University of Newcastle
School of Geography, Politics and Sociology

AN INFORMATION STATISTICS APPROACH TO ZONE DESIGN IN THE GEOGRAPHY OF HEALTH OUTCOMES AND PROVISION

by

Konstantinos Daras

December 2005

SUBMITTED IN ACCORDANCE WITH
THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

The candidate confirms that the work submitted is his own and that appropriate credit has been given where reference has been made to work of others.

ABSTRACT

Social scientists and policy makers are usually faced with boundary changes in administrative areas over time. The control of spatial issues deriving from boundary changes is even more important when they affect the organisation and allocation of resources in a national health system. In addition, the problem becomes more acute when the health organisations analyse sensitive data using geographies constructed to serve other administrative purposes. In recent literature, the modifiable nature of areas is reflected in the modifiable areal unit problem (MAUP) and widely acknowledged frameworks for geographical analysis are developed targeting to overcome this problem. The aim of this research is to suggest and develop methodologies supporting the health related studies to provide valuable decisions.

In order to achieve this aim the following research objectives have been developed. In this thesis, the crucial objective is to identify how geographical problems are related to health policies exploring available methodologies and suggesting solutions derived from informative statistic measures to unresolved practical issues. Consequently, an automated computer system developed formulating these problems in graph theory context and utilising their components through object oriented algorithms. The test and evaluation of the system is applied in a series of case studies investigating the effects of MAUP in various geographies and aggregation levels. The overall objective provides strategies and valuable practice for using the system as well as suggesting areas of health research that may benefit from the methodology. In the final chapter, the thesis concludes with a summary of findings and limitations for the suggested methodology providing an outline of the research directions for further work into the spatial issues in relation to health research.

ACKNOWLEDGEMENTS

A PhD thesis is a highly personal achievement, but it would never be possible without interaction and support of other people. First of all, I would like to thank my supervisor, Dr. Seraphim Alvanides, who has been an excellent supervisor throughout my studies. I would like also to thank Prof. Chris Brunsdon for his perfect quote "*stop writing programming language and start writing English language*" and his valuable advice as co-supervisor at my early years of my PhD study. Many thanks go to Dr. Ian Gregory for the historical datasets analyzed in Chapter 7. I acknowledge also the support of the Hellenic State Scholarship Foundation (HSSF) for the financial support throughout my research.

In addition, I would like to give many thanks to my friend Nikolaos and Stamatis for their patience and understanding throughout my studies as well as my friends from the AIKIDO club in University of Newcastle for their professional dancing figures during Joe's courses. Unfortunately, there were people that affected most from my research journey in Human Geography and I am referring to my family and friends in Greece. I would like to thank all of them and each one individually for their enduring support all these years.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

Introduction

1.1 Introduction

Geographical boundaries and places have appeared to cause confusion and arguments since the evolution of civilisation. In *Physics book IV* by Aristoteles, "…a *'place'* may be assigned to an object either primarily because it is its special and exclusive place, or mediately because it is *'common'* to it and other things, or is the universal place that includes the proper place of all things", while in his work *On the heavens*, Aristoteles defined the relationship between place and boundaries mentioning that "place means the boundary of that which encloses it" (Lukermann, 1961 p.201). Although Aristoteles' concept is relative in the way he describes his thoughts, what becomes clear is that place and boundaries were of great importance to be understood throughout human history. This is also highlighted in the health related studies by Hippocrates in 400 B.C. where the crucial role of 'place' in epidemics was discussed introducing the effects of natural boundaries in health incidents (Buck and PAHO, 1988).

Today, we still challenge ourselves to explain our surrounding environment and to a greater extent to control it by building natural and conceptual boundaries. In this attempt geographers show the way to understand the world but since the "quantitative revolution" in Geography in the 1960s and the reaction to the "qualitative revolution" since the 1970s, the academic discipline seems to be divided into two groups. For this reason, since the 1960s, debates concerning the dominance of one group against another have taken place in the research community (Johnston et al., 2003; Hamnett, 2003a; Hamnett, 2003b). What is becoming increasingly obvious, though, is that quantitative and qualitative approaches are two sides of the same coin. Both approaches elaborate theoretical and empirical agendas in which, theoretical research is of major importance because without theory the empirical analysis can be directed into mere empiricism, while theoretical research without related empirical analysis can provide mere theoreticism (Johnston et al, 2003; Hamnett, 2003a; Hamnett, 2003b; Unwin, 2005). In this thesis, we aim to balance the theoretical and empirical aspects of entirely quantitative approaches investigating the effects of boundaries in health research.

As a result, concepts related to geographical space and boundaries should be based on strong theoretical and empirical grounds. The study of boundaries and their enclosures is a vital geographical task, because it encapsulates our perceptions and representations of the real word, though these undoubtedly influence how we try to understand our environment.

1.2 Context and Objectives

Social scientists and policy makers are usually faced with boundary changes in administrative areas over time. The control of spatial issues deriving from boundary changes is even more important when they affect the organisation and allocation of resources in a national health system. For example, the research community has often it criticised the methodologies adopted from the Department of Health (DoH) in the UK to support the targeting of funding for primary health care. The problem becomes more acute when health organisations analyse sensitive data using geographies constructed to serve other administrative purposes. In geographical literature, the modifiable nature of areas is reflected in the modifiable areal unit problem (MAUP) and widely acknowledged frameworks of geographical analysis are developed in order to overcome this problem. Although there is a respectable literature concerning the MAUP effects and methods to tackle them in geographical analysis, the research of MAUP effects in health related studies is limited to a small number of publications mainly focusing on one side of the problem: the scale effect. As a result, useful findings that could derive from different aggregations of areal units are not properly investigated.

The aim of this research is to suggest and develop methodologies that will support health related studies to provide valuable decisions. In order to achieve this aim the following research objectives have been developed. In this thesis, the crucial objective is to identify how geographical problems are related to health policies by exploring available methodologies and suggesting solutions derived from informative statistic measures to unresolved practical issues. Consequently, an automated computer system will be developed formulating these problems in the graph theory context and utilising their components through object oriented algorithms. Testing and evaluation of the methodology will be carried out in a series of case studies, investigating the effects of the MAUP for various geographies and aggregation levels. The overall objective is to provide strategies and valuable practice for using the system, as well as to suggest areas of geographical health research that may benefit from the methodology.

In order to highlight the main targets of this thesis a summary of the above research objectives has been put together:

1. Examine how spatial problems associated with areal units are treated in geographical research related to health issues.
2. Formulate these problems using up to date theoretical concepts and techniques such as graph theory and object-oriented design.
3. Develop methodologies for tackling the identified problems in the form of an automated computer system.
4. Implement methods for measuring the most informative aggregation level in a zone design concept.
5. Test and evaluate the proposed methods in three case studies, targeting different scale levels and health related issues.
6. Provide potential areas of health policy that may benefit from the methodology, while suggesting further research improvements.

1.3 Thesis Structure

This thesis consists of eight chapters that can be further organised into three major sections; a schematic representation of the thesis organisation is presented in Figure 1.1 illustrating how the sections and chapters are linked. In Chapter 1 of the first section, we introduce research areas that this thesis aims to investigate, referring to our main objectives. A presentation of the whole thesis structure is also introduced, presenting the research directions where this study intents to focus. An extended literature review is undertaken in Chapters 2 and 3. In Chapter 2 an overview of the National Health Service (NHS) organisation in England and Wales is discussed, highlighting the frequent reforms of health administrative areas that involve rearrangement of their boundaries. In addition, one of the most important health issues faced by developed countries today, health inequalities, is briefly examined here, as a preamble to further discussion in Chapter 7. Furthermore, commonly used health and deprivation indices are presented discussing the advantages and disadvantages of using such indices for developing public health policy. On the other hand, the limited use of GIS and spatial methods in NHS bodies is discussed, making a case for their utilisation in health strategies. In this thesis, Chapter 2 describes the structure of public health in the UK and common utilities of monitoring the public health, while in Chapter 3 we investigate how the factor 'place' affects various health related studies. For a comprehensive look of spatial implications in health research, in Chapter 3 we discuss the role of boundaries in selected health related studies. In addition, we explore three widely used techniques: classification, clustering and regionalisation, for managing spatial features. Finally, the ecological fallacies and spatial difficulties such the MAUP effects that influence the results of health related studies are presented, focusing on areal unit related aggregation processes.

The second section consists of Chapter 4, where the issues raised in Chapter 2 and 3 are tackled in a new aggregation system based on zone design principles (Openshaw, 1984). In Chapter 4, the zone design system is gradually formulated borrowing concepts from the graph theory, while it is implemented by means of object oriented organisation. In particular, advanced contiguity methods are implemented here overcoming boundary related issues, while the selection of appropriate scale in a study area is statistically defined by measuring the goodness of fit of a model in relation to the number of its parameters (Akaike Information Criterion). In addition, a collection of zone design components is developed providing different methods of initial aggregation and securing the contiguity stability of each output zone. The latter methods of controlling contiguity stability highlight improvements in the form of faster methods for speeding up time consuming processes. New optimisation methods focusing on the homogeneity improvement of zones are also presented describing the spatial modelling process of each algorithm. In the final part of the chapter, we introduce four techniques that aim to assist zone design system in constructing more shape constraint output solutions.

The third and final section of the thesis consists of three case studies used to evaluate the performance of the zone design system implemented in Chapter 4. In each case study, we develop health related methodologies for tackling issues identified in the literature and later implemented in Chapter4. As shown in Figure 1.1, each case study aims to explore the characteristics of these methodologies at a different aggregation level highlighting the importance of a zone design system in health related studies. In this thesis, the selection of different scales was decided according to the crucial effects of the scale parameter in both geographical and health related datasets.

The first case study (Chapter 5) deals with the investigation of a statistical informative aggregation level for exploring the relationships between the traffic accidents experienced by children aged 0-17 years old and deprivation determinants in Tyne and Wear. To identify a suitable scale level in terms of goodness of fit we adopted the minimisation of Akaike Information Criterion methodology as suggested by Nakaya (2000) and further developed in Chapter 4. As a result, the case study uses the zone design system to provide two different zoning solutions focusing on the homogeneity of zones in terms of the children traffic accident rates and the Townsend Index. The novelty of this case study lies in the fact that the aggregation level of analysis is provided by advanced statistical measures of goodness of fit. In addition, the detailed level of analysis provides a useful mean for identifying the effects of the MAUP in communities with similar socioeconomic characteristics.

The second case study (Chapter 6) addresses the problem of zone design in a health policy related context. The suggested methodology reconstructs “Health Authorities” for England and Wales using census datasets in 1991 and 2001. In this case study, all the wards in England and Wales are aggregated based on three objective functions providing homogeneous output zones by optimising the Limiting Long Term Illness (LLTI) incident rates. To evaluate the new zoning solutions, the case study compares the different aggregation levels and output zones investigating the relationship between LLTI incidents and deprivation determinants.

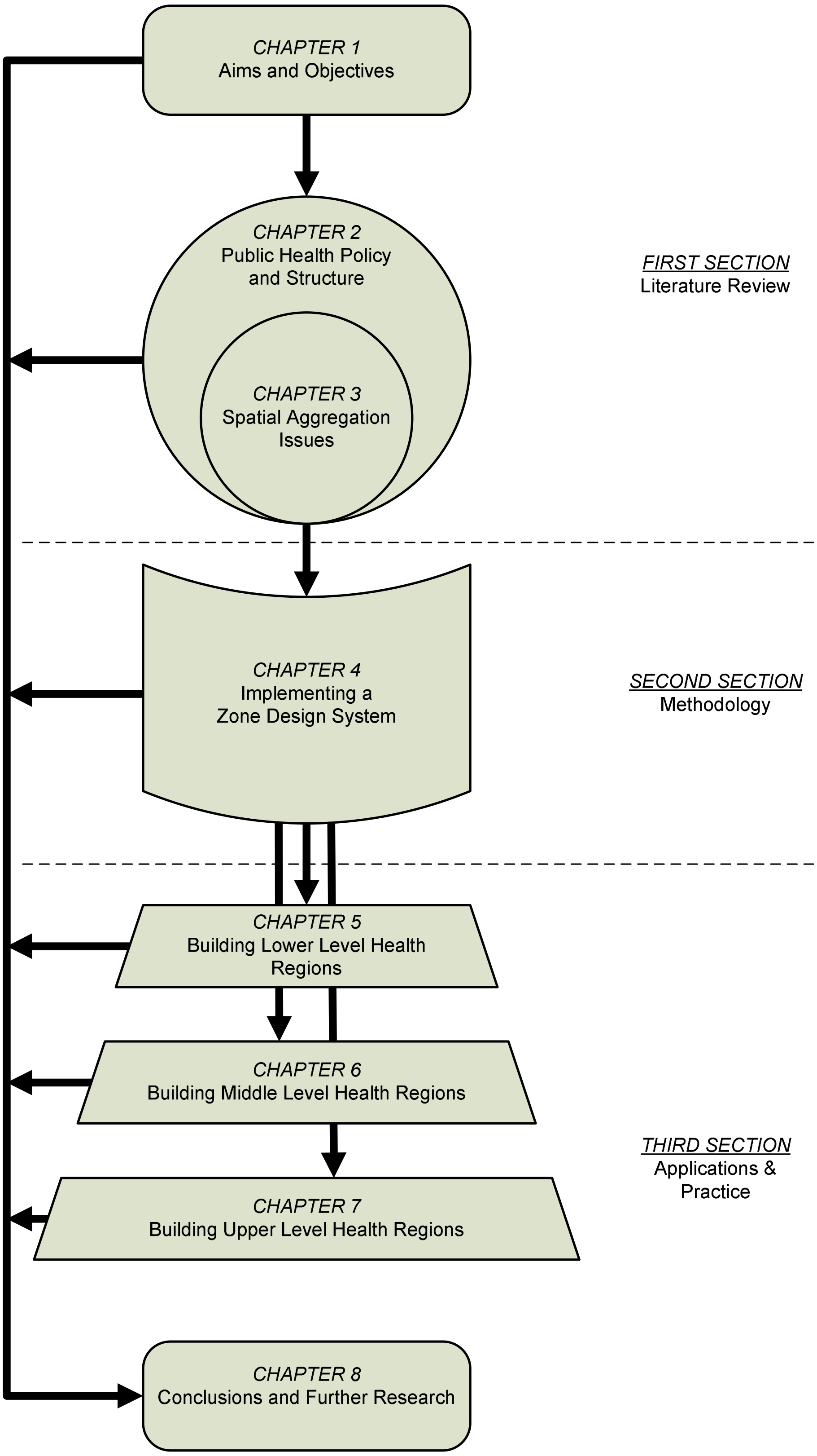


Figure 1.1: Schematic overview of the thesis structure

The final case study (Chapter 7) tackles a research question related to the analysis of spatial patterns over time. Using four representative dates in the history of infant mortality and the Local Government Districts in England and Wales, we design new output zones equivalent to the county level, aiming for homogeneous zones of infant mortality rates and similar base population. The four inequality measures have assisted in analysing all available administrative geographies and the new zoning solutions investigating the infant mortality and health inequalities throughout the period 1911-1971. This methodology highlights the contribution of zone design in explaining health related patterns in time series datasets.

The third section concludes with Chapter 8 presenting a critical evaluation of the research findings derived from all three case studies. In addition, a summary of the available components of the new zone design system is presented, highlighting the methods developed in this thesis. In the final part of the chapter, directions for further development of the zone design methodology are provided, suggesting possible combination of zone design concepts with multilevel techniques in relation to health research.

CHAPTER 2

Public Health Policy and Structure

2.1 Introduction

The first outbreak of cholera stroke Britain in 1831 and a series of epidemics emerged in 1849 and 1854 hitting the most populated towns like London (Champion, 1993; Snow, 1849). These outbreaks forced the British government, under the Poor Law Amendment Act in 1834, to establish the Poor Law Unions for recording vital health data. The Poor Law Unions became the authorities for compulsory registration of births, deaths and marriages. The registration started three years later and very soon the Unions observed many health phenomena taking place in their territory, such as high infant mortality (Freeman, 1968).

Between 1841 and 1851, Britain's population increased by 13.8%, while a remarkable growth in towns with special resources and locational characteristics was revealed. For example, towns classified as "watering places", "manufacturing", "mining and hardware" and "seaports" saw their population increasing by 25.6%, 23.8%, 23.4% and 21.9% respectively (Freeman, 1968). Overcrowding in towns caused social problems and a decline of their public health profiles. In 1848, the first Public Health Act set up a General Board of Health as a central authority, with the power of creating local boards following a ratepayers' petition or where death rates exceeded 23‰ (Smellie, 1946). Many boroughs turned to sanitary authorities under the Public Health Act and reserved their old status. The Poor Law Unions were subsequently divided into urban and rural sections, according to the Public Health Act of 1872.

A comment about the mutability of administrative areas during the period from 1830s onwards was made at the 1927 enquiry from a witness on behalf of the Ministry of Health, "*Personally, I do not want any questions of town planning complicated by any* *question of borough extensions*" (Freeman, 1968). Obviously, during that period borough changes were an extremely difficult task and most were completed according to political or empirical acts. Opposition by a person or a social group in policies concerning change of a social status is something that we face even today. Changes in administrative areas affect people's lives and, even if these changes are socially accepted the financial implications a government is called to tackle are costly and politically sensitive. Notwithstanding the financial barriers, the quality and quantity of health services till nowadays are not equally distributed in space. Often, urban populations receive better services than rural populations, thus increasing the health inequalities among areas (Haynes, 1987).

On the other hand, 'Airs, Waters, Places' written by Hippocrates in 400 B.C. (Buck, 1988) reminds us the importance of the factor 'Place' in health for the last twenty four centuries. Throughout his empirical work, Hippocrates reported that health incidents are strongly related to the surrounding environment. Today, the relationships between the health and the environment of people can be explained in more detail than ever. However, it is vital for any government to monitor the health picture of citizens and target areas of need, using appropriate methods and analytical tools.

2.2 UK Health Administrative Areas

The National Health Service of the UK was set up on 5 July 1948 under a very weak economy (Levitt and Wall, 1984). The wounds of the Second World War were still warm and many side effects such as under nutrition, limited building materials, shortage of fuel and housing in companion with the economic crisis in the USA predetermined a rough environment for the establishment of the new health system. Since 1948, the main aims of the NHS have been to increase the physical and mental health for all citizens by promoting health, preventing ill-health, diagnosing injury/disease, treating injury/disease and caring for those with long-term illness and disability, based on the need of citizens rather than their ability to pay. The NHS is accountable to the Parliament and managed by the Department of Health (DoH). The DoH supports the government in improving the health picture of the population in consultation with the

Health Authorities. The NHS was first structured based on three administrative divisions; the executive councils (117), the hospital boards (49) and the local health authorities (152). The overall coordination and planning at the regional level were undertaken by regional hospital boards, of which there were 14 in England and Wales and 5 in Scotland as shown in Figure 2.1.

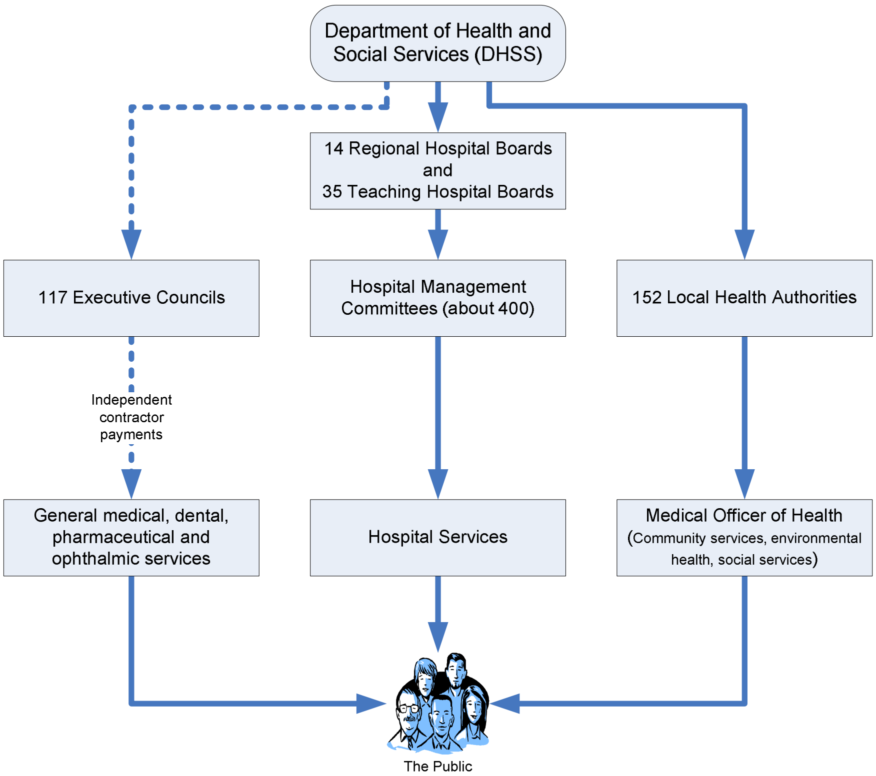


Figure 2.1: Organisation of health services in England in 1948.
(Source: Open University, 1985: p.5, as cited in Haynes, 1987)

These three branches of the health services caused critical coordination problems highlighting its administrative inconvenience. The lack of communication between the general practitioners, hospitals and community health services had unpleasant effects on the patients' health services. In 1974, increasing demands for restructuring of the health services came under consideration resulting in the 1974's reorganisation of the health system. The new administrative structure consisted of three hierarchical levels: regions, areas and districts (Figure 2.2). The Regional Health Authorities (RHAs) were established to provide, advise and carry out policies at the regional level. The population of each RHA ranged between 2 and over 5 million. RHAs were divided into area health authorities, whose geographical boundaries corresponded to the counties and local government districts in metropolitan counties. As Haynes (1987) argues, there was a debate concerning the use of geographical units that reflect the local distribution of ill health, but the advantage of having common boundaries for both health services and local governance prevailed in the end.

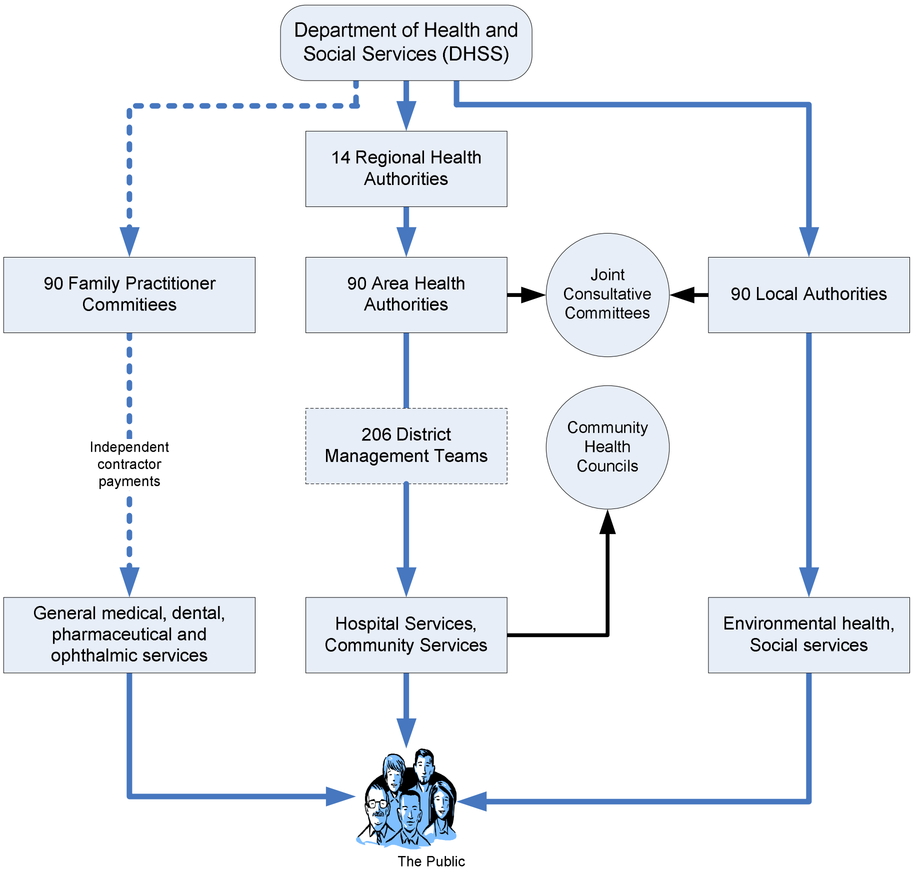


Figure 2.2: Organisation of health services in England in 1974.
(Source: Open University, 1985: p.7, as cited in Haynes, 1987)

Over the last two decade the NHS administrative organisation has consisted of two or three strategic levels and various changes resulted in different allocation of roles and responsibilities. From 1982 to 1996 a structure of Regional and District Health Authorities (RHAs and DHAs respectively) existed in England, but they were abolished by the Health Authorities Act in 1995 (HMSO, 1995). They were replaced by Regional

Offices and Health Authorities, but they have also been abolished and replaced by a new threefold structure of 8 Regional Offices (ROs), roughly 100 Health Authorities and 481 Primary Care Groups as shown in Table 2.1.

Twenty years after the establishment of Regional and District Health Authorities in 1982 the government in consultation with the Department of Health decided to restructure the NHS organisation. The lifted NHS according to the "NHS Plan" is implemented to simplify the communication between the NHS and patients, as highlighted by the Department of Health.

> "In the future, care and treatment will be redesigned around their [patients'] needs"
> (DoH, 2000)

Table 2.1: Changes of the NHS organisation structure since 1982 (England and Wales only).

	1982 to 1996	1996 to April 1999	April 1999 to March 2002	April 2002 to June 2003	July 2003 onwards	
Level I	Regional Health Authorities (14)	Regional Offices (8)	Regional Offices (8)	Directorates of Health and Social Care (4)	N/A	England
	NHS Wales Department (NHSWD)	NHS Wales Department (NHSWD)	NHS Wales Department (NHSWD)	NHS Wales Department (NHSWD)	N/A	Wales
Level II	District Health Authorities (192)	Health Authorities (~100)	Health Authorities (~100)	Strategic Health Authorities (~100)	Strategic Health Authorities (28)	England
	Health Authorities (5)	Health Authorities (5)	Health Authorities (5)	Health Authorities (5)	NHSWD Regional Offices (3)	Wales
Level III	N/A	N/A	Primary Care Organisations (481)	Primary Care Organisations (~481)	Primary Care Organisations (300 + 3)	England
	N/A	N/A	Local Health Groups (22)	Local Health Groups (22)	Local Health Boards (22)	Wales

The new structure for health administration on 1 July 2003 consists of 28 Strategic Health Authorities (SHAs) for England and 3 Regional Offices (NHSWD) for Wales which are constituted from 303 Primary Care Organisations (300 Primary Care Trusts and 3 Care Trusts) and 22 Local Health Boards (LHBs) respectively. The Primary Care Trusts (PCTs) and LHBs work with local authorities and other agencies that provide health and social care locally and they are responsible for understanding the needs of their local community. According to the NHS Plan the PCTs and LHBs are now at the centre of the NHS and receive 80% of the NHS budget. As they are the organisations closer to the citizen, the target of this plan is an effective health and social care system with less bureaucracy and rapid financial support for the needs of the local community.

But how can this plan tackle the socio-economic or geographical variations and inequalities in health care? A positive view of these issues has been noted by Macara (2002). He argues that the processes of decision-making are becoming more overt and more widely shared than in the past, as a result of the system's paternalism replaced by partnership in health care. For example, the Health Action Zones (HAZs) are partnerships between the NHS, local authorities, the voluntary and private sectors and local communities (DoH, 2002). HAZs were introduced in England in 1998/99, based on the 26 areas of highest deprivation and poor health. The aim was to identify new methods for tackling health inequalities and modernizing services, with the expectation that these methods would provide the national standards for the whole country. Geographically, most of the HAZs were based on health authorities covering over a third of the population of England (Bauld et al., 2000). HAZs were intended to last for at least five years, with the possibility of a seven-year life, although in late 2002 it was announced that funding would come to an end in March 2003. As a result, the original aims of extensive look at health policy issues and development of appropriate methodology were narrowed down to a four-year pilot program subject to key political interference (Halliday and Asthana, 2005).

2.3 Inequalities in Health

During the 1980s and 1990s, inequalities in health widened while the UK government has repeatedly expressed its intention to tackle these inequalities (Townsend and

Davidson, 1992; Whitehead, 1992). According to two reports, "Inequalities in Health" (HMSO, 1998) and "Our Healthier Nation" (HMSO, 1999), following the restructuring of the NHS (DoH, 2000) the UK government has set health policy strategies at both the national and the local levels aiming to reduce the health inequality. In terms of the national targets, by 2010 the gap in life expectancy between the five health authorities with the lowest life expectancy and the population at national level should be reduced by at least 10%. In addition, the gap in rates of infant mortality between lower social groups and the rest of the population needs to be reduced by at least 10%. At local level, the Health Improvement and Modernisation Plan (HIMP) partners should identify potential targets or outcomes for reducing health inequalities. The HIMP partners consist of NHS bodies, health, city, county and district councils, advisory groups and voluntary organisations. Moreover, the report "Our Healthier Nation" set the reduction of four indicators by the end of 2010: accidents by 40%, suicides by 20%, cancer by 20% and coronary heart diseases and strokes by 40%. The accomplishment of the health inequality targets needs to be measured with appropriate methods. In the Health Divide report (Whitehead, 1992) suggested three measurement requirements for an efficient inequality measure, as first proposed by Wagstaff et al. (1991). The measure should reflect the socioeconomic dimension to inequalities in health and the experience of the entire population and also be sensitive to changes in the distribution of the population in different socioeconomic groups. Furthermore, Whitehead (1992) identifies the need for accurate health data and effective monitoring tools for measuring inequalities and for allocating resources.

Nonetheless, recent studies uncover the existence of widening health inequalities in the UK utilising methods with strong theoretical ground. For example, Shaw et al. (2005) stated that analysing recent data for life expectancy (2001-3), inequalities between the whole of England and the five local authorities with the lowest life expectancy have actually increased. Their analysis employed two widely acknowledged inequality measures: the Gini coefficient for income inequality and the Slope Index of Inequality (SII). The Gini coefficient measure is the ratio of the area under the Lorenz curve to the area under the diagonal on a graph of the Lorenz curve (Lorenz, 1905). The SII is given by the regression coefficient in a simple regression of the health indicator of a socioeconomic group (e.g. ward) on the ranking of that group according to a deprivation indicator (Low and Low, 2004). Moreover, Shaw et al. (2005) suggested a misrepresentation of health inequality targets as they compare the worst groups with the national average rather than considering the entire distribution.

Tackling health inequalities implies the adaptation of appropriate methods for monitoring local and national gaps. As Low and Low (2004) argue, most social researchers are concerned with inequalities between regions ignoring the within regions health gaps. In their study, they measure the local health inequalities of Primary Care Trusts (PCTs) and Local Strategic Partnerships (LSPs) in the Sunderland area suggesting the existence of health gaps. On the other hand, at the national level Shaw et al. (2005) state that any improvement of living standards for the poorest social groups does not seem to reduce health inequalities as long as wealth inequalities have continued to grow.

2.4 Common Indices for Public Health Policy

The improvement of the social profile of a society on subjects such as health care, poverty, deprivation, social inequalities and income distribution whilst controlling costs is one of the most crucial tasks of a government. A government policy for improvement of living standards takes into account numerous socio-economic indicators. Many researchers for governmental or research purposes have developed and they still develop a variety of deprivation indices attempting to capture and measure deprived rural or urban areas in need. As Bartley and Blane (1994) argue, deprivation indices measure the proportion of households, usually in small geographical areas, with low living standards or high need for services. In addition, they suggest that it is important for deprivation indices to be understood and evaluated in terms of the purpose for which they are being used and the validity in which the socioeconomic characteristics are represented. Most deprivation indices are calculated for geographical areas, rather than individuals, and they tend to be based on purely statistical methods, overlooking spatial parameters. In same cases, such statistical methods have resulted in policies for sensitive areas with mixed populations, but failing to target the deprived people in wealthier wards (Townsend et al., 1989). In addition, most deprivation indices are constructed using decennial census data. As a result, the deprivation measures represent deprived areas in the year of the census, but the indices are less reliable as the census data becomes dated.

Despite the above limitations, deprivation indices are very important measures for public health policy as long as same spatial issues are considered. In the research community, the use of deprivation indices is widely accepted with an extensive range of applications according the relevant literature. For example, Dummer et al. (2000) based their research on deprivation scores from census data to examine the increasing inequality of stillbirth risk with social class and deprivation in England and Wales. Their investigation focused on two census levels: enumeration districts in Cumbria and county districts in England and Wales. Using the Townsend, Department of the Environment (DoE), and Jarman scores, they suggested that inequality in stillbirth risk in Cumbria has fallen significantly since 1966, while at the country level there was significant inequality in stillbirth risk in all time periods. In the UK, the Neighbourhood Renewal Unit (NRU) identifies deprived neighbourhoods based on an index of multiple deprivation (DETR, 2000) while Lee et al. (1995) suggested the Breadline Britain index for identifying the most deprived wards at the national level. The Breadline index estimates statistical weights for the component indicators providing an easily interpretable index of deprived areas.

As Sloggett and Joshi (1994) reiterate the point of Townsend et al. (1989) they note that "*for maximum effectiveness, health policy needs to target people as well as places*". Obviously, lack of interest to people or places could influence policy decisions, thus aggravating existing social problems. For example, in 1995 the deprivation payments to general practitioners by the UK DoH suffered from various limitations, including under enumeration: under-counting of homeless people and refugees in the deprivation measures for census wards resulted in misallocation of resources (Majeed et al., 1996). Similarly, Crayford et al. (1995) estimate the changes in deprivation payments made to general practitioners, by calculating the Jarman index for the smaller geographical unit of the census enumeration district, rather than ward. They suggest that the Jarman index could be more sensitively and appropriately applied to calculate the deprivation payments that practices receive using the census enumeration districts. Furthermore, Mackenzie et al. (1998) indicate that each organisation has preferences for different indices highlighting possible influences in the resource allocation decisions. They argue that the Townsend Material Deprivation Score is used commonly by Health Authorities while the Department of Health tends to use the Jarman underprivileged area score for allocation of additional payments to general practitioners.

The Jarman Underprivileged Area Score (UPA8), described as "*probably the most ubiquitous*" (Lee et al., 1995 p.23) index, was developed to measure the general practice workload and to support the targeting of funding for primary health care (Jarman, 1983; Jarman, 1984). It is developed based on a survey of GPs' subjective expressions of the social factors most likely to affect their workload. The score consists of eight variables which are individually weighted by a sample of London GPs. The weights of the eight variables are 3.34 for unemployment, 2.88 for overcrowding, 3.01 for lone parents, 4.64 for children under 5 years old, 6.62 for elderly living alone, 2.50 for ethnicity, 3.74 for low social class and 2.68 for residential mobility. However, Jarman index has been criticised as an index for detecting inner-city deprivation because it includes factors such as overcrowding and ethnicity (Davies, 1998). Talbot (1991) has extended this criticism by stating that the index is strongly biased towards London in the proportion of the population classified as deprived. As a result the index fails to deprivation in the north of England as the Thames regions benefit at the expense of peripheral regions.

On the other hand, the Townsend index (Townsend, 1987) is an unweighted combination of four census variables measuring material deprivation. The four variables target the lack of access to good housing (overcrowded conditions), the shortage of material possessions, the lack of access to car and unemployment. It was originally constructed from the 1981 Census datasets using the above variables to represent income, wealth, living conditions and material resources of the population. The Townsend index is a summation of standardised scores for each indicator. Although all the variables are represented as percentages the unemployment and overcrowding indicators are transformed using the logarithmic transformation $F(x) = ln\ (x + 1)$ to provide as normal as possible distributions. Townsend scores greater than zero indicate high material deprivation, while scores lower than zero represent the more affluent areas. Until recent years the Townsend index was a popular choice for researchers exploring morbidity and mortality in different social groups; for example, Townsend et al. (1989) investigated health inequalities in the north of England. Townsend scores can be reconstructed for areas of interest using the relevant variables extracted from the 1991 or 2001 census tables. However, it should be noted that the 1991 and 2001 indicators do not represent to the same degree the underlining deprivation because of changes in many social characteristics. For example, in 1981 the use of car ownership as an income indicator could reflect different social classes, but in later censuses car ownership is a weak indicator as a result of car availability even for underprivileged people. Furthermore, Adams et al. (2004) promote the use of Townsend index for enumeration districts instead of wards for measuring the payments to general practitioners. Nevertheless, the Townsend index is considered the best and most effortlessly implemented indicator of material deprivation currently available.

Furthermore, numerous indices have been developed attempting to capture the social profile of citizens. In table 2.2, a collection of the most popular indices is presented together with the indicators for each index. Generally, these indices can be categorised into three groups: those for investigating the health needs of the population; indices of deprivation; and those for identifying social exclusion (Tunstall and Lupton, 2003). In addition to the Jarman index, the Carstairs 1991 and Arbuthnott 1999 indices are also used for measuring health needs, developed with the Scottish health data and characteristics.

The Carstairs index was developed for the analysis of Scottish health data and it was based on four variables taken from the 1981 Census. The aim of this index was to measure material deprivation throughout Scotland using the census data. The Carstairs index is very similar to the Townsend index in terms of the selected indicators as three of the indicators (unemployment, no car and overcrowded households) are the same. The only difference is the use of social class instead of housing tenure used in the Townsend index. Carstairs and Morris (1989) state the social class provides a better indicator of material deprivation than housing tenure, arguing that people of a lower social class and also unemployed, reflect those with limited material resources (Carstairs and Morris, 1989). Moreover, they argued that housing tenure was less relevant in Scotland as the proportion of housing in public sector was higher than in England (Morris and Carstairs, 1991). However, the Carstairs index has been widely used to identify the relationships between deprivation and health (Carstairs and Morris, 1990).

Another health related index, the Arbuthnott index was developed by the Scottish Executive Health Department (SEHD) to provide an effective formula for distributing funding between health authorities in Scotland (SEHD, 2000). This index is based on four indicators: standardised mortality ratio (SMR) among people under the age of 65; unemployment; proportion of the elderly population claiming income support; and households with two or more indicators of deprivation (such as unemployed or permanently ill head of household, overcrowding, large households, lone parent families, and households of elders). The standardisation of indicators is achieved using the z-score method, while all the indicators have equal weight.

Additional indices focusing on deprived areas have been developed the last decades in the UK. For example, the DoE 1981 and 1991 indices were created by the Department of the Environment. Their main difference from the previous indices is the use of actual numbers instead of percentage or rates. As a result, areas with small numbers of cases are less weighted than larger areas providing less reliable measures. More recently, the Office of the Deputy Prime Minister (ODPM) developed more complex indices like the Index of Multiple Deprivation (IMD) 2000 and 2004. Using the IMD 2000, Adams and White (2005) argue that the geographical proximity of patients to general practitioners is greater in deprived than affluent wards in both rural and urban areas. Subsequently, the IMD 2000 has been updated to the new IMD 2004 following the release of the latest census denominators. The IMD 2004 contains seven domains which relate to income deprivation, employment deprivation, health deprivation and disability, education, skills and training deprivation, barriers to housing and services, living environment deprivation and crime (ODPM, 2004a; ODPM, 2004b). Each of these domains is weighted 22.5%, 22.5%, 13.5%, 13.5%, 13.5%, 9.3% and 9.3% respectively. The IMD 2004 is the most recent deprivation index in the UK developed at the Super Output Area (SOA) level. The SOAs are the latest 2001 census geographies and they are aggregation products of the lowest census geography, the Output Areas (OAs).

Although socioeconomic data and deprivation indices have been extensively used to monitor human activities and experiences, the extensive use of socioeconomic data for producing unified deprivation scores has been criticised by Smith et al. (2001). They support that indices of social fragmentation such as the Congdon measure (Congdon, 1996; Whitley et al., 1999) can relate better to suicide mortality than a generic deprivation index. Thus, different area based indices deriving from socioeconomic characteristics can reflect to a different extend mortality rates from different causes. Moreover, they state that area based indices should be developed, based on strong theoretical ground as well as related to specific health problems. Although Smith et al. (2001) support the development of particular indices for specific health issues, they fail to acknowledge potential problems deriving from the area boundaries.

2.5 GIS in Health Organisation and Health planning

The geographical analysis of health variation and its characteristics has been widely acknowledged for its importance in public health. The Geographical Information Systems (GIS) can enlighten health policy issues by mapping health phenomena and helping in understanding local patterns of disease. With respect to the geographical dimension, national surveys (in 1991 and 2001) observed an increase of uptake and use of GIS within all Health Authorities/Health Boards and Primary Care Organisations (PCOs) (Higgs et al., 2003; Smith et al., 2003). Higgs and Gould (2000) note that studies undertaken in the past confirmed that GIS technology was not really launched in the NHS. However, the uptake and use of GIS within NHS organisations is concentrated at the Heath Authority level, while few PCOs have personnel using GIS (DoH, 2001b; DoH, 2001c). The growing awareness of the value of GIS (Moon and Gould, 2000; More and Martin, 1998) and the recognition of the importance of spatial analysis in health care (Bryant, 1998) are leading the NHS to a new information era (DoH, 2001a).

The use of spatial analysis and GIS to identify local areas for primary health care planning has been discussed by Bullen et al. (1996). Their research used visualisation tools in GIS to identify nested hierarchies of localities for the management of primary health care in West Sussex. Their primary data consisted of five geographical layers: official boundaries, neighbourhoods, key service points, natural boundaries and population flows to GP practices. Using these layers, they subjectively aggregated the study area based on two different hypotheses: first by constraining the aggregation process with DHA boundaries; and second by relaxing the constraints in advance of the population flows to GPs. In addition, they constructed localities utilising a *k*-means cluster procedure with optimum patient flows directed to GP practices. Certainly, the study by Bullen et al. shows the potential of GIS in health regionalisation while Higgs and Gould (2001) mention the importance of using GIS in strategic health planning contexts referred by most of health-related studies.

In summary, GIS is an important tool for exploring spatial characteristics of disease incidence and relating them to socioeconomic determinants. Its ability to highlight problems related to particular georeferenced health data such as the selection of scale, small numbers problem, modifiable areal unit problem (Openshaw, 1984), ecological fallacies (Jones and Duncan, 1995) and appropriate geographical aggregation for spatial analysis (Alvanides and Openshaw, 1999) may benefit health care planning. Therefore, researchers working in Public Health should be aware of these issues relating to health data when investigating health variations and deciding on allocation of health resources.

2.6 Discussion and Conclusions

This chapter introduced three major topics related to public health: changeability of health administrative boundaries from 1834 to date, use of indices and other measures to monitor deprivation and inequalities, and importance of GIS tools in health policy. We selected the above topics as representative areas for identifying possible issues related to health policy debates. Although, in the health community the importance for more objective and efficient methods of handling health area changes is acknowledged, the complexity of boundary definition and the issues resulting from the changes of health administration areas over time suggest further investigation. Indices and inequality measures provide valuable estimators for health planning, however, the limited consideration of boundary effects and the use of administrative areas designed for other purposes instead of health policy should be further investigated in a spatial context.

On the other hand, GIS tools can be an asset for identifying and tackling health related spatial issues. A number of studies, mostly developed by academics, have used GIS to study disease patterns, health variations and to explore possible factors of underlining map patterns. For example, Openshaw et al. (1987) used the Geographical Analysis Machine (GAM) to study leukaemia clusters. GAM is a spatial analytical tool that investigates the study area for hidden clusters measuring cases placed within circles of varying size and position. Sabel and Gatrell (1998) provided a kernel density estimation tool to model patterns for patients with motor neurone disease in Lancashire and Cumbria. In addition, Diggle et al. (1990) adapted the k-function to investigate the expected cases within a given distance from a landfill site. However, the majority of epidemiological studies analyze health data based on administrative areas, such as county districts (Stiller and Boyle, 1996) and census wards (Dickinson and Parker, 1999). Therefore, it is pertinent to explore thoroughly the available aggregation methods and the issues derive from their relation with the public health. In the next chapter, issues related to aggregation processes are presented, discussing applications and methodologies developed in the health research community.

Table 2.2: Indicators used in major indices of health need, deprivation and social exclusion

	Index	Income							Economic					Housing					Health					Education						Other								
		No car	*Low income*	*Home rented*	*Low social class*	*GDP*	*Income inequality*	*Consumption*	*Unemployment*	*Economic inactivity*	*Non-earning*	*Employment*	*Unfilled vacancies*	*Home overcrowded*	*Home lacking in amenities*	*Unsuitable home*	*Mortgage arrears*	*Vacant homes*	*Mortality*	*Limiting long term illness*	*Health behaviour*	*Low birth weight*	*Mental health*	*Qualifications*	*16-17s not in education*	*School exclusions*	*Literacy*	*Numeracy*	*Other measures*	*Household composition*	*Crime*	*Ethnicity*	*Social interaction*	*Derelict land*	*Access to services*	*Deprivation*	*Mobility*	*Savings*
Health Need	**Jarman**				■				■					■																■		■					■	
	Carstairs 1991	■			■				■					■																								
	Arbuthnott 1999		■						■										■																	■		
Deprivation	**Townsend**	■		■					■					■																								
	Green 1994	■		■					■	■														■														
	DoE 1981			■					■					■	■															■		■						
	DoE 1991	■	■						■	■				■	■	■			■					■	■						■			■				
	ILD 1998	■							■					■	■																							
	IMD 2000								■	■	■					■			■	■				■	■						■				■			
	SADI 1988		■											■										■						■								
	Breadline	■		■	■				■											■																		
	Scotdep	■			■				■					■																								
	Scotdep 1998	■	■						■		■			■	■			■	■			■		■	■						■							
	SocDep		■						■											■										■								
	MatDep	■												■	■																							
	CASE	■		■	■				■											■										■								
	Bradford	■	■						■		■			■	■															■		■						
	Oxford			■					■	■				■																		■						
Social Exclusion	**Burchardt et al.**		■	■					■																								■					■
	Rahman et al.		■				■	■	■			■		■			■		■	■	■	■	■	■	■	■				■	■		■					
	DSS 1999		■								■	■				■			■		■			■	■		■	■										
	Crompton, Blair					■			■	■			■							■																		

Source: Tunstall and Lupton (2003), Appendix 1: Table A1, pages 30-31.

CHAPTER 3

Spatial Aggregation Issues

3.1 Introduction

While in Chapter 2 we presented issues arising from the changeability of health administrative areas in the UK, in this Chapter we explore some valuable concepts concerning the dynamic nature of boundaries which affect the majority of health studies, to a certain extent. In addition, the interference of boundaries with underlying health or socioeconomic patterns is discussed presenting three analytical techniques and representative health related empirical studies for each approach. Taking into account the aggregation scale, we provide an overview of existing ecological fallacies related to individuals and groups of people, while the modifiable areal unit problem is explained in both its scale and zoning effects. Subsequently, widely used aggregation methods are discussed highlighting the advantages and disadvantages of each method. Finally, we stress the spatial issues that affect health studies, while providing additional suggestions for appropriate methodologies in the context of automated aggregation.

3.2 Boundaries, places and patterns

3.2.1 Boundary definition and boundary types

Boundaries, from a geographical point of view, are formed by lines on a map which reflect areas (regions) with common characteristics. At the same time boundaries carry information about the kind of border separating areas. These regions are usually identified in terms of their human and physical particularities such as ethnicity or climate. For example political boundaries are very important as they define regions under the control of a governmental organisation. Often political boundaries cross physical boundaries so that regions (or counties, or cities) share access and control of natural resources or apply specific policies according the local needs. Moreover, social boundaries are more difficult to be created because usually these types of boundaries apply restrictions to resources and activities.

However, boundaries do not only define regions but also provide information about them. For example a boundary could be a physical entity such as a river or a mountain (Figure 3.1.b) that really disconnects regions and human activities. On the other hand, a boundary could join regions and link activities between people such as a bridge or a road (Figure 3.1.a). A boundary connecting two regions while separating two other regions is a more complex form of structure. For example, a rail line connects two towns but is likely to split various smaller areas on its crossing (Figure 3.1.c).

Boundaries do not only represent the connectivity (or lack of it) between two or more regions, they also specify strong or weak associations derived from the studied research question. Although, administrative boundaries could take various forms, their existence is result of human mannerism in which the human mind tries to conceptualise its surrounding environment applying natural or subjective limitations.

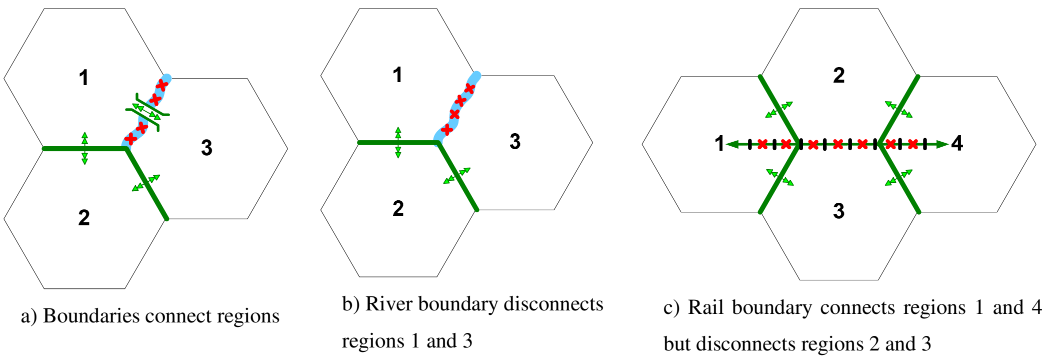


Figure 3.1: The threefold nature of boundaries

The structure of boundaries has fascinated geographers to develop administrative boundaries for public and governmental use. The development of administrative boundaries according to the principles of Hierarchical Spatial Reasoning by Eagleson *et al.* (2002) and their later study on the incorporation of differences in the urban and rural regions (Eagleson *et al.*, 2003) are examples that identify problems associated with boundary systems. Similarly, studies by Mineter (2003) who studied vector topology under parallel processing and by Theobald (2001) who noted the advantages and disadvantages of topological and non-topological structures attempted to minimise the processing time and simplify topological structures. An interesting and unique research during this decade was attempting to compare local-level migration and commuting flows for data collected in the 1981 and 1991 British censuses overcoming the changes through the time of the boundaries in the small areal units by Boyle and Feng (2002). Moreover, Norman et al (2003) recognise the importance of boundary specification and they highlight the lack of research on difficulties created by unstable '*zonal*' boundaries over time. Undoubtedly, boundary changes are as crucial as the Modifiable Areal Unit Problem, the 'ecological fallacy' and other varieties of geographical or social issues.

3.2.2 The realisation of boundaries and place in health

Geographers have long recognised the importance of boundary specification and their dynamic nature. Likewise, over the last decade in the spatial epidemiology research a considerable increase of spatial analysis studies have been attained in spatial epidemiology research, but have been little attention paid to boundary effects.

One close look of the boundary effects has been taken by Lawson et al. (1999). They examine boundary effect problems within case events and tract count data according to two schemes. The schemes are applied to the example of mortality from gastric cancer in the Tuscany region of Italy. The results suggested a better reflection of the underline structures and they support the need of taking into account the boundary effects in disease mapping. An interesting example of boundary and pattern analysis in health science is presented by Jacquez and Greiling (2003) comparing patterns in standardised morbidity ratios, calculated from New York State Department of Health data, to geographic patterns in overall predicted risk from air toxics using the USEPA National Air Toxics Assessment database. In their research, pattern recognition played a crucial role for delivery of valuable results and the boundary changes have been analysed according the local boundary and sub-boundary analysis (Oden et al., 1993) and the boundary overlap analysis (Jacquez, 1995).

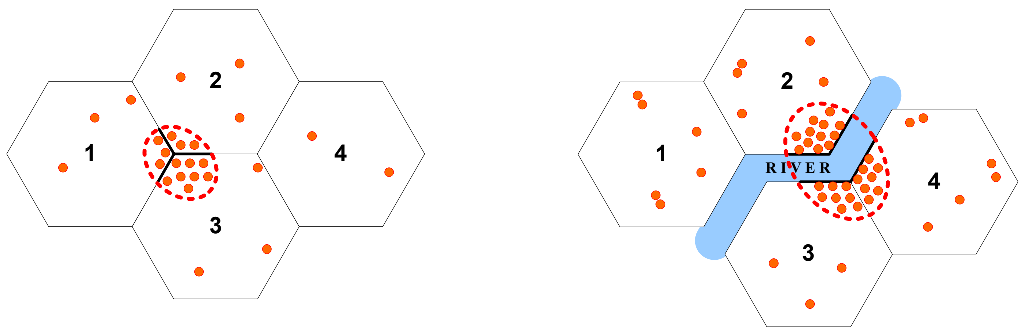
Moreover, it has been argued that "tying data to points based on grid references rather than fixed areas allows greater flexibility in managing boundary changes" (McVey and Baker, 2002). On the other hand, the boundary-free representations, such as population density surfaces, provide considerable advantages compared to zonal models (Martin, 1996), while different types of representation such as squares and hexagons are able to tackle problems referring to points falling on the gridlines (Alvanides et al., 2001).

However, it would be interesting to mention at this point, the close relationship between the administrative boundaries and the place effects. The place effects are contextual or ecological factors that influence individual sensitivity to disease. The article by Duncan et al (1998) illustrates a generic framework that can be used to tackle a number of different questions of interest to health researchers. As Duncan et al. (1993) and Boyle and Willms (1999) suggest place effects associated with large administrative areas are relatively small and scientists should be cautious about using administrative boundaries as sampling frames for testing hypotheses about place effects.

3.2.3 The dynamic nature of boundaries and health patterns

Administrative boundaries are governmental products that are created for supplementary aid in policy making and resource allocation. In contrast to administrative boundaries, physical boundaries usually increase the complexity and difficulties for social policy applications. Although, the natural boundaries are visible and their changes are easily recorded, the administrative boundaries are most of the time invisible entities and their changes are hard monitored. In addition, the dynamic nature of these two types of boundaries effect considerably the society lifestyle and they are close related to environmental and social patterns such as affluence or deprivation.

In health sciences, it is important for a researcher to capture the clinical picture of a study area and to perform analysis for the implications of boundary structures. For example, a health pattern of an affected area according to postcode information can fall between two or more regions (Figure 3.2.a), so when the postcode information is transferred to the regional level the pattern of incidents mislays its intensity. In addition, the physical borders are important to be analysed because clusters of interest are often separated by physical elements (Figure 3.2.b). A river boundary with an industry installation on the river's bank could be a serious factor for an illness outbreak at both sides of the river. As mentioned by Gatrell (2001) overlooking such boundary effects and focusing on the administrative regions can mislead researchers and produce biased results.



a) Hidden pattern across the boundaries of regions 1, 2 and 3

b) Hidden pattern separated by a river

Figure 3.2: Natural or administrative boundaries often split health patterns.

However, the discovery of health patterns using administrative regions is important in a policy making context, when the clinical picture of the population studied is less changeable than an illness phenomenon. For this reason, every decade the census collects information in national level and updates the data with annual population projections. Undoubtedly, at the end of decade period the errors concerning population projections are reaching the maximum value and produce less reliable information to the research community.

3.3 Classification, Clustering and Regionalisation

In this section, three closely related analytical techniques are introduced: classification, clustering and regionalisation. Often classification is improperly referred as clustering because both methods are used in many scientific fields under many different names. This section clarifies the differences between classification and clustering providing a brief explanation of their uses, focusing on the medical and geographical sciences. In addition, an investigation in the concept of regionalisation is presented and the procedures are compared in relation to their spatial elements.

Before examining the three techniques, a number of definitions are necessary to form a basic terminology as they are discussed in literature (Alvanides, 2000; Grigg, 1965). First, some sort of *unit* needs to be specified as the basic entity for data collection. A unit may represent information of an individual person (object, animal) or a group of individuals (group of objects or animals). It is also defined in different scientific fields as *operational taxonomic unit (OTU)*, *object*, *data unit* or *observation*. A unit can be represented as a *spatial*, *tabular* or even *conceptual object*. The different unit types according to their geographical aspect are:

- *Aspatial* unit: defined by non-locational information. An aspatial unit is usually appeared as tabular information that describes this unit.

- *Spatial* unit: defined by geographical and non-locational information. A spatial unit may take the form of *positional* (point) or *areal* (polygon) unit according to the object being represented.

Although, in aspatial units the tabular information could become extremely complicated and difficult to manipulate, the research attention of this thesis is upon the spatial units, focusing on areal units. In Figure 3.3, an abstract description of spatial units is illustrated. The spatial units are represented either as positional (points) or as areal (polygons). Both spatial representations are possible to group in larger units. In terms of positional units the grouping process is equivalent to cluster analysis providing groups of points with specific characteristics. On the other hand, areal units can be grouped into larger units (*zones*) and can refer to either compact areas or non contiguous areas. In this study, the use of term zone is assigned to contiguous grouped areal units. Furthermore, the grouping approach of areal units is usually based on cluster analysis, while zone design procedures are used for grouping zone units respectively. A discussion on existing zone design systems will follow in later sections of this chapter exploring their limitations and highlighting their advantages.

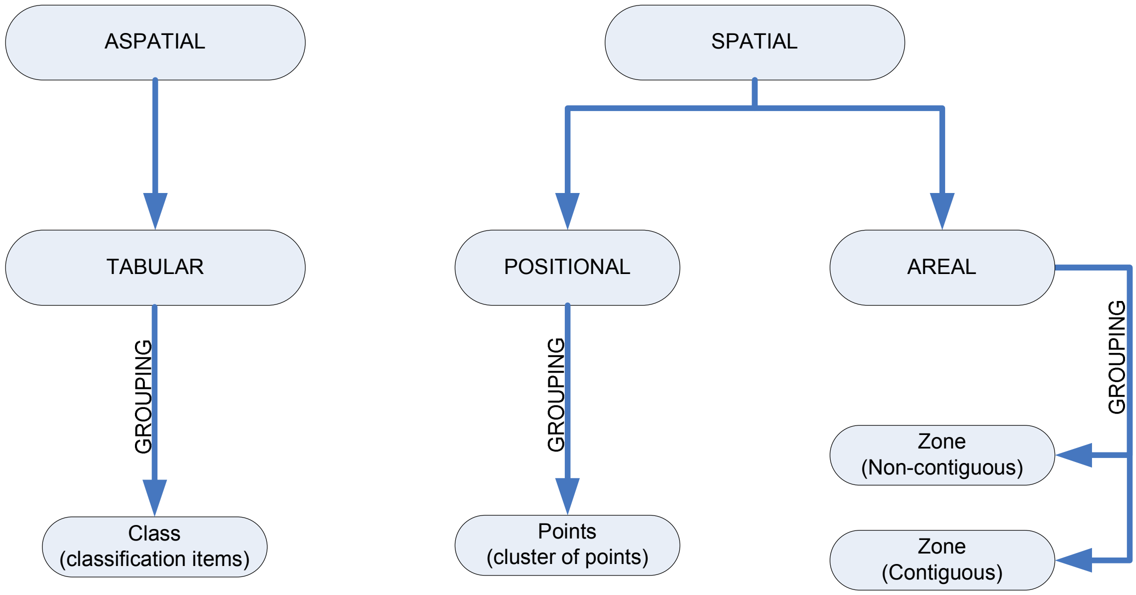


Figure 3.3: The different unit types according to their geographical aspects

3.3.1 Classification

Classification as defined by Anderberg (1973:p2) is "the process or act of assigning a new item or observation [tabular unit] to its proper place in an established set of categories [classes]". In a classification process, the essential attributes of each class are known in advance. This means that the researcher identifies the appropriate classes in advance and the classification process associates any given unit to the selected classes. A simple example of classification is introduced in Figure 3.4. There are six colourful items and three boxes; each of a single colour (yellow, red, green). The classification task here is defined as the process of assigning the items according to their colour to the same coloured box. In addition, if there was a white coloured item for the same assigned boxes then the classification procedure could be adjusted according to a new rule such as the brightness of item.

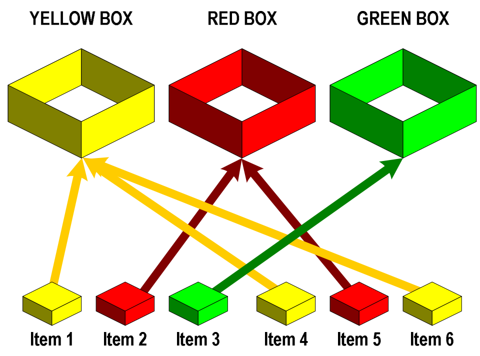


Figure 3.4: Classification example

Formal classifications of animals and plants date back to the Greek philosopher Aristotle in 4th century BC, but a formal categorisation was introduced by Linnaeus in 1753 and it is the basic structure of modern classification systems. A comparison of classified units in species and geographical taxonomies is presented in Figure 3.4. The Figure 3.5 shows a hierarchical classification for both taxonomies providing a more general description of the species and houses as the level of hierarchy is increased such as country and kingdom respectively. The computational advantage of the nineteenth century improved the taxonomy standards in various disciplines emerging several studies focused on classification improvements. Grigg (1965) considered classification as a necessary preliminary in most sciences and Harvey (1969) suggested classification to be used as a tool for realisation and understanding of the study problem but not as an end in itself.

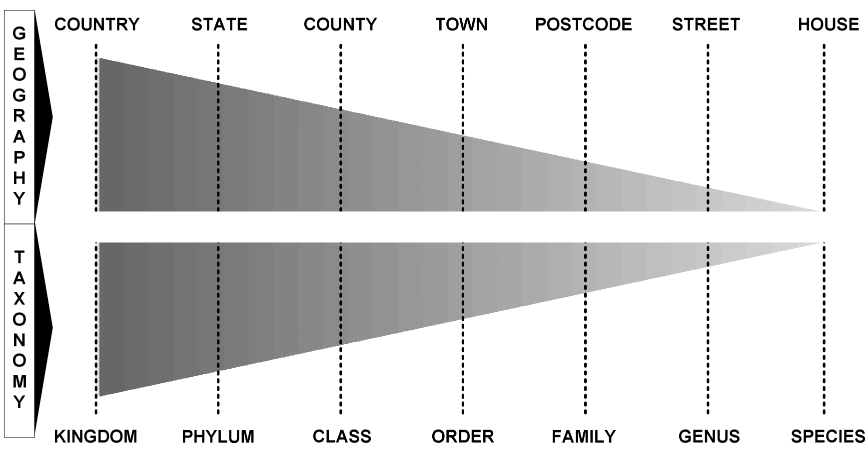


Figure 3.5: Comparison of classified units in species and geographical taxonomies.

In health science one of the most important classification applications is the International Classification of Diseases (ICD). The World Health Organisation (WHO) publishes a disease classification almost every decade from 1900 onwards (CDPHE, 2001). The ICD is used to classify diseases and other health problems recorded on many types of health and vital records including death certificates and hospital records (WHO, 1992).

In addition, the Office for National Statistics (ONS) developed the Area classification of Great Britain using the local and health authorities as spatial units. This classification provides a general indicator of the socioeconomic characteristics of each Local Authority District and Health Authority in Great Britain. Using information collected at the 1991 Census it classifies each authority into one of 6 Families, 12 Groups and 34 Clusters on the basis of 37 separate socioeconomic variables. The Health Authorities have been only classified into Families and Groups. The variables for the classification were derived from the 1991 Census data. In more detail, the ONS used the basic population, employment, socio-economic and household characteristics. Moreover, the first time recorded information concerning the ethnic groups and the limiting long term illness have been included as potential significant variables. As Wallace and Denham (1996) noted, the ONS Area Classification is an indicator of socio-economic similarity and difference between areas. It is used extensively for resource allocation and performance management purposes by the NHS Executive illustrating deprived areas in Great Britain.

3.3.2 Cluster analysis

Cluster analysis is an important area of applications for a variety of fields such as in the machine learning, pattern recognition, and statistics and it is close related to classification. While in classification the number of classes was given, in cluster analysis little or nothing is known about the group structure. All that is available is a collection of units whose group membership is unknown. The grouping (clustering) of data items, which are similar to each other, is based on specific criteria. The aim in this case is to develop such a group structure that fits the units optimising the selected criteria. For example, an exploration of a given set of items with different colours (red, yellow and green) using the cluster analysis is possible only if certain criteria have been defined. In Figure 3.6, the six items can placed in *k* boxes according the targeted criteria. Therefore, if the criterion is to identify the colour of each item and place it in the appropriate box then three boxes are identified. On the other hand, if the criterion is to separate the cold from the warm colours then two boxes are required. It is obvious from the example that the cluster analysis is possible to suggest classification groups and usually researchers use simultaneously both classification and clustering approaches during their studies.

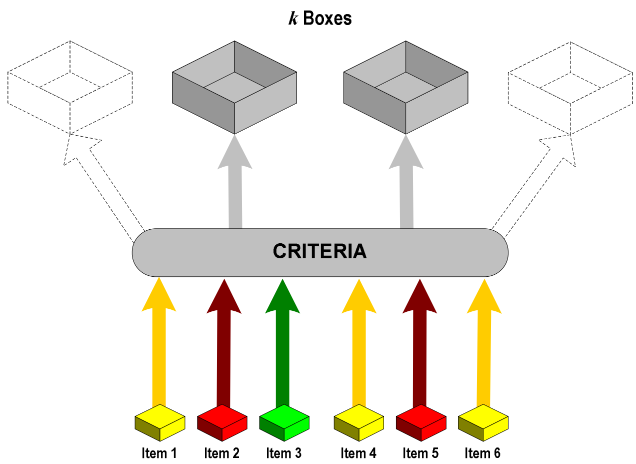


Figure 3.6: Clustering example

One of the most popular methods is the *k*-means clustering (also known as Forgy's method or MacQueen's algorithm). The *k*-means algorithm finds locally optimal solutions minimising the sum of the distances squared between each data unit and its nearest group centre "distortion" (Duda and Hart, 1973; Fukunaga, 1990). The *k*-means algorithm is formed spatial and aspatial according the studied needs. The aim of the *k*-means algorithm is to group M units with N dimensions into K clusters so that the within cluster sum of squares is minimised. The N dimensions are representing the available attributes assigned to the M units. The within cluster reduction of sum of squares is granted by swapping M units from one cluster to another (Hartigan and Wong, 1979). Although, the k-means algorithm requires a given number of K clusters, there are hierarchical clustering methods where during the process two clusters gradually broken down into smaller and smaller K clusters providing a hierarchical classification. Frequently, such clustering algorithms are used for cluster identification in spatial databases.

Especially, in large spatial databases an application has to cover the following requirements: minimal requirement of domain knowledge to determine the input parameters, discovery of clusters with arbitrary shape and good efficiency on large databases. The combination of these three requirements is rarely supported by a clustering algorithm. Kaufman and Rousseeuw (1990) suggested two basic categories of clustering algorithms: partitioning and hierarchical. Generally, a partitioning algorithm requires some domain knowledge concerning the number of clusters, which is not available in most of the cases. On the other hand, a hierarchical algorithm requires only a termination condition to control when the merge or division process should be terminated. The DBSCAN cluster algorithm is one of the few that can accomplice the above three requirements in combination (Ester et al., 1996). Three applications using point spatial databases presented by Sander et al. (1998) on the fields of remote sensing, molecular biology and astronomy, supported the advantage of DBSCAN for quick and precise cluster capturing. Halkidi et al. (2002b; 2002a) adapted the theoretical base of DBSCAN algorithm and they developed a new approach for Non-Point objects Clustering (NPClu). The NPClu algorithm is based on the minimum bounding rectangles vertices (MBRs) of two-dimensional objects. The MBR is the smallest rectangle that surrounds a selected shape. The algorithm explores possible clusters using the MBRs vertices introducing a spatial clustering procedure.

Since the 1980s the analysis of disease clustering has generated considerable interest in the area of health science. A substantial number of aspects of the clustering analysis have been studied in cluster detection of diseases starting from the detection of cholera disease in London by Snow. An interesting review of clustering detection methods presented by Lawson and Kulldorff (1999) extends the definitions of cluster analysis in disease detection by Besag and Newell (1991). Besag and Newell identified two types of cluster analysis: *focused* and *general* methods, according to the characteristics of each approach. Focused methods are used to assess the clustering around a predefined point, like a nuclear installation, requiring the number and the location of clusters. General methods are more sophisticated aiming at investigating whether clustering effects occur over the study area. In addition, general approaches can be split even further in two categories according to the focus of the method, such as the *tendency* and *location of cluster*. In terms of tendency of cluster approach, the algorithm investigates cases located close to each other, without any consideration as to where they occur. On the other hand, the location of cluster method explores the study area attempting to identify existing clusters.

Table 3.1: Methods for searching location of clusters and detecting a tendency to cluster.

Reference	**Purpose**
Turnbull et al (1990)	It searches for the location of clusters. Originally designed for population counts. MC simulation is needed to evaluate the significance of method.
Kulldorff and Nagarawalla (1995)	It searches for the location of clusters. Originally designed for population counts. It is constructed via the likelihood ratio test and evaluated using the MC simulation.
Besag and Newell (1991)	Detects a tendency to cluster. Originally designed for population counts. It requires a fixed number of cases and it is applicable only for rare diseases. Also it evaluates the significance of method using MC simulation.
Tango (1995)	Searches for the location of clusters and detect a tendency to cluster. it is designed for population counts. The evaluation of model is based on χ^2 approximation with good respond for small number of cases.
Cuzick and Edwards (1990)	Detects a tendency to cluster using a sample of cases and controls. MC simulation is used to evaluate the significance of method.
Anderson and Titterington (1997)	Detects a tendency to cluster using a sample of cases and controls. The statistical significance of this method requires advanced MC integration and simulation.

Moreover, a valuable comparison of general tests for spatial clustering introduced by Tango (1999) presented a brief description of six tests. Table 3.1 lists the six methods providing the purpose of each test and references for further interest. A common feature of these methods is the use of a Monte Carlo (MC) simulation to identify their distributions, excluding Tango's approach in which the ordinary χ^2 approximation is applied. The above cluster examples use generally aspatial datasets or positional units (points) such as, population counts and incidents. However, these cluster detectors are inappropriate for use with areal units, because they analyse the existence of clusters according to individual point information. As a result, the point clustering method leads to non contiguous zones in comparison to zoning procedures where the zones are strictly contiguous. Therefore, it is clear that clustering methods are not suitable for use in studies where areal units and their relationships predetermine important functions. An exception needs to be noted here for the NPClu algorithm by Halkidi et al. (2002a; 2002b) because of its promising nature for further research as an areal unit clustering method.

3.3.3 Regionalisation

Regionalisation was defined by Grigg (1965) as a special form of classification. The main difference is that while classification usually refers to aspatial units and their class allocation according to their characteristics, regionalisation is a spatial process where the spatial units are grouped together following predetermined spatial rules and restrictions. In particular, regionalisation involves the grouping of spatial units based on certain characteristics, structural properties and relationships with neighbouring areas. Using a simple example of twelve spatially related items (Figure 3.7), a region building method aggregates the areal units into three or four zones optimising the applied criteria and following the spatial rules and restrictions. On the other hand, the classification method provides for both spatial constructions the same number of classes (three classes) as it focuses on the attributes of areal units ignoring their spatial structure. It is obvious that the spatial relationships of the areal units in collaboration with the criteria needed to be satisfied provide different results compared to a classification method.

During the "quantitative revolution" (from the late 1950s to 1970s), there was an increasing recognition of the importance of region building in Human Geography (Johnston et al, 1993). As a result, numerous approaches and algorithms were suggested for tackling combinatorial and redistricting problems (Cliff and Haggett, 1970; Haggett, 1965; Haggett et al., 1977; Lankford, 1969). Cope (1971) identified three oversimplified elements involved in any regionalisation problem: a set of areal units (spatial areas), a number of targeted criteria and a grouping process. His observations provided the starting point for investigating these geographical problems and later researchers like Taylor (1973) and Openshaw (1984) investigated a more specific form of region building known as political districting problem. In the UK, political redistricting is performed by neutral government commissions. The commissions should provide electoral geographies following a given set of rules. The rules target natural boundary limitations as well as equal populations within regions. A selected regionalisation process is often under heavy criticism as it is possible to become favour to a political party providing additional parliamentary seats. Therefore, the computational algorithm of political redistricting should be expressed as a series of well-defined steps with respect to the aggregation criteria.

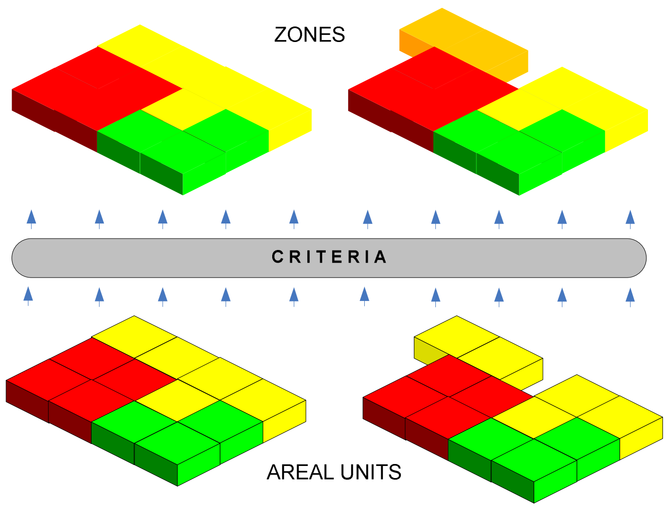


Figure 3.7: Regionalisation example

However, the geographical limitations and rules provide additional complexity in the region building approaches highlighting spatial and aspatial issues. In the following sections derived problems from aggregation processes are discussed investigating the consequences in health studies.

3.4 Fallacy effects in health studies

3.4.1 The realisation of ecological and other fallacies

Subsequent analysis of social areas has been carried out using data from Census Enumeration Districts (EDs), city wards, and other administrative units. When a relationship between two social groups (for example, ethnic minority and high infant mortality) is implicitly or explicitly inferred, the inference is often closely related to the *ecological fallacy*. This fallacy term has been generally defined, as the assumption that aggregated data of heterogeneous individuals may not be appropriate to draw inferences about individuals within the population. The ecological fallacy was first observed in sociological research. Robinson (1950) was the first researcher to identify the phenomenon by referring it as the *ecological correlation problem*. Alker (1969) based on Robinson's findings explored other inferential fallacies. He suggested that in addition to fallacies of disaggregation, there are also fallacies of aggregation. As a result, he created a typology of eight forms of ecological fallacy: *ecological fallacy*, *individualistic fallacy*, *cross-level fallacy*, *universal fallacy*, *selective fallacy*, *contextual fallacy*, *cross-sectional fallacy* and *longitudinal fallacy* according to the level of analysis and the direction of inferences between different levels of analysis. A brief demonstration of these fallacies is presented here.

The *ecological fallacy* is the attempt to conclude information, concerning an individual from related characteristics observed at an aggregated level. This fallacy was observed by Robinson (1950) and discussed extensively in 1970s. The *individualistic fallacy* is actually the opposite of ecological fallacy and it is the attempt to infer characteristics of an aggregated level according to the observations concerning an individual. The *cross-level fallacies* are generated when characteristics of a specific group of individuals are related to propose relationships of whole population. The *universal fallacy* and the *selective fallacy* are closely related to the cross-level fallacies. The universal fallacy refers to problems of generalising characteristics from a sub-sample and the selective fallacy represents a specific sub-sample as generalised characteristic. According Alker's description "If the sub-samples are randomly extracted, only ordinary problems of statistical inference exist, but nonrandomly selected subgroups raise the likelihood that other variables, perhaps associated with the partitioning relationship, destroy any such easy correspondences"(Alker, 1969: p80). The *contextual fallacies* take place at the same level of analysis, the context or social structure could change the strength or form of statistical characteristics. The *cross-sectional* fallacy and the *longitudinal* fallacy are related to temporal changes in a region and assume that relationships can be derived from social patterns observed in certain moments in time. For example, such temporal changes could be the move of different addresses for a considered percentage of people during the study period. Even though, the use of longitudinal data at certain moments in time maybe extremely helpful in social studies if the accompanied fallacies are controlled.

The above fallacies, in particular the first three, have been discussed in the research community in the 1970s mainly by Giggs (1973); Gudgin (1975); and Johnston (1976). At that period, Susser (1973) argued that *ecological* might not be the appropriate term for describing the fallacy deriving from data aggregation and he preferred the much more precise term *aggregative fallacy*. However, the use of such term should be careful as it refers to one direction effects appearing in aggregations of individuals or groups of people.

In recent years geographers and statisticians termed the aggregation effects as ecological fallacy (Tranmer and Steel, 1998) and most of them have rediscovered aggregative fallacies in their analysis (Openshaw, 1984a; Wrigley, 1990). Tranmer and Steel (1998) presented a statistical model used to explain the effects of aggregation on variances, covariances and correlations. The "adjusted" model is able to correct statistics in aggregate level by using individual information describing the local homogeneity of an aggregated areal unit. This approach has been discussed in a series of articles (Holt et al., 1996; Wrigley et al., 1996) and was expanded by Tranmer and Steel (2001) to estimate variance and covariance components in a three-level adjusted model using aggregate data (EDs and Wards) and individual social characteristics. Despite the positive contribution of this method to theoretical and analytical understanding of the aggregation and ecological effects, it is required to be noted that the above approach is based on two assumptions. First the aggregated variables need to be identified and explained according to the relationship between the individuals and the aggregated level, and second the selection of these variables either derives from previous research or by applying statistical measures. These assumptions could be carriers of ecological fallacies latent in the datasets and the levels of the study. Therefore a careful consideration of the above concealed effects is needed in any undertaken case study.

3.4.2 Fallacies in Health Geography

In general ecological studies analyse data at the aggregate level rather than the individual level. The use of aggregate levels into studies intends to provide a confidentiality barrier between the researcher and the individual's private information as well as to decrease the complexity and size of data. In health geography, the use of such ecological data to make inference about individuals is a very important issue as the aggregate level holds information about groups of individuals ignoring special characteristics of individuals. For example, a specific policy in an affluent area can be disadvantageous to those individuals with very low income or via versa. Therefore, the health researchers and organisations should be aware of enclosed fallacies in their analysis of aggregate levels because such an assumption can provide misallocation of health resources with a further problematic health policy. There is a large epidemiological literature concerning the ecological bias and fallacies focusing in epidemiological phenomena and structures (Greenland and Morgenstern, 1989; Greenland and Robins, 1994; Piantadosi et al., 1988; Richardson et al., 1987).

A clarification of the three most common varieties of fallacies; the ecological, individualistic and cross-level from the epidemiological perspective are discussed below with epidemiological examples taking into account the effects of ecological fallacy. As mentioned above the ecological fallacy refers to inferences about individuals based on characteristics at an aggregated level (Figure 3.8.a). A result of such fallacy is a misunderstanding of studied area and in addition problematic health policy. The ecological fallacy is well defined according to the transition of health policy from the aggregated level (community) to the individual and classified in three categories by Blakely and Woodward (2000): a direct cross level effect, cross level effect modification and an indirect cross level effect. The *direct cross level effect* is referred to the aggregated level that specific policies directly affect an individual (Figure 3.8.b). For example, a policy applied in a deprived community directly affects the health of individual. The *cross level effect modification* is the attempt to modify the effect of social group on an individual's health following conclusions derived from aggregated level, and the *indirect cross level effect* is defined as policy applied to a social group that indirectly affects an individual's health from aggregate level characteristics.

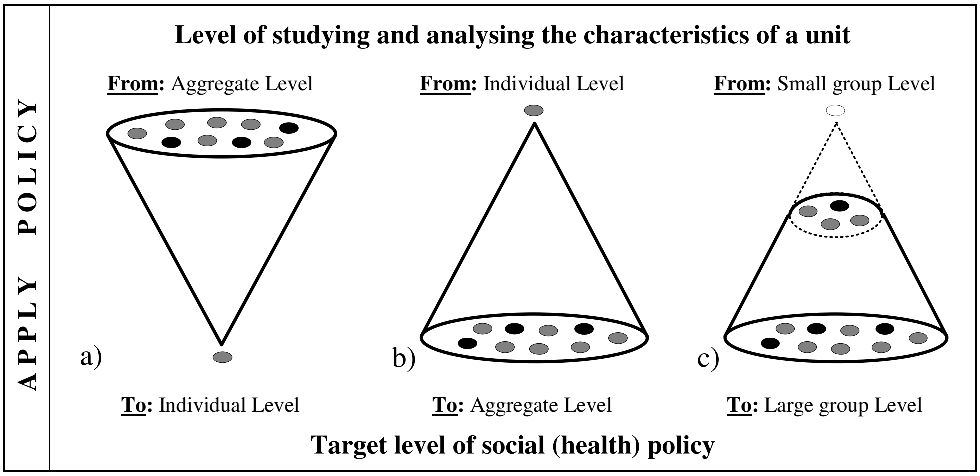


Figure 3.8: Three types of fallacies: a) Ecological, b) Individualistic and c) Cross-level fallacy

A demonstration of ecological fallacy and a methodology for reducing ecological bias taking into account different population datasets at the aggregate level are presented in the article by Lancaster and Green (2002). They performed regression analyses of limiting long term illness at both aggregated (electoral ward) and individual level (sample of anonymised records). Their analyses compared several measures of deprivation including the separate variables which make up the indices for examination of their effectiveness in explaining rates of illness. In addition, they constructed the deprivation scores at the individual as well as at the ward level. As a result, they suggested that the appropriate choice of socioeconomic variables and consideration of the age structure of the population can provide a single aggregate deprivation index that will explain most of the variation in rates of illness across the study region. However, the use of administrative areas like wards encloses issues concerning the homogeneity of the studied population within wards as well as the variation of population between wards.

The individualistic fallacy is illustrated by Wen et al. (2001). They present a case study, in which hospital level measures determine a reduction in the rate of normal appendix removal by delaying surgery, whereas the individual level measures suggest the opposite. This study emphasises that biased inference can occur in both individual and ecological level and it is important that the measures and the analysis at these levels must be selected based on the prior knowledge of the study area.

A valuable statistical tool in health related research is multilevel modelling. The multilevel modelling approach is based on the idea that different aggregate levels and their ecologies should be modelled simultaneously (Jones, 1991b). The concept behind multilevel modelling is to develop models in ecological levels reflecting the study area and then combine them into a general model (Jones and Duncan, 1995). As Jones and Duncan (1996) claimed, it is possible to conduct simultaneous analysis of several ecological levels such as individual and aggregated levels. Multilevel modelling is a multi-regression approach that estimates the models at the different scales determining the parameters of the lower scale model by the higher scale model. A multilevel model is structured in a hierarchical order and it is calibrated by shrinkage estimators which can provide more efficient information comparing to the usual ordinary-lest squares. Although, most of the empirical studies utilise the multilevel approach, the choice of the aggregated level is limited to existing administrative areas. As a result, the multilevel modelling reflects to a greater extent relationships between socioeconomic determinants when the aggregated geographies are well defined. In this thesis, we would not further investigate the multilevel modelling as it is beyond the scope of this research.

3.5 The Modifiable Areal Unit Problem (MAUP)

3.5.1 An overview of the MAUP

As Johnston et al. (1993) mentioned, the Modifiable Areal Unit Problem (MAUP) is a special form of ecological fallacy, associated with the aggregation of data into areal units for geographical analysis. The MAUP takes place when information of a low aggregated level is grouped into a higher aggregated level. The term modifiable reflects the changeable nature of areal units and zones when used to capture and describe spatial phenomena. Openshaw (1984b) set down the nature of MAUP providing an extended description. The MAUP is one of the most important problems that geographers and statisticians have to face when they deal with aggregation issues. The appropriate method of analysis has to address the questions: how the MAUP is involved in the analysis of case study aggregate level and how the MAUP can be tackled. A series of article researching the impact of the MAUP has been published since Gehlke and Biehl (1934) first noted the problem. They observed that the correlation coefficients are increasing as the aggregation level increases.

On a different kind of study field measuring regression and correlation parameters, Yule and Kendall (1950) firstly investigated the effects of MAUP using agricultural datasets concerning the counties of England. Although their dataset did not relate to any socioeconomic population characteristics, their research on the correlation coefficients between wheat and potato yield per acre was a confirmation of Gehlke and Biehl's results. However, since 1930s, many researchers attempted to understand the nature of MAUP using different approaches. For instance, Fotheringham and Wong (1991) on multivariate linear and logistic regression; Flowerdew and Amrhein (1989) on Poisson regression; Horner and Murray (2002) on excess commuting; and a collection of papers from Openshaw (1977b; 1977a). A milestone on the MAUP understanding is Arbia's (1989) research on spatial aggregation analysis. He suggested that the MAUP is directly connected with the degree of homogeneity of variables within and across zones. Expanding more this suggestion Holt, Steel and Tranmer discussed how areal homogeneity attached with the MAUP proposing the intra-area correlation approach to explain the MAUP (Holt et al, 1996; Tranmer and Steel, 2001). The intra-area correlation measure is extensively used to multilevel analysis for measuring the within and between variation of aggregation levels (Jones, 1991a).

3.5.2 Defining scale and zoning effects

Openshaw and Taylor (1979: p128) were the first to term the scale effect by Gehlke and Biehl (1934) and the effects of different methods of aggregation by Blalock (1964) as the "*modifiable areal unit problem*". The effects of MAUP were discussed by Openshaw (1984b) and are now known to researchers as the *scale effect* and the *zoning effect.* The *scale effect* is the variation in results that can be obtained when data for one set of areal units are progressively aggregated into fewer and larger units. For example, when census enumeration districts are aggregated into wards, districts, or other administrative units the results change with increasing scale. The *zoning effect* is the variation in numerical results arising from the grouping of small areas into larger units. This means that when an aggregation of enumeration districts into an aggregated level equivalent to wards occurs then it is possible to create a variety of output zones at the same level. As Wong and Amrhein (1996) mentioned "*researchers have to deal with the scale effect more frequently than the zoning effect*" because most of them usually work with specific aggregated levels and they use administrative areas that have been already specified. However, dealing only with one side of the problem is not the appropriate approach and it is important that zoning effect analysis takes place for suitable results.

Figure 3.9 presents the fallacies and the effects that affect an aggregation or even a disaggregation approach. In this example, the study area consists of 20 areal units and the aggregation occurs for 5, 2 and 1 output zones. At this point, it is important to prompt that an amount of error has surcharged the areal units' dataset (ecological fallacy) during the transmition of the individual level characteristics to the areal unit level. By aggregating the areal units into 5 or 2 output zones the scale effect is present. Furthermore, the variation within the 5 output zones is smaller that the variation within the 2 output zones. As mentioned above, areal units can be aggregated into various outputs zones on the same aggregate level. The figure shows three alternative output zones for each created aggregation level. On the top of the figure there is the level of a single output zone; this is the most aggregated level providing only a summary of the individual characteristics and it is the most heterogeneous level.

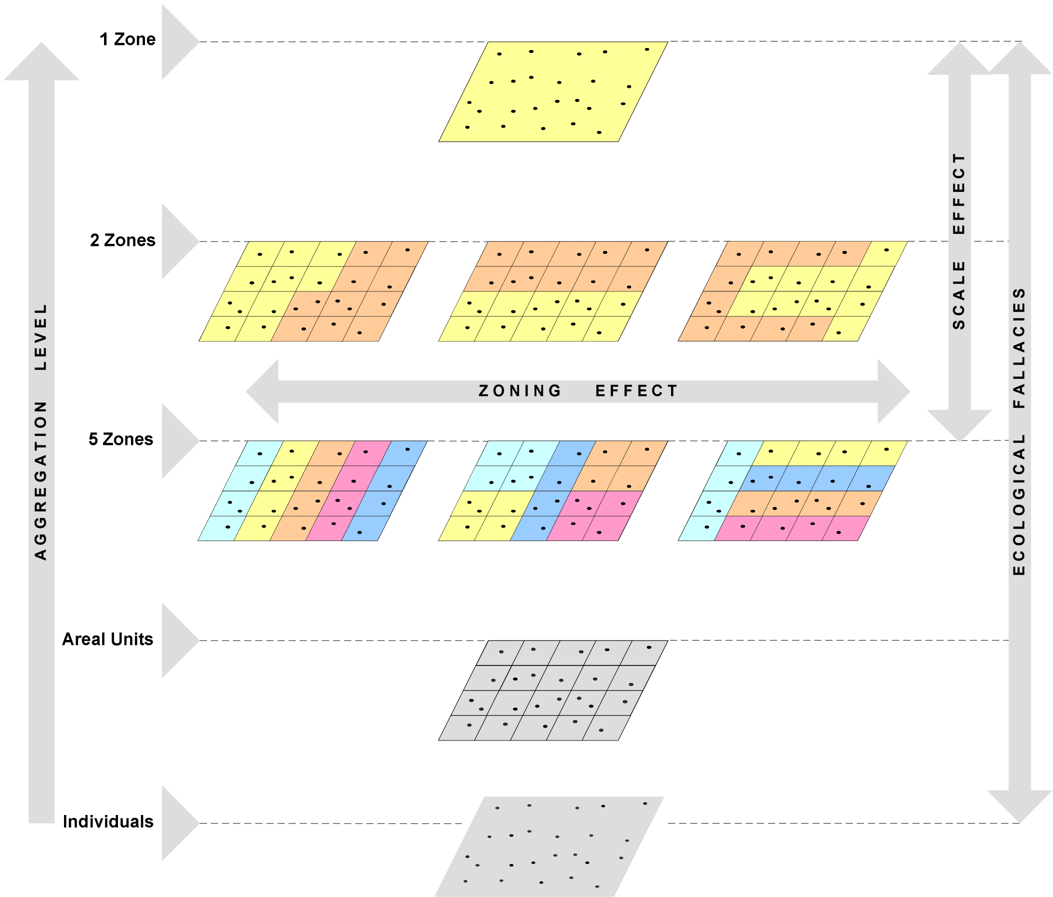


Figure 3.9: The effects of Modifiable Areal Unit Problem

More recent research by Cockings and Martin (2005) outlines the application of zone design techniques for the creation of aggregated areas in environment and health studies. Their empirical study of the relationship between deprivation and limiting long-term illness in the county of Avon is developed taking into account the scale and aggregation effects. The finding of increasing correlation between deprivation and morbidity as the aggregation level increases, follows previous findings in the literature (Gehlke and Biehl, 1934; Openshaw, 1984b). Their suggestion is that zone design tools provide further systematic exploration of scale and aggregation effects and can be beneficial in health studies.

3.6 Aggregation approaches

The manipulation of ecological fallacies and MAUP during an aggregation process has been constantly investigated by means of subjective and automated methodologies. In this section, we focused on automated aggregation methods providing an overview of aggregation systems with their information and characteristics. The advantages and disadvantages of each system will be discussed with reference to spatial and statistical aspects. The selection of the following systems is based on two basic aggregation criteria: the output zones are continuous and specific criteria are used to define the grouping procedure.

3.6.1 Zone Design Systems

The original automated zone design program (AZP) of Openshaw (1977c) was developed to search for advanced ways of aggregating areal units and explore the MAUP effects. AZP is a computational procedure, which seeks to optimise objective functions such as minimum or target population sizes, zone compactness and social homogeneity, while grouping a large number of areal units into a smaller number of output areas (zones). In mid 1990's, the original AZP procedure was updated to a ZDES system based on the work of Openshaw and Rao (1995). Workstation hardware improvements and wide availability of geographical data, dramatically extended and improved the capabilities of the ZDES system (Openshaw and Alvanides, 1999). In ZDES some constrains are implicit in the system and can never be violated such as the contiguity constrain. Holding the contiguity intact throughout the aggregation process, the system secures an unbroken and not overlapped adjacency for each output zone. There is a variety of design and handling constrains such as *tabu search* and *simulated annealing*, that introduce new ways of aggregating areas using heuristic criteria for improving the performance of the algorithm by allowing detrimental moves in order to explore more areal units. The core of the system is a FORTRAN procedure that is fed information from an Arc/INFO environment. As a result the ZDES provides a complex environment for further development and it is dependent to limited old programming methods.

However, the AZP system and its variations have been used in many empirical studies providing valuable findings for correlation models (Openshaw, 1977b), spatial interaction models (Openshaw, 1977a), investigating zone design criteria for regression models (Openshaw, 1978a) and applying size and shape constraints (Openshaw, 1978b). Recently, the ZDES has been suggested for policymaking and exploring spatial flow data sets (Alvanides, 2000).

3.6.2 Automated Zone Matching system (AZM)

In general the AZM system was developed for the needs of 2001 Census in UK. The original automated zone design procedures by Openshaw were adopted for the creation of 2001 Census output areas by Martin et al. (2001). The development of zone design system was implemented in visual basic programming language and it is able to provide homogeneous and equal populated zones as well as shape constraints (Martin, 1998). In detail, the user can set a maximum and minimum threshold of population within zones in case the target is to provide equal populated zones. On the other hand the compactness of zones is provided by calculating the squared perimeter divided by area method. This way the each compact zone targets a similar value to the value of circle (12.56).

In addition, the method for measuring homogeneity within output areas was introduced using an intra-area correlation measure for the maximisation of social homogeneity. The maximisation of intra-area correlation objective function measures the similarity of values within any area of interest (Holt et al., 1996). As the correlation values increases in the output zones the greater homogeneity is archived within each zone. Theoretically the maximum value of intra-area correlation is d = 1.0 and it can be computed for a single category or for multi-category variables. In practice, any intra-area correlation value above 0.05 implies a reasonable degree of homogeneity (Tranmer and Steel, 1998).

Furthermore, a couple of social studies funded from the ESRC supported the need of this tool suggesting methodologies for constructing the 2001 Census output areas (Clarke et al., 1999; Martin et al., 1999). In a health context, Cockings and Martin

(2005) suggested the AZM system as an important tool for exploring the scale and aggregation effects. Also, they proposed the automated output zones as more appropriate health geographies than the existing administrative areas. Although the AZM system has been developed lately, it is obvious that the system mainly implemented around the needs of 2001 Census providing limited range of objective functions and constraints.

3.6.3 Spatial Analysis in a GIS Environment (SAGE)

The Spatial Analysis in a GIS Environment (SAGE) was developed in the Arc/INFO environment providing a collection of statistical tools such as box plots, histograms and scatter plots, for region building (Haining et al., 1996; Wise et al., 2001). SAGE also includes facilities for constructing and modifying a wide range of weight matrices, and making adjustments for variable base populations. Accordingly, the research is able to provide different weights in each variable set as well as in criteria concerning the aggregation process. For example it is possible the SAGE system to provide zones with strong compactness in expense of homogeneity and via versa.

In detail, the SAGE system supports three criteria: homogeneity, equality and compactness. The homogeneity criterion targets homogeneity within zones minimising the statistical variation of selected variables within each zone. The equality criterion intends to provide similar zones to one another in terms of the sum of one variable like population. Finally, the compactness criterion controls the output shapes of zones measuring the within zones variance of the *X* and *Y* coordinates of the area centroids. The above three criteria construct a super criterion which is defined as the sum of the weighted functions for the three criteria:

$$f_0 = w_H f_H + w_E f_E + w_C f_C \qquad (3.1)$$

where, f_H, f_E and f_C are the objective functions for homogeneity, equality and compactness respectively and w_H, w_E, w_C are weights specified by the user concerning the importance of each criterion. A k-means clustering procedure is the core of the SAGE system and produces output zones rapidly. Despite its very strong statistical tools for socio-economic analysis, the initial allocation of areal units to zones is the main weakness of this system. The original algorithm after the seeding of *k* areal units as the cores of the targeted zones uses the objective function (equation 3.1) to archive a good initial aggregation. Therefore the system easily trapped in local optima and possible better solutions are not investigated. A series of methods have been developed to try and reduce this problem such as manual selection and loading predefined zones. However, the SAGE system developed targeting faster aggregations in expense of optimal solutions. In addition it is important to point out here that the output zones are products of one aggregation run and it is expected every time the algorithm executed to produce different outputs with high variation between the final objective function scores (Wise et al., 1997). As a result the aggregation effects aforementioned in earlier sections of this chapter are not extensively explored providing a good solution but not the optimum one.

Although the SAGE system suffers from a limited aggregation algorithm the additional statistical tools such as the generalised and local Moran and Getis statistics (Cliff and Ord, 1981; Getis and Ord, 1992) provide a statistically advanced environment for aggregation processes. Recent research by Ceccato et al. (2002) used the SAGE system to explore offence patterns for residential burglary, theft of and from cars, and vandalism in Stockholm city. Their findings suggest a noticeable shift in both geographical patterns and their association with socioeconomic variables. In addition, the limitation of SAGE system in terms of exploration of MAUP effects is clear here as the researchers run the SAGE system creating a number of different output zones for further manually comparison. Moreover, the scale effect was not investigated as the aggregation level was subjectively defined to 119 output zones. However, the statistical environment of the SAGE system provides a valuable analytical tool in the hand of geography researchers.

3.6.4 Traffic Analysis Zones (TAZ)

An interesting tool in transportation planning analysis, the Traffic Analysis Zones (TAZ) is presented in Ding's article (1998). His attempt was to examine the impact of aggregate levels on transportation combining land use and transportation information.

The TAZ system was developed in an ARC/INFO environment and the main procedures built on FORTRAN programming language. It was based on eight criteria: one type of land use in each individual zone is preferred minimizing the intra-zone trips, homogeneity, compatibility of output zones with census data, uniqueness and completeness, prohibit of islands creation, equity of trip generation, convexity and compactness. The homogeneity criterion is involving the calculation of a homogeneity index (I^h) which is a weighted Euclidean distance formula:

$$I_{ij}^{h} = \sqrt{\sum_{r} w_r (X_{ir} - X_{jr})^2} \qquad (3.2)$$

where I_{ij}^{h} is the homogeneity index between the adjacent pair of polygons i and j; w_r is the weight for attribute r; and X_{ir} , X_{jr} are the values of attribute r in polygons i and j respectively. The summary of weights for attribute r is: $\sum_{r} w_r = 1$

As mentioned in Ding's article, there are some issues unsolved. First is the absence of compactness criterion and the second is the lack of flexibility in dealing with conflicting criteria in the TAZ algorithm. Nevertheless, the TAZ system is developed for aggregating transportation areas investigating parameters (Ortuzer and Willumsen, 1994) such as homogeneity with respect to the socioeconomic characteristics, compactness of zones' shapes, respect of administrative boundaries, and respect of physical geographic separators placed on territories and exclusiveness covering special transportation problems like no doughnut.

3.6.5 Travel to Work Algorithm (TTWA)

The Travel to Work Algorithm developed by Coombes et al. (1982; 1986) is another aggregation method based on the maximisation of the inherent tendency of interaction data to link nearby areas. As Coombes (2000) noted the TTWA algorithm routinely produces contiguous regions without using an explicit contiguity constraint, relaying upon the optimisation of its objective function. Moreover, explicit contiguity information is not accessible in the TTWA system because the areal units of a study are replaced by their centroids. Vital for the employment of TTWA system is the construction of the Synthetic Data Matrix (SDM). The SDM is a cumulative overlap of matrices in which the classification information of areal units is binary expressed. For example, if the results from three binary matrices were collated then the value in each cell of the SDM would vary from 0 (for any pair of areas which were not in the same class according to any of those matrices) up to 3 (for any pair of areas which all three matrices had the same class).

Eurostat (1992) recommended that the TTWA method, known also as the European Regionalisation Algorithm (ERA) should be seen as the standard for defining local labour-market areas in European countries. Recently, an empirical study using the ERA method suggested by Shortt et al. (2005) for defining GP catchments in Northern Ireland. The GP catchment's areas should reflect important information such as levels of accessibility to surgery without overestimate or underestimate medical service areas. In their study, the SDM and ERA approaches are formalized to create an optimal set of non overlapping zones according to pre-defined population size and self-containment criteria. Also, they support that the produced set of zones provides compact, robust and highly self-contained catchments.

On the other hand, the SDM/TTWA approach appears to have need of considerable labor work especially at the stage of SDM's preparation. Even though, in practice the TTWA has been applied in various EU projects (Casado-Diaz, 1996; Sforzi et al., 1997) seems to be impossible for the system to archive maximum homogeneity withdrawing all the other constraints such as compactness and population size. The TTWA system provides an excellent framework for certain aggregation problems like the GP's catchment areas but sometimes the homogeneity of zones is most important than their compactness. Furthermore, the whole system provides a single set of zones suggesting a sort exploration of aggregation effects.

3.6.6 Multi-Objective Zoning and AggRegation Tool (MOZART)

The MOZART tool developed by Guo et al. (2000) is adopting basic zoning criteria from the TAZ system and unifies them under the graph partitioning approach. The multi-level technique by Karypis and Kummar (1998a; 1998b) of partitioning graphs has been implemented in this tool introducing a very fast algorithm for aggregating areal units. Their study focused on the equal population and the compactness of each output zone and it is applied on a small sample of areal units (577) producing 10 aggregate levels of 10, 20… 100 zones. Although algorithmically the tool provides a strong aggregation framework the weak geographical and statistical development of this tool is an important missing in this study. However, its potentiality is not reflected to empirical studies in literature limiting our critical view in the Guo et al. paper. For example, the tool was not evaluated in a real aggregation problem with its special geographical features such as rivers. In addition, it is obvious that the single set of zones the system produces every time the algorithm runs explores only one of the numerous possible aggregations suggesting limited investigation of aggregation effects.

3.6.7 Informative geographical aggregation

An informative approach for tackling the MAUP effects has been proposed by Nakaya (2000). He proposes a new methodology to select appropriate aggregation scale using the Akaike Information Criterion (AIC). The AIC is an estimator, which evaluates statistical models by taking into account the principle of parsimony. This principle states that the better model is the model whose goodness of fit is not only higher but also whose number of parameters is smaller (Parzen et al., 1998; Sakamoto et al., 1986).

Nakaya's approach applied in Tokyo Metropolitan Area datasets concerning standard mortality ratios of elderly males aged 65 to 84 years from the 1990 census and 1990 Vital Statistics of Japan. The statistical unstable rates of the area were tested by Poisson distribution analysis assuming that the number of the deaths is occurring randomly and independently. Accepting the Poisson distribution of death rates, he created a Poison dummy model (PDM) as suggested by Choynowski (1959) to statistically assess the incidence rate map evaluating the PDMs' goodness of fit by deviance and AIC criteria. A simulation of random aggregation suggested running the algorithm 200 times for each zone and the more informative aggregation level was targeted around the 50 zones derived from 262 areal units. In Figure 3.10, EAIC is the expected AIC that provides a aggregated model in terms of AIC while it holds the significance level of deviance. At the stage where the zone number is large, the random aggregation was able to attain deviance within the confidence interval. When we move to the aggregation level with fewer zones and the minimum AIC estimator then the deviance of aggregation is outside the confidence intervals, suggesting that an aggregation solution can bias some true variation. However, the AIC's ability to measure the model's goodness of fit in relation to additional information for the used number of model parameters can be very useful at the stage of defining the most informative scale level of a study area (Nakaya, 2000).

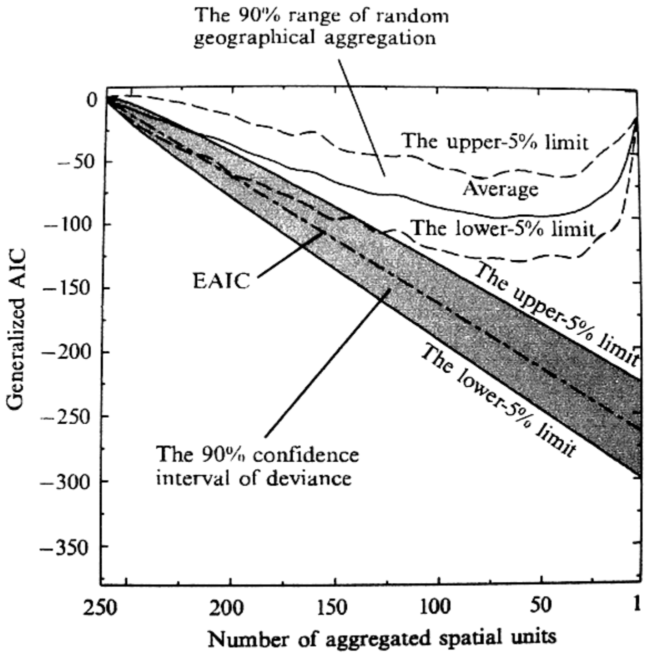


Figure 3.10: The scale relation of the expected EAIC, AIC and its confidence intervals.

Source: Nakaya (2000), page 98.

3.7 Discussion and Conclusions

In this chapter we investigated the most common spatial issues that have been or could be addressed in health research. As mentioned in Chapter 2, boundaries can have a considerable effect on policy decisions for health related issues; here we highlighted the dynamic nature of boundaries in relation to existing empirical studies from health and geography related research. This research drew attention to three popular techniques for analysing health related datasets: classification, clustering and regionalisation, discussing the benefits and limitations of each technique, in addition to relevant research studies in health and other research disciplines. Two vital analytical issues:

ecological fallacies and MAUP effects were discussed, focusing on their importance in health related studies, as well as providing their influence by spatial entities. In particular, we discussed the effect of ecological and MAUP issues in research dealing with both individuals and groups, highlighting the need for understanding these problems in order to achieve better health policy and allocation of health resources.

To enable the implementation of new methods and strategies related to health issues, in the last section of this chapter, we introduced available automated aggregation tools highlighting their advantages and disadvantages. Existing aggregation applications exhibit a range of limitations, referring to weak boundary manipulation and difficulties on extensive exploration of issues in different scales and boundary configurations. In addition, the complexity and size of available datasets, in combination with dated algorithms for tackling the problem, limit their use by academic researchers. As a result, a combination of the innovative characteristics of existing aggregation methods is required, providing appropriate methodologies for health related studies. In the next chapter, the majority of issues discussed here is formalised by means of graph theory, developing new strategies for tackling methodological limitations and introducing a new health focused aggregation method for the construction of specific health related geographies.

CHAPTER 4

Implementing a Zone Design System

4.1 Introduction

In the previous chapter, we discussed the importance of automated aggregation systems presenting their advantages and limitations. We concluded that a range of spatial issues appear to affect the majority of the aggregation systems reviewed, resulting in less reliable zoning solutions. On the other hand, although some of the available zone design systems can manipulate the scale and zoning effects, they also suffer from limitations concerning the manipulation of special boundary issues, the high time consumption of their algorithm process, as well as dependency on other software products. In this chapter, we use as a guide the principles of the original Automated Zone Procedure by Openshaw (1977a), while we enhance all the components of a zone design system using innovative approaches based on the graph theory and object oriented strategies. Moreover, we suggest the use of Akaike Information Criteria as a valuable estimator of the aggregation scale and provide a methodology for identifying the most statistically informative aggregation level.

4.2 Graph Theory in Zone Design

4.2.1 Zone design characteristics

Since the development of the original AZP (Openshaw, 1977b), new aggregation systems have been implemented expanding the capabilities and uses of zone design, such as the ZDES (Openshaw and Alvanides, 1999) and the AZM (Martin, 2003). In Chapter 3, an overview of available aggregation systems was presented, focusing on the innovative parts of each method and highlighting potential issues using each system.

The majority of these systems are aiming to aggregate *A* areal units to *Z* zones, while maintaining the contiguity of zones and optimising specific functions such as similarity within each zone (Figure 4.1). Both contiguous zones and objective functions are the main properties of a zone design system as originally described by Openshaw (1976).

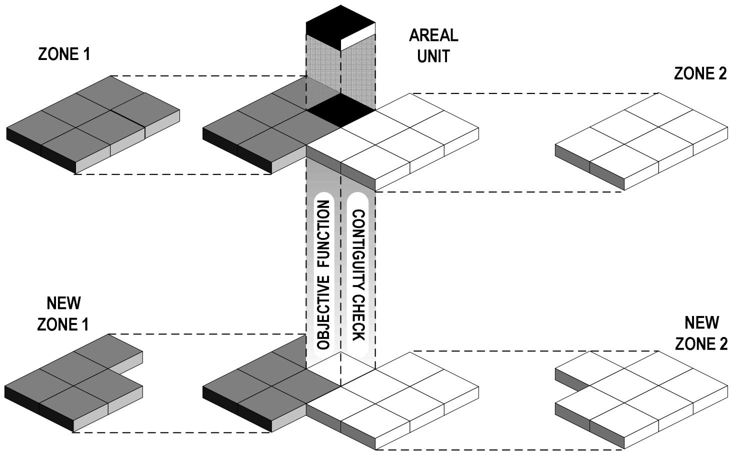


Figure 4.1: The two basic characteristics of a zone design.

One of the most widely used methods for optimising functions is the steepest descent or greedy algorithm (Luenberger, 1973). Given a function *F(x)*, the steepest descent optimisation targets the direction in which *F(x)* is locally optimised. This method is possible to proceed along two directions: minimising *F(x)* or maximising *F(x)*. Although maximisation of *F(x)* is feasible, the direction of minimising *F(x)* is the most common implementation of a steepest descent algorithm. For example, if we want to construct a method of equality in a set of units *M* then a steepest descent function could be formulated as the minimisation of differences between the available subsets *N* of the given set *M*. A generic formulation of such a function is:

$$F(x) = \sum_{i=2}^{m} \left| N_1 - N_i \right| \qquad (4.1)$$

where, m is the number of subsets N in the set M and N_i is the i^{th} subset of M.

In a zone design context the way to proceed from an existing aggregation to a better one is by swapping areal units at the borders of the zones, while optimising an objective function. During these swaps, it is possible for one zone to lose its contiguity therefore, a method of holding contiguity intact is essential. For example, Openshaw's AZP system tackled this problem by tracing an adjacency matrix using the depth first search algorithm (DFS). The algorithm is discussed later in this chapter, in more detail under a theoretical and practical context. One of the most serious difficulties in zone design is the approach adopted to maintain zone contiguities. The method should be as simple as possible avoiding complicated structures. That may lead to an exponential increase of processing time, during the iterative zone design procedure.

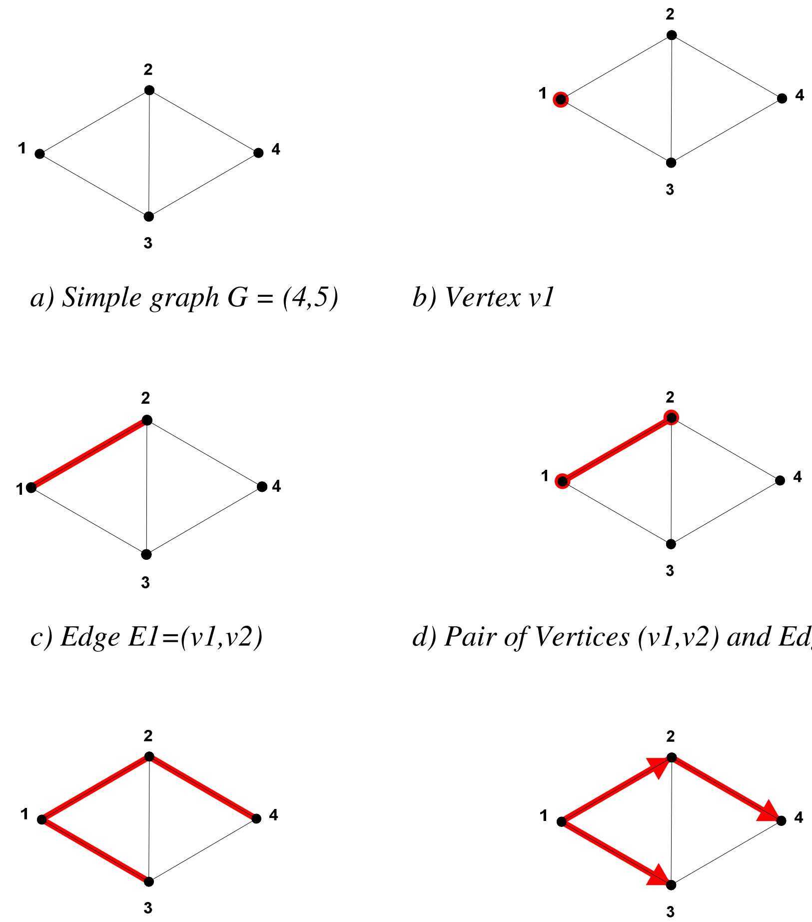
In Chapter 3, additional zone design properties were identified as equally important, such as the initial aggregation algorithm. The initial aggregation is the starting point for a zone design system. An initial aggregation targeting directly the criteria is avoided as the main zone design procedure is likely to be trapped into local optima and end the process providing an inadequate solution. Openshaw (1977a; 1978) suggested the use of an initial random algorithm (IRA) focusing on the principle of contiguous zones, as an appropriate first aggregation. It is obvious that the initial aggregation may offer a crucial advantage or disadvantage in the zone design process, according to the algorithm implemented. However, various methods have been suggested for the construction of an initial grouping (Taylor, 1973) where it becomes apparent that the selection of initial grouping should be based on the individual characteristics of each study.

Although the above three characteristics of a zone design system (objective function, contiguity check algorithm and initial aggregation) are structurally important, additional criteria perform special tasks expanding the capabilities of zone design system. Such criteria could be: the construction of compact zones in terms of shape formation, enrichment of adjacency information in studies with special geographical characteristics and opportunity to perform spatial analysis at the boundary features of study area. Evidently, each criterion applied to zone design performs a constraint to the output optimum solution with an additional increase of processing time. Therefore extensive use of criteria should be avoided if the study does not require such constraints, while thorough organisation of the zone design methodology is essential before each aggregation process.

4.2.2 Concepts of Graph Theory

In the following section some basic terminology on the graph theory will be valuable for the reader to become familiar with the terms and be able to understand the suggested new methods as they are described later in this chapter. One of the pioneers in graph theory was the Swiss mathematician Leonhard Euler. In 1736, he solved the problem of Konisberg bridges by turning land area into vertices (nodes) and the bridges into edges (links). The problem was a very simple question concerning an interesting puzzle of the seven bridges and four land areas in the Prussian city: is it possible for someone to visit each of the four areas of land and return to her/his area of origin crossing each bridge once and only once? Turning the bridges and lands into a network, called a *graph G* in mathematical terminology Euler expressed the complex physical structures of the city into handful mathematical elements. The study of networks is called *graph theory* and can be applied to a range of research problems that can be reconstructed into vertices and edges and analyzed as such.

Mathematically, a *graph* $G=(V,E)$ (Figure4.2.a) consists of a set of *vertices V* (Figure4.2.b), and a set of *edges E* (Figure4.2.c). Each edge E_i corresponds to a pair of vertices (Figure 4.2.d). There are two types of graphs: the *undirected graph* (Figure 4.2.e) where the pairs of vertices are unordered and represented as simple line segments and the *directed graph* (Figure 4.2.f) where the pairs are ordered and represented as arrows pointing from one vertex to another. A sequence of consecutive edges in a graph is called a *path* and the *length of the path* is the number of edges traversed. In Figure 4.2.e the path highlighted in red is constructed by the edges $E_1(3,1)$, $E_2(1,2)$ and $E_3(2,4)$ and its length is 3 according the three edges traversed. For each pair of vertices, if there is a path from one to another, the graph is *connected* and if the graph provides two or more paths then is *disconnected*. In addition, more complicated graph systems are possible to be constructed such as multigraphs or weighted graphs. A *multigraph* is a graph with multiple edges between the same vertices and a *weighted graph* has value (*weight* or *cost*) associated with every edge in the graph. The *weight of a path* in a weighted graph is the sum of the weights of the traversed edges.



a) Simple graph G = (4,5) *b) Vertex v1*

c) Edge E1=(v1,v2) *d) Pair of Vertices (v1,v2) and Edge E1*

e) Undirected graph *f) Directed graph*

Figure 4.2: Basic graph components

In addition, a graph *H* is sub-graph of *G* (or $H \leq G$) if $V(H) \subseteq V(G)$ and every edge of *H* is also an edge of *G*. The similarity of two graphs *H* and *G* is called *isomorphism* ($H \rightarrow G$) and if the isomorphism exists ($H \cong G$) then all theoretic properties of both graphs are identical. For further exploration of the extent of isomorphism between graphs three basic relationships between vertices and edges are introduced here based on the literature for network analysis (Dijkstra, 1959; Garrison, 1968; Haggett and

Chorley, 1969; Hayes, 2000; Tidswell, 1976) providing structural information on graphs:

a) There exists a minimum number of edges needed to connect all vertices and this can be calculated by equation 4.2.

$$C_{\min} = v - 1 \tag{4.2}$$

where, v is the number of vertices and graph G is minimally connected with $C_{\min}$ vertices.

b) The possible ways the vertices in a graph may be connected is defined as the factorial number of vertices divided by 2:

$$C_{\max}^{v} = \left(\frac{v!}{2}\right) \tag{4.3}$$

where, v! is the factorial number of vertices and it is divided by 2 because each edge is traversable in both directions.

and c) There is a maximum number of edges providing a *complete connectivity* without any edge becoming redundant and this can be calculated by equation 4.4.

$$C_{\max}^{e} = \frac{v^2 - v}{2} \tag{4.4}$$

Furthermore, an indication of how well vertices are connected in a particular graph can be provided by three measures: the cyclomatic number, the beta index and the efficiency of the network (graph).

The *cyclomatic number* relates the number of edges, vertices, and sub-graphs and it is expressed by equation 4.5.

$$C_{Cycl} = e - v + g \tag{4.5}$$

where, *e* is the number of edges, *v* the number of vertices and *g* the sub-graphs.

The *beta index* is one of the most important indices for measuring connectivity. It is formulated by dividing the number of edges by the number of vertices as follows:

$$I_{\beta} = \frac{e}{v} \tag{4.6}$$

where, *e* is the number of edges and *v* is the number of vertices.

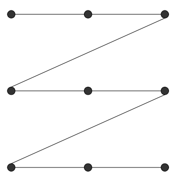
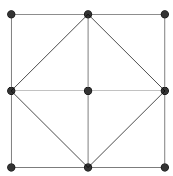
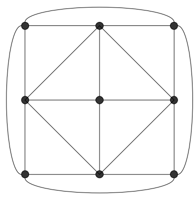
The values of beta index I_{β} vary according to the type of the graph. Where the graph contains more edges than vertices then the index value is more than 1, showing that the graph G has strong connectivity. Where both edges and vertices are equal then the value is 1 suggesting an average connectivity whilst a graph *G* with many vertices has value less than 1 showing a weak connectivity. According to the literature, the beta index is considered as a sensitive indicator of growth in a graph.

The *efficiency of the network* can be expressed as percentage, comparing the observed number of edges with the maximum possible edges (from equation 4.3), as shown in equation 4.7:

$$G_{Efficiency}(\%) = \frac{e}{C^{e}_{\max}} \times 100 \tag{4.7}$$

where, *e* is the number of edges and *v* the number of vertices. Any value in excess of 100% indicates additional edges in the graph which may be removed without detriment to the efficiency of the graph as a whole. The three connectivity measures are compared in Table 4.1 using nine vertices connected in three different ways. The first graph *G(9,8)* is a minimum connected graph and represents the lower level of connectivity with 0, 0.9 and 22.22% cyclomatic number, beta index and graph efficiency, respectively. As long we enrich the graph G connectivity by adding more edges the three different measures will improved measuring better connectivity within the graph (Table 4.1).

Table 4.1: The threefold measure of connectivity in three graphs with same number of vertices (V=9).

Graph	Attributes		Connectivity measures		
Graph *G=(V,E)*	Vertices (*V*)	Edges (*E*)	Cyclomatic Number (*Cycl*)	Beta Index (I_{β})	Efficiency of Network $G_{Efficiency}$ (%)
	9	8	0	0.9	22.22%
	9	16	8	1.7	44.44%
	9	20	12	2.2	55.55%

Graph theory involves a valuable bank of notions concerning the graph structure and the characteristics of each element in a graph. The short exploration of graph theory here is important when it is associated with structural objects of a spatial aggregation such as areal units and zones. The above terms can be linked with spatial characteristics and provide an advanced methodology for efficient spatial algorithms. For example, graph *G* can be the representation of a study area, each vertex *V* the centroid of an areal unit with edge *E* corresponding to the relationship between two areal units. When areal units are aggregated to zones, then the graph of the study area can be separated in sub-graphs equivalent to the number of zones. In this research, spatial problems are treated as relationships (edges) between areal units (vertices). A good example of this way of handling spatial information is the migration flows that are recorded between neighbouring areal units. In this case, a graph representation would be equivalent to a multigraph *Gm* with directed edges. Moreover, the weighted graphs could be areal units connected with a specific weight of connectivity or weight based on social interaction between areal units. The link of aggregation problems with graph theory describes the aggregation problems from a mathematical point of view and introduces areas for study such as compactness and contiguity stability that were difficult or sometimes impossible to handle with the conventional approaches.

4.2.3 Zone Design as a graph theory problem

Zone design as defined earlier provides a very broad concept for the needs of this thesis. The use of graph theory is appropriate here to provide an enhanced formulation for each spatial object. As a result, terms such as areal units, zones, boundaries, and other zone design properties are mathematically defined here using the graph terminology.

In this thesis, *areal units* are defined as vertices *V* of a given graph *G* according to the graph theory concepts. Graph *G* represents the whole area of study and when we are referring to its vertices, it is termed as $V(G)$. By substituting areal units with graph vertices, the connectivity between areal unit pairs (known as *contiguity*) can be represented by edges *E* of graph *G* (or $E(V)$). The existence of edge *E* defines *connectivity* between two areal units with possible additional information stored on each edge such as the direction of a migration flow or the strength of a boundary. The concept of strength of boundaries is introduced at a later stage in this Chapter. On the other hand, the use of flows between areal units extends the field of the current study into topics of migration analysis that are not explored in this thesis.

A *zone* can be represented by a sub-graph (*Hi*) of graph (*G*) and always for each $\forall Hi \in G$, *Hi* is smaller than *G* ($Hi < G$). It follows that given a set of sub-graphs (H1,.., Hi) can be further defined the *initial aggregation* of zone design if only $\forall Hi \in G$, the $V(Hi) \subseteq V(G)$ and every edge of *Hi* is also an edge of *G*. In addition, the $H_1 \cup ,\ldots, \cup Hi$ should be equal to graph *G.* Having both conditions right then the initial aggregation is feasible to be defined as the union of all sub-graphs $(H1 \cup ,\ldots, \cup Hi)$.

According to the graph theory, the optimisation of an *objective function* is expressed as a minimisation of a given function $F_{0,}$ where $F_0(x) \rightarrow 0$, $\forall\ Vi \in Hi$ and x is the vector of attribute values of all vertices. For example, each vertex (Vi) in the sub-graphs (Hi) can carry one or more variables (x). Therefore applying in sub-graphs *(Hi)* a homogeneity function, it is possible the minimisation of within variation in terms of the selected variables. The use of the same principle can be applied for the definition of *weighted boundaries* although here the minimisation of a given function $F_{0,}$ where $F_0(x) \rightarrow 0$, is for each $Ei \in Hi$ and x is the stored information of each edge Ei. It follows that the two basic elements of a graph, vertices and edges, are considered as structurally dynamic forms of objects adapting easily exogenous information.

Furthermore, the concept of *contiguity stability* of zone is important to be formulated according to the graph terminology. Therefore, a *contiguous zone* is characterised by a subgraph Hi when the relocation of an edge $Ei \in Hi$ to an adjacent sub-graph Hn does not influence the connectivity of the remaining edges in Hi. It is expected that Hi and Hn share a number of borders, in other words they share a number of vertices $V(G)$. In addition, the relocated edge Ei should be selected from the vertices $V(G)$ that connect the edges $Ei(Hi)$ and $En(Hn)$ of two neighbouring subgraphs (Hi, Hn).

The definitions introduced here are vital for the understanding of graph theory in a zone design context. The way that graph theory defines the spatial properties of zone design provides hierarchical structured elements with comprehensive characteristics. As a result, zone design here embraces the old features of previous research and expands them using concepts of graph theory.

4.3 Building an O-O Zone Design system

Describing the zone design components using graph theory terms provides a strong base for the later stages of zone design implementation. In this thesis, the development of a zone design system has been based on Object Oriented architecture (O-O). The O-O design is a very efficient way to construct natural or conceptual environments using concepts that later programming set of tasks can implement. According to the literature

(Budd, 2001; Graham, 2001; Rumbaugh et al., 2004), the O-O environment consists of the following basic concepts. *Objects* are the natural and conceptual elements that exist in the universe; trees, houses, roads, even people are objects under the O-O conceptualisation. The use of object as the basic unit in the O-O architecture and at a later O-O programming stage provides well organised structures. An object has state and behaviour. The *state* of an object is a set of characteristics that describe it. For example, the state of an aggregation process is defined by characteristics such as the zones, compactness and objective functions. The *behaviour* of an object is its actions which affect either itself or other objects. Therefore, in the example of aggregation the behaviour of the object 'aggregation' is consisted of a set of action such as the zoning. States and behaviours in O-O programming languages are termed '*properties*' and '*methods*' respectively. The grouping of all states and behaviours under an object is termed *encapsulation*. In addition, a collection of objects are capable of interacting each other and this kind of communication is known as *message*. For example an aggregation is not possible to perform the behaviour of zoning without a contiguity controller, who interacts with the zoning following a certain number of actions for securing the contiguity stability of zones. The general formulation of states and behaviours to all objects belonging in the same category is expressed by a *class*. A class creates identical objects such as a car factory makes cars based on the blueprints of designer. Moreover, every instance or state in a class could be itself a class and the parental class is called *metaclass*. An essential concept in O-O programming is the ability of a class to inherit specific states and behaviours from its superior class. *Inheritance* provides a powerful mechanism for organising and structuring software programs. Although additional O-O concepts are available, in this study the above terms are the most important for the reader to become familiar with the O-O terminology and how it is applied in the zone design system.

4.3.1 Experimenting in O-O programming

The development of a new zone design system based on contiguity principles, graph theory and O-O programming was a challenging process as during the implementation a number of parameters were taken into account. The new system should perform better in terms of aggregation time process, formulated under the graph theory and implemented according to the O-O architecture. Therefore, a beta version was essential to examine if such a system is feasible. The beta version of zone design was developed in Visual Basic 6.0 using the Map Object 2.0 library of ESRI. The Map Object library is a collection of classes that perform spatial analysis tasks. The developer can import a range of different file formats such as Arc/INFO coverages and shape files, map the datasets and carry out spatial analysis according to the needs of each study. Moreover, the development of zone design system focused on an O-O system handling the zone design components as objects. Therefore, we introduced two spatial levels during the process of aggregation. The lower level consists of the areal units which provide the finest geographical units and the upper level is the output zones deriving from the execution of the zone design algorithm. The hierarchical organisation of a zone design system according to the O-O principles is archived by structuring three classes. These classes describe three identified objects: an areal unit, a zone and the output aggregation. In Figure 4.3, the hierarchical organisation of zone design classes is illustrated showing at the tree root the class that describes the characteristics of each aggregation. In addition, the 'areal unit' class is partly involved in the spatial description of 'zone' objects, while 'zone' objects provide the spatial information of 'aggregation' class.

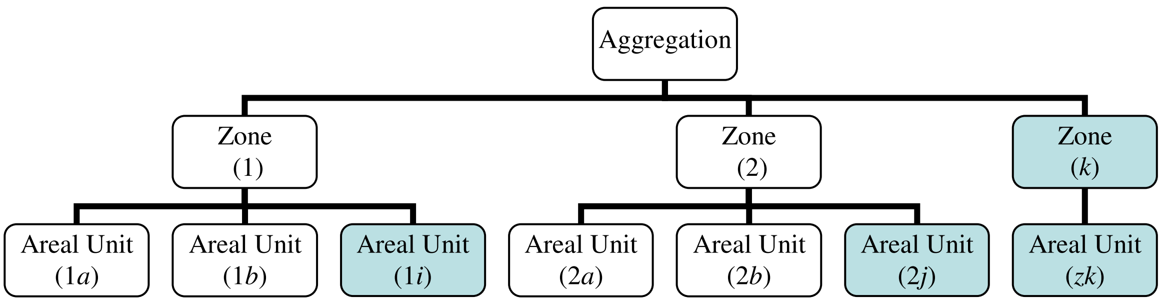


Figure 4.3: The hierarchical organisation of zone design objects

The beta version of the zone design system adopted the above hierarchical structure of classes describing the zone design components into each class feature. Table 4.2, introduces the properties and methods of each class. Class 'AreaA', for example, constructs objects that represent the areal units of a study. The produced object carries a unique identification number, a collection of the neighbouring areal units representing the local adjacency information and the available data variables referring to the areal unit. Furthermore, class 'ZoneZ' creates objects that consist of a group of contiguous areal units storing adjacency information of neighbouring zones and areas as well as tabular information at the zone level. In addition, 'ZoneZ' is also a metaclass as it retrieves adjacency information from the 'AreaA' classes. Finally, class 'Aggregation' is the primary class in the zone design system, because it consists of all the characteristics of the aggregation process and the aggregation process itself.

Table 4.2: The structure of classes as developed in the beta version of zone design system.

Class Name		***Description***
AreaA	***Properties***	
	AreaId	*Internal area ID*
	Cont_Areas	*Neighbourhood areas*
	Data_Vars	*Data variables assigned to area*
ZoneZ	***Properties***	
	ZoneId	*Internal zone ID*
	Areas	*Collection of 'AreaA' objects*
	Cont_Zone	*Neighbourhood zones*
	Cont_Areas	*Neighbourhood areas*
	Data_Vars	*Data variables assigned to zone*
Aggregation	***Properties***	
	Zones	*Collection of 'ZoneZ' objects*
	TargetZones	*Number of the target zones*
	TargetPopulation	*Number of the zone target population*
	TotalPopulation	*Zone's total population*
	MaximumRuns	*Maximum number of runs*
	MaximumIterations	*Maximum number of iterations*
	IdleRuns	*Accepted idle runs*
	IdleIterations	*Accepted idle runs*
	OldBestScore	*Old Objective Function score*
	NewBestScore	*New Objective Function score*
	ContiguityControl	*Selected method of contiguity control*
	ObjectiveFunction	*Selected Objective Function*
	Methods	
	ExportZones	*Exports the available set of zones*
	ReadZones	*Imports a set of zones from ASCII or SHP files*
	ReadPropertiesFile	*Imports zone design settings from an ASCII file*
	RandomZoning	*Builds a random aggregation based on the IRA method*
	Zoning	*Performs the main aggregation process*
	WritePropertiesFile	*Exports zone design settings from an ASCII file*

During the development of the zone design system, each algorithm was evaluated based on specific small datasets of less than 500 basic areal units. The beta system archived excellent performance at this level, using the basic zone design procedure. Additional investigation of its potential limitations was essential as the next step of this research was to expand the capabilities of the system by implementing advanced aggregation criteria. In a case study presented at the GISRUK conference (Alvanides et al., 2001), the zone design approach was adopted to produce output zones that minimise the variation of population mixing in Cumbria County. The beta version of zone design was used to aggregate 9,375 thiessen polygons and 1,192 enumeration districts into 171 homogeneous zones. The number of output zones was decided upon the existing 171 census wards in Cumbria. Even though the output zones were derived without difficulties, an exponential increase of processing time was observed especially during the aggregation of thiessen polygons. In-depth exploration of the zone design process highlighted the problematic areas: although, the theoretical implementation of the zone design system was correct, in practice the ‘collection’ object of VB programming language was incapable of performing aggregations with more than 2,000 areal units.

Moreover, during the above case study it became aparent that more efficient objective functions are needed. Regularly, the use of areal units results in islands with different kind of adjacencies, such as separation of areal units by natural boundaries (rivers). One of the most important issues of the study was the answer to a very simple question: *How many output zones are needed for a particular problem and why?* In fact, this question is the most difficult part of a zone design system (Openshaw, 1977b) because it has been noted that as the size of areal units increases the measure of association tends to increase. In addition, as larger areal units are used, the area tends to contain a more heterogeneous population and descriptive statistics become less representative of the population. On the other hand, if the smallest possible areal units are used in an attempt to ensure homogeneity of the population, the analysis is likely to suffer from small number effects. Therefore, a new approach of comparing the variance within zones at different aggregation levels is essential. Furthermore, alternative methods of initial grouping and shape compactness could provide better results, targeting specific characteristics of a study.

4.3.2 Areal units TO Zones system (A2Z)

During the evaluation of the beta zone design system, we collected valuable information concerning different aspects of the aggregation process as noted in the previous section. One of the most crucial issues was an increase of the processing time as the number of areal units increased. To a certain extent this is expected as many areal units provide more complicate and hard to solve aggregations but in the beta version of zone design the increase of processing time was extremely high, consuming computer resources. In the final version of zone design (A2Z), we tackled the problem by using alternative programming methods. In detail, the problem was targeted using a specific object of VB language. The object known as 'collection' is widely used in VB but in cases with many items it becomes significantly slow to access, during the retrieval of its stored information. To tackle this problem, the use of an 'array' instead of a 'collection' object is employed in the contiguity algorithms of the system, saving valuable processing time that can be used for exploration of further optimal aggregation solutions.

Securing that the A2Z system operates efficiently under aggregation problems with more than 2,000 areal units, we realised that additional zone design components can provide essential aid during the aggregation process targeting specific issues. As a result, the A2Z system has been developed to manipulate complicated adjacency relationships before the main zone design procedure performs the aggregation task. In addition, statistical measures of goodness of fit together with zone design system provide a valuable approach of informative selection of the appropriate aggregation level. During the zone design process one of the most time consuming functions is the one controlling for contiguity stability. In the A2Z system, a range of different approaches have been developed providing fast contiguity checks, especially in large datasets. Although, zone design performance has been improved considerably, additional optimisation functions were necessary to explore new ways of designing homogeneous zones. Furthermore, the A2Z provides tools to control for compactness of zone shapes using techniques based on the adjacency and the graph information of output zones. More detailed description of the above methods is presented in the following sections, providing additional information into alternative approaches as suggested in the relevant literature.

4.4 Advanced contiguity construction

4.4.1 Build Region Contiguities (BRC)

One of the limitations of previous versions of zone design systems was the handling of complex geographical features. For example, in Scotland there are many lochs and islands resulting in geographies with complicated adjacencies and the aggregation of these areal units by means of old zone design systems was almost impossible. The Build Region Contiguities (BRC) method was developed to tackle this problem and it is based on the concept of 'regions' as used in ESRI's specifications for spatial structures (ESRI, 1999). A *region* is a group of areal units recorded as a single feature having natural or arbitrarily assigned boundaries. This kind of spatial construction often appears during aggregation tasks, especially when natural object separate subjective areas (e.g. social areas) into multi-part polygons.

In order to express the region approach mathematically, the problem is illustrated as a given disconnected graph *G'*. According to the zone design principles, the areal units of each study should provide a connected graph. Therefore, instead of using the disconnected areal units the BRC method constructs a connected graph *G* using regions as features. The resulting graph *G* has fewer vertices *V(G)* than those vertices *V(G')* of the disconnected graph *G'* because each vertex *V(G)* consists of many vertices *V(G')*. As a result, graph *G* is better connected and consequently the aggregation problem becomes smaller than in graph *G'*.

For a better understanding of the BRC method, a hexagonal tessellation of 372 areal units with 5 underlying patterns (Figure 4.4.a) is tested. The example has been overlaid by two rivers resulting in the spatial partitioning of 24 areal units. The underlying clusters of three classification values have been crossed by the rivers for the evaluation of the BRC method. At this point the use of Arc/INFO software is necessary for producing the region-built coverage. The resulting coverage should carry the pattern classification information because the analysis takes place at the region level and any information stored in the partitioned areal units is not considered. The next step is to perform the aggregation based on a selected objective function. In this example, the aggregation process is repeated for three different aggregations: four, five and six output zones respectively, forcing the system for an exhaustive solution. The homogeneity objective function has been selected as the appropriate function in order to capture the hidden patterns; further explanation of this objective function will be introduced in a later section of this Chapter. There are five patterns, as illustrated in Figure 4.4.a, but due to the interference of physical features (rivers), traditional aggregation systems identify eight patterns. Exploring Figures 4.4.b, 4.4.c and 4.4.d it becomes apparent that the BRC method effectively overcomes the aforementioned problem for any aggregation level. In all three aggregations, the new method overcomes the problem of split areal units, resulting in clearly identified patterns that reflect the original dataset (Figure 4.4.a).

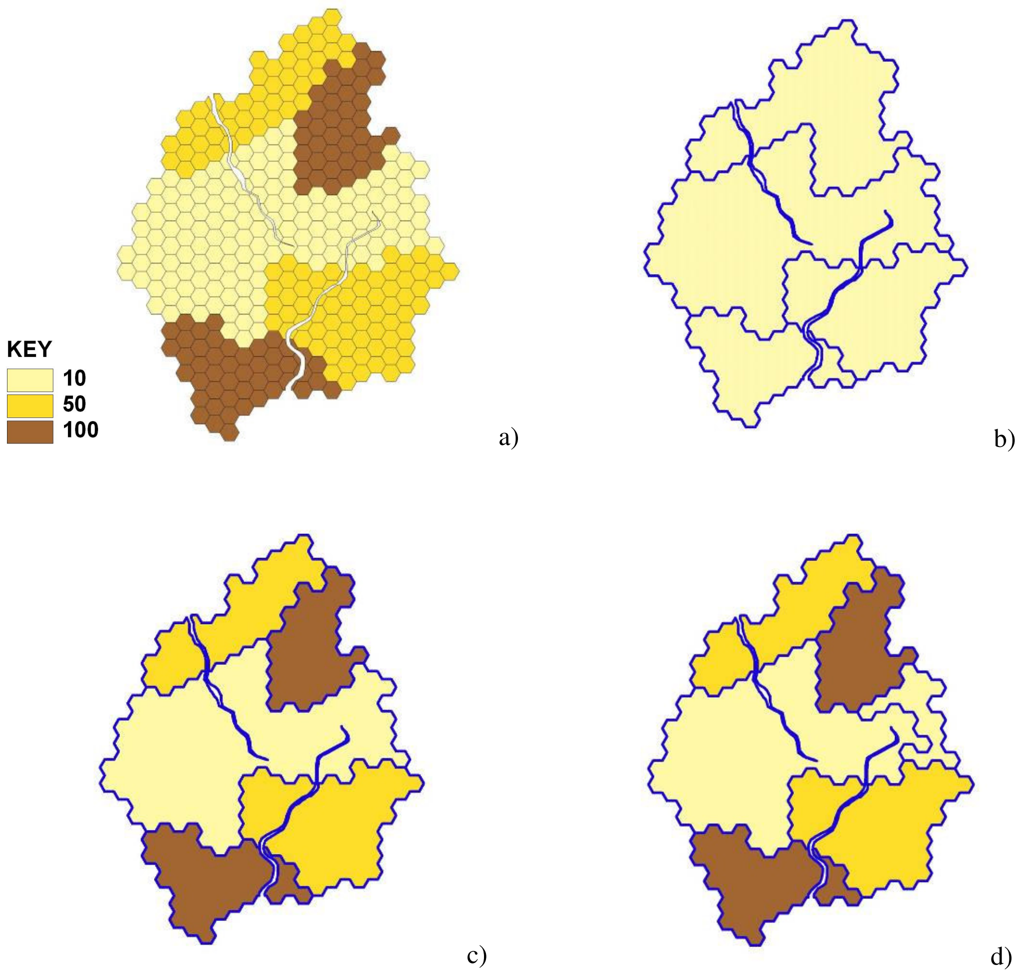


Figure 4.4: The Build Region Contiguities method applied on a hexagonal tessellation (a) building four (b), five (c) and six zones (d).

In the A2Z system, the BRC method handles the regions of a study area as areal units. The system first identifies all the regions and then traces the spatial relationships between them. The produced adjacency file is free of the spatial problems concerning rivers or other natural features that usually separate tessellation structures. Treating areal units as regions overcomes special and spatial features of areal units and at the same time provides a significant decrease in the number of areal units derived from larger datasets. The BRC method could be applicable using a region-built coverage as specified by ESRI and it is extremely important especially in datasets with high detailed geographical boundaries. Although, there is an issue concerning the reliance of system on ESRI products, the BRC method can also accept classification information on top of the basic topology using the areal units instead of regions.

4.4.2 Connect non-Connected Areas (CnCA)

In the previous method the focus was placed on areal units that represent not only one area but a group of separated areas using the concept of regions. Although the BRC method performs well on detailed datasets partitioned by physical structures, there are cases where the concept of regions is not applicable. For example, in cases where islands are linked to the mainland the BRC method solves partially the problem by grouping islands with areal units in the mainland. However, if such links are to be established using features from the existing transport network such as bridges and ferry lines then the BRC method is inappropriate.

Mathematically, the problem is illustrated either as a given disconnected graph G aiming for a new connected graph G' or, alternatively, as a given connected graph G targeting an enriched graph G' in terms of connectivity between certain edges Ei. In the former case, new vertices Vn provide links between isolated edges Ei and the rest of graph edges Em. Each vertex Vn connects a pair of edges Ei and Em improving the graph connectivity until all isolated edges Ei become connected within graph G. In the latter case, the objective is to improve the existing connectivity of a graph G by adding vertices Vn. Such a task builds or increases the connectivity of selected edges Ei. In both cases, the vertices $V(G) < V(G')$ and the edges $E(G) = E(G')$ suggest a better connected graph G'.

The Connect non-Connected Areas (CnCA) method tackles the above issue by taking into account features of existing transport networks. There are various spatial connections according to the topological concepts of relations between lines and regions (Egenhofer and Mark, 1995). In this approach, a simple intersection of line edges with regions is recommended as appropriate spatial relation. The intersection should be possible at both ends of the line (e.g. bridge) with both regions it connects. If such an intersection exists then the CnCA method updates the adjacency file adding a new contiguity between the intersected regions.

In Figure 4.5.a, an example of hexagonal areal units is used to explore the effectiveness of the CnCA method in the A2Z system. Each polygon has a value of 10, 50 or 100 and the whole tessellation hides five clusters. In addition, there are two islands and a river in the dataset therefore traditional zone design systems (ZDES and AZP) are incapable of aggregating the areal units. At a first glance, the example is likely to provide eight clusters because of the existence of natural boundaries. But as the ferry lines and the bridges of the river are attached to the dataset, the isolated areas become neighbours with the larger tessellation of hexagons. The CnCA approach updates the adjacency file of this study area using the linear features of the transport network. As a result, the A2Z system utilises the new adjacency information and aggregates the polygons into four, five and six homogeneous zones. In Figure 4.5.b, zone design built four output zones capturing the patterns accurately even if the number of zones is lower than the actual clusters of the example joining the clusters with values of 50 and 100.

Aggregating the tessellation into five output zones, the system identifies the five clusters without difficulties. In Figure 4.5.d, an aggregation of areal units into six output zones is illustrated. Over again, the produced zones capture the patterns accurately by splitting the pattern with values of 10 into two output zones, as the targeted zones are six and the available patterns are five. The common characteristic of these different scales is the ability of the CNCA method to inform the A2Z system of the additional relationships between regions.

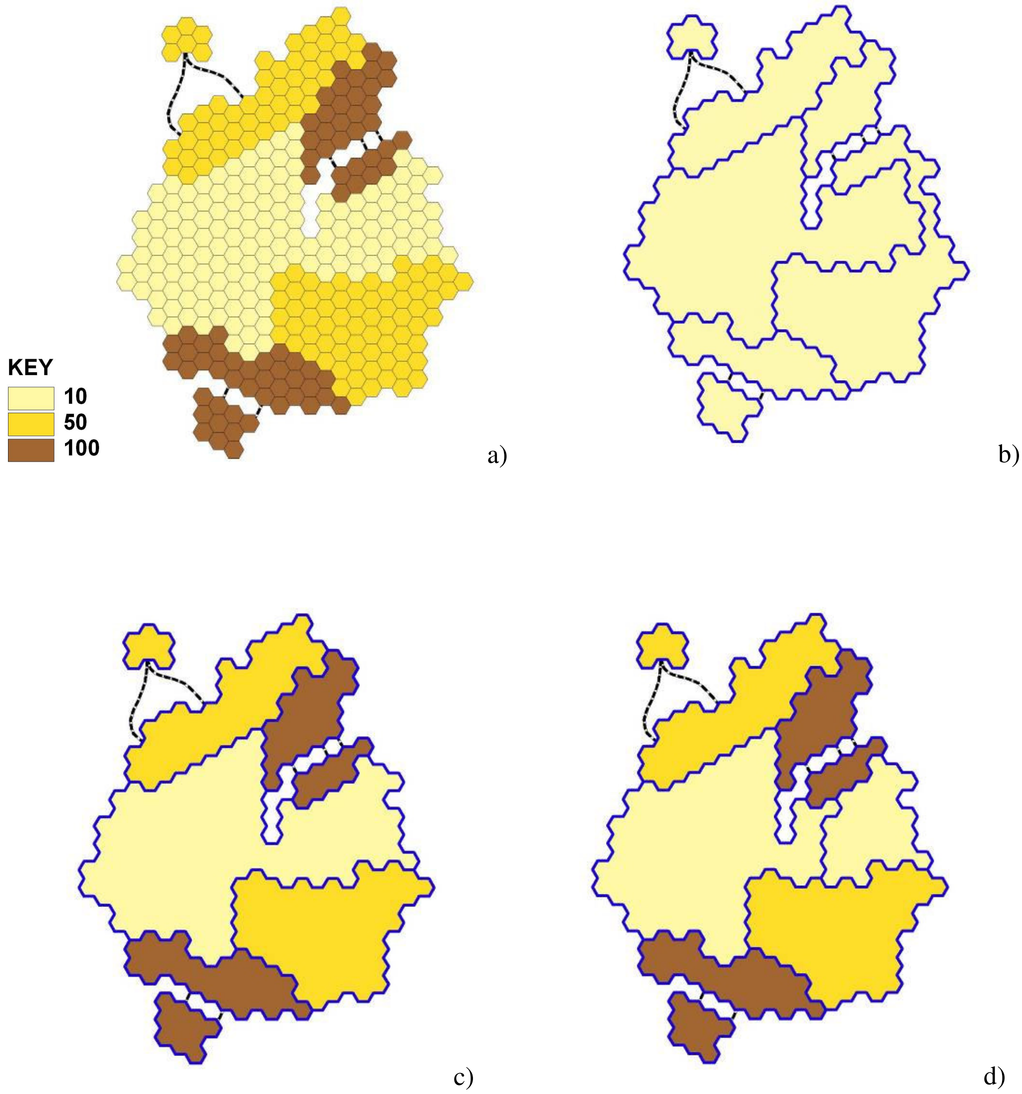


Figure 4.5: The Connect non-Connective Areas method applied on a hexagonal tessellation (a) building four (b), five (c) and six zones (d).

4.4.3 Weighted Adjacency (WA)

In the previous sections, specific issues concerning the construction of the adjacency file have been introduced. Both BRC and NCA methods effectively tackle spatial difficulties, such islands and natural boundaries. However, according to the graph theory and the structure of the A2Z system, it is also possible to construct weighted adjacencies between the areal units. Certainly, the way that the weights are defined and applied on the boundaries is very important. For example, Boots (2001) suggested a methodology for identifying the 'strongest' boundaries between areal units and at the same time to explore the effect of scale in the strength of each boundary. The weight of each boundary was calculated by measuring the Getis local statistics (Fotheringham, 1997; Getis and Ord, 1996) of an areal unit kernel in which its boundary is included. In general, it is recognised that boundary detection is influenced by scale, especially in vegetation boundaries of ecological studies (Fortin et al., 2000; Fortin and Edwards, 2000).

A more straightforward approach is to weight the boundaries according to their length or by measuring the number of intersections between the boundaries and existing transport networks. To test the relationship between length and intersections of road network we used the road network of the Tyne and Wear County in England (Figure 4.6) at two different scales: the 113 wards and the 719 super output areas. The road network was measured first as a whole, including motorways, A roads, B roads and minor roads and then as B and minor road network. To investigate if there is any relationship between length of boundaries and number of roads intersecting these boundaries, we measure the correlation coefficients at both SOA and ward levels. In table 4.3, the length of boundary is significant at the 0.01 level in all different network classification and at both SOA and ward levels. At SOA level, the whole road network is significant to length boundary at 0.538 comparing to 0.517 of B and minor road network. Slightly worst, at ward level the length boundary is correlated to whole road network at 0.443 and to B and minor road network at 0.429. The correlations at both network classifications perform similar while in terms of scale, the effect seems to be slightly better at SOA level.

Table 4.3: Correlation between length boundary and road networks at SOA and ward levels

	SOA boundary length	**Ward boundary length**
Whole road network	0.538**	0.443**
B & minor road network	0.517**	0.429**
Number of boundaries	2115	303

** Correlation is significant at the 0.01 level (2-tailed)

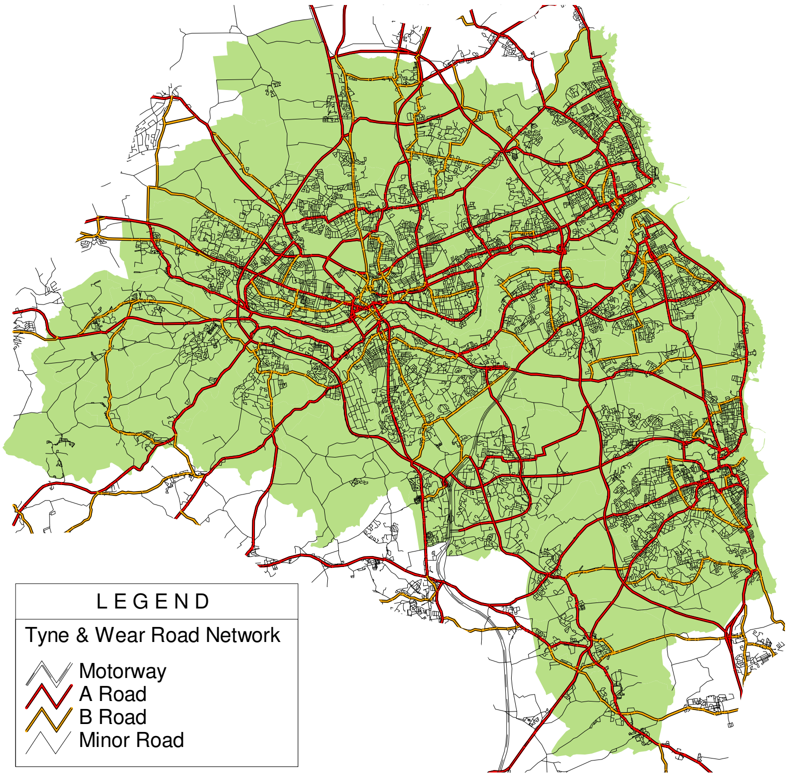


Figure 4.6: The road network in Tyne and Wear

Although the correlations are significant at the 0.01 level, the coefficients are not high enough to support the case that there is a strong relationship between length boundary and road network intersection counts. Therefore, we further explored this relationship using boxplot graphs. In Figures 4.7.a and 4.7.b, the boxplots show a small increase of intersection counts as the length of boundary increases at SOA level but in both network classifications there exists a large number of outliers suggesting a very weak relationship. Investigating the boxplots at ward level (Figures 4.7.c and 4.7.d), there is no clear relationship between boundary length and road network counts. In this case, there are less than ten outliers per graph showing a better distribution but the overall boxplot results suggest a random association. The conclusion of this experiment is that there is no relationship between boundary length and road network counts. As a result, the use of boundary length for measuring the connectivity between areas (SOAs or wards) is inappropriate. Therefore when there is need for boundary weights we suggest the measure of intersected transport network features with the underlined area boundaries.

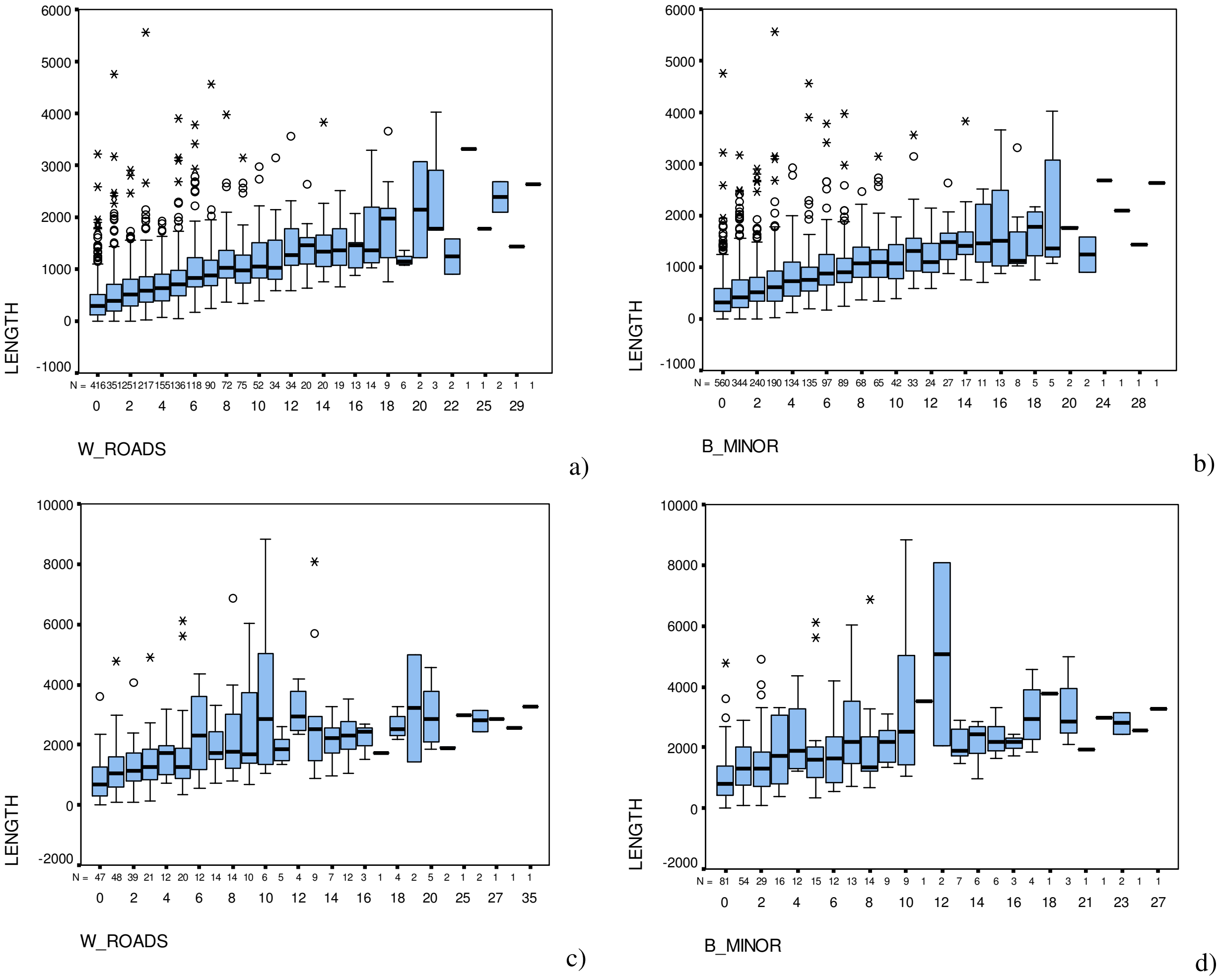


Figure 4.7: Relationship between boundary length and road network intersections: a), b) for whole and partial road network at SOA level respectively, c) and d) for whole and partial road network at ward level respectively.

It can be argued that each approach of weighted boundaries definition is target different issues and therefore the appropriate solution should be specific to each case study. However, it is important for an aggregation system to take into account boundary weights during the aggregation process as the use of boundary information provides better representation of natural or subjective issues. As a result, measures of weights can be associated to the adjacency file producing boundary weights for each areal unit. Mathematically, each 'boundary weight' is defined as the value W stored in a given edge E, $\forall E \in G$. Having all the edges $E(G)$ a value ranged from 0 to 1, where 0 and 1 represent disconnected and connected edges respectively , it is possible to construct sub-graphs H with homogeneous weights W by minimising the differences between the mean of weights W_{mean} (members of a sub-graph H) and the weights W within H. The Weighted Adjacency (WA) function is formulated as:

$$F_{WA} = \min \sum \frac{|W_{mean} - W|}{N} \tag{4.8}$$

where, N is the number of edges E in a sub-graph H.

In a zone design context, the boundary weight is represented by the adjacency weight value between two areal units. The function is targeting an optimum homogeneity of weights included in each zone by minimising the within zone weight differences. The F_{WA} function is useful when the identification of zones with similar boundary weighted values is required. In cases where the output zones are required to have boundaries with the lowest possible weight values, meaning strong connectivity within zones, then we maximise the average of weights W in each zone. This can expressed as follows:

$$F^{*}_{WA} = \max \frac{\sum W}{N} \tag{4.9}$$

Although, functions F_{WA} and F_{WA}* may provide useful functionality in the A2Z system, it is also noted that the use of weights during the aggregation process results in additional constrains. Therefore, careful consideration of research needs is advisable, before any weighting functions are applied to a case study.

In Figures 4.8, we measure the B and minor road intersections with each boundary at ward level in Tyne and Wear. Using the F_{WA}* functions (equation 4.9), the A2Z system builds a new aggregation targeting 20 zones. In this example, we disable the objective function to avoid complications deriving from the optimisation of variables, thus performing only the F_{WA}* function optimisation. However, in the A2Z system the user is also able to apply the weight function in parallel with the optimisation of other criteria. In Figure 4.8, the produced zones do not cross any boundaries without connections, respecting the constraints we applied. In addition, the resulting aggregation of the F_{WA}* function (equation 4.9) creates output zones that reflect strong internal connectivity and this is confirmed by the boundaries of each zone as they all consist of weakly weighted or disconnected borders.

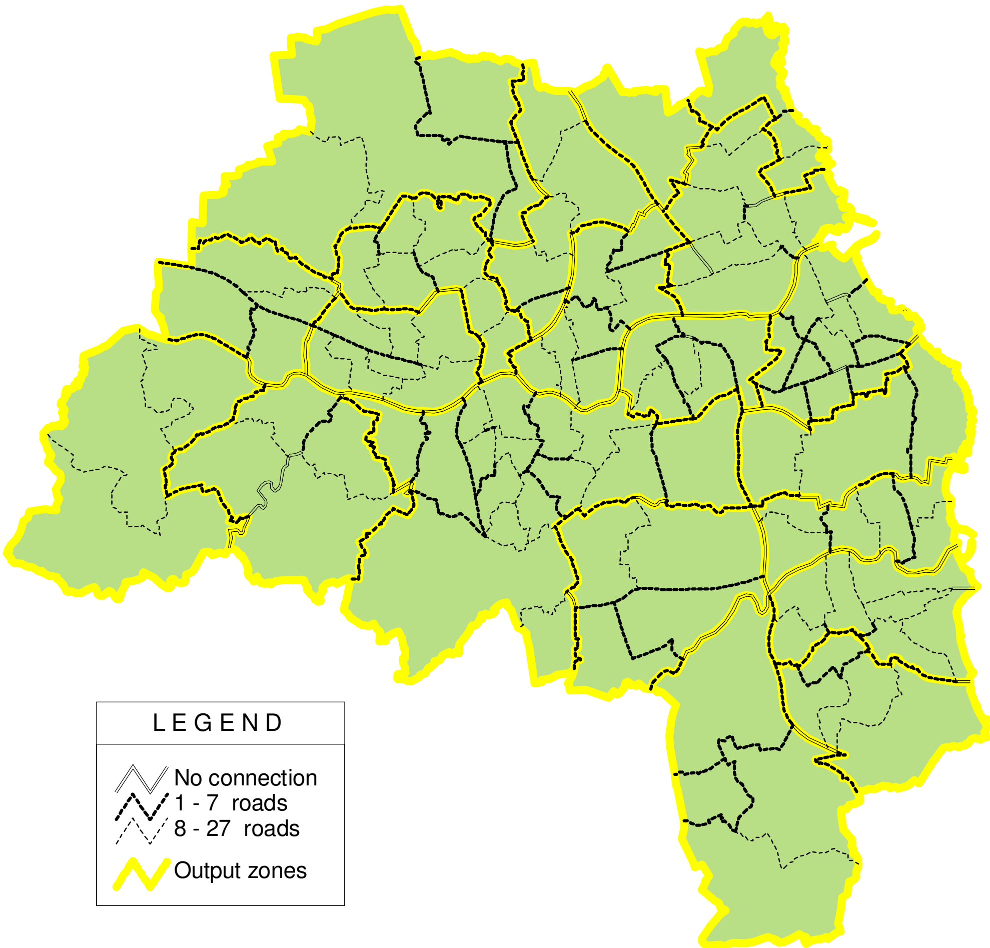


Figure 4.8: Output zones resulting from areal units weighted by the intersections between the ward boundaries and the B and minor road network in Tyne and Wear.

The WA approach extends the capabilities of A2Z in a large number of social and health issues that may require a more sophisticated methodology in aggregation problems. For example, during the latest UK NHS plan the DoH requested from each Health Authority to propose changes for a new organisation as Strategic Health Authorities. The government has established the following criteria for the construction of proposed Strategic Health Authorities: should serve populations of about 1.5 million on average, should be broadly aligned with clinical networks and should be conterminous with an aggregate of Local Authorities and not cut across Government

Office boundaries (DoH, 2001a). It is clear that all the above criteria are effectively provided by the WA method and the A2Z system. Therefore, the use of such tools is proposed here for consultation purposes, in addition to subjective knowledge (DoH, 2001b; DoH, 2001c).

4.5 Informative selection of zones

One of the questions that previous zone design systems were incapable of addressing is *how many output zones are needed for a specific aggregation problem*. Most zone design systems require the exact number of output zones or the output from an earlier aggregation. In this section, a methodology for zone size selection is presented using the informative approach suggested by Nakaya (1996; 2000) and expanding this approach through a zone design perspective. His methodology was based on the minimisation of the Akaike Information Criterion (AIC) running a procedure of sequential aggregation for a small dataset of 262 areal units. The aggregation level with the minimum AIC value has been suggested as the appropriate model because the goodness of fit is higher and the number of parameters is smaller according to the *principle of parsimony*. In addition, the proposed methodology explores two variations of AIC: the Bayesian Information Criterion (BIC) and the corrected AIC (AIC^C).

4.5.1 Zone selection using A2Z

The A2Z system is capable of identifying and proposing the most informative aggregation level using both the zone design algorithm and the Akaike Information Criterion. Nakaya's (2000) approach investigated all the possible aggregations by comparing the AIC values. In a small dataset such the 262 areal units of the metropolitan area of Tokyo, it is relatively straightforward to implement approach; however, the proposed aggregation procedure has not been evaluated when the number of areal units increases significantly.

The methodology presented here applies the zone design procedure for ten different aggregations, in an attempt to identify where the minimum information criterion is positioned. The ten aggregation levels are selected by starting from two zones and increasing to an additional 10% of total areal units. The segmentation of areal units assists the system to identify a range of aggregations that include the aggregation with the minimum information criterion. After identifying the range with the desired aggregation level the above steps are repeated for this range producing a sub-range. The final range in this iterative process consists of ten segments increasing by one zone, with one of these segments expected to produce the appropriate aggregation level.

For a better understanding of this process, the graph in Figure 4.9 illustrates the information criterion scores from 2 to N zones according a generic manner. The continuous line shows the information criterion values for all possible aggregations. If the number of areal units (A) is equal to ten then point Q is the aggregate level with the minimum information criterion, otherwise the algorithm selects the range between the segments Ax0.5 and Ax0.3 for further investigation. The range between these segments includes the minimum information criterion therefore an extensive exploration within this range using the previous steps can target the aggregation level with the minimum information criterion. An interesting condition the algorithm could face is when the desired aggregation level could be hidden between 2 zones and the first segment (Ax0.1). As the dashed line shows minimum information criterion recorded in the first segment (X point). In this case, further analysis will be focused in the range 2 to Ax0.1 attempting to locate the appropriate aggregation level.

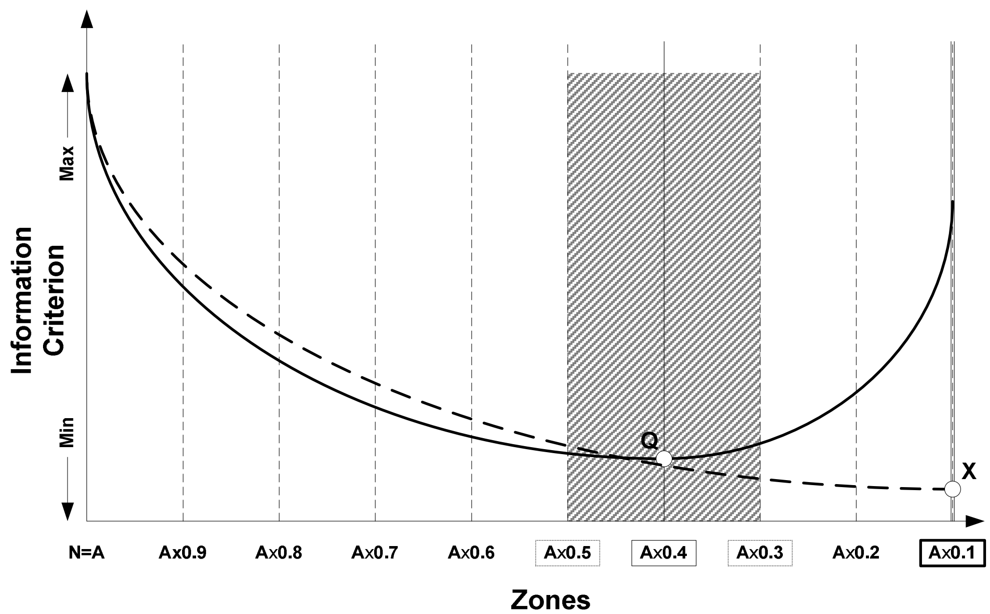


Figure 4.9: The selection of aggregation level according the minimisation of an information criterion.

4.5.2 Information Criteria

The above methodology supports any information criterion that follows the principles of parsimony which suggests that the best model should provide the highest goodness of fit with the minimum possible number of parameters. In addition, the selection of criterion needs to be validated according to the complexity of datasets investigating the effect of parameters in each criterion. In the following section, three criteria will be discussed and explored under the geographical aggregation perspective.

The information theoretic criterion based on the maximum likelihood principle was introduced by Hirotogu Akaike (Akaike, 1973) utilising a valuable criterion for comparison of statistical models. The Akaike Information Criterion (AIC) evaluates statistical models by taking into account the principle of parsimony which suggests as the best model, the one with the higher goodness of fit and smaller number of parameters. The AIC for a model (*m*) is defined as:

$$AIC_m = -2\hat{\ell}_m + 2p_m \tag{4.10}$$

where, $\hat{\ell}_m$ is the maximum *log*-likelihood of model (*m*) and p_m is the number of model parameters.

The use of formula 4.10 in a zone design system is proposed here as an appropriate method to compare aggregation outputs (*A* and *B*) considering the goodness of fit ($\hat{\ell}_A$ and $\hat{\ell}_B$) and the number of zones (*Z*). Here the number of parameters is identical to the number of zones because according Nakaya's research (Nakaya, 1996) the estimated values of maximum log-likelihood in a Poisson Dummy Model are calculated taking into account the observed rates in each zone. The formulas for the aggregation outputs *A* and *B* are as follows:

$$AIC_A = -2\hat{\ell}_A + 2Z_A \tag{4.11}$$

and

$$AIC_B = -2\hat{\ell}_B + 2Z_B \tag{4.12}$$

where, $\hat{\ell}_A$ and $\hat{\ell}_B$ the maximum log-likelihood of aggregation levels A and B, while Z_A and Z_B are the number of zones in each aggregation solution.

The aggregation solution with the minimum information criterion between two aggregation solutions A and B is the best one as its goodness of fit is high while its number of zones is small.

Since Akaike introduced the information criterion formula, a series of alternative criteria have been developed targeting less biased approaches. In this research the focus pointed on two variations of AIC: the corrected AIC (Hurvich and Tsai, 1989; Sugiura, 1978) and the Bayesian Information Criterion (Burnham and Anderson, 1998).
Both variations of AIC are based on the following generic formula substituting the *k* parameter:

$$AIC_{GenericFormula} = -2\hat{\ell} + kp \tag{4.13}$$

The construction of the classic AIC in formula 4.10 is possible for *k* = 2 As Hurvich et al. (1998) suggested, the corrected AIC (AIC^C) is less biased than the classic AIC. The main advantage of $\mathrm{AIC^C}$ is the way model parameters are calculated. The number of observations (*n*) is considered for the parameter estimation as follows:

$$AIC^C = -2\hat{\ell}_m + 2p\left(\frac{n}{n-p-1}\right) \tag{4.14}$$

or equivalently,

$$AIC^C = AIC + \frac{2p(p+1)}{p-n-1} \tag{4.15}$$

In this case, the *k* parameter of generic formula 4.13 is equal to $\frac{2(p+1)}{p-n-1}$.

The number of observations (n) in a zone design context, is the number of areal units (A) of study area. The formulation of equation 4.14 can be expressed as:

$$AIC^{C} = -2\hat{\ell} + 2Z\left(\frac{A}{A-Z-1}\right) \tag{4.16}$$

At this point it is important to mention that the above AIC formulas are for the cases of maximum likelihood estimation, but different estimations, such as least squares estimation (OLS) are beyond the scope of this thesis. Even though, the formula of AICc is derived from the assumption of Gaussian modelling and not Poisson modelling, the indicator is empirically useful for modelling (Nakaya et al, 2005).

The Bayesian Information Criterion (BIC) or Schwarz's Bayesian Criterion (SBC) is the second variant of AIC introduced here (Schwarz, 1978). The BIC estimator is applicable only if the assumed probability distribution is correct (Nakaya, 2001) and the complexity of the true model does not increase with the size of the data set (Burnham and Anderson, 1998). The BIC estimator can be expressed as follows:

$$BIC = -2\hat{\ell} + p\log(n) \tag{4.17}$$

As argued by Nakaya (2001:351) the BIC estimator is more biased (in terms of its parameters) than the classic AIC, but he notes that 'the bias could be negligible in large samples'. The argument by Burnham and Anderson (1998) that the BIC estimator can be used when the complexity of the true model does not increase with the size of the data set is in contrast with Nakaya's argument. In addition, Fotheringham et al. (2002) comment on both estimators without clearly concluding if the AIC is a better criterion than BIC.

The A2Z system supports all three estimators: AIC, AIC^{C} and BIC expanding the statistical abilities of classic zone design systems. Moreover, in the following sections and chapters, various comparisons of estimators will explore their behaviour according to small and large scale geographical case studies.

4.6 Initial Aggregation methods

One of the most important parts of any combinatorial optimisation method, and zone design system in particular, is the initial aggregation of areal units. An initial aggregation algorithm generates *Z* contiguous zones from *A* areal units (Openshaw, 1977a). As mentioned before, the appropriate selection of initial aggregation is beneficial for the main aggregation process providing quick and optimum results. Therefore, in this section, three different approaches are explained: the Initial Random Aggregation (IRA), the Initial Predefined Aggregation (IPA) and the Initial Predefined Zones (IPZ).

The IRA algorithm developed by Openshaw (1977b) provides a high degree of randomisation to ensure that the resulting aggregations are different during the iterations. In the A2Z system, the algorithm implemented according Openshaw's subroutine written in Fortran (Openshaw, 1977a). In comparison to Openshaw's IRA algorithm, the new IRA algorithm is implemented with O-O principles providing a quick random aggregation. The advantage of this approach is the use of objects instead of matrices avoiding the sustained sequential processes. For better understanding of modified IRA algorithm in A2Z system, it is recommended someone to go through the Table 4.2 where the zone design objects are described. The initial aggregation algorithm involves the following steps:

Step one: Random selection of *Z* areal units (*'AreaA'* objects)
Step two: Add *'AreaA'* objects into *Z 'ZoneZ'* objects
Step three: Copy the values of each *'Cont_Areas'* list into *'Cont_Zone'* list
Step four: For each *'ZoneZ'* object do

Start loop

a. Random selection of an *'AreaA'* object from the *'Cont_Areas'* list on condition that it is not member to another *'ZoneZ'* object.
b. Add selected *'AreaA'* object to the *'Areas'* collection
c. Update the *'Cont_Zone'* and *'Cont_Areas'* contiguity lists

End loop

Step five: Check if there is *'AreaA'* object without *'ZoneZ'* membership:

True: Invalid initial aggregation (possible isolated areal units in dataset)

False: Copy *'ZoneZ'* objects into *'Aggregation'* object.

The IPA algorithm is an alternative approach of initial aggregation utilising an initial set of areal units as the cores of later zones. The steps of IPA algorithm are the same as the IRA's approach with only difference in the first step. Instead of a random selection of Z areal units, it is replaced by a predefined set of Z areal units. Consequently, the user is authorised to subjectively specify the cores of each zone providing a user-biased basis for initial aggregation. For example, the use of IPA algorithm could subjectively expose a part of dataset into many zone cores providing more zones in certain areas rather the IRA algorithm could do. This could be the case of datasets with urban and rural areas. In this situation, the user could select more zone cores inside an urban area resulting to fewer cores in rural areas.

Although IRA and IPA algorithms are random aggregations with different starting points for zone cores, it may be desirable to use an existing zone set as initial aggregation. In the A2Z system this is termed the Initial Predefined Zones (IPZ) method and it can be beneficial when the goal is to improve an existing aggregation. However, the use of predefined zones as initial aggregation provides limited ranges of alternative aggregations because the zone design procedure is likely to be trapped in local optima. In general, the IPZ approach should be adopted when the related study is seeking a better rearrangement of neighbouring areal units between existing output zones.

4.7 Contiguity stability controls

The most important difference between a zone design system and a clustering system is the ability of the former to control the stability of contiguity during the aggregation process. In zone design system a fundamental part is to build zones as groups of areal units that are contiguous. To this end, many researchers developed interesting methods such as the matrix approach by Openshaw (1977b). In this section, four methods for securing contiguity stability are introduced. First, the Depth First Search (DFS) and

Breadth First Search (BFS) methods are described explaining the way to control contiguities using techniques deriving from the field of graph theory. At this point it should be mentioned that the adjacency matrix method as proposed by Openshaw's AZP is a variant of the depth first search routine. A disadvantage of Openshaw's routine is that the adjacency information is stored in a matrix whereas the DFS and BFS developed here are based on object oriented structures of stack and queue data strategies providing more dynamic structures to enhance the efficiency of algorithm. In addition, the advantages and disadvantages of the Switching Point method (Macmillan and Pierce, 1994) are presented describing the way this method handles the contiguity stability. The Switching Point method was not implemented in the A2Z system but we provide a description for comparative purposes. Finally, the Perimeter's Contiguity Stability method is presented here as a more appropriate solution for tackling the computer processing demands of large aggregation problems, compared to previous attempts.

4.7.1 Depth First Search (DFS)

The DFS is a traversal algorithm, consisting of a virtual 'travel' across a given network (graph) aiming to explore all the nodes of the graph. The result of this trip is a spanning tree between the adjacent vertices. The first vertex that DFS algorithm starts is the root point and the depth of the spanning tree depends on the graph size (Figure 4.10).

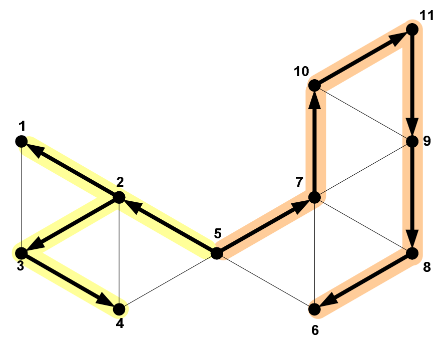
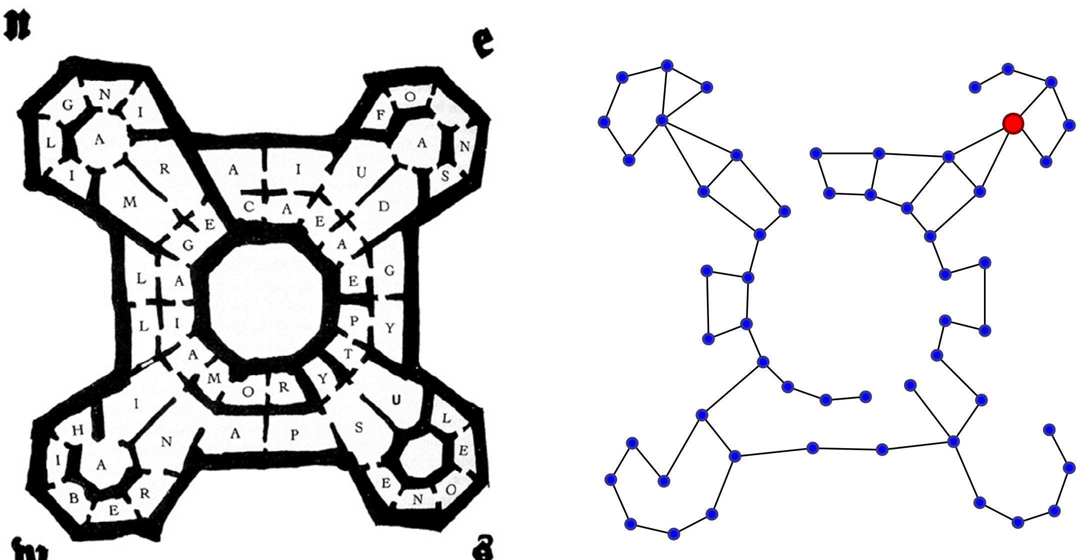


Figure 4.10: The DFS Tree starting from the 5th vertex

An interesting example of a DFS traversal algorithm can be found in 'The name of the rose' by Umberto Eco (1984). When William and Adso are trapped in the labyrinth, William has the idea of marking the paths with signs to find the way out.

"*To find the way out of a labyrinth, William recited, there is only one means. At every new junction, never seen before, the path we have taken will be marked with three signs. If, because of previous signs on some of the paths of the junction, you see that the junction has already been visited, you will make only one mark on the path you have taken. If all the apertures have already been marked, then you must retrace your steps. But if one or two apertures of the junction are still without signs, you will choose any one, making two signs on it. Proceeding through an aperture that bears only one sign, you will make two more, so that now the aperture bears three. All the parts of the labyrinth must have been visited if, arriving at a junction, you never take a passage with three signs, unless none of the other passages is now without signs.*" (Eco, 1984: p 176)



a) b)

Figure 4.11: The library labyrinth according Eco's book and the graph representation a) the plan and b) the graph of library.
Source: (Eco, 1984)

In William's description of how they are going to solve the labyrinth's mystery he explores the graph of the library (Figure 4.11.a and Figure 4.11.b) by putting marks on the junctions (vertices) and the paths (edges). The starting point of their exploration was room A in the east tower of the building as William describes "*The room, as I said, had seven walls, but only four of them had an opening …*" (Eco, 1984: p 169). The explanation of the DFS algorithm and how it is applied in the A2Z context will be introduced using the example of Eco's library graph.

The idea behind depth first search is the following. We walk through the library trying to enter new apertures whenever we can. The first time we visit a room, we put a mark there, and we continue from another aperture. When we arrive at a room that already has a mark, we return through the same aperture from which we came and try another aperture. If all the apertures leading from the room have already been visited, then we return through the aperture from which we first entered. We always try to explore new apertures, we return from the aperture from which we first entered a room only if we tried all other apertures. This approach is called *depth first search* because we first try to visit new apertures going deeper into the library. The resulting spanning tree is presented in Figure 4.12 and consists of the starting point (S) which is the root of the tree and tree junctions (N_1, N_2 and N_3).

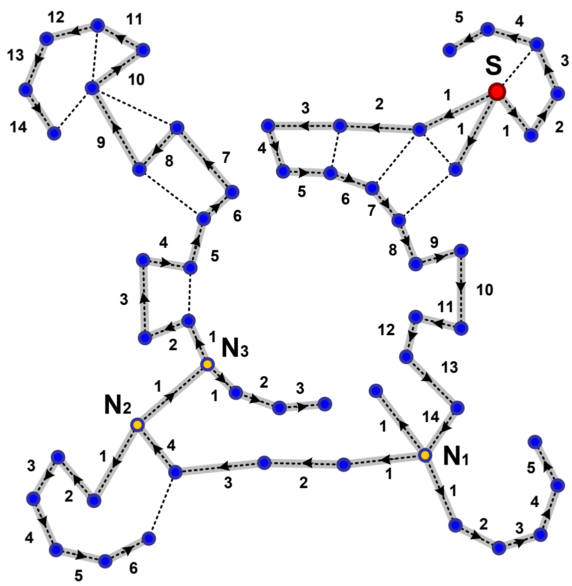
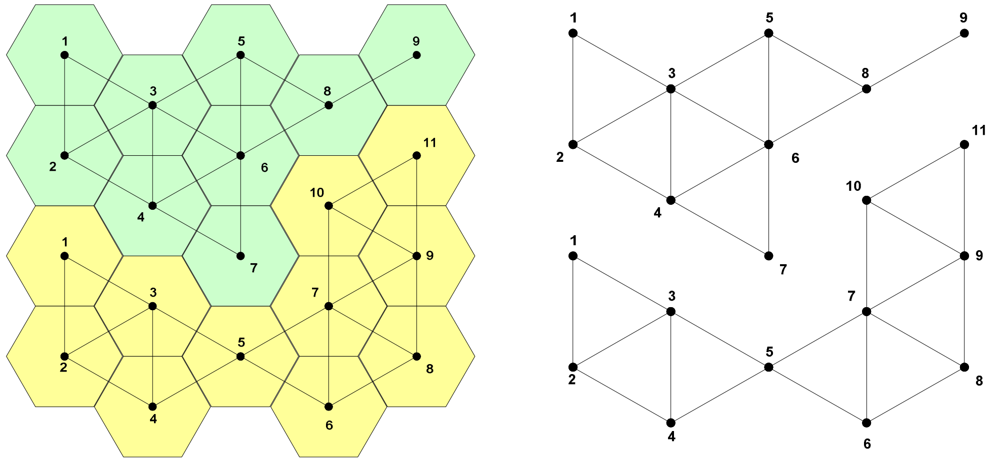


Figure 4.12: The library's graph exploration using the DFS algorithm.

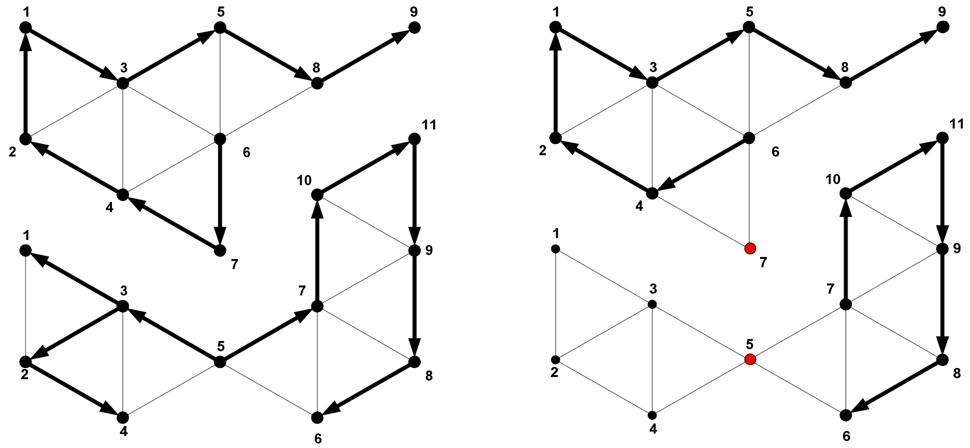
The way the DFS divides the graph and its adaptability to recursive algorithms makes it very useful. The implementation of such an algorithm is based on the stack (First In First Out) method which means that any item (vertex) inserted to the stack will be derived in the same order. Building the graph is not necessary in the zone design because there is topological structure behind the source files which provides the graph indirectly. However, it is important in a zone design system to be able to control the contiguity stability of each zone. For this reason the use of DFS for building a special spanning tree (*DFS tree*) is important. The DFS tree is created after each move of a single areal unit to a new zone. If the number of the vertices differs by one then the zone keeps its contiguity stability otherwise the zone will be broken in two or more zones. The division of zone in two or more subgroups is not acceptable in a zone design system because the contiguity stability of zones is a fundamental principle and should be preserved during the zone design process.

In the following example we describe the way that the DFS algorithm works in the A2Z system (Figure 4.13.a and Figure 4.13.b). The sample areal units are grouped in a green and a yellow zone. In this example our aim is to move an areal unit (7^{th} for the green zone & 5^{th} for the yellow zone) to another zone. Before the selected areal unit changes zone the number of areal units per zone is identified. The number of areal units will provide the information for later comparisons with the number of areal units included in the DFS tree. With the help of the DFS algorithm it is possible to build the DFS trees for the zones and to re-count the areal units per zone. If the before and after totals for each zone differ by one areal unit, then the selected areal unit can move to the other zone without any consequences to the contiguity stability of the zone. In Figure 4.13.c and Figure 4.13.d, areal unit 7 has been selected to move from the green to the yellow zone. According to the DFS method, green zone has 9 areal units in the tree. After the remove of areal unit 7 the DFS tree is created with root point a random areal unit (in this case is areal unit 6). The counting of DFS tree vertices returns 8 areal units which is acceptable because the algorithm removed only one areal unit and the contiguity stability of the green zone was not broken. On the other hand, suppose areal unit 5 has been selected to move from the yellow zone to the green zone. Selecting again a random areal unit (e.g. areal unit 7) in the yellow zone, the DFS method creates the spanning tree. In this case, the number of areal units in the yellow zone before the move of areal unit 5 is 11, while the spanning tree consists of only 6 areal units suggesting the collapse of contiguity stability in the yellow zone.



a) Sample areal units and zones

b) The graphs for the two zones



c) The DFS Trees before the areal unit's move

d) The DFS Trees after the areal unit's move

Figure 4.13: Example of DFS algorithm for 2 zones and 20 areal units.

The DFS algorithm is implemented in the A2Z system based on the above approach and consists of six steps as follows:

Step one: Create an array (*Array A*) and add the areal units and their contiguities belonging to the selected zone. Do not include areal unit (*A*) that is proposed to be removed.

Step two: Create two empty arrays (*History* and *Areas*)

Step three: Get the first item (*i*) of '*Array A*' array and add it in the '*History*' array.

Step four: Get the contiguities of the '*i*' item and add each areal unit in the '*History*' and '*Areas*' collections.

Step five: Repeat until '*Areas*' collection is empty.

Start of Loop

a. Get the first item (*j*) of the '*Areas*' collection.

b. Get the contiguities of the '*j*' item.

c. Check if every contiguous areal unit is stored in the '*History*' collection.

If the answer is negative then add the areal unit in the '*History*' and '*Areas*' arrays.

d. Remove the first item '*j*' from the '*Areas*' array.

End of loop

Step six: If the '*History*' array and the number of areal units of the zone before the remove are equal then the *contiguity* of the zone is not broken.

4.7.2 Breadth First Search (BFS)

The Breadth First Search (BFS) is another traversal algorithm able to trace a network producing a spanning tree (*BFS tree*). The BFS algorithm traverses the graph level by level and seems like a more organized order than the DFS (Figure 4.14). It is implemented similarly to the non-recursive implementation of DFS, but the stack (First In First Out) in DFS is replaced by a queue (First In Last Out) in BFS. This means that an inserted areal unit at the first place in the queue will be accessed last.

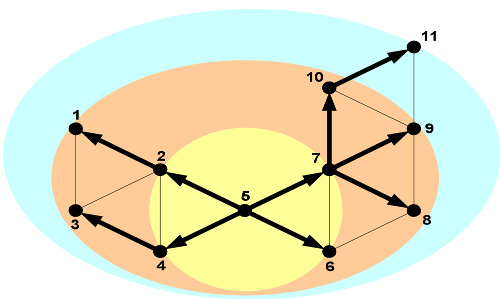


Figure 4.14: BFS tracing a graph level by level

Using the graph from the library example (Figure 4.11) and starting the BFS tracing from vertex S we produce an entirely different spanning tree with many more junctions ($N_1 \ldots N_8$), compared to the DFS algorithm presented in the previous section. The BFS tree is broader and its branches are shorter than a DFS tree. The advantage of the BFS approach is that the time needed to reach the lowest node of the spanning tree from the root the tree is shorter than using a DFS tree. For example, reaching junction N_8 from the root of the spanning tree (S) involves traversing 20 edges with the BFS algorithm, compared to 28 edges with the DFS algorithm (Figure 4.15). In a zone design implementation the increasing number of junctions is a disadvantage for the BFS algorithm because the memory of the system stores more information for the junctions. The increase of stored information signifies high memory consumption and declining the performance of the system. In case the memory capacity is sufficient to support the algorithm then it performs better than the DFS as the spanning tree has shorter branches.

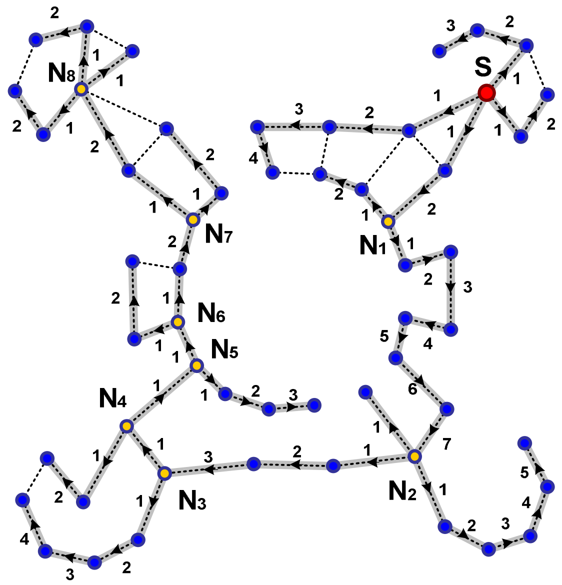


Figure 4.15: The library's graph exploration using the BFS algorithm.

The BFS algorithm is implemented in the A2Z system and consists of six steps as follows:

Step one: Create an array (*Array A*) and add the areal units and their contiguities belonging to the selected zone. Do not include areal unit (*A*) that is proposed to be removed.

Step two: Create two empty arrays (*History* and *Areas*)

Step three: Get the first item (*i*) of '*Array A*' and add it in the '*History*' array.

Step four: Get the contiguities of the '*i*' item and add each areal unit in the '*History*' and '*Areas*' collections.

Step five: Repeat until '*Areas*' collection is empty.

Start of Loop

a. Get the last item (*j*) of the '*Areas*' collection.

b. Store the address of the item (*j*) in a temporary variable (*TempVar*)

c. Get the contiguities of the '*j*' item.

d. Check if every contiguituous areal unit is stored in the '*History*' collection. If the answer is negative then add the areal unit in the '*History*' and '*Areas*' arrays.

e. Remove the item from the '*Areas*' array with address equal to 'TempVar'.

End of loop

Step six: If the '*History*' array and the number of areal units of zone before the remove are equal then the *contiguity* of the zone is not broken.

4.7.3 Switching Point (SP)

A different approach to handle the contiguity stability of zones has been proposed by Macmillan and Pierce (1994). The switching point method relies on the topological characteristics of the areal unit and its relationship with the neighbouring areal units. The basic principle is that any areal unit can be removed from the donor zone when the areal unit has two switching points (vertices) in common with the donor zone. A switching point is a vertex with specific behaviour depending on the zone label left and right of each boundary of the areal unit. A routine traces clockwise (or anticlockwise) all the boundaries of the selected areal unit and when two consecutively arcs change left-right zone label status then the between vertex is marked as a switching point. The focus on the areal unit tracing provides speed advantage (Alvanides and Macmillan, 1997) compared to the adjacency matrix because the contiguity checking is not affected by the size of the problem, as it only examines the selected areal unit and its neighbours.

For a better understanding of the switching point method, an example is presented in Figures 4.16 and 4.17 covering two cases: an areal unit that can be removed and an areal unit that cannot be removed. In Figure 4.16, areal unit 7 is selected to be removed from zone A. The SP algorithm records all the vertices and boundaries of the areal unit, including information of the zones to the left and right of each segment. The starting point is a random vertex on the boundary of areal unit 7, in this case the vertex shared between 7, 8 and 10. Using an anticlockwise trace we select the segment shared between areal units 7 and 10. At this stage it is recorded that the status of this segment is a separation between the two zones (A and B). Stepping on the next segment (between areal units 7 and 6) the status of this segment changes to a border inside zone A. When two sequential segments change their status from border between two zones to border inside the same zone, then the common vertex (included in areal units 7, 6 and 10) is labelled as a switching point. On the other hand, if there is no change of status in two sequential segments then the common vertex remains a simple vertex without any effect. This process ends when all segments of selected areal unit have been traced. If the number of switching points is two then removal of areal unit is permitted (Figure 4.16). However, if more than two switching points have been encountered then contiguity collapses (Figure 4.17).

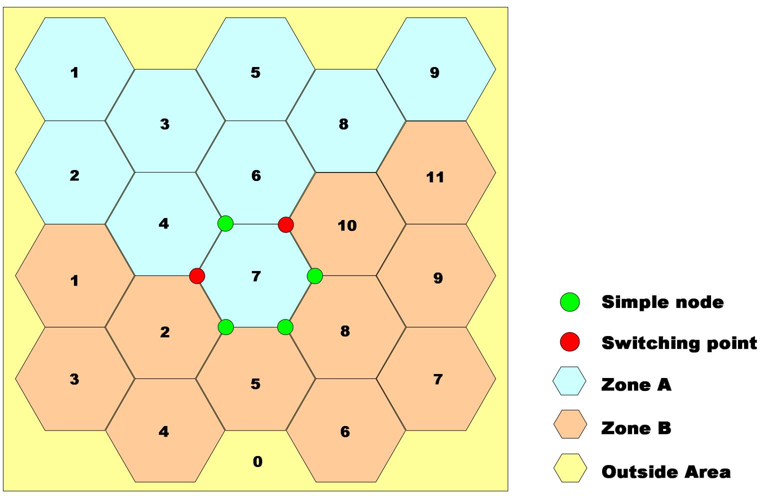


Figure 4.16: The Switching Point method for the selected areal unit 7 of zone A

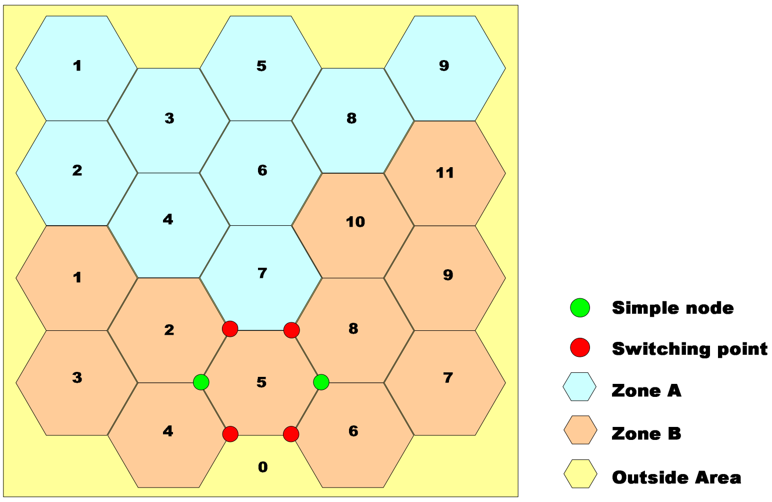


Figure 4.17: The Switching Point method for the selected areal unit 5 of zone B

One complication of this method has been illustrated by Macmillan and Pierce concerning the inability of switching points to cope with island topological structures. In their example three island cases (Figures 4.18.a; 4.18.b and 4.18.c) were introduced arguing that the problem is of minor concern. In fact, the island structures may result from areal unit areas manipulated as regions or connectivity between areal units with physical boundaries and they are a frequent phenomenon, especially when the analysis takes place in a large study area. Another complication arising from the use of switching points is the presence of an 'outside area' (see Figure 4.17) to control for the contiguity of zones located in the outer area of dataset. For example, in Figure 4.17, if the 'outside area' does not exist the SP algorithm counts only two switching points and approves the move of areal unit 5 to the zone A, resulting to collapse the contiguity stability in zone B. Although the SP algorithm performs a faster contiguity check than the AZP algorithm, the spatial limitations of SP algorithm suggest its use only in simple aggregation problems.

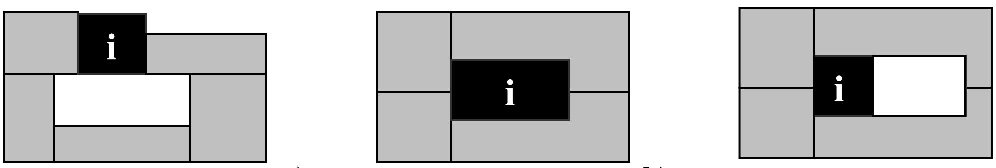


Figure 4.18: The island complication of Switching Point method

Source: Macmillan and Pierce (1994): p 230

4.7.4 Perimeter's Contiguity Stability (PCS)

The AZP algorithm and the DFS and BFS methods can secure the contiguity stability of output zones, but they are more time consuming compared to the switching point method. The disadvantage of these approaches is that the algorithms trace the whole network of a selected zone and when the zones consists of many areal units then this process becomes extremely slow. In contrast, the switching point method operates only if its spatial limitations have been tackled and traces only the features of selected areal unit and its neighbours holding the processing time almost equal for every contiguity check. In this section, the Perimeter's Contiguity Stability method (PCS) is introduced as an innovative method for securing contiguity stability of zones without the limitations on the geographical structure of datasets, while performing the contiguity check in a sort time period. The development of A2Z system is based on an object oriented architecture that supports the PCS method to control the contiguity of zones focusing only in the areal units that comprise the perimeter of each zone (Figure 4.19).

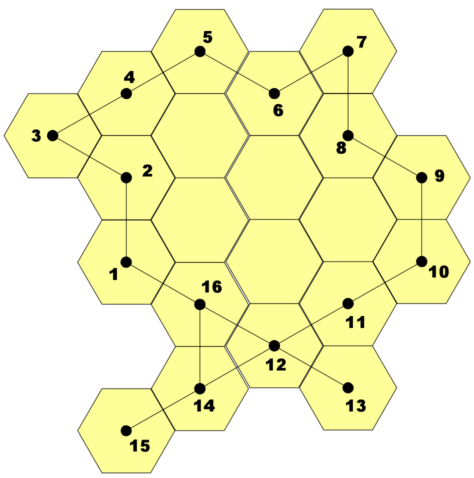


Figure 4.19: The Perimeter's Contiguity Stability method using graph theory

The PCS method is based on the structure and the retrieved information of the object "areal unit". The object stores information about the status of every areal unit concerning its position. When the areal unit is in adjacency with other zones a positive note is stored in the object description otherwise the object remains unchanged. The marked areal units comprise the perimeter of a selected zone and the produced contiguity network becomes much simpler especially for problems where the number of areal units per zone is large. The shrinking of the contiguity network results in faster tracing, using the DFS and BFS methods for controlling the contiguity stability. Moreover, the PCS method is an early step for preparing an effective and small contiguity network for the needs of DFS or BFS algorithm.

Examples of accepted and problematic contiguity are illustrated in Figures 4.20.a and 4.20.b, respectively. In the first example, areal unit 12 is selected for removal. The DFS (or BFS) algorithm randomly selects an areal unit (1) as the root of the DFS (BFS) tree. From areal unit 1 the algorithm traces the whole contiguity network passing from fifteen areal units excluding the removed areal unit 12. The resulting DFS (BFS) tree suggests that there is no effect on the contiguity stability of the zone and, as a result, the selected areal unit 12 is removed successfully. The second example (Figure 4.20.b) demonstrates the case of removing the critical areal unit 14. Again the DFS (or BFS) algorithm traces the contiguity network starting from areal unit 1, but this time it reaches only fourteen areal units. In this case, the algorithm fails to include areal unit 15 in the DFS (or BFS) tree and as a result, does not permit areal unit 14 to be removed.

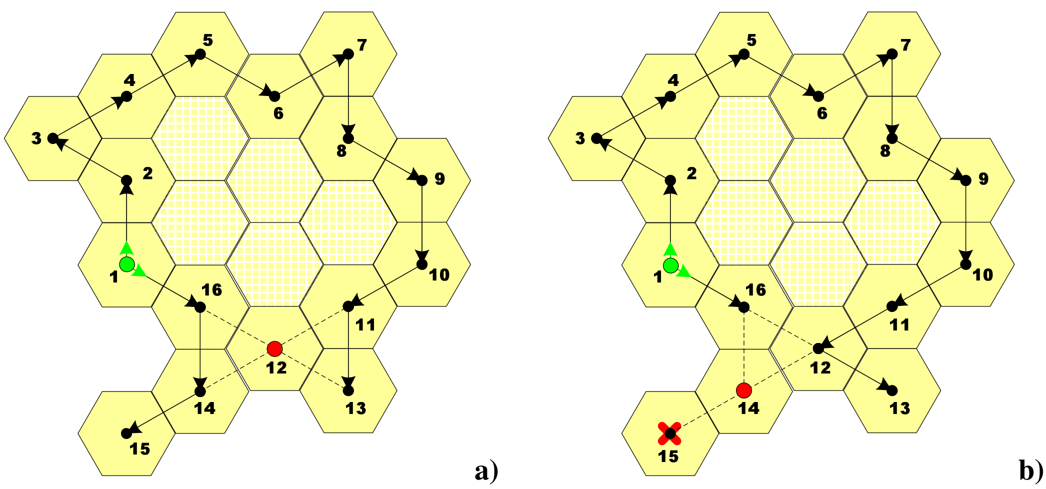


Figure 4.20: An example of the PCS method a) remove of areal unit 12 b) remove areal unit 14

4.8 Optimisation techniques

A fundamental part of zone design algorithm is the implementation of objective functions that can optimise specific criteria (Openshaw, 1978). A number of objective functions have been implemented, reflecting different spatial problems such as equal populated and homogeneous zones. This section presents four objective functions; homogeneity, k-homogeneity, chi-square and deviance, for optimising the homogeneity within zones. In this thesis, we implemented only homogeneity objective functions as in health related studies the resulted homogeneous zones provide better ground for further statistical analysis (Cockings and Martin, 2005).

4.8.1 Homogeneity

The homogeneity function is a simple form of k-means clustering model and it is based on the calculation of attribute distance between two observations. Depending on how the distance is calculated the function model can result functions: Euclidean distance, Manhattan distance and power distance as described in this section (Robinson, 1998). Euclidean distance expressed in equation 4.18 is the distance between two observations *A* and *B* resulting from the sum of squared differences of their *x, y* coordinates. Structuring a non-geometric space, the notion of distance does not represent anymore a measure of geometrical link such as the original Euclidean distance but highlights the differences of attributes (equation 4.19).

$$d_E = \sqrt{(x_A - x_B)^2 + (y_A - y_B)^2} \tag{4.18}$$

$$d_E = \sqrt{\sum_{i=1}^{n} (x_{iA} - x_{iB})^2} \tag{4.19}$$

where, x_{iA} and x_{iB} are the values of observations *A* and *B* respectively for variable *i* in a n-dimensional attribute space (*n* variables in the dataset). The generic formula of Euclidean distance d_E (equation 4.18) can be formulated according to the number of variables.

In the zone design context, the homogeneity function is structured in a single dimension attribute space and the distance indicates the sum of squared differences of two observations using one variable (one dimension). In practice, the distance is calculated between the attribute value of each areal unit (x_a) in zone z and the mean of zone $\bar{x}_z$. Therefore, in zone design the distance for an areal unit a is defined as:

$$d_E^a = \sqrt{(\bar{x}_z - x_a)^2} \tag{4.20}$$

where, $\bar{x}_z = \dfrac{\sum x_{a \in z}}{n_z}$ and n_z is the number of areal units in zone z. The objective function for homogeneity is then calculated as the sum of the distances ($\sum_a d_E^a$), expressed as:

$$OF_{Homogeneity} = \min(\sum_a d_E^a \Big/ Z) \tag{4.21}$$

where, Z is the number of zones in the model. The minimisation of distances between the mean of zones and their areal units produce homogeneous output zones consisting of areal units with similar values for the selected variable.

Other variations of homogeneity are available in the zone design system differing on the method of distance calculation. The Manhattan distance (or block distance) is similar to the Euclidean distance but the squared differences are replaced with absolute differences of observations. The mathematical definition of Manhattan distance (d_M) is expressed as:

$$d_M^{AB} = \sum_{i=1}^{n} |x_{iA} - x_{iB}| \tag{4.22}$$

and in the zone design system it is defined as:

$$d_M^a = \left|\bar{x}_z - x_a\right| \tag{4.23}$$

In general, the zone design system supports both homogeneity functions providing two ways for measuring differences between the zones mean and areal unit attributes. In practice, the system copes faster with the Manhattan distances in relation to Euclidean distance because the process to find the absolute value of a number is calculated much easier than to find the square value of same number. However, the optimisation of an objective function using Manhattan distances provides sufficient improvement compared to the Euclidean distance where the attribute differences shrink between the observations.

Theoretically speaking, the above attribute distances can be formulated using a generic formula, extending equation 4.18 as follows:

$$d_P^{AB} = \sqrt[m]{\sum_{i=1}^{n} \left|x_{iA} - x_{iB}\right|^k}$$

(4.24)

where, m and k are the two powers. If both powers are equal to 2 then the distance is identical to Euclidean distance (equation 4.18). In addition, if both powers are equal then the distance is identical to Manhattan distance (equation 4.22).

In the zone design context, equation 4.24 can be expressed as:

$$d_P^a = \sqrt[m]{\left|\bar{x}_z - x_a\right|^k} \tag{4.25}$$

where, m and k are the two powers inserted by the user. The powered distance is then calculated for the whole model and according to the objective function (equation 4.21) areal units with similar values formulate homogeneous zones. Although, the implementation of a homogeneity function using the distances derived from the generic formula is possible, this research concentrates only in the Euclidean and Manhattan distances.

4.8.2 k-Homogeneity

The *k*-homogeneity function (or similarity function) is based on the homogeneity function as defined earlier. It creates zones with theirs areal units grouped according to equally weighted distances of *k* variables (Johnston, 1976). The *k* variables should be weighted using an appropriate standardisation method such as signed χ^2 statistics or z scores (Simpson, 1996). The standardisation of variables is essential as the use of raw data such as percentages are likely to produce results concentrated on the variables with the highest values (percentages). For example, using two un-standardised indicators A and B with indicator A taking values between 0% - 10% (e.g. unemployment) and indicator B taking values between 30% - 40% (e.g. no car), the indicator B is likely to have three times more influence than A. The standardised indicators compromise equal magnitude during the calculation of the *k*-homogeneity function for all indicators and for this reason it is strongly suggested to standardise indicators before they are used.

The *k*-homogeneity objective function can be expressed mathematically as:

$$d_{iz} = \sqrt{(\overline{x_{1z}} - x_{1iz})^2 + (\overline{x_{2z}} - x_{2iz})^2 + ... + (\overline{x_{kz}} - x_{kiz})^2} \qquad (4.26)$$

where, d_{iz} = multivariate distance of areal unit '*i*' of zone z

$\overline{x_{1z}}, \overline{x_{2z}}, ..., \overline{x_{kz}}$ = mean of each indicator *k* per zone '*z*'

$x_{1iz}, x_{2iz}, ..., x_{kiz}$ = current value of indicator *k* for areal unit '*i*'

The *k* squared differences $(\overline{x_{1z}} - x_{1iz})^2, (\overline{x_{2z}} - x_{2iz})^2, ..., (\overline{x_{kz}} - x_{kiz})^2$ represent different indicators *k* and measure the multivariate distances (d_{iz}) between areal units (*i*) and the mean value of each indicator ($\overline{x_{1z}}, \overline{x_{2z}}, ..., \overline{x_{kz}}$) per zone (*z*). From the sum of distances per zone ($\sum_{i=1}^{n_z} d_{iz}$) equation 4.26 calculates a local objective function score (dz_j) for each zone (equation 4.27) and the results feed into equation 4.28, for the calculation of the global objective function score (*BS*).

$$dz_j = \sum_{i=1}^{n_z} d_{izj} \tag{4.27}$$

where, dzj = local distance of each zone z_j

d_{iz} = distance of areal unit '*i*' from mean of zone z_j

n_z = number of areal units of zone *z*

$$BS = \frac{\sum_{j=1}^{Z} dz_j}{Z} \tag{4.28}$$

where, BS = distance for an iteration

dzj = local best score of each 'j' zone (z)

Z = number of zones

An example of this approach is graphically represented in Figures 4.22.a and 4.22.b. These Figures contain two zones (Z_1 and Z_2) and an areal unit (*i*) that is to be allocated to one of the zones. Each point (1, 2, 3…) represents the values of 'Indicator A' and 'Indicator B', for each areal unit, in a 2-dimentional attribute space.

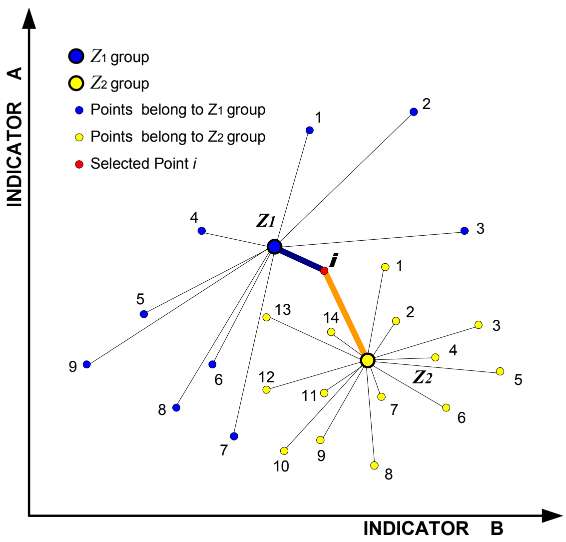


Figure 4.22.a: Graphical representation of two indicators (A and B) - move of selected point 'i' to group Z_1.

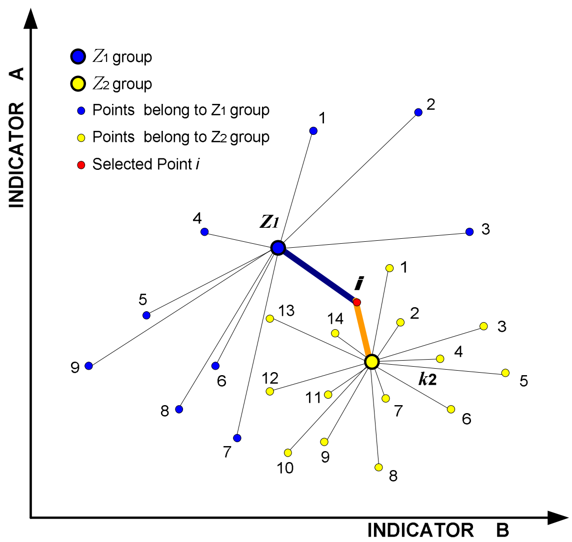


Figure 4.22.b: Graphical representation of two indicators (A and B) - move of selected point 'i' to group Z_2.

Areal unit '*i*' is about to be allocated to one of the two zones Z_1 and Z_2. As Figure 4.22.a shows, areal unit '*i*' is closer to the mean of zone one ($\overline{Z_1}$), so it should receive label '1' because the algorithm is trying to minimize the distances between each areal unit and the zone's mean. In Figure 4.22.b, we have a different result because areal unit '*i*' is further away from the mean of zone Z_1, compared to the mean of zone Z_2. Therefore, the areal unit is allocated to zone Z_2.

Furthermore, the development of Manhattan distance is available in the k-homogeneity function expanding it according to the needs of the research. For example, the use of Manhattan distance stretches the differences between observations, compared to Euclidean distance, while the objective function minimises the differences between zone means and their areal units (Figure 4.23).

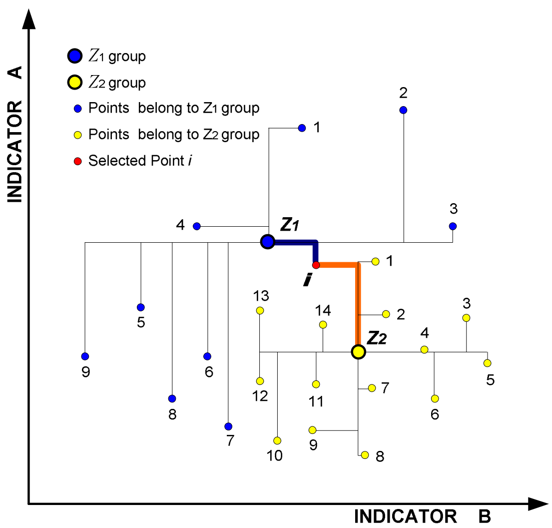


Figure 4.23: Graphical representation of two indicators (A and B) using Manhattan distances

4.8.3 Chi-Square

The chi-square goodness-of-fit test is widely used to evaluate if a sample of data came from a population with a specific distribution. It can be used with discrete distributions such as the binomial and the Poisson (Greenwood and Nikulin, 1996). The chi-square test is applied to grouped data and its value depends on the way data is aggregated. In general, this value measures the variance of data within groups and the smaller value between different groupings represents the best available fit of a model. Usually, the chi-square test is defined for the hypothesis test, where the null hypothesis (H_0) is that the data has the same distribution as the model and the alternative hypothesis (H_a) is that they does not. For the chi-square goodness-of-fit computation, the data are divided into *N* groups and it is defined as:

$$\chi^2 = \sum_{i=1}^{N} \frac{(\mathrm{O}_i - \mathrm{E}_i)^2}{\mathrm{E}_i} \tag{4.29}$$

where *Oi* is the observed incidence for group *i* and *Ei* is the expected incidence for group *i*.

In zone design, the chi-square goodness-of-fit test focuses on population datasets. Therefore, the observed incidence *Oi* represents people with special social and health needs while the expected incidence *Ei* reflects the expected people in need for areal unit *i* using a Poisson Dummy Model (Choynowski, 1959). For the Poisson Dummy Model (PDM), the a_z incident rate for zone *z*, is defined as:

$$a_z = \frac{\sum_{i \in z} O_i}{\sum_{i \in z} B_i} \tag{4.30}$$

where *Bi* is the size of population in each areal unit *i*.

According to Nakaya's (1996) study, the Poisson Dummy Model is defined as a Poisson regression model specified by dummy parameters:

$$O_i = a_z B_i + \varepsilon_i \tag{4.31}$$

where a_z is the dummy parameter denoting the incident rate of zone *z* to which areal unit *i* belongs, and $a_z B_i$ is the expected value of incident in areal unit *i*:

$$E_i = a_z B_i \tag{4.32}$$

The PDM's degrees of freedom *df* for the tested aggregation is the subtraction of the total number of areal units *Z*, from that of zones *Z*:

$$df = N - Z \tag{4.33}$$

In the A2Z system, the chi-square test is implemented as an objective function with the target of minimising the value of the chi-square test ($F(\chi^2) \rightarrow 0$) throughout the zone design process. In other words, the chi-square objective function minimises the variance within the zones providing more homogeneous zones. The chi-square objective function is implemented as follows:

Step one: For a given aggregation, calculate the incident rate a_z of each zone *z* (equation 4.30).

Step two: Estimate the expected number of incidents E_i of each areal unit *i* (equation 4.32).

Step three: Calculate the chi-square value at the aggregated level (equation 4.29).

The chi-square test with the Poisson distribution can be the solution for cases where the incident rates are unstable especially in less populated areas (Choynowski, 1959). Consequently, the transformation of the chi-square test into a zone design objective function provides a useful component for population based aggregations.

4.8.4 Deviance

Deviance or Likelihood ratio tests (LRTs) have been used to compare the goodness-of-fit of two nested models (Nelder and Wedderburn, 1972). The LRT test of two nested models *0* and *z* is defined as:

$$LRT = -2\ln\left(\frac{L_z(\hat{\theta})}{L_0(\hat{\theta})}\right) \tag{4.34}$$

where, $L_z(\hat{\theta})$ and $L_0(\hat{\theta})$ are the two likelihood functions of models *z* and *0,* with model *z* having fewer parameters than model *0*. Following the PDM definition in the previous section it is feasible to associate model *z* with zones and model *0* with areal units.

However, the LRT test can also be presented as a difference in the log-likelihoods and this is often more practical as they can be expressed in terms of deviance. As *log(A/B) = logA – logB*, equation 4.34 can be expressed as:

$$\begin{aligned} LRT &= -2\big(\ln(L_z) - \ln(L_0)\big) \\ &= -2\ln(L_z) + 2\ln(L_0) \\ &= -2(\ell_z - \ell_0) \end{aligned} \tag{4.35}$$

where, ℓ_z and ℓ_0 denote the log-likelihood of model *z* and the full model respectively. The full model completely fits the sample data with zero degrees of freedom and the estimated incidents E_i are equivalent to the observed incidents O_i (Nakaya, 2000). The log-likelihood of the full model is formulated as:

$$\ell_0 = \sum_i (-O_i + O_i \ln O_i) \tag{4.36}$$

while the log-likelihood of the aggregation model *z* with $df = N - Z$ degrees of freedom is defined as:

$$\ell_z = \sum_i (-E_i + O_i \ln E_i) \tag{4.37}$$

In the zone design context, the LRT is possible to provide an alternative of the chi square objective function producing homogeneous zones. The LRT for a true model *z* approximately follows the chi square distribution of *df* degrees of freedom. Therefore, minimisation of the LRT value should result to a decrease of within zone variation. The LRT algorithm consists of the following steps:

Step one: For a given aggregation, calculate the incidents rate a_z of each zone *z* (equation 4.30).

Step two: Estimate the expected number of incidents E_i of each areal unit i (equation 4.32).

Step three: Calculate the LRT value at the aggregated level (equation 4.36) using the measures of log-likelihood for the aggregation level (ℓ_z, equations 4.37) and the areal unit level (ℓ_0, equations 4.36).

4.9 Shape constraint methods

A valuable component of the zone design system is the control of shape compactness for the output zones. Compactness of produced aggregations is desirable because it provides better identification and cohesion of output zones. As Morrill (1987 p.249) suggested "*Geographers, who have a particular concern for territorial measurement, know not to expect too much from a compactness criterion*". Although, the compactness of the output zones is a result of both the shape constraint and the objective function optimisation, in this section two compactness methods are introduced. local spatial dispersion and squared perimeter divided by area methods are discussed here, identifying possible disadvantages for each approach. Moreover, the local spatial dispersion method has been implemented in the A2Z system in addition to two new approaches developed using graph theory measures: contiguity constraint and advanced contiguity constraint.

4.9.1 Local Spatial Dispersion

The minimisation of local spatial dispersion (LSD) by Alvanides and Openshaw (1999) is equivalent to a type of location–allocation problem. This method controls the shape compactness by calculating the distance between the centroids of areal units in each zone and their output zone centroid. Generally, the LSD algorithm is developed using the geometrical features of areal units and zones. For example, for a given aggregation (Figure 4.24) the LSD measure is calculated by measuring the Euclidian distances between the centroid of each areal unit i and the centroid of its zone z. Mathematically, it is expressed as follows:

$$LSD = \sum_{i \in z} \sqrt{(\bar{x}_z - x_i)^2 - (\bar{y}_z - y_i)^2} \Big/ n_z \tag{4.38}$$

where $\bar{x}_z$ and $\bar{y}_z$ are the coordinates of the centroid of zone z, x_i and y_i are the coordinates of the centroid of areal unit i and n_z is the number of areal units in zone z.

During the zone design procedure the areal units constantly change zone membership while attempting to achieve an optimum solution. Therefore every time such a change occurs, it is necessary to recalculate the zone centroid. Traditionally, the algorithm would have to dissolve the areal units and then calculate the centroid of each zone in GIS related context. It is obvious that the dissolve task would increase the processing time of the zone design procedure.

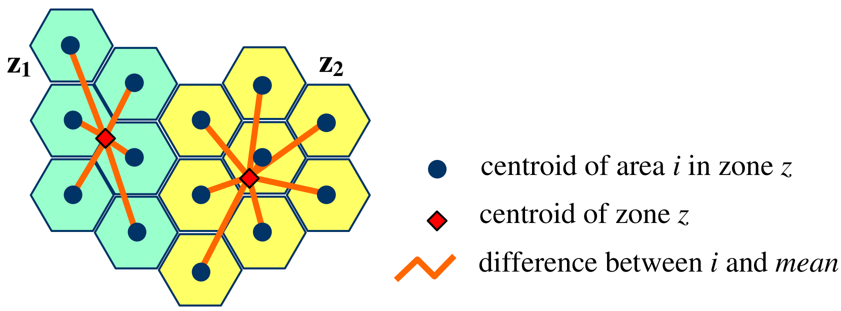


Figure 4.24: An example of LSD using two output zones z_1 and z_2.

Consequently, in the A2Z system the LSD approach is implemented using only the centroid coordinates of each areal unit. Instead of dissolving the areal units, the developed LSD approach derives the coordinates of each zone by calculating the mean of the coordinates of the areal unit centroids in zone z:

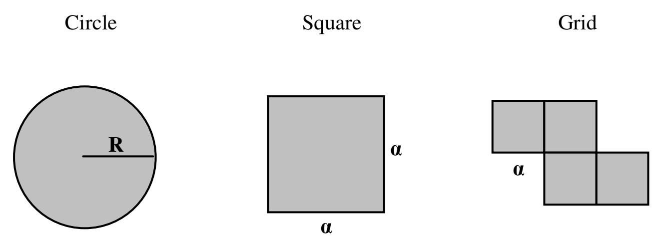
$$\bar{x}_z = \frac{\sum_{i \in z} x_i}{n_z}, \text{ and } \qquad \bar{y}_z = \frac{\sum_{i \in z} y_i}{n_z} \tag{4.39}$$

The zone centroid coordinates are then used in equation 4.38 providing the final LSD measure for the selected zone. The minimisation of all LSD measures during the zone design procedure results in spatially compact output zones.

4.9.2 Squared Perimeter divided by Area

A geometrical alternative to the local spatial dispersion constraint was proposed by Prof. D. Martin to the Office for National Statistics (2001) for the design of the 2001 Census Output Areas. The minimisation of the squared perimeter divided by the area of a zone effectively compares the ratio of each shape with the optimum ratio of a circle. As shown in Figure 4.25, a circle (ratio 12.56) will always be more compact than any other shape. Although, in terms of calculation it is very simple, during the implementation of this approach it is necessary to dissolve the areal units within every zone. As mentioned earlier such a task is costly in processing time. In addition, at higher aggregation levels with small number of zones the perimeter has large values as there are many small areal units, resulting in detailed geometry and large perimeter values. A possible solution to tackle this problem is to apply a generalisation method at the higher aggregation levels but is likely to result in additional transformations of the dataset. Of course such a method is not recommended in a zone design system because the generalisation process is very time consuming and adding it to the aggregation process will increase dramatically the processing time. As a result, in the A2Z system the squared perimeter divided by area method is not applicable. Although the approach is not used in the zone design, we should note here that it is an excellent measure for comparison between existing aggregation solutions.

Shape



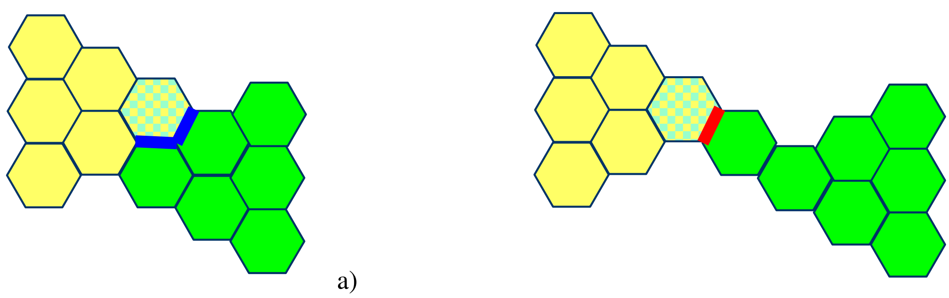
		R = 1		α = 1		α= 1
Perimeter	**2πR**	**6.28**	**4α**	**4**	**10α**	**10**
Area	πR^2	**3.14**	α^2	**1**	$4\alpha^2$	**4**
Ratio	$\frac{(2\pi R)^2}{\pi R^2}$	**12.56**	$\frac{(4a)^2}{a^2}$	**16**	$\frac{(10a)^2}{4a^2}$	**25**

Figure 4.25: Ratios for three different shapes

4.9.3 Contiguity Constraint (CC)

The following approach takes into account the contiguity information of the areal units and builds a rule that applies to each area the system examines. The Contiguity Constraint (CC) method can reduce the time process of shape control calculation because it focuses on the selected areal unit and the two contesting zones (local shape control). Moreover, it selects the areal unit that is considered for a move and recodes all the neighbour areal units. If the number of the neighbours is above the number the user has set as a rule then it is acceptable to move the areal unit to the target zone, otherwise, the areal unit remains in the donor zone.

For better understanding of the CC algorithm the following example is described with a rule that accepts changes of an areal unit only if it has more than one neighbour per zone. In Figure 4.26.a there are two zones: the yellow hexagons depict the donor zone and the green hexagons comprise the target zone. The checked areal unit (hexagon) between the two zones belongs to the donor zone (yellow) and we examine if it can be transferred to the target zone (green) by satisfying the rule set for the CC algorithm. In Figure 4.26.a the algorithm records two neighbour for each zone and the areal unit is moved to the target zone. On the other hand, in Figure 4.26.b a different configuration of zones cannot satisfy the rule of two neighbours and therefore the areal unit remains in the donor zone.



a) b)

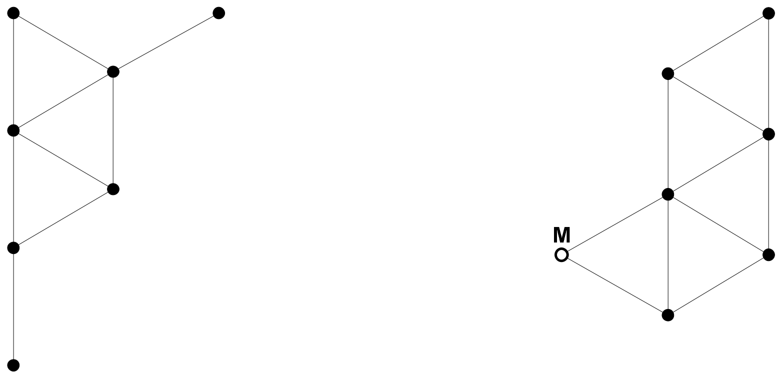
Figure 4.26: The CC method using a simple example: a) the areal unit does not change zone (2 neighbours) and b) the areal unit changes zone (4 neighbours)

The measure of compactness by the CC method is introduced here as an alternative approach of shape constraint. The advantage of this algorithm is the speed of compactness calculation especially for large dataset. As mentioned above the CC algorithm does not use geometrical information, only the contiguity information of areal units. The CC method provides a mild compactness constraint helping the system to minimise the effect of gerrymandering shapes, as it is only applied in neighbouring areal units.

4.9.4 Advanced Contiguity Constraint (ACC)

In this section, the Advanced Contiguity Constraint (ACC) is introduced as a new method for compactness measurement, combining the simplicity of the CC method with basic graph analysis techniques. The connectivity measures as suggested in the first section of this chapter are widely used in network efficiency for transport systems (Briggs, 1972; Garrison, 1968; Wegener, 1994). The idea introduced here is to use the connectivity measures of graphs as indicators for the compactness of zones. The ACC method can be constructed by calculating the *beta* index (equation 4.6) and the index of *network efficiency* (equation 4.7) for an advanced control of compactness during the aggregation. The ACC method analyses the graph of each zone by comparing the indices before and after the move of an areal unit. If the zone improves the indices then the move of the areal unit is accepted by the system otherwise the aggregation process continues to the next pair of zones.

For a better explanation of the ACC method, the graphs G_A and G_B of Figure 4.27 represent two neighbouring zones (*A* and *B*) before the move of areal unit (*M*) from zone *B* to zone *A,* while Figure 4.28 illustrates the graphs of zones *A* and *B* after the reallocation of areal unit *M*. The graph measures of G_A (7,9) and G_B (7,11) in table 4.4 provide a first sight for the structure of each graph. Figure 4.27 shows a more open structure for G_A compared to G_B and the lack of compactness is confirmed by the indices of compactness and network efficiency (Table 4.4). In the example, areal unit *M* (vertex *M*) is selected to be moved from zone *B* to zone *A*. The selected vertex, according to visual examination, provides a more compact graph structure for both zones if allocated to zone *A* (Figure 4.28). Subsequently, the Beta index for zone A is improved from 1.29 to 1.5, while retaining the same percentage of network efficiency.



$G_A(7,9)$
(ZONE A)

$G_B(7,11)$
(ZONE B)

Figure 4.27: The graph representation of the two zones before the move of areal unit M

$G_A(8,12)$
(ZONE A)

$G_B(6,9)$
(ZONE B)

Figure 4.28: The graph representation of the two zones after the move of areal unit M

Furthermore, the use of both indices (Beta and Network efficiency) introduces cases where one index remains the same while the other index is improved after adjusting the zones. Such cases in the A2Z system are standardised for advantageous compactness. The possible combinations between the Beta index and the network efficiency measure are:

- Improvement of beta index and network efficiency measure: the system accepts the change of a selected areal unit because both criteria are fulfilled.
- Worsening of beta index and improvement of network efficiency measure, where the system accepts the move of an areal unit, because the efficiency of the network is more informative than the beta index.
- Improvement of beta index and no change to the network efficiency measure: the system accepts the change of selected areal unit even if the network efficiency percentage remains unchanged.

The above rules combine both measures helping the A2Z system to decide if an areal unit is possible to be moved or not aiming for more compact zones.

Table 4.4: The graph measures of GA and GB before and after the adjustment of vertex *M*.

Measure	Before areal unit (M) move		After areal unit (M) move	
	$G_A(7,9)$ (ZONE A)	$G_B(7,11)$ (ZONE B)	$G_A(8,12)$ (ZONE A)	$G_B(6,9)$ (ZONE B)
Minimum edges	6	6	7	5
Complete connectivity	21	21	28	15
Cyclomatic Number	3	5	5	4
Beta Index	1.29	1.57	1.5	1.5
Efficiency of Network	43%	52%	43%	60%

The ACC method results in compact zones using less computational power than other methods such as local spatial dispersion and squared perimeter divided by area because it does not involve the geometrical elements of a zone. Moreover, the structure of the A2Z system can incorporate further compactness methods based on more advanced graph analysis. For example additional flexibility such as weighted compactness for rural and urban areas or even availability to perform shape constraint till the system achieves a predefined compactness level.

4.10 Discussion and Conclusions

In this chapter, the principles behind the zone design system were introduced in an attempt to bring together all its different components in relation to the spatial issues discussed in Chapters 2 and 3. The presented methodologies follow the principle of the original AZP formulated by Openshaw (1977a) while the innovative characteristics of other aggregation applications (Chapter 3) were taken into account throughout the implementation of the proposed A2Z system. The aggregation issues highlighted in Chapter 3 were formulated using graph theory and the final implementation of the new zone design system was developed using object oriented structures. Furthermore, important problems relating to the spatial characteristics of areal units in geographical space have been tackled by constructing advanced contiguity adjacencies while innovative use of boundaries was suggested, utilising weighted adjacencies for the representation of natural or conceptual barriers. An additional novelty of this thesis is the selection and implementation of the Akaike Information Criterion for investigating the most informative aggregation level as the level for further exploration of the MAUP effects. In this chapter, we presented a methodology for implementing a goodness of fit estimator in a zone design context.

Finally, we explained the components of zone design system relating to the initial aggregation, control of contiguity stability, objective functions and shape compactness. The newly developed components of zone design as well as the proposed methods for tackling spatial issues are put to work in a variety of aggregation problems from the health research domain. In the following three chapters, we evaluate the three health related case studies addressing and tackling key challenges relevant to the MAUP effects.

CHAPTER 5

Building Lower Level Health Regions

5.1 Introduction

In the previous chapter, the characteristics of the new zone design system (A2Z) were detailed introducing a variety of methods related to spatial issues deriving from the literature, as discussed in Chapters 2 and 3. This chapter will evaluate the new system at small scales investigating the impact of aggregation in health related data. The lower level refers to the scale size of the areal units and in this study the focus is on what is widely regarded as neighbourhood level. In addition, the newly proposed methodology for identifying the most informative aggregation level is tested using traffic accidents of children in Tyne and Wear. Thus, we will provide an automated methodology for selecting the aggregation level in which analysis is going to take place.

Accidental injury in young children is a crucial cause of mortality, morbidity and disability in developed countries. In the UK a considerable proportion of health care resources are allocated to reduce children accidents (Towner et al., 1993). One of the most persisting unintentional injuries is the traffic accidents affecting 1-19 years old children. In the early 1990s the number of child pedestrian fatalities in Britain relative to child population, was considerably larger than the EU average (DfT, 2003). The UK government has set a clear target of 40% reduction of accidents by the end of 2010. In this study, we will investigate the relationship between the traffic accidents of children and the deprivation in the Super Output Areas (SOAs) of Tyne and Wear.

Furthermore, the A2Z system will construct two new aggregation sets at the AIC proposed level targeting homogeneous zones in terms of the children traffic accident rates and the Townsend deprivation index respectively. During the zone design process, the A2Z will control the shape compactness of zones in a very weak manner because the aim of this study is to maximise the differences between zones. This way, it will be possible to clearly identify the zones with high rates or scores.

5.2 Children's' traffic accidents

The majority of studies on childhood injury have concluded that there is higher risk among children from deprived areas. For example, Abdalla et al. (1997a; 1997b) and Abdalla (1997) noted higher rates of casualties in deprived than affluent areas of Lothian region in Scotland. Their research conducted at census output areas level calculating the distances between the output area of residents and the accident locations. Although, their findings appeared to be related with the class of road and social status of output areas, they did not explore the aggregation effects of their study. On the other hand Sharples et al. (1990) came into the similar conclusion using individual data records in Northern Regional Health Authority and the social status of children involved in fatal accidents at ward level.

Additionally, attempts to explain social differences in traffic accidents have suggested a strong correlation between the social class and injury mortality. Christie (1995) stated that the risk of children traffic accidents is strongly class related. She indicated that children in the lowest social classes are over four times more likely to be killed than children in highest social classes. Her results were based on mortality rates and socioeconomic characteristics in the UK at individual or household level. Similar studies suggested that the social gradients in injury mortality exceed those for any other cause of death in young people, and the inequalities between different socioeconomic groups are higher in relation to child pedestrian deaths than all causes of death in children (Jarvis, 1999; Roberts and Power, 1996).

In the UK the majority of studies are conducted using accident and socioeconomic data at individual and household level. Recent findings have shown high injury rates of children in households with single parents (Judge and Benzeval, 1993; Roberts and Pless, 1995) while a strong association to traffic accidents have been seen in poor housing conditions including type, quality, and tenure of housing as well as overcrowding (Alwash and McCarthy, 1988; Runyan et al., 1992). The general conclusion from these studies suggested that deprivation determinants have strong relationships to traffic accidents among children. Generally speaking this may be true but it could be an example of ecological fallacy as the high rates of accidents in deprived areas could be result of risk factors on top of household deprivation (Reading et al., 1999).

On every occasion, traffic accidents investigated at area-based dataset, the studies were based on areas created to serve administrative purposes such as local authorities and wards. These arbitrary administrative areas are very heterogeneous in terms of socioeconomic characteristics. As Cockings and Martin (2005) mentioned, when the research aim is to test a hypothesis or to explore spatial patterns of disease then the maximisation of internal homogeneity is required. To address these problems, a few epidemiological studies have constructed manually social homogeneous geographical areas. For example, Propper et al. (2004) suggested the construction of equally populated neighbourhoods of around 500 people aggregating the census 1991 enumeration districts. Moreover, Reading et al (1999) produced 'social areas' by adjacent enumeration districts with similar Townsend Index values and proportions of the population aged 0-4 years. Both examples were based on visual inspection and the researchers' knowledge of the area. Although the use of manual aggregation is possible in small study areas, the construction of social zones becomes unfeasible as the number of areal units increases.

Furthermore, the selection of number of zones is usually based on the knowledge of researcher of the studied area and is prone to human fault. In the current study, the basic geographies are automatically aggregated by means of the A2Z system constructing geographies with similar Townsend Index scores. The AIC method is used to define the most informative aggregation level combining the variance of traffic accidents and the degrees of freedom of each aggregation as suggested in Chapter 4. Both the A2Z system and the AIC method provide a variety of automated aggregations using advance statistic measures to explore and select the appropriate aggregation level. In addition, we examine the significance of children traffic accidents at two aggregation levels: super output areas (SOAs), output zones based on Townsend Index scores. While the comparison of social indicators and the exploration of MAUP at each aggregation level provide valuable information concerning the characteristics of dangerous areas.

5.3 Datasets and statistical methods

The study area covers the county of Tyne and Wear in England as defined at the time of the 2001 Census. The county of Tyne and Wear has total population of roughly one million people with 237,645 children aged 0-17 years old and it is consisted of five local districts: Newcastle upon Tyne, Gateshead, North and South Tyneside and Sunderland (Figure 5.1). In the table 5.1, the population in each local district according the 2001 Census tables are listed providing the number of children aged 0-17 years old.

Table 5.1: The total population and children aged 0-17 years old in Tyne and Wear.

Authority	Children (aged 0-17 yrs)	Total Population
Newcastle upon Tyne	56,540	259,536
Gateshead	41,599	191,151
North Tyneside	40,998	191,659
South Tyneside	34,672	152,785
Sunderland	63,836	280,807
Tyne and Wear	237,645	1,075,938

Source: (UK National Statistics, Census 2001)

The event data were supplied by the Tyne and Wear Police/Local authority providing information on the exact location where the incident took place and limited information on people's social background or home postcode. The dataset involves children 0-17 years old experiencing a traffic accident in the period 1996-2001 in Tyne and Wear County. In this study, we investigate a five year period instead of single year tackling this way the problem of small numbers.

Altogether 6,111 children were involved in traffic accident during the period between 1996 and 2001, of which 0.6% were fatal incidents, 13.8% were serious accidents with possibly hospital attendance and 85.6% were slight incidents. In the analysis, the event data was the total children incidents without any severity filtering, as the level of analysis was very low and the risk of small number area effect was present. Although, Walsh and Jarvis (1992) have suggested that epidemiological studies of accidents should not use slight severe injuries but those recorded by hospitals, the later datasets are protected by confidentiality regulations with authorisation from the ethic committee of each organisation. As far as this study is concerned, 85.6% of children accidents are slight severe injuries and its possible exclusion limits the dataset to roughly 880 cases in whole Tyne and Wear making the data unsuitable for use. Thus, we included the slight severe injuries into our analysis knowing that the finding will reflect more the less serious than the fatal and severe accidents.

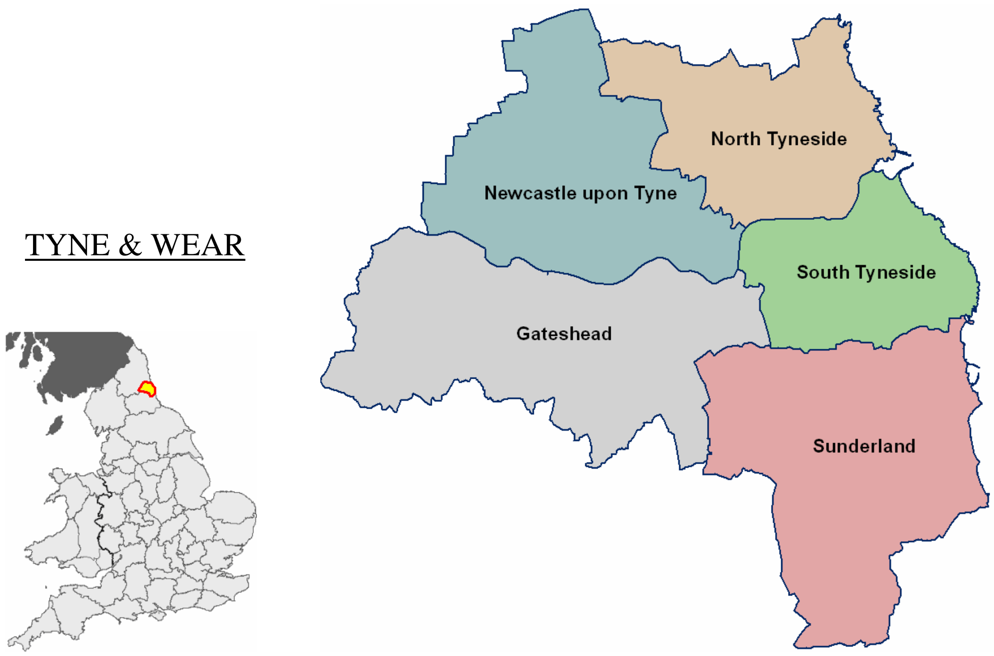


Figure 5.1: Map of local districts in Tyne and Wear County.

We have used five social determinants to investigate the socioeconomic profile in relation to the traffic accident on children. The social determinants are provided by the 2001 Census datasets (Table 5.2). The first four indicators: overcrowding, unemployment, owner occupancy and car availability are used to construct the Townsend material deprivation index (Townsend et al., 1988). The Townsend index ranges from negative values, indicating affluent areas to positive values for deprived areas. The strong relation of the Townsend index to the health status of population has been discussed in Morris and Carstairs (1991) as a general measure of deprivation. We selected the 2001 Census determinants because the analysis should be as much coherent as possible with the period of accident events. Therefore, the 1991 census determinants can not used as they are too dated.

Table 5.2: The indicators used in the current case study (Census 2001)

Table	Indicators	Numerator Variables	Denominator Variable
KS019	Overcrowding	4	1
KS009	Unemployed	5	2+3+4+5+6
KS018	Owner Occupancy*	2+3	1
KS017	Car availability	2	1
KS020	Lone Parents	11+12	1

* Calculated as follows: 100 – ((KS0180002 + KS0180003/ KS0180001)*100)

Our study area consisted of the 2001 census super output areas (SOAs) whose socioeconomic variables are calculated by summarising the output areas (OAs) information within each SOA. SOAs were chosen as the basic geographical unit because they are one level above output areas (OAs) in the UK for which census data are available. The selection of SOAs instead of OAs was necessary because on each OA there were few accident cases resulting in small number effects. There are 719 SOAs in Tyne and Wear and their population varies between 1,000 and 2,000 people. Descriptive statistics in the SOAs level of Tyne and Wear show an average of 40.8% households not owned by the residents and 41.0% households not owning a car with standard deviations of 24.2 and 17.2 respectively. In addition, an average of 6.0% households is overcrowded reaching a maximum of 33.6% in areas close to Newcastle upon Tyne city centre. Similarly, unemployment in Tyne and Wear scores an average of 8.4% with maximum values of 33.9% next to Tyne River in the Newcastle city centre and in South Tyneside.

Although the population of SOAs is between 1,000 and 2,000 people providing almost equally populated areas, there is an average of 329 children aged 0-17 years old per SOA reaching 726 children in some areas. It is obvious that the children population is not equally distributed within Tyne and Wear County as it varies between 10 and 726 children per SOA. Investigating the cases of children traffic accidents, the distribution of data is ranged from 0 to 49 accidents with additional two outlier SOAs of 72 and 102 accidents. Figure 5.2 shows the frequencies of children accidents without the two outliers for better graph representation. The distribution of accidents is squished to the left side with a mean value 9 while the standard deviation value is 7.5. Investigating the distribution of accidents, we can identify four groups of SOAs. The two groups ranged from 0 to 9 and from 10 to 17 accidents reflect the majority of SOAs while the last two groups referred to SOAs with many accidents including the outliers. The Figure 5.3 illustrates the above classification providing information about the SOAs with children traffic accidents in terms of actual numbers. It can be seen that Newcastle's and Gateshead's SOAs provide high numbers of accidents. However, the actual number of accidents can be misleading as two same large numbers of accidents may be referred to different population bases. For example, if there are 30 cases in an urban area with 300 people it will be less serious if those 30 cases recorded in a rural area with 90 people.

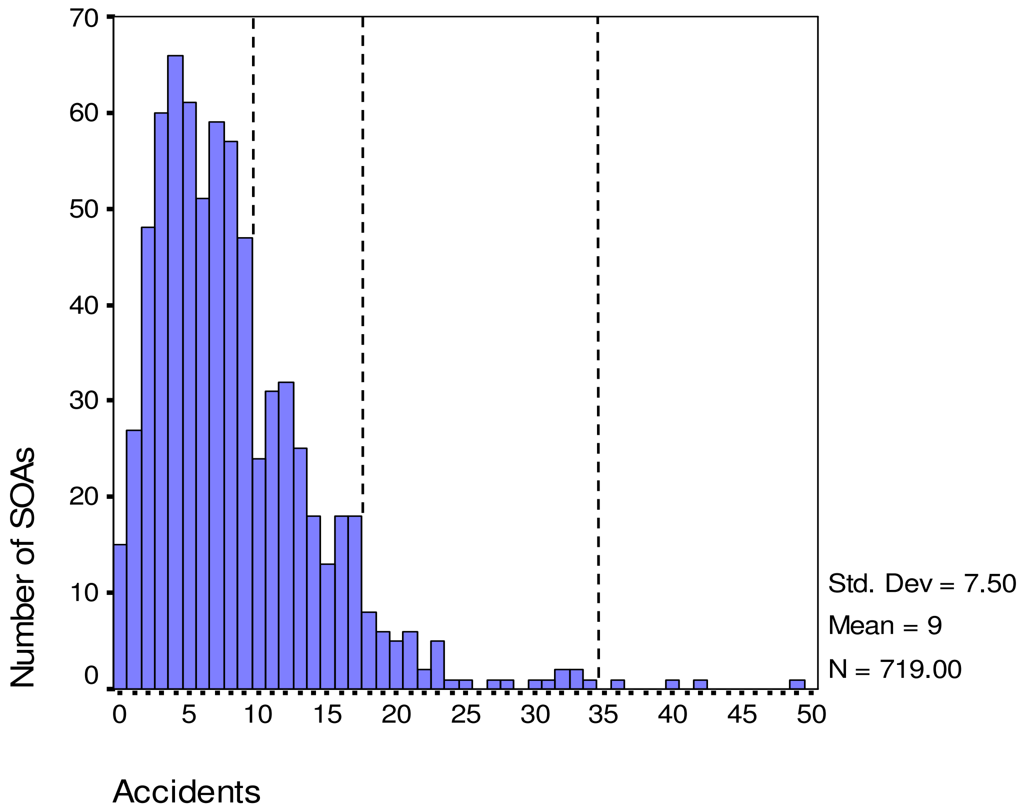


Figure 5.2: Frequencies of children accidents at SOAs level.

Therefore, we transform the accident cases to accidents rates using the following definition. The children traffic accident rate of an area z is defined as:

$$CTAR_z = \sum_{i \in z} y_i \Big/ \sum_{i \in z} B_i \tag{5.1}$$

where, y_i is the number of children traffic accidents involving children and B_i is the number of children aged 0-17 years old population for each SOA i.

In Figure 5.4, most of SOAs are low rated while the SOAs at the border of Newcastle and Gateshead score higher with the City of Newcastle to experience the highest rates of 1.06 and 1.14. A careful examination of the study area shows that the specific two areas suffers from many children traffic accidents but the children may not be resident of this problematic area as the data are reporting incident places without any information of home address of child. Therefore, the accident rate indicates how dangerous an area is rather how many children resident in this area involved in traffic accidents.

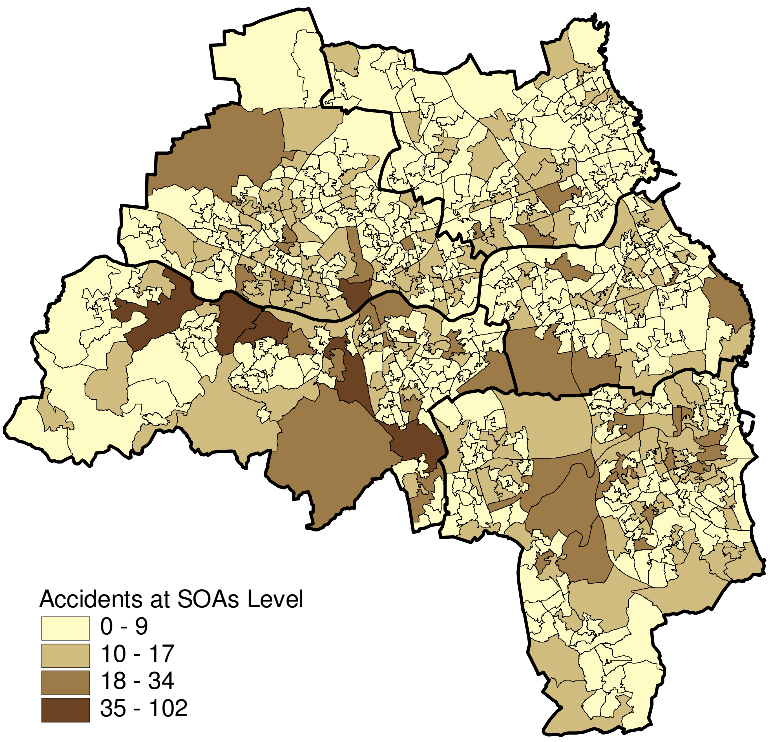


Figure 5.3: The cases of children accidents at SOAs level.

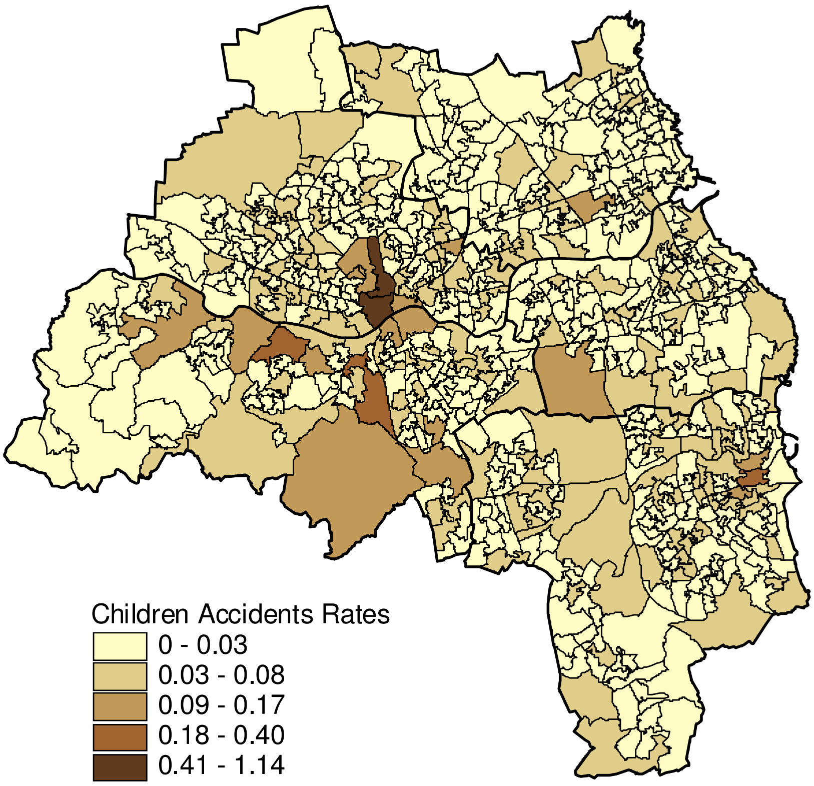


Figure 5.4: Children traffic accident rates at SOAs level.

The relationship of selected deprivation indicators with the rates of traffic accidents at SOA and ward level are explored presenting the most significant determinants for accidents in Tyne and Wear. Table 5.3 shows the 'overcrowding households' indicator as the most strongly correlated determinant to children traffic accidents with r = 0.286 and 0.571 at the SOA and ward levels respectively. The unemployment indicator at the SOA level is not significant while the same indicator at ward level becomes significant with r= 0.370. This is an example of the well-known scale effect whereas the correlation coefficient tends to increase as the level of geographical aggregation increases (Gehlke and Biehl, 1934). In addition, a negative correlation appears in the lone parent indicator at SOA level changing to a non-significant determinant at the ward level. Although the literature suggests that lone parent households are a strong determinant of children accidents, in the case of Tyne and Wear it seems that lone parents' households are not correlated with children traffic accidents, at least at the areal level of SOAs and wards.

Table 5.3: Correlations with children traffic accidents aged 0-17 years old at SOA and Ward level.

	Children Traffic Accidents aged 0-17 yrs old					
	SOA Level			*Ward Level*		
	r	*Sig.*	*N*	*r*	*Sig.*	*N*
No Car	0.155**	0.000	719	0.443**	0.000	113
Overcrowding	0.286**	0.000	719	0.571**	0.000	113
Not owner occupied	0.150**	0.000	719	0.443**	0.000	113
Unemployment	0.066	0.078	719	0.370**	0.000	113
Lone Parent	-0.101**	0.007	719	0.034	0.723	113

** Correlation is significant at the 0.01 level (2-tailed)
* Correlation is significant at the 0.05 level (2-tailed)

5.4 Zone Design characteristics

In this research, the zone design plays a central role as it is used for identifying the most informative aggregation level by means of AIC as well as for constructing homogeneous zones in terms of accident rates and Townsend index. The utilisation of zone design system to investigate the statistical variation of accidents at all possible aggregation levels can be a time consuming process. Therefore, here the method of segments is used providing a quick identification of the suggested aggregation level as described in Chapter 4. In detail, the A2Z system aggregated the 791 SOAs until the zoning solution scores the minimum AIC value (Figure 5.5). Also, an important asset of zone design, the objective function, set to optimise the zoning systems using the deviance function as suggested in Chapter 4. For each aggregation process, the zone design executes the optimisation for 10 runs of 300 iterations each or until the objective function does not improved any more.

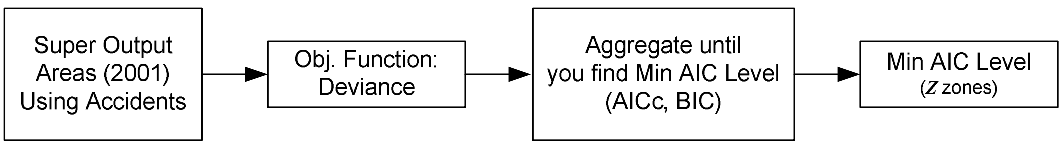


Figure 5.5: The diagram for identifying the most informative aggregation level.

Then, the number of zones (z) which are statistically appropriate for further investigation according the AIC is used to provide the scale in which an extensive aggregation will take place. As the Figure 5.6 shows, the areal units of this study

(SOAs) grouped twice using the suggested scale level. Both aggregations are based on the k-Homogeneity objective function as introduced in Chapter 4. In the first grouping, the zone design process targets the maximisation of homogeneity within zones in terms of accident rates, while the second aggregation aims to build homogeneous zones using the Townsend Index. The grouping process for both aggregations is repeated for 100 runs of 300 iterations each investigating an extended range of grouping solutions at the same scale.

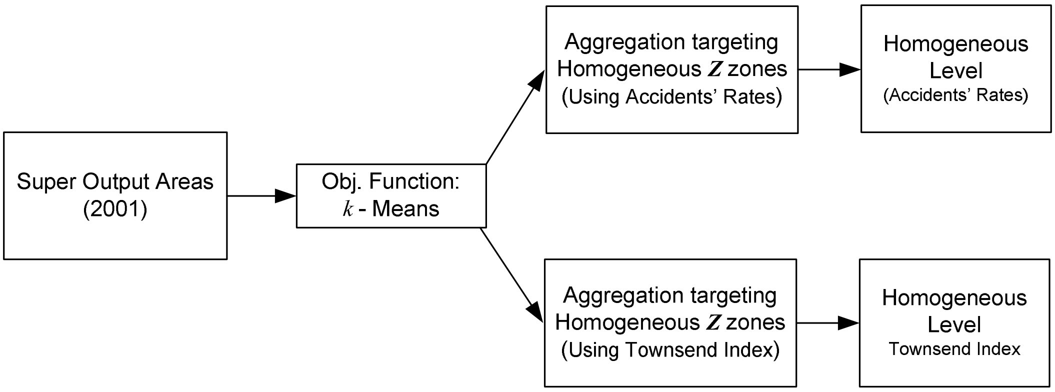


Figure 5.6: The diagram of aggregation for providing homogeneous zones at predefined scale level.

5.5 Selecting the most informative aggregation level

As this thesis strongly supports the idea of constructing geographies with criteria focused on health related characteristics rather using administrative areas constructed to tackle other social issues, it is important to investigate the appropriate aggregation level of this case study, using advanced statistical measures of goodness of fit at each scale. The methodology of selecting the most statistically informative aggregation level by means of AIC criterion as discussed in Chapter 4, is extensively applied in this case study. Every aggregation level is constructed by the A2Z system using the SOAs as basic units and optimising the homogeneity of children traffic accident rates.

The objective function for this case study is the *deviance* or *likelihood ratio* statistic measure. As the deviance measures the variance within the zones, a minimisation of deviance is equivalent to homogeneous constructed zones. The deviance of an aggregation *A* is calculated using the equation 5.2.

$$D_A = -2(l_A - l_0) \tag{5.2}$$

where, l_A is the log-likelihood of aggregation *A* and l_0 is the log-likelihood of the aggregation with zero degrees of freedom. For this case study the aggregation level 0 is equivalent to the SOA level. The calculation of log-likelihood at aggregation A is defined as:

$$l_A = \sum_i (-E_i + O_i \ln E_i) \tag{5.3}$$

where, E_i is the estimated number of children traffic accidents in *i* SOA and it is calculated as follows:

$$E_i = CTAR_z \cdot B_i \tag{5.4}$$

To calculate the log-likelihood of SOA level (l_0), equation 5.3 is converted by replacing the estimated children traffic accidents (E_i) with the observed values (O_i):

$$l_0 = \sum_i (-O_i + O_i \ln O_i) \tag{5.5}$$

By minimising the deviance at a selected aggregation level, the zone design system produces similar children accident values within each zone. The CC method of shape constraint is used providing a slight compactness rule that prevents extremely complicate shapes as discussed in Chapter 4. The A2Z system randomly generates *z* zones every time one of the 20 runs starts and improves the aggregation output for 300 iterations per zone.

At first stage, the SOAs were aggregated to a sequence of equally divided intervals of 2, 72, 144, 216, 288, 360, 500 and 719 zones and we measured the chi-square, deviance, AICc, AIC and BIC at each aggregation level. As the target of this method is to identify the minimum information criterion, we should select one of the three available criteria to use for further exploration. Hurvich et al. (1998) suggested the corrected AIC (*AICc*) is less biased than classical AIC, as the advantage of AICc is concentrated in the parameters calculation. In Table 5.4, the AICc suggests the need of closer investigation at the segments between the 113 and 216 zones. At the second stage the A2Z system aggregated 123,133,153,163,173,183 and 193 zones repeating the calculations of goodness of fit measures.

In Figure 5.1, the values of three criteria (AIC, AICc and BIC) are graphed suggesting the aggregation level of 183 zones as the most appropriate and informative level as all of them reach their minimum value. In depth aggregation at the segment between 173 and 193 could provide the optimum number of zones, but as it is mentioned at previous chapters the zone design system explores a number of possible solutions optimising certain criteria. This nature of zone design is responsible for the deviation between the informative measures and the theoretical curve (Figure 4.9: Chapter 4). In this case study, the 183 zones are representing the aggregation level in which the significance of selected social indicators is evaluated. Additionally, the AICc measure seems to increase the differences between aggregated levels as the parameters are calculated differently in each criterion. In Chapter 4, the AIC, AICc and BIC criteria have been extensively explained under a zone design context. In practice, Figure 5.7 illustrates the performance of the three criteria showing a rapid increase of AICc values after the 183 zones in comparison to AIC and BIC criteria.

At the bottom of Table 5.4, the measures of goodness of fit at ward level imply a high variation of children accident values in comparison with the 113 aggregated zones of A2Z system. Both chi-square (4318.41) and deviance (3000.92) measures show that the ward level is statistically equivalent to the homogeneous zones between 2 and 72 aggregation levels. As it is obvious, the aggregation levels of 2 to 72 zones are high scaled geographies with strong social heterogeneity within and between zones.

Comparing the ward level with 113 homogeneous zones, the table illustrates strong differences of variance supporting the comment that the ward level may not be suitable for analysis of health related phenomena such as the children traffic accidents. On the other hand, the suggested level of 183 zones scores far better at AIC, AICc and BIC criteria with -16438.85, -16312.97 and -16282.07 values, while at ward level the values are -14174.27, -14131.68 and -14077.46 respectively. Worth noted here is the SOAs level where the degree of freedom is zero representing the full model (l_0) as mentioned earlier in this section. The AIC, AICc, and BIC values are extremely high as the parameters of the full model reach the maximum value (719). Consequently, the two issues of selecting the appropriate aggregation level and constructing homogeneous zones based on social determinants have been addressed here using the statistical information criteria in a zone design context.

Table 5.4: Measuring the goodness of fit at various automated aggregation levels.

Number of Zones	Chi-Square	Deviance	df	AIC	AICC	BIC	1st Stage	2nd Stage
2	8462.93	3511.96	717	-13885.22	-13885.21	-13883.51	■	
72	1402.95	1236.16	647	-16021.02	-16004.75	-15959.34	■	
113	1175.32	1066.22	606	-16108.96	-16066.37	-16012.15	■	
123	1026.89	966.19	596	-16188.99	-16137.72	-16083.61		■
133	1010.13	869.79	586	-16265.39	-16204.46	-16151.45		■
144	855.06	786.20	575	-16326.98	-16254.23	-16203.62	■	
153	770.58	729.95	566	-16365.24	-16281.83	-16234.16		■
163	775.16	718.18	556	-16357.00	-16260.67	-16217.35		■
173	664.01	647.14	546	-16408.04	-16297.57	-16259.83		■
183	**619.36**	**596.33**	**536**	**-16438.85**	**-16312.97**	**-16282.07**		■
193	606.79	588.15	526	-16427.04	-16284.40	-16261.69		■
216	611.03	571.05	503	-16398.13	-16211.39	-16213.08	■	
288	538.91	519.56	421	-16305.62	-15918.50	-16058.89	■	
360	482.16	474.10	359	-16207.08	-15481.05	-15898.66	■	
500	438.95	438.99	219	-15962.19	-13664.03	-15533.83	■	
719	0.00	0.00	0	*FM*	*FM*	*FM*	■	
Ward Level (113)	4318.41	3000.92	606	-14174.27	-14131.68	-14077.46		

FM is the full model where the number of zones is same to areal units.

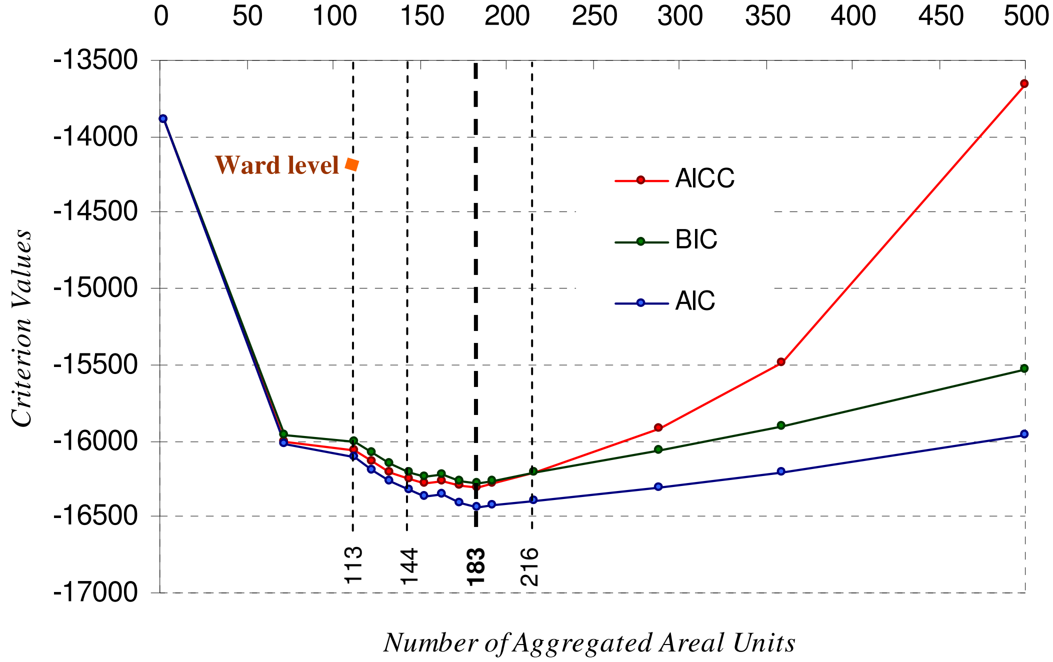


Figure 5.7: The performance of AICc, AIC and BIC measures as the number of zones increases

5.6 Results

From the above statistical measures, the level of 183 zones has been selected for further analysis involving the construction of 183 homogeneous zones using the Accident rates (Z_{ACC}) and Townsend Index scores (Z_{TI}). In Figure 5.8, the 183 homogenous zones constructed by Children Traffic Accidents rates (Z_{ACC}) are showing an increase of traffic accidents in areas with high density road network. As expected, the density of road network is determined by the population density and it is obvious that urban areas such as the Newcastle City are experiencing more traffic accidents than rural areas. However, there are two zones in Gateshead and one in Sunderland with high rates between 0.12 and 0.23. Concentrating in the Newcastle area, there is one zone with 1.08 accident rate. At first glance, it is seems to be a mistake as there are more incidents than the children population. A careful examination of the study area shows that the specific area suffers from many children traffic accidents but the children may not be resident of this problematic area as the data are reporting incident places without any information of home address of child. However, an explanation for the improper value of the rate could be the accident record file. This case study is based on the five years accident record and cases of children having more than one accident are possible (especially for slight severe accident). Therefore, the accident rate indicates how dangerous an area is rather how many children, residents of this area, involved in traffic accidents.

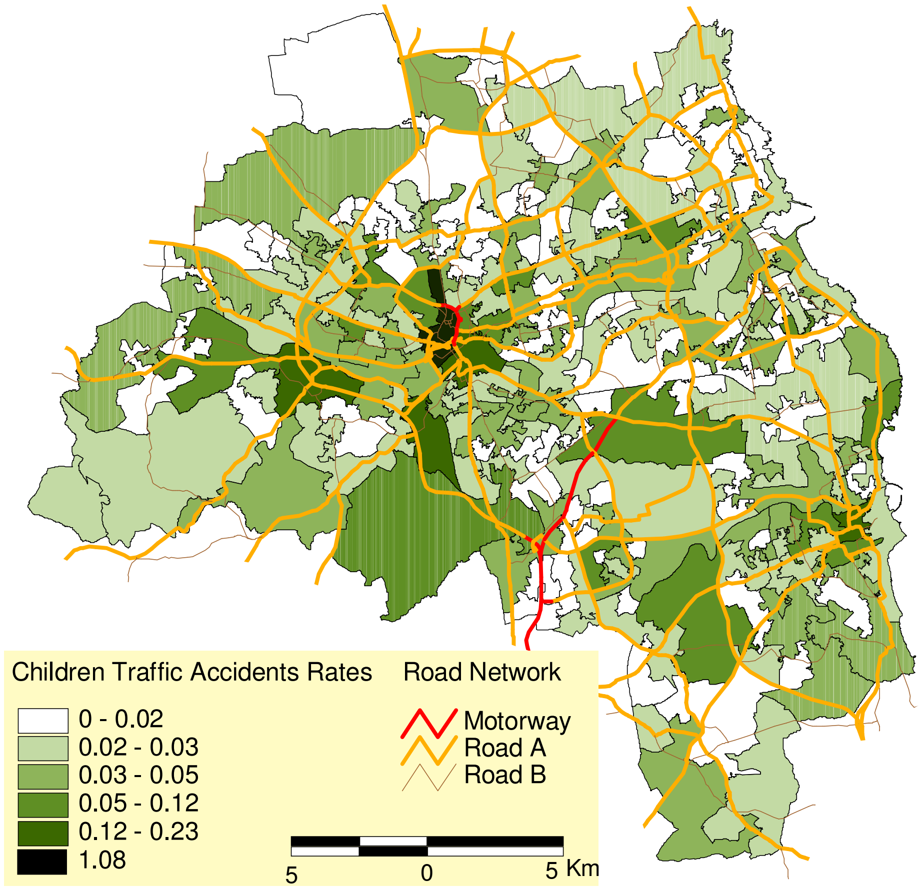


Figure 5.8: Map of Children Traffic Accidents in percentages using 183 homogenous zones by accident rates (Z_{ACC}) in Tyne and Wear.

The maps in Figure 5.9 show the Townsend Index scores in Tyne and Wear at SOA level. Also, the boundaries of the 183 homogeneous zones by children traffic accident rates are presented. Some direct findings show high deprivation values in Newcastle area especially at the areas adjacent to Tyne River and in Sunderland area. As expected, the 183 homogeneous zones by Townsend Index scores (Z_{TI}) is capturing the variations of deprived areas better than Z_{ACC} zoning solution. However, Z_{ACC} output zoning illustrates an almost equally homogenous aggregation level suggesting strong relationships between the children accidents and the deprived areas.

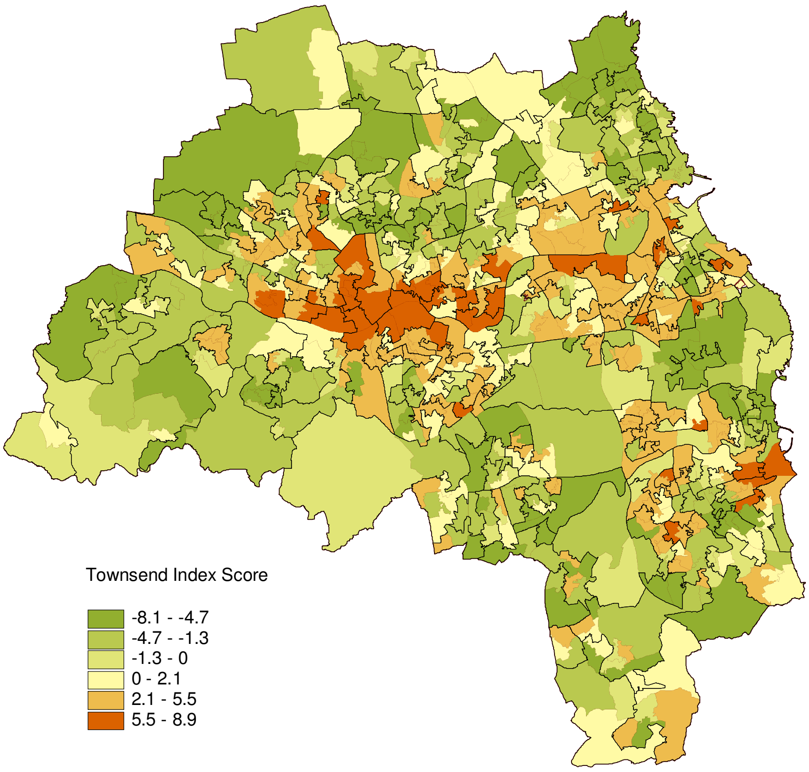


Figure 5.9: Map of Townsend Index scores at SOA level with 183 zones constructed by similar Children Traffic Accident rates.

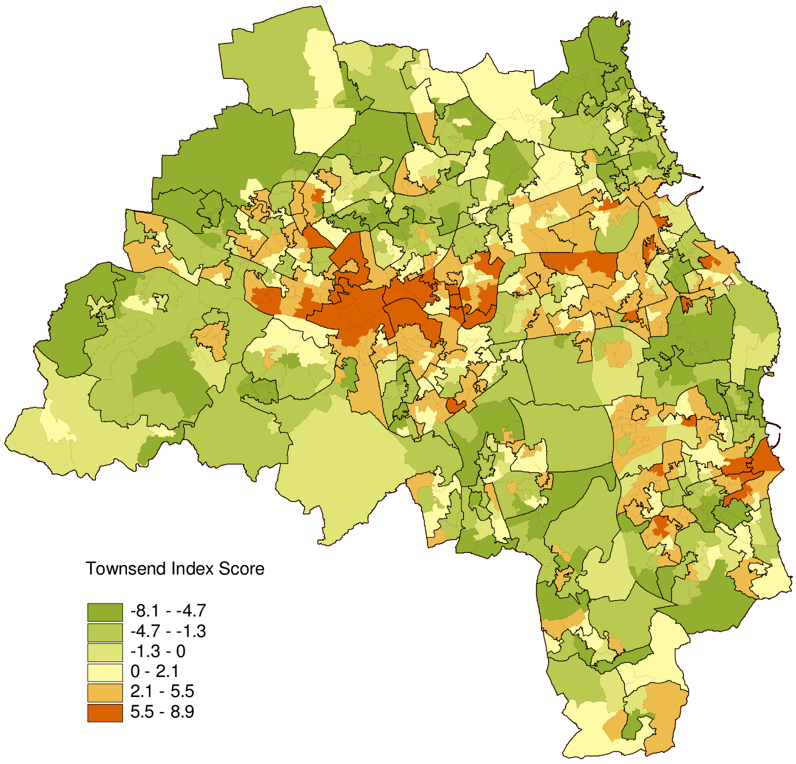


Figure 5.10: Map of Townsend Index scores at SOA level with 183 zones constructed by similar Townsend Index scores.

To statistically explore if both aggregations Z_{ACC} and Z_{TI} are strongly related the type of relationship between children traffic accident rates, indicators of material deprivation and Lone parent household determinant is investigated using exploratory analysis. In Table 5.4, the visual findings from the maps are statistically confirmed at both aggregated zones as the significance of children accidents with the determinants result in similar correlation coefficients. The 'overcrowding' determinant remained is now more strongly correlated at the Z_{TI} than Z_{ACC} zones with correlations of 0.474 and 0.425 respectively. The 'no car' and 'not owner occupied' determinants are significant at the 0.01 level but they are not as strongly related with children accidents as the 'overcrowding' indicator. Furthermore, the 'unemployment' determinant is not significant at any aggregation while the 'lone parent' indicator gives negative correlation suggesting that the single parent households are not related to children accidents. According to Table 5.5, the Z_{TI} zones provide better geographies than Z_{ACC} zones as there was noticeable improvement of correlations between children accidents and determinants, especially for the significant determinants at the 0.01 level.

Table 5.5: Correlations with children traffic accidents aged 0-17 years old at 183 zones level.

	Children Traffic Accidents aged 0-17 yrs old					
	183 zones (Z_{ACC})			***183 zones (Z_{TI})***		
	r	***Sig.***	***N***	***r***	***Sig.***	***N***
No Car	0.202**	0.006	183	0.234**	0.001	183
Overcrowding	0.425**	0.000	183	0.474**	0.000	183
Not owner occupied	0.209**	0.005	183	0.239**	0.001	183
Unemployment	0.081	0.275	183	0.098	0.186	183
Lone Parent	-0.175*	0.018	183	-0.085	0.252	183

** Correlation is significant at the 0.01 level (2-tailed)
* Correlation is significant at the 0.05 level (2-tailed)

For both suggested zoning solutions, their correlations are weaker than ward level as listed in Table 5.2. This is result of the increase from 113 wards to new 183 zones. As Gehlke and Biehl (1934) mentioned the correlation coefficient tends to increase as the level of geographical aggregation increases, therefore the differences between the correlation at wards and suggested zones levels reflect the 70 additional zones in the Z_{TI} and Z_{ACC} zoning solutions. The Z_{TI} zones obtain more unbiased estimates of correlation because the aggregation process was based on explanatory variables and as Blalock

(1964) mentioned the aggregation of target variables often causes severe biases of correlations.

Furthermore, the Figure 5.11 illustrates the children traffic accident rates at SOA level with the boundaries of 113 wards in Tyne and Wear. It can be seen that the wards are very heterogeneous in terms of accident rates and this is reflected in the measures of variance discussed earlier. In addition, local clusters of SOAs with high accident rates are separated by ward boundaries resulting in heterogeneous wards. Therefore, we strongly encourage the use of informative statistics for investigating the scale level, while the zone design system provides homogeneous zones with specific characteristics.

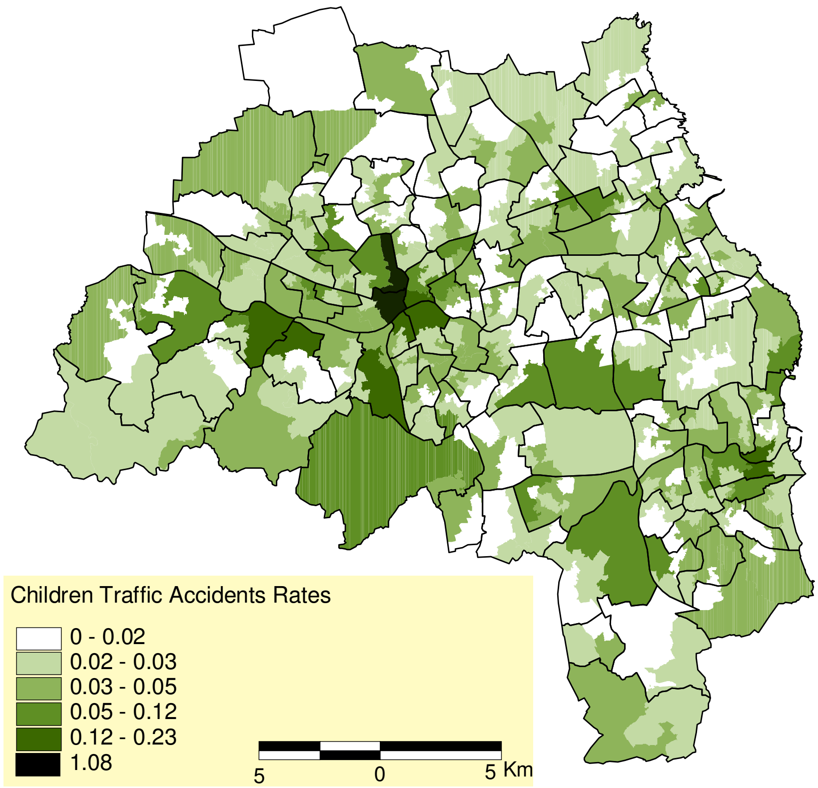


Figure 5.11: Map of children traffic accidents rates at SOA level with the boundaries of 113 wards in Tyne and Wear.

5.7 Discussion and Conclusions

We have shown an increase of correlation coefficient as the scale changes from SOA to ward in the area of Tyne and Wear. According to Gehlke and Biehl (1934) the correlation coefficient increases as the aggregation level increases and this is clearly presented in Table 5.3. Despite the changes of scale and aggregation, overcrowding, households without car and owner occupancy determinants are all significant at the 0.01 level in relation to children accidents. An interesting finding is the negative correlation of the lone parent household determinant at Z_{ACC} level, while in ward and Z_{TI} level it becomes insignificant. This findings is in contrast with the relevant literature, but it reflects the nature of data that has been used in this study as most of the literature studies use hospital admission data (Reading et al. 1999) or longitudinal datasets (Daras et al., 2005) at the household level. Hospital admission data usually consists of information about the home address of children and the severity of the incident, but does not provide the location of an accident (Stewart-Brown et al., 1986). This is a limitation of the available datasets in the UK, as there is linkage between incident and hospital admission data.

In this case study, a methodology for selecting the appropriate number of zones was proposed using advanced statistical measures such as AIC, AICc and BIC. In detail, we suggested the minimisation of AICc or similar criterion targeting the appropriate aggregation level (scale effect) and the Deviance or chi-square goodness of fit measures to explore the quality of different aggregations at the same level (aggregation effect). Applying the methodology in Tyne and Wear, we concluded that the most statistically informative level is the one consisting of 183 zones. The construction of homogeneous zones by Townsend Index scores and its comparison with the homogeneous zones by Children Traffic Accident rates exhibit significantly strong relationship of accidents with deprived areas. The next chapter will explore the aggregation issues at a middle level by means of another empirical study focusing on limiting long term illness in England and Wales.

CHAPTER 6

Building Middle Level Health Regions

6.1 Introduction

The effect of spatial scale is acknowledged in the research community as one of the most serious challenges the researchers face. As mentioned in the literature review of this thesis the scale effect in health geography is closely related to ecological fallacies. However, it is also known that the way data is arbitrarily aggregated and scaled can have an effect on the interpretation of results. Usually, health studies employ administrative areas as the units of analysis and occasionally the scale effect is determined as a crucial factor that affects the results of a study. On the other hand, the aggregation effect tends to be neglected by the researchers and the aggregated output zones are taken for granted using various aggregation methods such as clustering. Acceptance of such output geography may be granted only if the aggregation process is exhaustive providing all the alternative output zones. For small aggregation problems the exhaustive analysis can be easily implemented, but when the study area becomes larger involving complicated criteria of large number of areal units, such exhaustive algorithms are not possible to be implemented by the means of a personal computer.

In this chapter, the zone design investigates the impact of MAUP in middle scale geographies related to health and deprivation determinants. The middle level refers to the scale size of the basic areal units and in this study wards are grouped into larger zones. The use of A2Z system is suggested as a method of optimising criteria for the construction of homogeneous zones at health authority level. Three objective functions: k-means, chi-square and deviance are used to build homogeneous zoning solutions in terms of Limiting Long-Term Illness (LLTI) and their variations are compared with the existing health authority level using the correlation measures between LLTI, Townsend Index and the Townsend Index's four indicators. Furthermore, the MAUP is explored, focusing on the aggregation effect, using the within and between statistical variation of variables (analysis of variance). The zone design approach is proposed here as a more appropriate basis for analysis than any pre-existing census aggregation targeting health related zones.

6.2 Overview

The importance of geography in the social indices reports is strongly suggested by researchers, while in general spatial issues such as scale effects are introduced at a later stage for comparative purposes. The absence of spatial consideration in health studies is noticeable at policies in resource allocation affecting or excluding specific categories of groups and individuals. Moreover, the incorrect choice of resource allocation formulas for example the increase of general practitioners funds in deprived areas has been strongly criticised (Carr-Hill and Sheldon, 1991). As Townsend (1979; 1987) noted, the deprivation is a wider concept of poverty and it could be defined as disadvantaged areas with lack of material resources. Despite acknowledgement that spatial issues, such as the MAUP and deprived areas should be explored in conjunction, very few studies observe the geographical parameters using an entire spatial overview of the study area. Furthermore, literature suggests a close relationship between deprivation and households with health needs across England and Wales (Boyle, 1993; Boyle et al., 1998). Empirical evidence points out that the areas with high LLTI incidents rates are strong related to deprived areas, while the increase of material deprivation conceals possible increase of LLTI incidence.

In England and Wales, the census datasets are the most complete and accurate information that concerns the whole population, while additional socioeconomic indicators are recorded. The census reports the collected information in census geographies providing a conductive ground for researchers to investigate socioeconomic characteristics at the national or local level. Such an example is the empirical study by Cockings and Martin (2005) where the relationship between deprivation and limiting long-term illness in the county of Avon is investigated comparing the correlations of LLTI, Townsend Index and Townsend Index's indicators. Their study concentrated on the MAUP effects at a range of aggregations using enumeration districts (EDs). They constructed equally populated zones for a variety of aggregation levels using the automated zone procedure. While detailed discussion of the scale effect was reported comparing the correlations between LLTI and Townsend index scores, their research explores the aggregation effect into an insufficient extent by minimising the population threshold to 80% of the target value at four representative scales. However, their findings suggested an increase of correlation between deprivation and morbidity as the size of zones increases (Figure 6.1). This appears to be in agreement with the results suggested by Openshaw and Taylor (1979).

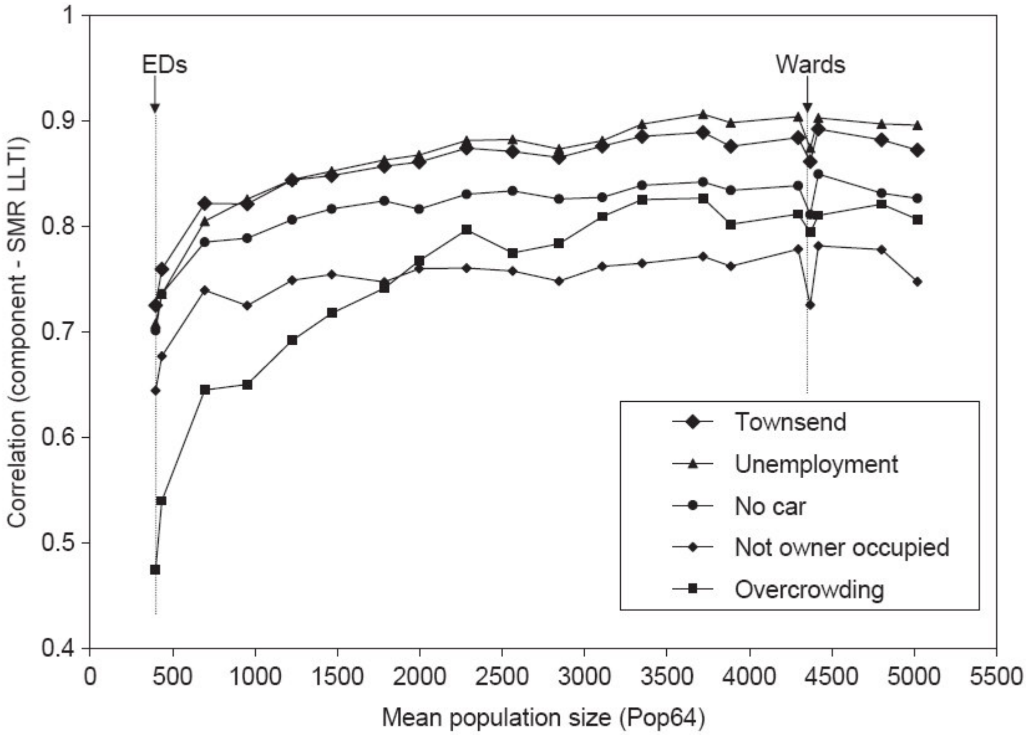


Figure 6.1: Correlation between Townsend deprivation score and Townsend components and SMR LLTI 0–64 by mean zone (Pop64) size. <u>*Source*</u>: Cockings and Martin (2005), p.2739

Similarly, the design of homogeneous health administrative areas in terms of limiting long-term illness proposed by Daras and Alvanides (2005), demonstrating some of the potential capabilities of A2Z as a spatial analytical tool, aggregating roughly 10,000 wards in England and Wales. In their study, the 1991 census wards aggregated into 195 zoning solutions exploring the variability of boundary changes with and without shape constraints. Part of this research is adopted in this case study for the evaluation of objective functions at constraint and free aggregation environment. To investigate further the above findings, a supplementary investigation of aggregation and scale effects and the use of three different homogeneity functions are explored using the 1991 and 2001 census datasets in England and Wales.

6.3 Data and statistical methods

The census deprivation information and geographies have been collected using the CasWeb and UK Borders services respectively. In particular, CasWeb provided the deprivation indicators for the Townsend Index and the limiting long term illness cases for both England and Wales Censuses in 1991 and 2001 (Table 6.1). The construction of Townsend Index has a theoretical background as its four variables were not selected at random but for what they represented. Therefore, despite the fact that both the measurement and reporting of these variables has changed since 1991, the 2001 the census measures should still provide valuable information concerning the relationship of deprivation and LLTI.

Table 6.1: The census variables used as Townsend and LLTI indicators

Census	Table	Indicators	Numerator Variables	Denominator Variable
1991	SAS 23	Overcrowding	3+4	1
	SAS 08	Unemployed	78+232-240 -241-87	12+166-174 -175-21
	SAS 20	Owner Occupancy	1-2-3	1
	SAS 20	Car availability	10	1
	SAS 12	Limiting long-term illness	1	2 (SAS 01)
2001	KS019	Overcrowding	4	1
	KS009	Unemployed	5	2+3+4+5+6
	KS018	Owner Occupancy*	2+3	1
	KS017	Car availability	2	1
	KS008	Limiting long-term illness	2	1

* Calculated as follows: 100 – ((KS0180002 + KS0180003/ KS0180001)*100)

Statistically, the Townsend Index is calculated by summarising the z scores of each variable, while a z score statistic is used for standardising the indicators. The z score represents the number of standard deviations between the percentage value of an area and the mean value for all areal units. It is negative when the value of area is below the mean and positive when it is above. Mathematically, it is formulated as:

$$z_i = (y_i - \overline{y})/SD \qquad (6.1)$$

Where z_i is the *z* score for a given area *i*, y_i the percentage value of the area, $\overline{y}$ is the mean value for all areal units in the dataset and *SD* is the standard deviation for all areal units. In this case study the mean and standard deviation are calculated at the ward level covering England and Wales. Additionally, the unemployment and overcrowding variables are used in Townsend Index as logarithmic values. If there are areal units with zero values then the two determinants can not be calculated as logarithm of zero does not exist. The areal units with zero values are tackled by adding +1 at the percentage values (v_i) as follows in equation 6.2.

$$L_i = Log(v_i + 1) \qquad (6.2)$$

As a result, the z value of problematic areal units becomes zero which is equal to the mean of all areal units.

Moreover, the LLTI variable has been standardised following equation 6.3. The methods of standardising variables using percentages and z-scores have limitations as Simpson (1996) suggested. For example, the small number effect could become a serious issue as areas with extremely small numbers of cases and population can produce high rates and misleading statistical results. The aforementioned difficulty can be tackled by identifying the problematic areas with small number effects and excluding them from the statistical and spatial operations. In this study, the LLTI variable has not been standardised for sex and age groups. However, it is recommended that further research on this subject should be targeted on specific social groups.

LLTI (%) = (*LLTI cases in Households* / *Total residents in Households*) x *100* (6.3)

In addition, the 1991 and 2001 census geographies are used in this study covering the study area of England and Wales. Census wards were selected as the basic areal units and health authorities as the target aggregation level for this study. In detail, the datasets of the 1991 census geographies consist of 9,504 wards and 195 District Health Authorities (DHAs) in England and Wales, while the 2001 census geographies involve 8,796 wards and two different geographies for Health Authorities (HAs) constructed before and after April 2002 (100 and 50 respectively).

In general, census wards for a decade remain unchanged in term of their boundaries as they are specifically designed for carrying the census variables. On the other hand, HAs are frequently redrawn, depending on various health policies and initiatives (Daras and Alvanides, 2005). During the Census 2001 enumeration and before the data became available to the public, the Department of Health reconstructed the 100 HAs to 50 Strategic Health Authorities (SHAs). As the 2001 Census was carried out on 29 April 2001 (National Statistics, 2005) this research has to identify the appropriate scale that provides the best results. Although, it is looks more appropriate to use the HAs rather the SHAs as the HAs were abolished in April 2002, it is also important to investigate the appropriate scale using statistical measures. For further statistical investigation of health authorities' differences, we compare the SHAs and the HAs with two simple measures of variance (chi-square, deviance) and three informative criteria (Figure 6.2). As their degrees of freedom are different the measures of variance are not appropriate for a comparison concerning datasets in different scale levels because the degrees of freedom change. For this reason the evaluation of health authorities is concentrated in the three informative criteria which provide a joined form of deviance and degrees of freedom as introduced in Chapter 4 and extensively used in the identification of the most informative aggregation level in Chapter 5.

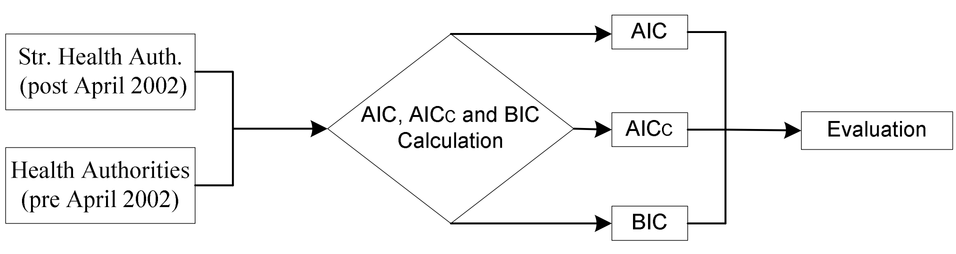


Figure 6.2: Evaluation of SHAs and HAs by means of three informative criteria

In Table 6.2, the large number of degrees of freedom in contrast with the small number of zones indicates that both aggregation levels (HAs and SHAs) are not statistically informative enough. However, such results are expected as the study explores aggregation phenomena at the district level where the differences within each zone are highest. The Chi- Square and Deviance measures of variance can not provide a clear suggestion if the 100 HAs are a better geography for statistical analysis. On the other hand, the use of information criteria such as the AIC, AICc and BIC suggests that HAs are the most informative level with the AIC, AICc and BIC scoring the lower values of -730,188, -730,185 and -729,993 respectively.

Table 6.2: Statistical characteristics of each aggregation level for 2001 Census datasets

Aggregation Level	Number of Areal Units	Number of Zones	Degrees of Freedom	Chi- Square	Deviance	AIC	AICc	BIC
Wards	8796	8796	0	0	0	*FM*	*FM*	*FM*
Health Authorities	8796	100	8696	299762	302468	-730188	-730185	-729993
Strategic Health Authorities	8796	50	8746	343154	344742	-688013	-688013	-687916
England & Wales	8796	1	8795	585218	578955	-453899	-453899	-453897

FM is the full model where the number of zones is same to areal units.

6.4 Designing Health Authorities in England and Wales

After the statistical and spatial correction of datasets, an investigation was implemented to study the parameter needs of the aggregation process. Such a piloting attempt is appropriate to identify the most suitable aggregation parameters which provide the most advantageous use of computer processing as zone design consumes mainly processor power. In the current research, one more reason to recognise the most efficient settings is the large number of areal units that could expand the process time of aggregation without a considerable improvement of results. Using the A2Z system, the 1991 census data was aggregated to 195 zones using a *k*-means objective function. The selection of *k*-means as a pilot function was proposed as it has a more complicated structure than the other two functions: deviance and chi-square. Thus, the zone design system utilises the most time consuming function for archiving the optimum solution. The pilot aggregation was carried out twice: first, the system aggregated the areal units using a shape constrain; then, the shape constrain was removed providing additional flexibility during the aggregation process. Both aggregations were repeated for 20 runs and 200 iterations computing roughly 30,000 boundary changes per run. The number selection of runs and iterations selected based on empirical knowledge concerning the strength of computer processor and the position where the objective function is incapable to improve the objective score further.

In Figure 6.3, the graph indicates that the homogeneity score before the aggregations is 101.36. For unconstrained zoning (green line) homogeneity improves rapidly as the number of the iterations increases reaching 79.65 for 200 iterations. On the other hand, the contiguity constrain method (red line) blocks any improvement of homogeneity after the 100th iteration. In general, the shape constraint method reduces the evolution of the homogeneity algorithm at the expense of the objective function score, compared to the unconstrained approach. Mapping the results of each aggregation and comparing them with the DHAs, it is obvious that the unconstrained method is providing extremely complicated shapes intending to build more homogeneous zones (Figure 6.3.c). On the other hand the shape constrained zones present similar shapes with DHAs with same noticeable changes at the zone boundaries (Figure 6.3.b).

The pilot results indicate the unconstrained zones as inappropriate outputs in such a large dataset. The use of a shape constraint would be beneficial for this case study as the output zones preserve compact shapes and in addition their homogeneity increases in comparison with the DHAs. These pilot aggregations assist the methodological structure of this study identifying the appropriate zone design setting in terms of processing time, use of constrains and particular characteristics of such large datasets.

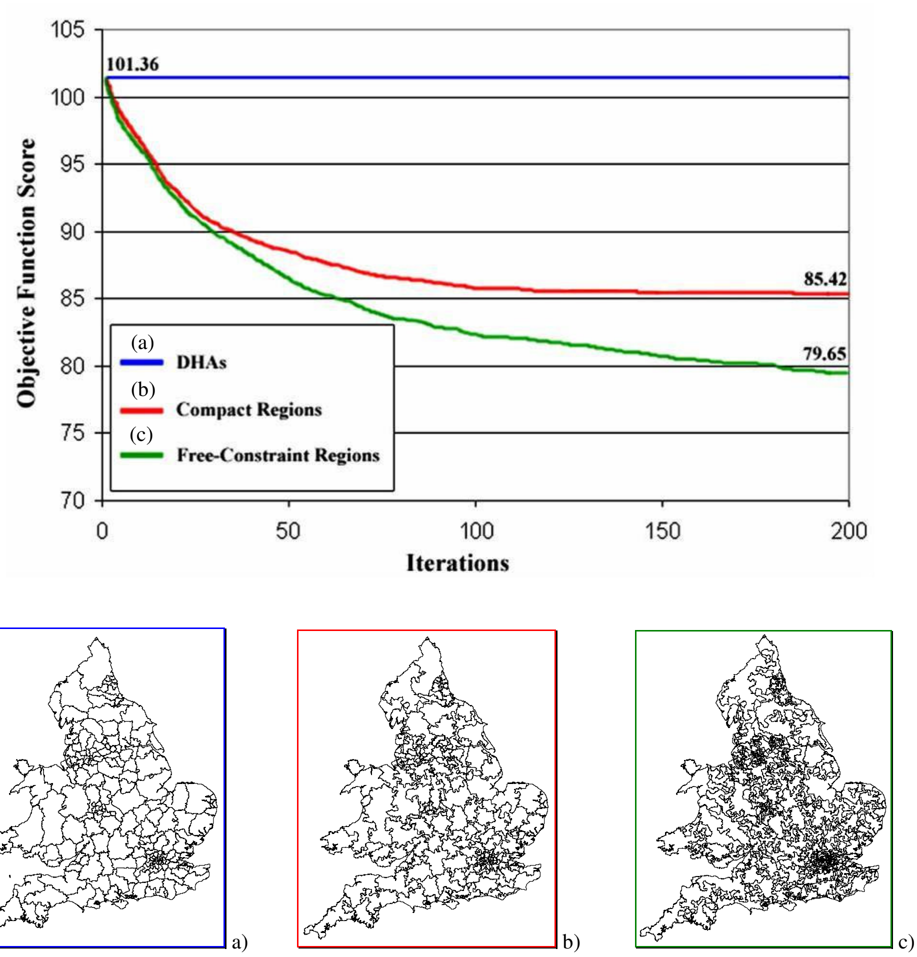


Figure 6.3: The improvement of LLTI homogeneity according the objective function score and compactness, for: a) DHAs, b) compact zones and c) free-constraint zones.

Subsequently, the A2Z system and its parameters for use in the current case study could be illustrated in the following steps:

Step 1 Aggregation of A areal units into Z zones, where $Z < A$.

In the 1991 census dataset: A = 9,504 wards and Z = 195 DHAs.

In the 2001 census dataset: A = 8,798 wards and Z = 100 HAs.

Step 2 *Parameters set:*

Objective Function: a) k-Means, homogeneity with LLTI

b) Chi-square, homogeneity with LLTI

c) Deviance, homogeneity with LLTI

Shape Control: Advanced Contiguity Constraint (ACC)

Runs = 20; Iterations= 200; Idle Iteration= 25%

Step 3 Upload health geographies (DHAs or HAs) as starting zones.

Step 4 Select zone *i*, where $i = 1 \ldots Z$.

Step 5 Identify a set of areal units bordering on members of zone *i* that could be moved into zone *i* without destroying the internal contiguity of the donor zone(s).

Step 6 Randomly select an areal unit from this list and if there is a local improvement in the current value of the objective function or a move that is equivalently as good as the current best then make the move, update the list of candidate areal units, and return to step 4 or repeat step 5 until the list for zone *i* is exhausted.

Step 7 Repeat steps 4 – 6 until the maximum number of iterations (200) is exceeded or no further improvements can be made; this is at the idle iteration of 25%, where the zone design will finish if there is no further improvement of the objective function (Openshaw and Rao, 1995).

Step 8 Repeat steps 3 – 7 until the maximum number of runs (20) is exceeded.

As described above, the A2Z system aggregates the 1991 and 2001 census wards reconstructing the 195 DHAs and 100 HAs respectively. In this study the output zones have been created based on the LLTI according to two criteria: homogeneity and compactness. The criteria of homogeneity were based on the k-means, chi-square and deviance functions. For the k-means function the aggregation was carried out using percentages of LLTI while for the chi-square and deviance functions the zones were based on the observed and expected LLTI rates.

As the Figure 6.4 illustrates, the zone design system creates three zoning solutions for each census dataset. In each zoning solution we used a different homogeneity function and the advanced contiguity constraint for more compact zones as introduced in Chapter 4. The new zoning solutions were produced targeting to maximise homogeneity of LLTI percentages within each zone. For further statistical analysis of new zoning systems the available census information at ward level summarised and recorded to the new dissolved zone solutions. In addition the Townsend Index scores and the percentages of each variable measured at the new zoning solutions, DHAs and HAs levels.

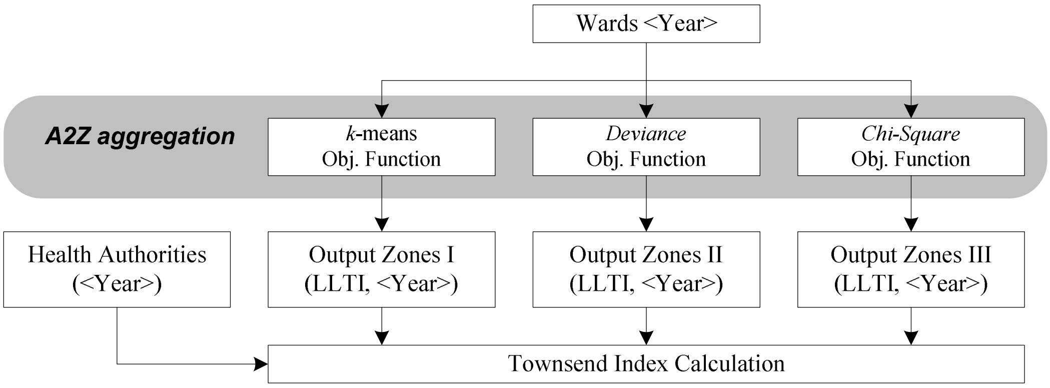


Figure 6.4: Generic organisation diagram of the case studies.

Moreover, the most homogeneous zoning solution for each year (i.e. 1991 & 2001) was identified by measuring the variance of each zoning with two functions chi-square and deviance. The zoning solution with high homogeneity in terms of LLTI is compared with the respective health authority and census wards exploring the scale effect while the comparison of health authority with the three new zoning systems provides valuable information concerning the aggregation effect.

As Fotheringham (1997: p.157) suggests, exploratory analysis provides useful information to "*aspects of data or a modelling technique that might be worth investigating in more detail*." Thus, exploratory analysis of the relationship between LLTI, Townsend Index and Townsend material indicators was carried out in SPSS and ArcView for all the Census geographies including the new output zones created by the zone design system.

6.5 Results and Analysis

6.5.1 Evaluating the different objective functions

Prior to the exploratory analysis, the most homogeneous zoning solutions were evaluated using two common variance functions: chi-square and deviance. Both functions are explained in detail in Chapter 4 providing the theoretical background and how they were used in zone design as functions or alternatively as measures of variance in the new output zones. The calculation of chi-square and deviance was carried out in

ArcView using an Avenue script developed to statistically support the case studies of this thesis, measuring five statistical methods: chi-square, deviance, AIC, AICc and BIC. In Table 6.3, the 1991 new zoning solutions *I*, *II*, *III* and Health Authorities are compared illustrating a clear improvement of zone design against the 195 Health Authorities. Both deviance and chi-square values at HAs level are considerable higher than the values at the new aggregations suggesting more variation within HAs than the new zoning solutions. This was partly expected as all three objective functions are constructed for optimisation of maximum homogeneity within zones. Concentrating on the three new zoning systems, zoning *I* has the minimum value for all variance measures. The chi-square has a value of 183,938 and the deviance 183,160, while zoning *II* produces higher values of 196,813 and 196,799 respectively. The worst homogeneity values were recorded for zoning *III*, with chi-square 211,115 and deviance 211,098.

Table 6.3: Statistical comparison of 1991 output zones I, II, and III using variance measures.

Aggregation Level	Number of Areal Units	Number of Zones	Degrees of Freedom	Chi- Square	Deviance
Wards 1991	9504	9504	0	0	0
Zones I_{1991} (*k*-means)	9504	195	9309	183938	183160
Zones II_{1991} *(Chi-Square)*	9504	195	9309	196813	196799
Zones III_{1991} *(Deviance)*	9504	195	9309	211115	210689
Health Authorities	9504	195	9309	238928	238992
England & Wales	9504	1	9503	471191	461790

Repeating the same calculations for the 2001 data and output zones, the best score in terms of homogeneity was produced by zoning *III*. In Table 6.4 zoning solution III has the lowest chi-square value of 251,451 and deviance 253,717 compared to the other two zoning solutions *I* and *II*. In addition, solution *I* seems to perform worst, with a chi-square value of 256,267 and deviance 258,283. The values of chi-square and deviance for all zoning systems in the 2001 dataset produce small differences in terms of goodness of fit. This is explained by the large aggregation scale of the case study as 8,796 areal units are grouped into 100 zones and even after applying zone design the heterogeneity inside the zones is still strong.

Table 6.4: Statistical comparison of 2001 output zones I, II, and III using variance measures.

Aggregation Level	Number of Areal Units	Number of Zones	Degrees of Freedom	Chi- Square	Deviance
Wards 2001	8796	8796	0	0	0
Zones I_{2001} (*k*-means)	8796	100	8696	256267	258283
Zones II_{2001} (*Chi-Square*)	8796	100	8696	252208	254798
Zones III_{2001} (*Deviance*)	8796	100	8696	251451	253717
Health Authorities	8796	100	8696	299762	302468
England & Wales	8796	1	8795	585218	578955

As a result of both statistical comparisons, the 1991 zoning solution *I* and the 2001 solution *III*, based on the *k*-means and deviance objective functions respectively provide the most homogeneous zones. During the aggregation process, the zone design system stored the objective function scores for further exploration of the optimisation path each zoning solution followed. The produced graphs have been standardised using percentages at the y axis instead of raw numbers. Although standardisation distorts any information concerning the most homogeneous zone, it makes it possible to compare and identify the way different algorithms improved their scores. As Figure 6.5 shows, the combined graphs of the three objective functions have been constructed for the 1991 and 2001 new zoning system.

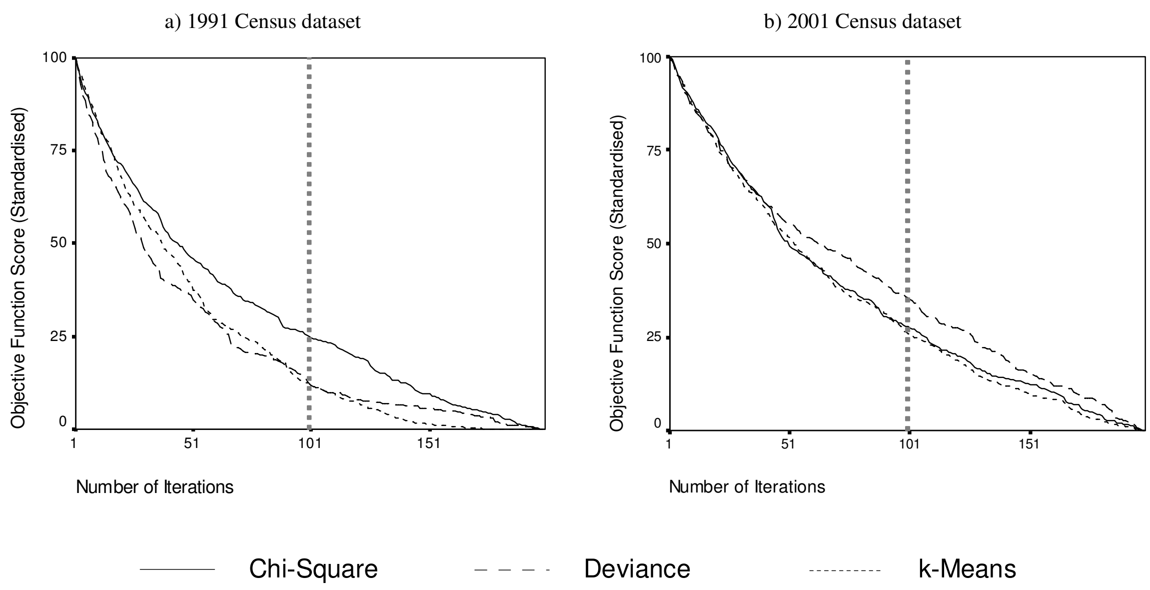


Figure 6.5: The performance of objective functions for 200 iterations using standardised scores

In Figure 6.5.a, the deviance and k-means objective function scores optimise the criterion faster than the chi-square method. In addition to the previous statistical comparison (Table 6.2) the *k*-means zones (solution *I*) provide the most homogeneous geographies suggesting a link between the way optimisation occurs and the output zones. On the other hand, in the 2001 graph the optimisation process performed in a similar way for all three functions presenting more consistent improvement of scores than the 1991 objective functions. Visual comparison of both graphs suggests that extreme abnormalities of the objective function scores during the aggregation process do not result in satisfactory final zoning solutions. As this comment is derived from the best six optimisations, it is suggested that future in depth research on a larger number of optimum solutions at the same aggregation level should investigate this observation. In addition, performance differences in the 1991 graph are most likely to be affected from the aggregation level the analysis took place. While the 1991 wards were grouped in 195 output areas, the 2001 wards were aggregated in 100 output areas. Although the zone design system can group the 1991 wards into 100 output areas, we favoured the number of health authorities at each census period because of changes in the definitions of census variables. The different aggregation levels as well as the different number of basic areal units prohibit any further analysis. This is expected as the output zones depended, to some extent on the basic areal units used (Openshaw and Taylor, 1981: p.60). Furthermore, in both Tables 6.3 and 6.4 the zoning solutions based on deviance as the objective function should attain the best result in terms of deviance, but Table 6.3 shows that *k*-means provides the optimum result. This can be explained by the early steps in zoning process such as the initial aggregation method and by the selected aggregation settings.

6.5.2 Measuring scale relationships

In order to investigate the relationship between the LLTI, Townsend Index and Townsend's indicators, exploratory analysis was carried out providing maps, descriptive information and correlations for all census and new geographies. The 1991 and 2001 census wards and health authorities are compared with the optimum zones in terms of within LLTI homogeneity intending to identify spatial changes across England and Wales.

Exploring the 1991 maps (Figure 6.6) it is obvious that increase of scale provides less detailed areas. Figures 6.6.a, 6.6.b, 6.6.c and 6.6.d illustrate the variation of LLTI for administrative areas in England and Wales using the quantile method with five categories. The darker shades in the map represent higher percentages of LLTI cases. At first glance, Figure 6.6.b clearly highlights the two well known Health Authorities with very high LLTI rates: Mid-Glamorgan in Wales and Durham in England. Figure 6.6.a uses the same visualization method to highlight more detailed patterns given the smaller size of wards, with the DHA boundaries appearing as lines. By visually comparing the two maps, it becomes apparent that DHAs are very heterogeneous in terms of LLTI rates. For example, a very clear pattern of high LLTI concentration for wards in West and Mid-Glamorgan, South Wales, has been divided amongst five DHAs. The proposed zoning solution *I* (*k*-means) is illustrated in Figures 6.6.c and 6.6.d using the same visualisation method. As expected, the new geographies capture better the LLTI incidences joining areas with similar LLTI values. Statistically speaking, the average of LLTI cases at ward and DHA level are very similar with means of 12.52% and 12.90% respectively, and it is noted that it increases slightly as the aggregation increases. However, the values are more spread about the mean at the ward level, with a standard deviation of 3.64 compared with 2.46 at the DHA level. In Wales, the LLTI values at DHA level are very high while the output zones correctly map high values of LLTI in the most populated areas with the rural area in the centre of Wales being close to average LLTI values. Moreover, the lower incidence of LLTI at ward level spreads throughout England and Wales whereas at DHA level, it is concentrated around the greater London area. These geographical variations of LLTI are an example of the MAUP tackling the scale effect as explored in the current thesis and higher standard deviation of a ratio at finer spatial unit naturally occurs due to the small number problem.

Similarly, the LLTI incidence in 2001 has been mapped at the ward and HA level investigating changes or identifying repeated characteristics over time. In Figure 6.7, it can be seen that areas in England and Wales with high incidence of LLTI are similar to the maps in 1991. Also noticeable is the increase of LLTI incidence in the whole of England and Wales with lower percentages of LLTI concentrated around the greater

London area. The average of LLTI cases for the ward and HA levels are 18.17% and 18.71% respectively illustrating a slight increase, similar to the 1991 dataset. In addition, the spread about the mean is larger at ward level with a standard deviation of 4.74 compared to 3.22 at HA level.

The investigation of the incidence of LLTI across England and Wales in 1991 and 2001 suggests that LLTI is subject to spatial scale. For further exploration of these variations, the following exploratory analysis enables statistical conclusions to be drawn based on the indicators of the Townsend Index and the overall Townsend Index itself. As suggested by Gehlke and Biehl (1934), the correlation coefficient tends to increase as the level of geographical aggregation increases. In Table 6.5, the 1991 variables result in higher correlation coefficients as the aggregation level increases except from overcrowding correlations for DHAs and zoning system *I* (*k*-means) where suggest a decrease of correlation coefficient as the zone size increases. However, the new zoning system (*I*) increases all correlations providing zones with better relationships to LLTI indices while the overcrowded value of 0.207 at DHA level improves dramatically to a value of 0.269 at zoning *I, very close to the value at ward level (0.285).* All 1991 determinants are significant at $p < 0.01$ level with the 'unemployment' and 'no car' indicators scoring the highest correlations.

The earlier statement by Gehlke and Biehl (1934) applies to the 1991 variables and it is also supported by the findings in other relevant studies (Cockings and Martin, 2005; Openshaw and Taylor, 1979), although in the 2001 wards, health authorities and suggested zoning solution the correlations tend to perform out of the ordinary findings. At the 2001 ward level, the census variables are significant to LLTI except from the 'overcrowding' variable. Also, noticeable is the small correlation between 'Not owner occupied' and 'LLTI'. The increase of aggregation level proposes negative correlation between the 'overcrowding' and 'LLTI' variables, while the 'not owner occupied' variable is not significant for either HAs or zoning *III*. Once more, the 'unemployment' variable increases the correlation coefficients as the scale increases. It is evident from the analysis that the 'overcrowding' variable is not strongly related to LLTI incidences in the 2001 data. As the correlations of determinants expected to follow the empirical findings from other studies, we also list the correlations of deprivation and LLTI at district level. From Table 6.5 it can be seen that there are two indicators (no car, not owner occupied) and the Townsend Index which demonstrate a decrease of correlation coefficients as the zone size increases. Moreover, the new zoning solution *III* provides better correlations than the HAs level reaching similar values to district level.

Comparing these findings with the results of study by Cockings and Martin (2005), we can see that although there is an increase of correlations at lower scales (close to EDs), close to ward level their correlation measures perform in some cases similar to our findings. In addition, at their highest level the correlation coefficients tend to decrease suggesting that our inconsistent findings between wards, districts and HAs may be products of spatial phenomena utilised at such high aggregated scales. A possible explanation can be the different number of areal units per zone as we analysis a national dataset with the wards as basic areal units while Cocking and Martin study datasets at local level in AVON area. Therefore, further research should investigate this phenomenon exploring a wide range of scales, basic areal units and indicators.

Table 6.5: Correlations with LLTI in England and Wales, 1991 and 2001 Censuses.

	LLTI – 1991 Census			**LLTI – 2001 Census**			
	Wards	**DHAs**	***OZ (I)***	***Wards***	**Districts**	**HAs**	***OZ (III)***
No Car	0.695** Sig. 0 N= 9504	0.723** Sig. 0 N= 195	0.728** Sig. 0 N= 195	0.527** Sig. 0 N= 8796	0.425** Sig. 0 N= 375[Я]	0.357** Sig. 0 N= 100	0.401** Sig. 0 N= 100
Overcrowding	0.285** Sig. 0 N= 9504	0.207** Sig. 0 N= 195	0.269** Sig. 0 N= 195	**0.018** Sig. **0.094** N= 8796	-0.169** Sig. 0 N= 375[Я]	*-0.295*** *Sig. 0.003* *N= 100*	*-0.302*** *Sig. 0.002* *N= 100*
Not owner occupied	0.400** Sig. 0 N= 9504	0.442** Sig. 0 N= 195	0.436** Sig. 0 N= 195	0.294** Sig. 0 N= 8796	0.145** Sig. 0 N= 375[Я]	**0.094** Sig. **0.352** N= 100	**0.159** Sig. **0.114** N= 100
Unemployment	0.648** Sig. 0 N= 9504	0.722** Sig. 0 N= 195	0.760** Sig. 0 N= 195	0.578** Sig. 0 N= 8796	0.610** Sig. 0 N= 375[Я]	0.610** Sig. 0 N= 100	0.644** Sig. 0 N= 100
Townsend Index	0.578** Sig. 0 N= 9504	0.579** Sig. 0 N= 195	0.613** Sig. 0 N= 195	0.449** Sig. 0 N= 8796	0.308** Sig. 0 N= 375[Я]	**0.197*** Sig. **0.050** N= 100	**0.254*** Sig. **0.011** N= 100

** Correlation is significant at the 0.01 level (2-tailed)
* Correlation is significant at the 0.05 level (2-tailed)
[Я] does not include Isles of Scilly district

6.5.3 Measuring aggregation relationships

For further examination of MAUP, except from the scale effect, this research also concentrated on the aggregation effect, investigating the correlations at the DHA level, including DHAs, zoning systems *I*, *II*, *III* and analysing the variance of LLTI, no car, overcrowding, not owner occupied and unemployment variables. The analysis of variance (F) or one-way ANOVA is a procedure that compares the spread between the zone means with the spread of values within each zone. The use of F value can provide a useful homogeneity measure formulated as follows:

$$F = \frac{V_B}{V_W} \qquad (6.4)$$

where, V_B is the between zones variance and V_W is the variance within each zone.

The use of the F value can be used to demonstrate the general changes of variance in a dataset with the same degrees of freedom (df) but it can not indicate which of the means (or zones) is responsible for these changes. In addition, it is noted that the use of F values in small datasets will produce misleading results due to the assumption of a normal distribution. However, this is not the case here as the dataset is large enough to provide unbiased results.

In Figure 6.8, the new zoning solutions (maps b, c and d) and the DHAs (map a) are visually compared using percentages of LLTI incidents. The maps categorised into five groups using the quantile method where the lighter brown colours reflect to areas with small percentages of incidents while the darker colours related to areas experiencing high percentages of LLTI incidents. Although all maps show South Wales, Liverpool, Sheffield and Durham areas with high LLTI percentages, the most important here is the boundary comparison between the new zoning solutions because it is possible to visually compare the different objective functions. The new zoning systems operate to very similar manner targeting to capture the concealed patterns. In Figure 6.9, almost identical findings are present. The high rated areas do not dramatically change and the objective functions produce similar zones in terms of LLTI homogeneity. Although both Figures 6.8 and 6.9 suggest that all three objective functions operate to a similar manner, the correlations between deprivation and LLTI as well as the analysis of variance can provide a clearer picture.

In Table 6.6, the three new zoning solutions and the DHAs are compared searching high correlations between the LLTI incidences, Townsend Index and Townsend indicators. Although the zoning systems *II* and *III* (chi-square and deviance) suggest that the correlations between the LLTI, 'no car' and 'not owner occupied' indicators are better than those of the zoning *I* (*k*-means), the zoning solution *I* provides an overall improvement of correlations. In general, the new zoning systems provide better correlations between LLTI and deprivation measures suggesting strong relationships. Following Table 6.3, the zoning *I* (*k*-means) is the most homogeneous in terms of LLTI incidences. Therefore, the homogeneity of deprivation determinants in comparison to LLTI incidences is presented in Table 6.7. The zoning solution *I* improves also the homogeneity of deprivation determinants with the 'unemployment' and 'no car' indicators to improved more than others with *F* values 39.2 and 33.1 respectively. In addition, the zoning *II* minimises the variation within zones to a lesser extend, while the solution *III* has the worst *F* values in comparison to zonings *I* and *II*.

Table 6.6: Correlations with LLTI at DHA level in England and Wales.

	LLTI – 1991 Census			
	DHAs	***k*-means**	**X^2**	**Deviance**
No Car	0.723** Sig. 0	0.728** Sig. 0	0.749** Sig. 0	0.742** Sig. 0
Overcrowding	0.207** Sig. 0	0.269** Sig. 0	0.256** Sig. 0	0.244** Sig. 0
Not owner occupied	0.442** Sig. 0	0.436** Sig. 0	0.482** Sig. 0	0.480** Sig. 0
Unemployment	0.722** Sig. 0	0.760** Sig. 0	0.752** Sig. 0	0.743** Sig. 0
Townsend Index	0.579** Sig. 0	0.613** Sig. 0	0.614** Sig. 0	0.607** Sig. 0

N = 195

** Correlation is significant at the 0.01 level (2-tailed)
* Correlation is significant at the 0.05 level (2-tailed)

Table 6.7: Analysis of Variance at DHA level in England and Wales.

	LLTI	Overcrowding	Unemployment	Not Owner Occupancy	No Car	df	Sig.
DHAs	37.8	21.1	32.0	12.2	24.6	194	.000
Zones II$_{1991}$ *(Chi-Square)*	53.7	22.6	38.2	14.0	31.7	194	.000
Zones III$_{1991}$ *(Deviance)*	48.3	22.0	35.9	13.4	29.3	194	.000
Zones I$_{1991}$ *(k-means)*	61.1	22.6	39.2	14.6	33.1	194	.000

N = 9504, Significant at p < 0.001

In Table 6.8, the correlations between the LLTI and deprivation listed using the 2001 census information. All the new zoning systems improve the correlations in comparison to the HAs level. Although, the correlations are very similar at the new aggregations, it can be seen that the zoning *I* provides the highest correlations to all deprivation indicators and Townsend Index scores. On the other hand, the Table 6.4 suggested that best fitted aggregation was the zoning solution *III* and this can be verified in Table 6.9 where the variation within zones in terms of LLTI is highest at zoning *III*. However, the Table 6.9 shows that the zoning solution *I* also can maximise the homogeneity of LLTI and deprivation determinants to a great extent.

Table 6.8: Correlations with LLTI at HA level in England and Wales.

	LLTI – 2001 Census			
	HAs	***k*-means**	**X^2**	**Deviance**
No Car	0.357** Sig. 0	0.426** Sig. 0	0.413** Sig. 0	0.401** Sig. 0
Overcrowding	*-0.370*** *Sig. 0*	*-0.295*** *Sig. 0.003*	*-0.291*** *Sig. 0.003*	*-0.302*** *Sig. 0.002*
Not owner occupied	**0.094** Sig. **0.352**	**0.181** Sig. **0.072**	**0.177** Sig. **0.078**	**0.159** Sig. **0.114**
Unemployment	0.610** Sig. 0	0.648** Sig. 0	0.648** Sig. 0	0.644** Sig. 0
Townsend Index	**0.197*** Sig. **0.050**	0.270** Sig. 0.007	0.266** Sig. 0.007	**0.254*** Sig. **0.011**

N = 100

** Correlation is significant at the 0.01 level (2-tailed)
* Correlation is significant at the 0.05 level (2-tailed)

Table 6.9: Analysis of Variance at HA level in England and Wales.

	LLTI	Overcrowding	Unemployment	Not Owner Occupancy	No Car	df	Sig.
HAs	69.5	65.4	59.6	18.7	40.7	99	.000
Zones $\mathbf{II_{2001}}$ *(Chi-Square)*	94.5	61.9	67.5	19.7	45.3	99	.000
Zones $\mathbf{III_{2001}}$ *(Deviance)*	95.8	62.2	68.5	20.1	45.6	99	.000
Zones $\mathbf{I_{2001}}$ (*k*-means)	93.7	62.2	68.7	20.0	45.8	99	.000

N = 8796, Significant at p < 0.001

6.6 Conclusions

This case study aimed to explore the relationship between deprivation and LLTI indices aggregating 1991 and 2001 ward geographies into fewer zones equivalent to the number of health authorities in respective date. In 2001 census wards, we had to decide which aggregation is most appropriate for our study between the 100 HAs and 50 SHAs. Although, a subjective look at the 2001 dataset indicates HAs as a better geography than SHAs comparing their number of zones and the degrees of freedom, we supported this indication using the statistical information criteria AIC, AICc, and BIC. The use of such criteria in problems concentrated to the choice of scale level provides a valuable utility for deciding the appropriate level of analysis.

Before the extensive investigation in MAUP effects and the relationships between deprivation and LLTI incidents, we utilised a pilot aggregation process exploring the performance of the most demanding objective function (*k*-means) in terms of process need as well as the best possible settings of zone design system for the whole dataset of roughly 10,000 wards in 1991 census. In addition, it has investigated the way the zone design system performs with and without shape constraints. Using the suggested parameters, we produced three new zoning systems per census date focusing on three different objective functions for homogeneity optimisation. To select the new zoning solution with the best goodness of fit, we measure the variation within zones in all three zonings. We found that in 1991 dataset the aggregation using the *k*-means function performs better than the deviance and chi-square while in the 2001 dataset all three objective functions have similar performances.

Exploring the scale effect and correlations between LLTI and deprivation, this study demonstrates two different phenomena in the census datasets. In 1991, the resulted correlations increase as the zoning size increases following the findings from other studies (Gehlke and Biehl, 1934; Openshaw and Taylor, 1979). In contrast, the 2001 correletions have a rather unusual performance with the three determinants (no car, not owner occupied and Townsend Index) providing a decrease of correlation values as the aggregation level becomes higher. Comparing the later finding to Cockings and Martin (2005) study, we identify also to a lesser extent the decrease of correlations in their results. Furthermore, the decrease of correlations seems to take place after the ward level suggesting a dependence of zones on the basic areal units used in the research (Openshaw and Taylor, 1981: p.60). In this study, we believe that an extended research of how determinants correlated to LLTI indices at the highest aggregation levels (e.g. districts) is needed to investigate spatial or statistical factors affecting the correlation coefficients. In addition, the new zoning solutions provide considerable stronger correlations at the level of health authority and they reflect the unemployment and no car characteristics to a greater extend in both studied dates.

In the next Chapter, we will investigate the infant mortality and health inequalities in England and Wales at the period from 1911 to 1971, exploring aggregation effects at regional level (Local Government Districts).

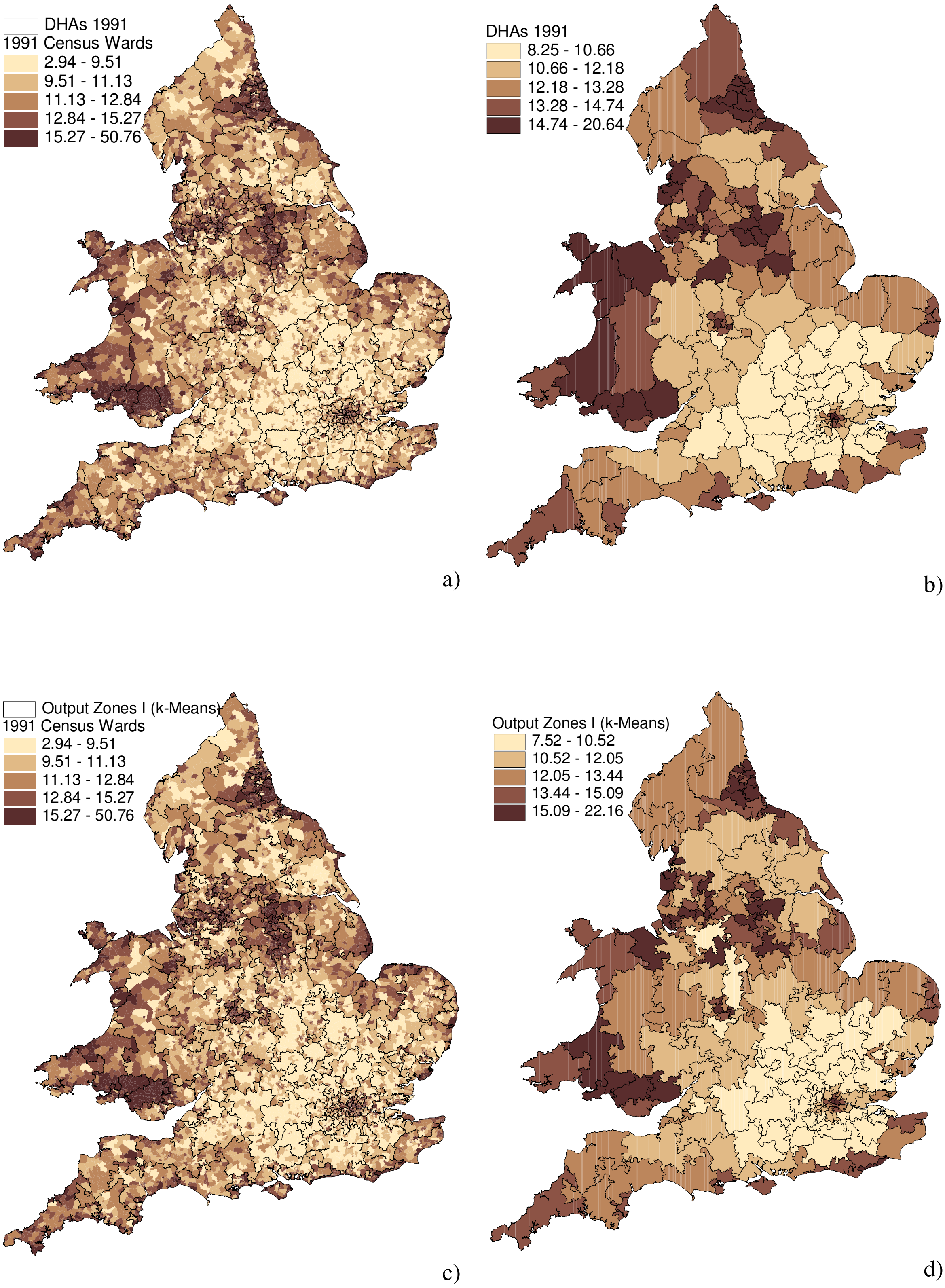


Figure 6.6: Incidence of LLTI (%) in 1991: a) Wards with DHAs' boundaries, b)District Health Authorities (DHAs), c) Wards with new zone boundaries and d) Output Zones I (*k*-Means)

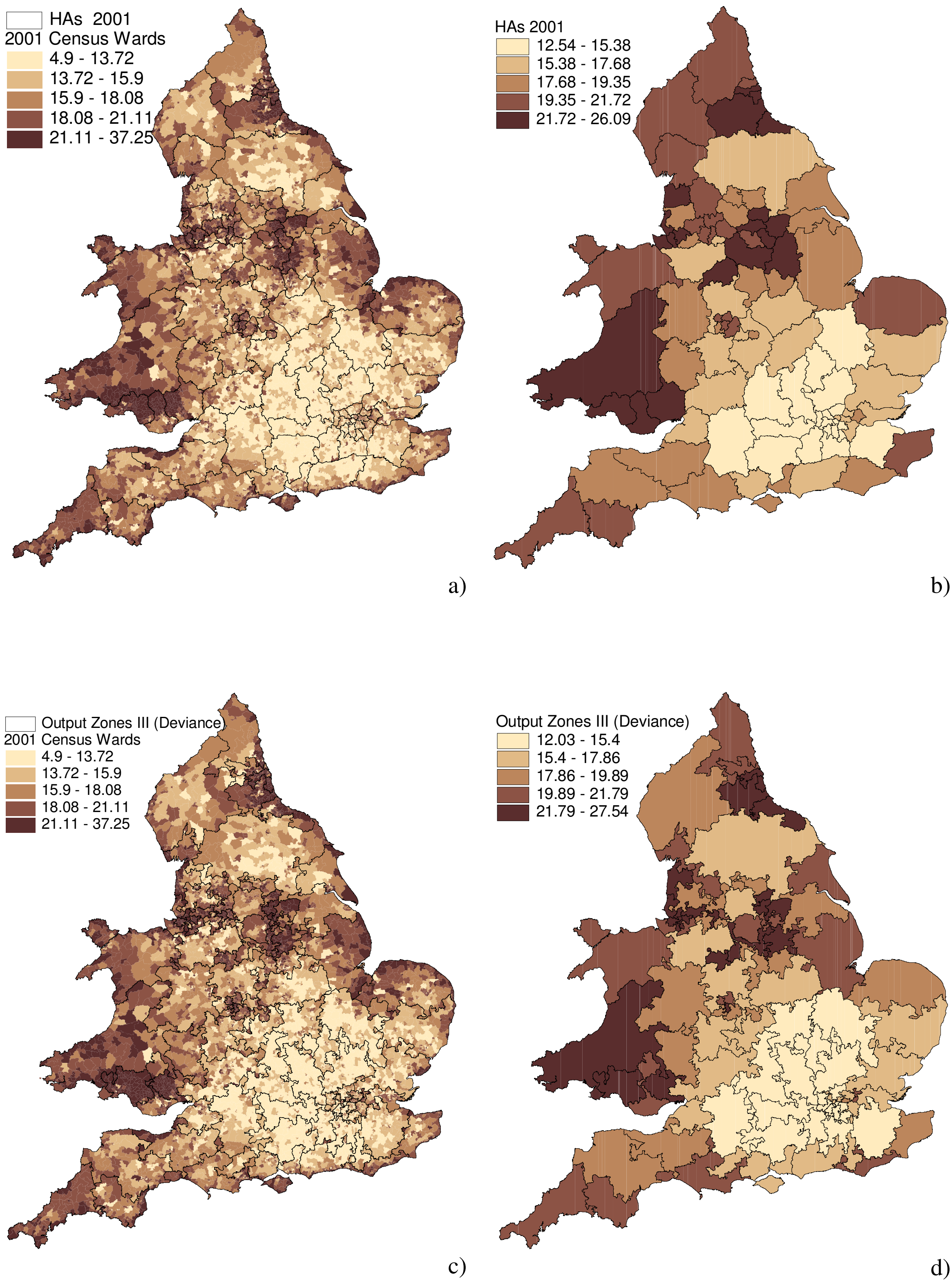


Figure 6.7: Incidence of LLTI (%) in 2001: a) Wards with HAs' boundaries, b) Health Authorities, c) Wards with new zone boundaries and d) Output Zones III (Deviance)

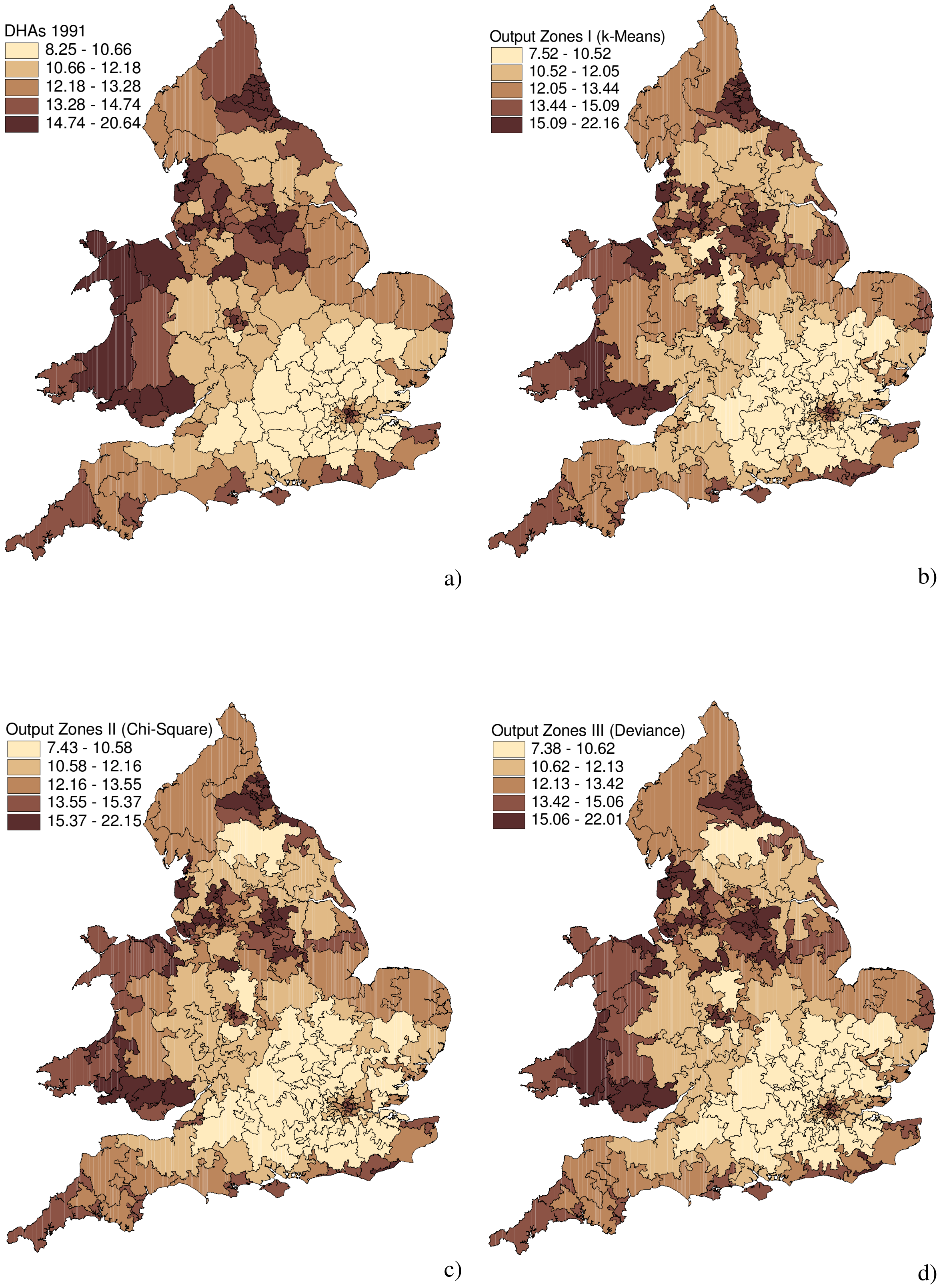


Figure 6.8: Incidence of LLTI (%) in 1991: a) District Health Authorities (DHAs), b) Output Zones I (*k*-Means), c) Output Zones II (Chi-Square) and d) Output Zones III (Deviance)

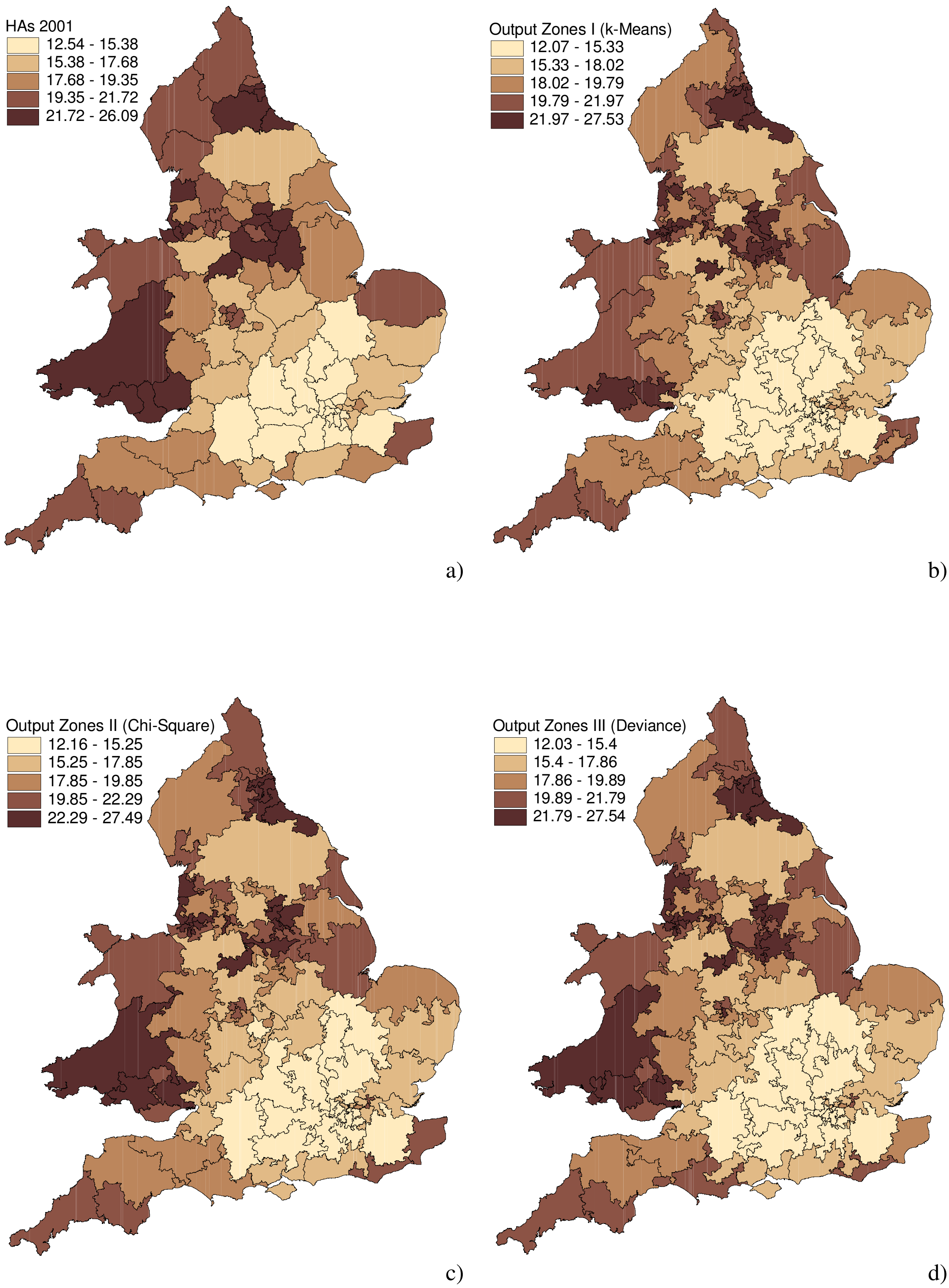


Figure 6.9: Incidence of LLTI (%) in 2001: a) Health Authorities (HAs), b) Output Zones I (*k*-Means), c) Output Zones II (Chi-Square) and d) Output Zones III (Deviance)

CHAPTER 7

Building Upper Level Health Regions

7.1 Introduction

While in the previous Chapter the focus was directed to propose zoning solutions at a middle level exploring the MAUP effects, this chapter examines the use of zone design system in highly aggregated health regions at a county level. At this level the heterogeneity within zones reaches high values providing an unstable environment for analysis. Therefore, the zone design system selected to construct homogeneous zoning solutions targeting similar infant mortality rates and base populations.

An overview of infant mortality in England and Wales over the last century is presented, focusing in findings of the period 1911 -1971 where infant mortality rates have rapidly declined. The aim of this research is to identify regional inequalities in infant mortality. In particular, the changing patterns of infant mortality over the twentieth century are investigated by means of zone design system. As a result, new homogeneous regions are constructed using optimisation criteria based on two variables: infant mortality rate and estimated population. Additionally, we examine how the differences between the best and worst areas during that period have risen, as noted by Gregory et al (2001).

Furthermore, the present study concentrates in inter-regional variations during the period of declining of infant mortality and investigates such variations at a higher geographical scale aggregating local districts into fewer zones. The proposed number of zones is equal to the number of counties in the studied period. The selection of counties as the appropriate level of analysis was based on literature findings in which the majority of them were operated at county level. Thus, it is possible to compare results of this study with existing findings related to infant mortality and health inequalities in

England and Wales. In addition, inequality at the local district and county level is compared using four inequality measures: coefficient of variance, variance of logarithms, Gini coefficient and Theil's Entropy Index (Lee, 1991; Congdon et al., 2001). Finally, the advantages and disadvantages of using various geographies and geographical levels in a study are explored using four different geographies: Local Government Districts, Counties, output zones with homogeneous regions by infant mortality and output zones with homogeneous regions by estimated population.

7.2 Existing findings and aims of case study

Infant mortality is generally accepted as one of the most essential demographic measures of well-being in society. It is a particularly sensitive indicator and remains a useful mean of picturing poverty and inequalities in health, especially in developing countries. Infant mortality is universally defined as the death of infants under one year old and generally is expressed as:

$$IMR = (Infant\ Deaths\ /\ Infant\ Births) \times 1000 \qquad (7.1)$$

In England and Wales, infant mortality has been extensively researched and compared with other socio-economic and environmental factors, providing a series of debates and literature over the last century. For example, the Black Report (Townsend and Davidson, 1992) on health inequalities in the United Kingdom explored a variety of social variables, including infant mortality. It found higher infant mortality rates to be related to indicators such as low income households, overcrowded, unhygienic houses and single parent households, resulting in 39 recommendations of crucial steps by the government to tackle health inequalities. In this section, seminal geographical studies of infant mortality using data series of the last century are discussed highlighting important findings about the relationship of infant mortality with social and environmental factors and the composite structure of inequalities in England and Wales.

One of the most interesting interpretations of infant mortality has been given by Beaver (1973). He suggested that cow's milk accelerated the decline of infant mortality during the eighteenth and nineteenth centuries. The reduction of rate was partly associated with agricultural and commercial developments whilst cows' milk became generally available both in countryside and town. The argument in his paper that animal milk has played a crucial role in the reduction of infant deaths is widely accepted by the research community. In addition, many researchers have also focused on social indicators such as mother's health, household's characteristics and social environment. For instance, the urban environment, income and fertility factors were investigated by Watterson (1988) using the 1911 Census dataset. He supported that infant mortality decline accelerated in urban areas and he noted that the decline was advanced by high or regular income areas, whereas the fertility indicator had only a relatively small effect.

The causes of rapid infant mortality decline in England and Wales during 1861 – 1921 have been discussed by Woods et al (1988; 1989). They explored the variations of infant mortality using the registration districts of 1911 and factors such as type of environment and social class. Their research provided some evidence supporting the decline of infant deaths and suggesting that the rapid decline of infant mortality was related to social classes. Moreover, their findings in urban places where the infant mortality was highest indicated a precipitous decline of infant deaths. During the 1890s, where the weather conditions especially in summer season was favourable for epidemic diarrhoea the urban areas with pure sanitary provision countered a high infant mortality rate. According to their research, when they discounted the cases of diarrhoeal diseases, infant mortality appeared to have declined continuously from the late 1880s.

In contrast, the existence of a relationship between infant deaths and social class has been identified by many researches, such as Haines (1995). His research focused on the relationship of infant mortality with four indicators: income, social class, rural/urban and occupation in household. Haines examined three types of social classifications such as the Registrar General's social class grouping used in the 1951 Census (proposed by Stevenson (1928)) and his findings suggest that the influence of social class in the decline of infant mortality was positive. Furthermore, he noted that infant mortality declined more rapidly among the higher social classes during the period of 1890s and 1900s and also that income and urbanisation were more successful in explaining the level of infant deaths as opposed to occupational indicators. In addition, he suggested that even if the absolute differences in infant mortality were smaller, after about 1911 relative inequality persisted supporting the findings by Lee (1991). Lee's research was focused on the regional inequalities and he suggested that during the period 1881 until 1931 there was a continued increase in inequality as infant mortality in the healthier regions fell more rapidly. His analysis was based on four different measures of inequality: coefficient of variation, variance of logarithms, Gini coefficient and Theil's Entropy Index. Both Lee (1991) and Woods (1997; 2000) suggested that there were inter-regional variations during the decline of infant mortality. A recent study by Congdon et al (2001) based on regional district level uses a variety of inequality measures identify variations in infant mortality in relation to differences between urban and rural areas.

As we have seen above infant mortality provides a useful indicator of measuring health inequalities and well-being in society. Consequently, the aim of this case study is to introduce the zone design approach as a valuable tool of analyse health data at county and local district level. Additionally, the aggregation effects of administrative and output zones are explored using the inequality measures suggested by Lee (1991). As a result, the existing variations between and within regions are further investigated providing a valuable insight of infant mortality and health inequalities in the period examined here.

7.3 Definition of Spatial Units and Zone Design rules

7.3.1 Data issues and spatial unit choice

The selection of datasets was based on the findings of the literature review and the key dates of infant mortality and inequality in England and Wales for the period after 1900s as presented by Beaver (1973) and Lee (1991). In Figure 7.1 Beaver's diagram is illustrated with additional representative dates. It shows the post 1900s rate decline of infant mortality (red line) and additional information such as the periods of First and Second World War (grey bars) and the selected datasets dates (dashed lines). Ideally, this research would consider continuous time series of datasets but the information is only available through the Great Britain Historical Geographical Information System (Gregory and Southall, 1998). Their sources of producing such datasets were mainly census records carried out every ten years. The two World Wars interrupt the collection of many statistical series and there was no 1941 census. According to these restrictions, the current study is concentrated in four dates: 1911, 1928, 1951 and 1971 using Local Government Districts (LGDs) which consisted of county, municipal and metropolitan boroughs and urban and rural districts. As Gregory et al (1991) mentioned the LGDs were subdivisions of Registration Districts (RDs) and the urban areas were designated as boroughs or urban districts while the rest of RDs became rural districts. The number of LGDs varies according to the year, starting with about 2060 LGDs in 1911 and declining to about 1450 LGDs in 1971.

Furthermore, the datasets cover the 1881 – 1931 and 1931 – 1971 phases of inequality change as reported by Lee's (1991) findings. He supported that during the period of 1881 to 1931 inequality increased as result of the faster infant mortality decrease in healthier regions, while in the period of 1931 to 1971 the continuous decline of mortality rates converged the decrease of inequality.

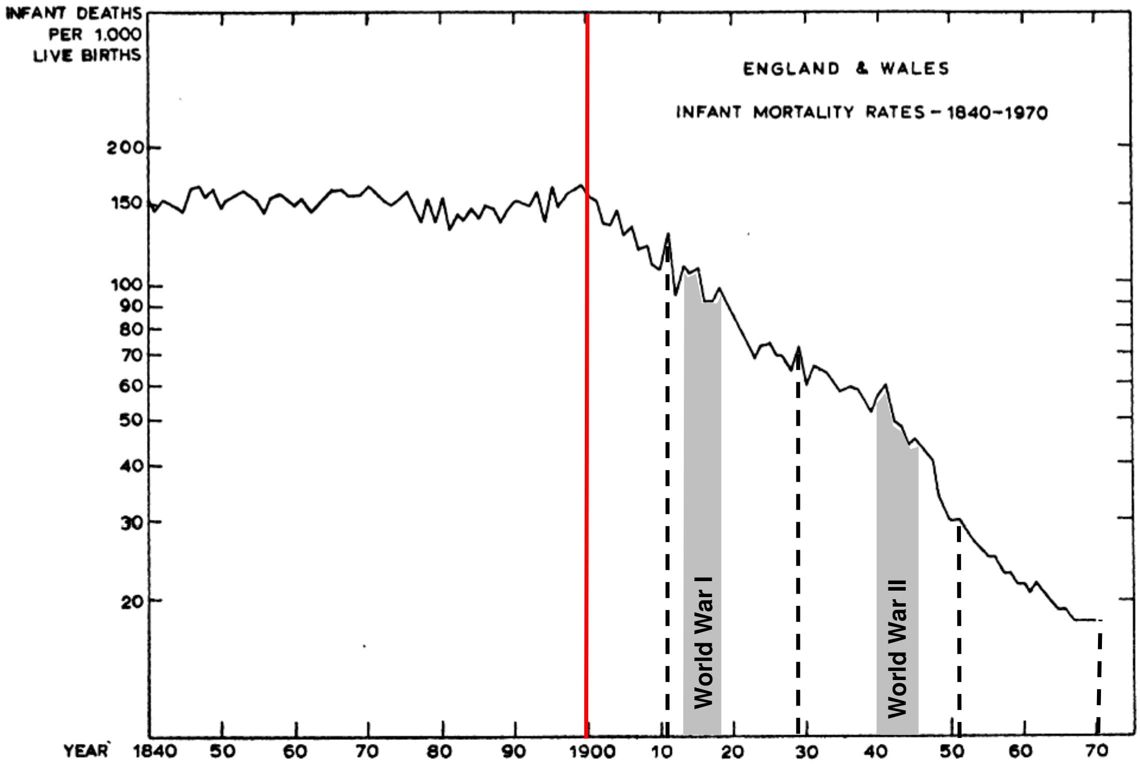


Figure 7.1: Infant Mortality Rates during 1840-1970 in England and Wales.

(Source: Beaver (1973): p. 244)

In this research, each dataset includes three variables: infant deaths under one year old, infant births and estimated population. A first observation of datasets indicates the appearance of some small numbers in infant deaths and births in the 1911 and 1928 datasets. It is suspected that this major issue of small number effects in both years is related to the completeness of recording and registration of births in the UK (Woods, 1997; 2000). The use of such data can potentially mislead a researcher overstating the extent of mortality rate. In addition, extreme mortality rates may reflect small populations at risk and important spatial patterns could be hidden using simple mapping techniques (Besag and Newell, 1991). The Figure 7.2.a and 7.2.b show the existence of high infant mortality rates at the LRD level. In addition, the small number effects in these datasets have been tackled using a dissolve method concentrating in areas with birth population less than 10 infants. In the current study, the number of problematic areas was about ten areas for each dataset (1911 and 1928). In detail, the approach of merging a small populated area with a large populated area was selected as the most appropriate and less time consuming. This approach selects the areas with population less than 10 infants and groups them with large adjacent zones to meet the maximum threshold of population. As a result the whole dataset is free of small numbers and problems related to them. Furthermore, the availability of datasets in Arc Info export format (e00) at first glance was ideal for compatibility reasons with zone design system. The transformation of export files to coverages, even with extremely high fuzzy tolerances, created coverages with numerous sliver polygons and additional editing was essential for the final datasets. Once the data was free of the small number effects and the geometrical errors, it was statistically accepted the calculation of infant mortality rates using the formula 7.1.

During the period from 1911 to 1928, the population characteristics of the UK dramatically changed. During that period, the 36 million estimated people increased into 49 million people in the UK with a decrease of child births from 870,000 to 780,000 and a rapid decrease of infant deaths from 115,000 to 15,000. As it is illustrated in Figure 7.3, the northern counties of the UK suffered most from infant mortality, suggesting a north-south divide. In addition, the distance of a county from the centre of

London appeared to reflect the mortality rates with the lowest values near London while highest values scored by the northern and western counties.

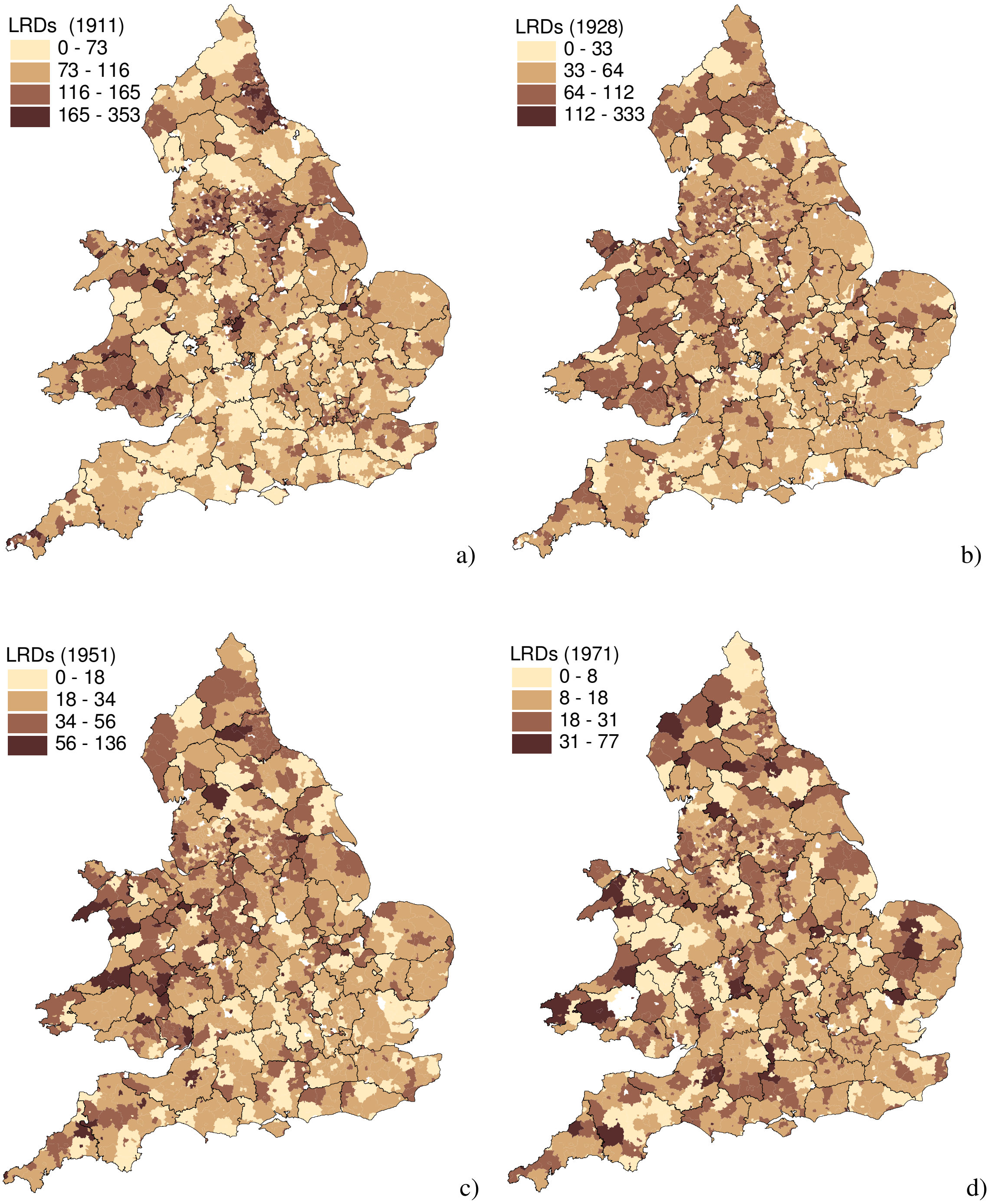


Figure 7.2: The infant mortality rates using Local Government Districts (LGDs) in: a) 1911, b)1928, c)1951 and d)1971.

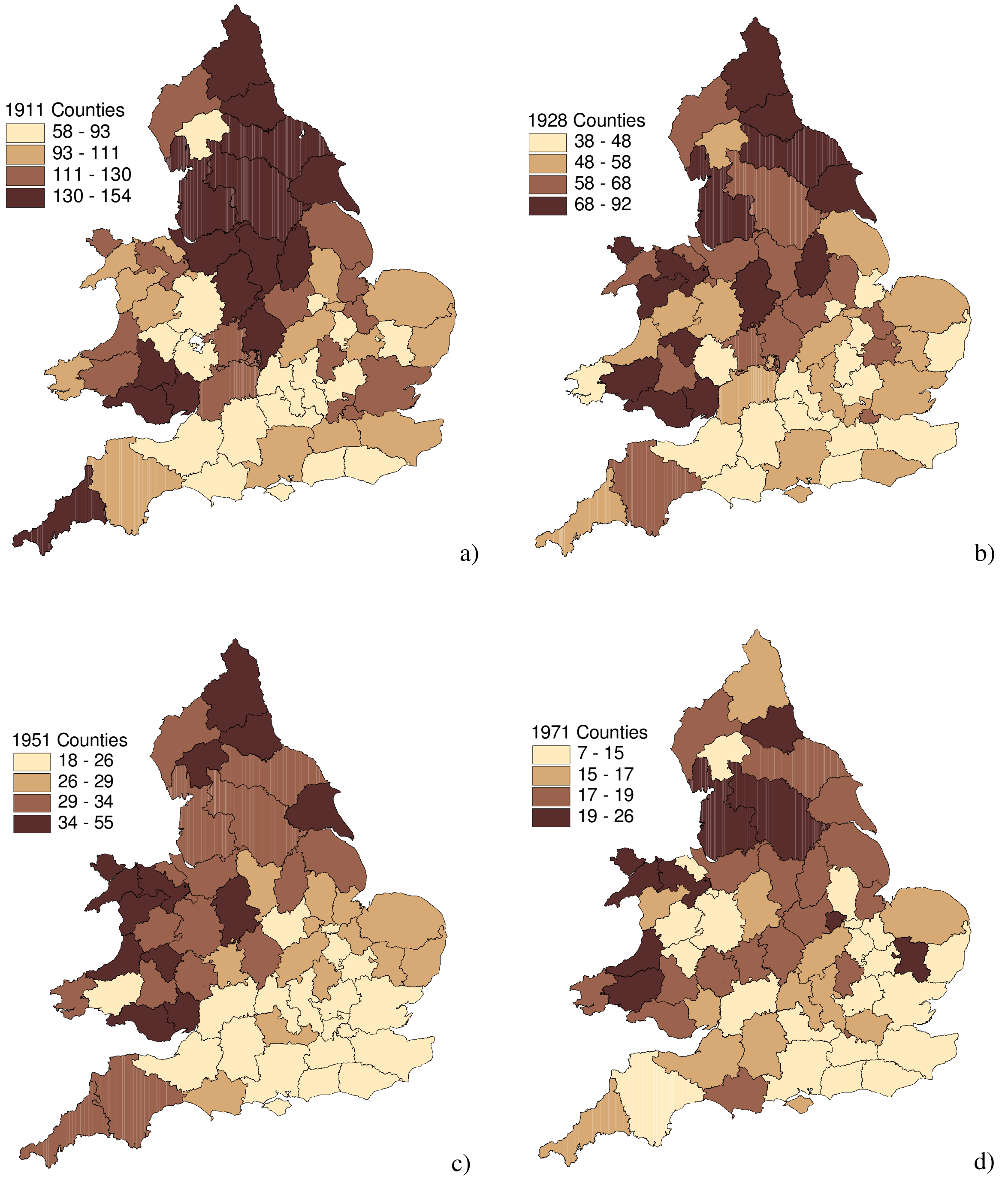


Figure 7.3: The infant mortality rates (‰) using the County level in: a) 1911, b) 1928, c) 1951 and d) 1971.

Comparing Figures 7.2 and 7.3, there is a strong heterogeneity of infant mortality rates within counties. While Figure 7.3 shows counties with high infant mortality rates, in Figure 7.2 the same counties illustrate high variations between local districts in each county. For example, Durham and Cardiff areas are well-known for their high infant mortality rates and all together are extremely heterogeneous. As Martin and Cockings (2005) suggest, homogeneity within health related zones is essential for any exploratory analysis of spatial patterns of disease. Therefore, the use of zone design system is suggested here for constructing homogeneous zones in terms of similar infant mortality rates and estimated population. Thus, the variation between zones is maximised and the within mortality rates and population differences are minimised.

7.3.2 Defining zone design rules

In a zone design context, it is important for the user to identify the rules that the system should follow and specify the criteria of the aggregation process. The zone design system in this study aggregated approximately 2000 Local Government Districts producing 62 output zones (equivalent to 62 counties in England and Wales). The system constructed homogeneous output geographies and for this reason the use of a homogeneity function (equation 4.21) was essential. The aggregation process started

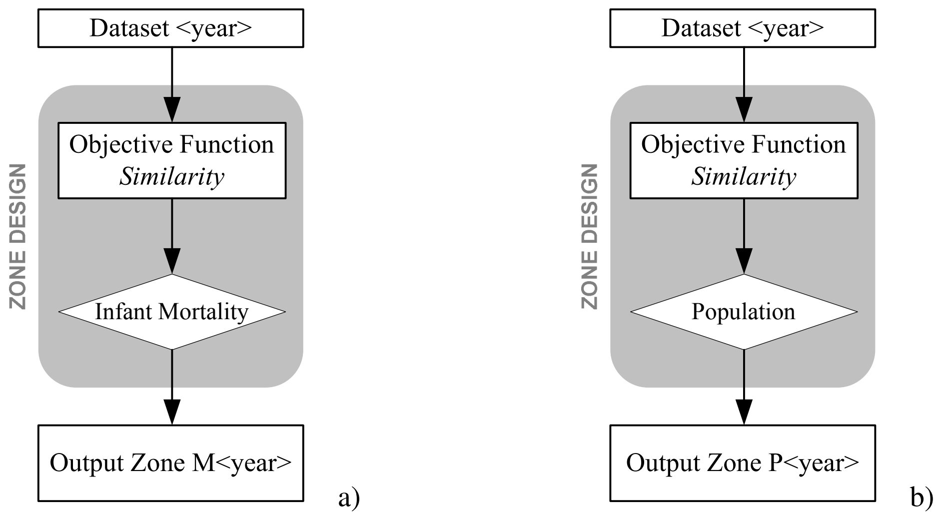


Figure 7.4: Generic diagram of zone design system using a) similar rates of infant mortality and b) similar population numbers (where <year> is 1911, 1928, 1951 and 1971).

each run from a predefined number of zones equivalent to the 62 counties. It was executed for 10 times and at each run the criterion was optimised to a maximum of 300 iterations or for 20 idle iterations on the homogeneity score. The above process was repeated twice for each study year, constructing output regions with similar infant mortality rates (Figure 7.4.a) and output regions with similar population numbers (Figure 7.4.b).

The aggregation of LGDs into 62 output homogeneous zones using the infant mortality rates groups the similar rates in each zone, highlighting the problematic areas. Figure 7.5 illustrates the infant mortality rates in 1911, 1928, 1951 and 1971 respectively at the output zone level. Generally speaking, there is a clear decrease of mortality rates throughout that period. Mapping the infant mortality rates using four categories (Figure 7.5.a), Durham, Manchester, Liverpool, Sheffield and Birmingham areas appeared to have the highest infant fatalities reaching the rate of 174 infant deaths per thousand births. In 1928 (Figure 7.5.b), the output zone patterns change dramatically, because of the decline of infant mortality in the whole UK, thus showing only Durham, Manchester and Liverpool with more than 75‰ infant loses, while the majority of northern regions and Wales have rates above the national average. In 1951, the concentration of infant deaths around Durham remained one of the highest rates nationally, in addition to two equally problematic areas in Wales: Cardiff and Caernarfon (Figure 7.5.c). However, infant mortality rates remained high in Wales and northern England, highlighting once again a north-south divide. The output zones of 1971 exhibit the smallest mortality rates ranging between 6‰ and 29‰. While the highest measures appeared more spread than previous years, including the aforementioned problematic areas.

In addition, groupings of LGDs targeting homogeneous zones in terms of similar population numbers are presented in Figure 7.6, illustrating the infant mortality rates for the new geographies. The aggregation of similar LGD populations is important for this study as the resulting zones reflect the density of population representing areas with rural or urban characteristics. In more detail, Figure 7.6.a shows Durham, Manchester, Liverpool, Sheffield and Birmingham areas scoring high mortality rates. Cardiff area appears to come across with equivalent high rates highlighting a potential mortality pattern in south Wales. In 1928, the new zones (Figure 7.6.b) present Durham, Cardiff, Manchester and Liverpool experiencing many infant deaths while in later years (1951 and 1971) the areas with severe infant mortality rates are similar to those have been created using similar mortality rates (Figure .7.5).

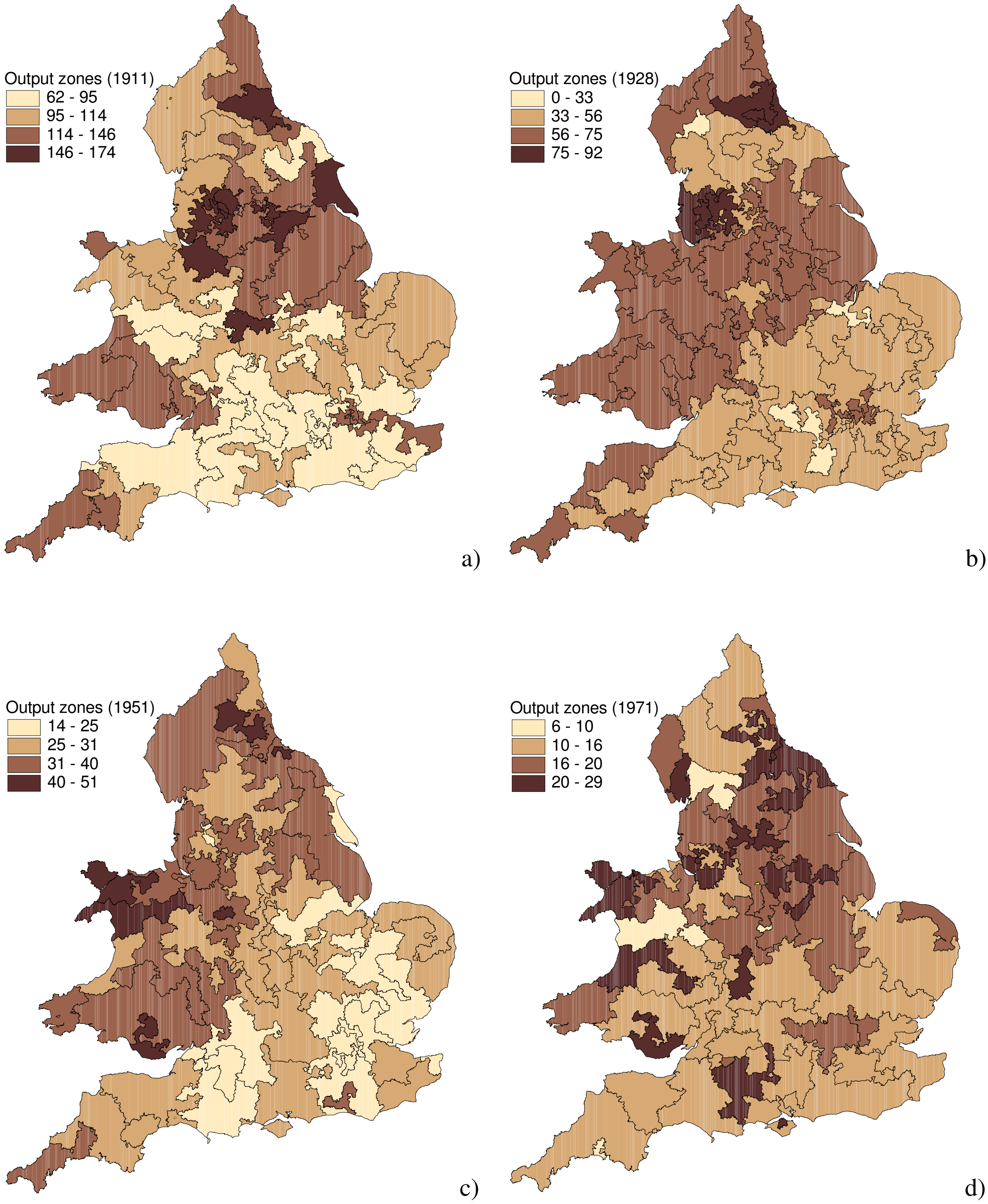


Figure 7.5: The infant mortality rates using the output geographies based on the similarity of infant mortality rates in a) 1911, b) 1928, c) 1951 and d) 1971.

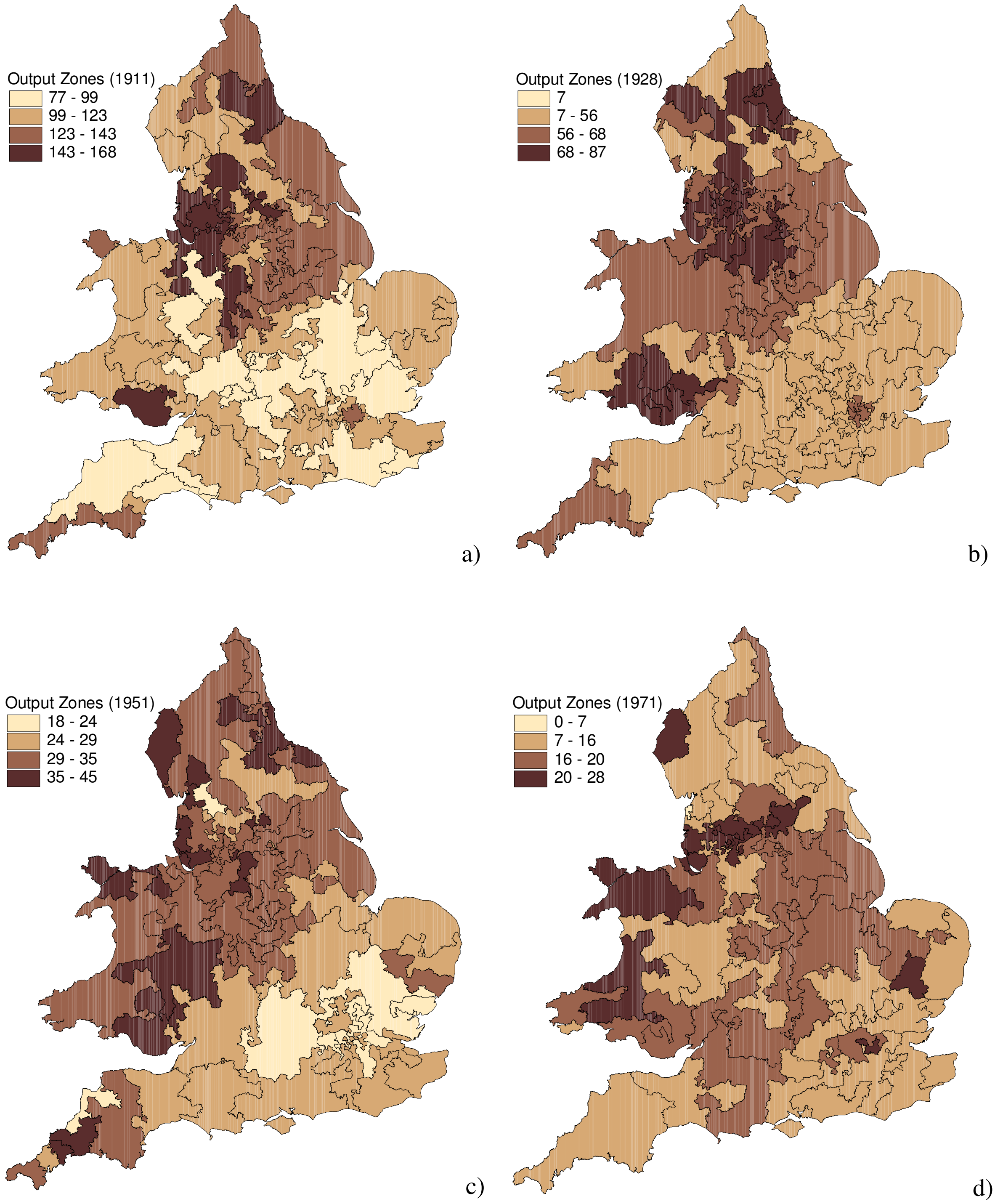


Figure 7.6: The infant mortality rates using the output geographies based on the similarity of population numbers in a) 1911, b) 1928, c) 1951 and d) 1971.

As a result, these eight new geographies suggest the existence of extreme infant mortality rates during the studied period in Durham, Cardiff, Manchester and Liverpool. The zones of similar population numbers highlighted the urban areas as the most likely to face high rates of infant mortality. It is also known that the areas around these four cities were heavily industrialised with people working at coal mines and generally poor life conditions. While the mapping of new homogeneous zones provides a better look at infant deaths in the UK compared to administrative counties, it is important for further investigation to be completed providing possible explanations about the health inequalities and their relation to infant mortality.

7.4 Methods of measuring changes in Infant Mortality

The exploration of inequality changes in infant mortality in the levels of: LGD, county, output region, based on similar infant mortality rates and output region with similar population values are measured by means of four inequality measures: coefficient of variation, variance of logarithms, Gini coefficient and Theil's entropy index (Allison, 1978; Lee, 1991; Cowell, 1995; Congdon et al, 2001). The mathematical definition and some strengths or even limitations of the above inequality measures is important to be mentioned at this point.

First, the coefficient of variation is sensitive to change throughout a given distribution and it is calculated from the standard deviation of the y_i, divided by the $\bar{y}$ mean of y_i's. Mathematically, it is expressed as:

$$\frac{\sqrt{\frac{1}{n}\sum\left(y_i-\bar{y}\right)^2}}{\bar{y}} \qquad (7.2)$$

where, y_i is the value for observation *i* and $\bar{y}$ is the mean of the *y's*.

The variance of logarithms is sensitive to changes at the lower ranges of a scale and it is formulated as:

$$\frac{\sum\left(\log\left(y_i/y*\right)\right)^2}{n} \qquad (7.3)$$

where, $y*$ is the geometric mean of the y's.

Alternatively, it could be expressed as:

$$\frac{\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2}{n} \qquad (7.4)$$

where $x_i = \log(y_i)$ and $\bar{x}$ denotes the mean of all x_i.

The Gini coefficient was developed as a summary measure of income inequality in society by Corrado Gini and it is associated with the plot of wealth concentration introduced a few years earlier by Max Lorenz (Lorenz, 1905). The Gini coefficient is particularly responsive to transfers affecting the middle values of a distribution (Atkinson, 1980) and it is defined as:

$$\frac{\sum_{i=1}^{n}\sum_{j=1}^{n}\left|y_i-y_j\right|}{n^2\bar{y}} \qquad (7.5)$$

where, y_i and y_j are the *i* and *j* observations, respectively.

As Lee mentioned, "Theil's entropy index is responsive to changes throughout the distribution and measures deviations from a state of equality in which each variable in the distribution has a share equivalent to its relative size" (Lee 1991: p60). The Theil's entropy index can be expressed as:

$$\frac{\sum_{i=1}^{n}y_i\log\left(y_i/\bar{y}\right)}{n\bar{y}} \qquad (7.6)$$

Although, the aforementioned measures are concerned with different changes throughout the distribution of a dataset, they are expected to provide similar findings especially when they measure the variation between the output zones.

7.5 Results of the analysis

The above inequality measures were implemented here as separate spatial tools using the ArcView GIS programming language, known as Avenue. The resulting algorithms are able to calculate the four inequality measures at both national and local level or in other words to measure the between and within variation of infant mortality rates at any given aggregation. However, the LGDs provide additional information concerning the rural and urban districts, therefore the measures should reflect the inequalities more accurately than previous studies that concentrated at a county level (Lee 1991).

In Table 7.1, the four inequality measures were applied to the LGD level using the urban/rural classification. The Coefficient of Variation measure shows increase of between LGDs inequalities from 0.4026 in 1911 increased to 0.6399 in 1971. In general this increase is recorded from the Gini coefficient and Theil's entropy index while the variance of logarithms suggests an increase until 1928 to a value of 0.0546, while in later years it suggests a decrease of inequality. The inequality measures at LDGs level using the classification of rural and urban areas are also listed illustrating the changes of inequality. All the measures except the variance of logarithms show in 1911 higher variation in urban areas compared to rural areas. Gradually, we see an increase of inequalities in both urban and rural areas with higher rate in rural areas. As a result, in 1971 the rural districts appeared with higher inequalities than the urban districts. However, this is not the case for the variance of logarithms measure as the urban districts retained high inequality values during the period of study.

In conclusion, table 7.1 suggests two main findings. Inequalities at LDG level increased considerably throughout the studied period for both urban and rural districts. On the other hand, a higher variation of infant mortality rates is measured in rural districts than urban districts. This is reflected in all four inequality measures showing in 1971 high inequalities in rural districts.

Table 7.1: The between areas four inequality measures using the urban/rural classification.

Inequality Measure		1911	1928	1951	1971
Coefficient of Variation	*All*	0.4026	0.5266	0.5329	0.6399
	Urban	0.3909	0.5475	0.5504	0.6364
	Rural	0.3675	0.4502	0.4950	0.6405
Variance of Logarithms	*All*	0.0378	0.0546	0.0350	0.0325
	Urban	0.0478	0.0643	0.0410	0.0333
	Rural	0.0277	0.0450	0.0290	0.0318
Gini Coefficient	*All*	0.1989	0.2263	0.2461	0.2618
	Urban	0.2009	0.2407	0.2434	0.2590
	Rural	0.1969	0.2119	0.2489	0.2646
Theil's entropy index	*All*	0.0312	0.0410	0.0437	0.0500
	Urban	0.0288	0.0432	0.0430	0.0494
	Rural	0.0210	0.0274	0.0428	0.0454

A further exploration of inequalities using the four inequality measures is provided investigating the zone effect at the county level. The homogeneous zones in terms of infant mortality rates and similar population numbers are compared to the 62 administrative counties. In Table 7.2, the variation of infant mortality rates between the counties shows an increase of inequalities until 1951, while in 1971 the measures suggest a decrease of inequalities. Again the variance of Logarithms provides different results suggesting an increase in 1971. However, using the homogeneous zones the measures reveal a story. Both aggregations illustrate high inequality values in 1928 and 1971 while in 1951 a slight decrease of inequality is recorded. The slight change is recorded to all the inequality measures providing a different perspective to the one presented by the county level. In summary, two main findings are derived from table 7.2. The county areas show a continuous increase of inequality measures until 1951 with a drop in 1971 and secondly, the use of homogeneous zones suggests a slight drop of inequality values in 1951.

Exploring the inequalities within the counties and zones, in table 7.3 the inequality measures at county level show a continuous increase of inequalities during the studied period. However, introducing the homogeneous zones it is obvious that they score lower inequality values than the counties, with better homogeneity of infant mortality rates in the zones constructed targeting similar population values. Furthermore, the homogeneous zones suggest increase of inequalities throughout the period while in 1951 a slight decline appears. Comparing Table 7.3 with Table 7.2, we can conclude that in 1951 there was a slight decrease of inequalities in England and Wales as it is supported by both tables and the four inequality measures.

Table 7.2: The between areas four inequality measures at county and output zones levels.

Inequality Measure		1911	1928	1951	1971
Coefficient of Variation	*County*	0.1952	0.2053	0.2180	0.1920
	Inf. Zone	0.2397	0.2874	0.2742	0.3660
	Pop. Zone	0.2252	0.2499	0.1893	0.2885
Variance of Logarithms	*County*	0.0078	0.0076	0.0079	0.0087
	Inf. Zone	0.0103	0.0249	0.0135	0.0231
	Pop. Zone	0.0089	0.0218	0.0070	0.0113
Gini Coefficient	*County*	0.1116	0.1155	0.1163	0.1052
	Inf. Zone	0.1322	0.1564	0.1480	0.1738
	Pop. Zone	0.1241	0.1302	0.1075	0.1512
Theil's entropy index	*County*	0.0084	0.0089	0.0097	0.0084
	Inf. Zone	0.0121	0.0149	0.0156	0.0257
	Pop. Zone	0.0106	0.0145	0.0078	0.0106

Table 7.3: The within areas four inequality measures at county and output zones levels.

Inequality Measure		1911	1928	1951	1971
Coefficient of Variation	*County*	0.3638	0.4867	0.4996	0.6312
	Inf. Zone	0.3188	0.4713	0.4296	0.5918
	Pop. Zone	0.3272	0.4020	0.3917	0.5378
Variance of Logarithms	*County*	0.0207	0.0221	0.0190	0.0146
	Inf. Zone	0.0162	0.0194	0.0147	0.0131
	Pop. Zone	0.0186	0.0185	0.0141	0.0136
Gini Coefficient	*County*	0.1823	0.2095	0.2194	0.2246
	Inf. Zone	0.1614	0.1955	0.1864	0.2072
	Pop. Zone	0.1685	0.1808	0.1796	0.2183
Theil's entropy index	*County*	0.0211	0.0315	0.0351	0.0423
	Inf. Zone	0.0164	0.0273	0.0243	0.0370
	Pop. Zone	0.0183	0.0254	0.0231	0.0404

7.6 Discussion and Conclusions

Our analysis has utilised zone design system to produce homogeneous aggregated zones at county level for an analysis of infant mortality and health inequalities during the period 1911 to 1971. Our finding, mapping the LGDs, counties and homogeneous zones, suggest that in Durham, Cardiff, Manchester and Liverpool infant mortality rates have been considerably high during this period. This is partly explained by the heavy industrialisation of those areas as the majority of national productivity and population was concentrated in these areas according the literature findings (Freeman, 1986). The mapping of counties and zones suggests a north-south division as noted by Gregory et al (2001). However, more interesting results from our inequality measures are as follows. At LGD level we have confirmed that health inequalities increased throughout the studied period. Analysing the LGDs rural-urban classification we found that both rural and urban inequalities had a significant increase while the rural areas increased at a higher rate taking the lead from urban areas in 1971.

At county level, our research explored the between and within variations of infant mortality at three aggregations. Firstly the administrative counties show an increase of inequalities during the period 1911 to 1951, while in 1971 the variation of infant mortality decreases. Investigating the homogeneous zones, our findings suggest that the dropping point of inequalities in England and Wales took place in 1951 and not in 1971 as suggested by Lee (1991). The different dropping points reported in Lee's and our results can be explained by the modifiability of boundaries (MAUP) and we suggest that further research is needed to understand the mechanisms behind the inequalities in England and Wales during the studied period. Certainly our methods and results are influenced by the original LGDs and the information attached on them as there are estimations of population in LCDs (Gregory et al, 2001) as well as spatial problems such as urban LGDs neighboured by rural LGDs shaping a donut formation. However, we believe that the above findings support the introduction of homogeneous zones in such research, highlighting the importance of zone design system as a supportive tool for explaining the MAUP effects in historical-medical datasets.

CHAPTER 8

Discussion and Conclusions

8.1 Introduction

This chapter summarises the results of this thesis bringing together theory and practice as presented in earlier chapters. First, it highlights what has been discovered through the evaluation of a zone design system in three case studies. The application of a zone design methodology at various scale levels illustrates its potential within health related geographical research. Second, this chapter presents a number of suggestions for tackling unsolved theoretical issues and places the suggested methodologies in the context of a wider health related perspective. In addition, it examines what has been achieved with respect to the aims and objectives of the research as discussed in Chapter 1. Finally, this chapter presents a discussion of further research possibilities, in order to identify where improvement in the methodologies of this research can be made. The chapter concludes with suggestions on how the research presented here can be developed further, by incorporating ideas such as the multi-level concept.

8.2 Summary of Case Studies

The three case studies presented here were purposely chosen to explore certain methodological concerns in the health research area. According to the latest NHS plan (DoH, 2000), the improvement of health figures such as children accidents and inequalities should be archived by well grounded methodologies. In this context, Chapters 5 and 6 investigated the socioeconomic census characteristics in relation to the traffic accidents experienced by children in Tyne and Wear and the LLTI in the whole of England and Wales respectively. Furthermore, Chapter 7 explored health inequalities in the England and Wales during the period 1911 to 1971, using historical data related to infant mortality incidents.

In this thesis, each case study targets a representative level of geographical/ administrative scale. Thus, aggregation issues as discussed in Chapter 3 were explored in different scales highlighting the importance of a zone design approach in health research. Throughout the case studies, the smallest possible basic areal units were aggregated up to the scale level where existing administrative areas are used by the majority of health researchers. This approach was adopted here to demonstrate the inadequacy of administrative geographies in health research, as each one of these administrative geographies has been constructed with specific objectives. A clear example is the 1991 EDs that were designed for collection of census information, but they have been repeatedly suggested for allocation of health resources by health researchers (Crayford et al., 1995; Mackenzie et al., 1998). The misuse of areal units was highlighted throughout the three case studies showing the advantages of methodologies adopting a zone design approach.

What became clear in all three case studies is that the methods introduced in Chapter 3 can partly cope with the scale, aggregation and topological problems faced in complex health environments. On the other hand, zone design can suggest solutions and assist in exploring health related research questions by investigating possibilities that are infeasible by existing GIS tools. Handling the changeability of health administrative areas in a zone design context can provide a methodological framework for explaining relationships between socioeconomic and health factors and supporting research related to the improvement of health services and allocation of resources. The case studies demonstrate the ability of zone design to effectively tackle the MAUP effects of health related datasets, while providing extensive exploration of feasible solutions.

In Chapter 5, we provided a complete methodology for investigating traffic incidents experienced by children. Using the information criteria estimators, the proposed method identifies the most statistical informative aggregation level, providing a valuable approach for measuring the goodness of fit and also taking into account the degrees of freedom of each model. The detection of the statistically most informative level can be extremely valuable in the health research community as it can be compared with alternative approaches, such as subjectively defined administrative areas. For further exploration of the suggested aggregation level, the relationships between children accidents and deprivation determinants studied by using the zone design system to produce health related zoning solutions. Although the introduced method investigated in depth the MAUP effects, further analysis is proposed, by including natural and social barriers as zoning constraints.

In Chapter 6, we used the zone design system to construct LLTI related health zones for the whole of England and Wales by evaluating different homogeneity objective functions. The comparison of the new output zones to the existing Health Authorities revealed two different phenomena in the 1991 and 2001 datasets. In the 1991 dataset the correlation coefficients between LLTI and deprivation determinants increase as the zone size increases, in accordance with the literature findings (Cockings and Martin, 2005; Openshaw and Taylor, 1979). However, in the 2001 dataset the correlations decrease as long as the scale increases. We suggest that the latter effect is the product of the available deprivation determinants which are not related to the LLTI incidents to the same degree as in the 1991 dataset. However, the new zoning solutions provided a considerable improvement of within zones homogeneity in comparison to health administrative areas. Therefore, we suggest that the zone design methodology can be a useful tool in the hands of health researchers for investigating the scale and zoning effects as well as for controlling the variance within and between zones.

In the last case study (Chapter 7), we investigated infant mortality and health inequalities in England and Wales during the period 1911–1971. The use of the zone design system for constructing homogeneous zones in terms of infant mortality rates and base population illustrated a clear north-south division, as noted by Gregory et al. (2001). The challenge of a time series dataset was accomplished with interesting findings concerning the dropping point of inequality during the studied period. While Lee (1991) supports that inequality started to decrease in England and Wales in 1971, our findings suggest that this happened earlier, in 1951. However, the different dropping point of inequality can be explained by the modifiability of boundaries of areal units. Using the methodology employed in Chapter 7, the zone design system was expanded to investigate changes of infant mortality throughout the studied period. Thus, health researchers can highlight possible temporal patterns in a study area using historical health datasets.

The zone design system is capable of producing numerous zoning solutions, while satisfying a range of prespecified criteria. However, researchers should further investigate such new aggregation outputs using existing statistical tools. For example, all three case studies were evaluated here using model-comparison indicators (AIC, deviance, chi-square) and statistical measures (ANOVA, coefficient of variation, variance of logarithms, Gini coefficient and Theil’s entropy index). The combination of zone design and available statistical measures can highlight patterns and relationships between health incidents and socio-economic indicators providing a valuable consultant for health policy and allocation of health resources.

8.3 Methodological progress on aggregation issues

The main objective of this thesis aimed to suggest new methodologies for supporting the health professionals and researchers in developing suitable health policies using statistical and spatial techniques in a zone design context. To achieve this objective, the health related and aggregation issues (discussed in Chapter 2 and 3) were demonstrated utilising zone design methods and statistical measures by means of graph theory and object-oriented programming. These new techniques were brought together in a proposed A2Z zone design system extending the features of dated aggregation systems such as ZDES (Alvanides and Openshaw, 1999) and AZM (Martin et al., 2001).

8.3.1 Resolving topological issues

The research presented here has used graph theory in order to formally define spatial issues related to the complexity and changeability of boundaries. In Chapters 2 and 3 the adjacency and connectivity between areal units was highlighted as an important factor when designing new geographies. The majority of existing approaches do not support complicated topological structures, while exceptions such as the methodology developed by Coombes (2000) require laborious preparation of datasets. Although, the DoH in the UK has repeatedly requested the use of boundary limitations throughout the design of health administrative areas (DoH, 2001a; DoH, 2001b) most health organisations perform this task manually using subjective information and criteria.

In order to provide an advanced zone design system capable of utilising boundary constraints, we identified three special topological issues that emerged between areal units. An areal unit may consist of disconnected spatial entities such as part of the mainland together with an island. In addition, areal units may be isolated in terms of their topological adjacency characteristics, but they can be connected by other means, like a bridge connecting two areal units separated by a river. Both these cases can affect the aggregation process as discussed in Chapter 4 and specific methods were developed as part of zone design system to tackle these issues. Moreover, in this study an advanced technique for controlling weighted boundaries was developed expanding the capabilities of the zone design method. Solving all three issues, the proposed aggregation system can perform more complicated aggregation tasks handling barriers and connections between areal units which until today were not applicable.

8.3.2 Identifying the suitable scale level

One of the most challenging issues addressed in Chapter 4 is the selection of an appropriate scale level using advanced statistical measures. The selection of the appropriate aggregation scale during the zone design process was encountered as the weak side of previous zone design systems. In this thesis we strongly support the use of statistical information criteria like the AIC to investigate which scale is statistically best fitted. Although the information criteria were successfully applied in Chapters 5 and 6 investigating the most informative aggregation level in both studies, we support that the information criteria should be part of an overall analysis of scale effects taking into account the objectives of each study. In addition, the use of three different information criteria, AIC, AICc, and BIC in both case studies operated in a similar manner, suggesting that datasets with large number of observations do not seem to affect the estimators. However, special treatment should be given in small datasets, because according to the literature the information criteria can provide biased measures related to parameter specification.

The selection of the most informative aggregation level using information criteria was implemented in this research adopting Nakaya's (2000) approach and extending his research in a zone design context. As a result, the new zone design system is capable of identifying the most statistically valuable aggregation level. As the methodology is based on health incident rates with a Poisson distribution, we support its use into similar health studies where the aggregation level is not given.

8.3.3 Improving the aggregation method

In Chapter 3 the zone design algorithms developed by Openshaw (1977b; 1978b) were introduced identifying issues related to limitations of the FORTRAN programming language. In the A2Z zone design system, the algorithm is structured following the original zoning algorithm, while adopting graph theory for defining the aggregation problems. Subsequently, the defined components of zone design were implemented using an object oriented language (Visual Basic 6). The combination of graph theory with object oriented structure of the new zone design improved dramatically all the parameters of system. While the original A2Z system was developed utilizing the licensed Map Object 2.0 library, it can be easily converted to an open source system as the Map Objects are used for visualization purposes only.

In addition, the A2Z system supports three initial aggregation methods, the initial random aggregation, the initial predefined aggregation and the initial predefined zones. The existence of these methods can be very useful as the researcher can adopt the best initial aggregation approach that fits the needs of her/his study. For example, in Chapter 6 the initial aggregation of predefined zones was selected as we sought to improve the homogeneity of the existing Health Authorities. Furthermore, we implemented three different algorithms for checking the stability of zones during the process, introducing the Perimeter's Stability Method as the most economic approach in terms of processing time. Using this approach, aggregation problems with a large number of areal units can be solved faster as the system is concerned only with the areal units at the boundaries of each zone. An example of its importance and sufficient handling was presented in the case studies of Chapters 6 and 7 with 10,000 and 2,000 areal units respectively, representing large aggregation problems.

8.3.4 Extending objective functions and spatial constraints

In this thesis, we concentrated on homogeneous zones as in Chapter 3 many researchers highlighted the importance of such geographies in health research (Cockings and Martin, 2005). Therefore the developed objective functions focused on the minimisation of variance within zones. The *k*-homogeneity function optimises the homogeneity within zones in terms of *k* standardised variables providing a very useful objective function when the aggregation has to take into account multiple socioeconomic determinants. However, deviance and chi-square functions were also developed using an incident variable and the base population. In Chapters 5, 6 and 7, we evaluated all three objective functions without identifying any reasons to be in favour of one of them. Although the objective functions performed in a similar manner we ought to stress the importance of adapting the objective function to the individual characteristics of each case study. In this thesis, we only used homogeneity functions without extending their complexity into other requirements, such as equal populations, because the objective to maximise within zone homogeneity.

In this thesis, graph theory provided a solid theoretical framework for the new zone design system. Its importance was demonstrated in Chapter 4 suggesting new strategies for controlling the compactness of zones. The Contiguity Constraint method applies a slight compactness rule assisting the zone design system to avoid gerrymandering shapes, while the Advanced Contiguity Constraint approach can provide more restricted compactness constrains. Both methods make use of graph theory elements such as edges and vertices. We consider here the use of graph theory as a potential asset in our research because complicate topological phenomena can be effortlessly implemented in this framework. In addition, further research in graph theory literature can improve or solve spatial problems that geographers have struggled to solve until today. An good example of such an unsolved problem is to calculate all the possible aggregation solutions in an aggregation process, taking into account the contiguity of areal units (Keane, 1975; Rossiter and Johnston, 1981).

8.4 Directions for further research

As stated throughout this thesis, the adoption of zone design methodology in health related studies can become a valuable tool supporting researchers to investigate health issues as well as to suggest better approaches for resource allocations. In chapters 5, 6 and 7, we attempted to provide methodologies in a zone design context demonstrating the importance of zone design system in health research. The overall aim of the dissertation has been achieved but it is recognised that more research should focus in the following topics.

In Chapters 2 and 3, we highlighted the need of an automated procedure related to health issues. Although Chapter 4 effectively tackled most of the technical aspects, there is still need for developing new methodologies to target specific health issues and evaluate the new approaches in a health related context. While Smith et al (2001) suggest developing indices for particular health issues, here we also support the need of new geographies related to health characteristics. The adoption of census administrative geographies (constructed for different purposes) by the health research community, have uncover many limitations (Cockings and Martin, 2005). The zone design methodology can play a strategic role constructing new geographies focused on the particular characteristics of each research project.

Even though the suggested aggregation system provides an advanced framework for aggregation analysis, we support further development of the system in the following components. It became obvious that objective functions are static elements of the system and every change requires programming skills to adjust the new characteristics. Therefore we believe that further development of zone design can provide to the user tools for constructing objective functions in a scripting format. As scripts here, we refer to certain commands which can be executed by the user with an overall result of constructing new objective functions focused on the study needs. In this direction, more advanced algorithms can be implemented based on graph theory. Alternative suggestions such the use of multi-threads/parallel technology (Openshaw and Schmidt, 1996) and genetic algorithms (Nakaya, 2000) can be implemented, resulting in further improvement of the processing time.

Another area of future research can be the amalgamation of zone design (Openshaw 1984a) and multilevel (Jones 1991a; 1991b) concepts for implementing methodologies that investigate relationships between different aggregation levels, while exploring the effects of the MAUP throughout these levels. Although Daras et al. (2005) used the zone design system to construct the appropriate levels for multilevel analysis, we believe that the implementation of an extended zone design methodology, which optimizes different hierarchical aggregation levels using multilevel models can provide the next generation of aggregation systems. Daras et al. (2005) suggested a set of different aggregation solutions reflecting the socioeconomic characteristics of families in Avon area, while investigating the neighborhood perceptions of each family as part of a project focusing on accidents experienced by children aged 0-4 years old. Other health related studies have also employed multilevel models studying health problems at different aggregation levels (Reading et al., 1999), however, such aggregation levels were laboriously designed based on subjective information. What we suggest here is the combination of zone design and multilevel concepts for developing objective methods for health related research.

8.5 Epilogue

In Chapter 1, we discussed the polarization of Human Geography in two groups: Quantitative and Qualitative and if, as geographers, we can favour one of them. In this thesis, we introduced relevant quantitative research in Human Geography in order to suggest innovative techniques appropriate for health related research purposes. Even though this thesis was accomplished with clear quantitative orientations, we support the comment by Johnston et al. (2003, p.160) that "polarization – into, in effect, quantitative versus qualitative, or extensive versus intensive, research – is not only misleading but also creates a dualism that is both unrepresentative of much social science research practice and potentially very limiting to its development,…". We hope that both groups of geographers will narrow this polarization within Human Geography in the future, promoting mixed methods and multidisciplinary research agendas in academia and the research community.

References

Abdalla, I. M. 1997, *Statistical investigation and modelling of the relationships between road accidents and social characteristics,* PhD Thesis, Napier University.

Abdalla, I. M., Barker, D. and Raeside, R. 1997a, "Road accident characteristics and socioeconomic deprivation", *Traffic Engineering and Control*, **38**, pp. 672-676.

Abdalla, I. M., Raeside, R., Barker, D. and McGuigan, D. R. D. 1997b, "An investigation into the relationships between area social characteristics and road accident casualties", *Accident Analysis and Prevention*, **29**, 5, pp. 583-593.

Adams, J., Ryan, V. and White, M. 2004, "How accurate are Townsend Deprivation Scores as predictors of self-reported health? A comparison with individual level data", *Journal of Public Health*, **27**, 1, pp. 101-106.

Adams, J. and White, M. 2005, "Socio-economic deprivation is associated with increased proximity to general practices in England: an ecological analysis", *Journal of Public Health*, **27**, 1, pp. 80-81.

Akaike, H. 1973, "Information theory and an extension of the maximum likelihood principle," in *Breakthroughs in statistics (1992), vol. 1*, Springer-Verlag, pp. 610-624.

Alker, H. 1969, "A Typology of Ecological Fallacies," in *Quantitative Ecological Analysis in the Social Sciences*, M. Dogan & S. Rokkan, eds., MIT Press, Cambridge.

Allison, P. 1978, "Measures of inequality", *American Sociological Review*, **43**, pp. 865-880.

Alvanides, S. 2000, *Zone Design methods for application in Human Geography,* PhD Thesis, University of Leeds.

Alvanides, S. and Macmillan, W. 1997, "A fast contiguity checking algorithm for use in zone design", GISRUK '97 Conference.

Alvanides, S. and Openshaw, S. 1999, "Zone design for planning and policy analysis," in *Geographical Information and Planning*, J. Stillwell, S. Geertman, & S. Openshaw, eds., Springer-Verlag, Berlin, p. 299.

Alvanides, S., Daras, K. and Dickinson, H. 2001, "*Designing spatial frameworks for epidemiological research: boundary-free measures and optimal zones of population mixing*", GISRUK '01 Conference.

Alwash, R. and McCarthy, M. 1988, "Accidents in the home among children under 5: ethnic differences or social disadvantage?", *British Medical Journal*, **296**, pp. 1450-1453.

Anderberg, M. R. 1973, *Cluster Analysis for Applications* Academic Press, Inc., New York.

Anderson, N. H. and Titterington, D. M. 1997, "Some methods for investigating spatial clustering, with epidemiological applications", *Journal of the Royal Statistical Society, Series A*, **160**, pp. 87-105.

Arbia, G. 1989, "Statistical effect of data transformations: a proposed general framework," in *The Accuracy of Spatial Data Bases*, M. Goodchild & S. Gopal, eds., Taylor and Francis, London, pp. 249-259.

Atkinson, A. B. 1980, *Wealth, Income and Inequality* Oxford University Press, Oxford.

Bartley, M. and Blane, D. 1994, "Appropriateness of Deprivation Indexes Must be Ensured", *British Medical Journal*, **309**, 6967, p. 1479.

Bauld, L., Judge, K., Lawson, L., Mackenzie, M., Mackinnon, J. and Truman, J. 2000, *Health action zones in transition: progress in 2000*, Department of Health, UK.

Beaver, M. W. 1973, "Population, Infant Mortality and Milk", *Population Studies*, **27**, 2, pp. 243-254.

Besag, J. and Newell, J. 1991, "The detection of clusters in rare diseases", *Journal of the Royal Statistical Society, Series A*, **154**, pp. 143-155.

Blakely, T. and Woodward, A. 2000, "Ecological effects in multi-level studies", *Journal of Epidemiology and Community Health*, **54**, pp. 367-374.

Blalock, H. M., 1964, Causal Inferences in Nonexperimental Research, Chapel Hill, NC, University of North Carolina Press.

Boots, B. 2001, "Using local statistics for boundary characterization", *GeoJournal*, **53**, 4, pp. 339-345.

Boyle, M. H. and Willms, D. J. 1999, "Place effects for areas defined by administrative boundaries", *American Journal of Epidemiology*, **149**, 6, pp. 577-585.

Boyle, P. and Feng, Z. 2002, "A method for integrating the 1981 and 1991 British census interaction data", *Computers, Environment and Urban Systems*, **26**, 2-3, pp. 241-256.

Boyle, P. J. 1993, "Modelling the relationship between migration and tenure", *Transactions, Institute of British Geographers*, **18**, pp. 359-376.

Boyle, P. J., Flowerdew, R. and Shen, J. 1998, "Analysing local-area migration using data from the 1991census: the importance of housing growth and tenure", *Regional Studies*, **32**, pp. 113-132.

Briggs, K. 1972, *Introducing Transport Networks* Univ.of London Press, London.

Bryant, J. 1998, "The importance of human and organisational factors in the implementation of computerised information systems: an emerging theme in European healthcare", *The British Journal of Healthcare Computing and Information Management*, **15**, 4, pp. 27-30.

Buck, C. and PAHO 1988, *The Challenge of epidemiology : issues and selected readings* Pan American Health Organization, Washington, D.C.

Budd, T. 2001, *An Introduction to Object-Oriented Programming* Addison-Wesley, Massachusetts.

Bullen, N., Moon, G. and Jones, K. 1996, "Defining localities for health planning: A GIS approach", *Social Science & Medicine*, **42**, 6, pp. 801-816.

Burnham, K. P. and Anderson, D. R. 1998, *Model Selection and Inference* Springer.

Carr-Hill, R. A. and Sheldon, T. 1991, "Designing a deprivation payment for general practitioners: the UPA(8) wonderland", *British Medical Journal*, **302**, pp. 393-396.

Carstairs, V. and Morris, R. 1989, "Deprivation and mortality: an alternative to social class?", *Community Medicine*, **11**, pp. 210-219.

Carstairs, V. and Morris, R. 1990, *Deprivation and health in Scotland* Aberdeen University Press, Aberdeen.

Casado-Diaz, J. M. 1996, "Mercados laborales locales: analisis preliminar del caso valenciano", *Revista de Estudios Regionales*, **39**, pp. 159-172.

CDPHE 2001, *New International Classification of Diseases (ICD-10): the History and Impact*, Colorado Department of Public Health and Environment, Denver, 41.

Ceccato, V., Haining, R. and Signoretta, P. 2002, "Exploring offence statistics in Stockholm City using spatial analysis tools", *Annals of the Association of American Geographers*, **92**, pp. 29-51.

Champion, J. A. I. 1993, *Epidemic disease in London* London : Centre for Metropolitan History.

Choynowski, M. 1959, "Maps based on probabilities", *Journal of the American Statistical Association*, **54**, pp. 385-388.

Christie, N. 1995, *Social, economic and environmental factors in child pedestrian accidents: A research review*, Transport Research Laboratory, Wokingham, Project Report 116.

Clarke, G. P., Alvanides, S., and Bailey, A. J., 1999, A Zone Design Tool for Use with 2001 Census Data, ESRC 2001 Census Development Programme,Award No: H507255153.

Cliff, A. D. and Haggett, P. 1970, "On the efficiency of alternative aggregations in region-building problems", *Environment and Planning A*, **2**, pp. 285-294.

Cliff, A. D. and Ord, J. K. 1981, *Spatial processes* Pion, London.

Cockings, S. and Martin, D. 2005, "Zone design for environment and health studies using pre-aggregated data", *Social Science & Medicine*, **60**, 12, pp. 2729-2742.

Congdon, P. 1996, "Suicide and parasuicide in London: a small-area study", *Urban Studies*, **33**, pp. 137-158.

Congdon, P., Campos, R. M., Curtis, S. E., Southall, H. R., Gregory, I. N. and Jones, I. R. 2001, "Quantifying and Explaining Changes in Geographical Inequality of Infant Mortality in England and Wales since the 1890s", *International Journal of Population Geography*, **7**, pp. 35-51.

Coombes, M. 2000, "Defining locality boundaries with synthetic data", *Environment and Planning A*, **32**, pp. 1499-1518.

Coombes, M. G., Dixon, J. S., Goddard, J. B., Openshaw, S. and Taylor, P. J. 1982, "Functional regions for the population census of Great Britain," in *Geography and the urban environment*, D. T. Herbert & R. J. Johnston, eds., Wiley, London, pp. 63-112.

Coombes, M. G., Green, A. E. and Openshaw, S. 1986, "An efficient algorithm to generate official statistical reporting areas: the case of the 1984 travel-to-work areas revision in Britain", *Journal of Operational Research*, **37**, 10, pp. 943-953.

Cope, C. R. 1971, "Regionalisation and the electoral districting problem", *Area*, **3**, pp. 190-195.

Cowell, F. A. 1995, *Measuring Inequality*, 2nd edn, Prentice Hall/Harvester Wheatsheaf, London.

Crayford, T., Shanks, J., Bajekal, M. and Langford, S. 1995, "Analysis from inner London of deprivation payments based on enumeration districts rather than wards", *BMJ*, **311**, 7008, pp. 787-788.

Cuzick, J. and Edwards, R. 1990, "Spatial clustering for inhomogeneous populations", *Journal of the Royal Statistical Society, Series B*, **52**, pp. 73-104.

Daras, K. and Alvanides, S. 2005, "Designing sustainable public health geographies in England and Wales," in *GIS for Sustainable Development*, M. Campagna, ed., Taylor & Francis Group, London.

Daras, K., Haynes, R., Jones, A. and Reading, R. 2005, "Neighbourhood zone design for a study of injuries to preschool children", RGS-IBG Annual International Conference, London, p. 35.

Davies, J. 1998, "Healthy Living Centres", *Health Services Journal*, **5**, pp. 1-5.

DETR, 2000, Index of Deprivation 2000: Regeneration Research Summary, London, Department of the Environment, Transport and Regions.

DfT 2003, *Child Road Safety: Archieving the 2010 Target*, Department of Transport, London.

Dickinson, H. O. and Parker, L. 1999, "Quantifying the effect of population mixing on childhood leukaemia risk: the Seascale cluster", *British Journal of Cancer*, **81**, pp. 144-151.

Diggle, P. J., Gatrell, A. C. and Lovett, A. A. 1990, "Modelling the prevalence of cancer of cancer of larynx in part of Lancashire: a new methodology for spatial epidemiology.," in *Spatial Epidemiology*, R. W. Thomas, ed., Pion, London, pp. 153-171.

Dijkstra, E. W. 1959, "A note on two problems in connection with graphs", *Numeriske Math*, **1**, pp. 269-271.

Ding, C. 1998, "The GIS-Based Human Interactive TAZ Design Algorithm: Examing The Impacts of Data Aggregation on Transportantion Planning Analysis", *Enviroment & Planning B:Planning and design*, **25**, pp. 601-616.

DoH, 2000, The NHS Plan: a plan for investment, a plan for reform, London, Department of Health: Stationary Office Cm 4818-I.

DoH, 2001a, Building the information core implementing the NHS Plan, London, Department of Health: Crown Copyright.

DoH, 2001b, Consultation on the proposal to establish a new Health Authority for Avon, Gloucestershire and Wiltshire, London, Department of Health: Crown Copyright.

DoH, 2001c, Consultation on the proposal to establish a new Health Authority for Somerset and Dorset, London, Department of Health: Crown Copyright.

DoH, 2002, Shifting the Balance of Power: The Next Steps, London, Department of Health: Crown Copyright.

Duda, R. O. and Hart, P. E. 1973, *Pattern Classification and Scene Analysis* Wiley, New York.

Dummer, T., Dickinson, H., Pearce, M., Charlton, M. and Parker, L. 2000, "Stillbirth risk with social class and deprivation: no evidence for increasing inequality", *Journal of Clinical Epidemiology*, **53**, 2, pp. 147-155.

Duncan, C., Jones, K. and Moon, G. 1993, "Do places matter: a multilevel analysis of regional variations in health related behaviour in Britain", *Social Science & Medicine*, **37**, pp. 725-733.

Duncan, C., Jones, K. and Moon, G. 1998, "Context, composition and heterogeneity: Using multilevel models in health research", *Social Science & Medicine*, **46**, 1, pp. 97-117.

Eagleson, S., Escobar, F. and Williamson, I. 2002, "Hierarchical spatial reasoning theory and GIS technology applied to the automated delineation of administrative boundaries", *Computers, Environment and Urban Systems*, **26**, 2-3, pp. 185-200.

Eagleson, S., Escobar, F. and Williamson, I. 2003, "Automating the administration boundary design process using hierarchical spatial reasoning theory and geographical information systems", *International Journal of Geographical Information Science*, **17**, 2, pp. 99-118.

Eco, U. 1984, *The name of the rose* Richard Clay (The Chaucer Press) Ltd, Suffolk.

Egenhofer, J. M. and Mark, M. D. 1995, "Modeling conceptual neighborhoods of topological line-region relations", *International Journal of Geographical Information Systems*, **9**, 5, pp. 555-565.

ESRI 1999, *Understanding GIS: The ARC/INFO Method of Unix and Windows/NT* ESRI Press, Redlands, CA.

Ester, M., Kriegel, H., Sander, J. and Xu, X. 1996, "*A density-based algorithm for discovering clusters in large spatial databases with noise*", Proc. 2nd Int. Conf. on Knowledge Discovery and Data Mining (KDD'96), Portland, pp. 226-231.

Eurostat 1992, *Study on employment zones*, Eurostat, Luxembourg, E/LOC/20.

Flowerdew, R. and Amrhein, C. 1989, "Poisson regression models of Canadian census division migration flows", *Regional Science Association*, **67**, pp. 89-102.

Fortin, M.-J. and Edwards, G. 2000, "Accuracy issues related to the delineation of vegetation boundaries," in *Spatial uncertainty in ecology*, C. Hunsaker et al., eds., Springer-Verlag, Berlin.

Fortin, M.-J., Olson, R. J., Ferson, S., Iverson, I., Hunsaker, C., Edwards, G., Levine, D., Butera, K. and Klemas, V. 2000, "Issues related to the detection of boundaries", *Landscape Ecology*, **15**, pp. 453-466.

Fotheringham, A. S. 1997a, "Trends in quantitative methods I: stressing the local", *Progress in Human Geography*, **21**, 1, pp. 88-96.

Fotheringham, A. S. 1997b, "Analyzing Numerical Data," in *Methods in Human Geography: A Guide for Students Doing a Research Project*, R. Flowerdew & D. Martin, eds., Longman, Essex.

Fotheringham, A. S. and Wong, D. W. S. 1991, "The modifiable area unit problem in multivariate analysis", *Environment and Planning A*, **23**, pp. 1025-1044.

Fotheringham, A. S., Brunsdon, C. and Charlton, M. 2002, *Geographically weighted regression: the analysis of spatially varying relationships* Wiley, Chichester.

Freeman, T. W. 1968, *Geography and regional administration : England and Wales: 1830- 1968* London : Hutchinson University Library.

Fukunaga, K. 1990, *Introduction to Statistical Pattern Recognition*, 2nd edn, Academic Press, New York.

Garrison, W. 1968, "Connectivity of the interstate highway system," in *Spatial Analysis*, B. Berry & D. Marble, eds., Prentice Hall, NJ, pp. 239-249.

Gatrell, A. 2001, "Geographies of health: an introduction", Blackwells, Oxford.

Gehlke, C. E. and Biehl, K. 1934, "Certain effects of grouping upon the size of the correlation coefficient in census tract material", *Journal of the American Statistical Association*, **supplement 29**, pp. 169-170.

Getis, A. and Ord, J. K. 1992, "The analysis of spatial association by use of distance statistics", *Geographical Analysis*, **24**, pp. 189-206.

Getis, A. and Ord, J. K. 1996, "Local spatial statistics: an overview," in *Spatial Analysis: Modelling in a GIS Environment*, P. Longley & M. Batty, eds., GeoInformation International, Cambridge.

Giggs, J. 1973, "The distribution of schizophrenics in Nottingham", *Transactions Institute British Geographers*, **59**, pp. 55-76.

Graham, I. 2001, *Object-Oriented Methods - Principles and Practice* Addison-Wesley, Harlow.

Greenland, S. and Morgenstern, H. 1989, "Ecological bias, confounding, and effect modification", *International Journal of Epidemiology*, **18**, pp. 269-274.

Greenland, S. and Robins, J. 1994, "Ecological studies - biases, misconceptions and counterexamples", *American Journal of Epidemiology*, **139**, pp. 747-760.

Greenwood, P. E. and Nikulin, M. S. 1996, *A Guide to Chi-Squared Testing* John Wiley & Sons, New York.

Gregory, I. N. and Southall, H. R. 1998, "Putting the past in its place: the Great Britain historical GIS," in *Innovations in GIS 5*, S. Carver, ed., Taylor and Francis, London, pp. 210-221.

Gregory, I. N., Dorling, D. and Southall, H. R. 2001, "A century of inequality in England and Wales using standardized geographical units", *Area*, **33**, 3, pp. 297-311.

Grigg, D. 1965, "The Logic of Regional Systems", *Annals of the Association of American Geographers*, **55**, pp. 465-491.

Gudgin, G. 1975, "The Distribution Of Schizophrenics In Nottingham: A Comment", *Transactions Of The Institute Of British Geographers*, **64**, pp. 148-149.

Guo, J., Trinidad, G. and Smith, N. "MOZART: A Multi-Objective Zoning and AggRegation Tool", Manila.

Haggett, P. 1965, *Locational Analysis in Human Geography* Arnold, London.

Haggett, P. and Chorley, R. 1969, *Network Analysis in Geography* St. Martins Press, New York.

Haggett, P., Cliff, A. D. and Frey, A. 1977, *Locational Methods* Arnold, London.

Haines, M. R. 1995, "Socio-economic Differentials in Infant and Child Mortality during Mortality Decline: England and Wales, 1890-1911", *Population Studies*, **49**, 2, pp. 297-315.

Haining, R., Ma, J. and Wise, S. 1996, "Design of a software system for interactive spatial statistical analysis linked to a GIS", *Computational Statistics*, **11**, pp. 449-466.

Halkidi, M., Batistakis, Y. and Vazirgiannis, M. 2002a, "Cluster validity methods: Part I", *SIGMOD Record 31*, **2**, pp. 40-45.

Halkidi, M., Batistakis, Y. and Vazirgiannis, M. 2002b, "Clustering validity checking methods: Part II", *SIGMOD Record 31*, **3**, pp. 19-27.

Halliday, J. and Asthana, S. 2005, "Policy at the margins: developing community capacity in a rural Health Action Zone", *Area*, **37**, 2, pp. 180-188.

Hamnett, C. 2003a, " Response to Johnston et al., Contemporary fiddling in human geography while Rome burns: has quantitative analysis been largely abandoned - and should it be?", *Geoforum*, **34**, pp. 163-164.

Hamnett, C. 2003b, "Contemporary human geography: fiddling while Rome burns?", *Geoforum*, **34**, pp. 1-3.

Hartigan, J. A. and Wong, M. A. 1979, "Algorithm AS136: A k-means clustering algorithm", *Applied Statistics*, **28**, pp. 100-108.

Harvey, D. 1969, *Explanation in Geography* St. Martin's Press, New York.

Hayes, B. 2000, "Graph theory in practice: Part I", *American Scientist*, **88**, 01, pp. 9-13.

Haynes, R. 1987, *The Geography of Health Services in Britain* Croom Helm Ltd., London.

Higgs, G. and Gould, M. 2000, "Commentary: Healthcare Commissioning, the modern NHS, and geographical information systems", *Environment and Planning A*, **32**, 11, pp. 1905-1908.

Higgs, G. and Gould, M. 2001, "Is there a role for GIS in the 'New NHS'", *Health and Place*, **7**, 3, pp. 247-259.

Higgs, G., Smith, D. P. and Gould, M. I. 2003, "Realising `joined-up' geography in the National Health Service: the role of geographical information systems?", *Environment and Planning C: Government and Policy*, **21**, 2, pp. 241-258.

HMSO 1995, *Health Authorities Act 1995 (c. 17)*, Her Majesty's Stationery Office: Crown Copyright, London.

HMSO 1998, *Inequalities in Health*, Her Majesty's Stationery Office: Crown Copyright, London.

HMSO 1999, *Saving Lives: Our Healthier Nation*, Her Majesty's Stationery Office: Crown Copyright, London.

Holt, T., Steel, D. G. and Tranmer, M. 1996, "Areal homogeneity and modifiable areal unit problem", *Geographical Systems*, **3**, pp. 181-200.

Horner, M. W. and Murray, A. T. 2002, "Excess Commuting and the Modifiable Areal Unit Problem", *Urban Studies*, **39**, 1, pp. 131-139.

Hurvich, C. M. and Tsai, C. L. 1989, "Regression and time series model selection in small samples", *Biometrika*, **76**, pp. 297-307.

Hurvich, C. M., Simonoff, J. S. and Tsai, C. L. 1998, "Smoothing parameter selection in nonparametric regression using an improved Akaike information criterion", *Journal of the Royal Statistical Society, Series B*, **60**, pp. 271-293.

Jacquez, G. M. 1995, "The map comparison problem : Tests for the overlap of geographic boundaries", *Statistics in Medicine*, **14**, pp. 2343-2361.

Jacquez, G. M. and Greiling, D. A. 2003, "Geographic boundaries in breast, lung and colorectal cancers in relation to exposure to air toxics in Long Island, New York", *International Journal of Health Geographics*, **2**, 4.

Jarman, B. 1983, "Identification of underprivileged areas", *British Medical Journal*, **286**, pp. 1705-1709.

Jarman, B. 1984, "Underprivileged areas: validation and distribution of scores", *British Medical Journal*, **289**, pp. 1587-1592.

Jarvis, S. 1999, "Why do we need action on injury in the UK?", *Public Health Forum*, **3**, pp. 1-5.

Johnston, R. J. 1976, "Areal studies, ecological studies, and social patterns in cities", *Transactions of the Institute of British Geographers*, **1**, pp. 118-122.

Johnston, R. J., Gregory, D. and Smith, D. M. 1993, *The dictionary of Human Geography* Blackwell.

Johnston, R. J., Hepple, L., Hoare, T., Jones, K. and Plummer, P. 2003, "Contemporary fiddling in human geography while Rome burns: has quantitative analysis been largely abandoned - and should it be?", *Geoforum*, **34**, pp. 157-161.

Jones, K. 1991a, *Multi-Level Models for Geographical Research* CATMOGS, Norwich.

Jones, K. 1991b, "The specification and estimation of multilevel models", *Transactions Of The Institute Of British Geographers*, **NS16**, pp. 148-160.

Jones, K. and Duncan, C. 1995, "Individuals and their ecologies: Analysing the geography of chronic illness within a multilevel modelling framework", *Health and Place*, **1**, 1, pp. 27-40.

Jones, K. and Duncan, C. 1996, "People and places: The multilevel model as a general framework for the quantitative analysis of geographical data," in *Spatial Analysis: Modelling in a GIS Environment*, P. Longley & M. Batty, eds., John Wiley & Sons, New York, pp. 79-104.

Judge, K. and Benzeval, M. 1993, "Health inequalities: new concerns about the children of single mothers", *British Medical Journal*, **306**, pp. 677-680.

Karypis, G. and Kumar, V. 1998a, *A Fast and High Quality Multilevel Scheme for Partitioning Irregular Graphs*, University of Minnesota, Minneapolis, 95-035.

Karypis, G. and Kumar, V. 1998b, *Analysis of multilevel graph partitioning*, University of Minnesota, Minneapolis, 95-037.

Kaufman, L. and Rousseeuw, P. J. 1990, *Finding groups in data* Wiley, New York.

Keane, M. 1975, "The size of the region-building problem", *Environment and Planning A*, **7**, pp. 575-577.

Kulldorff, M. and Nagarwalla, N. 1995, "Spatial disease clusters: Detection and Inference", *Statistics in Medicine*, **14**, pp. 799-810.

Lancaster, G. A. and Green, M. 2002, "Deprivation, ill-health and the ecological fallacy", *Journal of the Royal Statistical Society, A*, **165**, 2, pp. 263-278.

Lankford, P. M. 1969, "Regionalization: theory and alternative algorithms", *Geographical Analysis*, **1**, pp. 196-212.

Lawson, A. B. and Kulldorff, M. A. 1999, "A review of Cluster detection Methods," in *Disease Mapping and Risk Assessment for Public Health*, A. B. Lawson et al., eds., Wiley, New York.

Lawson, A. B., Biggeri, A. and Dreassi, E. 1999, "Edge effects in disease mapping," in *Disease Mapping and Risk Assessment for Public Health*, A. B. Lawson et al., eds., Wiley, New York.

Lee, C. H. 1991, "Regional Inequalities in Infant Mortality in Britain, 1861-1971: Patterns and Hypotheses", *Population Studies*, **45**, 1, pp. 55-65.

Lee, P., Murie, A., and Gordon, D., 1995, Area measures of deprivation: a study of current methods and best practices in the identification of poor areas in Great Britain, Birmingham, CURS, University of Birmingham.

Levitt, R. and Wall, A. 1984, *The Reorganised National Health Service*, 3rd edn, Croom Helm Ltd., London.

Lorenz, M. O. 1905a, "Methods of measuring the concentration of wealth", *American Statistical Association*, **9**, pp. 209-219.

Low, A. and Low, A. 2004, "Measuring the gap: quantifying and comparing local health inequalities", *Journal of Public Health*, **26**, 4, pp. 388-395.

Luenberger, D. G. 1973, *Introduction to Linear and Nonlinear Programming* Addison-Wesley, London.

Lukermann, F. 1961, "The Consept of Location in Classical Geography", *Annals of the Association of American Geographers*, **51**, 2, pp. 194-210.

Macara, A. 2002, "Managing for Health: Why Health Care?", *Health Care Management Science*, **5**, 4, pp. 239-242.

Mackenzie, I. F., Nelder, R., Maconachie, M. and Radford, G. 1998, "'My ward is more deprived than yours'", *Journal of Public Health Medicine*, **20**, 2, pp. 186-190.

Macmillan, W. and Pierce, T. 1994, "Optimization modelling in a GIS framework: the problem of political redistricting," in *Spatial Analysis and GIS*, A. S. Fotheringham & P. Rogerson, eds., Taylor & Francis, London.

Majeed, F. A., Martin, D. and Crayford, T. 1996, "Deprivation payments to general practitioners: limitations of census data.", *British Medical Journal*, **14**, 313, pp. 669-670.

Martin, D. 1996, "An assessment of surface and zonal models of population", *International Journal of Geographical Information Systems*, **10**, pp. 973-989.

Martin, D. 1998, "Optimizing census geography: the separation of collection and output geographies", *International Journal of Geographical Information Science*, **12**, pp. 673-685.

Martin, D., 2001, Geography for the 2001 Census in England and Wales, Office for National Statistics, Census Customer Services.

Martin, D. 2003, "Extending the automated zoning procedure to reconcile incompatible zoning systems", *International Journal of Geographical Information Science*, **17**, 2, pp. 181-196.

Martin, D., Tranmer, M., and Nolan, A., 1999, 2001 Census Output Areas, ESRC's 2001 Census Development Programme, Award No: H507255154.

Martin, D., Nolan, A. and Tranmer, M. 2001, "The application of zone design methodology to the 2001 UK census", *Environment and Planning A*, **33**, pp. 1949-1962.

McVey, E. and Baker, A. 2002, "Improving ONS spatial referencing – the impact on 2000 births and deaths data", *Population Trends*, **Spring**, 107, pp. 14-22.

Mineter, M. J. 2003, "A software framework to create vector-topology in parallel GIS operations", *International Journal of Geographical Information Science*, **17**, 3, pp. 203-222.

Moon, G. and Gould, M. 2000, *Epidemiology : an introduction* Open University Press, Buckingham.

More, A. and Martin, D. 1998, "Quantitative health research in an emerging information economy", *Health and Place*, **4**, 3, pp. 213-222.

Morrill, R. 1987, "Redistricting, Region and Representation", *Political Geography Quarterly*, **6**, 3, pp. 241-260.

Morris, R. and Carstairs, V. 1991, "Which deprivation? A comparison of selected deprivation indexes", *Journal of Public Health Medicine*, **13**, pp. 318-325.

Nakaya, T. 1996, *An Information Statistical Approach to the Modifiable Areal Unit Problem in Map Analysis,* PhD Thesis, Tokyo Metropolitan University.

Nakaya, T. 2000, "An information statistical approach to the modifiable areal unit problem in incidence rate maps", *Environment and Planning A*, **32**, 1, pp. 91-109.

Nakaya, T. 2001, "Local spatial interaction modelling based on the geographically weighted regression approach", *GeoJournal*, **53**, 4, pp. 347-358.

Nakaya, T., Nakase, K. and Osaka K. 2005, “Spatio-temporal modelling of the HIV epidemic in Japan based on the national HIV/AIDS surveillance", *Journal of Geographical Systems*, **7**, 3-4, pp.313-336.

National Statistics 2005, *Census 2001: Output prospectus*, Census Customer Services, Fareham.

Norman, P., Rees, P. and Boyle, P. 2003, "Achieving Data Compatibility over Space and Time: Creating Consistent Geographical Zones", *International Journal of Population Geography*, **9**, 5, pp. 364-386.

Oden, N. L., Sokal, R. R., Fortin, M. J. and Goebl, H. 1993, "Categorical wombling: Detecting regions of significant change in spatially located categorical variables", *Geographical Analysis*, **25**, pp. 315-336.

ODPM 2004a, *Better method of identifying deprived areas to help tackle disadvantage launched - Rooker*, Office of the Deputy Prime Minister, *2004/0109*.

ODPM 2004b, *The English Indices of Deprivation 2004 (revised)*, Office of the Deputy Prime Minister, London, 04NRU02094.

Openshaw, S. 1976, "A regionalisation procedure for a comparative regional taxonomy", *Area*, **8**, pp. 149-152.

Openshaw, S. 1977a, "Optimal zoning systems for spatial interaction models", *Environment and Planning A*, **9**, pp. 169-184.

Openshaw, S. 1977b, "Algorithm 3: a procedure to generate pseudo-random aggregations of N zones into M zones, where M is less than N", *Environment and Planning A*, **9**, pp. 1423-1428.

Openshaw, S. 1977c, "A geographical solution to scale and aggregation problems in region-building, partitioning and spatial modelling", *Transactions Of The Institute Of British Geographers*, **NS2**, pp. 459-472.

Openshaw, S. 1978a, "An empirical study of some zone-design criteria", *Environment and Planning A*, **10**, pp. 95-113.

Openshaw, S. 1978b, "An optimal zoning approach to the study of spatially aggregated data," in *Spatial Representation and Spatial Interaction*, I. Masser & P. J. B. Brown, eds., Martinus Nijhoff, Leiden.

Openshaw, S. 1984a, "The modifiable areal unit problem", *Concepts and Techniques in Modern Geography (CATMOG)*, **38**.

Openshaw, S. 1984b, "Ecological fallacies and the analysis of area census data", *Environment and Planning A*, **16**, pp. 17-31.

Openshaw, S. and Taylor, P. J. 1979, "A million or so correlation coefficients: three experiments on the modifiable areal unit problem," in *Statistical Applications in the Spatial Sciences*, N. Wrigley, ed., Pion, London, pp. 127-144.

Openshaw, S. and Taylor, P. J. 1981, "The Modifiable Areal Unit Problem," in *Quantitative Geography: A British View*, N. Whigley & R. J. Bennet, eds., Routledge, London.

Openshaw, S., Charlton, M., Wymer, C. and Craft, A. 1987, "A mark I geographical analysis machine for the automated analysis of point data sets", *International Journal of Geographical Information Systems*, **1**, 4, pp. 335-358.

Openshaw, S. and Rao, L. 1995, "Algorithms for re-engineering 1991 census geographies", *Environment and Planning A*, **27**, pp. 425-446.

Openshaw, S. and Schmidt, J. 1996, "Parallel simulated annealing and genetic algorithms for re-engineering zoning systems", *Geographical Systems*, **3**, pp. 201-220.

Openshaw, S. and Alvanides, S. 1999, "Applying geocomputation to the analysis of spatial distributions," in *Geographical Information Systems - Vol.1: Principles and Technical issues*, 2nd edn, P. A. Longley et al., eds., Wiley, pp. 267-282.

Ortuzer, J. D. and Willumsen, L. G. 1994, *Modelling Transport* John Wiley & Sons, New York.

Parzen, E., Tanabe, K. and Kitagawa, G. 1998, *Selected Papers of Hirotugu Akaike* Springer-Verlag, New York.

Piantadosi, S., Byar, D. P. and Green, S. B. 1988, "The ecological fallacy", *American Journal of Epidemiology*, **127**, pp. 893-904.

Propper, C., Jones, K., Bolster, A., Burgess, S., Johnston, R. and Sarker, R. 2004, *Local neighbourhood and mental health: evidence from the UK* Manchester, ESRC Ref: H333250042.

Reading, R., Langford, I. H., Haynes, R. and Lovett, A. 1999, "Accidents to preschool children: comparing family and neighbourhood risk factors", *Social Science & Medicine*, **48**, pp. 321-330.

Richardson, S., Stucker, I. and Hemon, D. 1987, "Comparison of relative risks obtained in ecological and individual studies: Some methodological considerations", *International Journal of Epidemiology*, **16**, pp. 111-120.

Roberts, I. and Pless, B. 1995, "Social policy as a cause of childhood accidents: the children of lone mothers", *British Medical Journal*, **311**, pp. 925-928.

Roberts, I. and Power, C. 1996, "Does the decline in child injury mortality vary by social class? A comparison of class specific mortality in 1981 and 1991", *British Medical Journal*, **313**, pp. 784-786.

Robinson, G. M. 1998, *Methods and Techniques in Human Geography* Wiley, Chichester.

Robinson, W. S. 1950, "Ecological Correlations and the Behavior of Individuals", *American Sociological Review*, **15**, pp. 351-357.

Rossiter, D. J. and Johnston, R. J. 1981, "Program GROUP: the identification of all possible solutions to a constituency-delimitation problem", *Environment and Planning A*, **13**, pp. 231-238.

Rumbaugh, J., Blaha, M., Premerlani, W., Eddy, F. and Lorensen, W. 2004, *Object-Oriented Modelling and design* Prentice Hall, New Jersey.

Runyan, C. W., Bangdiwala, S. I., Linzer, M. A., Sacks, J. J. and Butts, J. 1992, "Risk factors for fatal residential fires", *New England Journal of Medicine*, **327**, pp. 859-863.

Sabel, C. E. and Gatrell, A. C. 1998, "Exploratory spatial data analysis of motor neurone disease in north west England: beyond the address at diagnosis," in *GEOMED '97: International Workshop on Geomedical Systems*, L. Gierl et al., eds., Leipzig, Teubner-Verlag Stuttgart, pp. 58-69.

Sakamoto, Y., Ishiguro, M. and Kitagawa, G. 1986, *Akaike information criterion statistics* KTK Scientific Publishers, Tokyo.

Sander, J., Ester, M., Kriegel, H. and Xu, X. 1998, "Density-based Clustering in Spatial Databases: The Algorithm GDBSCAN and its Applications", *Data Mining and Knowledge Discovery*, **2**, 2, pp. 169-194.

Schwarz, G. 1978, "Estimating the dimension of a model", *Annals of Statistics*, **6**, pp. 461-464.

SEHD 2000, *Fair shares for all. Final report of the national review of resource allocation for the NHS in Scotland*, Scottish Executive Health Department, Edinburgh.

Sforzi, F., Openshaw, S. and Wymer, C. 1997, "Le procedura di identificazione dei sistemi locali del lavoro," in *I Sistemi Locali del Lavoro 1991*, F. Sforzi, ed., ISTAT, Rome, pp. 235-242.

Sharples, P. M., Storey, A., Aynsley-Green, A. and Eyre, J. A. 1990, "Causes of fatal childhood accidents involving head injury in Northern region, 1979-86", *British Medical Journal*, **301**, pp. 1193-1197.

Shaw, M., Smith, G. D. and Dorling, D. 2005, "Health inequalities and New Labour: how the promises compare with real progress", *British Medical Journal*, **330**, pp. 1016-1021.

Shortt, N. K., Moore, A., Coombes, M. and Wymer, C. 2005, "Defining regions for locality health care planning: a multidimensional approach", *Social Science & Medicine*, **60**, pp. 2715-2727.

Simpson, S. 1996, "Resource allocation by measures of relative social need in geographical areas: the relevance of the signed X2, the percentage, and the raw count", *Environment and Planning A*, **28**, pp. 537-554.

Sloggett, A. and Joshi, H. 1994, "Higher mortality in deprived areas: community or personal disadvantage?", *BMJ*, **309**, 6967, pp. 1470-1474.

Smellie, K. B. 1946, *A History of local government* London : Allen and Unwin.

Smith, G. D., Whitley, E., Dorling, D. and Gunnell, D. 2001, "Area based measures of social and economic circumstances: cause specific mortality patterns depend on the choice of index", *Journal of Epidemiology and Community Health*, **55**, pp. 149-150.

Smith, D. P., Gould, M. I. and Higgs, G. 2003, "(Re)surveying the uses of Geographical Information Systems in Health Authorities 1991-2001", *Area*, **35**, 1, pp. 74-83.

Snow, J. 1849, *On the mode of communication of cholera* J. Churchill, London.

Stevenson, T. H. C. 1928, "The Vital Statistics of Wealth and Poverty", *Journal of the Royal Statistical Society*, **91**, 2, pp. 207-230.

Stewart-Brown, S., Peters, T. J., Golding, J. and Bijur, P. 1986, "Case definition in childhood accident studies: a vital factor in determining results", *International Journal of Epidemiology*, **15**, pp. 352-359.

Stiller, C. A. and Boyle, P. 1996, "Effect of population mixing and socioeconomic status in England and Wales, 1979-85, on lymphoblastic leukaemia in children", *British Medical Journal*, **313**, pp. 1297-1300.

Sugiura, N. 1978, "Further analysis of the data by Akaike's information criterion and the finite corrections", *Communications in Statistics, Theory and Methods*, **A7**, pp. 13-26.

Susser, M. 1973, *Causal Thinking in the Health Sciences: Concepts and Strategies of Epidemiology* Oxford Univ. Press, New York.

Talbot, R. J. 1991, "Underprivileged areas and health care planning: implications of use of Jarman indicators of urban deprivation", *British Medical Journal*, **302**, pp. 383-386.

Tango, T. 1995, "A class of tests for detecting 'general' and 'focused' clustering of rare diseases", *Statistics in Medicine*, **14**, pp. 2323-2334.

Tango, T. 1999, "Comparison of general tests for spatial clustering," in *Disease Mapping and Risk Assessment in Public Health*, A. e. al. Lawson, ed., Wiley, New York, pp. 111-117.

Taylor, P. J. 1973, "Some implications of the spatial organisation of elections", *Transactions of the Institute of British Geographers*, **60**, pp. 121-135.

Theobald, D. M. 2001, "Topology revisited: representing spatial relations", *International Journal of Geographical Information Science*, **15**, 8, pp. 689-705.

Tidswell, V. 1976, *Pattern and Process in Human Geography* University Tutorial Press.

Towner, E., Dowswell, T. and Jarvis, S. 1993, "The Effectiveness of Health Promotion Interventions: a Literature Review," in *Reducing Childhood Accidents*, Health Education Authority, London.

Townsend, P. 1979, *Poverty in the United Kingdom* Penguin, Harmondsworth.

Townsend, P. 1987a, "Deprivation", *Journal of Social Policy*, **16**, 1, pp. 125-146.

Townsend, P. 1987b, *Poverty and labour in London: Interim report of a centenary survey* Low Pay Unit, London.

Townsend, P., Phillimore, P. and Beattie, A. 1988, *Health and deprivation: inequality and the North* Croom Helm, London.

Townsend, P. and Davidson, N. 1992, "The Black Report," in *Inequalities in Health:The Black Report and the Health Divide*, P. Townsend, N. Davidson, & M. Whitehead, eds., Penguin, London.

Tranmer, M. and Steel, D. G. 1998, "Using census data to investigate the causes of the ecological fallacy", *Environment and Planning A*, **30**, pp. 817-831.

Tranmer, M. and Steel, D. G. 2001, "Ignoring a level in a multilevel model: evidence from UK census data", *Environment and Planning A*, **33**, 5, pp. 941-948.

Tunstall, R. and Lupton, R. 2003, *Is targeting deprived areas an effective means to reach poor people? :an assessment of one rationale for area-based funding programmes* Centre for Analysis of Social Exclusion, London.

Turnbull, B. W., Iwano, E. J., Burnett, W. S., Howe, H. L. and Clark, L. C. 1990, "Monitoring for clusters of disease: Application to leukemia in upstate New York", *American Journal of Epidemiology*, **132**, 1, pp. 136-143.

Unwin, D. J. 2005, "Fiddling on a different planet?", *Geoforum*, **36**, pp. 681-684.

Wagstaff, A., Paci, P. and van Doorslaer, E. 1991, "On the measurement of inequalities in health", *Social Science & Medicine*, **33**, pp. 545-557.

Wallace, M. and Denham, C. 1996, "The ONS classification of local and health authorities of Great Britain", *Studies in Medical and Population Subjects*, **59**.

Walsh, S. and Jarvis, S. 1992, "Measuring the frequency of "severe" accidental injury in childhood", *Journal of Epidemiology and Community Health*, **46**, pp. 26-32.

Watterson, P. A. 1988, "Infant Mortality by Father's Occupation from the 1911 Census of England and Wales", *Demography*, **25**, 2, pp. 289-306.

Wegener, M. 1994, "Operational urban models: state of the art", *Journal of the American Planning Association*, **60**, pp. 17-29.

Wen, S. W., Demissie, K., August, D. and Rhoads, G. G. 2001, "Level of aggregation for optimal epidemiological analysis: the case of time to surgery and unnecessary removal of the normal appendix", *Journal of Epidemiology and Community Health*, **55**, pp. 198-203.

Whitehead, M. 1992, "The Health Divide," in *Inequalities in Health:The Black Report and the Health Divide*, P. Townsend, N. Davidson, & M. Whitehead, eds., Penguin, London.

Whitley, E., Gunnell, D., Dorling, D. and Smith, G. D. 1999, "Ecological study of social fragmentation, poverty and suicide", *British Medical Journal*, **319**, pp. 1034-1037.

WHO 1992, *ICD-10 : international statistical classification of diseases and related health problems*, 10th edn, World Health Organization, Geneva.

Wise, S., Haining, R. and Ma, J. 1997, "Regionalisation tools for the exploratory spatial analysis of health data," in *Recent Developments in Spatial Analysis*, A. Getis & M. M. Fischer, eds., Springer, Berlin, pp. 83-100.

Wise, S., Haining, R. and Ma, J. 2001, "Providing spatial statistical data analysis functionality for the GIS user: the SAGE project", *International Journal of Geographical Information Science*, **15**, pp. 239-254.

Wong, D. and Amrhein, C. 1996, "Research on the MAUP: Old wine in a new bottle or real breakthrough?", *Geographical Systems*, **3**, pp. 73-76.

Woods, R. I. 1997, "Infant mortality in Britain: a survey of current knowledge on historical trends and variations," in *Infant and Child Mortality in the Past*, A. Bideau, B. Dejardins, & H. P. Brigonoli, eds., Charendon Press, Oxford, pp. 74-88.

Woods, R. I. 2000, *The Demography of Victorian England and Wales* Cambridge University Press, Cambridge.

Woods, R. I., Watterson, P. A. and Woodward, J. H. 1988, "The Causes of Rapid Infant Mortality Decline in England and Wales, 1861-1921. Part I", *Population Studies*, **42**, 3, pp. 343-366.

Woods, R. I., Watterson, P. A. and Woodward, J. H. 1989, "The Causes of Rapid Infant Mortality Decline in England and Wales, 1861 - 1921. Part II", *Population Studies*, **43**, 1, pp. 113-132.

Wrigley, N. 1990, "ESRC and the 1991 Census", *Environment and Planning A*, **22**, pp. 573-582.

Wrigley, N., Holt, T., Steel, D. and Tranmer, M. 1996, "Analysing, modelling, and resolving the ecological fallacy," in *Spatial Analysis: Modelling in a GIS Environment*, P. Longley & M. Batty, eds., GeoInformation International, Cambridge, pp. 25-40.

Yule, G. U. and Kendall, M. G. 1950, *An Introduction to the Theory of Statistics* Griffin, London.