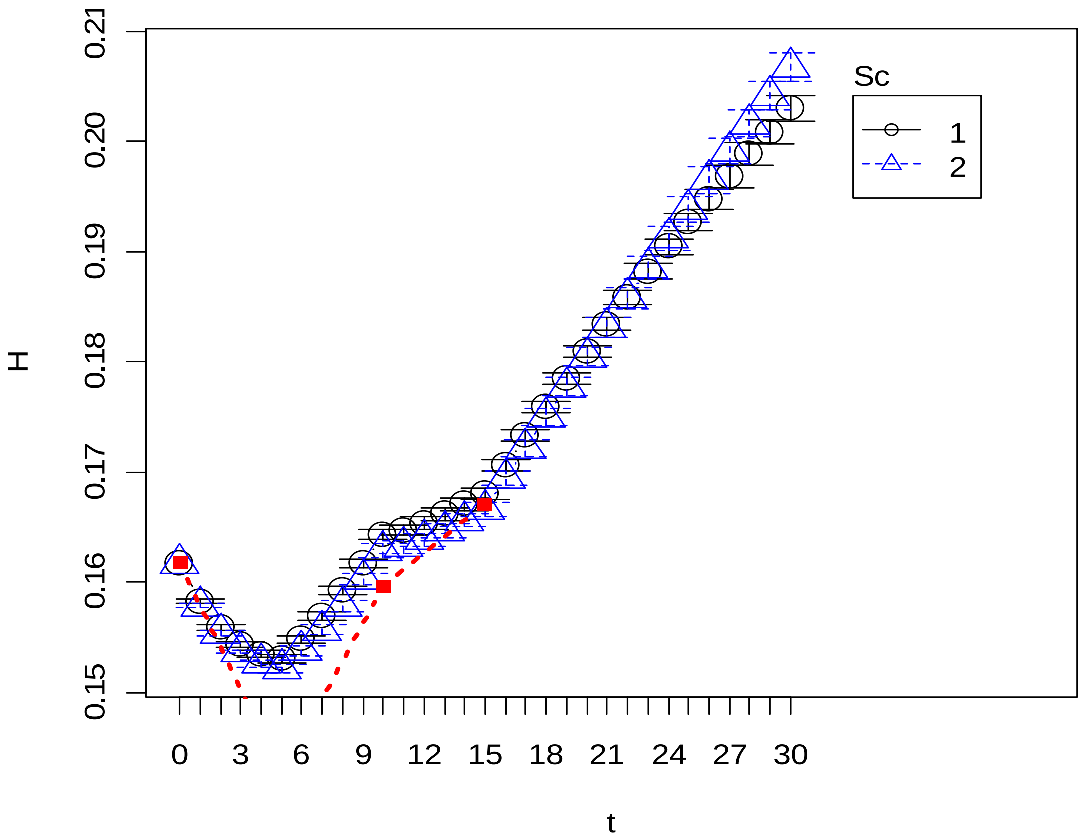


Figure 7.19 H-values Under Two Different Scenarios (MA)

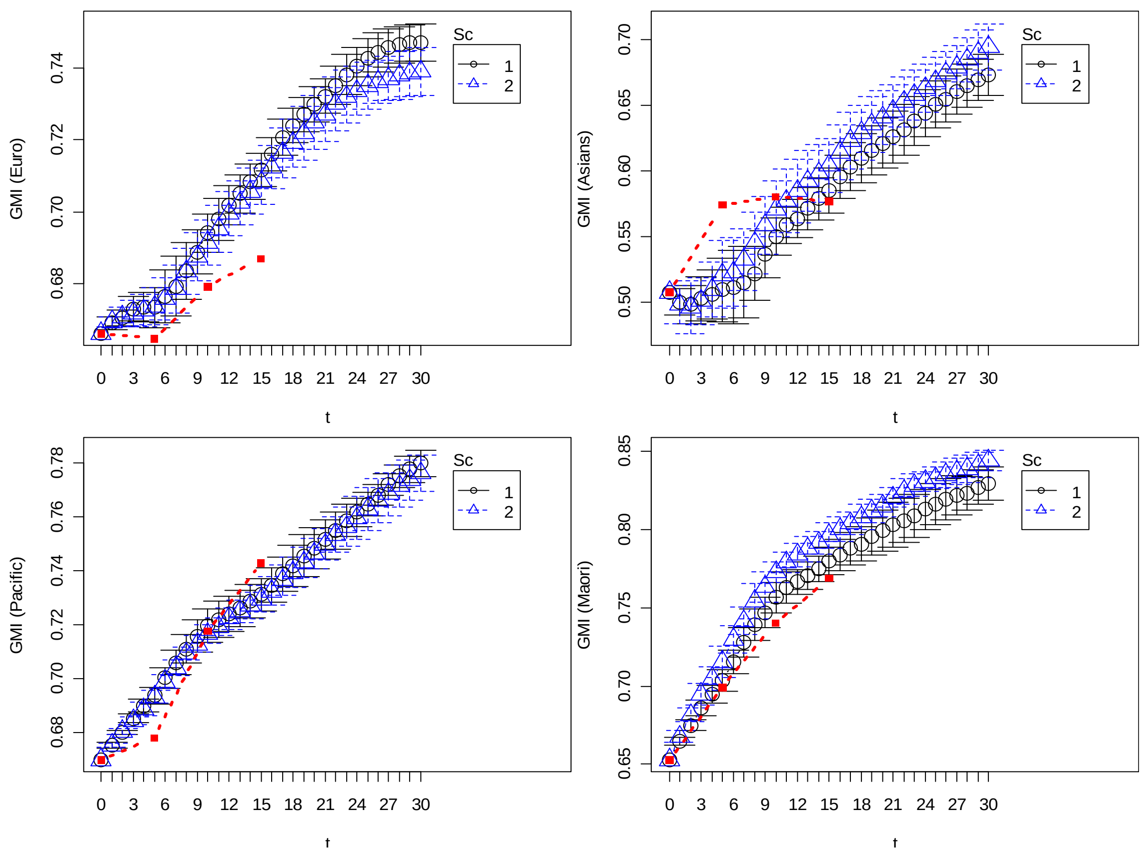


Figure 7.20 Selected GMI-values by Ethnicity Under Two Different Scenarios (MA)

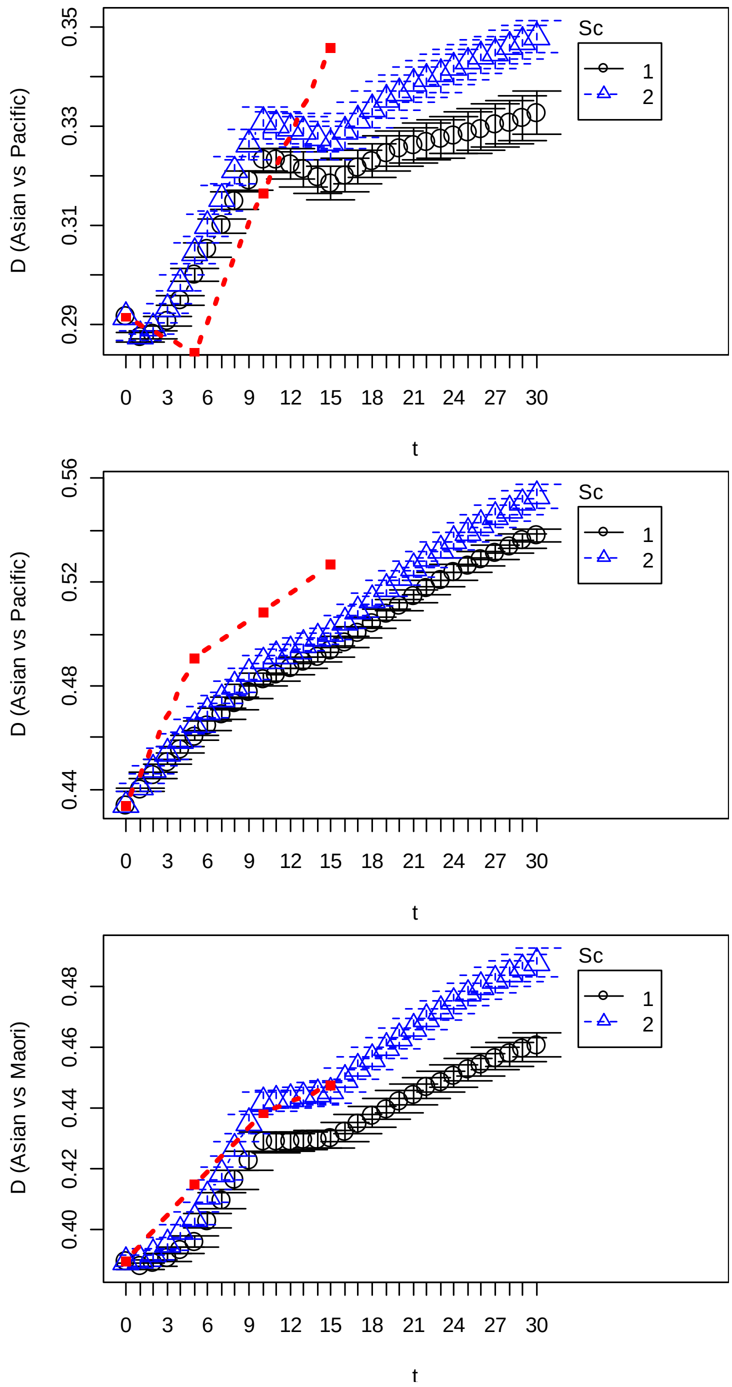


Figure 7.21 Selected D-values Under Two Different Scenarios (MA)

To some extent, it seems that keeping the same behaviour whilst being more mobile will accelerate the evolutionary transformation that would anyway take place without an additional dose of mobility. This suggests that mobility could be used as a sort of 'catalyser' with double-edged effects on segregation dynamics. That is, when the behaviours and trends are segregationist or 'integrationist', mobility can speed up the trending process. This does not mean there might not be other effects that could transpire in an urban system when a certain segment of the population suddenly becomes more mobile. (Further investigation would be necessary to properly evaluate this aspect.)

7.5.3 Influencing Housing Development and Vacancy Effects

It is sensible to think that vacancy rate – that is, the share of housing units that are unoccupied – influences segregation measures. Vacancy rates vary considerably from city to city and from one area to another within the same city. For example, based on U.S. Census Bureau data on rental vacancy rates for 75 largest metropolitan statistical areas[18] in the fourth quarter of 2008, the vacancy rate for Richmond (VA) was 23.7%, while it was only 3% for Boston-Cambridge (MA). The usual assumption by economists and housing policy analysts that supply and demand for rental housing are in balance (equilibrium) when the vacancy rate is about 5% (that is, higher is considered as oversupply and less as undersupply or housing shortage) for all places seems to be based on "rules of thumbs" – and therefore subject to uncertainty – as equilibrium vacancy varies among places and over time (Belsky, 1992).

The model presented in this thesis has this additional advantage over the most typical models (for investigating residential segregation where vacancy percentage is uniformly set to a fixed and often arbitrary value), as it employs realistic rates (optionally in stochastic manner with some level of control over the randomness variability as it was done here and explained in section 7.4 and table 7.3) that vary among different districts (TAs).

Governments and municipal/ local authorities have considerable powers to influence the level of housing vacancy, for example, by financing/ promoting (social) housing development projects among others.

In this section, we examine the effects of a housing shortage in certain localities (areas) on the segregation dynamics of the urban area. To do this, we set the vacancy

[18] http://www.census.gov/hhes/www/housing/hvs/prevqtrs.html

rate of the large TA of *Auckland* to only 2% (instead of the 7.8% previously set). The simulated scenario will be compared with the original case (realistic vacancy for *Auckland*), as summarized in table 7.6.

Table 7.6 Scenario Table: Comparing Different Vacancy Effects

Sc ID	Scenario Name	Operative Time	Vacancy rate
Sc1	Med (original vacancies)	t ≥ 1 (1991)	7.8 %
Sc2	Med (tighter vac. rate for AKL)	t ≥ 1 (1991)	2.0 %

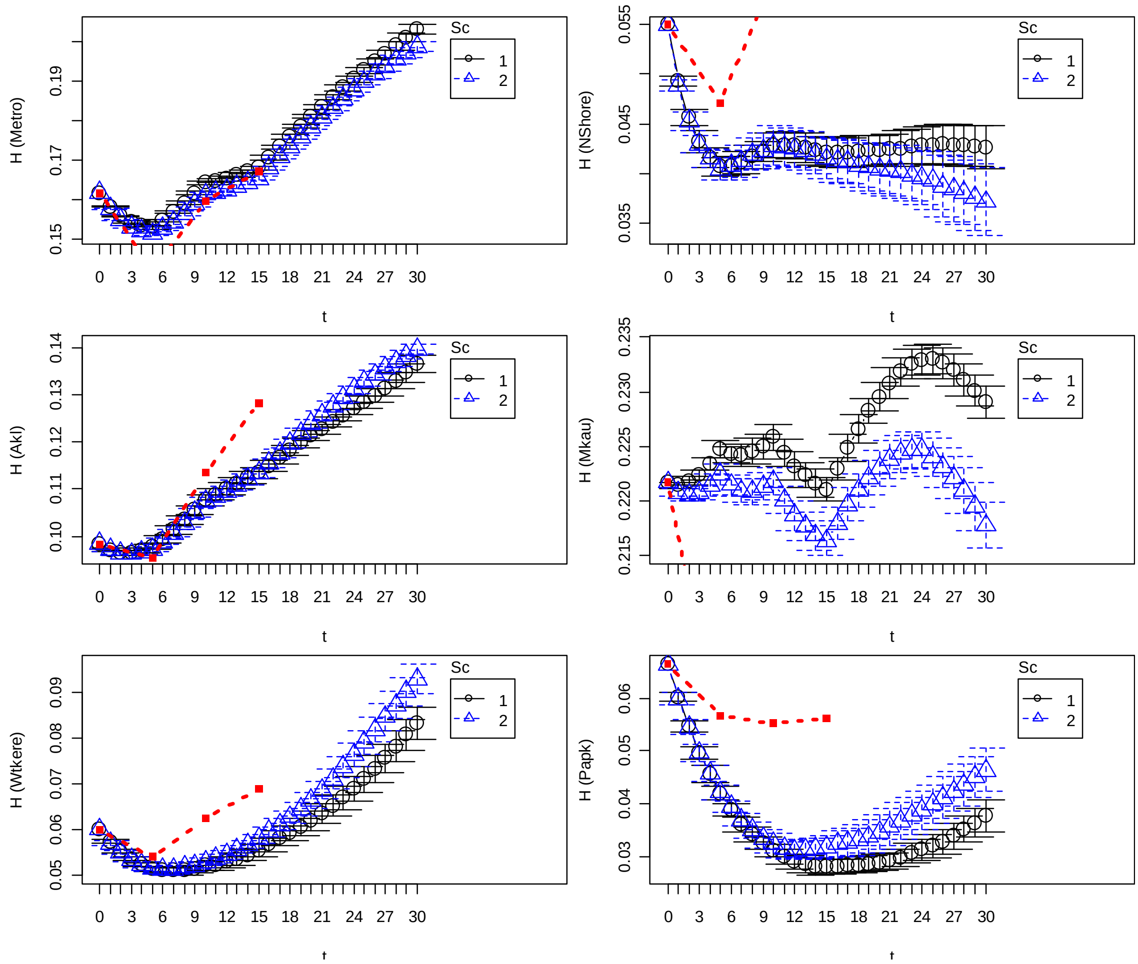


Figure 7.22 H-values Under Two Different Scenarios (MA and 5TAs)

More restricted housing availability at this large TA creates a new dynamic in the urban area. In term of Hs, the global value fell slightly (figure 7.22, top left), as it did in *Manukau* and *North Shore* but it actually increased in *Auckland*, as well as in *Waitakere* and *Papakura*.

By examining the population change, we can observe that the overall population in *Auckland* has effectively decreased, with the repercussions manifested in a swelling of population in all other TAs, as those moving were naturally looking for vacant spots in places where available stocks existed.

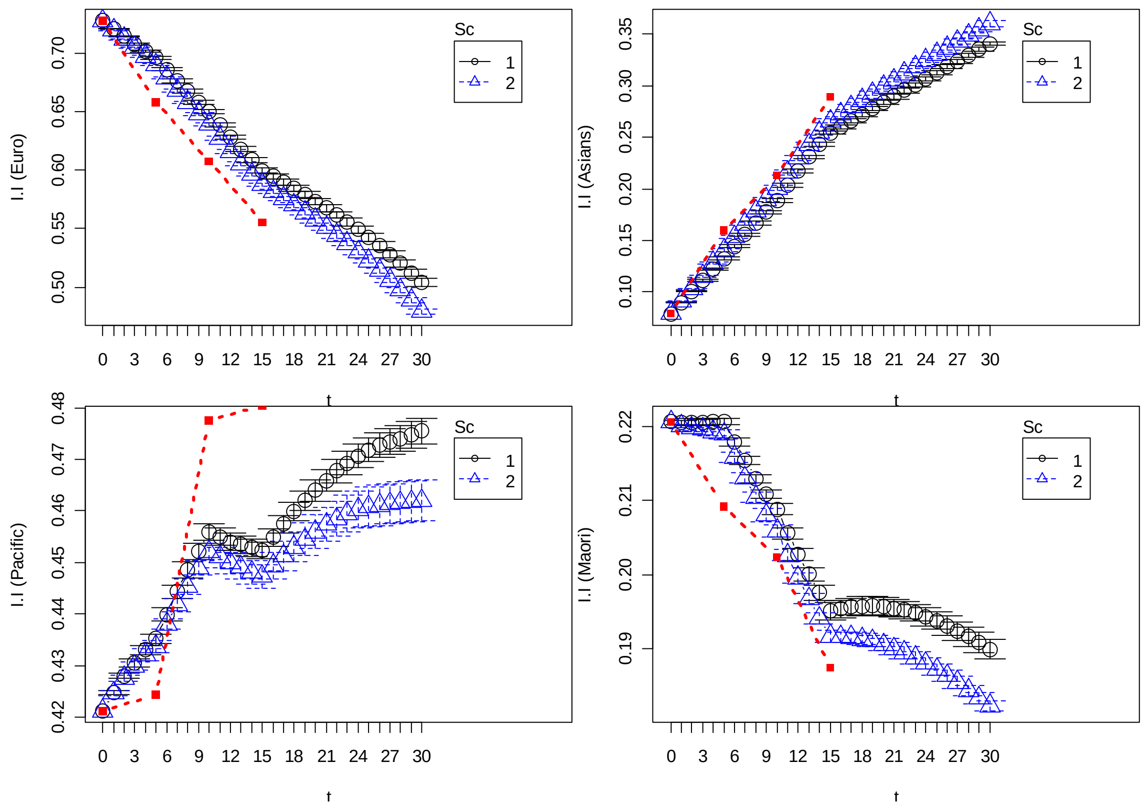


Figure 7.23 II_{xx}-values Under Two Different Scenarios (MKAU)

Although the tight vacancy in *Auckland* has caused the sizes of all its ethnic groups' population to decline (about 19% less population in total), the shares have not been the same. For example, the share of Asians is considerably larger (about 28%). Similarly, these shares were dissimilar for the receiver TAs too. *Manukau*, for example, received a larger share of Asians than any other ethnic group. This has resulted in a new distribution order, some elements of which can be seen in the isolation index of this TA. As can be seen in figure 7.23, the random chance for Maoris and Pacific people in particular (the largest minority group in this TA) to meet other ethnic-like people has deteriorated and the overall dynamic has caused the new aggregate weighted average deviation of each area unit's entropy from the entire *Manukau* TA entropy (H) to decline.

As can be seen in figure 7.24, this TA (*Manukau*) has also seen a rise in the numbers of AUs which are predominantly Asian.

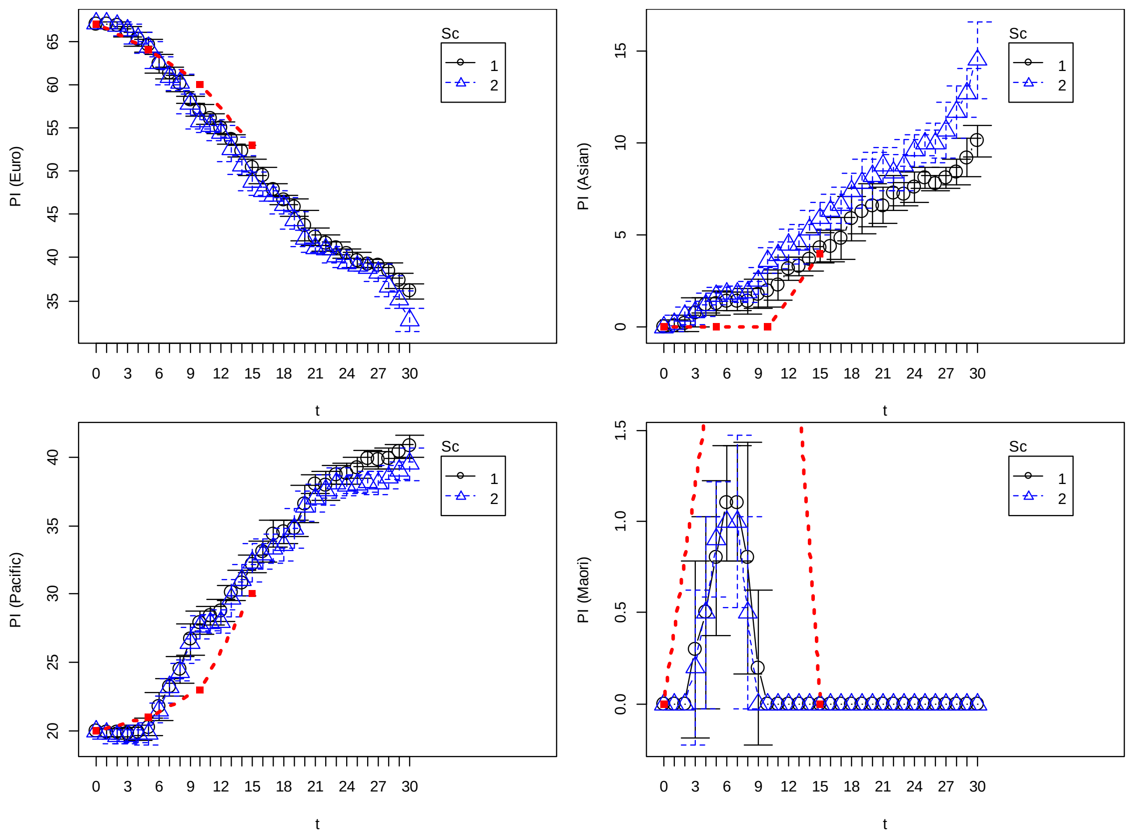


Figure 7.24 PI-values by Ethnicity Under Two Different Scenarios (MKAU)

The lessons from these experiments conducted in this modelling tool for this section suggest that housing availability (level of vacancy) plays a key role in the segregation dynamic by 'pulling' or 'pushing' prospective settlers towards or away from certain areas of the large urban system. The result of this 'indirect' influence on the distribution of population can be detrimental from the segregation point of view if housing development projects, for example, are concentrated in only specific areas/ neighbourhoods and they attract specific segments of (ethnic or other categories) the population more than what is ideally needed for the balanced distribution of that neighbourhood/ district.

7.6 Economic Change Effects Using a Proxy Factor

One of the questions which naturally come to mind regarding segregation is the effect of individual/ households earning (income) and the extent to which it plays a role in the dynamic and level of segregation.

The economic power that comes with higher revenue would allow those endowed householders a 'freedom of choice', that is, to choose a location and settle wherever they prefer in the large urban area. With the lack of 'income' variable in the current modelling framework, another variable was included which can be seen as the 'freedom of movement' factor and can be used as a proxy for 'economic power', income and 'freedom of choice'. This was done by specifying a probability of global locational move preference value, P(GL) [here noted at L], which determines stochastically a proximate percentage of those agents granted 'freedom of movement', seen as a financial capability to choose and settle in any part of the metropolitan area.

In the calibrated version, different restrictive L-values were assumed for ethnic groups (i.e. none with total freedom of movement, see table 7.1), giving Europeans a higher global movement percentage over Asians, who in turn had a higher percentage than Pacific or Maori people. Here, we completely eliminate this restriction for both Asians and Pacific people by allowing those internal/ domestic migrants (i.e. endogenous movers generated as a result of the turnover process) to move and settle anywhere in the metropolitan area, and examine the effects with the original setting (BaU). These scenarios are summarized in table 7.7.

Table 7.7 Scenario Table: Comparing Different Global Move Capability Effects

Sc ID	Scenario Name	Operative Time	$L_{2\&3}$
Sc1	Med (BaU: restricted move for Pacific & Asians)	t ≥ 1 (1991)	30%; 60%
Sc2	Med (global move freedom for Pacific & Asians)	t ≥ 1 (1991)	100%; 100%

Perhaps not surprisingly, the simulation outcome confirms that such economic power/ freedom is more segregationist-prone. We can see that the diversity of ethnic population distributed throughout entire urban area (measured by H) has deteriorated along with its constituent TAs, except *Papakura* (see figure 7.25). In this latter case, the financially liberated Asians and Pacific Islander agents – taking into account their preferences and other governing conditions (e.g. vacancy, etc.) – did not in our model find *Papakura* attractive enough for relocation.

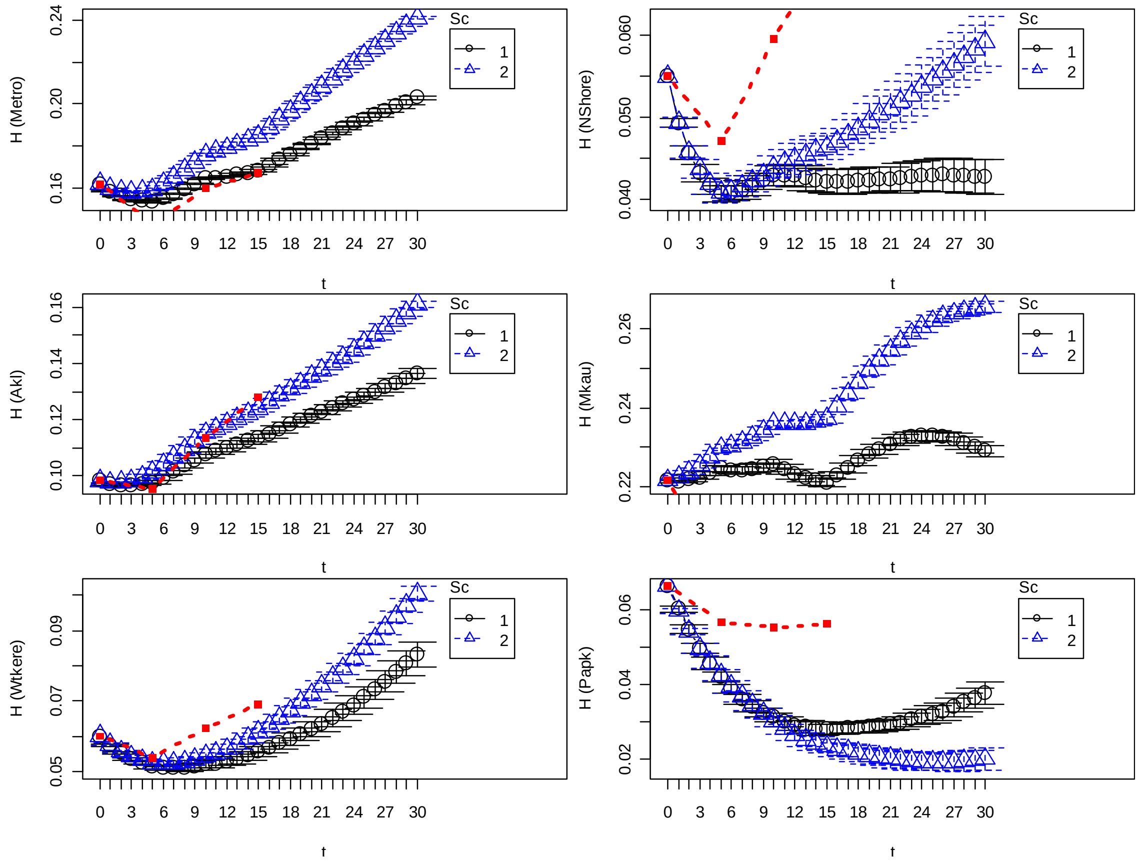


Figure 7.25 H-values Under Two Different Scenarios (MA and 5TAs)

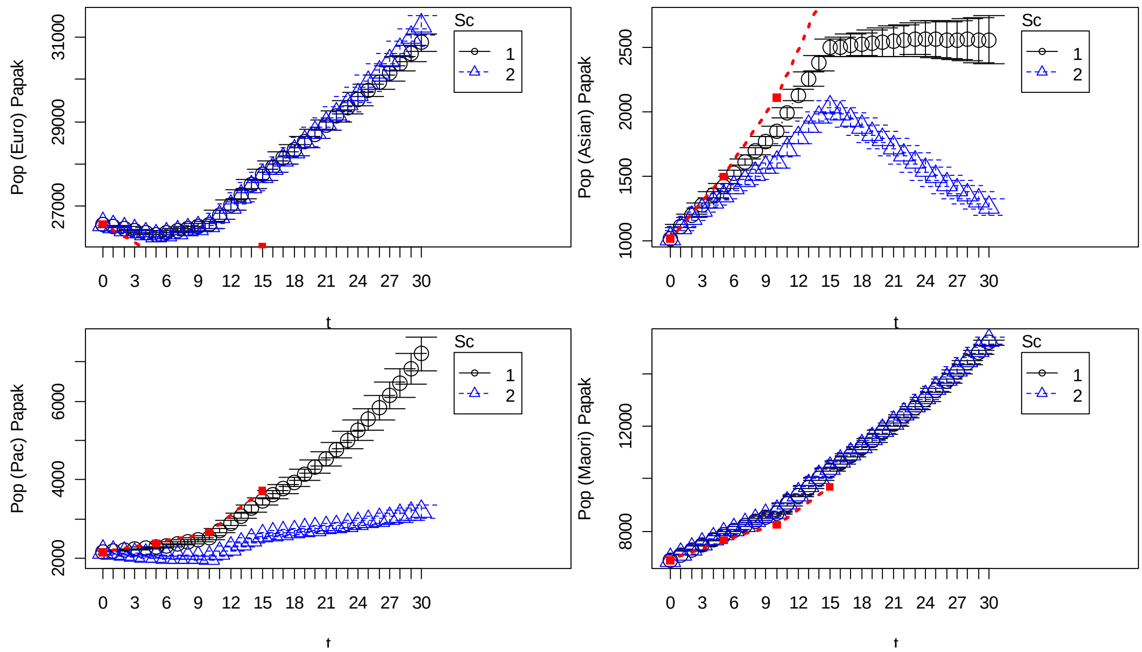


Figure 7.26 Population by Ethnicity Under Two Different Scenarios (*Papakura*)

By looking at the population of *Papakura* (figure 7.26) we can see that fewer numbers of Asians and Pacific people migrated to this TA under Sc2 scenario, something that certainly contributed positively (from a TA-based H level perspective) to the more balanced distribution of ethnic populations in at least some of the constituent AUs.

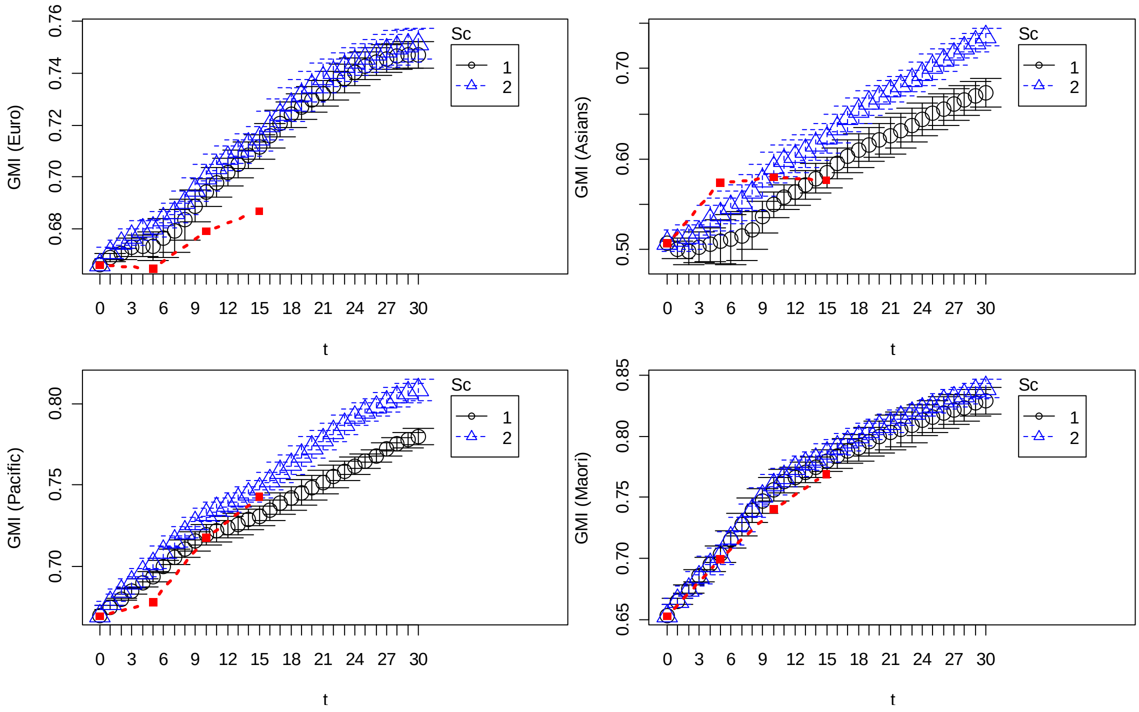


Figure 7.27 GMI-values by Ethnicity Under Two Different Scenarios (MA)

The increase in the segregation index level is not only limited to H, as other global measures such as D, I.I. and GMI (the only one shown here in figure 7.27) have also intensified as a result of the residential movements of these financially unrestricted agents.

There has also been some change in the visual characteristics of the urban area. For example, at the macro level, the numbers of area units dominated by Asians have increased under Sc2 (so that of *Auckland*). In other places, such as *Manukau*, the PI for Asians was slightly lower, as it was for Pacific people in *Waitakere* (see figure 7.28 and 7.29).

Perhaps the dilemma which has been observed in these experiments of having more prosperous householders living in an inevitably more segregated urban area may be attenuated (or eradicated) if the improved economic conditions are accompanied by a promotional (gradual) behavioural changes campaign (e.g. by media). Behavioural change will be examined in the next section.

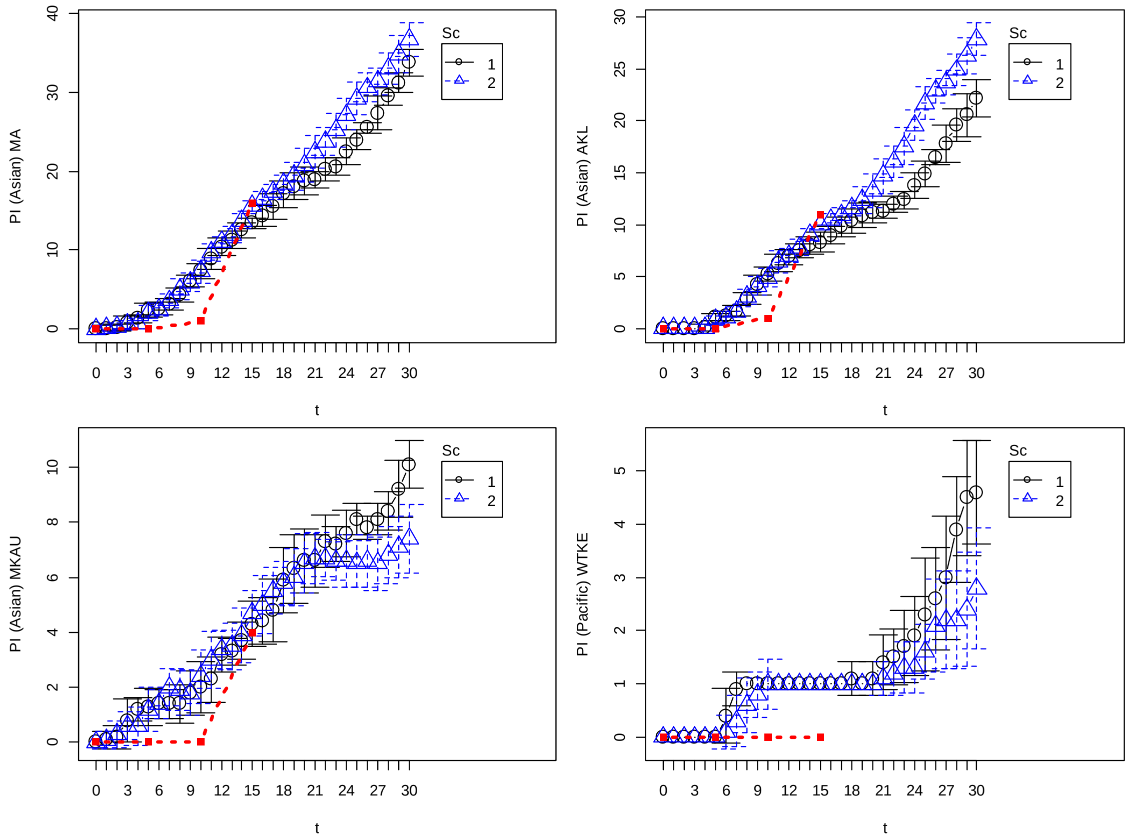


Figure 7.28 Selected PI-values Under Two Different Scenarios (MA/ TAs)

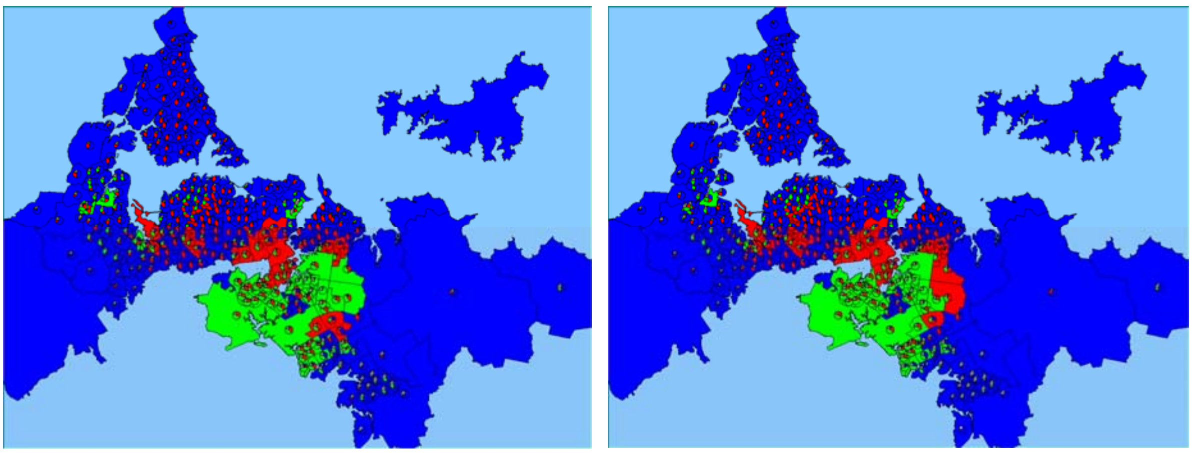


Figure 7.29 MA Snapshots, Sc1 vs. Sc2

7.7 Behavioural Change Effects Scenario

A 'classic' question in any Schelling type investigation is how the change in attitude and behaviour of agents vis-à-vis the presence of other ethnic groups in our neighbourhoods influences the dynamic and level of segregation. In other words, what would be the effects of more tolerant householders in choosing their residential locations in neighbourhoods with less (or ideally without) consideration for their racial compositions.

To examine this, we will reduce the intolerance level of Pacific people by a fourth of its original calibrated level after 2006 and compare the outcome with the BaU case, as summarized in table 7.8.

Table 7.8 Scenario Table: Comparing Different Tolerance Behaviour Effects

Sc ID	Scenario Name	Operative Time	T_3
Sc1	Med (BaU)	t ≥ 1 (1991)	0.12
Sc2	Med (lower tolerance for Pacific people)	t ≥ 16 (2007)	0.08

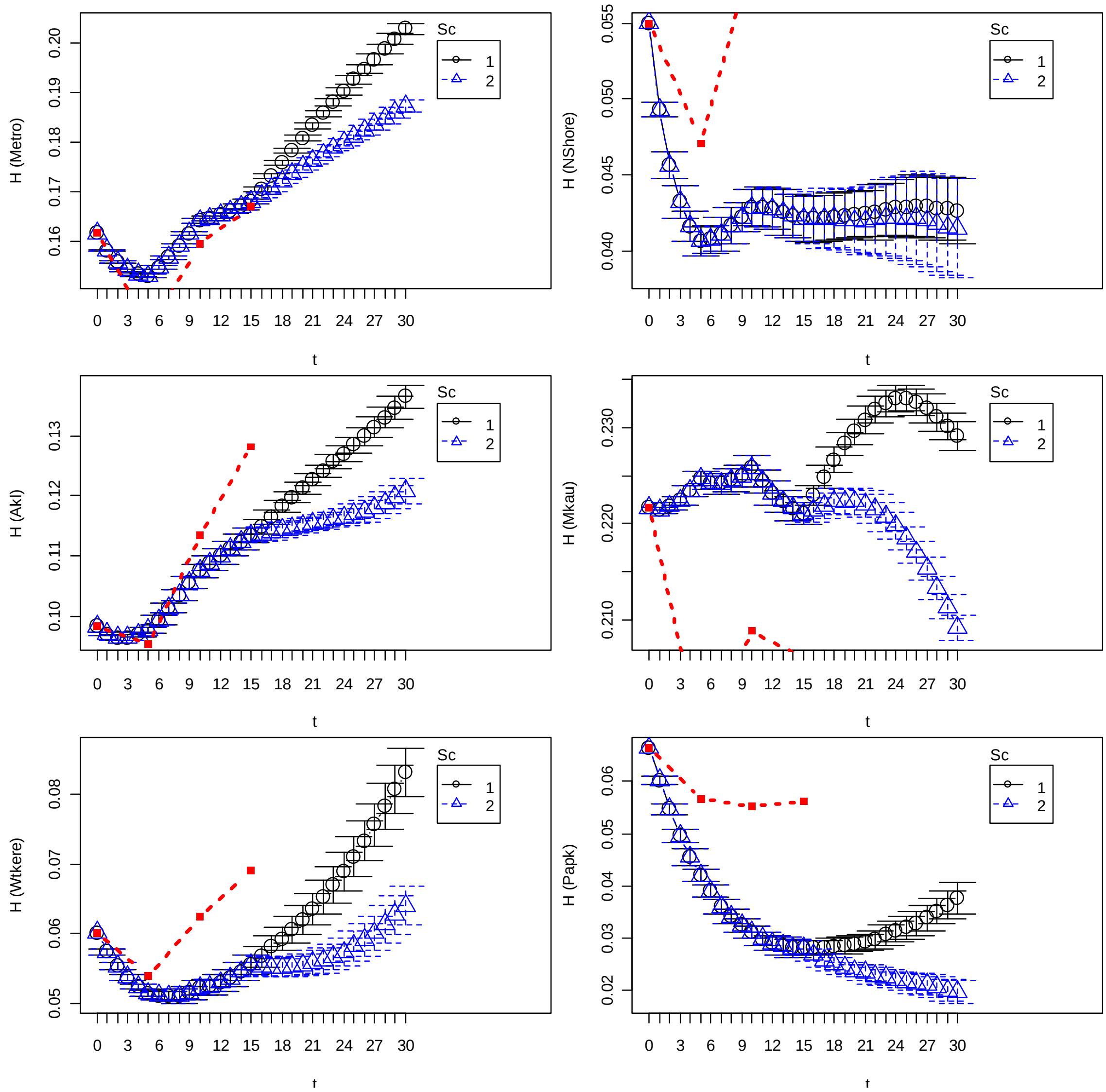


Figure 7.30 H-values Under Two Different Scenarios (MA/ 5TAs)

As expected the outcomes confirm the reduction of segregation measured by H at the macro geographical level, but also in practically all the meso-geographical constituents (albeit vaguely in *North Shore*, which incidentally sees a small boost of its Pacific population under Sc2), as can be seen in figure 7.30.

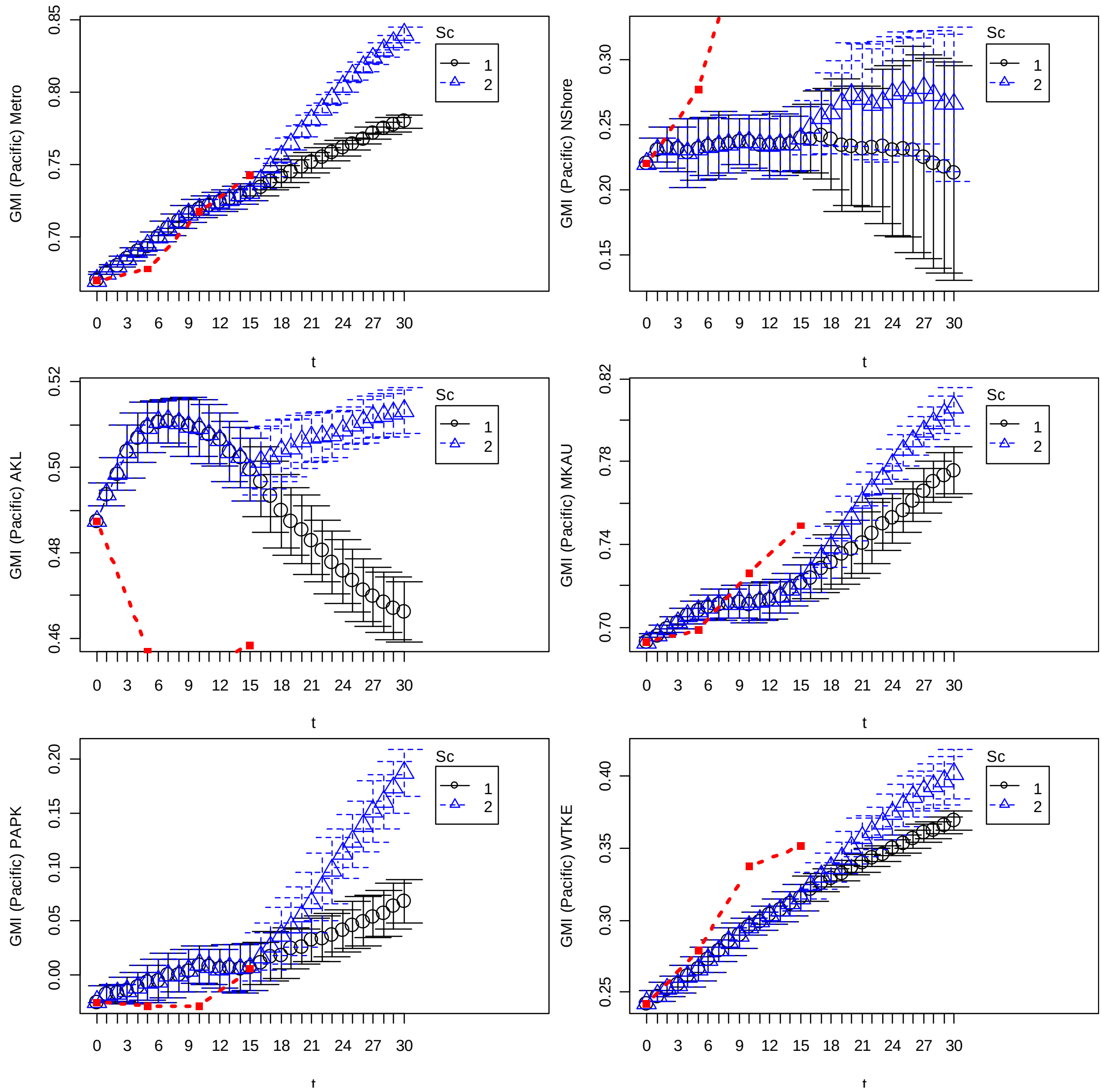


Figure 7.31 Pacific People GMI-values Under Two Different Scenarios (MA/ 5TAs)

However, unlike the experiments in the previous section where a rather more positive correlation between H and GMI was observed, the correlation in this case seems to be rather negative (see figure 7.31). This is because in the previous section the increase of H was engendered by the relocation of agents who had total freedom of movement. This freedom of choice for the preferred locations at the large macro urban scale also increased the positive autocorrelation for those non-restricted ethnic agents.

On the hand, even if the global entropy level (H) has declined as a result of more 'tolerant' behaviour (which made the entropy of each or some of the constituent area unit less divergent from the global entropy), Pacific people in this experiment can still more easily find a place that satisfies their more relaxed behaviour (preference), and this continues to produce a higher level of autocorrelation (see figure 7.31).

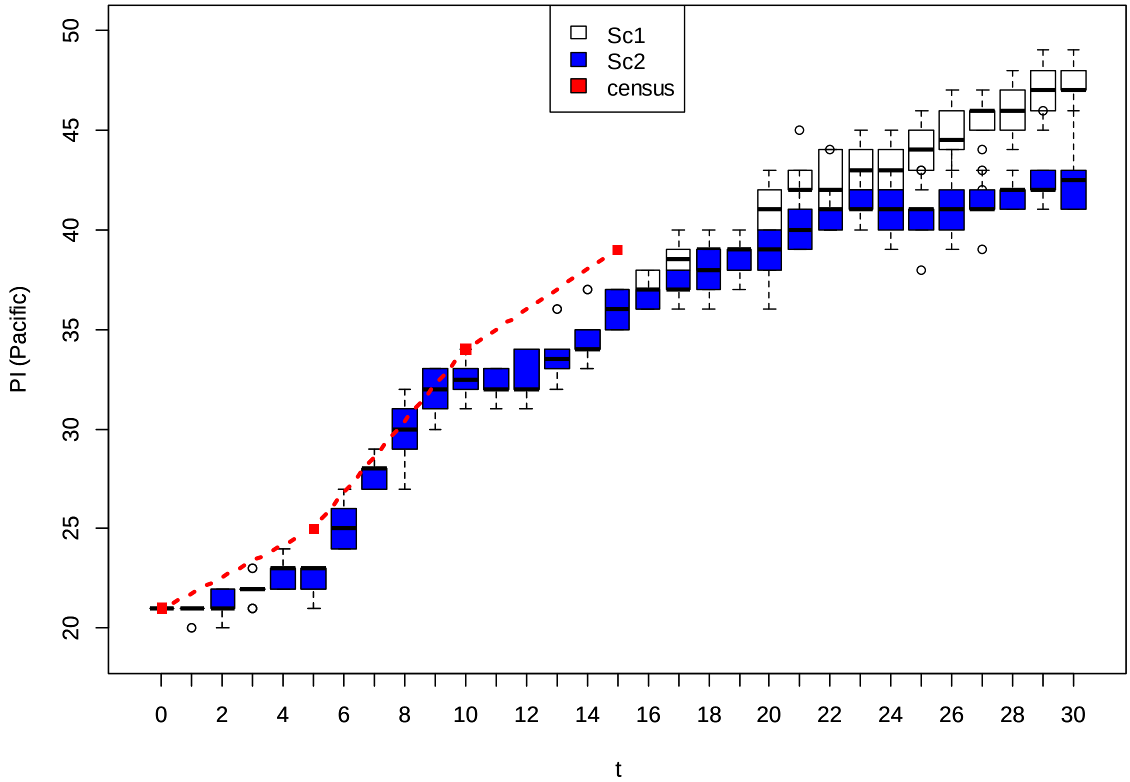


Figure 7.32 Pacific People PI-values Under Two Different Scenarios (MA)

The decline in the total number of area units where Pacific people were the predominant group before and after relaxing their T-value (see figure 7.32) is also another indication that changes in behaviour such as more open attitudes and views regarding their neighbours' attributes would be helpful in changing the urban mosaic of cities with fewer segregated neighbourhoods.

7.8 Discussion and Conclusion

This chapter presented a methodical approach for calibrating models where the patterns are important feature of the validation process and its success. It is particularly useful when the sweeping of the entire parameter space is not possible or practical, for example because of computational limitations. 'Multiple realizability' is probably one of the most challenging parts of the calibration and validation process. It may not be surprising that most models that aspire to validation stick with one single target (pattern) to match with their outputs. In this chapter, although we chose the global H as the main target pattern, other measures such as D, GMI, I.I and the visual map of the city (considering PI-values for the same purpose) were also taken into account. Moreover, these measures were not only available at the macro geographical level (i.e. MA), but also in all five meso-geographical constituent 'districts' (TAs). In general, this approach has a double-sided effect. From one side, it makes it almost impossible to have all the measures correspond to the real modelled world. (At least, with the models where the decision-making mechanism is a much simplified version of that in reality, something that arguably, models are supposed to be built for.) On the other side, the approach allows more insightful calibration/ validation as it points to the deficiency and imperfection of the realized outcomes, even if they have a high correspondence to the empirical-based global patterns (i.e. macro-level target). This often allows a remedying of the situation by selection of a 'better parameterization' among the multiple conceivable realizations (in a way, it can reduce the set of multiple realization space). This is even the case if the selected configuration may have sporadically less corresponding 'traits' as in its targeted pattern. This seems to be like settling on a compromise 'middle ground', which was done during the POC of the model in this chapter. Better matching was possible if the only target was the global H, but the final selected parameterization could also generate interesting (if not occasionally remarkable) patterns of D, GMI, I.I. and map the city's mosaic plurality – above those "path dependant" (Brown et al., 2005) area units/ regions – notwithstanding that the main target (H) was not a 'perfect' match.

In any case, with the improvement of the quality of the calibration (in general) as an open ended endeavour, it was possible to use the model for the objectives set in the thesis. Using several scenarios, we have shown in this chapter the potential of this model (even in its current state) as a policy-informing tool for questions relating to residential mobility, choice and segregation at different geographical scales. More cases were covered (and the situation was explained as to a lay third-person policy maker who may use and observe the results), instead of engaging in the ‘cause and effect’ relationship explanation. In a complex system – such as in residential segregation – the latter question is itself an important topic, although explaining the ‘causality’ in a complex system (even if possible in a reliable manner) is not often the main objective “as the interest of simulation is [often] at the side of observing the property of emergence” (Wu, 1999, p. 202). Nevertheless, recording additional data during the simulations and providing advanced statistics along with better and more sophisticated (data) visualization (ideally including ‘on-the-fly’ mode and integrated with the simulation framework) as intended for the future development of this model will certainly be a step in the right direction towards a better understanding of complex systems and phenomena using a modelling approach. Furthermore as complement, using complex-systems specific methods to analyze the joint influence of changes in multiple model parameters on model output can also potentially be used, including ‘customized’ regression based meta-modelling and Design of Experiments (Happe et al., 2006), time-dependent global sensitivity analysis (Ligmann-Zielinska & Sun, 2010) and web of causation (Krieger, 1994).

Towards this direction, it would be conceivable to use agent-based modelling frameworks as decision support systems/ tools to acquire valuable information about the modelled complex system, particularly when the criticism of the ABM approach based on its “absence of generality”, or “difficulty of estimation” fails, in the face of concrete evidence (Leombruni & Richiardi, 2005).

Chapter 8

8. Conclusions and Future Work

8.1 Summary and Conclusions

This thesis set out to

- measure, describe and explain past, present, and possible future patterns of residential location by ethnicity, by focusing on change in Auckland urban settings,
- develop a simulation-based model as an approach to understanding the dynamics of urban neighbourhood change,
- develop an approach to the transparent verification of simulation models of residential segregation using aggregate census data,
- examine the influences of some of the underlying drivers of residential location choice in the Auckland region through the period 1991 to 2006,
- investigate the linkages between the social micro and macro levels, and to operationalise these linkages in the simulation model,
- establish census data as potential 'test-beds' for future modelling-based research which can address theoretical issues in urban geography and sociology.

To this end, previous modelling approaches were examined so that a basis for choosing an improved modelling approach could be established. The subsequently proposed agent-based model was developed as a response to the limitations of a grid-based approach (in particular), whilst also embracing the vision of unifying agent-based modelling, microsimulation and geosimulation approaches under the same umbrella. The model was informed using census data and to validate the model the results (aggregate and spatial) of the simulated scenarios were compared (rather 'qualitatively') to the census data of the Auckland urban region from the period of 1991 to 2006. Several scenarios were simulated to examine the influence of various parameters on residential choice and its dynamic and the potential of the model for informing policy-related questions was demonstrated. The linkage between micro elements (e.g. behaviour at the individual level) and macro elements (e.g. higher level structures generated by the behaviour of agents at the individual level with a feedback loop to the lower level, reinforcing the behaviour indirectly, by limiting the residential choice), in a bottom-up, top-down interplay were discussed. This was done within three (micro/ meso/ macro) geographical scale perspectives.

A static statistical analysis approach was used to measure and describe the patterns of residential segregation in the Auckland metropolitan area. The simulations provided a mechanism to produce similar patterns (with some level of certainty based on the results of previous census years) and projected some 'information' about how the future patterns may look like (if initial conditions remain the same or changed). The study has therefore demonstrated the potential for using census data (along with the modelling approach) as a 'test-bed' to investigate theories of sociology, demography, economics, and urban geography in a rigorous and transparent environment.

This research is unique in a number of ways through its objectives, theoretical and conceptual frameworks, simulation modelling framework, the various methodologies used and through its empirical validation. It is the first Schelling-type model of its kind with microsimulation (MSM) and geosimulation features using aggregate empirical data (including realistic vacancy rates, whilst micro individual households remain at the core of its agent-based modelling approach). This allows various segregation dimensions to be measured both at meso and macro geographical scales. The micro agents decide where to relocate according to their multiple-level preferences. One of these preferences (global locational mobility) allows agents to express both global and local notions of a segregationist structure – in the sense that they are willing or (financially) able to relocate globally – or simply have local preferences regarding their preferred residential neighbourhoods. The global preference grants a freedom of mobility advantage which can be seen as a partial proxy for income. Agents in this model however do not have any knowledge of the 'global' emerging phenomenon (i.e. segregation) which arises as a result of individual local and global interactions, preferences and decision-making activities. The outcome is something that is sometimes referred to as 'weak emergence' (as opposed to 'strong emergence' where agents perceive it and are involved in its realization).

Furthermore, via its transition mechanism both synchronous and asynchronous modes are at play at each simulation cycle, that is; a turnover process, by generating moving agents along with a rule-based placement process (one agent after the other, in both local and global move preferences). As in microsimulations (MSM), 'micro' units (including individuals, households and groups in each area units) are identified by an important data characteristic (i.e. ethnicity) – which can be disaggregated from census data to the individual level – whilst the simulated population are usually 'aligned' with the aggregate empirical observations over census periods. Movements are generated by a probability based process (as in MSM), applied to each unit of ethnic group population in area units. On the other hand, as in agent-based models, moving agents use rule-based mechanisms to settle in fitting locations. This is the first facet of model 'hybridity'. As a second facet, the model also allows both realistic and artificial (grid-based) environments to be simulated in the same framework, a feature not previously used in other residential segregation models. Lastly, like the Schelling tipping model – it is a bounded-neighbourhood based model where certain activities such as immigration (of agents) occur exogenously to the model – whilst typical endogenous activities simultaneously take place in a checkerboard-like or realistic environment of multiple bounded-neighbourhoods.

The model is informed and validated by census data (available in the public domain) which have been altered to minimise the level of error ('noise') embedded in the data. It is also the first residential segregation model of Auckland (New Zealand).

Chapter 1 explained the motivations for embarking on this individual-based simulation modelling approach and established a set of research questions and objectives and gave the details of the structure of the thesis.

Chapter 2 described the old and ongoing residential segregation phenomenon and explained why residential segregation is a complex and multidimensional problem and why its determinants and causes are not well understood. Using the literature, the chapter provided details of different modelling approaches and exposed the shortcomings, vulnerabilities, and particularly the lack of empirical support of previous modelling endeavours.

Chapter 3 was dedicated to the description of various aspect of the research design. It clarified the hybrid modelling approach which embraced a combined agent-based modelling, microsimulation and geosimulation design. It also proposed a conceptual framework within which the scope of the research parameters was established, as well as a research model on which the model was deployed and used.

This prepared the ground for a description in chapter 4 of the actual design of the simulation model which was built around and shaped by data (O'Sullivan, 2004). Following loosely the ODD protocol (Grimm et al., 2006; Grimm et al., 2010), the description of the final model (*HAAMoS*), its characteristics, behaviour of agents, and its scheduling were presented. This chapter also compared the *HAAMoS* with other Schelling-type models. It was then demonstrated how this model can still exhibit the same basic behaviours that normally occur in simple Schelling and Fossett models.

Chapter 5 verified and exhibited the behaviour of the model (and the agents) and the extent to which it conforms to the behaviour of the original Schelling and one its most sophisticated extensions – the Fossett model. Prior to this, this chapter gave the details of the selected measures implemented in the model for gauging different dimensions of segregation.

Chapter 6 unpacked the empirical data and explained why it was necessary to delimit the data for this study. This was followed by the description of a variety of sources of noise in the data (in particular the effects of the 2006 New Zealander category surge and multiple ethnicity choice) and explained the approach to attenuate their detrimental effects. Therefore, the data used in this study was somewhat different from data used in other studies (see, for example, Johnston et al., 2009). Descriptive statistical analysis was applied to explore the census-based data. Tables, graphs, figures and measured values were essential elements to achieve the objective of describing the past and present patterns of residential locations in the Auckland region by ethnicity. These census–based measures formed the basis for an empirical benchmark set that was used in the simulated scenarios in order to validate the outcomes. A methodology was also proposed to identify the 'hottest' and 'coldest' spots, using as example the case of Asians as a variable of interest.

Chapter 7 proposed another methodology called pattern-oriented calibration (POC) for the calibration of the model. A series of realistic scenarios, including policy-oriented ones, were built, executed and the results were presented and discussed in this chapter. This was done using empirical values for the parameter (e.g. number of 'immigrants', vacancy rates, etc.). Although it was not always possible to obtain the exact patterns (locations) and values as generated in the real world and from multiple perspectives (on data), the model demonstrated its potential as an analytic and decision-informing tool. At the least it was demonstrated that the model has the ability to generate plausible changes similar to empirical residential distributions over time – even though detailed information about behaviour or detailed representation of decision making behaviour was not included.

Thus the overall goal of this study, the development of a potential 'test-bed' consisting of an agent-based model and census data for future modelling-based research which can address residential segregation questions in urban geography (and sociology, economy and the like) has been attained. However, with the lack of data on residential decision making behaviour in regard to ethnicity among the Auckland population, a credible calibration of behaviours in the real city experience remains a challenge.

8.2 Methodological Implications and Caveats

One of the methodological strengths of this study is its ability to simulate an entire population of a large metropolitan area without extensive use of computer resources (e.g. memory or CPU-time) or by deploying it on parallel, distributed, clustered or grid computers. This is achieved through the use of aggregate levels of data (population) and by assuming that at each time cycle; only a small percentage of the population would decide to change their residential locations. At that time, these normally dormant/ inactive 'agents' would become 'conscious' and act upon the realities of life related to the need to change their current homes. This rationale is supported by the behaviour of individuals/ householders in the real world so that unless pressing circumstances arise, people do not usually decide (out of the blue) to embark on the daunting task of searching for a house and relocating. Therefore, in many cases, simulating the entire population by creating instances of agent(s) who are alive and alerted, and physically occupy memory space but are idle most of the times would be an inefficient and costly (in terms of computer resources) way of modelling. Although today's computers are faster, cheaper and more powerful, simulation of computational agents at large scale is still a real issue and this 'aggregate' approach could be an interesting alternative method for many current problems.

This approach has both pros and cons and it will not be suitable for all types of problems and investigations. For example, it is not possible to continuously monitor the states of all individual agents, if this was needed in order to trigger some specific actions after one (or some of them) reached a certain threshold level. In the context of residential segregation, for instance, it will not be possible with this approach to embed a 'happiness' gauge, to monitor the satisfaction of agents about their current locations and allow them to relocate if their dissatisfactions reach certain levels or exceed their own individual limits. Relocations can alternatively be produced 'synthetically' using a probability-based turnover function, as has been used in the *HAAMoS* model. This can also be justified theoretically since in reality there exist many other external factors causing (quasi) forced (not totally voluntary) moves such as change of job (proximity to workplace), change in age of children when there is a desire to send them to high ranking schools in different locations, dissatisfaction with (or lack of) public transport service, new noisy or undesirable neighbour, to mention a few.

This feature also allows the operation of both synchronous and asynchronous modes at each iteration cycle of simulation. In the synchronous mode, dissatisfied agents decide to move (leave the neighbourhood) at the same time, something that is emulated by the turnover process. On the other hand, the placement of the agents is done in asynchronous (updating) mode (see, for example, Cornforth et al., 2005); that is, agents relocate individually, so that no two agents can move to the same neighbourhood with one single vacant spot left. In other word, once an agent is in the process of making a decision to move into a location, the departure or arrival of another agent does not affect his choice (the agent observes the changes in environment immediately after they happen). This may also be seen as a kind of 'adaptability' (though built in the model rule and not in agents) so that when an agent causes a change in the environment, the agents coming afterwards have to adapt to the new situation and act in a new (updated) environment.

Another implication of the methodological choice of using aggregate forms of individuals/ households at the bounded neighbourhoods is that the immediate neighbourhood (in the case of a grid environment whether Moore or Von Neumann) corresponds to the neighbouring area units (and not the neighbouring households as it would have been the case in the continuous neighbourhood environment). This means that the behaviour of the agents in this modelling environment may not be fully in line with the realistic behaviour of householders who decide to relocate to desirable locations (i.e. when the proportion of ethnic-alike people is satisfactory). For example in this setting, the individuals/ householders who consider settling in Mt. Albert (Central) area unit will evaluate the overall residential segregation indices of the surrounding area units of Point Chevalier (South), St. Lukes, Springleigh, Owariaka, Sandringham (West) and Walmsley. Investigating what percentage of population may decide about their residential choices in such a way was not part of this study and existing surveys seem not to have included this aspect/ question regarding the decision-making process. However, it will not be surprising to learn that fewer people (among those who consider ethnic percentage as a decisive factor) are inclined to consider such a large area (notwithstanding that this would depend on the size of the area units). So, it is important to bear in mind that that this methodological approach may not be (greatly) supported by empirical behaviours.

Although the current model implements only a single 'actor' (object) called the realtor, the design of the model allows the integration of more realtor agents into the modelling system with reasonably small programming efforts. This would enable the examination of discrimination at different levels; that is, it also allows studying the unequal distribution of residential locations and the effects of discriminatory behaviours of those agencies that have the power to influence this distribution. Moreover, by adding additional realtor agents, a more heterogeneous environment of interacting household and realtor agents would enable a further advantage of the ABM dynamic power in the modelling system to become evident. Since the realtor agents will always be 'alive' and present throughout the lifetime of the simulations (unlike household agents, because of their more manageable/ limited numbers), it would also be possible by implementing more sophisticated algorithms for them to exhibit adaptive learning behaviours. This way they can behave differently from each other and in some cases for example, they can even form or adapt new discriminatory strategies against certain ethnic groups. This would add a new dimension to the modelling framework (eventually evolving to a CAS framework) and provide a novel opportunity to investigate the effects of 'racial steering'.

With regard to topological issues, it is important to bear in mind that similar to no-toroidal grid-based environments, a realistic GIS-map simulation can potentially suffer from edge effects. The map is cut off from the rest of the geographical areas based on artificial/ administrative boundaries imposed by the SNZ. In reality these cut-off area units at the outmost edges of the map are connected to other geographical areas (area units), where population can move in and out. The structure of neighbourhoods in a

GIS-map (compared with a toroidal chess-board environment) is non-homogeneous (i.e. area units do not all have the same number of neighbours). However, by eliminating those areas which are out of the areas under study, the structure of the neighbourhoods will be even more non-homogeneous and potentially remain vulnerable to edge effects. It must be noted though that this phenomenon is somewhat mitigated by the Auckland region's particular topology which is surrounded by the sea.

Relevant to this topic is the suggestion in the recommendations for further research section that future extensions of this model should consider introducing a distance-based weight factor into the decision-making process. This is because agents will not rate (and relocate to) area units regardless of their distance from centres of 'business' activity and amenities. A more complete but sophisticated approach would involve utilizing techniques used in integrated land-use/ transport models (also known as Land-Use Transport Interaction (LUTI) or urban models) where land-using activities generate the demands for transport and new developments (housing, malls, job centres, etc.) which would emerge around newly built transport hubs/ stations along with the presence of developer, investor and producer agents as well as financial, labour, property, transport and product (goods and services) markets.

Another methodological feature of the model was its ability to allow the study of residential segregation in both inter and intra-urban and at three different geographical scales. Although for a variety of reasons this proved to be an advantage (including a more informed selection among multiple realisations resulting from a calibration with a global pattern as target), it can occasionally appear as a disadvantage for it would be difficult to validate a model which has been calibrated on a macro data/ measure and have it correspond to all the meso-based measures (i.e. perspectives on other smaller geographical constituents) too. It is true that this feature can occasionally irritate modellers as it often reveals the potential deficiencies, weaknesses or imperfections of the model during calibration and validation process (or at least the feeling that this would be an 'impossible mission'). On the other hand in the long term at least, the consequences of this pointing to weaknesses and deficiencies would lead to improved models if seen as opportunities to remedy them. At least theoretically (and arguably) it may be sensible to think that embodied agents in the model would eventually react in a similar manner to their real world bodies if they are provided with the same mechanism of thinking and sensing. This might though be the crossing of the ditch between the traditional use of models and the newer world of 'virtual reality' (e.g. artificial societies) as a fresh modelling paradigm.

Although the number of ethnic households could have been calculated by dividing the ethnic population by the national household size, the outcomes using the household instead of the individual population would not have been expected to be significantly different. This is partly because agents do not currently encapsulate any information related to household characteristics, which was not used in this study. However, as household characteristics (along with some lifecycle processes related to household structure) will be added to the model, it would become important to incorporate households in the model too.

It is important to note that most of methodological implications discussed in this section are not uniquely limited to the study of residential segregation and may be also relevant in other fields and territories where spatial agent based modelling simulation could be a potential solution.

8.3 Theoretical and Policy Implications

Like many other models using census data, the *HAAMoS* model is not immune to deficiency in theory or the shortcomings of data. The outcomes, analyses of results and their interpretations have been shaped by the various assumptions and suppositions (for instance about the behaviour of a specific group which may not necessarily have been empirically informed) that were used.

For example, Vasil and Yoon (1996) stressed that the single category of Asians has little meaning. In reality, there are 48 countries in Asia (including some in what is often referred to as the 'Middle East'). Excluding the Middle East (based on SNZ categorisation), the 2006 New Zealand Census of Population and Dwellings list includes the birthplace and ethnicities of people in Northeast, Southeast, and South Asia regions, totalling 27 Asian countries.

In today's New Zealand, this single categorical group may not be well explained through 'Orientalism' theory, that is, to enable "those who are not Occidental (i.e. Orientals) to be stigmatised as 'different' and to be kept at a 'safe' distance" (Sarup, 1994, seen in Pawson et al., 1996). It can though perhaps be explained by "historically based cultural mis(understanding), which also results in the stereotyping of Asian immigrants as a single identifiable group" (Pawson et al., 1996, p. 373).

Therefore, reading into the residential segregation of Asians (in particular) must be done carefully, since the diversity of people from Asia makes it by far the "greater diversity based on race, religion, culture, language and ways and values of life than Europe"(Vasil & Yoon, 1996, p. 5), for example.

As Bedford and Ho (2008) state, it would be

> […] unwise to think that discussions can go very far with a single category of Asian [since] diversity within this population is much too great. [In particular] and notwithstanding this warning about the irrelevance of a label of Asian for peoples with cultural links to a vast region stretching from the Middle East to Japan that Europeans have labelled as Asia, it remains common practice in New Zealand to refer to the country's Asian population as an entity. This is certainly the case with regard to projections of the country's population, especially when the recent ethnic projections show that the Asian population will grow much more rapidly over the next 20 years than the European, Maori or Pacific ethnic components (p. 1).

The second, third and further generations of the 'Asian' population may still choose the Asian category as ethnicity choice, but even among the first generation, special efforts to establish contact with the larger New Zealand society, especially the dominant group of Pakeha (European), and the effort to mix and mingle with them is documented (see, for example, Vasil & Yoon, 1996, p. 13). For instance, the concentration of Asians (e.g. Indians, Chinese and South Koreans) in specific neighbourhoods in the Auckland region is not as high as some other groups. In fact, their distribution is much wider than one might have expected/ assumed (Friesen, 2008).

Furthermore, as mentioned previously in this thesis, people can identify with more than one ethnic category when they complete the New Zealand Census question on ethnicity. This means that the ethnic categories comprise people of mixed ethnic origins and therefore these categories can potentially overlap.

Awareness of the extent of these issues – which at first may appear subtle – is particularly important when the goal of the model is to inform policy decisions.

The *HAAMoS* model has been built around the existing aggregate data. It is possible to test different theory-based hypotheses based on available parameters, although the model is extendable and therefore has greater potential for accommodating other types of investigations. To overcome the kind of problems explained above, data that are more refined are required. Notwithstanding, many questions suggested by theories about ethnic enclaves or ethnicity based residential location decision making can still be investigated by this model and their effects and trends into the future can be measured (quantitatively) and observed qualitatively. Therefore, the model (and its enhanced extensions) has the potential to be useful for testing different policy-oriented scenarios. Nonetheless, more work need to be done to ensure the reliability of the outcomes with relation to uncertainty (Lowell, 2007). This should normally include conducting separate uncertainty analysis in addition to sensitivity analysis.

It should also be kept in mind that in general "when treating space and time in its most disaggregate or local forms, microscopic models often do not perform to a level of `goodness-of-fit' comparable to conventional aggregate models" (Batty & Xie, 1997, seen in Wu, 1999). Although this model has the advantage of using more aggregate levels of data, it is still not a conventional aggregate model, but an ABM which is not immune to 'goodness-of-fit' challenges.

To build a model is relatively inexpensive for the potential benefits that could be realised. However, even with the successful introduction into the model of evolutionary characteristics (e.g. society, urban area and the like), simulation modelling may still not be the ideal tool for centrally engineering of society (Allen, 1997, p. 253). Moreover,

> [i]f the goal is to provide decision-relevant information to end-users and policy-makers, then their needs should help to define the focus of model development […]. What variables do policy-makers care about? What spatial scale and what time scale are of interest? The decisions of where to allocate computational resources (model resolution, number of model components, model complexity, number of ensemble members, number and type of scenarios and number of models) [in this thesis] were mostly pragmatic, guided by [research objectives,] scientific curiosity, experience from the past and computational constraints, rather than by an assessment of what would be most useful for the user and policy-maker (Knutti, 2008, p. 4660).

Although the main objective of this model was not to inform policies related to residential segregation, this study offers several promising policy related implications. First, the agent-based modelling approach shows encouraging potential to be used as a pertinent and promising tool for investigating residential segregation (or related/ similar topics). Moreover, empirical data and in particular census data can be successfully used to inform the model and serve to ease the uncertainties regarding future trends (projections). Many important individual characteristics related to residential choice (such as preferences) are not included in the census data. This can be complemented by empirical data obtained from survey studies, for example. Other aspects of the census data can be also included in the study, in particular economic variables (along with other socio-economic factors, if necessary complemented by other sources of data, such as median house prices or deprivation indices in area units) which were not used in this study. In such a way, this modelling framework would allow an investigation on a wide spectrum of different scenarios relevant to the dynamic of residential segregation. Perhaps the most salient outcome of this study is in its demonstration of a Schelling-type model which is capable of generating interesting and plausible changes in the residential distribution of a large population over time and at different geographical scales – without sufficient information about underlying social mechanisms or a detailed representation of decision making behaviour.

For a model to be used specifically for the policy-related analysis targeting existing social systems, an appropriate/ ideal approach would perhaps be to engage right at the beginning with stakeholders who are the "source of expert information concerning the particular social domains of concern", so that these "domain experts can provide descriptions of the goals and action of relevant actors as well as patterns and modes of interaction among them" (Moss, 2002, p. 7267). Nonetheless, determining whether this model (or its enhanced versions) is useful for such purpose "is much less a question of whether [it is] useful but more a question of how [it] can be useful (Louie & Carley, 2008, p. 26)".

8.4 Recommendations for Future Research

The residential choice model presented in this thesis has demonstrated how it can be used along with the empirical census data in a 'test-bed' environment for addressing residential segregation questions. But there remains much more to be done. As alluded to in the previous section, there are a number of future research avenues that could be pursued as possible extensions of this study.

These future works can fall under four topics: full/ exhaustive exploration of the current parameter space; or with extended data; enhancing/ improving the current version of the model; and the expansion of the descriptive statistical analysis methodology for identifying hottest and coldest spots.

Of these four, expanding the model capability with a new set of extended data and/ or enhancing the current version of the model would allow a more complete representation of the 'test-bed' framework and could thus address broader (theoretical and concrete) questions regarding residential segregation choice.

The current model allows many different scenarios to be built and investigated, most of which have not been presented in this thesis. In particular, investigating the emergence of segregation at meso and micro levels, combined simultaneous changes (behavioural or others), and the parallels between realistic and artificial environments' outcomes would offer a unique platform to look for potentially insightful findings.

Many aspects of the census data have not been used in this study. Broadening the residential segregation characterisation to include economic variables, such as household revenue, for example, would add another important dimension and realism to the system. Other supplementary data (not necessarily from the census), or micro-level census data which includes both geographical as well as individual or household characteristics would be essential for answering some specific questions relating to residential choice. For example, information about the price of houses in specific neighbourhoods combined with individual or household-based revenue can enable a series of interesting queries (economic scenarios) regarding the choice of residence and affordability. Similarly, with data on households' size, number of bedrooms, and the like, a series of socio-economic scenarios can be constructed. Although because of the lack of ethnicity-based data at the household level, individuals (from aggregate data) have been used to embody agents in this study (which was fine to demonstrate the proposed concept and model utility), using household agents would enhance the realism incorporated into the model. In general, better data would reduce the level of uncertainty in the quantitative aspects of the simulations and the conclusions drawn.

It would be interesting to allow both 'in-group' (e.g. 'homophily', as currently implemented) and 'out-group' (e.g. 'xenophobia') behaviours (preferences) to be enabled in future extensions of this framework to examine possible differences. Petrescu-Prahova (2009) make a distinction between these 'ethnocentric' and 'out-group avoidance' attitudes by arguing that "the former is associated with the clustering of households of the same ethnicity, the latter with the separation in space of households of different ethnicities"(p. 22). At the same time, allowing more explicit heterogeneity within the same group members would be justifiable.

Some small enhancements of the model would also allow broader investigations to take place. For instance, in the conceptual framework of the model, the realtor was considered a separate agent through which some forms of institutional discrimination such as 'racial steering' and 'manageralism' could be operationalized and investigated.

For this to become possible, several instances of the *Realtor* class should be created (much like household agents as instances of the *Household* class, although at much smaller numbers). In part, because of the already high magnitude of this model, it was decided to not to proceed with the realization of this feature. Therefore, there is currently only one realtor instance, making it impracticable to investigate these sorts of institutional discrimination questions (except for a central discriminatory policy through the realtor). For this reason, there was no need in this thesis to represent the realtor as an agent, but a part of the model mechanism (in implementation of the relocation process). The current model is a research development in progress and it is designed to be enhanced long after this thesis is completed. With the current flexible design, it would be possible to allow several realtors to be present throughout different neighbourhoods and interact with individual agents. Doing so could provide a novel platform for enabling investigations regarding discrimination at different scales, as well as an avenue for examining the effects of 'racial-steering'. Furthermore, these realtor agents could exhibit some forms of adaptability and learning behaviours (for instance, by using 'neural networks' techniques, 'reinforcement learning' [e.g. Q-learning, Roth-Erev], 'learning classifier system' or 'genetic algorithms'), so allowing more realistic behaviours to take place in the model. This can be seen as an enhancement of the current framework towards a CAS-based framework which might be more appropriate for the inherent complexity of residential segregation.

Currently, immigration and growth are treated as synonyms: knowing the realistic percentage of growth in an area, agents are created (with global mobility preference) and pooled before entering the world. Immigration should ideally be treated a subset of growth, and if known, natural birth increase should be deducted from immigration numbers.

Although realistic vacancy rates (at district levels) can currently be employed in the model, for projected years for which empirical data was not available, same (or average) rates were used as an alternative (with some controlled dose of randomness). For those projected years, a method (formula) that can dynamically calculate and adjust the vacancy in relation to the district (meso) or area unit (micro) growth would add both further realism and credibility to the model. For instance, Glaeser et al. (2006) found that when a city grows by 10%, its vacancy rate would decline by eight-tenths of one percent.

In the realist scenarios presented in this thesis, the realistic growth of the population sizes altered the proportional sizes of the groups' populations as the simulation progressed. As the proportion sizes of some ethnic groups change, determining and maintaining the intended level of tolerance would become a challenging task. A dynamic mechanism which will adjust and keep the tolerance level constant as a function of the proportional size of the population could be a part of the future plan for the extension of this model.

"[C]ities are complex organisms, evolving and changing according to local rules and conditions which manifest more global order across many scales and times"(Batty & Longley, 1994, p. 10). Future extensions of this model could include further life cycle and demographic processes (e.g. 'ageing', 'death') and life course events (e.g. 'household formation', 'marriage', 'divorce'). In general, the power of the modelling approach would be greatly (but perhaps arguably) enhanced if the evolution of this framework would gravitate towards an artificial society/ city model. The co-evolution dynamics in such a model would embrace the evolutionary nature of certain developments which involve human activities including land-use, transport routes or extensions, infrastructures, markets (financial markets, property markets, labour markets, transport markets and product markets) as well as the new commercial, business and service centres that would emerge as a result of dynamics between these components (even 3D-GIS urban representation could be a part of such system, allowing, for example, certain agents to choose residential apartments facing attractive amenities such as green spaces, parks, and sea views). This would make sparsely inhabited areas of the greater urban area more attractive for agents to move there naturally – whilst at the same time the nature of the simulated city/ society would remain "inherently unpredictable" (Batty, 2008a, p. 17) – in part, by its random growth and variance rates (Batty, 2008b, p. 769). Currently, there is higher probability for agents to visit locations with higher vacancies (something that is empirically defendable, as new developments attract potential tenants), but they are equally rated as long as they have a satisfactory racial composition. In other words, there is currently no mechanism in place to stop a potential global tenant to choose a very distant and isolated area unit over others, as long as other essential criteria are satisfied. This was partly the reason why the large districts of *Rodney* and *Franklin* were excluded from this study. Even without an artificial-city-based or CAS version of this framework, it would be possible to include a distance-based function in the weighting calculation of an agents' decision-making process. In any case, the simulation of the larger Auckland region including *Rodney* and *Franklin* should be part of any future studies.

Future models can also attempt to incorporate 'knowledge acquisition/ compiler' modules based on the previous experiences on the 'integration' of geospatial systems, GIS and artificial intelligence (AI) techniques and concepts (see, for example, Moore, 2000). This would improve various aspects of the simulation modelling framework, including calibration (POC) and 'error modelling', whilst inching towards delivering a better decision support modelling environment.

Continuation or expansion of the statistical methodology for identifying hottest and coldest spots, both at metropolitan area (macro-geographical) level and TA/ district (meso-geographical) level would provide fresh research perspectives and is therefore recommended. Consideration for using different spatial units as well as household numbers (instead of individuals) would potentially result in better empirically-compliant identifications.

A number of significant questions in regard to residential choice and residential segregation remain unanswered. What finally are the determinants of residential choice? Which factors are more important than others? Finding the answers to these broad questions was not part of the mandate of this study. However, the model framework presented here provides a promising avenue for future examination of these important issues.

Bibliography

Acevedo-Garcia, D., Lochner, K. A., Osypuk, T. L., & Subramanian, S. V. (2003). Future Directions in Residential Segregation and Health Research: A Multilevel Approach. *American Journal of Public Health, 93*(2), 215-221.

Agyei-Mensah, S., & Owusu, G. (2010). Segregated by Neighbourhoods? A Portrait of Ethnic Diversity in the Neighbourhoods of the Accra Metropolitan Area, Ghana. *Population, Space and Place, 16*, 499-516.

Alexander, T., Davern, M., & Stevenson, B. (2010). Inaccurate Age and Sex Data in the Census PUMS Files: Evidence and Implications. *National Bureau of Economic Research Working Paper No. 15703*.

Allan, J.-a. (2001). Review of the Measurment of Ethnicity: Classification and Issues. Statistics New Zealand.

Allen, P. M. (1997). *Cities and Regions as Self-Organizing Systems: Models of Complexity*. Amsterdam: Gordon and Breach Science Publishers.

Anselin, L. (1995). Local Indicators of Spatial Association—LISA. *Geographical Analysis, 27*, 93-115.

Anselin, L. (2005). *Exploring Spatial Data with GeoDa: A Workbook*: Spatial Analysis Laboratory, University of Illinois, Urbana-Champaign.

Anselin, L., Cohen, J., Cook, D., Gorr, W., & Tita, G. (2000). Spatial Analyses of Crime. In D. Duffee (Ed.), *Criminal Justice: Measurement and Analysis of Crime and Justice* (Vol. 4, pp. 213-262). Washington DC: National Institute of Justice.

Axtell, R. L., Epstein, J. M., Dean, J. S., Gumerman, G. J., Swedlund, A. C., Harburger, J., et al. (2002). Population Growth and Collapse in a Multiagent Model of the Kayenta Anasazi in Long House Valley. *PNAS, 99*(3), 7275-7279.

Aydinonat, N. E. (2007). Models, Conjectures and Exploration: An Analysis of Schelling's Checkerboard Model of Residential Segregation. *Journal of Economic Methodology, 14*(4), 429-454.

Balci, O. (1994). Validation, Verification, and Testing Techniques Throughout the Life Cycle of a Simulation Study. *Annals of Operations Research, 53*(1), 121-113.

Ballas, D., Clarke, G., Dorling, D., Eyre, H., Thomas, B., & Rossiter, D. (2005). SimBritain: A Spatial Microsimulation Approach to Population Dynamics. *Population, Space and Place, 11*, 13-34.

Ballas, D., & Clarke, G. P. (2001). Modelling the Local Impacts of National Social Policies: A Spatial Microsimulation Approach. *Environment and Planning C: Government and Policy, 19*(4), 587-606.

Banks, M. (1996). *Ethnicity: Anthropological Constructions*. London: Routledge.

Batty, M. (2005). Agents, Cells, and Cities: New Representational Models for Simulating Multiscale Urban Dynamics. *Environment and Planning A, 37*(8), 1373-1394.

Batty, M. (2008a). Fifty Years of Urban Modeling: Macro-Static to Micro-Dynamics. In S. Albeverio, D. Aundrey, P. Giordano & A. Vancheri (Eds.), *The Dynamics of Complex Urban Systems: An Interdisciplinary Approach* (pp. 1-20). Heidelberg: Physica-Verlag.

Batty, M. (2008b). The Size, Scale, and Shape of Cities. *Science, 319*, 769-771.

Batty, M., & Longley, P. (1994). *Fractal Cities: A Geometry of Form and Function*. London: Academic Press.

Batty, M., & Torrens, P. M. (2005). Modelling and Prediction in a Complex World. *Futures, 37*(7), 745-766.

Batty, M., & Xie, Y. (1994a). Modelling Inside GIS: Part 1. Model Structures, Exploratory Spatial Data Analysis and Aggregation. *International Journal of Geographical Information Science, 8*(3), 291-307.

Batty, M., & Xie, Y. (1994b). Modelling Inside GIS: Part 2. Selecting and Calibrating Urban Models Using ARC-INFO. *International Journal of Geographical Information Science, 8*(5), 451-470.

Becker, G. S. (1957). *The Economics of Discrimination*. Chicago: University of Chicago Press.

Bedford, R., & Ho, E. (2008). *Asians in New Zealand: Implications of a Changing Demography*. Outlook Series Paper 07. Wellington: Asia New Zealand Foundation.

Bell, W. (1954). A Probability Model for the Measurement of Ecological Segregation. *Social Forces, 32*(4), 357-364.

Belsky, E. S. (1992). Rental Vacancy Rates: A Policy Primer. *Housing Policy Debate, 3*(3), 793-813.

Benenson, I. (1999). Modeling Population Dynamics in the City: From a Regional to a Multi-Agent Approach. *Discrete Dynamics in Nature and Society, 3*, 149 -170.

Benenson, I., Omer, I., & Hatna, E. (2002). Entity-based Modeling of Urban Residential Dynamics: The Case of Yaffo, Tel Aviv. *Environment and Planning B: Planning and Design, 29*(4), 491-512

Benenson, I., & Torrens, P. M. (2004). *Geosimulation: Automat-based Modeling of Urban Phenomena*. England: Wiley.

Birkin, M. (2008). Hybrid Geographical Models of Urban Spatial Structure and Behaviour. In S. Albeverio, D. Aundrey, P. Giordano & A. Vancheri (Eds.), *The Dynamics of Complex Urban Systems: An Interdisciplinary Approach* (pp. 95-109). Heidelberg: Physica-Verlag.

Birkin, M., Clarke, G. P., & Clarke, M. (1996). Urban and Regional Modelling at the Microscale. In G. P. Clarke (Ed.), *Microsimulation for Urban and Regional Policy Analysis* (pp. 10-27). London: Pion.

Bobo, L., & Zubrinsky, C. L. (1996). Attitudes on Residential Integration: Perceived Status Difference, Mere In-group Preference, or Racial Prejudice. *Social Forces, 74*, 883-909.

Boman, M., & Holm, E. (2004). Multi-Agent Systems, Time Geography, and Microsimulations. In M.-O. Olsson & G. Sjöstedt (Eds.), *Systems Approaches and Their Application*.

Bonabeau, E. (2002). Agent-based Modeling: Methods and Techniques for Simulating Human Systems. *Proceedings of the National Academy of Sciences, 99*(3), 7280–7287.

Box, G. E. P. (1976). Science and Statistics. *Journal of American Statistical Association, 71*(356), 791-799.

Brade, D. (2003). *A Gerneralized Process for the Verification and Validation of Models and Simulation Results.* Unpublished Ph.D. dissertation, Universität der Bundeswehr München.

Brown, D. G., Page, S., Riolo, R., Zellner, M., & Rand, W. (2005). Path Dependence and the Validation of Agent-Based Spatial Models of Land Use. *International Journal of Geographical Information Science, 19*(2), 153 - 174.

Brown, L. A., & Chung, S.-Y. (2006). Spatial Segregation, Segregation Indices and the Geographical Perspective. *Population, Space and Place, 12*, 125-143.

Bruch, E. E., & Mare, R. D. (2006). Neighborhood Choice and Neighborhood Change. *American Journal of Sociology, 112*(3), 667.

Bruch, E. E., & Mare, R. D. (2009). Preferences and Pathways to Segregation: Reply to Van de Rijt, Siegel, and Macy. *American Journal of Sociology, 114*(4), 1181-1198.

Burgess, E. W. (2002). The Growth of the City. In G. Bridge & S. Watson (Eds.), *The Blackwell City Reader* (pp. 244-250). Oxford: Blackwell.

Caldwell, S. B., Clarke, G. P., & Keister, L. A. (1998). Modelling Regional Changes in US Household Income and Wealth: A Research Agenda. *Environment and Planning C: Government and Policy, 16*(6), 707-722.

Carley, K. M. (1996). *Validating Computational Models*. Working paper: Carnegie Mellon University.

Castle, C. J. E., & Crooks, A. T. (2006). *Principles and Concepts of Agent-Based Modelling for Developing Geospatial Simulations*. Working Papers Series (paper 110): UCL Centre for Advanced Spatial Analysis.

Cater, J., & Jones, T. (1989). *Social Geography: An Introduction to Contemporary Issues*. London: Edward Arnold.

Charles, C. Z. (2003). The Dynamics of Racial Residential Segregation. *Annual Review of Sociology, 29*(1), 167-207.

Clark, W. A. V. (1986). Residential Segregation in American Cities: A Review and Interpretation. *Population Research and Policy Review, 5*, 95-117.

Clark, W. A. V. (1987). Urban Restructuring from a Demographic Perspective. *Economic Geography, 63*, 103-125.

Clark, W. A. V. (1991). Residential Preferences and Neighborhood Racial Segregation: A Test of the Schelling Segregation Model. *Demography, 28*(1), 1-9.

Clark, W. A. V. (1992). Residential Preferences and Residential Choices in a Multiethnic Context. *Demography, 29*(3), 451.

Clark, W. A. V. (2006). Ethnic Preference and Residential Segregation: A Commentary on Outcomes from Agent-Based Modeling. *Journal of Mathematical Sociology, 30*, 319-326.

Cornforth, D., Green, D. G., & Newth, D. (2005). Ordered Asynchronous Processes in Multi-agent Systems. *Physica D: Nonlinear Phenomena, 204*(1-2), 70-82.

Courgeau, D. (2003). From the Macro-Micro Opposition to Multilevel Analysis in Demography. In D. Courgeau (Ed.), *Methodology and Epistemology of Multilevel Analysis: Approaches from Different Social Sciences* (Vol. 2, pp. 43-84).

Crooks, A., Castle, C., & Batty, M. (2008). Key Challenges in Agent-based Modelling for Geo-spatial Simulatation. *Computers, Environment and Urban Systems, 32*, 417-430.

Daude, E., & Langlois, P. (2007). Comparison of Three Implementations of Schelling's Spatial Segregation Model. In D. Phan & F. Amblard (Eds.), *Agent-based Modelling and Simulation in the Social and Human Sciences* (pp. 295-326). Oxford: Bardwell Press.

David, N. (2009). Validation and Verification in Social Simulation: Patterns and Clarification of Terminology. In *Epistemological Aspects of Computer Simulation in the Social Sciences* (pp. 117-129). Berlin: Springer-Verlag.

Davies, G., & Fagan, J. (2002). *Does Differential Social Organization in Immigrant Neighborhoods Reduce Risks of Violent Crime? Evidence from Panel Data in New York City*. Paper presented at the the annual meeting of the American Society of Criminology (ASC).

Dessalles, J. L., & Phan, D. (2005). Emergence in Multi-Agent Systems: Cognitive Hierarchy, Detection, and Complexity Reduction. In P. Mathieu, B. Beaufils & O. Brandouy (Eds.), *Artificial Economics: Agent-Based Methods in Finance, Game Theory and Their Applications* (pp. 147-159): Springer.

Editorial. (2004). Geosimulation: Object-based Modeling of Urban Phenomena. *Computers, Environment and Urban Systems, 28*, 1-8.

Edmonds, B., & Moss, S. (2004). From KISS to KIDS – An 'Anti-Simplistic' Modelling Approach. In P. Davidsson, B. Logan & K. Takadama (Eds.), *Multi-Agent and Multi-Agent-Based Simulation* (pp. 130–144): Springer.

Ferguson, N. (2008). *The Ascent of Money: A Financial History of the World*: Penguin Press.

Fleming, D. K. (1989). Identification of the Shipping District in New York, Houston and Seattle: 1956 and 1987. *Geoforum, 20*(4), 469-485.

Fossett, M. (2006a). Ethnic Preference, Social Distance Dynamics, and Residential Segregation: Theoretical Explorations Using Simulation Analysis. *Journal of Mathematical Sociology, 30*, 185-274.

Fossett, M. (2006b). Including Preference and Social Distance Dynamics in Multi-Factor Theories of Segregation. *Journal of Mathematical Sociology, 30*, 289-298.

Fossett, M., & Waren, W. (2005). Overlooked Implications of Ethnic Preferences for Residential Segregation in Agent-Based Models. *Urban Studies, 42*(11), 1893-1917.

Friesen, W. (2000). Tangata Pasifika Aotearoa: Pacific Populations and Identity in New Zealand. *New Zealand Population Review, 26*(2), 105-125.

Friesen, W. (2008). *Diverse Auckland: The face of New Zealand in the 21st Century?* Outlook Series Paper 06. Wellington: Asia New Zealand Foundation.

Galster, G. (1988). Residential Segregation in American Cities: A Contrary Review. *Population Research and Policy Review, 7*, 93-112.

Gibson, C. (1968). The Neighborhood Health Center: the Primary Unit of Health Care. *American Journal of Public Health, 57*(7), 1188-1191.

Gilbert, N. (2006). When Does Social Simulation Need Cognitive Models? In R. Sun (Ed.), *Cognition and Multi-Agent Interaction: From Cognitive Modeling to Social Simulation* (pp. 428-432). Cambridge Cambridge University Press.

Gilbert, N., & Troitzsch, K. G. (1999). *Simulation For the Social Scientist*: Open University Press.

Glaeser, E. L., Gyourko, J., & Saks, R. E. (2006). Urban Growth and Housing Supply. *Journal of Economic Geography, 6*(1), 71-89.

Goering, J. (2006). Shelling Redux: How Sociology Fails to Make Progress in Building and Empirically Testing Complex Causal Models Regarding Race and Residence. *Journal of Mathematical Sociology, 30*, 299-317.

Grimm, V., Berger, U., Bastiansen, F., Eliassen, S., Ginot, V., Giske, J., et al. (2006). A Standard Protocol for Describing Individual-based and Agent-based Models. *Ecological Modelling, 198*(1-2), 115-126.

Grimm, V., Berger, U., DeAngelis, D. L., Polhill, J. G., Giske, J., & Railsback, S. F. (2010). The ODD Protocol: A Review and First Update. *Ecological Modelling, 221*(23), 2760-2768.

Grimm, V., & Railsback, S. F. (2006). Agent-Based Models in Ecology: Patterns and Alternative Theories of Adaptive Behaviour. In F. C. Billari, T. Fent, A. Prskawetz & J. Scheffran (Eds.), *Agent-Based Computational Modelling: Applications in Demography, Social, Economic and Environmental Sciences* (pp. 139-152). Heidelberg: Physica Verlag.

Grimm, V., Revilla, E., Berger, U., Jeltsch, F., Mooij, W. M., Railsback, S. F., et al. (2005). Pattern-Oriented Modeling of Agent-Based Complex Systems: Lessons from Ecology. *Science, 310*, 987-991.

Gross, J. S. (2005). Business Improvement Districts in New York City's Low-Income and High-Income Neighborhoods. *Economic Development Quarterly, 19*(2), 174-189.

Haining, R. (1990). *Spatial Data Analysis in the Social and Environmental Sciences*. Cambridge: Cambridge University Press.

Happe, K., Kellermann, K., & Balmann, A. (2006). Agent-based Analysis of Agricultural Policies: An Illustration of the Agricultural Policy Simulator AgriPoliS, its Adaptation and Behavior. *Ecology and Society, 11*(1).

Hayden, D. (2003). *Building Suburbia: Green Fields and Urban Growth, 1820-2000*. New York: Pantheon Books.

Herrigel, G. (2000). Large Firms and Industrial Districts in Europe: Deregionalization, Re-regionalization, and the Trasformation of Manufacturing Flexibility. In J. H.

Dunning (Ed.), *Regions, Globalization, and the Knowledge-Based Economy* (pp. 286-302). Oxford: Oxford University Press.

Howe, T. R., Collier, N. T., North, M. J., Parker, M. T., & Vos, J. R. (2006). *Containings Agents: Contexts, Projections, and Agents.* Paper presented at the Proceedings of the Agent 2006 Conference on Social Agents: Results and Prospects.

Hudson, J. R. (1988). Pioneer Artists and Middle Class Settlers: Lifestyle Variables and Neighborhood Succession. *Journal of Family and Economic Issues, 9*(3), 250-263.

Hutchinson, J., & Smith, A. D. (Eds.). (1996). *Ethnicity*: Oxford University Press.

Johnston, R., Poulsen, M., & Forrest, J. (2004). Ethnic Residential Segregation in New Zealand: A Comparative Approach. *British Review of New Zealand Studies 14*, 7-36.

Johnston, R., Poulsen, M., & Forrest, J. (2009). Evaluating Changing Residential Segregation in Auckland, New Zealand, Using Spatial Statistics. *Tijdschrift voor Economische en Sociale Geografie, 102*(1), 1-23.

Kaplan, A. (1998). *The Conduct of Inquiry: Methodology for Behavioral Science*: Transaction Publishers.

Knopp, L. (1997). Gentrification and Gay Neighborhood Formation in New Orleans: A Case Study. In A. Gluckman & B. Reed (Eds.), *Homo Dconomics: Capitalism, Community, and Lesbian and Gay Life* (pp. 45-64). New York: Routledge.

Knutti, R. (2008). Should We Believe Model Predictions of Future Climate Change? *Phil. Trans. R. Soc. A, 366*(1885), 4647–4664.

Kotze, N. J., & Van Der Merwe, I. J. (2000). Neighbourhood Renewal in Cape Town's Inner City: Is It gentrification? *Journal of Family Ecology and Consumer Sciences, 28*, 40-46.

Krieger, N. (1994). Epidemiology and the Web of Causation: Has Anyone Seen the Spider? *Social Science & Medicine, 39*(7), 887-903.

Krink, T. (2004). Complexity and the Computing Age: Can Computers Help Us to Undrestand Complex Phenomena in Nature? In K. v. K. Niekerk & H. Buhl (Eds.), *The Significance of Complexity: Approaching a Complex World Through Science, Theology, and the Humanities*. Aldershot: Ashgate.

Laurence, R. (1994). *Roman Pompeii: Space and Society*. New York: Routledge.

Laurie, A. J., & Jaggi, N. K. (2003). Role of "Vision" in Neighbourhood Racial Segregation: A Variant of the Schelling Segregation Model. *Urban Studies, 40*(13), 2687-2704.

Law, A. M. (2009). *How to Build Valid and Credible Simulation Models.* Paper presented at the Proceedings of the 2009 Winter Simulation Conference.

Lee, S., & Leigh, N. G. (2005). The Role of Inner Ring Suburbs in Metropolitan Smart Growth Strategies. *Journal of Planning Literature, 19*(3), 330-346.

Lefebvre, B. (2004). Justice Dept. Accuses Detroit Bank of Bias in Lending. *The New York Times*, (20 May 2004).

Leombruni, R., & Richiardi, M. (2005). Why Are Economists Sceptical about Agent-based Simulations? *Physica A, 355*(1), 103–109.

Li, W. (2006). Introduction: Asian Immigration and Community in the Pacific Rim. In W. Li (Ed.), *From Urban Enclave to Ethnic Suburb: New Asian Communities in Pacific Rim Countries*. Honolulu: University of Hawai'i Press.

Lieberson, S. (1980). *A Piece of the Pie: Blacks and White Immigrants Since 1880*. Berkeley: University of California Press.

Lieberson, S. (1981). An Asymmetrical Approach to Segregation. In C. Peach, V. Robinson & S. Smith (Eds.), *Ethnic Segregation in Cities* (pp. 61-82). London: Croom Helm.

Ligmann-Zielinska, A., & Sun, L. (2010). Applying Time-dependent Variance-based Global Sensitivity Analysis to Represent the Dynamics of an Agent-based Model of Land Use Change. *International Journal of Geographical Information Science, 24*(12), 1829-1850.

Logan, J. R., & Zhang, W. (2004). Identifying Ethnic Neighbourhoods with Census Data. In M. F. Goodchild & D. G. Janelle (Eds.), *Spatially Integrated Social Science* (pp. 113-126): Oxford.

Logan, J. R., Zhang, W., & Alba, R. D. (2002). Immigrant Enclaves and Ethnic Communities in New York and Los Angeles. *American Sociological Review, 67*(2), 299-322.

Louie, M., & Carley, K. (2008). Balancing the Criticisms: Validating Multi-Agent Models of Social Systems. *Simulation Modeling Practice and Theory (special issue on Simulating Organizational Processes)*, 42-105.

Lowell, K. E. (2007). At What Level Will Decision-Makers Be Able to Use Uncertainty Information? In L. Oxley & D. Kulasiri (Eds.), *MODSIM07 International Congress on Modelling and Simulation* (pp. 2666-2672).

Mahdavi, B., O'Sullivan, D., & Davis, P. (2007). An Agent-based Microsimulation Framework for Investigating Residential Segregation Using Census Data. In L. Oxley & D. Kulasiri (Eds.), *MODSIM07 International Congress on Modelling and Simulation* (pp. 365-371).

Mahdavi, B., O'Sullivan, D., & Davis, P. (2009). *Operationalisation of Micro-Macro in an Empirically Informed Agent-based Microsimulation Model for Investigating the Emergence of Segregation using Constrained Census Data.* Paper presented at the International Conference on Emergence in Geographical Space: Concepts, Methods and Models, Paris, France.

Massey, D. S., & Denton, N. (1989). Hypersegregation in U.S. Metropolitan Areas: Black and Hispanic Segregation Along Five Dimensions. *Demography, 26*(2), 373-391.

Massey, D. S., & Denton, N. A. (1988). The Dimensions of Residential Segregation. *Social Forces, 67*(2), 281.

Merz, J. (1991). Microsimulation - A Survey of Principles, Developments and Applications. *International Journal of Forecasting, 7*, 77-104.

Mickelson, R. A. (2006). Segregation and the SAT. *Ohio State Law Journal, 67*, 157-199.

Milano, K. W. (2008). *Remembering Kensington & Fishtown: Philadelphia's Riverward Neighborhoods*. Charelston: History Press.

Miller, J. H., & Page, S. E. (2007). *Complex Adaptive Systems: An Introduction to Computational Models of Social Life*. Princeton: Princeton University Press.

Moore, A. (2000). Geospatial Expert Systems. In S. Openshaw & R. J. Abrahart (Eds.), *GeoComputation* (pp. 127-159). London: Taylor & Francis.

Moore, A. (2011). Geographical Vector Agent-based Simulation for Agricultural Land-use Modelling. In J. Marceau & I. Benenson (Eds.), *Advanced Geosimulation Models* (pp. 30-48). Sharjah: Bentham Science.

Moss, S. (2002). Policy Analysis from First Principles. *PNAS, 99*(3), 7267-7274.

Moss, S. (2008). Alternative Approaches to the Empirical Validation of Agent-Based Models. *Journal of Artificial Societies and Social Simulation, 11*(1).

Moss, S., & Edmonds, B. (2005). Towards Good Social Science. *Journal of Artificial Societies and Social Simulation, 8*(4).

Muth, R. (1969). *Cities and Housing: The Spatial Pattern of Urban Residential Land Use*. Chicago: University of Chicago Press.

Nathans, H. S. (2003). *Early American Theatre from the Revolution to Thomas Jefferson: into the Hands of the People*. Cambridge: Cambridge University Press.

O'Sullivan, D. (2001). Graph-Cellular Automata: A Generalised Discrete Urban and Regional Model. *Environment and Planning B: Planning and Design, 28*, 687-705.

O'Sullivan, D. (2004). Too Much of the Wrong Kind of Data: Implications for the Practice of Micro-Scale Spatial Modelling. In M. F. Goodchild & D. Janelle (Eds.), *Spatially Integrated Social Science* (pp. 95-107): Oxford University Press.

O'Sullivan, D. (2009). Changing Neighborhoods–Neighborhoods Changing: A Framework for Spatially Explicit Agent-Based Models of Social Systems. *Sociological Methods & Research, 37*(4), 498-530.

O'Sullivan, D., Macgill, J., & Yu, C. (2003). *Agent-based Residential Segregation: A Hierarchically Structured Spatial Model.* Paper presented at the Agent 2003 Conference on Challenges in Social Simulation, Chicago, IL.

Openshaw, S. (1983). *The Modifiable Areal Unit Problem*: GeoBooks.

Openshaw, S., & Taylor, P. J. (1979). A Million or so Correlation Coefficients: Three Experiments on the Modifiable Areal Unit Problem. In N. Wrigley (Ed.), *Statistical Methods in the Spatial Sciences* (pp. 127-144). London: Pion.

Openshaw, S., & Taylor, P. J. (1981). The Modifiable Areal Unit Problem. In N. Wrigley & R. J. Bennett (Eds.), *Quantitative Geography* (pp. 60-74): Routledge & Kegan Paul.

Oreskes, N., Shrader-Frechette, K., & Belitz, K. (1994). Verification, Validation, and Confirmation of Numberical Models in the Earth Sciences. *Science, 263*(5147), 641-646.

Pancs, R., & Vriend, N. J. (2007). Schelling's Spatial Proximity Model of Segregation Revisited. *Journal of Public Economics, 91*, 1-24.

Pawson, E., Bedford, R., Palmer, E., Stokes, E., Friesen, W., Cocklin, C., et al. (1996). Senses of Place. In R. L. Heron & E. Pawson (Eds.), *Changing Places: New Zealand in the Nineties*: Longman Paul.

Pearce, D. (1979). Gatekeepers and Homeseekers: Institutionalized Patterns of Ratial Steering. *Social Problems, 26*, 325-342.

Petrescu-Prahova, M. (2009). *Beyond Schelling: The Role of Attitudes toward In-group and Out-group Members in Ethnic Residential Settlement and Segregation*. Working Paper no. 91, Center for Statistics and the Social Sciences, University of Washington.

Portugali, J. (2000). *Self-Organization and the City*: Springer.

Reardon, S. F. (2006). A Conceptual Framework for Measuring Segregation and Its Association with Populaion Outcomes. In J. M. Oakes & J. S. Kaufman (Eds.), *Methods in Social Epidemiology* (pp. 169-192). San Francisco: Jossey-Bass.

Reardon, S. F., Farrell, C. R., Matthews, S. A., O'Sullivan, D., Bischoff, K., & Firebaugh, G. (2009). Race and Space in the 1990s: Changes in the Geographic Scale of Racial Residential Segregation, 1990-2000. *Social Science Research, 38*(1), 55-70.

Reardon, S. F., & Firebaugh, G. (2002). Measures of Multigroup Segregation. *Sociological Methodology, 32*, 33-67.

Reardon, S. F., Matthews, S. A., O'Sullivan, D., Lee, B. A., Firebaugh, G., Farrell, C. R., et al. (2008). The Geographical Scale of Metropolitan Racial Segregation. *Demography, 45*(3), 489-514.

Reardon, S. F., & O'Sullivan, D. (2004). Measures of Spatial Segregation. *Sociological Methodology, 34*(1), 121-162.

Rephann, T. J., & Holm, E. (2004). Economic-Demographic Effects of Immigration: Results from a Dynamic Spatial Microsimulation Model. *International Regional Science Review, 27*(4), 379-410.

Robinson, S. (1997). *Simulation Model Verification and Validation: Increasing the Users' Confidence.* Paper presented at the Proceedings of the 1997 Winter Simulation Conference.

Rosenblum, D. (1995). Overcoming Stigmas: Lesbian and Gay Districts and Black Electoral Empowerment. *Howard Law Journal, 39*(1), 149-200.

Sakoda, J. M. (1971). The Checkerboard Model of Social Interaction. *Journal of Mathematical Sociology, 1*, 119-132.

Sander, R., Schreiber, D., & Doherty, J. (2000). Empirically Testing a Computational Model: The Example of Housing Segregation. In D. Sallach & T. Wolsko (Eds.), *Proceedings of the Workshop on Simulation of Social Agents: Architectures and Institutions* (pp. 108-116). Chicago: Argonne National Laboratory.

Sanders, L. (2007). Agent Models in Urban Geogrpahy. In D. Phan & F. Amblard (Eds.), *Agent-based Modelling and Simulation in the Social and Human Sciences* (pp. 147-167). Oxford: Bardwell Press.

Sargent, R. G. (2008). *Verification and Validation of Simulation Models.* Paper presented at the Proceedings of the 2008 Winter Simulation Conference.

Sarre, P., Phillips, D., & Skellington, R. (1989). *Ethnic Minority Housing: Explanations and Policies*: Avebury.

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

Schelling, T. C. (1971a). Dynamic Models of Segregation. *Journal of Mathematical Sociology, 1*, 143-186.

Schelling, T. C. (1971b). On the Ecology of Micromotives. *The Public Interest, 25*, 61-98.

Schelling, T. C. (1972). A Process of Residential Segregation: Neighborhood Tipping. In A. H. Pascal (Ed.), *Racial Discrimination in Economic Life*: Lexington Books.

Schelling, T. C. (1987). *Micromotives and Macrobehavior*. New York: Norton.

Sills, D. L. (Ed.). (1997). *International Encyclopedia of the Social Sciences* (Vol. 13). New York: Macmillan.

Skvoretz, J. (2006). Introduction to Special Edition on "Ethnic Preferences, Social Distance Dynamics, and Residential Segregation: Theoretical Explanations Using Simulation Analysis" by Mark Fossett. *The Journal of Mathematical Sociology, 30*(3), 181 - 184.

Smith, D. P. (2005). 'Studentification': The Gentrification Factory? In R. Atkinson & G. Bridge (Eds.), *Gentrification in a Global Context: the New Urban Colonialism* (pp. 72-89). New York: Routledge.

Statistics New Zealand. (2004). *Report of the Review of the Measurement of Ethnicity*.

Statistics New Zealand. (2005). *Statistical Standard for Ethnicity*.

Statistics New Zealand. (2006a). *QuickStats About Culture and Identity: 2006 Census*.

Statistics New Zealand. (2006b). *QuickStats About New Zealand's Population and Dwellings: 2006 Census*.

Statistics New Zealand. (2006c). *QuickStats About Pacific Peoples: 2006 Census*.

Statistics New Zealand. (2007a). *QuickStats About Maori: 2006 Census*.

Statistics New Zealand. (2007b). *Survey of Dynamics and Motivation for Migration in New Zealand: March 2007 Quarter*.

Sterman, J. D., Jr., E. J. R., Oreskes, N., Belitz, K., & Shrader-Frechette, K. (1994). The Meaning of Models. *Science, 264*.

Sugrue, T. J. (1996). *The Origins of the Urban Crisis : Race and Inequality in Postwar Detroit*: Princeton University Press.

Tomes, N. (1978). A "Torrent of Abuse": Crimes of Violence between Working-Class Men and Women in London, 1840-1875. *Journal of Social History, 11*(3), 328-345.

Vasil, R., & Yoon, H.-k. (1996). *New Zealanders of Asian Origin*. Institute of Policy Studies: Victoria University of Wellington.

Vencatasawmy, C. P., Holm, E., Rephann, T., Esko, J., Swan, N., Öhman, M., et al. (1999, September 3-7 1999). *Building a Spatial Microsimulation Model.* Paper presented at the 11th European Colloquium on Quantitative and Theoretical Geography, Durham, England.

Wedam, E. (2003). The 'Religious District' of Elite Congregations: Reproducing Spatial Centrality and Redefining Mission *Sociology of Religion, 64*(1), 47-64.

Werker, C., & Brenner, T. (2004). *Empirical Calibration of Simulation Models*. Papers on Economics and Evolution # 0410. Jena: Max Planck Institute for Research into Economic Systems.

White, M. J. (1983). The Measurement of Spatial Segregation. *The American Journal of Sociology, 88*(5), 1008-1018.

Wilson, F. D., & Hammer, R. B. (2001). Ethnic Residential Segregation and Its Consequences. In A. O'Connor, C. Tilly & L. Bobo (Eds.), *Urban Inequality: Evidence from Four Cities* (pp. 272-303). New York: Russell Sage Foundation.

Windrum, P., Fagiolo, G., & Moneta, A. (2007). Empirical Validation of Agent-Based Models: Alternatives and Prospects. *Journal of Artificial Societies and Social Simulation, 10*(2).

Wolf, T. D., & Holvoet, T. (1998). Emergence Versus Self- rganisation: Different Concepts but Promising When Combined. In S. A. Brueckner, G. D. M. Serugendo, A. Karageorgos & R. Nagpal (Eds.), *Engineering Self-Organising Systems: Methodologies and Applications*. Berlin: Springer-Verlag.

Wong, D. W. S. (1997). Spatial Dependency of Segregation Indices. *Canadian Geographer, 41*(2), 128–136.

Wooldridge, M., & Jennings, N. R. (1995). Intelligent Agents: Theory and Practice. *The Knowledge Engineering Review, 10*(2), 115-152.

Wu, B. M., Birkin, M. H., & Rees, P. H. (2008). A Spatial Microsimulation Model with Student Agents. *Computers, Environment and Urban Systems, 32*, 440-453.

Wu, F. (1999). GIS-based Simulation as an Exploratory Analysis for Space-time Processes. *Journal of Geographical Systems, 1*(3), 199-218.

Zhang, J. (2004a). A Dynamic Model of Residential Segregation. *Journal of Mathematical Sociology, 28*(3), 147-170.

Zhang, J. (2004b). Residential Segregation in an All-Integrationist World. *Journal of Economic Behavior and Organization, 54*, 533-550.

Zhang, J. (2009). Tipping and Residential Segregation: A Unified Schelling Model. *forthcoming in Journal of Regional Science*.

Appendices

Appendix A: Nomenclature

..C	Confidential value
ABM	Agent-Based Model/ Modelling
ABS	Agent-Based Simulation/ Systems
AAMoS	Aggregated Agent-based Microsimulation of Segregation
AI	Artificial Intelligence
AKL	Auckland territorial authority/ district
AMoS	Agent-based Microsimulation of Segregation
AU	Area Unit, *(see also CAU)*
BaU	Behaviour as Usual
CA	Cellular Automaton/ Automata
CAS	Complex Adaptive System
CAU	Census Area Unit, *(see also AU)*
CTA	Census TA; interchangeably used by TA or district *(see also TA)*
COMPASS	Centre of Methods and Policy Application in the Social Sciences
CSV	Comma Separated Value
CAT	Census Territorial Authority, *(see also TA)*
D	Dissimilarity index
ESDA	Exploratory Spatial Data Analysis
Euro	Europeans (ethnicity)
GAL	GenePix Array List
GIS	Geographic/ Geographical/ Geospatial Information System(s)
GMI	Global Moran's I
GUI	Graphical User Interface
H	Information theory index
HAAMoS	Hybrid Aggregate Agent-based Microsimulation of Segregation
HES	Hybrid Environment Shapefile
HH/ hh	Household
II/ I.I	Index of Interaction/ Isolation
JDK	Java Development Kit
KIDS	Keep It Descriptive Stupid
KISS	Keep It Simple Stupid
LISA	Local Indicator of Spatial Associations (*see also LMI*)
LMI	Local Moran's I (*see also LISA*)
LQ	Location Quotient
LUTI	Land-Use Transport Interaction
MA	Metropolitan Area; interchangeably used by UA (*see also* UA)
Mao	Maoris (ethnicity)
MAS	Multi-Agent Simulation/ Systems
MAUP	Modifiable Areal Unit Problem

MKAU/ MKU	Manukau territorial authority/ district
MoSC	Modelling Social Change (project)
MSM	Microsimulaiton
NEI	Not Elsewhere Included
NRP	Noise Reduction Process
NShore/ NSH	North Shore territorial authority/ district
NZ	New Zealand
ODD	Overview, Design concepts, and Details
OOD	Object-Oriented Design
Pacific/ PAC	Pacific People/ Islanders (ethnicity)
PAPK/ PAK	Papakura territorial authority/ district
PI	Plurality Index
POC	Pattern-Oriented Calibration
POM	Pattern-Oriented Modelling
PPD	Pre-Package Data *(see also TBD)*
RR3	Random Rounding (Rule) to base 3
REPAST	Recursive Porous Agent Simulation Toolkit
SES	Socio-Economics
SNZ	Statistics New Zealand
SV	State Variable
TA	Territorial Authority; interchangeable with district *(see also CTA)*
TBD	Table Builder Data *(see also PPD)*
UA	Urban Area; interchangeable with MA *(see also* MA)
UML	Unified Modeling Language
VOI	Variable(s) Of Interest
WTKE/ WTK	Waitakere territorial authority/ district

Appendix B: Variable Notation

Variables and parameters can be set using the *HAAMoS* GUI panel or through the automated execution support, that is, the parameter file designed for batch mode run, as explained briefly in section 4.2.1, along with an example in appendix E. This is how the scenarios are enacted. Denotations, definitions and the range of variables/ parameters along with the measures and other functions are described in the following table (next page).

There is also a column that explains whether (and to what extent) these are empirically informed. The values for many parameters can be directly obtained from census (or other survey) data. For example, the number of area units and territorial authorities are precisely known. The values for many other parameters are 'indirectly' (often fairly accurately) obtained from census aggregate data. However, some values are not available at the micro-geographical scale. For instance, census data on unoccupied dwellings, which are used for setting the vacancy rates (percentages), are only available at the district (territorial authority) level. Therefore, at the area unit level, the applied vacancy rates are practically the census-based estimations. Other values, such the population growth percentage or immigration numbers can be roughly estimated from the census data.

In any case, the model can also be used like any other typical model (although here in both artificial and realistic environments) for exploration or experimentation using synthetic (no empirical) data.

Those variables considered 'state variable' (SV) are identified by ✔ in separate columns. State variables are those low-level variables describing elementary properties of the model's entities and cannot be deduced from any other low-level state variable (Grimm et al., 2006, p. 118).

This table can be used as a reference when quick information about variables, parameters or other notations found in equations or other types of expression (e.g. tables of scenario settings) are needed.

Table of Denotation, Definition and Range of Variables & Parameters

Denotation	S V	Definition	Value/ Range/ Scale	Empirically informed?
G	✓	Number of ethnic **g**roups; Maximum four: g_1/g_E :European (Blue); g_2/g_A: Asian (Red); g_3/g_P: Pacific (Green); g_4/g_M: Maori (Grey)	[1-4]	Census
h $\mathbf{h_g}$	✓	individual/ **h**ousehold agent; h_E/h_1 : European (g_1); h_A/h_2 : Asian (g_2); h_P/h_3 : Pacific (g_3); h_M/h_4 : Maori (g_4)	OBJ	Census (indirectly, from aggregate)
u $\mathbf{u_s}$	✓	Spatial **u**nit; **u** alone is an area unit home to many **h**. The exact location of a specific **h** in **u** in this model is unspecified (unknown) $\mathbf{u_s}$ is a spatial unit/ area such a territorial authority (TA) or entire MA s = AKL, NSH, WTK, MKU, PAK (for TA) or MA (for entire metropolitan area)	OBJ	***Realistic***: Census ***Grid-based***: restricted to cell based model
U		Total numbers of area units in the entire urban/ metropolitan area; It can also refer to the entire urban/ metropolitan area as one large spatial unit (= $\mathbf{u}_{MA}$)	4 – 316	Census
t	✓	Tick time of the simulation; $t=1 \equiv t_1$	0…k	---
$\mathbf{t_s}$	✓	Stop tick time of the simulation	0…k	Estimated (Census)
N(u)		a list of **n**eighbours of **u** excluding **u** itself, defined as $[n_i(u) \in N(u)$: i=1… k]) - **N(u)** set size and content depends on choice of N_{type} MR: MOORE V-N: Von-Neumann	$[n_i(r) \in N(r)$: i=1 … k] MR(rad=1): $k \in 3,5,8$; V-N (rad=1): $k \in 2,3,4$	***GIS-based*:** Census ***Grid-based*:** restricted to cell based topology
T (T_E, T_A, T_P, T_M) (T_1, T_2, T_3, T_4)	✓	- In-group/co-ethnic contact preference threshold (be surrounded by like/ similar-people) named **t**olerance; Minimal percentage (fraction) of in-group contact necessary in order to relocate to a new location **r;** its value can be different for each ethnic group. T-subscripts **E**, **A**, **P** and **M** refer to European (g_1), Asian (g_2), Pacific (g_3) and Maori (g_4)	0.0...1.0	Estimated

P-PERSIT$_g$	✓	A probability like value which indicates how persistent an (household) agent is in searching for a desirable location in the same area unit before looking further at another area unit (elsewhere). The value is compared with uniformly distributed random number in order to decide whether it is time to move further. All the members of the same ethnic group have the same P-PERSIT value. g = E, A, P, M/ 1, 2, 3, 4 (Euro, Asian, Pacific, Maori)	0.0…1.0	Imprecisely
L	✓	- Locational move preference of a moving agent; Two choices: 'global' in large geographical area (e.g. MA or TA) OR 'local' within immediate neighbourhood; It is assigned based on local-global probability for mobility **P(GL)** (although **P(M)** also indirectly plays a role). All new comers to the world by default have GLOBAL preference, although it is possible to change this in the model from entire MA to specific TAs	LOCAL, GLOBAL	Imprecisely
P(GL)$_g$	✓	- **P**robability of **g**lobal **l**ocational move preference for a given ethnic group; For simplicity, in chapters 5 and 7, **L** notation is used to refer to P(GL)$_g$ - Using a binomial distribution, **P(GL)$_g$** values are used to randomly calculate the number of agents willing to move globally. This would be a subset of migrant population size calculated by turnover process; g = E, A, P, M (Euro, Asian, Pacific, Maori)	0.0…1.0	Estimated
P(M)$_g$	✓	- **P**robability of **m**obility (turnover) for a given ethnic group; - Using a binomial distribution, **P(M)$_g$** will be used to randomly calculate the number of agents forced to relocate based on AU ethnic population size g = E, A, P, M/ 1, 2, 3, 4 (Euro, Asian, Pacific, Maori)	0.0…1.0	Estimated (from Census and SNZ survey)
isGATA	✓	'is Growth Agent TA' oriented is a boolean parameter to indicate whether the growth of population (enabled by creation of new agents) is TA oriented (or not): meaning these particular agents will only search for suitable locations within the TA for which the growth has been calculated	TRUE - FALSE	Implicitly
MIN$_V$	✓	- **Min**imum number of vacancies (used for initialization only at t=0) Can be used when the population size of AU is very small	0… Z^+	Estimated
V$_{pc}$ (t)	✓	TA-based vacancy percentage applicable to related spatial units at time **t**	0 … R^+	Census (TA level)
v(u, t)	✓	Number of **v**acancies in the micro area unit **u** and at time **t**	0 … Z^+	Census

v(t)		Number of **v**acancies in the meso/ macro spatial area at time **t**	0 … Z^+	Census
Y_0	✓	**I**nitial **y**ear (for initialization of the model at t=0, when census data is used)	1991, 1996, 2001, 2006	Census
φ_g (u, t)		Population of group **g** at time **t** at the given area unit **u** area unit (AU) is smallest micro- geographical area in the model	0 … Z^+	Census
φ(u, t)		[total] Population of the area unit **u** at time **t** area unit (AU) is smallest micro- geographical area in the model	0 … Z^+	Census
Φ_g (t)		[total] Population of group **g** in given meso/ macro spatial unit at time **t** Spatial unit: *Meso*: territorial authority (TA); *Macro*: metropolitan area (MA)	0 … Z^+	Census
Φ (t)		[total] Population (all groups) in given meso or at macro spatial unit at time **t**; i.e. $\sum \Phi_g(t)$, for g = 1 to G (total number of ethnic group simulated) Spatial unit: *Meso*: territorial authority (TA); *Macro*: metropolitan area (MA)	0 … Z^+	Census
π_g (u, t)		Proportion of group **g** in micro area unit **u** at time **t**	0 … 100	Census (@ t=0)
Π_g (t)		Proportion of group **g** in meso/ macro geographical area at time **t** Geographical area: *Meso*: territorial authority (TA); *Macro*: metropolitan area (MA) - In two groups setting, for example, **Π_A(t=0)** =10 (or , **Π_2(t=0)** =10) indicates the composition of group#2 (Asian) at initial stage (prior to start of simulation) is 10% and is the minority group (10/90); In general $\sum \Pi_i(t)$, i = 2 …G = 100	0 … 100	Census (@ t=0)
MX_F_i	✓	**Max**imum number of **in**flow (immigrants) of agents which are coming to the metropolitan area for the first time at each time step (simulation cycle)	0… Z^+	Census (Estimated)
MX_F_o	✓	**Max**imum number of **o**utflow agents (emigrants) which are leaving definitely the metropolitan area at each time step (simulation cycle)	0… Z^+	Estimated
$P(F_i)_g$	✓	- **P**robability of **in**flow of new agents; - Probability that of total **MX_F_i** , this specified percentage would randomly determine the exact number of each new settler group simulated in the model; The sum of the probabilities for each group must be equal to 1. For example, in two groups case, **$P(F_i)_{g1} + P(F_o)_{g2} = 1$**. In general $\sum$**$P(F_i)_g$** = 1, g = 2 …G	0.0…1.0	Census (Estimated)

$P(F_o)_g$	✓	- **P**robability of **o**ut**f**low agents (leaving MA) - Probability that of total **MX_F_o**, this specified percentage would randomly determine the exact number of each leaving ethnic group simulated in the model The sum of the probabilities for each group must be equal to 1. For example, in two groups case, **$P(F_o)_{g1} + P(F_o)_{g2} = 1$**. In general $\sum$**$P(F_o)_g$** = 1, g = 2 …G	0.0…1.0	Estimated
λ_{VISIT}	✓	Average number of AUs an agent will **visit** (inspect) before relocating; For GLOBAL movers this mean is used in a poisson distribution to randomly obtain the maximum number of visits; for LOCAL movers, it is set to maximum number of their immediate neighbours	0…U [GLOBAL], 0…N(u).size() [LOCAL]	Imprecisely
σ	✓	Standard deviation value influencing how accurately an agent can guess (estimate) the co-ethnic proportion of a neighbourhood for eventual relocation (currently the same value is used for all groups). Larger values would make the stochastic guessing less accurate than real value (i.e. makes the random guessing variation to oscillate further from the mean)	0.01…0.1	Imprecisely
RS	✓	**R**andom **n**umber generator **s**eed to replace default value. Using the same seed makes the reproducibility possible	Z^+	N/A
$D_{x,y}$ (t)		Measure: Dissimilarity index between two groups of **x, y** at time **t** at meso (TA) or macro (MA) level; $D_{x,y} = D_{y,x}$	0.0 – 1.0	N/A
H (t)		Measure: Information theory index at time **t** at meso (TA) or macro (MA) level	0.0 – 1.0	N/A
$I.I_{xx}$ (t)		Measure: Isolation index among group **x** member at time **t** at meso (TA) or macro (MA) level	0.0 – 1.0	N/A
$I.I_{xy}$ (t)		Measure: Exposure index between members of groups **x** and **y** at time **t** at meso (TA) or macro (MA) level; $I.I_{xy} \neq I.I_{yx}$	0.0 – 1.0	N/A
GMI_g (t)		Measure: (Global) Moran's I index for group **g** at time **t** at meso (TA) or macro (MA) level	-1.0 – 1.0	N/A
PI_g (t)		Measure: Plurality index where group **g** forms the plurality bloc at time **t** at meso (TA) or macro (MA) level	0 – 316	N/A

Acronym in the table header: **SV**: State Variable; the terms Household and Individual agents are used interchangeably

Appendix C: Prototypes

C.1 *AMoS*

The *AMoS* (Agent-based Microsimulation of Segregation) prototype (see figure C.1) is built on a cell-/ grid-based layout representing an artificial city ('world'). Based on 'continuous' (local) notion of neighbourhood – combined with 'bounded' (district) characteristics – it is to some extent similar to the Fossett's SimSeg model (Fossett, 2006a).

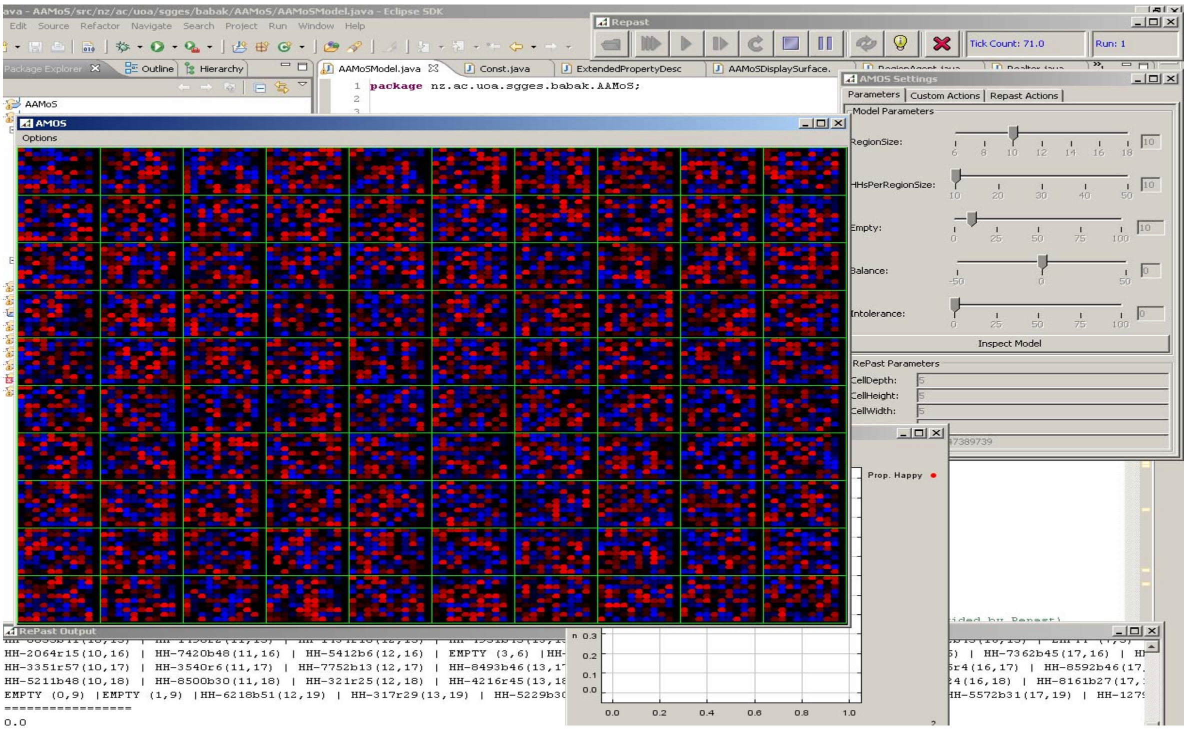


Figure C.1. Snapshot of *AMoS* Prototype

The concepts of bounded and continuous neighbourhoods have been explained in the literature (see O'Sullivan, 2009, for example). Figure C.2 illustrates the topology of the *AMoS* program and illuminates the concept of bounded and continuous neighbourhoods. Each of the small rectangles (presenting an area unit/ neighbourhood) identified by yellow lines (in figure C.1), is a bounded neighbourhood, depicted as larger rectangles in the top layer of figure C.2. Bounded neighbourhoods can be seen as a container/ block (or collection of locations) populated by a number of agents (households). Without this bounded neighbourhood, the simulated world represented by a single grid (lattice) would be a single neighbourhood with multiple continuous cells (locations), depicted in the lower layer of figure C.2 (comparable to the entire *AMoS* grid without those yellow rectangles). The majority of agent-based and cellular-based models use this type of neighbourhood (i.e. continuous).

In general, two (adjacent) locations are considered neighbours if they correspond to the model-specific definition of being neighbours. In a two-dimensional square lattice (rectangular grid) topology employed in cellular automata and most grid versions of agent-based models, the two most widely used are *Moore* and *von Neumann* neighbourhoods (see figure C.3).

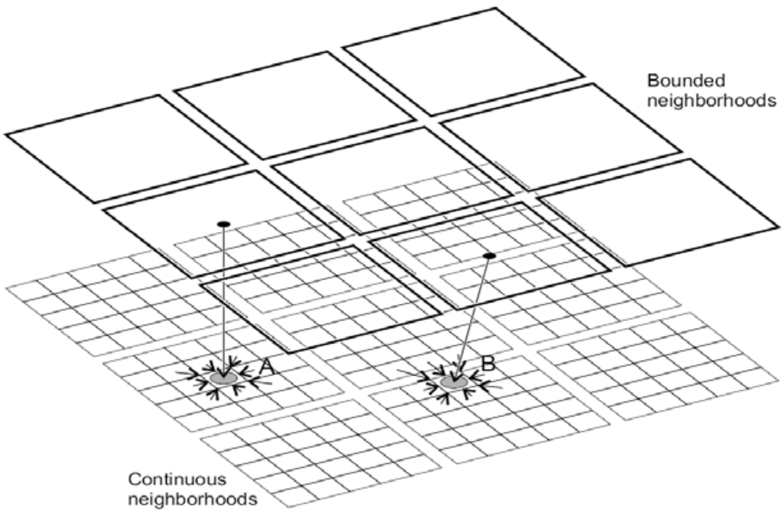


Figure C.2. Bounded and Continuous Neighbourhoods (form O'Sullivan et al., 2003)

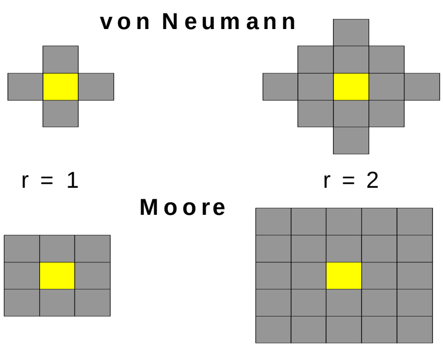


Figure C.3. Moore and von Neumann Neighbourhoods

The von Neumann neighbourhood comprises four cells surrounding a central cell (yellow), orthogonally. The Moore neighbourhood comprises eight cells surrounding a central cell, both orthogonally and diagonally. The Moore neighbourhood is the most used form of neighbourhood, particularly in its radius (**r**, or 'Chebyshev distance') equal to 1, sometimes referred to as 3×3.

In a continuous neighbourhood (where the grid world can be considered the only neighbourhood), local neighbourhoods (dynamically defined in an ad hoc manner based on a cell of interest, whether defined by Moore or von Newmann), will often overlap with each other. In a bounded neighbourhood, "locations are indirectly neighbours to one another via their common location inside a bounded neighbourhood", "whereas a location in a different neighbourhood [district] has a completely different set of neighbours" (O'Sullivan, 2009, p. 506).

The *AMoS* program and its layout could accommodate the entire census area units (CAU/ AU) of the Greater Auckland (about 370 of them) and the inhabited households within each of their boundaries. However, the 'instantiatation' of each household agent requires an allocation of computer memory and resources for about 400,000 estimated households (as alive/ active objects). This implied an exceedingly slow execution of the program (which started to manifest after the creation of only about 60,000 agents).

With limitations regarding the availability of census data at micro-granular scale as well as with typical desktop computer power, the need for developing a new prototype arose. The new prototype model could reflect the high level granularity of the available data whilst allowing the linkage between micro-level household (behaviours) and macro-level aggregate outcomes. The coarse-granular data could be accommodated by an 'encapsulated' entity of aggregate form of information on a large individual population. This was the rationale for the creation of an aggregated version of *AMoS* prototype (named *AAMoS*) which will be presented in the next section.

C.2 *AAMoS*

AAMoS (Aggregated Agent-based Microsimulation of Segregation) was developed based on the knowledge and experience gained from previous prototype (*AMoS*) and in response to the observed limitations (e.g. computational power and resources, data format and availability, etc.).

In this model individual/ household agents are created in an 'on-the-fly' fashion (i.e. created when needed) – at the time they decide to relocate to a new location (neighbourhood). Their individual lifetime ends (by freeing the allocated memory) once they settle in a new neighbourhood. In such way, they cannot be continuously traced out during the entire execution of the simulation program, except for the short time that they are alive. In other words, unlike the previous prototype model (*AMoS*), they are not entirely 'instantiated' at the beginning of the program, and do not occupy a uniquely defined space on the grid (at a specific neighbourhood). The rationale behind this approach is that such a lack of individual traceability should not be a barrier to the description of the aggregate behaviour of the collection (i.e., neighbourhood populations).

AAMoS can simulate up to four ethnic groups, so that their proportional sizes within each area unit alongside vacant spaces are painted in their respective colour (blue, red, green, and black, as well as grey for vacancies) for visual inspection, as can be seen in figure C.4.

Another feature of *AAMoS* (not yet enabled in its successor *HAAMoS*) is an elaborate mechanism to allow initialisation of simulation in multiple stochastic ways. This is done by specifying values for certain probability variables which could influence clustering, isolation and unevenness dimensions, as well as a drop-down list for selecting a 'direction' for the random effect throughout the 'world' (e.g. North, South, West, East, Northwest, Northeast, Southeast, Southwest, Centre). For instance, it would be possible to create some random clustering in the centre of the grid-based city.

AAMoS was a successful evolution in corresponding more closely to the needs and objectives of this study as well as a better adaption to the constraints imposed by data and computational resources/ power. In fact, in comparison with the previous prototype, *AAMoS* is a more fully-grown (developed) and fully functional model, sharing many of the characteristics and capabilities of its successor *HAAMoS*.

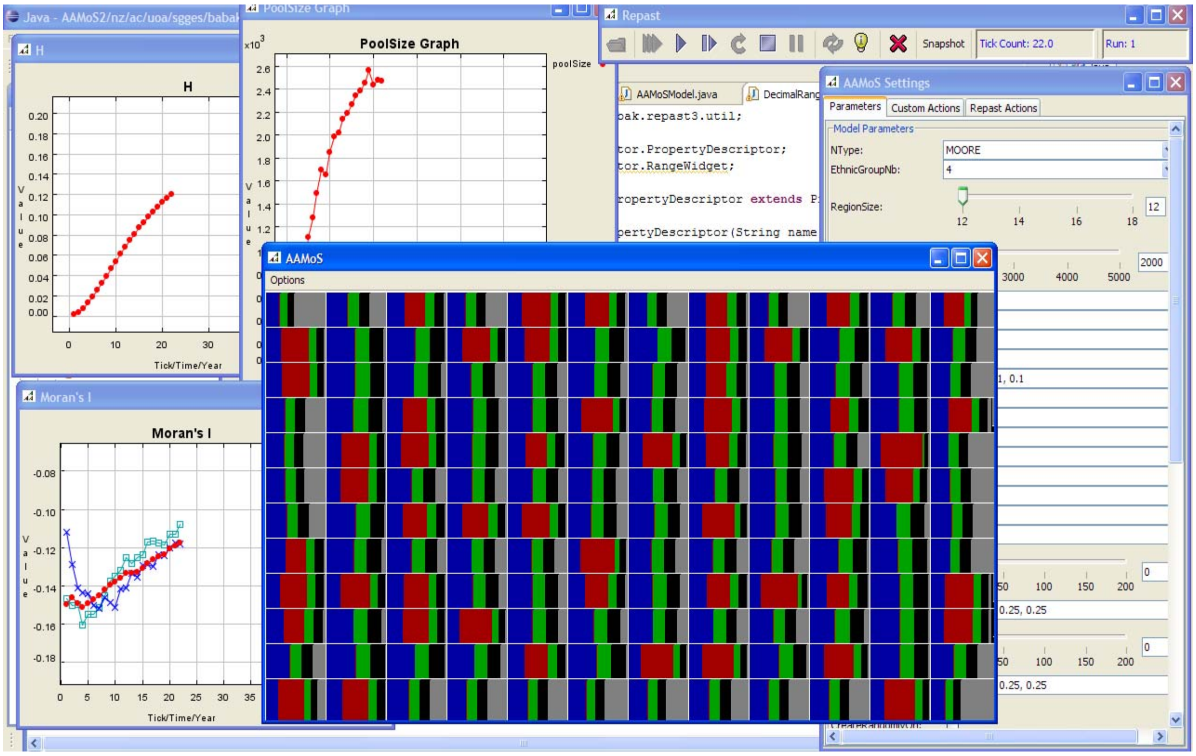
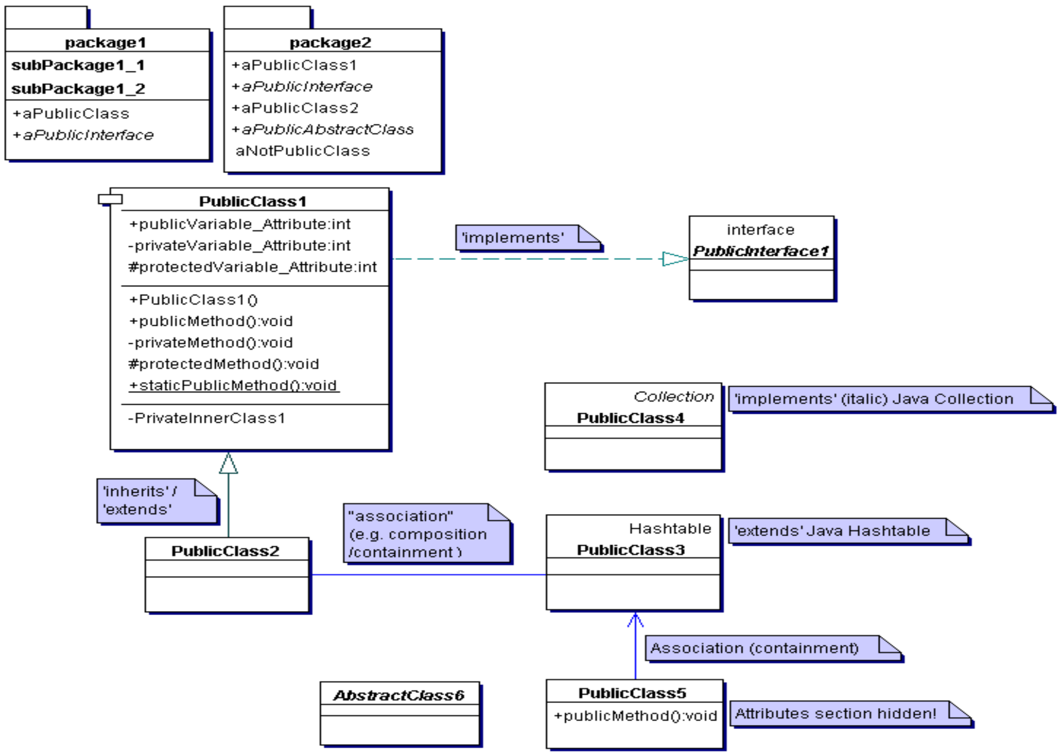


Figure C.4. Snapshot of *AAMoS* Program

Appendix D: UML Diagrams

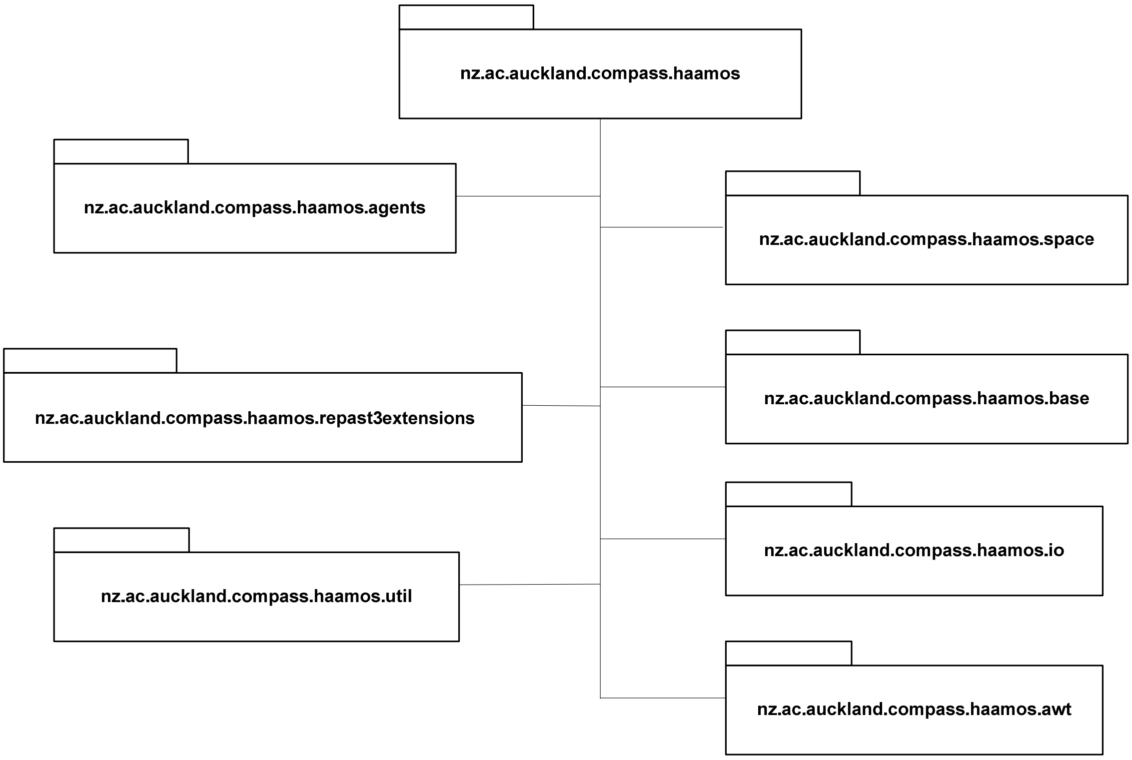
D.1 Guide for Reading Class Diagrams



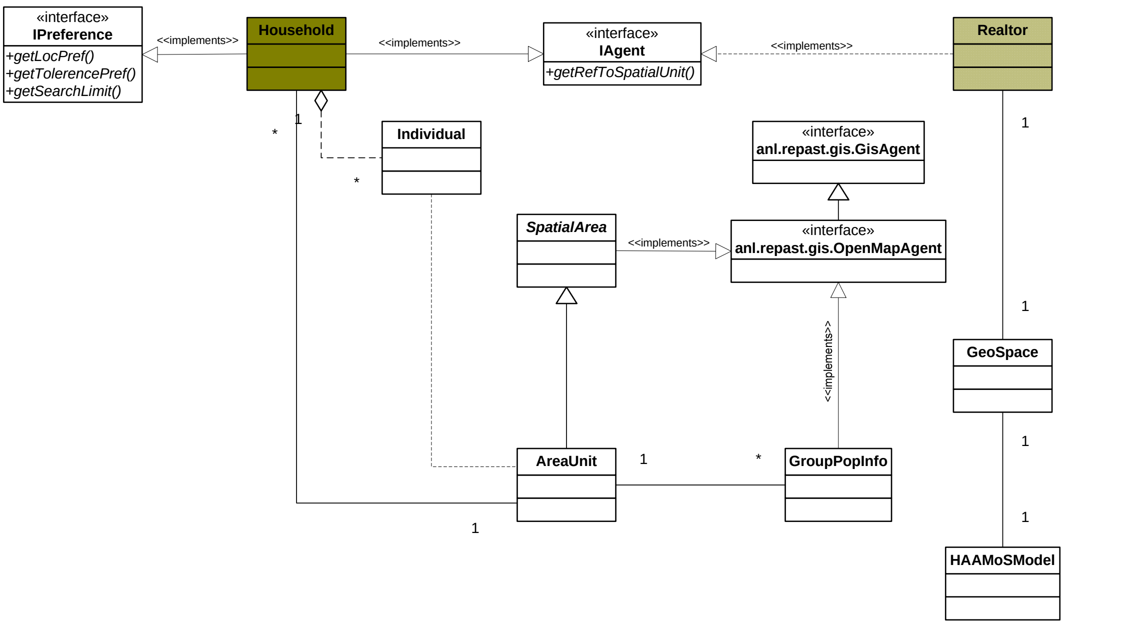
Guide for Reading Class Diagrams

The above figure shows symbols used in other UML diagrams shown in this appendix to describe the *HAAMoS* model. Notes (in blue) explain the meaning of symbols or relationships.

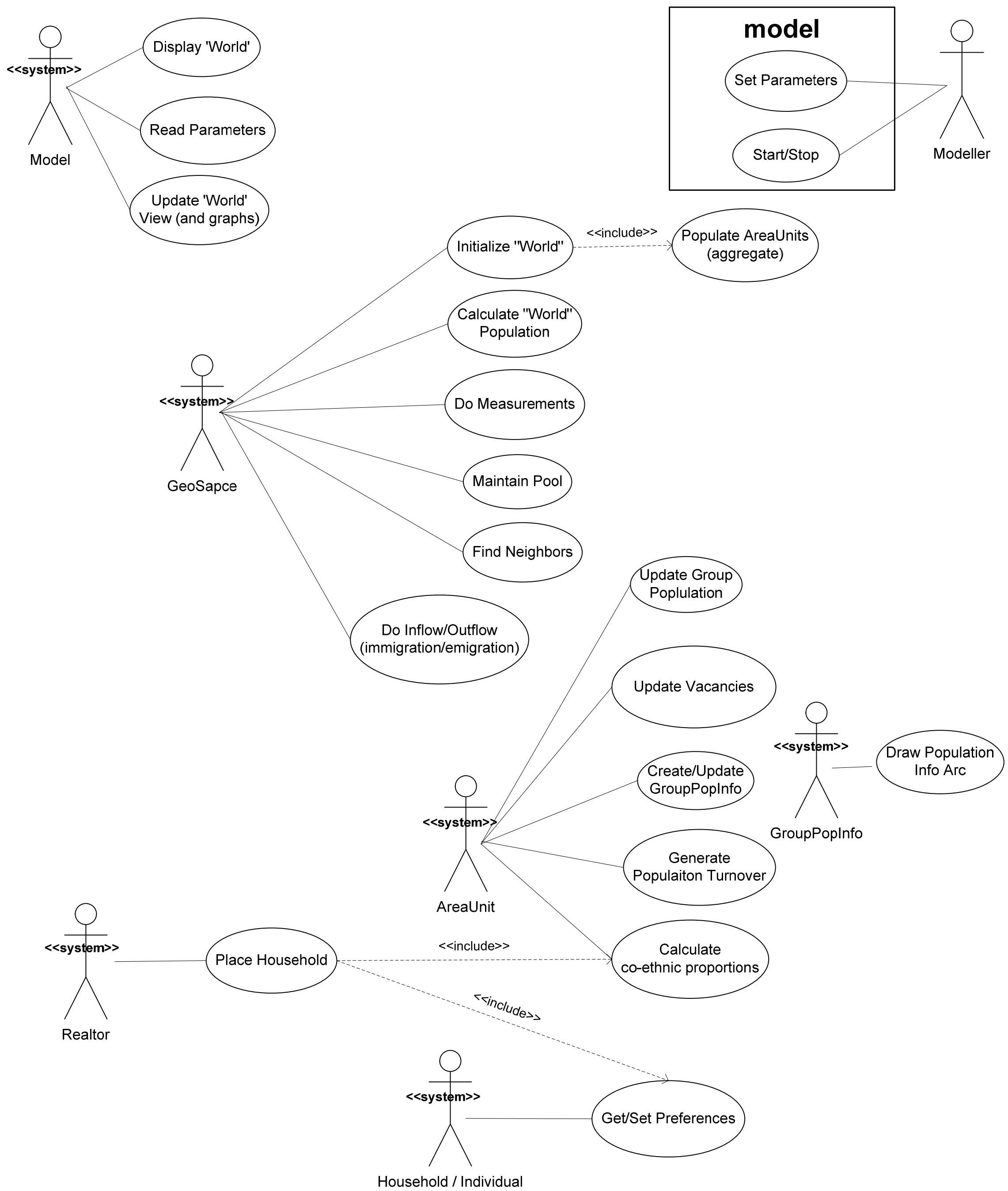
D.2 *HAAMoS* Diagrams



The above diagram shows the main packages of the *HAAMoS* model.



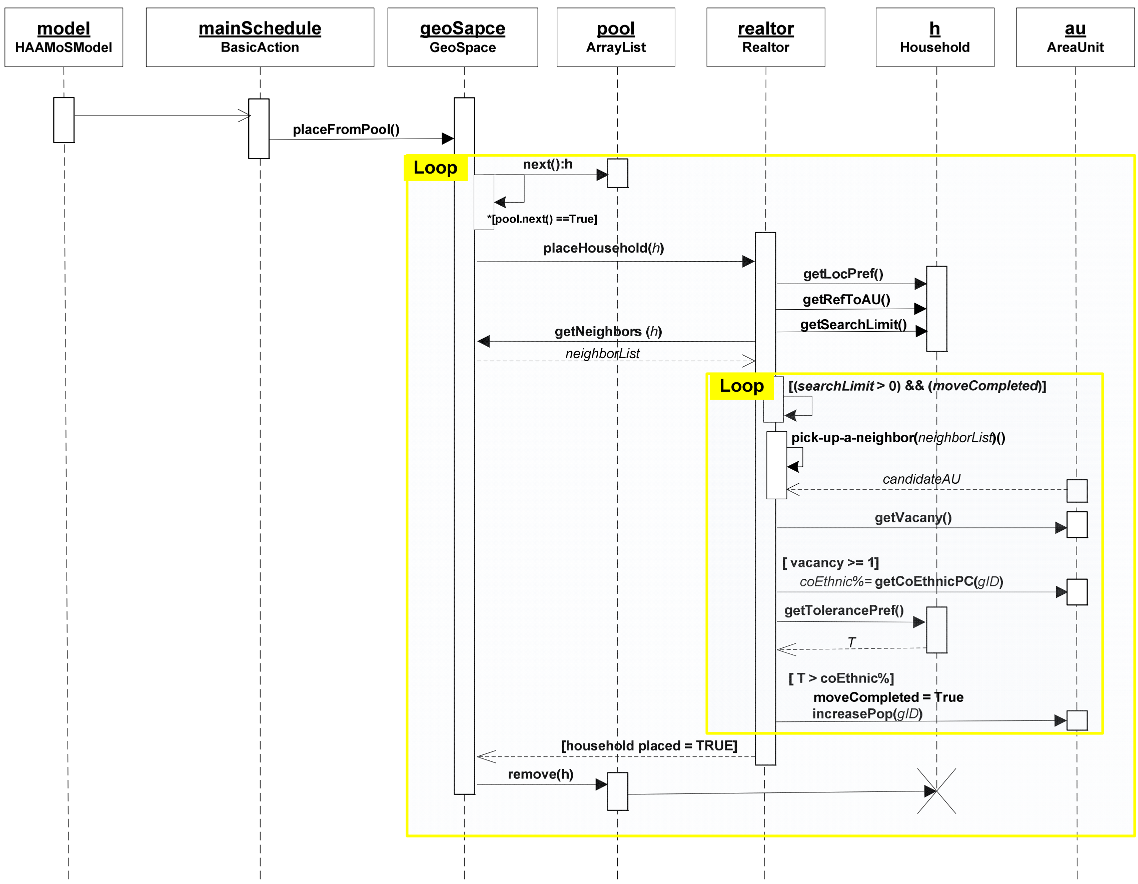
The above diagram shows the key classes and their relationships in the *HAAMoS* model.



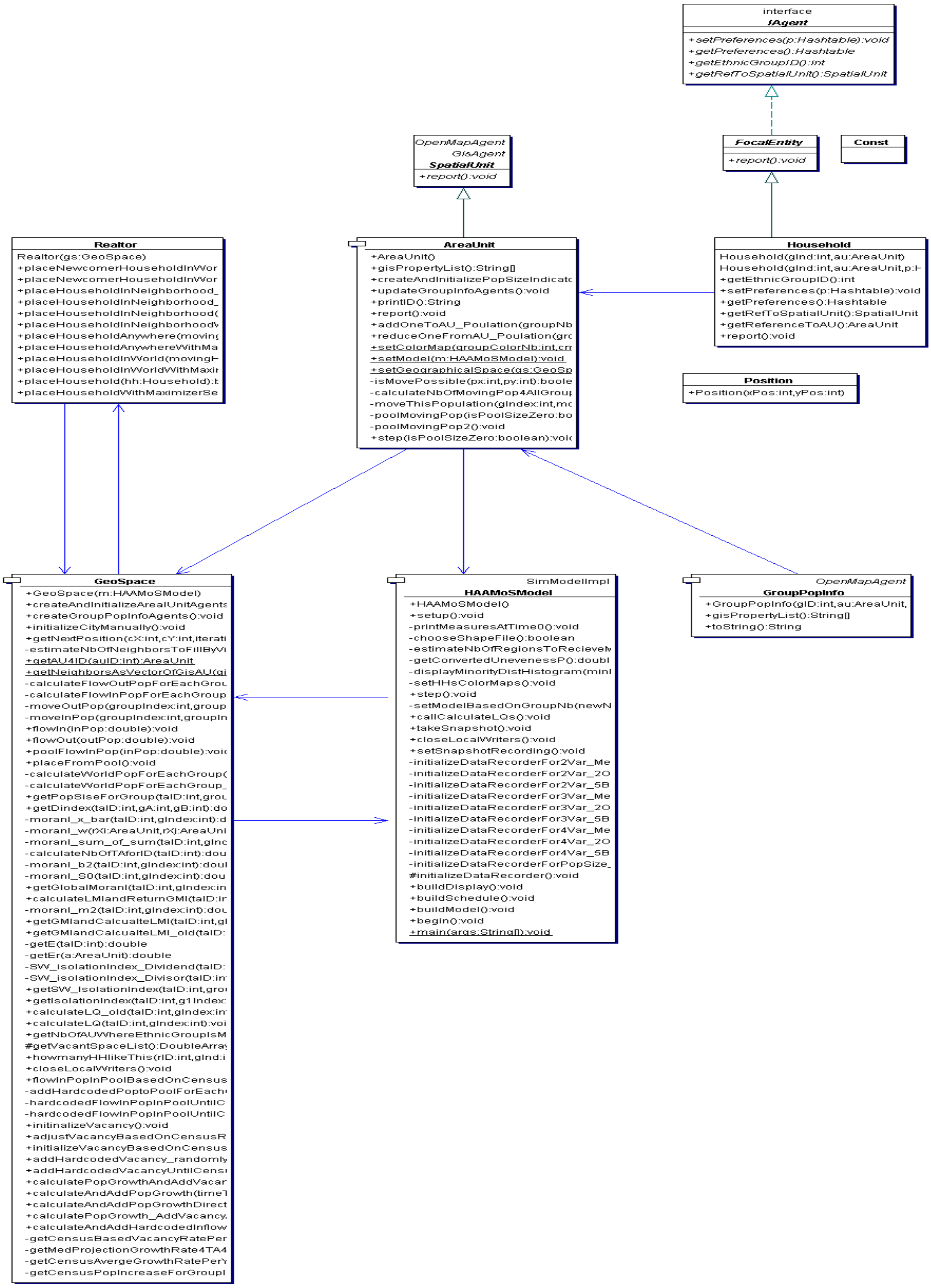
The above diagram is an illustration of the model behaviour core, as a whole and as a system, described by some of its major actors along with some of their key tasks (as use-cases[19]). This diagram shows how the (system) actors may use other actor(s) to

[19] Use-cases describe (or rather depict) system behaviour from actors' standpoints (*who* can do *what*). They can be viewed as a kind of usage scenario (similar to task analysis) where functional description of a system and its major processes are often graphically described through the lens of users and interactions. In software or systems engineering the description often portrays the system behaviour as it responds to external (out of the system) and internal requests.

perform their tasks. The tasks are described in an informal and non-exhaustive manner. Nevertheless, the diagram provides a good representation of the major actors involved in the model along with their roles



This sequence diagram shows the interaction of different actors in the placement of a household from the pool for an agent with local move preference.

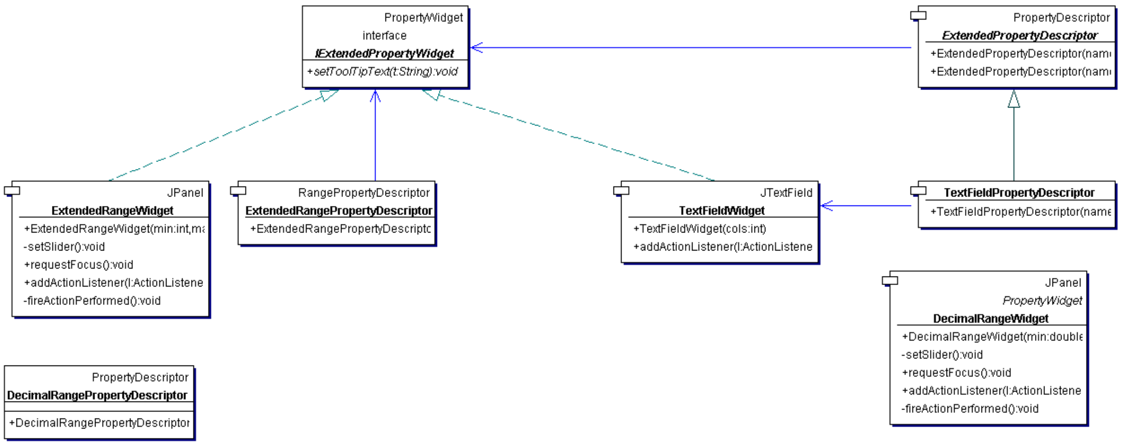


Class Diagram (*core* package)

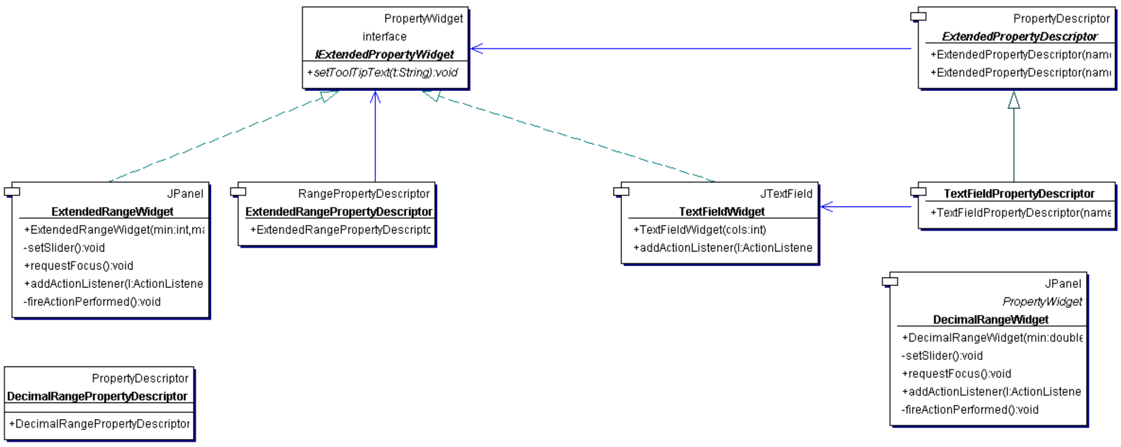
Because of the great number of methods and variables (attributes) contained in each class, only methods are displayed (attributes, properties and inner classes are hidden in this diagram).

D.2.1 Class Diagrams of Extended Repast Classes

Many Repast classes have been extended to include missing capabilities or to build new functionalities, necessary for the efficient development of *HAAMoS*.



Class Diagram (*Repast3extension.util* package)



Class Diagram (*Repast3extension.gis.display* package)

Appendix E: Batch Run Example

The parameter file has a specific syntax and format for instructing a customized sweep of the parameter space.

```
/* Batch Run Example */

runs: 1
TolerancePrefG2 {
    start: 0.1
    end: 0.5
    incr: 0.1
}

TolerancePrefG1{
   set:0
}

RngSeed {
  set: 1234
}

StopAtTick {
   set: 50
}
```

Example of Batch Parameter File Syntax

The simple example shown above instructs the execution of the model with fixed initial conditions including a random seed number (*RngSeed* = 1234) and a co-ethnic contact preference specification for group 1 (named *TolerancePrefG1 = 0,* i.e. no discriminatory behaviour). Each run of a simulation stops at a specified time tick (*StopAtTick* = 50). The simulation will start with the group 2 tolerance preference (*TolerancePrefG2*) equal to 10% which will be increased up to 50%, each increases the tolerance preference by 10%. In total 5 different runs will be simulated. More sophisticated batch runs usually include nested loops.

Appendix F: SNZ Census Data Confidentiality Rules

Rounding Rule (RR3)

Based on this rule, all counts of individuals, families, households and dwellings must be randomly rounded to base 3 (i.e. counts are multiples of 3). After each count has been rounded, totals must be separately treated in the same way (i.e. randomly rounded to base 3). Under the random rounding process, all table cell values, including row and column totals, are rounded as follows:

Zero counts and counts that are already multiples of three are left unchanged. Other counts are rounded to one or other of the two nearest multiples of three. For example, an original count of 17 would be rounded to 15 with a probability of 1/3 and rounded to 18 with a probability of 2/3, since $15 \times 1/3 + 18 \times 2/3 = 17$.

Derivations Rule

All derivations from counts e.g. percentages and ratios must be derived from the randomly rounded counts. This excludes totals and subtotals, as these are independently randomly rounded. Once a derivation has been calculated, there will be no further random rounding applied to the derived data.

Mean Cell Size Rule (MCSR)

Mean cell size is calculated as the total unrounded subject population in a geographic area divided by the number of categories (excluding the totals) for that same geographic area.

The number of categories in the denominator is determined by the number of categories in the standard census variables. The mean cell size for individual geographic areas must be greater than two.

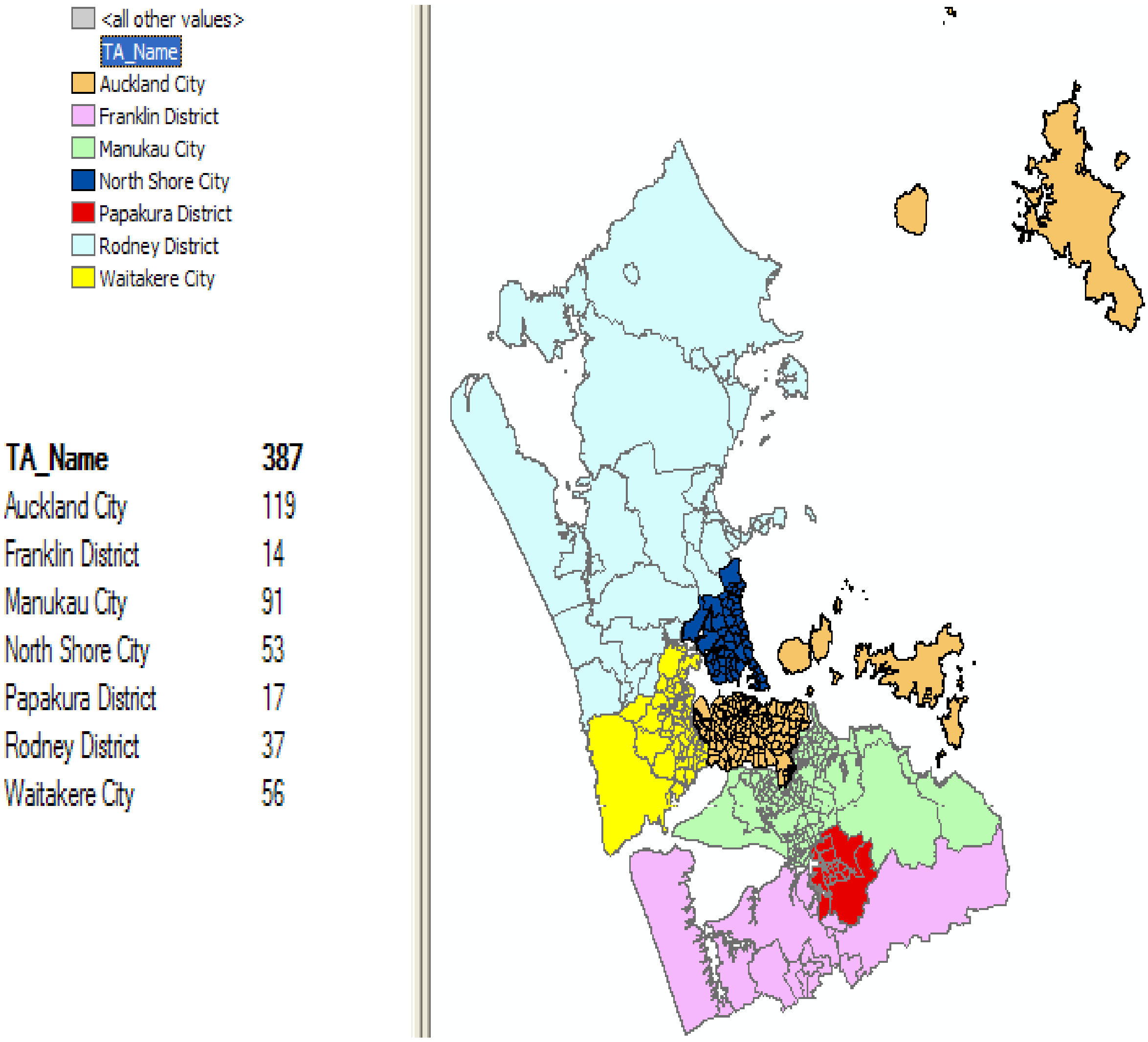
Tables may contain '*marginal tables*', which will appear as groups of totals embedded in the main table. Each of these tables will need to be assessed against this rule as a table in its own right. Often, the main table will fail the rule but the marginal tables will pass it. A separate mean cell size calculation is required for each marginal table. This additional calculation ensures that any table is treated consistently, whether it appears in a larger table, or on its own.

A marginal table uses combinations of variables from the main table. For example, in a 3-dimensional table for one geographic area, and the variables age, sex and ethnicity, there could be 2-dimensional tables with that geographic area and each of: age and sex; sex and ethnicity; and age and ethnicity. There could also be 1-dimensional tables for that geographic area and each of: age; sex; and ethnicity. Each of these 2-dimensional and 1-dimensional tables are the marginal tables of the original 3-dimensional table.

When the mean cell size rule suppresses large cells, or the rule cannot be applied to output, a threshold rule may be applied instead. Cells at or below the threshold value of 5 are suppressed, and higher counts are released.

Appendix G: Additional Data, Map and Info

G.1 Complete Auckland Region Map



Complete Map of Auckland Region with 387 Area Units [20]
Area Units Designated as "Area Outside TA" are Excluded

---

[20] One lower than maximum of 388 (i.e. 399 minus 11 areas outside TA).

G.2 Information on Area Units

Area Units Included in 'Area Outside TA'

AU-Code	AU	TA-Code	TA	G-TOT	Confidential (..C)
520803	Bays-Waiheke Island	999	Area Outside TA	15	Yes
616002	Inlet-Port Fitzroy	999	Area Outside TA	96	Non- (Only in PPD ver.)
616100	Cape Barrier	999	Area Outside TA	0	Yes
617101	Inlet-Kaipara Harbour South	999	Area Outside TA	0	Yes
617200	Inlet-Takapuna Head	999	Area Outside TA	0	Yes
617501	Oceanic-Auckland Region East	999	Area Outside TA	33	Yes
617602	Oceanic-Auckland Region West	999	Area Outside TA	3	Yes
617605	Inlet-Manukau Harbour	999	Area Outside TA	3	Yes
617703	Tamaki Strait	999	Area Outside TA	33	Yes
617800	Inlet-Waiuku River	999	Area Outside TA	0	Yes
617901	Inlet-Waitemata Harbour	999	Area Outside TA	1008	Non

Area Units Containing ..C (confidential/ low) Values

AU-Code	AU	TA-Code	TA	G-TOT	(..C) description
520804	Tidal-Waiheke Island	007	Auckland City	12	
615800	Mokohinau Island	007	Auckland City	0	
615900	Little Barrier Island	007	Auckland City	27	
616001	Kaikoura and Rangiahua Islands	007	Auckland City	48	
616200	Rakitu Island	007	Auckland City	18	
616300	Browns Island	007	Auckland City	0	
616400	Aiguilles Island	007	Auckland City	0	
617400	Inlet-Hobson Bay	007	Auckland City	3	
617502	Tidal-Great Barrier Island	007	Auckland City	6	
617503	Tidal-Motutapu and Browns Islands	007	Auckland City	0	
617606	Tidal-Manukau Harbour North	007	Auckland City	0	
617704	Tidal-Eastern Bays	007	Auckland City	33	
617902	Tidal-Waitemata Harbour	007	Auckland City	0	
617903	Auckland City-Marinas	007	Auckland City	153	01 (only PPD)
521902	Middlemore	008	Manukau City	570	06 (only PPD)
522605	Half Moon Bay Marina	008	Manukau City	6	
523107	Burswood	008	Manukau City	4404	91
523109	Dannemora	008	Manukau City	7383	91
617604	Tidal-Manukau Harbour	008	Manukau City	0	
617702	Tidal-Tamaki	008	Manukau City	9	
505911	Gulf Harbour Marina	004	Rodney District	81	06 (PPD), 91&06 (TBD)
617102	Inlet-Kaipara River	004	Rodney District	0	

Area Units with Overall Lowest Population among 399

35 Area units with overall lowest population among original 399 area units in the Auckland Region are shown in the table below. The first 27 of them (which have even lower overall population) contain confidential values (..C), entirely or partially for certain census year. Eliminating these 27 will bring the number of maximum usable area units to 372. By eliminating all these 35, only 364 area units will remain for possible usage.

AU_ID	AU_Name	TA_ID	TA_Name	G-TOT	..C
615800	Mokohinau Island	007	Auckland City	0	Y
616100	Cape Barrier	999	Area Outside TA	0	Y
616300	Browns Island	007	Auckland City	0	Y
616400	Aiguilles Island	007	Auckland City	0	Y
617101	Inlet-Kaipara Harbour South	999	Area Outside TA	0	Y
617102	Inlet-Kaipara River	004	Rodney District	0	Y
617200	Inlet-Takapuna Head	999	Area Outside TA	0	Y
617503	Tidal-Motutapu and Browns Islands	007	Auckland City	0	Y
617604	Tidal-Manukau Harbour	008	Manukau City	0	Y
617606	Tidal-Manukau Harbour North	007	Auckland City	0	Y
617800	Inlet-Waiuku River	999	Area Outside TA	0	Y
617902	Tidal-Waitemata Harbour	007	Auckland City	0	Y
617400	Inlet-Hobson Bay	007	Auckland City	3	Y
617602	Oceanic-Auckland Region West	999	Area Outside TA	3	Y
617605	Inlet-Manukau Harbour	999	Area Outside TA	3	Y
522605	Half Moon Bay Marina	008	Manukau City	6	Y
617502	Tidal-Great Barrier Island	007	Auckland City	6	Y
617702	Tidal-Tamaki	008	Manukau City	9	Y
520804	Tidal-Waiheke Island	007	Auckland City	12	Y
520803	Bays-Waiheke Island	999	Area Outside TA	15	Y
616200	Rakitu Island	007	Auckland City	18	Y
615900	Little Barrier Island	007	Auckland City	27	Y
617501	Oceanic-Auckland Region East	999	Area Outside TA	33	Y
617703	Tamaki Strait	999	Area Outside TA	33	Y
617704	Tidal-Eastern Bays	007	Auckland City	33	Y
616001	Kaikoura and Rangiahua Islands	007	Auckland City	48	Y
505911	Gulf Harbour Marina	004	Rodney District	81	Y
616002	Inlet-Port Fitzroy	999	Area Outside TA	96	N
617903	Auckland City-Marinas	007	Auckland City	153	N
506620	Kawau	004	Rodney District	432	N
506616	Matheson Bay	004	Rodney District	459	N
520900	Islands-Motutapu, Rangitoto, Rakino	007	Auckland City	492	N
521902	Middlemore	008	Manukau City	570	N
505803	Waiwera	004	Rodney District	942	N
617901	Inlet-Waitemata Harbour	999	Area Outside TA	1008	N

Nine Non-Confidential Area Units Eliminated from Study

AU_ID	AU_Name	TA_ID	TA_Name	G-TOT	..C
617903	Auckland City-Marinas	007	Auckland City	153	N
506620	Kawau	004	Rodney District	432	N
506616	Matheson Bay	004	Rodney District	459	N
520900	Islands-Motutapu Rangitoto Rakino	007	Auckland City	492	N
521902	Middlemore	008	Manukau City	570	N
505803	Waiwera	004	Rodney District	942	N
505400	Leigh	004	Rodney District	1629	N
512802	Armour Bay	006	Waitakere City	1731	N
521000	Great Barrier Island	007	Auckland City	3963	N

Area Units and Their Altered Connectivity Weight in GAL File

AU Affected in GAL	Neighbouring #		New Added
AU_ID *(name, [TA])*	**Old**	**New**	**AU_ID** *(name, [TA])*
506800 (*Helensville*, [*Rodney*])	1	2 (+1)	506641 (*Parakai* [*Rodney*])
506641(*Parakai* [*Rodney*])	2	3 (+1)	506800 (*Helensville*, [*Rodney*])
520801 (*Waiheke Island, [Auckland]*)	0	2 (+2)	514101 (*Akl Harbourside, [Auckland]*) 509500 (Stanley Bay, *[North Shore]*)
514101 (*Akl Harbourside, [Auckland]*)	4	5 (+1)	520801 (*Waiheke Island, [Auckland]*)
509500 (Stanley Bay, *[North Shore]*)	1	2 (+1)	520801 (*Waiheke Island, [Auckland]*)

Units Consisting of List (of Polygons)
[not displaying a dynamic pie-chart]

AU_Name	TA_Name
Army Bay	Rodney
Tauhoa-Puhoi	Rodney
Cape Rodney	Rodney
Algies Bay-Mahurangi	Rodney
Parakai	Rodney
South Head	Rodney
Kaukapakapa	Rodney
Muriwai Beach	Rodney
Glamorgan	North Shore
Karekare	Waitakere
Rosebank	Auckland
Waiheke Island	Auckland
Turanga	Manukau
Beachlands-Maraetai	Manukau
Ambury	Manukau
Favona North	Manukau

G.3 TA-based Projections

The SNZ projections (presented in the tables below) are given for each ethnicity and each area using different fertility (birth rate), mortality, migration, and inter-ethnic mobility assumptions. The terms 'low', 'medium', and 'high' do not correspond to probabilities, but merely indicate the combination of assumptions. For example, the medium projection series uses medium fertility, medium mortality, medium migration, and medium inter-ethnic mobility for each ethnic group and each area.

The SNZ projections are given for four ethnic groups in seven districts (TA) of the Auckland region and here are matched with their counterparts in this model (thesis) from selected area units within the same districts. The values for selected area units for this thesis are lower, since many area units have been eliminated for this study. Moreover, data have been adjusted for this study (e.g. for multiple ethnicity choice effect). Despite different values for variables in this thesis, calculated real annual changes are usually close to SNZ's values. This is despite lack of values for the 1991 Census year. So the average annual change under the 'real' column for SNZ values are based on 1996-2006 changes, whilst for this thesis they are based on 1991-2006 changes. There are however two exceptions: changes for Europeans in *Papakura* (0.53 for SNZ vs -2.52 for Thesis) and Maoris in *Auckland* (-2.57 for SNZ vs 0.60 for Thesis), can be explained by the elimination of certain area units, and/ or adjustment of data.

European Change in Each TA: Past (SNZ vs. Thesis) & SNZ Projections

						Change: %AVG/1Y	SNZ Projection: %AVG/1Y (2006-21)		
TA		**1991**	**1996**	**2001**	**2006**	**Real**	**High**	**Med**	**Low**
North Shore	SNZ (All)	N/A	154,600	157,100	163,500	0.83	1.3	0.4	-0.6
	Thesis (53)	134,122	134,870	138,229	144,291	0.85			
Waitakere	SNZ (All)	N/A	126,100	125,400	128,000	0.65	1.3	0.3	-0.8
	Thesis (55)	102,244	102,937	102,797	103,969	0.64			
Auckland	SNZ (All)	N/A	253,100	250,500	256,200	0.71	1.5	0.6	-0.5
	Thesis (104)	211,177	209,236	212,010	217,250	0.79			
Manukau	SNZ (All)	N/A	155,800	150,600	153,400	-2.68	0.7	-0.6	-2
	Thesis (87)	131,832	129,550	126,841	129,145	-2.69			
Papakura	SNZ (All)	N/A	31,500	30,900	32,100	0.53	1.3	0.3	-0.8
	Thesis (17)	26,578	25,866	25,546	26,056	-2.52			

Asian Change in Each TA: Past (SNZ vs. Thesis) & SNZ Projections

TA		1991	1996	2001	2006	Change: %AVG/1Y Real	SNZ Projection: %AVG/1Y (2006-21) High	Med	Low
North Shore	SNZ (All)	N/A	16,800	26,100	41,900	0.96	5.3	3.9	2.2
	Thesis (53)	5,322	13,951	21,500	34,852	0.99			
Waitakere	SNZ (All)	N/A	12,600	20,000	32,400	0.93	5.6	4.2	2.6
	Thesis (55)	5,200	9,847	15,668	25,090	0.93			
Auckland	SNZ (All)	N/A	52,200	76,600	109,100	1.07	4.8	3.3	1.6
	Thesis (104)	24,067	41,564	60,358	86,320	1.09			
Manukau	SNZ (All)	N/A	30,000	46,000	75,700	1.04	5.6	4.2	2.6
	Thesis (87)	12,894	24,070	37,287	60,400	1.05			
Papakura	SNZ (All)	N/A	1,900	2,700	4,000	0.67	5.5	4	2.2
	Thesis (17)	1,014	1,494	2,111	3,033	0.66			

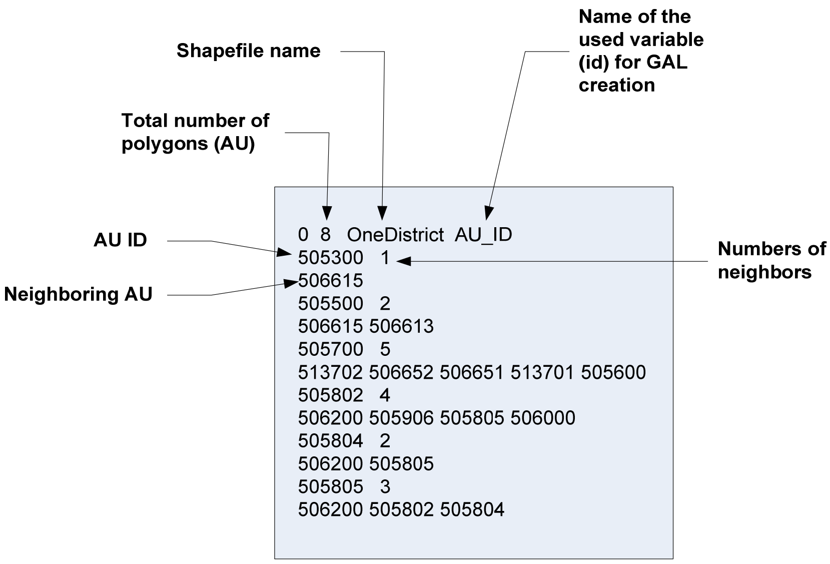
Pacific Change in Each TA: Past (SNZ vs. Thesis) & SNZ Projections

TA		1991	1996	2001	2006	Change: %AVG/1Y Real	SNZ Projection: %AVG/1Y (2006-21) High	Med	Low
North Shore	SNZ (All)	N/A	5,800	6,600	7,300	0.63	2.7	1.8	0.8
	Thesis (53)	3,532	4,743	5,502	5,994	0.68			
Waitakere	SNZ (All)	N/A	20,600	26,300	30,500	0.85	3.8	2.9	2
	Thesis (55)	13,423	16,066	20,558	23,431	0.85			
Auckland	SNZ (All)	N/A	52,400	54,400	56,600	0.74	1.8	0.6	-0.8
	Thesis (104)	39,702	40,421	43,045	45,013	0.77			
Manukau	SNZ (All)	N/A	66,500	82,300	99,800	1.00	3.8	3	2
	Thesis (87)	44,335	49,169	63,473	75,625	0.99			
Papakura	SNZ (All)	N/A	3,100	3,500	4,900	0.65	6.1	5.3	4.3
	Thesis (17)	2,154	2,375	2,658	3,719	0.63			

Maori Change in Each TA: Past (SNZ vs. Thesis) & SNZ Projections

TA		1991	1996	2001	2006	Change: %AVG/1Y Real	SNZ Projection: %AVG/1Y (2006-21) High	Med	Low
North Shore	SNZ (All)	N/A	13,300	13,300	14,200	0.57	2.5	1.4	0.2
	Thesis (53)	8,162	11,278	11,177	11,561	0.72			
Waitakere	SNZ (All)	N/A	22,200	24,000	26,200	0.74	2.8	1.9	0.9
	Thesis (55)	14,459	17,719	18,978	20,038	0.78			
Auckland	SNZ (All)	N/A	35,800	33,000	34,900	-2.57	1.9	0.8	-0.3
	Thesis (104)	25,604	28,115	26,363	26,808	0.60			
Manukau	SNZ (All)	N/A	47,800	50,000	53,900	0.79	2.6	1.7	0.7
	Thesis (87)	34,386	36,523	38,665	41,127	0.80			
Papakura	SNZ (All)	N/A	9,800	10,600	12,500	0.69	3.3	2.3	1.1
	Thesis (17)	6,906	7,648	8,236	9,675	0.70			

G.4 GAL File Example



GAL File Example

G.5 The Ethnicity Question in Different Census Questionnaires

7 Which ethnic group do you belong to?
Tick the box or boxes which apply to you

15 New Zealand European
16 New Zealand Maori
17 Samoan
18 Cook Island Maori
19 Tongan
20 Niuean
21 Chinese
22 Indian
23 Other *(such as Dutch, Japanese, Tokelauan)*
Please state

1991

10 Tick as many circles as you need to show which ethnic group(s) you belong to.

- NZ Maori
- NZ European *or* Pakeha
- other European → Which of these groups?
  - English
  - Dutch
  - Australian
  - Scottish
  - Irish
  - other → Print your ethnic group(s)
- Samoan
- Cook Island Maori
- Tongan
- Niuean
- Chinese
- Indian
- other (such as *FIJIAN, KOREAN*) → Print your ethnic group(s)

1996

11 Which ethnic group do you belong to?
Mark the space or spaces which apply to you.

- New Zealand European
- Māori
- Samoan
- Cook Island Maori
- Tongan
- Niuean
- Chinese
- Indian
- other (*such as* DUTCH, JAPANESE, TOKELAUAN). Please state:

2001

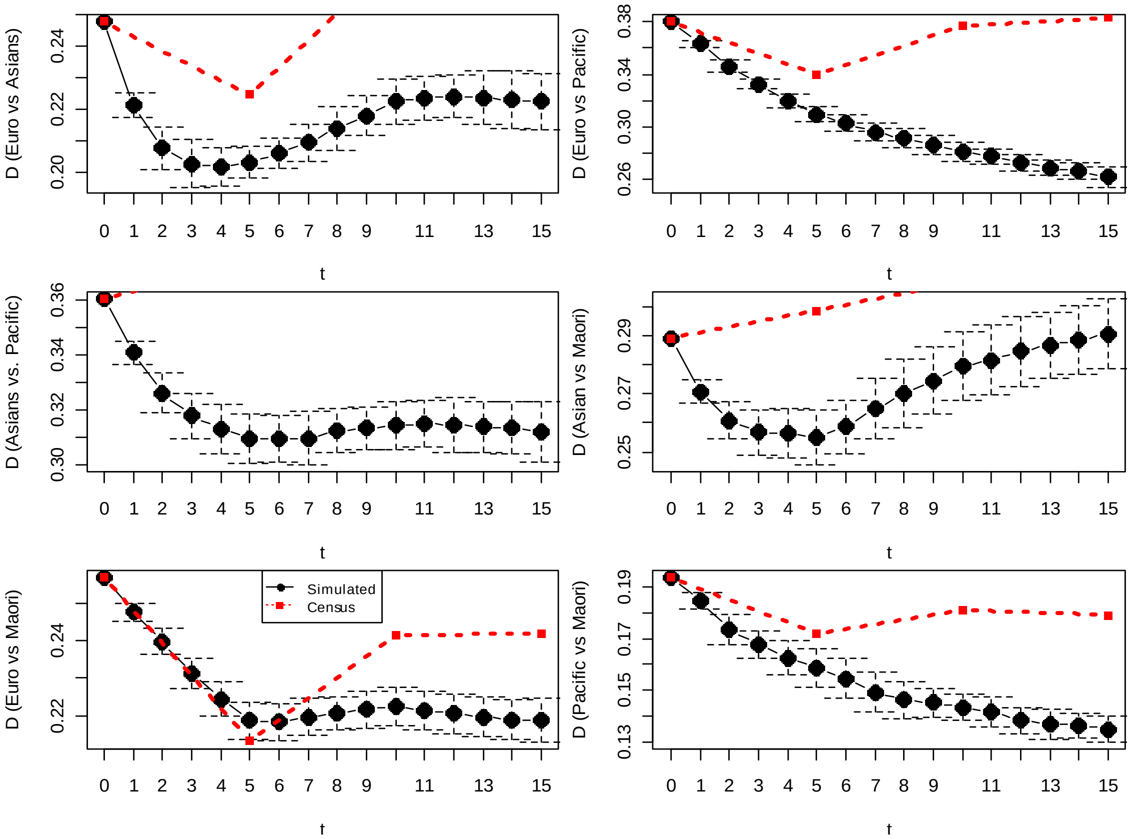
11 Which ethnic group do you belong to?
Mark the space or spaces which apply to you.

- New Zealand European
- Māori
- Samoan
- Cook Island Maori
- Tongan
- Niuean
- Chinese
- Indian
- other such as *DUTCH, JAPANESE, TOKELAUAN*. Please state:

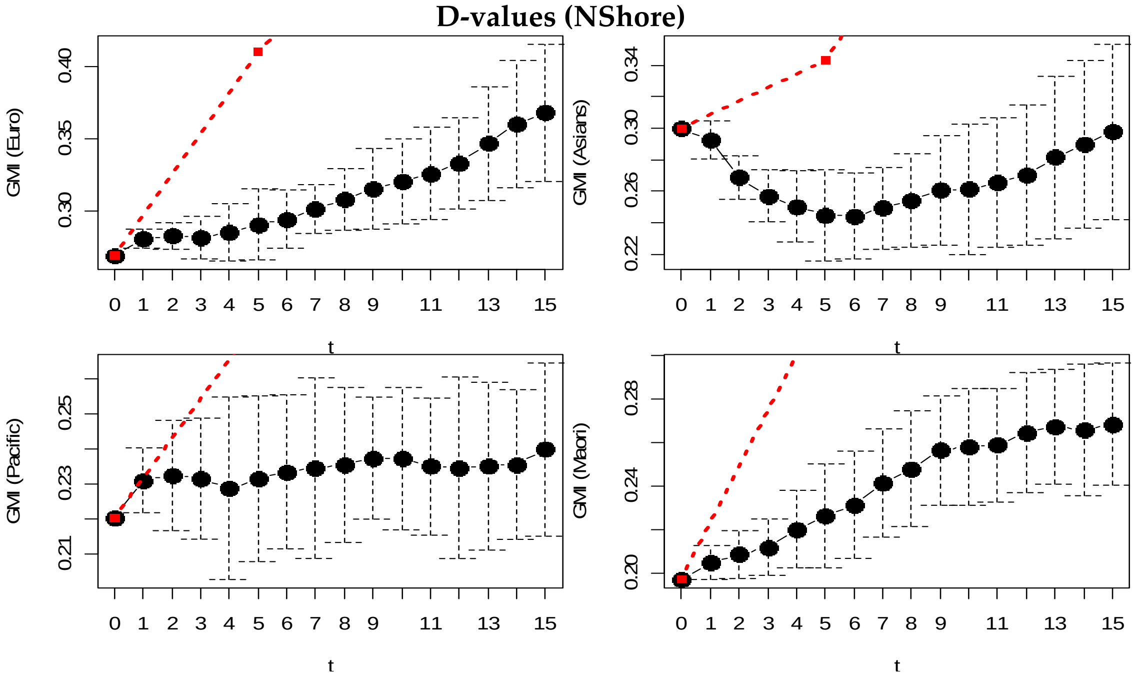
2006

Appendix H: Additional Simulation Outputs

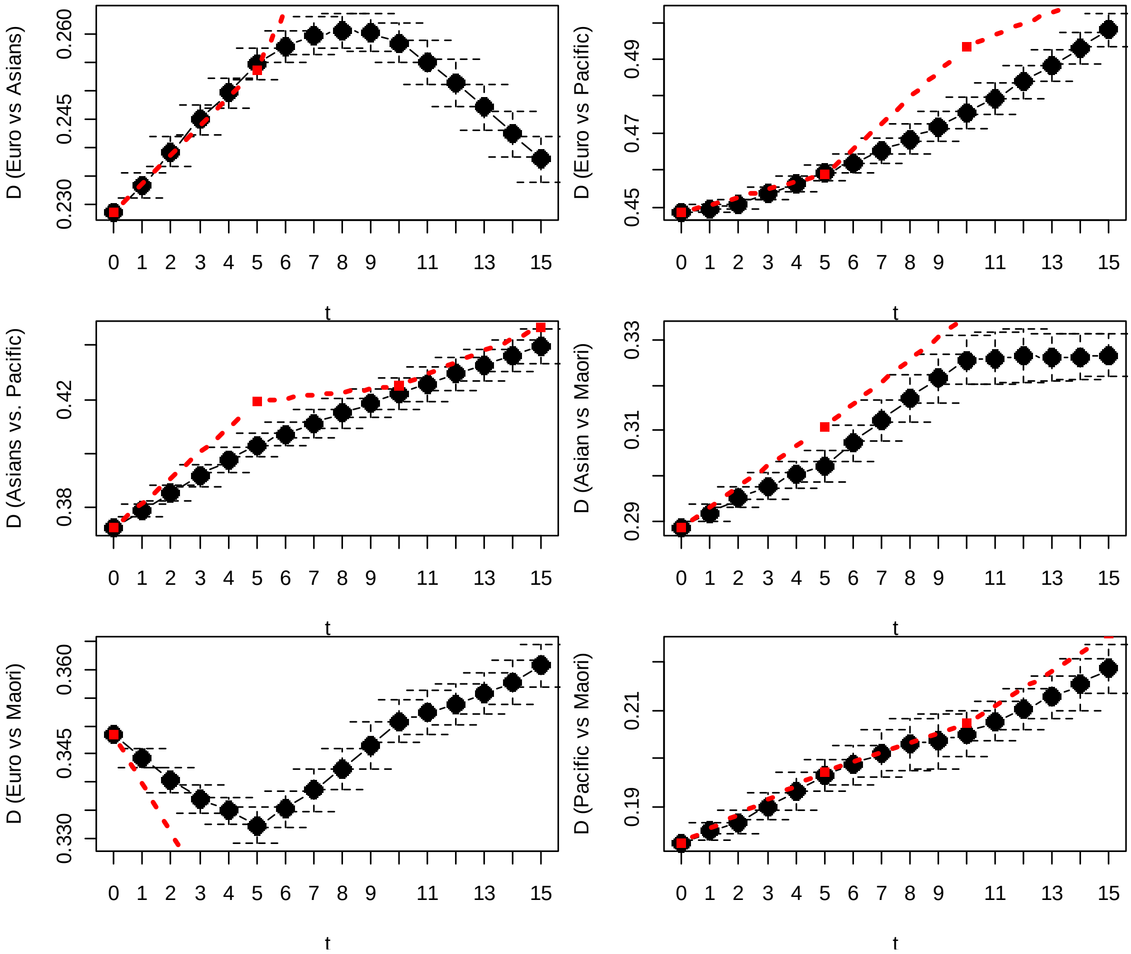
H.1 *POC of the Model Outputs (TAs)*



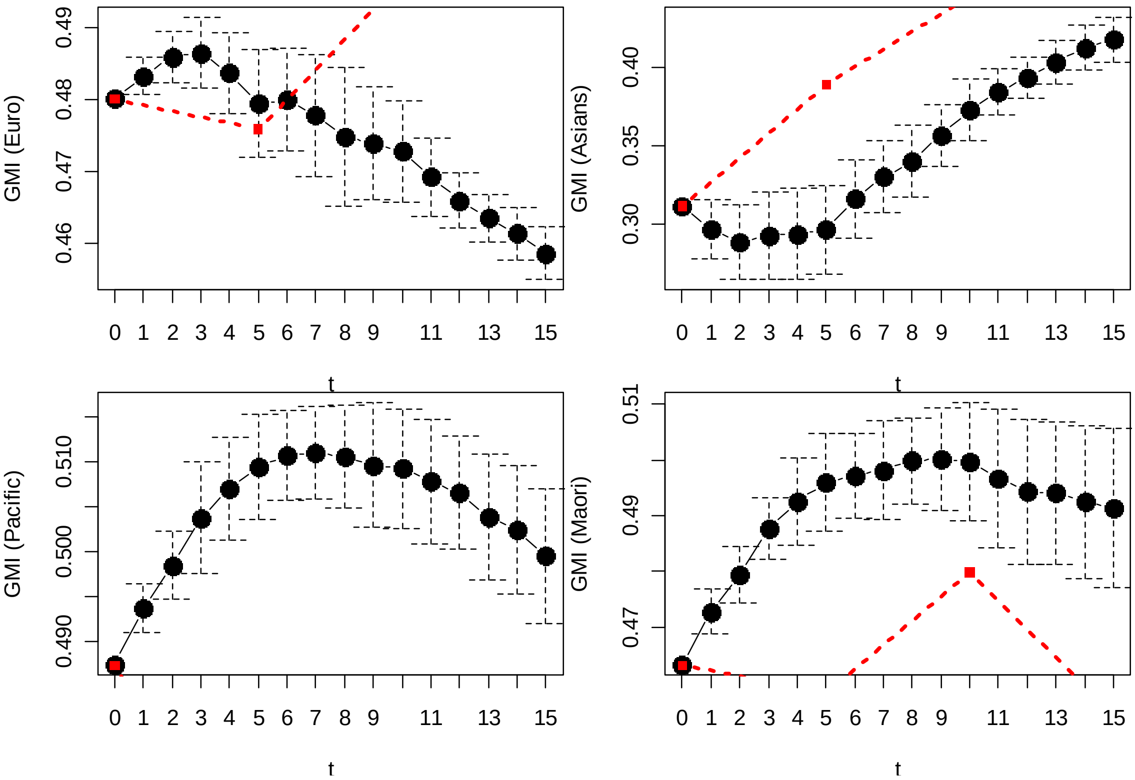
D-values (NShore)



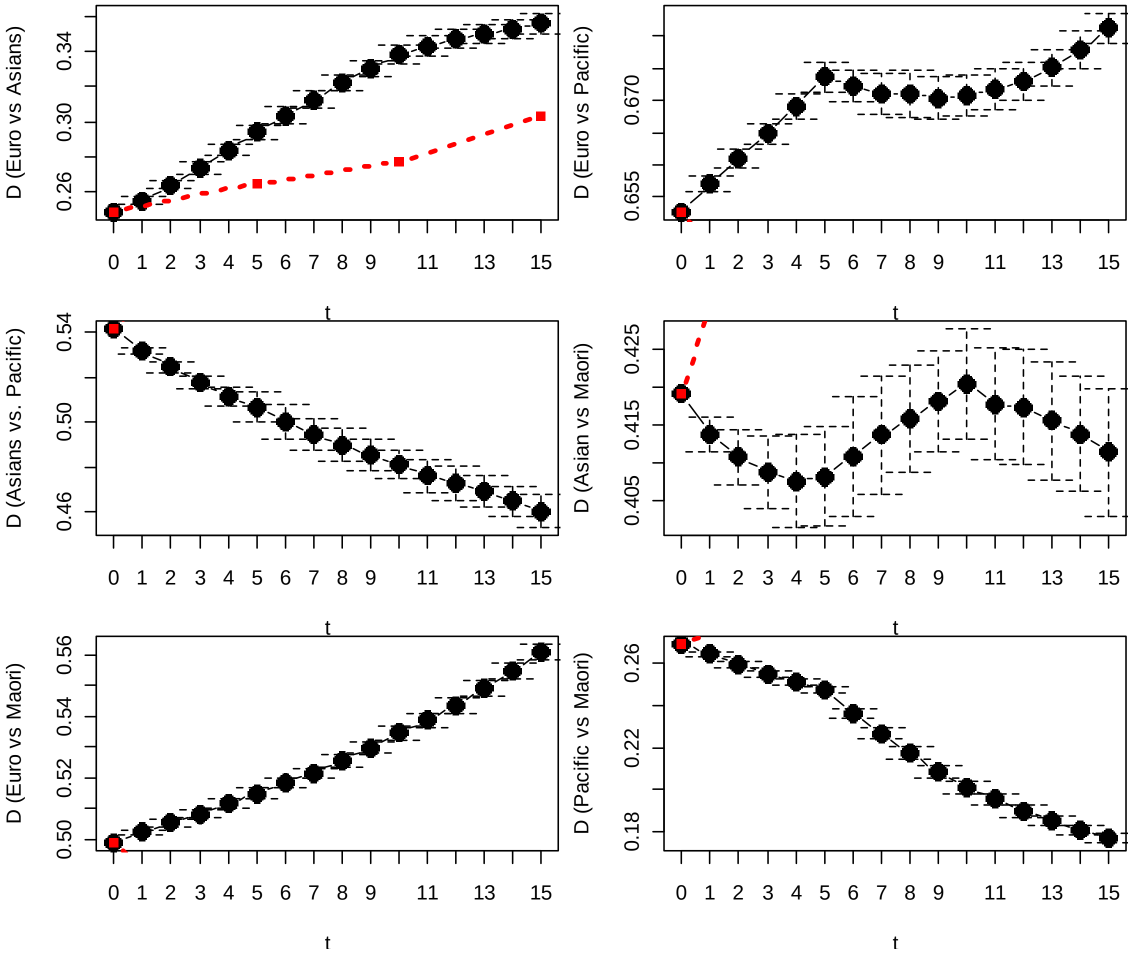
GMI-values (NShore)



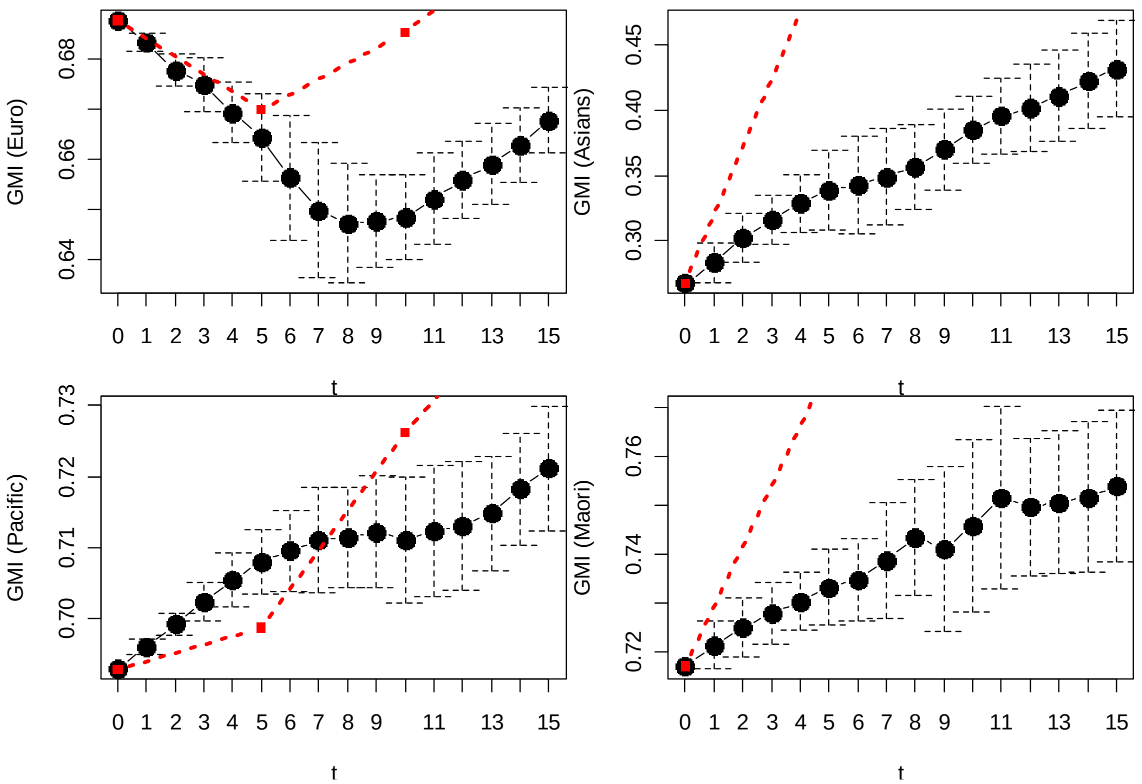
D-values (AKL)



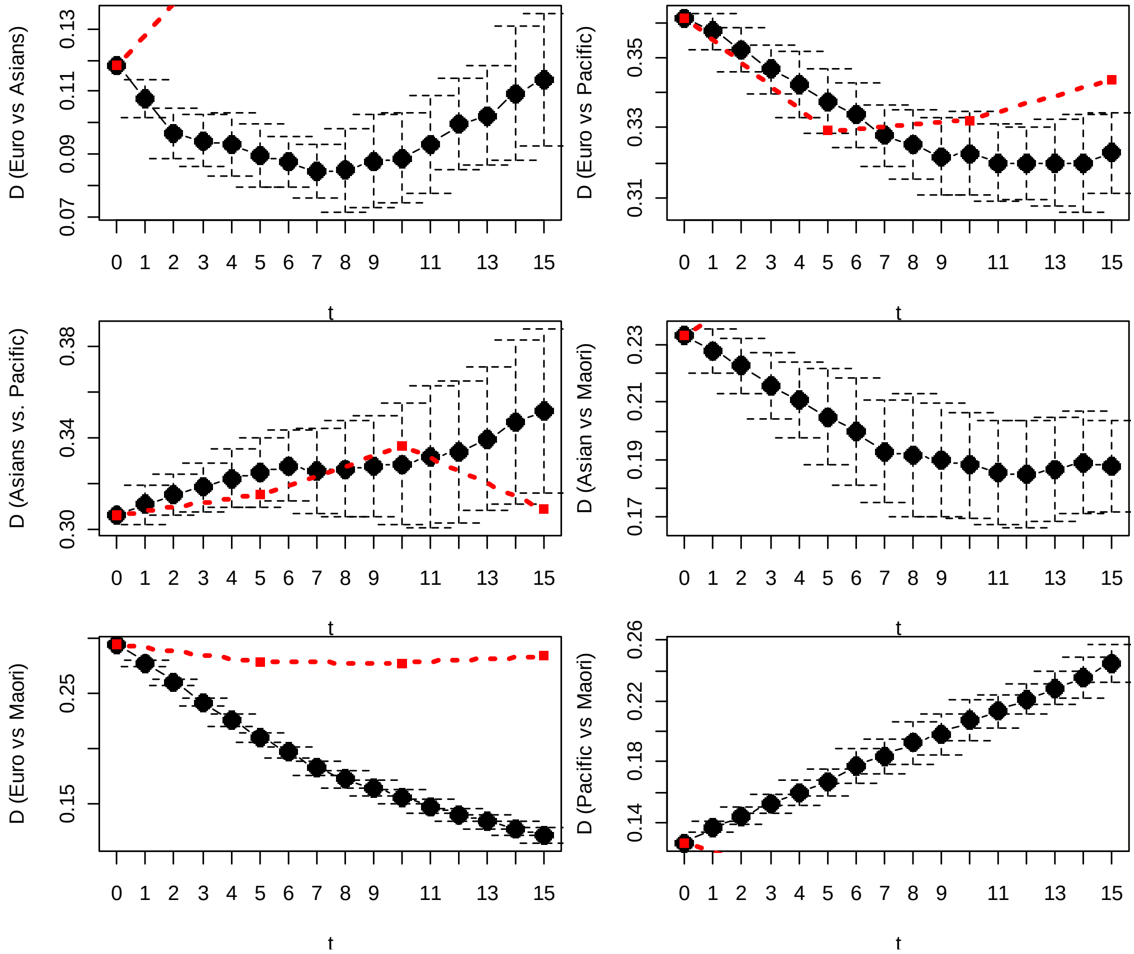
GMI-values (AKL)



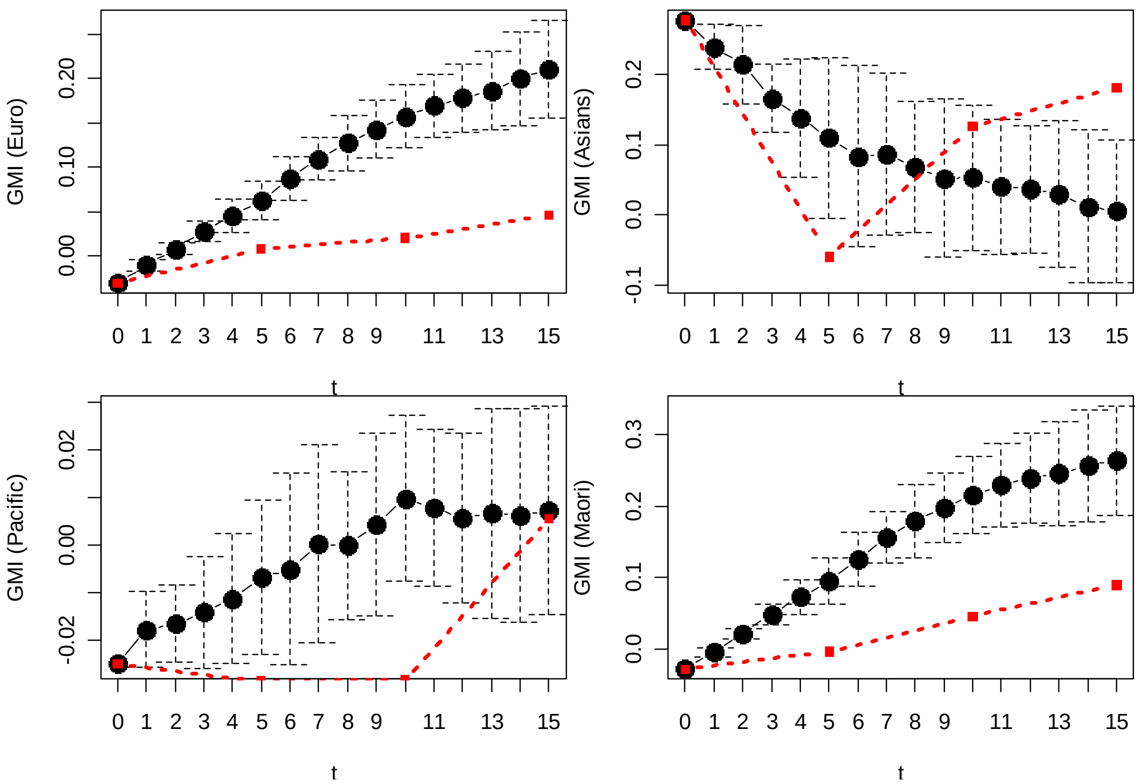
D-values (MKAU)



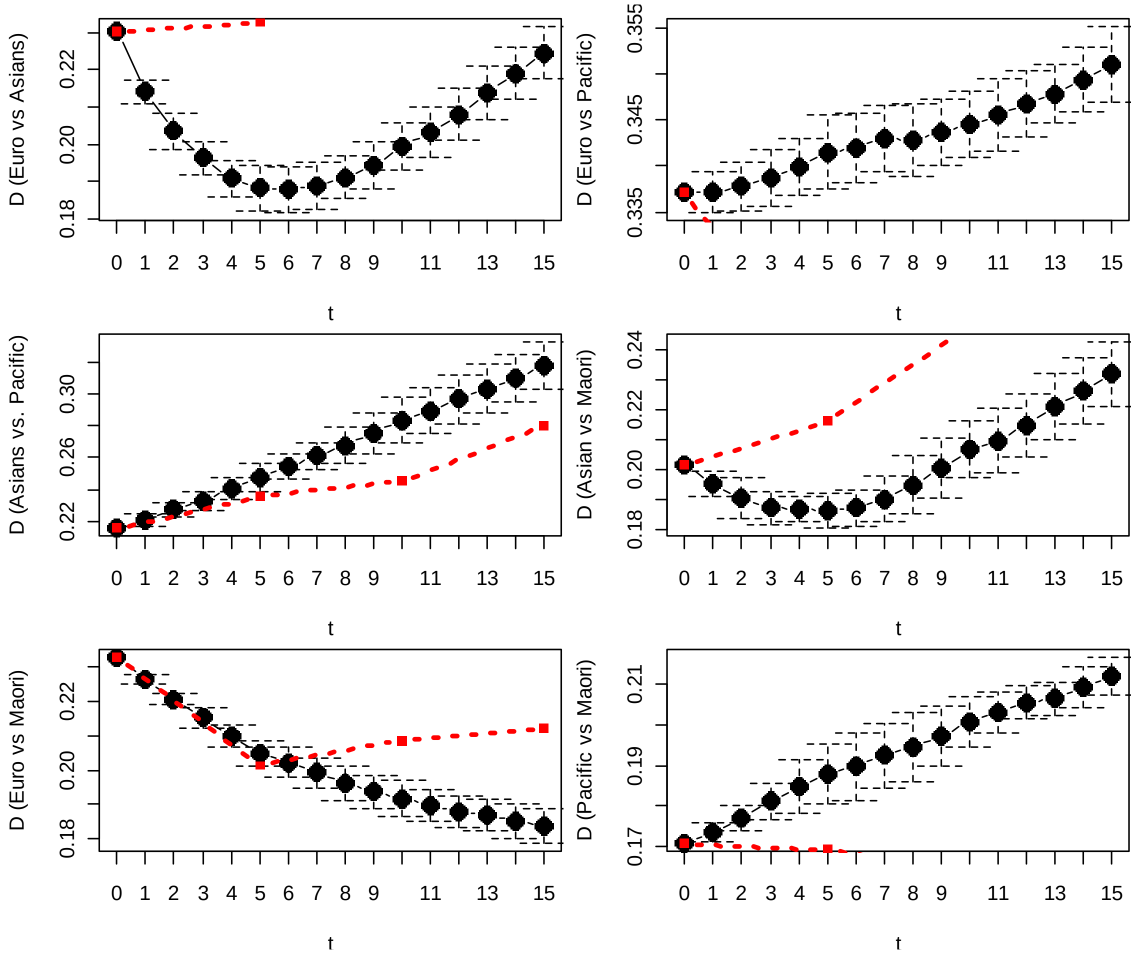
GMI-values (MKAU)



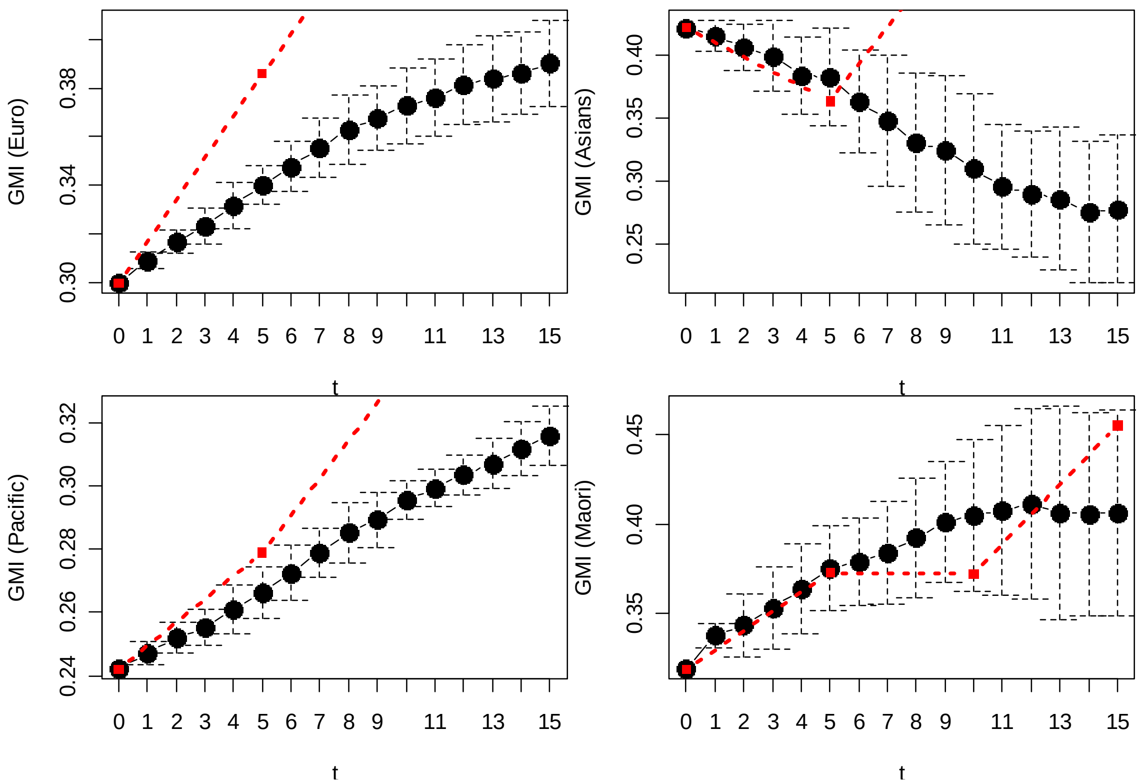
D-values (PAPK)



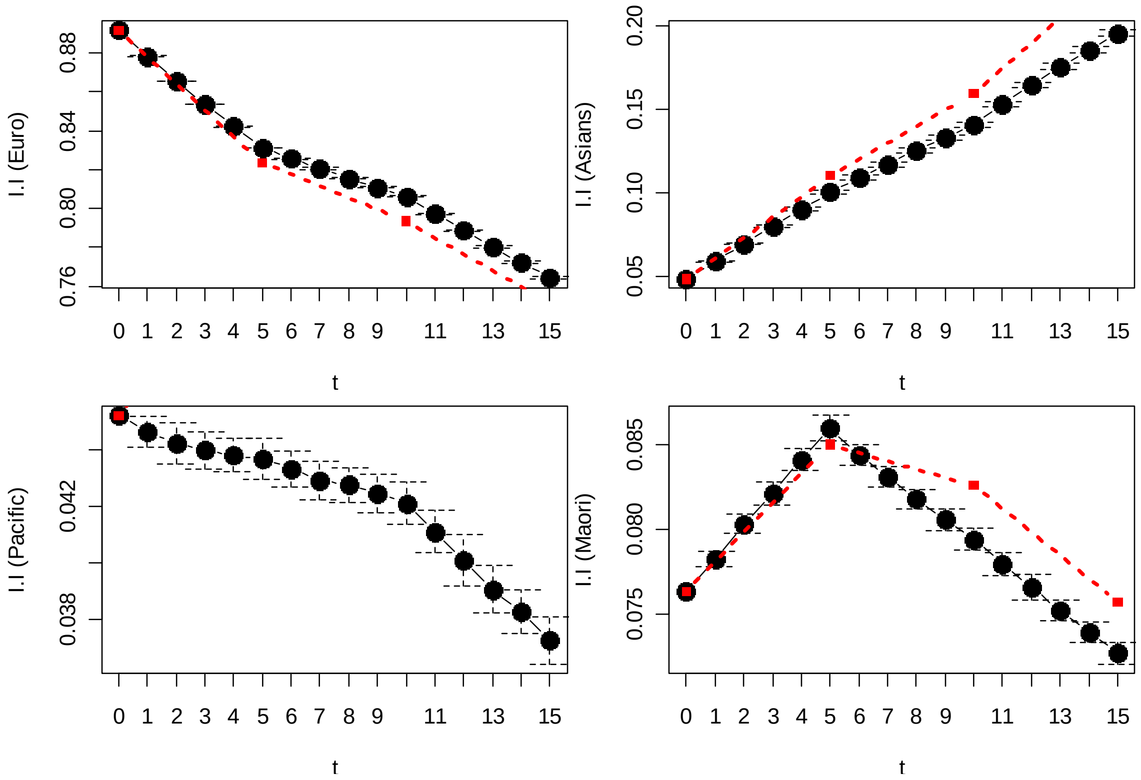
GMI-values (PAPK)



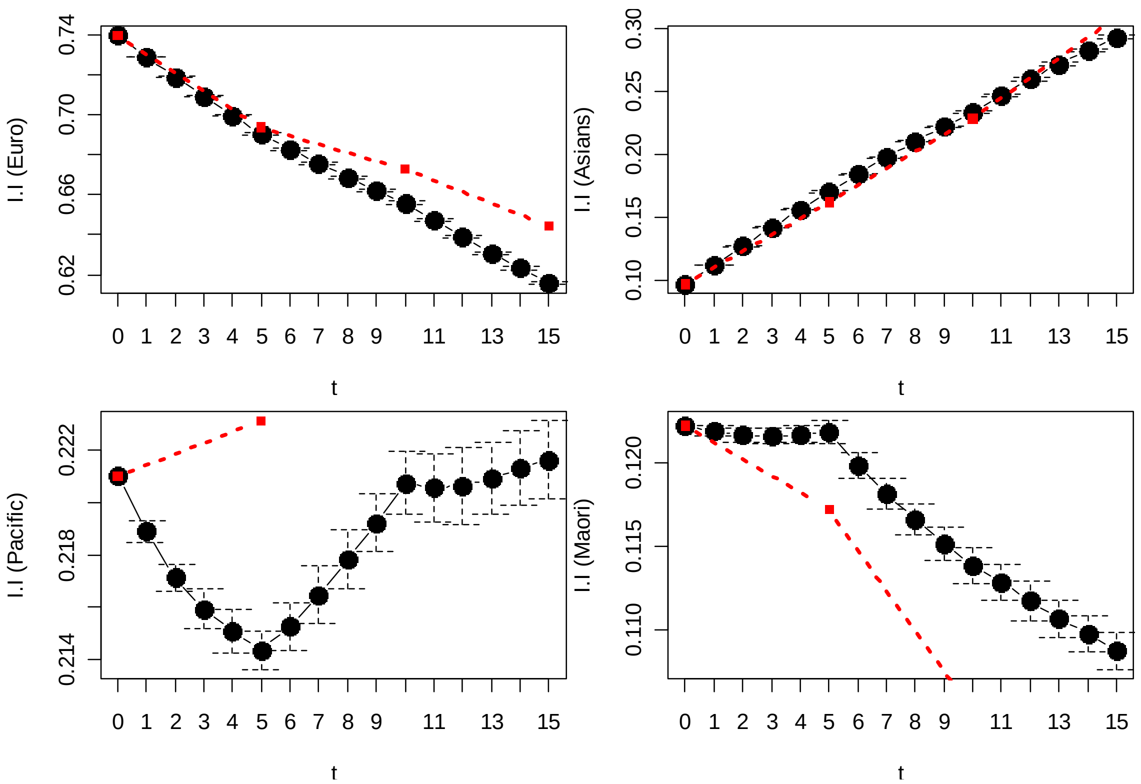
D-values (WTKRE)



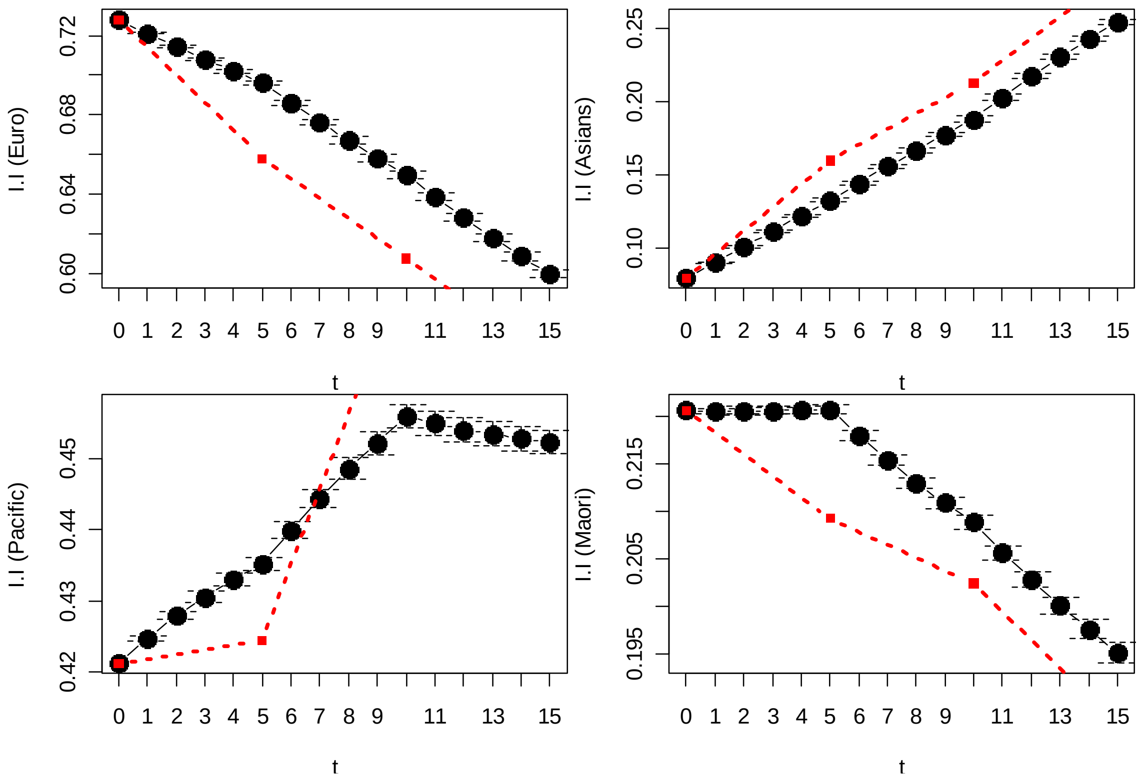
GMI-values (WTKRE)



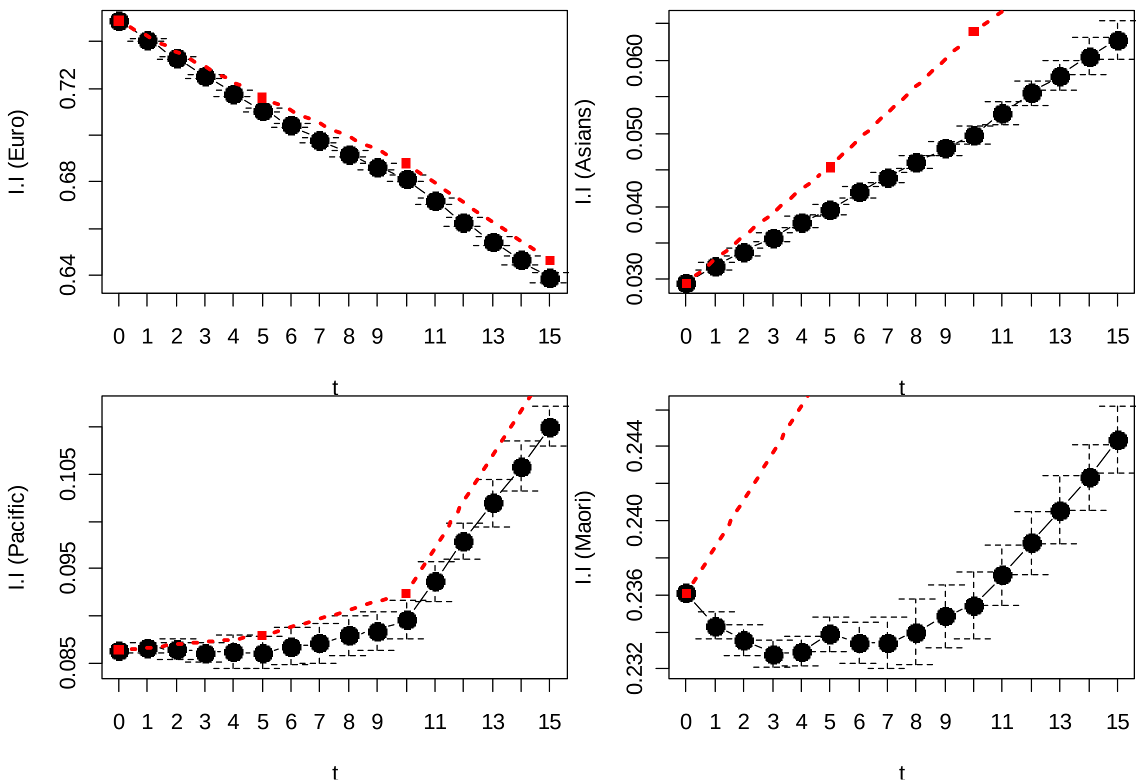
II_{xx}-values (NShore)



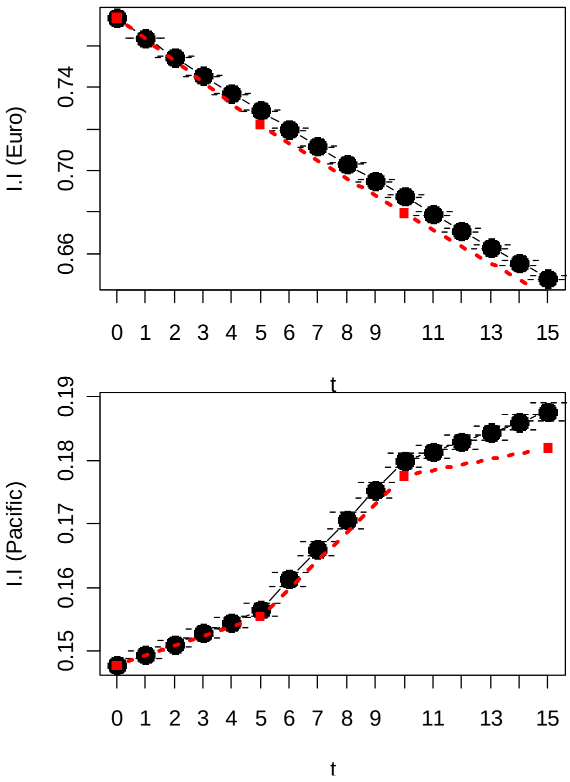
II_{xx}-values (AKL)



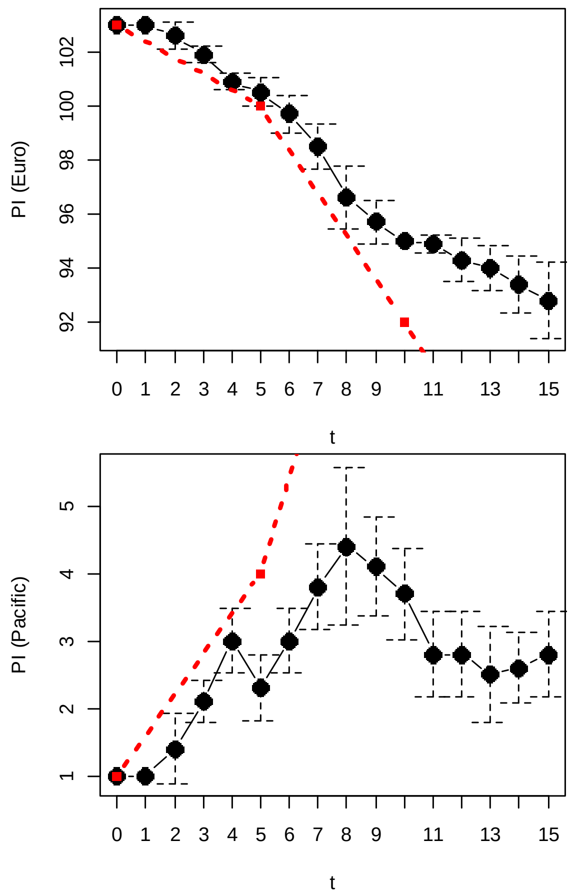
$\mathbf{II_{xx}}$**-values (MKAU)**



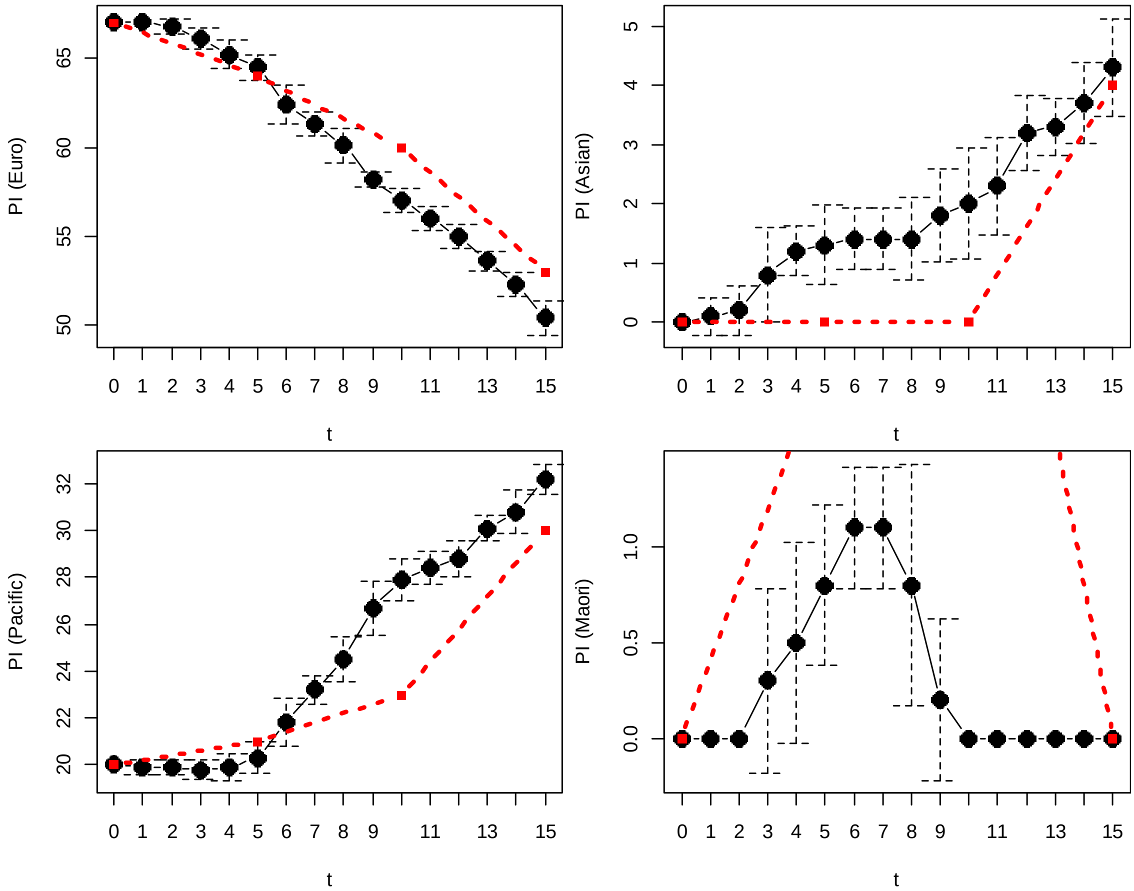
$\mathbf{II_{xx}}$**-values (PAPK)**



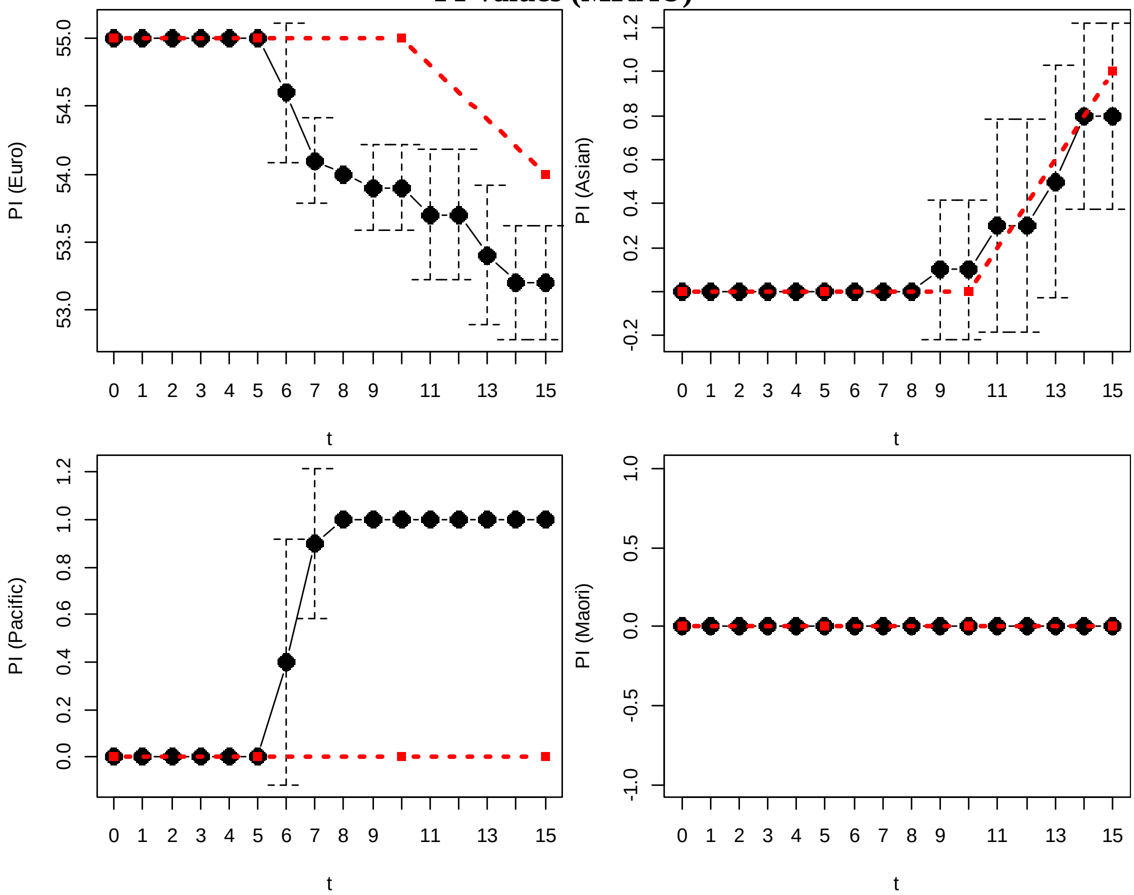
II_{xx}-values (WTKE)



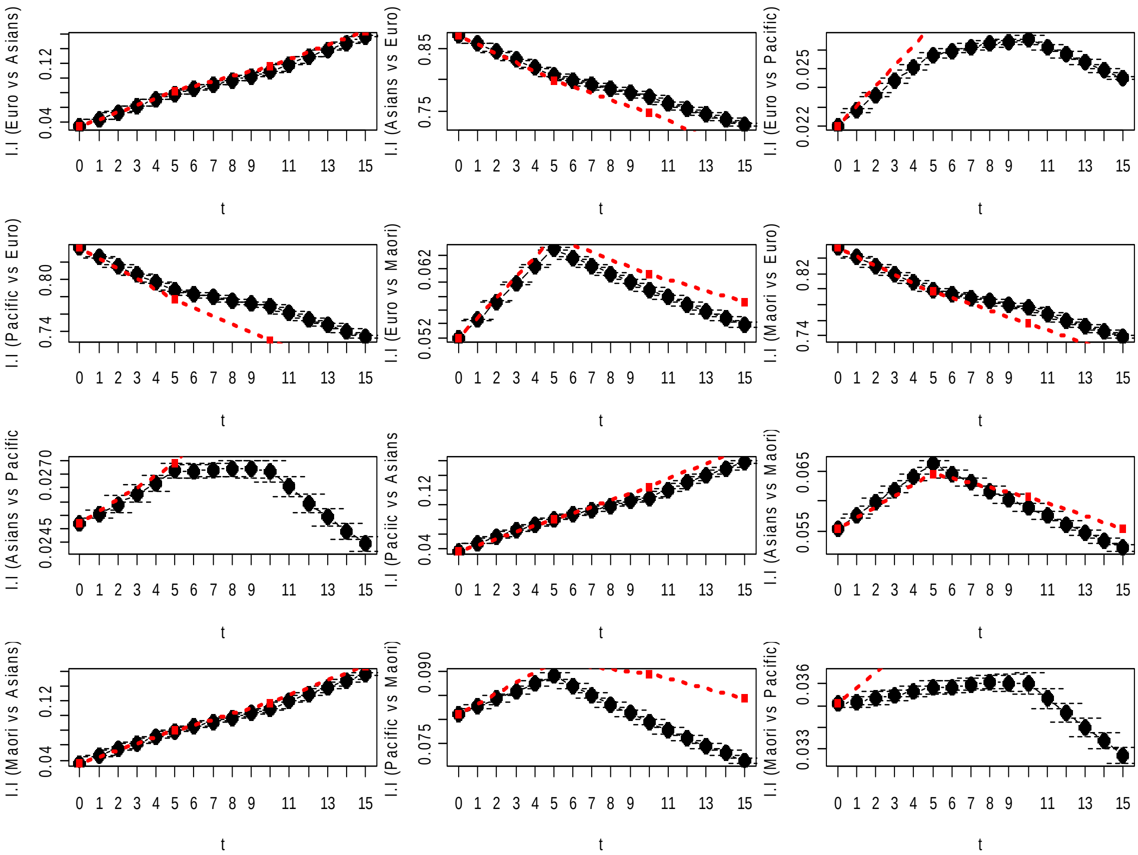
PI-values (AKL)



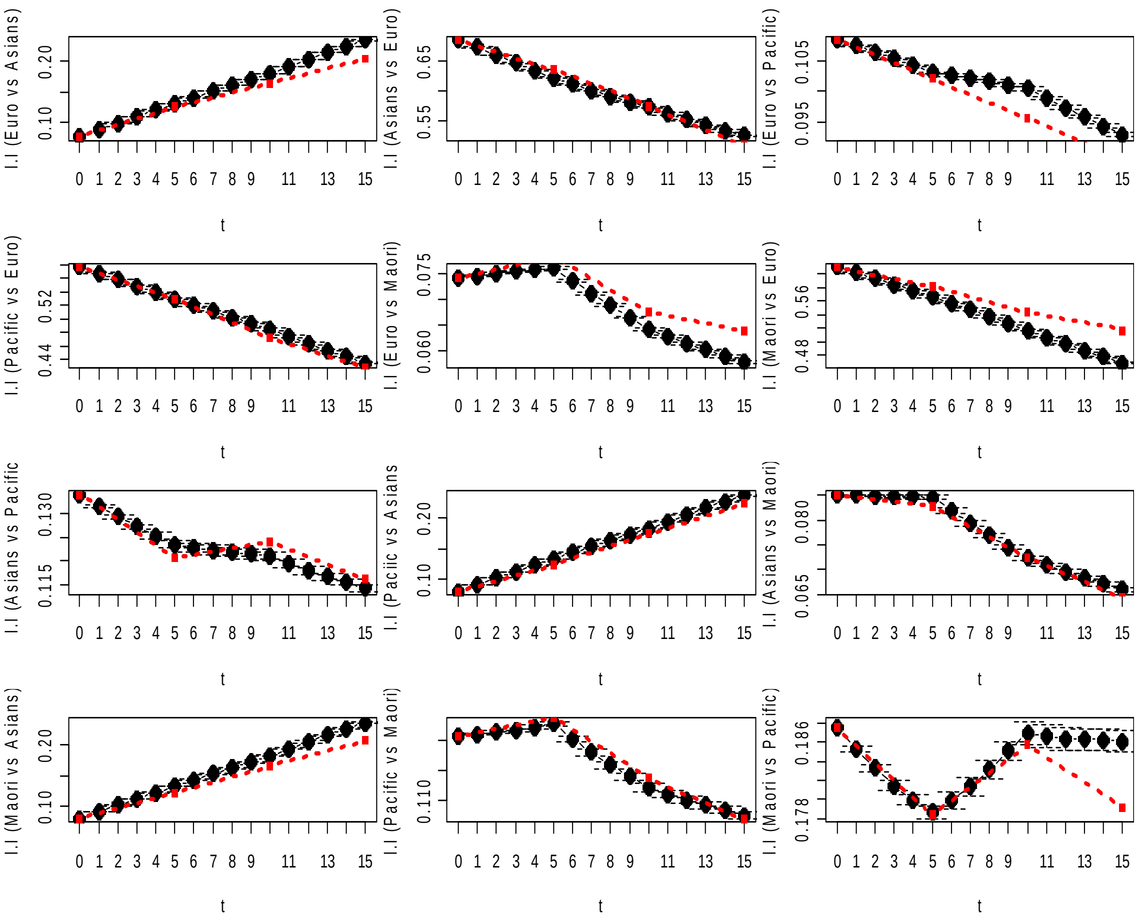
PI-values (MKAU)



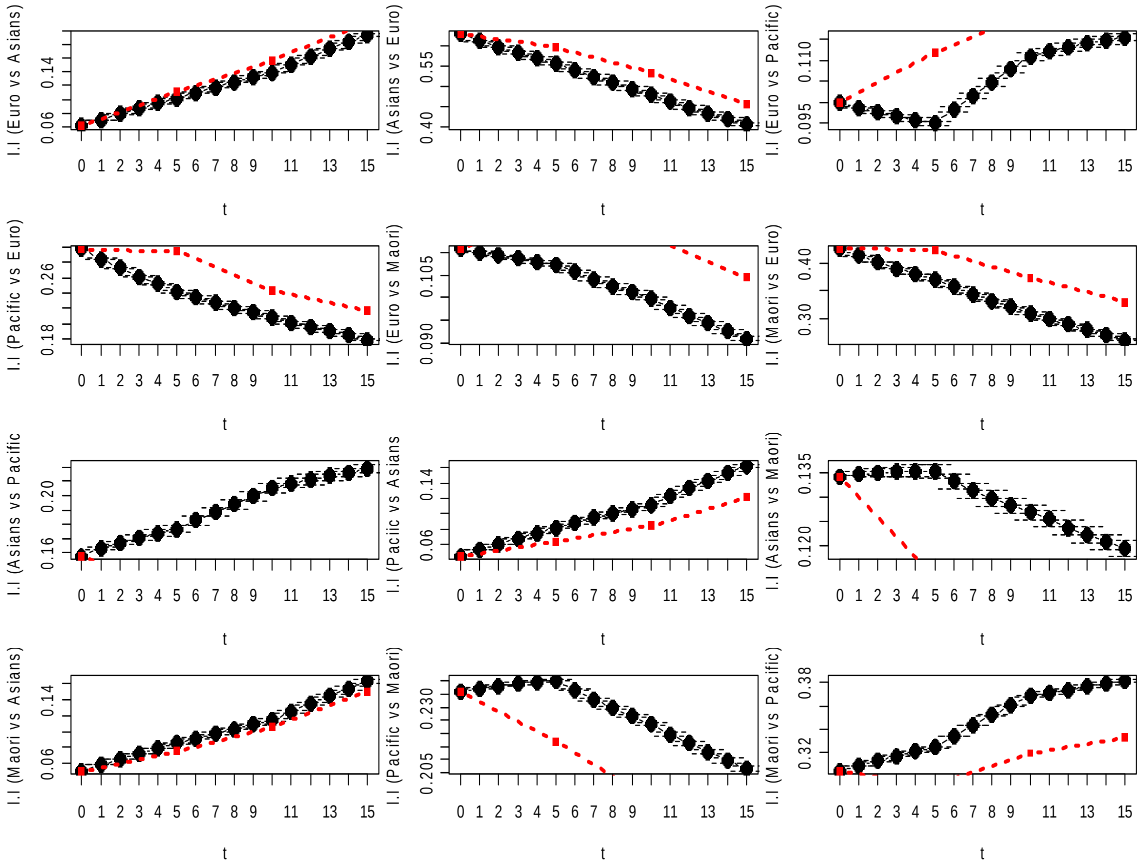
PI-values (WTKE)



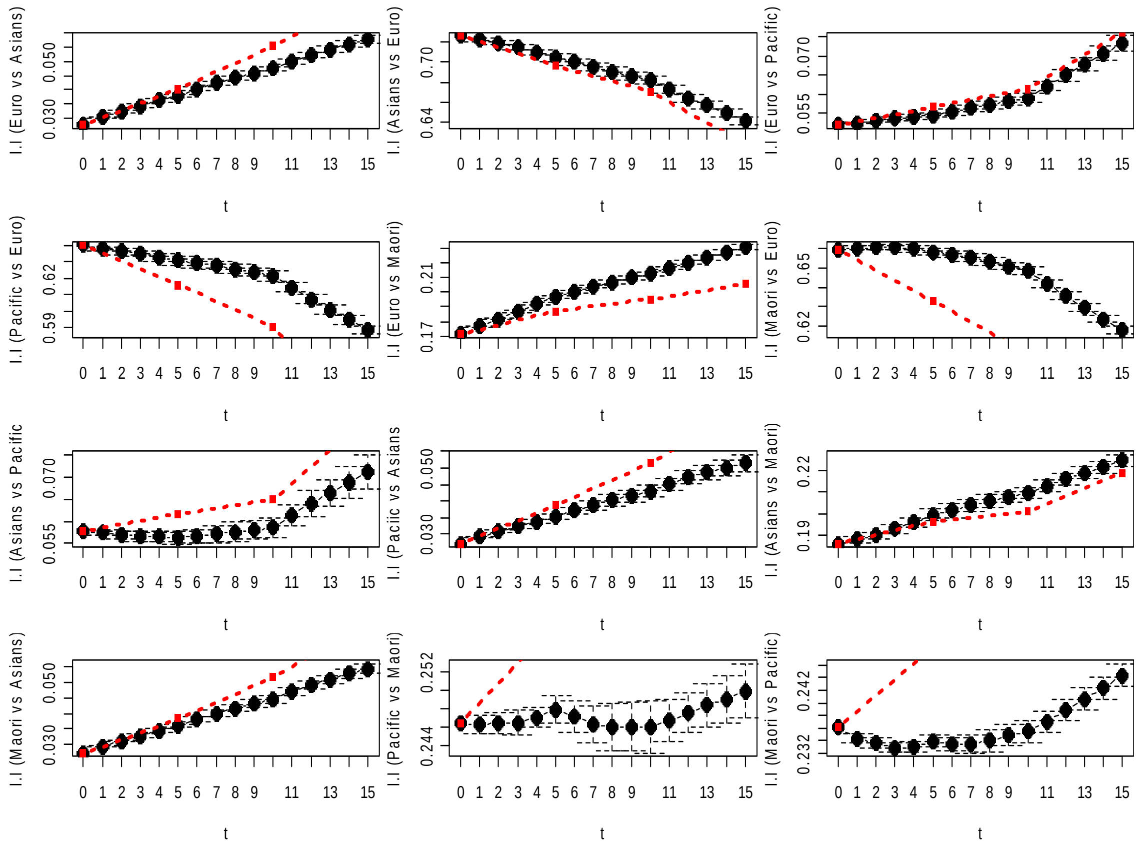
II_{xy}-values (NShore)



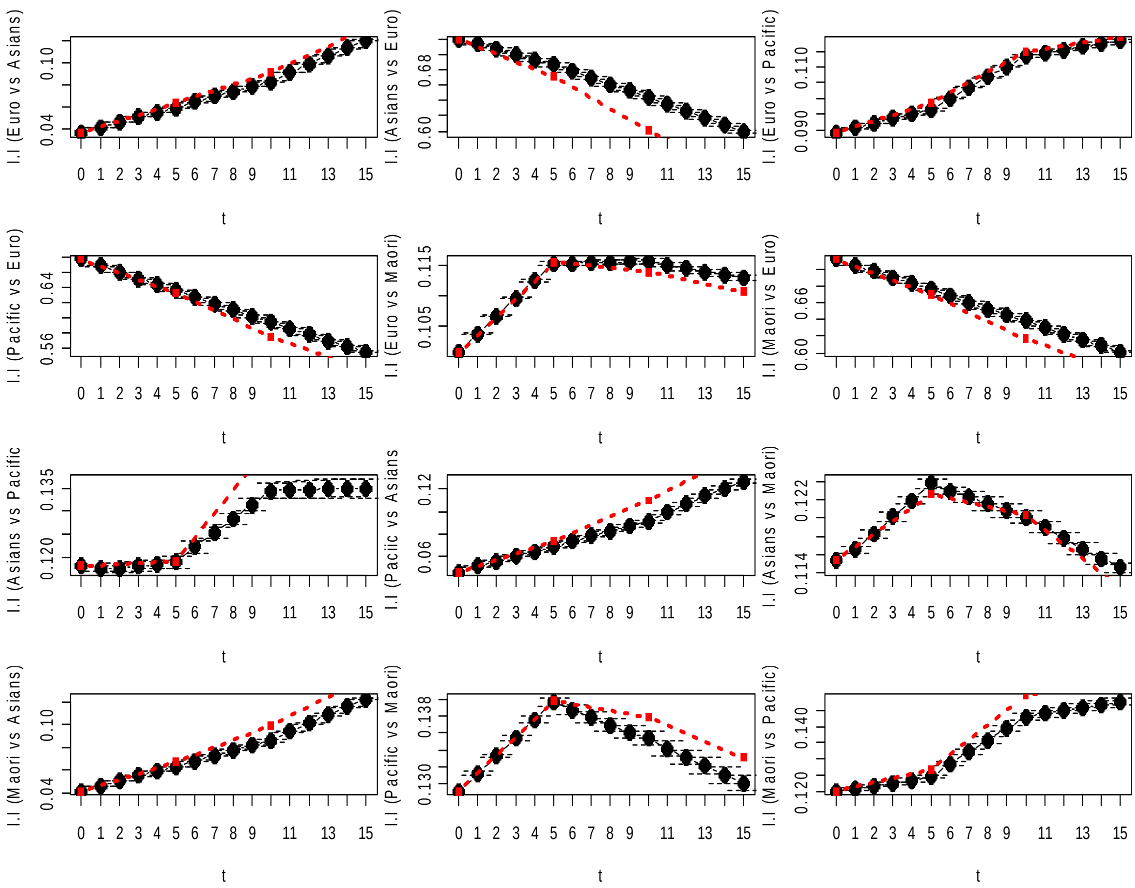
II_{xy}-values (AKL)



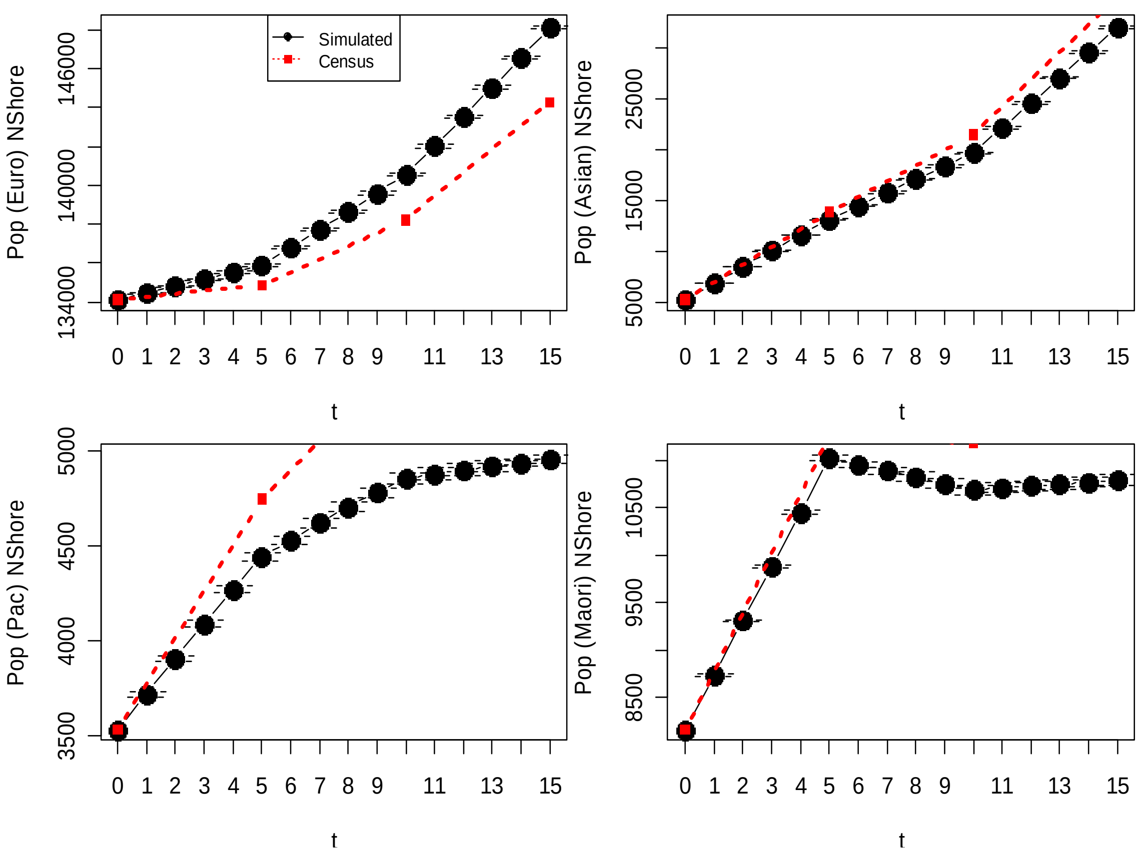
II$_{xy}$-values (MKAU)



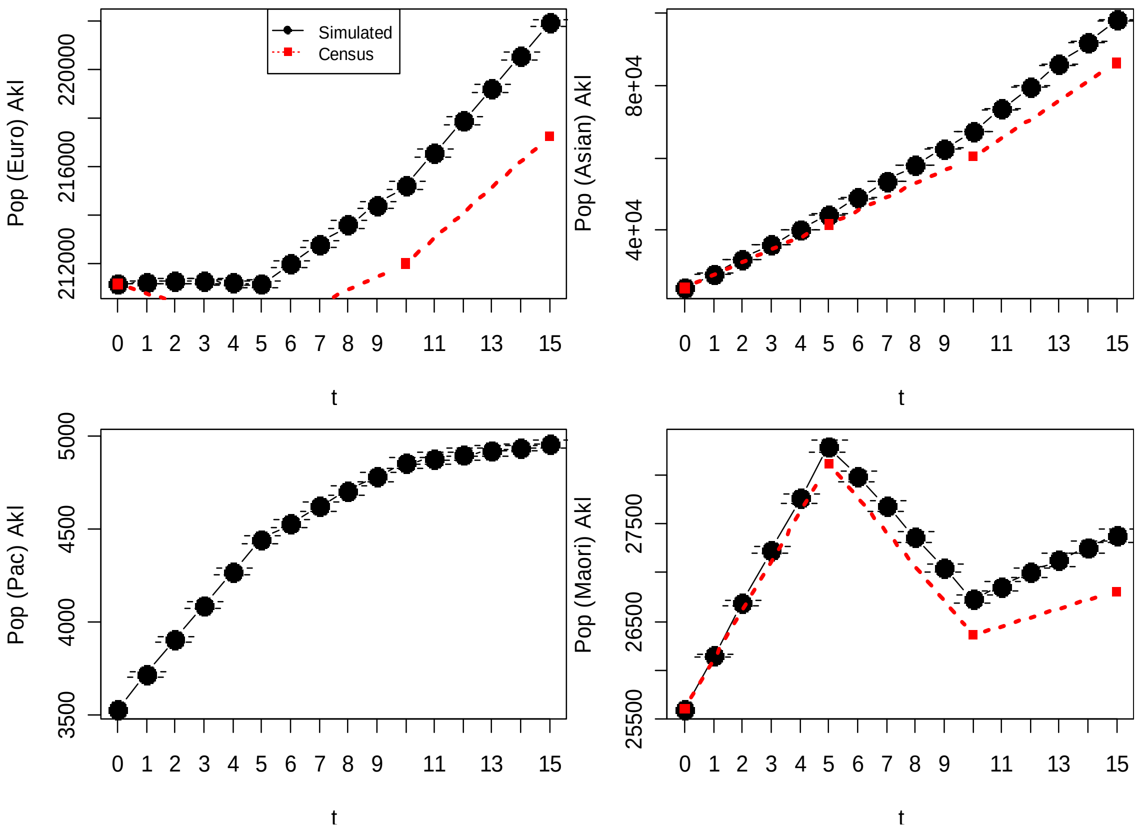
II$_{xy}$-values (PAPK)



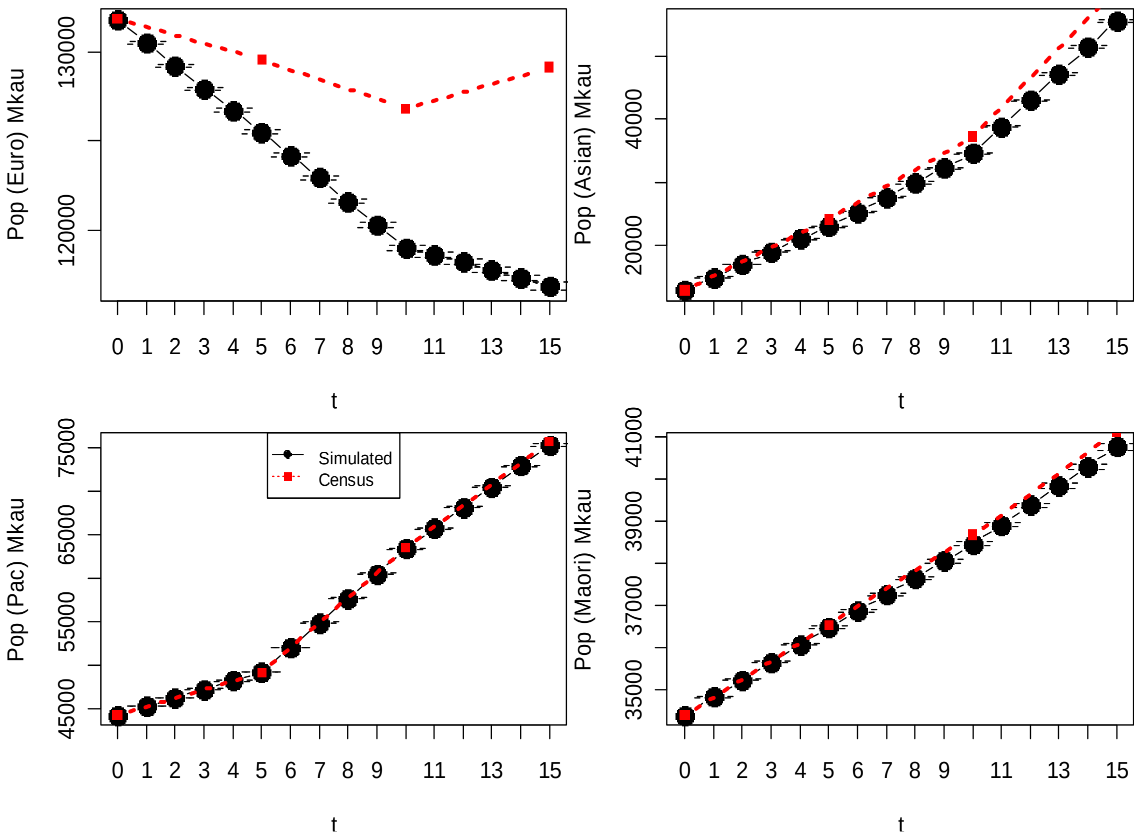
II_{xy}-values (WTKE)



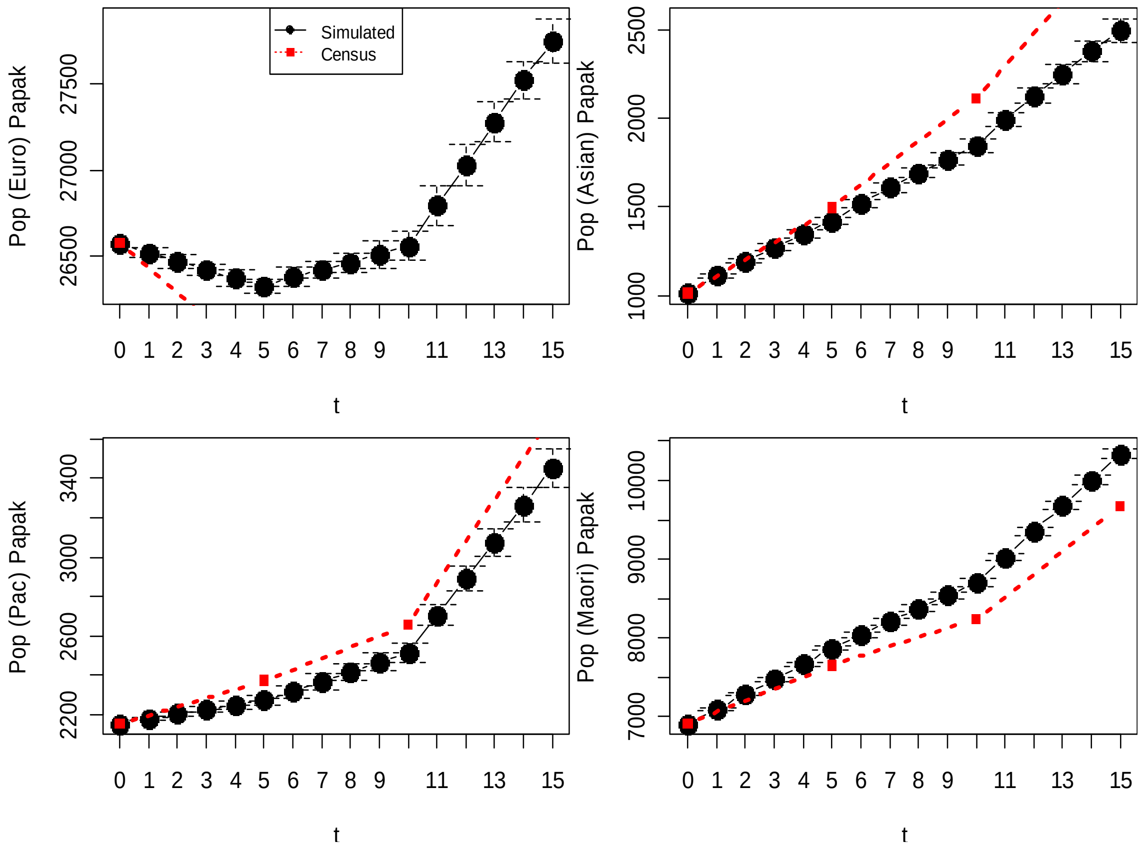
Pop (NShore)



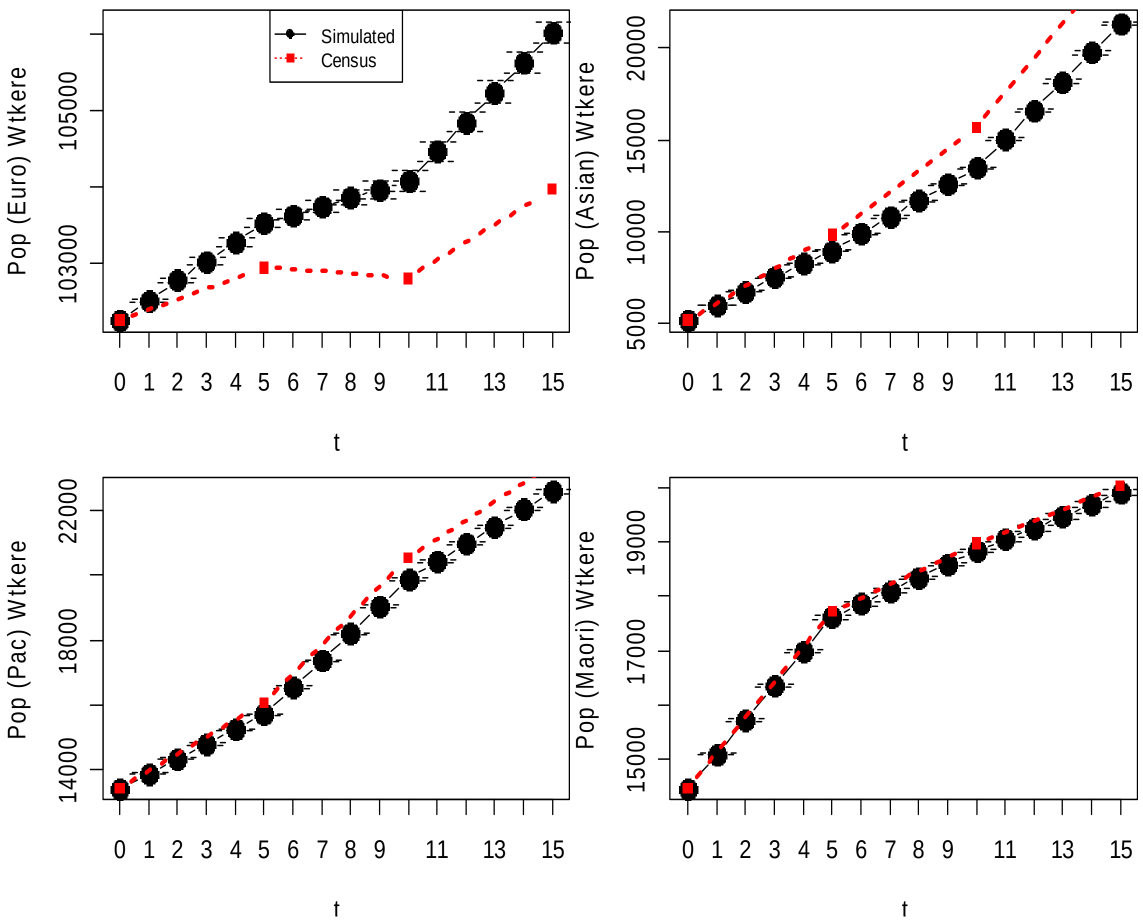
Pop (AKL)



Pop (MKAU)



Pop (PAPK)



Pop (WTKE)

Appendix I: Accompanying CD-ROM Content

I.1 *HAAMoS* Model

I.1.1 Source Files

This section contains the source codes (java files) of the model.

I.1.2 Javadocs

This section contains *javadocs* documents (html format) of the model.

I.1.3 Shapefiles and Parameter Files

This section contains shapefiles for both five and seven territorial authorities (TAs), as well as those used in Schelling and Fossett experiments.

I.1.4 Installation/ User Guide

This section contains materials on the installation and usage of the model.

I.2 Data

This section contains census data (e.g. PPD, TBD, mobility, vacancy, etc.).

I.3 Data Analysis

This section contains the R source codes.

I.4 Required Programs

This section contains the programs needed in order to install and run the model. These include Repast 3.1, JDK, Eclipse and additional libraries. These can alternatively be downloaded from the websites of their respective developers.