Spatio-thematic Accuracy in the Evaluation of the English Safer Cities Programme

Thesis submitted for the degree of

Doctor of Philosophy

at the University of Leicester

by

Ho Chung LAW

BSc (Hons), BA (Hons), C. Psychol.

Department of Geography

University of Leicester

June 1999

Abstract

The Safer Cities Programme in England as a whole implemented over 3,600 crime prevention schemes in 20 cities between 1988-1995 (total costing £30 million). The large-scale evaluation of the Programme's impact on domestic burglary has estimated that, overall, schemes of the Safer Cities Action reduced burglaries by 56,000 and were cost-effective (a saving of about £31 million). Using two cities: Bristol and Coventry within the Safer Cities Programme as a case study, this research aims to explore some of the accuracy issues in the GIS processing involved in the evaluation. This thesis a) describes how spatio-thematic accuracy can be estimated using Monte Carlo and dasymetric methods within the context of the Safer Cities Programme Evaluation; b) thereby provides a precise quantitative statement on the errors involved in the geographical data processing; and c) examines how spatial errors may affect the conclusion of the Evaluation using multi-level modelling. On average, the results show that the overlay method used in the Evaluation has over-estimated the household counts by 3.6% and 5% for Bristol and Coventry respectively. Subsequently, the Safer Cities Programme Evaluation has underestimated the action intensity by –0.8 and –9% and the burglary risk by –7% and –5% (for Bristol and Coventry respectively). Multi-level modelling shows that the mean errors due to the spatial interpolation estimated by the Monte Carlo dasymetric method are –1.5%, 2.3% and 0.7% for Coventry, Bristol, and the two cities combined respectively. In all cases, these are well within the standard errors generated by the overlay method. It is concluded that spatial and thematic errors have no significant impact upon the conclusions of the Safer Cities Programme Evaluation. However, spatial analyses show that potential burglary hot spots might have been missed as a result of such errors in crime pattern analysis. The analysis of the error distribution shows that a geographical area would have a higher error rate if it has: dense population; is near the city centre; or has an irregular geographical boundary. The implications in GIS applications, and crime prevention for decision and policy makers are discussed.

Ho LAW

To Mum & Dad

Julie and Elizabeth

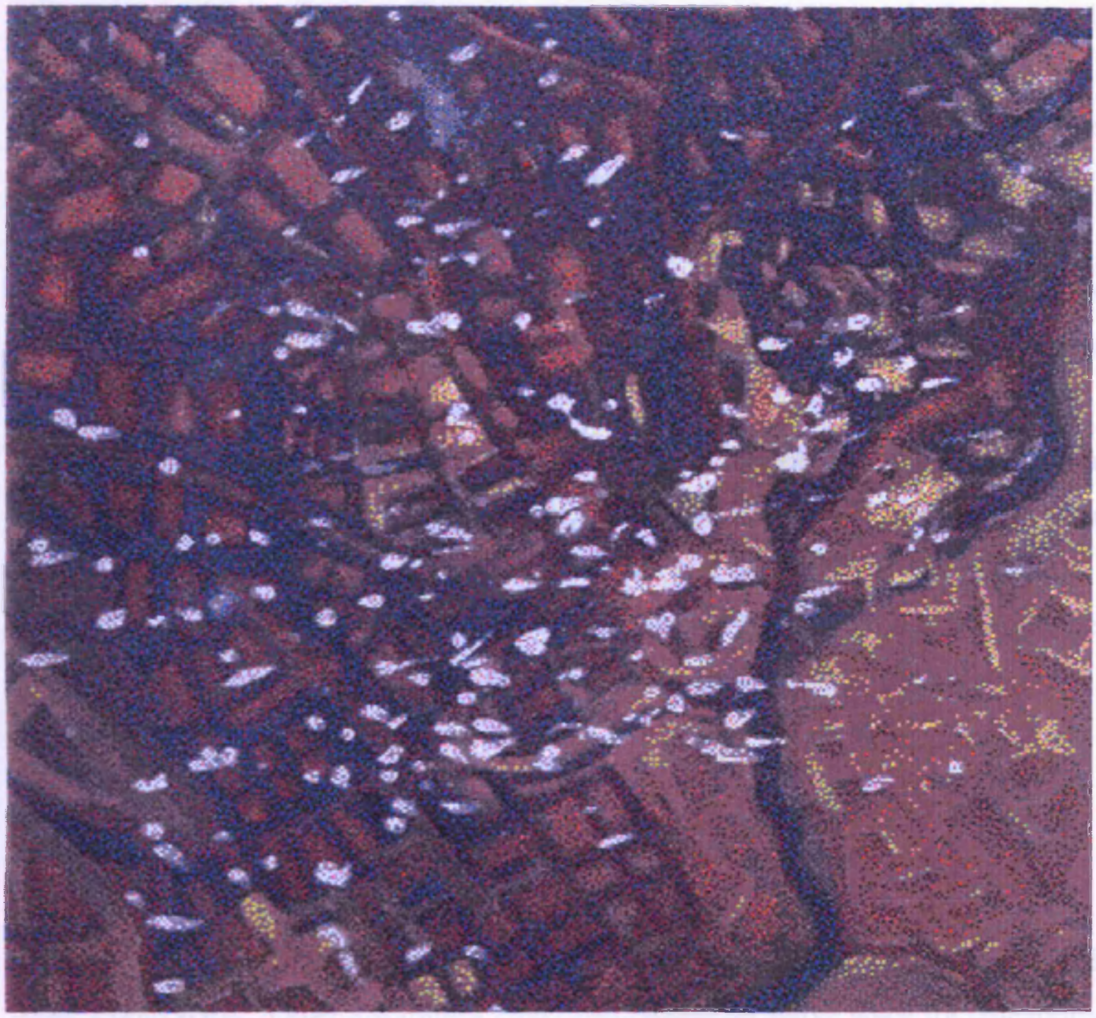


Plate 1: Monte Carlo simulation on Bristol satellite image 1995. Oil on canvas. 20" x 40".

Preface

My journey

As a child, influenced by my Chinese culture, I thought that one would find a meaning through education. Such a search took me away from my homeland (Hong Kong, a former British colony) to England. Later, I was told that getting a PhD would be the highest academic achievement. The award was perceived as a marker of one's destiny on the journey of searching for a meaning, whatever that meaning might be. My journey has taken twenty-two years. During this time, I got married, have a daughter, trained as a psychologist, work as a scientist, and in my own time, practise as an artist.

The PhD itself has taken me 5 years and 11 months to complete. It took me one year to find out what to do, more than one year to work out how to do it and gather all the data sets required, and another year to implement the methodology. Surprisingly, it has taken two and half years to write it all up. The writing up is a far more dynamic process than I expected. It has enabled me to discover areas of omission, which required further analyses. In other words, the writing up of this thesis is an integral part of the research process.

Expression: one/ I / we; active voice / passive voice

Traditionally, a passive voice is encouraged within a scientific discipline to emphasise one's objectivity. Other disciplines tend to encourage an active voice for readability. The style of writing follows a fashion of the culture, which changes in cycles. Ultimately it is a matter of personal taste. It depends on the object or subject to be emphasised in the sentence. A PhD thesis emphasises one's originality. Since this research is based on the work carried out by a whole research team which I was a member of, there is an occasion that is difficult to disentangle between the two. So the subject 'we' is used to refer to the evaluation team as a whole within the context of the Safer Cities Programme. The subject 'I' is used to refer to the work carried out by me. For a general statement I use 'one' or a passive voice when the object is the subject for discussion.

Art, science and process

I use scientific knowledge to inform my work of art, and my artistic practice to articulate my scientific research. For instance the satellite image of Bristol was painted in oil on canvas and the Monte Carlo simulation was performed by throwing dots of white paint at random on the canvas from a distance (Plate 1). The process enabled me to consolidate my thought on the methodology on this research (though its implementation in Lisp-Stat was somewhat tricky).

While writing up the thesis, I was invited to review *the two minute aeroplane factory* exhibition by Chris Burden at the Tate in London. The installation consisted of an automatic assembly line to manufacture and fly model aeroplanes at the rate of one every two minutes. The planes were made of balsa wood, tissue paper, rubber bands and plastic. The process of science and technology was laid bare in front of spectators, and regarded as a work of art. This further blurred the boundary between art and science. It emphasised the process. In a similar vein, the process of doing this research has been laid bare for the reader to examine. While there is a strict definition about what science or scientific practice is, there is no formal definition of art. If one defines art as what an artist produces, *this thesis is a work of art.*

Coming home

The completion of this thesis coincides with the first international symposium on Spatial Data Quality, which happens to be in Hong Kong this summer. Although I (with my family) have been back to Hong Kong many times, going home to present the summary of this research at the symposium has a symbolic meaning to me. It is a classic Chinese story, which depicts a young man who left home in search of education and the highest position with the emperor, through a long and difficult journey, finally achieved the success and returned home. The famous legend *Butterfly Lovers* (*Liang Shan Bo and Zhu Ying Tai* set in the fourth century in Southern China) is one of the variance derived from such a story line. The story has a good ending. It is with such a note that I would like this preface to end, and the thesis to begin. I hope the reader would enjoy tracing some of the steps I took during my journey of this research.

Acknowledgements

The case study described in this thesis was based on a large-scale research project on Safer Cities Programme Evaluation by the Home Office Research & Planning Unit, now called Research Development & Statistics Directorate (RDS) between 1988-1995. The PhD research to study the spatio-thematic accuracy in the evaluation of the English Safer Cities Programme has been sponsored by the Home Office. The evaluation itself involved some 30 person-years effort. Sometimes it is difficult to differentiate between work and PhD research and those who involved whom I should acknowledge. Here I shall only briefly mention a few people who come to my mind at the time of writing.

For the Evaluation of the Safer Cities Programme, there is a long list of acknowledgements in the attached *Home Office Research Study* and will not be reiterated here. I should mention two people whose contributions are particularly relevant to the context of this study. In particular, I am indebted to Dr Paul Ekblom for his scoping & scoring principle in the Evaluation design, his advice on multi-level modelling, and his continuous support and encouragement throughout this project. I would like to thank David Howes of the North West Regional Research Laboratory, the Lancaster University (who since then has moved to Buffalo, USA to pursue his PhD study) who developed the scoping & scoring program in ARC/INFO AML.

For my PhD research, I am most grateful to Professor Peter Fisher of Leicester University who has supervised this research with enthusiasm. I would like to thank Dr Alan Stratchan (Leicester University) for his constructive (and sometimes critical) comments at the initial stage of my research. Especially, I would also like to thank Dr Chris Brunsdon of Newcastle University who inspired and supplied me with the Lisp-Stat programming environment for error modelling.

Doing a part time PhD with a full time job is hard work. It is even harder with a family commitment on the top. I should thank my wife Julie and daughter Elizabeth who have been so supportive during this period, in particular, Julie in her MA course of study in Library and Information Studies who helped me search over thousands of up-to-date literature for my research topic.

I would also like to thank the following people during my journey of this PhD research:

- Dr Phil Emmott (my former colleague, now at Department of Education & Employment) who commented on the statistics of this research.
- Dr David Godfrey (RDS, Home Office) who commented on the Monte Carlo simulation and the testing of the random number generator.
- Dr Peter Grove (RDS) who provided a critique on the methodology and presented three insightful lectures on statistics as part of the continuous professional development for the operational research group.
- Bill Hickin (Leicester University) - converted the Landsat image files into IDRISI format.
- Dr Alex Hirschfield (Liverpool University) - provided numerous and valuable references of his research in application of GIS within socio-demographic context of criminology.
- Dr Chris Kershaw (RDS) - provided the references of his past research on remote sensing.
- Tim Read, a senior researcher from the former Police Research Group which has now become part of RDS, who provided me some further references on crime pattern analyses.
- Dr John Wenzhong Shi (Assistant Professor, Hong Kong Polytechnic University) - provided the references of his past research on error modelling, and commented on this research during my visit.

Although not directly relevant to this research, the following people who I have found particularly supportive during my visit to the University are:

- David Orme and Kate Moore who always provided technical help and support no matter how trivial the matter might be.
- Dr Clare Madge who raised my awareness on cultural geography.
- Anthony Pither (Director of Music) who organised regular Wednesday lunchtime concerts which enriched the texture of my cultural life.

As a whole I found the University of Leicester and in particular the Geography Department a creative and stimulating environment which has made me feel at home throughout my PhD research.

Ho LAW

Spatio-thematic Accuracy in the Evaluation of the English Safer Cities Programme

Contents

Appendices

Chapter One
Introduction

The subject of this research is to explore some of the accuracy issues in the application of Geographical Information systems (GIS) within the specific context of the Safer Cities Programme Evaluation. GIS is now a mature technology widely used in industry across different disciplines. Thus this study would not seek to define what GIS is but focus on the issue of spatial data operations within the context of criminology in general and a crime prevention initiative in particular, namely the Safer Cities Programme. It is assumed that the readers would be familiar of what GIS is. For those who would like to follow up the history and definitions of GIS, they are advised to refer to the literature (see McHarg, 1969; Berry, 1987; de Man, 1988; Carter, 1989; for historical development; and for debate on the concept of GIS, Maguire, 1991; Goodchild, 1990; Openshaw, 1991).

In recent years, practical applications of GIS stretch from decision analyses to decide on the location of supermarkets, and fire stations (Sträng, 1996) to the analyses of the spatial distribution of crimes and their preventive action. Since the Home Office in Britain is regarded as the 'Law and Order Department' within which this case study is placed, this research is primarily concerned with the evaluation of crime preventive action using GIS (Law and Ekblom, 1996). "Reduction in crime, particularly juvenile crime, the fear of crime, and maintenance of good order" - is the number one aim (of the seven) in the Home Office's new statement of purpose (Home Office Annual Report, 1998, also see Appendix 1). As one will see later, the aims of the Safer Cities Programme supported this. The impact of GIS technology is not limited within the boundary of geography as a discipline, but also other disciplines such as social science and economics within a wider context. However, in order to focus the domain of this research, this chapter overviews specifically the impact of GIS in general upon crime pattern analyses and policy making which are relevant to the concerns of the Home Office. GIS has been seen by both the Home Office and police forces as an important crime analytical tool for understanding crime problems and the development of crime prevention measures. The question, 'Can crime pattern analysis make a significant contribution to reducing levels of crime?' has been one of the major pre-occupations for the Home Office as exemplified by the Safer Cities Programme upon which this case study is based.

As an introduction, this chapter locates the specific context of this research within a global context of GIS applications in crime prevention. This chapter examines the theoretical importance of the topic. Section 1.1 describes the impact of GIS technology upon the general context of crime prevention and policy making. Section 1.2 briefly introduces some of the particular concerns on the accuracy issues in the GIS processing which is to be further explored in this research. Section 1.3 outlines the structure of this thesis.

1.1 Impact of GIS technology on criminology and policy making

In the last fifteen years, there have been increasing uses of GIS in crime reduction programmes around the globe. Ecological research in criminology has received a revived interest as a result of using GIS to handle aggregational data (Bursik, 1988, Farrington *et al.*, 1993; Sampson and Groves, 1989). Shaw and McKay (1972) have used GIS to study delinquency patterns in Chicago neighbourhoods. In general, studies of the scenes of crime in terms of their area characteristics have formed a major theme of empirical research in criminology (Brantingham and Brantingham, 1991). The advances of GIS technology have enabled a revival of the formal application of ecological concepts to the analysis of crime problems and, more generally, to understanding social changes and conflicts in cities (the concept known as Chicago School in 1930s, see Park, 1936).

On the operational level, most police forces in Britain over the years would have experimented with 'pin-in-the-wall' type maps. However, such a task is time consuming, and with limited flexibility, difficult to maintain. GIS's unique ability to overlay separate data sets makes it an excellent tool for identifying factors related to the multidimensional, multifaceted crime problem (Rich, 1995). GIS enables crime problems in particular areas to be accurately identified. For example it provides an answer, in seconds, to questions like, "Where did the burglaries with an entry via the back door in 'A' Division occur over a weekend?" This in turn assists in the development of efficient preventive measures to combat the problems identified. Furthermore, rather than rely solely on the police feeling of where the problems lie, it provides firm, factual evidence for the existence of problem areas. Even if the problems identified are already known to the police, the analysis provides a much more suitable basis for decision making. Applications of GIS can thus lead to more efficient policing by focusing attention on clearly identified problems and enabling limited human resources to be used more effectively (Houghton and Berry, 1989; Ratcliffe and McCullagh, 1998).

In the 1980s when GIS was relatively uncommon and expensive, staff in the Home Office already examined the range of micro-computer tools which might assist with crime pattern analysis; and in conjunction with Staffordshire Police, developed a prototype crime analysis system which could be used by a police force in an operational environment (Houghton and Berry, 1989).

More recently, there have been numerous studies reported which claim to use GIS for crime pattern analyses and policy decision aids. For example, Brunsdon (1989) developed an expert system using Bayes' Theorem to analyse crime hot spots over time such as the hot spots shown in Figure 1.1. Johnson *et al* (1997) suggest that the geographical location of repeat victimisation may contribute to burglary 'hot spots'. Systems that show not only the patterns of crimes on maps but also the relationships between levels of crime and the social, demographic and physical variables have also been reported (Hirschfield *et al*, 1995a & b; Hirschfield and Bowers, 1997; Bowers and Hirschfield, 1999). However, the study is inconclusive as it is based only on the recorded crime data available on Merseyside covering a twelve-month period. In contrast, the Safer Cities Programme evaluation reported in Chapter 2 is based on the recorded crime data available for 16 English cities and boroughs covering a six-year period (Ekblom *et al* 1996a).

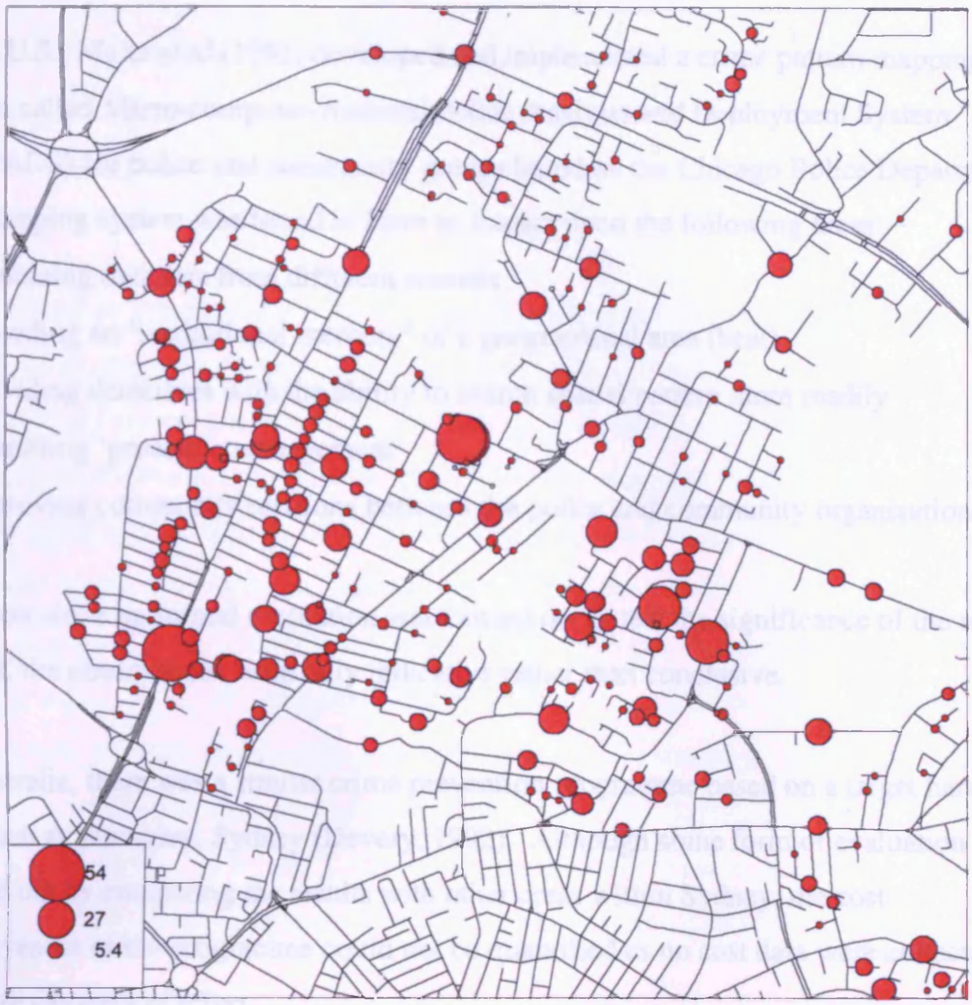


Figure 1. 1: Burglary counts displayed in one of the Safer Cities.

At the same time as GIS development, there is a shift on the emphasis of policing from police-based approach to community based crime prevention known as problem oriented policing (Goldstein, 1990; Rossmo and Fisher, 1993; Spelman and Eck, 1987). The implementation of this approach has used GIS to map different forms of crime profiles. There is also an increase demand from the policy and decision-makers for these kinds of crime reduction programmes (usually funded from central government) to be the subject of vigorous evaluation.

In the U.S., Maltz *et al* (1991) developed and implemented a crime pattern mapping system called Micro-computer-Assisted Police Analysis and Deployment System (MAPADS) for police and community groups based on the Chicago Police Department. The mapping system was found to have an impact upon the following areas:

- combining data sets from different sources
- providing an 'institutional memory' of a geographical area (beat)
- providing detectives with the ability to search spatial pattern more readily
- permitting 'proactive management'
- improving community relations between the police and community organisations.

However since no formal evaluation was carried out to test the significance of the above impact, the observations were only indicative rather than conclusive.

In Australia, there was a similar crime prevention programme based on a target hardening approach at Waverley, Sydney (Devery, 1992). Although some form of evaluation was carried out by comparing the results with other areas within Sydney, the cost effectiveness of the programme could not be quantified as no cost data were collected to measure amounts of effort.

From the above, one can see that using GIS to help implement crime prevention programmes is not unique to England but a common trend across the globe. However, the Safer Cities Programme Evaluation was probably the largest and most comprehensive evaluation of a crime prevention programme in this century both in terms of its size, and scope. The Programme cost £30 million and its Evaluation took more than 30 person-years to complete. Chapter 2 describes the context of the Safer Cities Programme, and the outcome of its Evaluation. For a more comprehensive account of the Evaluation, see the attached Home Office Research Study (Ekblom *et al*, 1996a). The summary of results and conclusions is also published separately as the *Home Office Research Findings 42* (Ekblom, *et al* 1996b). The purpose of this research is not to substantiate the findings of the Evaluation, but to take the opportunity, using the Safer Cities Programme as the context, to explore some of the accuracy issues in GIS processing (see Chapter 4).

1.2 GIS and spatial accuracy issues

There is now an emphasis from government and public sector bodies such as the Inter-departmental Group for Geographical Information (IGGI) and the Association of Geographical Information (AGI) respectively on making the most efficient use of geographical information by sharing data. However, although the use of geographical data is rapidly increasing, our understanding of associated data processing error, especially for the integration of multiple spatial data sets (which is a typical application of GIS as exemplified by the Evaluation of the Safer Cities Programme) lags far behind (Lunetta *et al.*, 1991). As the spatial data sets become a common source for many applications, accurate data handling using GIS become increasingly important. Indeed, based on a literature review and telephone interviews with thirty people from various public organisations, Rich (1995) identified the data quality issue as the most serious obstacle to the increased application of GIS for crime control and prevention.

The process of applying GIS to a crime prevention programme and its evaluation is problematic. It involves combining different kinds of data sets that have been acquired at different scales and to different levels of precision. Uncertainty about the accuracy in GIS processing becomes critical for this kind of data integration (Chrisman, 1984, Tomlin, 1991). However, despite the advance of GIS technology, and its impact upon the wider community, error handling in GIS is still a relatively neglected area especially for the 'end-users' who are not GIS experts. Although future GIS may have modules addressing the problems of error, few incorporate such modules today (Drummond and Ramlal, 1992; Goodchild, 1995). Spatial accuracy is a major concern in GIS applications and it demands priority for research initiatives (Goodchild and Gopal, 1989; Lunetta *et al*, 1991; Star *et al*, 1991).

A comprehensive understanding of the nature of geographical information is crucial in the data manipulation process and for the success of GIS as a whole. We need to address questions such as: What are the consequences of bringing together different data sets which may be collected at different scales and times? How accurate is the overlay of different geographical features during the GIS processing? Usually, neither the GIS vendors (developers) nor the users are willing to take an extra effort to address issues of accuracy. This research 'travels the extra-mile' and attempts to explore some of these issues further.

1.3 Structure of the thesis

Conceptually, the structure of this thesis follows the classic research process:

1. reviewing literature (Chapter 2, 3)
2. formulating aims and objectives (Chapter 4)
3. developing methodology (Chapter 5)
4. implementing the method (Chapter 6, 7)
5. presenting and discussing the results (Chapter 8-11) with a conclusion (Chapter 12)

However, the large scale of the Safer Cities Programme Evaluation, and both the multi-disciplinary and inter-disciplinary nature of this research have made the task of assessing the GIS accuracy challenging. In a sense I am attempting to 'ride two horses' at the same time: crime prevention practices and the GIS methodology. Each is an inter-disciplinary area. The literature review reflects this dual-concern. There are three blocks of literature review scattered around the first half of the thesis. Two of these are about the context of the case study. This is due to the dual nature of the context, that is, the Safer Cities Programme evaluation (Chapter 2) and the issues within it: the application of GIS (Chapter 3). In particular the GIS accuracy issue was singled out as one of the most important areas for detailed analyses (Chapter 5).

From an initial literature review of spatial accuracy issues, a *classification of error groups* is formulated which provides a framework for the subsequent exploration (Chapter 3). Chapter 4 sketches the outline of my research by mapping the general framework developed in Chapter 3 upon the specific context of the case study described in Chapter 2. It describes the *aim, objectives, and scope* of this research. A further literature review was carried out to formulate the *methodology* used in this research (Chapter 5). In particular, it provides a substantive review of the methods used in spatial error handling. Chapter 6 provides a quick but *preliminary assessment* of the accuracy in the Evaluation while Chapter 7 describes the detailed *implementation* of the more accurate but labour-intensive method. Chapters 8-10 present the *results* in a step by step fashion and show how the household counts, action intensity and the outcome of the burglary risk are affected by the accuracy assessment. Chapter 11 reports *testing* to see whether the results described above are significant or not by referring back to the context of the Safer Cities Programme Evaluation and examines how the results may affect its conclusion. Finally Chapter 12 provides the *conclusion* of this research, discusses its implication and indicates future research. Figure 1.2 shows the dynamic structure of this study which readers may find useful as a guide.

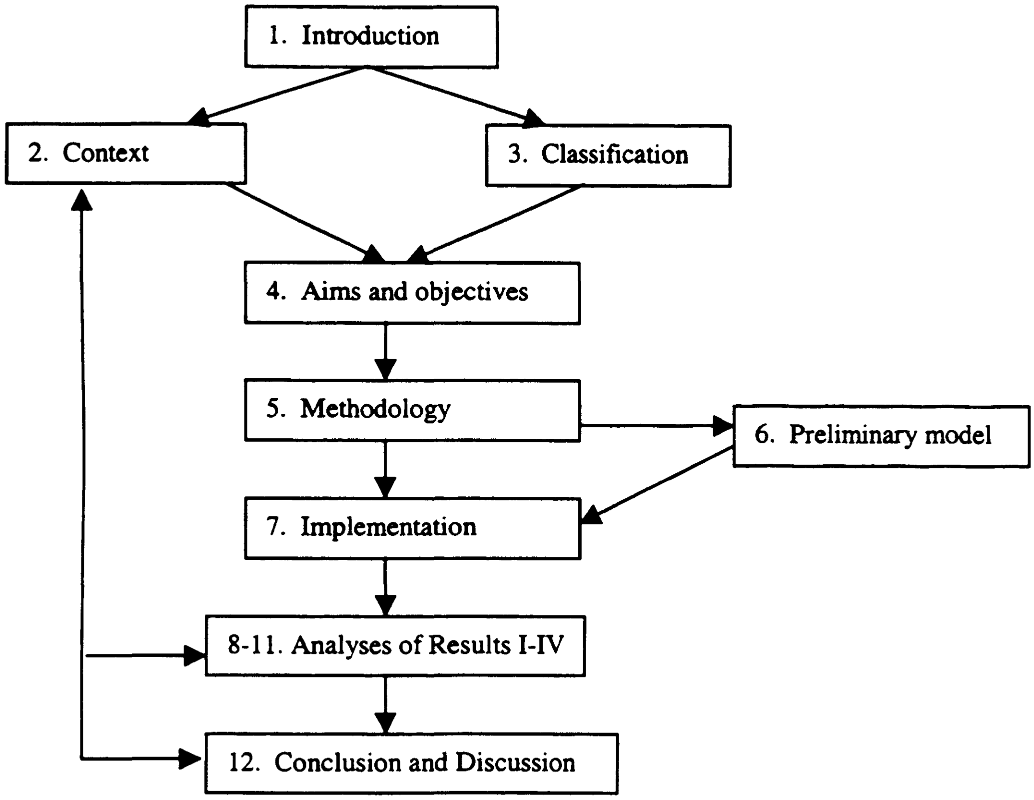


Figure 1. 2: Structure of the thesis

The next step is to examine the historical precedence for the Evaluation of the Safer Cities Programme upon which the case study is based. This is provided within the terms of references of Chapter 2.

Chapter Two

GIS within the context of the Safer Cities Programme Evaluation

The details of the Evaluation of Safer Cities Programme has been published in Home Office Research Study (164) on the *Safer cities and domestic burglary* and included in this thesis as addendum (Ekblom, *et al* 1996a). This chapter provides a summary description of the Evaluation. This is necessary as it provides detailed terms of reference and the historical precedence for this case study. Many aspects of the Programme are glossed over, but certain issues, in particular the technical details, are expanded. First, Sections 2.1 and 2.2 describe the historical context of the Phase one Safer Cities Programme and its evaluation, respectively. The rest of the chapter covers the evaluation methods from several perspectives. It starts with a general description of the evaluation strategy (Section 2.3); then more specifically the Scoping and Scoring principles used which are central to both data analyses and GIS processing in the evaluation. Section 2.5 describes the various sets of data used, the conceptual linkages between the data which centre on a spatial database, and the practical linkages and storage in terms of hardware and software arrangements. Section 2.6 describes the selection of a specific crime type (domestic burglary) for the Evaluation and this research. The outcome of the Evaluation of the Safer Cities Programme is summarised in Section 2.7 for domestic burglary. Finally Section 2.8 discusses the issues of spatial accuracy in the Evaluation which this research attempts to address.

2.1 The context of the Safer Cities Programme in England

Phase one of the Safer Cities Programme was inaugurated in 1988 and wound up in Autumn 1995. Altogether, it cost about £30 million, including £8 million administrative costs. Substantial levered-in funds were also obtained from other sources. The Safer Cities Programme was set up by the Home Office as part of the British Government's wider programme: *Action for Cities*, initiated by the former Prime Minister Margaret

Thatcher to deal with the multiple social, physical and economic problems of some of the larger urban areas. The objectives of Safer Cities Programme were to:

1. *reduce crime;*
2. *lessen fear of crime; and*
3. *Create safer cities within which economic enterprise and community life could flourish.*

Most Safer Cities initiatives were based locally. These were developed from the 1980s concept that crime is best tackled at the local level. The initiatives also adopted a 'partnership' or multi-agency approach to crime prevention as exemplified by an earlier programme, the so-called 'Five Towns' initiative (Liddle and Bottoms, 1992; and Ekblom *et al*, 1996a).

Twenty English cities or boroughs (the so-called *Safer Cities* for short) were chosen for Phase one of the Programme implementation on the basis of the annual crime statistics collected by the Home Office. Four of the twenty cities were selected as 'pilot projects' to establish working procedures for the implementation of the Safer Cities Programme. The subsequent 16 cities of the above were included in the evaluation. These were: Birmingham; Bradford; Bristol; Coventry; Hartlepool; Hull; Islington; Lewisham; Nottingham; Rochdale; Salford; Sunderland; Tower Hamlets; Wandsworth; Wirral; and Wolverhampton (See Figure 2.1 for map).

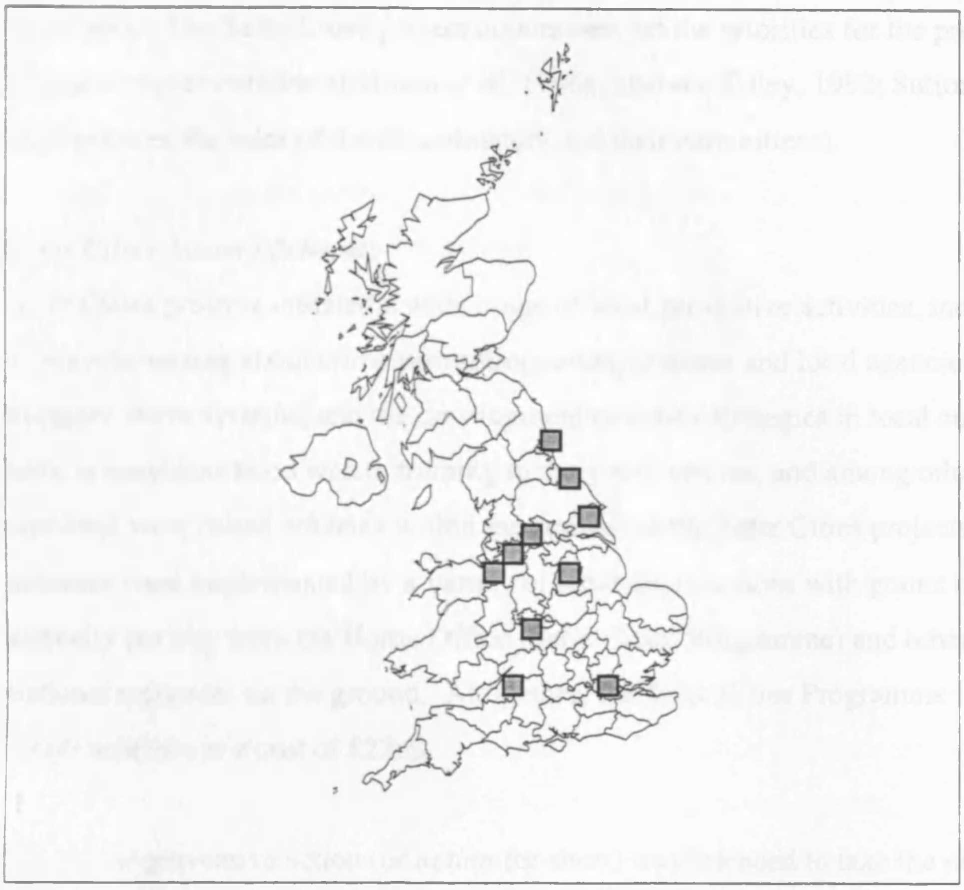


Figure 2. 1: Safer Cities in England

Safer Cities in the county areas (each symbol may consist of more than one Safer City)

A local group was set up, which consisted of a co-ordinator and a small team, in each of the Safer Cities. The term *Safer Cities project* was used to refer to such local group and their activities in the context of the Programme. Each co-ordinator was guided by a steering committee representing local government, police, probation, voluntary bodies and commerce. The Safer Cities project co-ordinators were recruited locally and drawn from a wide range of backgrounds, including police, social work, probation and local government. They were given some initial training and support from professionals in the Home Office and elsewhere. They were also provided with an initial 'crime and social profile' of their local areas, by the Research and Statistics Department of the Home Office (now known as Research Development and Statistics Directorate, RDS). These profiles included a beat-by-beat picture of recorded crime rates and were time-consuming to produce. The Safer Cities project committees set the priorities for the project and oversaw implementation (Ekblom *et al*, 1996a; also see Tilley, 1992; Sutton, 1996 for a discussion of the roles of the co-ordinators and their committees).

Safer Cities Action / Schemes

Safer Cities projects initiated a wide range of local preventive activities, including awareness-raising about crime prevention among citizens and local agencies such as burglary alarm systems, and the development of safety strategies in local communities such as neighbourhood watch, training in the youth centres, and among others. These activities were called *schemes* within the context of the Safer Cities projects. These schemes were implemented by a variety of local organisations with grants up to £250,000 annually per city from the Home Office (Safer Cities Programme) and other local or national resources on the ground. Altogether, the Safer Cities Programme initiated some 3,600 schemes at a cost of £22m.

The crime preventive action (or *action* for short) was intended to take the rational, problem-oriented approach developed in the 1980s and 1990s (Tilley, 1993b; Laycock and Tilley, 1995; Sutton, 1996). The term 'scheme' is usually used to describe the crime preventive activities within the context of the Safer Cities Programme. The term 'action' is used specifically to describe the nature of the preventive action within the context of its Evaluation. This 'preventive process' involves the following steps:

1. Identify local crime patterns by analysing crime data and the contextual information.
2. Set objectives.
3. Adopt appropriate preventive measures (tailor-made rather than off-the-shelf).
4. Implement action.
5. Evaluate what has been done.
6. Make changes where necessary.

The schemes deliberately addressed a wide range of crime problems using various methods. The crime problems were classified according to the Home Office's catalogue of crime statistics. These included: assault, domestic violence, sexual assault, domestic burglary, commercial burglary, criminal damage, robbery, fraud, vehicle-related theft, shop theft, theft on person and other theft.

The majority of schemes focused on crime reduction (Programme Objective 1), while others addressed fear of crime (Programme Objective 2). Some schemes focused on the city as a whole (eg, through publicity campaigns, information initiatives such as crime prevention buses, or multi-agency programmes), and some schemes focused on vulnerable individuals, groups of homes, particular institutions (such as schools and clubs), or particular localities (eg, housing estates, car parks or city centres).

Preventive methods can be classified as either 'situation' or 'offender-oriented'. The situation preventive action comprised measures such as better security hardware, alarms, improved lighting, and surveillance measures. The offender-oriented schemes covered youth work, holiday play schemes, credit unions, adventure playgrounds, employment advice, even morality plays in schools. Out of the 3010 schemes classified, 30% of the proximal components are offender-oriented, and 70% situational (for more detailed classification, see Law and Ekblom, 1996).

2.2 The evaluation of the Safer Cities Programme

Within the British socio-political context at that time, the Safer Cities Programme coincided with the Government's Financial Management Initiative, and was subject to scrutiny in terms of value for money (cost-effectiveness). The Programme impact evaluation (required by the Treasury) was carried out by a team of social researchers and scientists within the former Research and Planning Unit (part of the Home Office RDS). I was one of the scientists in the Evaluation Team responsible primarily to implement the evaluation strategy using GIS. The evaluation strategy was designed by the team leader Dr Paul Ekblom (Principal Social Researcher).

Evaluation of the Safer Cities Programme is not to be confused with the local *process evaluation* conducted by the co-ordinators themselves. The latter was for ensuring that at least a minimal assessment was made of each scheme funded. The local process evaluations were part of the conditions of grant (see Youell's, 1993, evaluation guide). As such, they were carried out at a number of levels, for different purposes. The standards of the local process evaluations varied and were difficult to compare. For example, Tilley (1993a) evaluated a number of 'themes' such as Safer Cities schemes using CCTV in car parks and domestic burglary. The success which projects had in fostering local community safety strategies was also assessed (Tilley, 1992). This was to facilitate the continuation of local co-ordinated crime prevention after the Safer Cities projects closed as planned. Process evaluation assembled good practice information, used detailed retrospective case studies of ten selected burglary schemes was complementary to the much larger-scale Programme evaluation (see Tilley, 1993a; and Tilley and Webb, 1994 for details of the process evaluation).

In contrast to the process evaluation, the focus in the RDS study was on the impact of the Safer Cities Programme as a whole (and hence so-called *impact evaluation*). We attempted to answer two key questions in terms of the Programme Objectives:

1. Was there any change in crime and fear of crime in the Safer Cities project cities?
2. If so, to what extent can this change be attributed to the effects of Safer Cities actions, as opposed to other causes?

This was an extremely challenging task in terms of linking measures of *Safer Cities action* to measures of *outcome*, and collecting and integrating the wide range of data sets. Changes in crime were likely to be influenced by many local factors, and by background trends at city and national level. All these could mask any impact of Safer Cities and thus needed to be taken into account as much as possible. Furthermore, many schemes were small in resource terms, or spread thinly over large areas. Their individual impact was likely to be modest. It was best to consider a large number simultaneously. However, with the *conventional* single-scheme study, the location of the scheme, its timing, target crime problem (eg burglary) and target victim population (eg elderly residents) are all known in advance, a 'bespoke' evaluation can be designed around these parameters. By contrast, an 'industrial-scale' impact evaluation would require a general-purpose approach, given the number, variety and size range of the schemes (for example setting up youth activities, improving the security of doors and windows on entire council estates, funding all-female taxi services, or installing street lighting outside a retirement home). (See Ekblom, 1990; Ekblom *et al*, 1994; Ekblom and Pease, 1995; Polder, 1992; and Junger-Tas, 1993 for wider discussions of the difficulties of evaluating crime prevention initiatives.)

2.3 The Programme impact evaluation strategy

In answering the questions set out in the above (Section 2.2), we linked measures of preventive action to measures of outcome (crime surveys of residents, recorded crime statistics) at the small area level and looked for change in outcome differentially associated with the presence of action. We had to cope with the major difficulty of *not* knowing in advance where and when the schemes were to be implemented within the Safer Cities. To minimise the risks of delivering inconclusive findings, and to conduct a 'fair test' which balanced the risk of mistakenly reporting success of Safer Cities against that of mistakenly reporting failure, the strategy devised was ground-breaking in several ways, and as far as we know, represent a pioneering approach (Ekblom, 1992; Ekblom and Pease, 1995, Law and Ekblom, 1994a & b; Ekblom *et al*, 1994; 1996a). This involved several steps:

1. Devise in advance of knowing where local schemes were to be implemented, a strategy for sampling small areas within each city which would give a good chance of hitting the eventual preventive action.

2. Combine diverse sources of data covering different territorial units (EDs, beats, 'neighbourhoods') using a GIS and its associated relational database.
3. Retrospectively, when the location of action was known, compare changes in crime in areas where there had been more action with areas with less or none (see *internal comparison*).
4. Compare changes in the Safer Cities with similar cities not in the Programme (see *external comparison*).
5. Consider the collective impact of a large number of schemes simultaneously (rather than conducting a series of single-scheme studies) to increase the likelihood of detecting the weak signal of the 'Safer Cities effect' against background noise in the form of strong local random fluctuation of crime levels.
6. Take a global view of Safer Cities Programme impact through a 'dose-response' analysis incorporated within statistical modelling (the dose = the input of action within each small area; the response = change in outcome measure, such as the incidence of burglary, in the same small area).
7. In support of Step 6, develop ways of identifying which scheme had the potential, if it worked, to affect which outcome measure (see '*scoping*').
8. Quantify the input of preventive action per measurement site, that was potentially detectable by the relevant outcome measure (see '*scoring*').
9. Use *multi-level modelling* to explain variation in outcome measures such as crime victimisation or fear; in particular to take separate account of individual explanatory factors (such as survey respondents' age or tenure) and area explanatory factors (including the score of preventive action and also contextual factors such as unemployment).

External comparison

Areas receiving Safer Cities action at some point over the three-year period were compared with those which did not. A set of nine (carefully matched) *comparison cities* was also examined to provide a picture of more general national trends in similar urban areas, over six years (including three years prior to the Safer Cities Programme). These cities were matched to Safer Cities equivalents by four *'family groups'* taken from a classification of local authority districts based on the 1981 Census (Craig, 1985). They were also selected for comparability of total recorded crime rates over the period 1986-90.

To reduce cost and effort, areal unit-level data for the comparison cities were not collected. Instead, their *city-level* annual figures were used to construct two 'indicators' of crime rates. A *global indicator* was based on all comparison cities, with burglary incidence risk weighted to adjust the population composition by family group (in the comparison cities) to the composition in the Safer Cities. There was, however, considerable variation in crime trends observed between the family groups. Therefore, a *family indicator* was calculated separately for each family group of Safer Cities, based on the appropriate comparison cities. The comparison cities for the crime statistical analyses include: Barnsley, Burnley, Hackney, Haringey, Liverpool, Manchester, Oldham, Leeds and Southwark.

Internal comparison

For internal comparison, the outcome measure (such as the burglary counts) was converted into *incidence rates* per 100 households. For crime statistics, the standard beat maps were digitised by professional consultants (GDC Ltd). The spatial data were merged with population data from the 1991 Census on GIS by means of simple overlay operations. In effect, each beat was 'tiled' with the Census data from the EDs which most closely approximated its territory. This process also enabled us to link to the beats contextual data from the Census and its derivatives (such as the Index of Local Conditions, which is a set of measures of deprivation; DOE 1995). (Multi-level modelling is explained in detail later in the Methodology, Chapter 5).

2.4. Scoping and Scoring principles

The scoping and scoring principles mentioned earlier (Step 7 and 8 in Section 2.3) are not only essential for quantifying the Safer Cities action, but also central to the use of GIS in the Evaluation. Furthermore, they provide a way of removing some of the irrelevant factors that derive from the vicissitudes of measurement, which otherwise would obscure relationships in the statistical modelling. This section re-describes the principles in detail.

Scoping

Scoping operates on the data to determine whether a given preventive scheme, *if it works*, can be said to have the potential to influence a given outcome measure. There are four dimensions to scoping (Ekblom *et al,* 1994):

- *Space* - is a given scheme located close enough to a survey sampling point for the outcome measure (eg incidence of burglary) taken at that point to be potentially influenced?
- *Time* - is the timing of the action such that it was able to exert an influence on the outcome measure?
- *Problem* - space and time considerations apart, is there a plausible cause-effect relationship between the nature of the action in a given scheme, and the crime or fear problem covered by the outcome measure in question?
- *Subgroups* - is a given scheme directed towards a particular population subgroup - eg women? Does the outcome measure cover this subgroup (eg in a women-only question in the survey)?

To implement the scoping principle, we specified each outcome measure in terms of a Zone of Detection (ZD) whose dimensions correspond to the four aspects just outlined, and each scheme in terms of a Zone of Influence (ZI) characterised by the same four dimensions, and looked for the zone of overlap (ZO). To be in scope, a scheme had to overlap with the outcome measure on all four dimensions simultaneously. Figure 2.2 shows the relationship between ZI, ZD, and ZO.

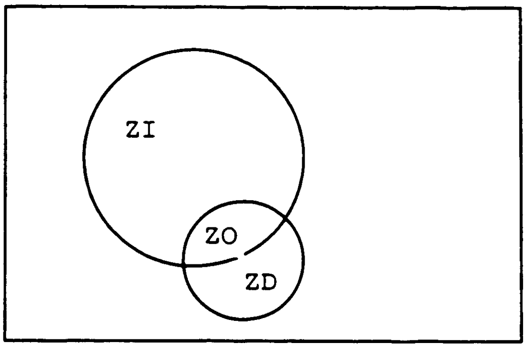


Figure 2. 2: An idealised diagram representing the scoping process

We chose the smallest feasible units for defining the space and time dimensions of ZD, ZI and ZO - namely the 1991 Census Enumeration District (ED) and the month, for the Survey, and beat and year for the crime statistics.

Scoring

Our outcome measures cover for example the risk of being burgled over a particular time period, averaged over a group of respondents sampled in a particular geographical area (e.g. an ED as the ZD). The risk is averaged as a population estimate per 100 *households* in the ZD, for burglary, say; or per *individual* in the ED, for assault. Generalising, either households or individuals are the *unit at risk of victimisation*. To link this outcome measure to dosage in a conceptually tight way suitable for statistical modelling, we devised a score representing the dose of input of preventive action received, for the *same* units - i.e. the *input per unit at risk of victimisation.*

Since we did not know exactly *which* household or individual has received any input from a given scheme, we worked with *averages* on this side of the equation too. These averages operated both over the populations and households contained within spatial territories, and over the time periods of measurement. On the *spatial/geographic* side, a particular scheme's ZI might be much bigger than the ZD with which it overlapped, and the ZD would then only receive a proportion of the total input. Therefore, we needed to adjust the input score down to reflect the share-out of the input over the wider area. Likewise, only part of the ZD might be affected by the scheme. In this case we had to dilute the input further to reflect the reduced probability that any particular respondent received the action. Both geographical share-out and dilution were calculated using population data from Census Small Area Statistics relating to the ZI, ZO and ZD. For instance, if a scheme was directed at women only (eg sexual harassment), then we would use the adult female population as the base; if at households (eg residential burglary), then we would use households. On the *temporal* side, we again had share-out and dilution to adjust for since for example a scheme might only have started up halfway through the measurement period (see Ekblom *et al*, 1994 for details). Altogether, then, we adjusted the original total financial input (A) to a scheme by factors representing geographic *share-out* of action and *dilution* of measurement, and temporal share-out and dilution i.e. the action score (S) = A x Share-out x Dilution / temporal share-out and dilution.

Since Share-out = N_{ZO} / N_{ZI} and Dilution = 1/ N_{ZD}

where

N_{ZO} is population-based in the Zone of Overlap

N_{ZI} population-base in the Zone of Influence

N_{ZD} population-base in the Zone of Detection

by substituting the values of the share-out and dilution into the above, Equation 2.1 shows scoring calculations.

$$S = (A \times N_{ZO} / N_{ZI} / N_{ZD})^{-T^*} \qquad \textbf{(2.1)}$$

where

S is action score

A amount of action in £

T* number of years the action took place which is subject to the same principle of share-out and dilution but in temporal scale

The contract for implementing the scoping and scoring principles within a GIS (using ARC/INFO) was awarded to North-West Research Regional Laboratory (NWRRL) at Lancaster University in consultation with the Home Office Evaluation Team. The scoping and scoring system was developed using the ARC Macro Language (AML), a high level language which allowed not only the compilation of ARC/INFO command sequences but also the use of a range of programming constructs such as looping and logical operations as well as the creation of menu interfaces. The final version of the scoping and scoring system was eventually transferred onto the Home Office GIS, updated and managed by me.

2.5. The structure of the spatial database, data sets, hardware and software

The realisation of the evaluation strategy described earlier (Section 2.3, 2.4) required the use of state-of-the art computing centred around a GIS, and the integration of a large number of different data sets, both spatial and non-spatial units (Ekblom *et al,* 1994; and Law and Ekblom, 1994a & b). This involves linking the following different types of measure (over time and at the small-area level such as enumeration districts in up to 30 cities):

1. action data (what preventive schemes are located where);
2. outcome data (including monthly recorded crime totals per police beat over a 5 year-period, and before- and after- surveys of people's perception of safety and experience of victimisation); and
3. covariate data (principally from the Census Small-Area Statistics at area level and individual level - survey demographics).

Outcome data

The recorded crime data (as the outcome measure for the Programme Objective 1, Section 2.1) were collected for up to twelve major offence categories per beat, from 14 of the 16 Safer Cities evaluated from 1987, the year before the Safer Cities Programme began, to 1992 (there were problems with data supply in the other two, Wandsworth and

Islington). Beats were on average about ten times the area and the population of the EDs used in the survey. The before- and after- surveys of people's perception (as the outcome measures in terms of Programme Objectives 1-3) was carried out by MORI Ltd under a contractual arrangement managed by the Home Office Evaluation Team.

In terms of data analyses, all three types of data were linked together through multiple regression within multilevel modelling – a relatively new statistical technique which enables simultaneous examination of changes in crime risk over time in households, localities and cities (Ekblom *et al,* 1993 and Chapter 5 for detail). The Safer Cities action and covariate measures were used to explain variation in the outcome measures (such as the risk of being a burglary victim). The various data sets were linked relationally, most through INFO within ARC/INFO. Figure 2.3 shows these relations, Figure 2.4 the computing arrangements. The relations were described using the so-called EAR (Entity-Attribute-Relationship) model (Everest, 1986). Composite keys (attributes common to different data sets) enabled action, outcome and context data to be linked.

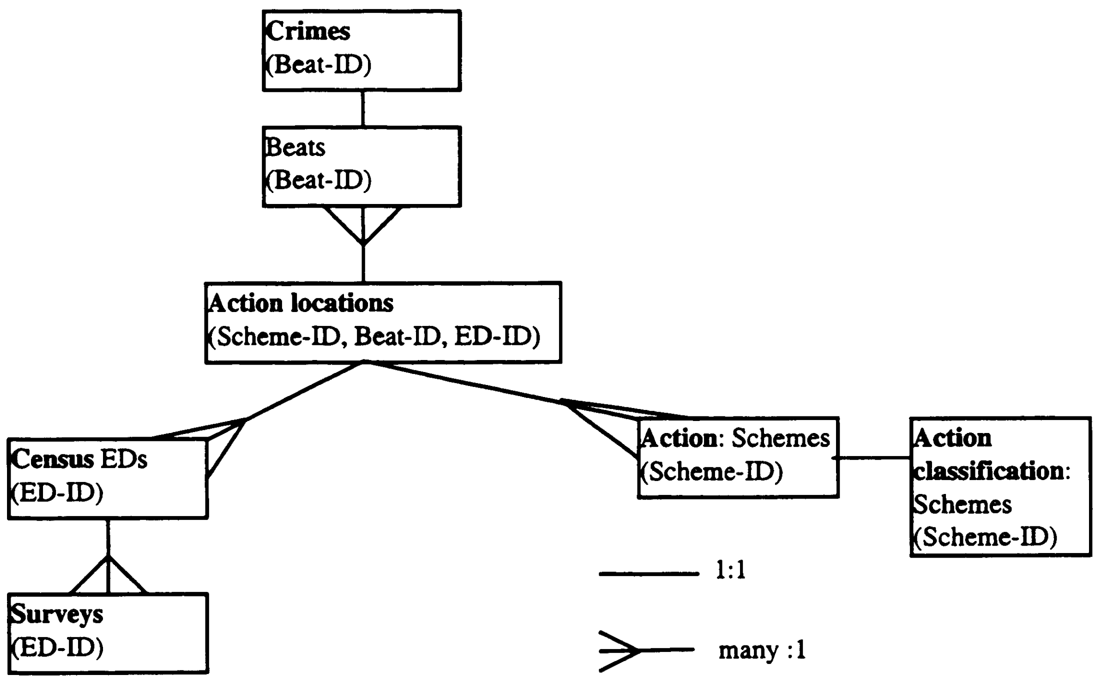


Figure 2. 3: The E-R diagram of the spatial database

Entities and attributes of the various data sets (with key attributes underlined) include:

- **Action:** Schemes (<u>scheme-ID</u>, dates, financial input, implementation status, target crime type, target victim type etc.)
- **Action Classification:** Schemes (<u>scheme-ID</u>, methods, mechanisms etc.)
- **Action location:** Action locations (<u>scheme-ID</u>, <u>beat-ID</u>, <u>91ED-ID</u>)
- **Crimes:** Crimes (<u>beat-ID</u>, month, crime_types ..., area, perimeter)
- **Surveys:** Survey (respondent-ID, <u>ED91-ID</u>, wave, postcode, respondent demographics, responses....)
- **Census:** 1991 EDs (<u>ED91-ID</u>, counts)

Two topological data sets were also required to make various spatial linkages:

- <u>ED-91 boundary data</u> bought commercially
- <u>Police beat boundary data</u> digitised bespoke

A compete set of the Entity-Attribute Tables is listed in Appendix 2 in INFO format. The architecture of the computer arrangements for data storage and manipulation is shown in Figure 2.4. The action, action location, Census and crime data were transferred from various sources into an INFO database using ARC/INFO on a VAXstation 4000.60 under VMS. I developed a custom-built, user-friendly, front-end system for data manipulation and retrieval with the consultant support from ESRI-UK Ltd. I further developed a menu-driven classification system to categorise the action data within ARC/INFO. The Census data were originally stored on the VAX under VMS and processed by means of C91 on a personal computer (PC DEC 386), before transfer into INFO. The survey data was handled using SPSS on the VAX. Reassignment from 1981 to 1991 ED was done through importing and exporting of files to NWRRL's own GIS. The surveyed EDs were represented in INFO for extracting demographic covariates from the Census, and for scoping and scoring; likewise the various time periods covered by the surveys. Scoping and scoring, as described earlier, was carried out within ARC/INFO and all the data sets were brought together for statistical modelling using the ML3 software on another personal computer.

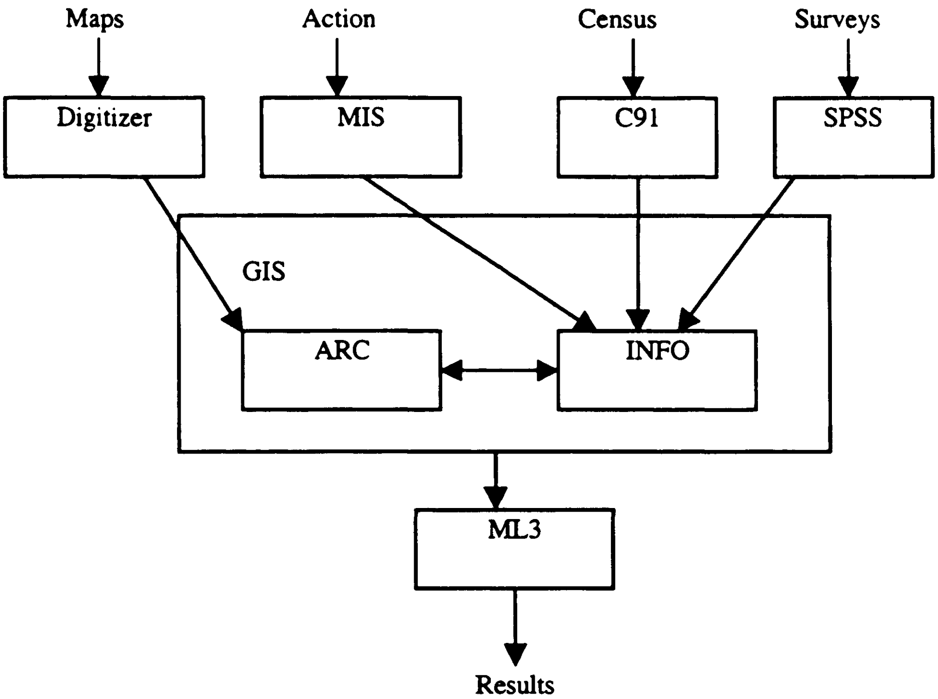


Figure 2. 4: System architecture for the Evaluation

2.6 Selection of a crime type for the Phase One evaluation: Burglary schemes

To maximise its possible measurable impact on that crime type, we selected a particular crime type for Phase one reporting of the Safer Cities Programme. To select a particular crime type for detailed analyses, the following factors were considered:

- the frequency of co-ordinators targeting
- the development of preventive practice
- the locality (localised effect).

The simple descriptive statistics of the action using the Safer Cities Management Information System would provide a picture of the above factors. Figure 2.5 shows that, by 1995, of the 2,300 Safer Cities schemes in all 16 cities, over half were targeted on dwellings. Furthermore, a third of the schemes were targeted on burglary (Figure 2.6). The values represent total specified funds in £ thousands spent on that category.

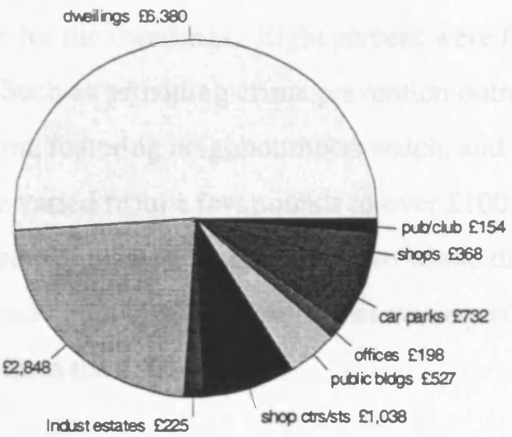


Figure 2. 5: Physical targets of the Safer Cities action (based on Ekblom *et al*, 1996a)

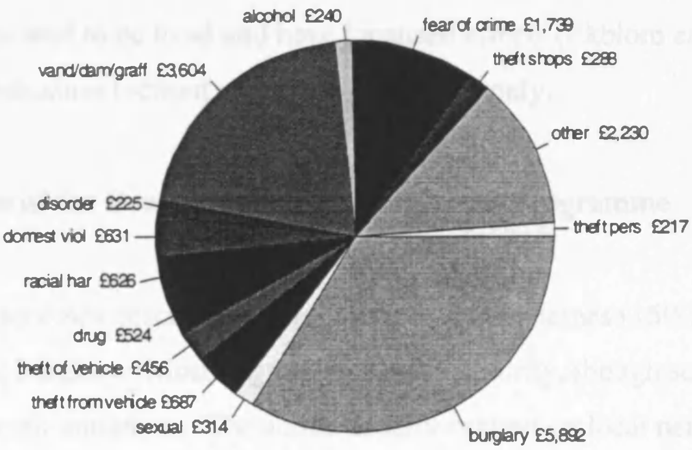


Figure 2. 6: Target crime types of the Safer Cities action (Ekblom *et al*, 1996a)

Altogether, some 500 schemes were targeted on domestic burglary. Some £4.4m of Safer Cities funds were spent (excluding the further levered-in funds or in-kind assistance from other local or national sources). Nearly 300 schemes were targeted on domestic burglary at the local level. A further 62 schemes, such as publicity campaigns, were targeted at city level. Of the local schemes, three-quarters focused on domestic target hardening. This included door, window and fencing improvements, entry systems, and security lighting around individual houses or blocks. A number of weaknesses were tackled together in a security package for the dwellings. Eight percent were focused on community-oriented action. Such as providing crime prevention outreach workers, raising awareness of prevention, fostering neighbourhood watch, and property marking. The amount spent per scheme varied from a few pounds to over £100,000. The areas which schemes covered ranged from single blocks of flats to whole districts. On average about 5,200 households were covered. Geographically this was equivalent to 26 Enumeration Districts (EDs) from the 1991 Census.

Schemes such as publicity campaigns targeted at city level would not be the subject of the evaluation as their 'thin spread' was unlikely to have had significant impact that was measurable locally. Thus burglary on dwellings was chosen for detailed analyses because co-ordinators often targeted it, preventive practice is relatively well-developed, and burglary schemes tend to be local and have localised effects (Ekblom *et al*, 1996a). Thus the Phase one evaluation focused upon burglary action only.

2.7 The outcome of the Evaluation of the Safer Cities Programme

Out of the 3600 schemes described above, just over 500 schemes (15%) were set up to prevent domestic burglary. Most upgraded physical security, though some mounted community-oriented initiatives. The action usually centred on local neighbourhoods or estates. The results of nearly 300 of the schemes showed that the Safer Cities Action reduced burglary and was cost-effective. Simply implementing action in a police beat reduced local risks by nearly 10%. Physical security measures against burglary seemed to work independently. Community oriented activities required reinforcement with action against other types of crime in general. While the overall cost of each burglary prevented ranged between £300 and £900 in low and high-crime areas respectively, the average financial cost of a burglary to the state and the victim was about £1100 (estimated). In total, this represents an estimated saving of £31 million from the 56,000 burglaries prevented.

Reduction in burglary risk was greater where there was more intense burglary action but to achieve these bigger falls cost disproportionately more. 'Marginal cost' estimates per extra prevented burglary ranged from about £1,100 in the highest risk areas to about £3,300 in the lower risk ones. In monetary terms extra expenditure was justified only in high-risk areas. However, it would be better to increase the action from low to moderate (or intensive) to prevent displacements of both crime types and geographical locations. In the areas with low intensity action, it seemed either that some burglaries were displaced to nearby neighbourhood areas or that burglars switched to other property crime. Furthermore (according to the surveys), people were more worried than before if they were aware but it was low level action. When action was moderate (or intensive), not only did the adjacent areas benefit from reduction in both burglary and other crime, but also that people's perceptions of their quality of community life improved especially where action was most intensive.

At a 10% incidence level of risk (equivalent to the average prevalence risk in the survey) *the mere presence of Safer Cities burglary action seemed to reduce the risk of burglary by about seven percent.* On the marginal impact, *given the presence of action at the average intensity* (£3.57 per household), *for an additional £1 of action the risk of burglary fell by a further 0.8%.* Step and marginal-intensity effects combined showed an *overall reduction of some ten- percent* at the average action intensity. (For further information, see Ekblom *et al*, 1996a.)

2.8 Issues of Spatial Accuracy in the Safer Cities Programme Evaluation

Although the results of the Evaluation shows that the Safer Cities Programme has an effect upon crime reduction such as burglary risk, the complexity of the data manipulation involving GIS operations in the evaluation process raised several data quality issues.

First, the data collection of such large-scale research was problematic. There were missing data in some of the Safer Cities and data input was not an error-free process for such a large amount of data sets.

Second, in order to produce beat-level data for the analysis from the Census and the Index of Local Conditions, the beat boundaries were digitised from beat maps, then overlaid with the smaller EDs using ARC/INFO. However police beats usually bore no relationship to other administrative territories such as wards or EDs. The boundaries of beats and their constituent EDs do not always match to form a one to many relationship (Figure 2.7).

Figure 2. 7: An example of Beat-ED overlay operation in one of the Safer Cities (Coventry, West of the city centre, beat boundary – purple, map scale: 1 cm = 0.3 km)

Third (and similar to the second) the actions and survey interviews were linked relationally via ED-ID using GIS. However, the survey originally used 1981 EDs as a spatial unit (because the Before Survey was carried out prior to the 1991 Census), while the rest of the data used the 1991 EDs. It is well known phenomenon that Census boundaries change. This required transformation of the 1981ED-IDs to 1991ED-IDs. Initially this was achieved by a look-up table between the two supplied by OPCS, but we found the cut-off points between their overlays insufficiently accurate for our purposes, and the fact that the relationship between 1981 and 1991 EDs was often not 'one-to-one' but 'many-to-many' raised the question of spatial accuracy of these data transformations.

Fourth, the police beat boundaries changed over time (for example, Bristol). Fortunately the change tended to be the amalgamation of smaller beats into a larger beat (the so-called super-beat for the purpose of spatial database management). The beats varied widely in size and population, with averages of 230 hectares and over 2,200 households. 'True' beats averaged 180 hectares and 1,700 households; superbeats 600 hectares and 6,300 households. The fact that beat boundaries change might seem a simple administrative task and to have required no more than keeping the spatial database up-to-date. However this did raise the question of the basic unit of analyses for crime statistics. It is related to a well-know problem - called the Modifiable Areal Unit Problem (MAUP).

The MAUP is well documented by Openshaw (1984), and Openshaw and Taylor (1979). It is composed of two problems: 1) scale problem 2) aggregation problem (Openshaw, 1984). The scale problem is the variation in results when data from one unit (so called zone) are aggregated into fewer and larger units (so called regions) for analysis, for examples, when EDs are aggregated into Wards, Districts, and Counties, or in our case, EDs into Beats. The aggregation problem is the variation in results due to the possible alternative combinations of areal units of analysis at equal or similar scales (for example, when the number of units is constant).

A further problem is that of areal interpolation since the ability to define alternative aggregations of the same area means that statistical measures need to be interpolated from one aggregation to another (Openshaw and Taylor, 1979; Goodchild and Lam, 1980; Lam, 1983; Openshaw, 1984; Fotheringham, 1989; Flowerdew and Openshaw, 1987; Goodchild *et al*, 1993). Thus, MAUP may be called MAUS - the Modifiable Areal Unit Syndrome since it consists of more than one problem as well as their related implications such as ecological fallacy problem (Law and Fisher, 1995). An ecological fallacy occurs when conclusions based on aggregate zonal (or grouped) data is applied to the individuals within that zone assuming they are homogeneous when they are not. The ecological fallacy due to the MAUP known as 'aggregation bias' is also well reported in the field of criminology (Langbein and Lichtman, 1978).

The MAUP also highlights a basic spatial representational issue. It raises the question as to how many zones a given region should represent in order to present a meaningful map. For example, Monmonier and Schnell (1984) show how different maps could be presented from the same sets of data by adjusting different selections of choropleth class intervals. While there may be no GIS error involved, there is the question of interpretation errors.

Thus if users were not aware of all of the above issues, the problems of the spatial accuracy might remain hidden. These might cast doubts upon the validity and reliability of the conclusion of the Evaluation. These issues call for further research in the spatial accuracy of the data processed for the analyses of the Evaluation which this study addresses. These issues will be further explored in a more specific way and form the scope of the research presented in the thesis (Chapter 4). First, more general issues on spatial accuracy will be examined in the next chapter.

2.9 Chapter summary

This Chapter has described a £30 million Programme called Safer Cities between 1988 and 1995, and its evaluation by the Home Office RDS. The evaluation strategy involving so-called scoping and scoring principles, and the use of GIS, have also been described. Domestic burglary on dwellings was chosen for the impact evaluation. The evaluation has showed that the Safer Cities Programme reduced burglary and was cost-effective. Simply implementing action in a police beat reduced local risks by nearly 10%. In total, this represents an estimated saving of £31 million with 56,000 burglaries prevented. However, the complexity of the data manipulation involving GIS operations in the evaluation process raised several data quality issues which may affect the certainty of the conclusion of the evaluation. These issues will be explored further in the rest of this thesis.

Chapter Three
Issues and classification of spatial errors

Before a comprehensive account of the errors can be quantified for the evaluation of the Safer Cities Programme, it is necessary to examine all the possible sources of error in the GIS processing in general. To develop a system for the complete classification of spatial data quality is beyond any single research study. This chapter examines the sources of error and uncertainty in the geographic information. In doing so an attempt has been made to create a classification system for error. This is intended to provide a framework for the case study and leads to a standardised way to specify the quality of spatial data in the future. There are several reasons why there is particular value in producing a common classification of error. First, it would be helpful in promoting research on error links between disciplines. As the GIS community grows (as described in Chapter 1), researchers and practitioners from different disciplines may use the same terms to mean different things. Agreeing on a common terminology would have a unifying force across the GIS community as well as assisting understanding between different disciplines such as Geography, Statistics and Social Sciences. Secondly, an agreed classification would focus more attention on error issues, encouraging the design of GIS and administrative spatial data sources which minimise errors and enable them to be measured. Third, the classification would encourage GIS users to monitor errors over time.

The ideal goal of the classification is to promote a common framework throughout the GIS community. This in turn will promote a common language on error issues and enhance understanding of this area between different research applications to be more readily appreciated, and encourage research from each to be more widely applied. However within the scope of this case study, what is necessary is to outline different usage of error processes and define a framework to enable similarities in methodologies (Chapter 6). The general issues on error and policy within the context of the U.K. government are examined in Section 3.2. A general classification framework (which is applicable to the specific context of this case study) is developed by examining the relevant literature review (section 3.3 - 3.4). Seven spatial data quality components and a software engineering model are critically evaluated for the purpose of this research in Sections 3.5 and 3.6. Finally, a general classification system for errors is proposed in Section 3.7. First, however, to clarify the concept of error, the definition of a few key terms in error description is required (Section 3.1).

3.1 Definitions

Definition is the necessary first step before any measurements, analyses and visualisations of spatial data quality can be made (Taylor, 1995). It is helpful to start with the concept of total *quality* and establish a clear statement of definitions of errors. This is not to equate 'quality' with 'error', but to explore their interrelationship. Although high quality data need not be error-free, erroneous data would imply poor quality of the data. In general, *Quality* is defined as "the totality of features and characteristics of a product or service that bear on its ability to satisfy stated or implied needs" (ISO 8402, 1986). It relates to the terms: *Fitness for use* (Grady, 1993; Moellering, 1984a) or fitness for purpose (Ralphs, 1993). In particular, *data quality* is defined as: "The totality of features and characteristics of a data set that bear on its ability to satisfy a stated set of requirements" (ISO 8402). According to Brassel *et al* (1995) *Fitness for use* of a data set and for an application can be defined as: "The totality of features and characteristics of a data set that bear on its ability to satisfy a set of requirements *deemed to be appropriate to an application.*" [Italic added.] Data quality information is usually provided by the producer of a data set. It should be constantly updated rather than evaluated once in the life time of a data set. Fitness for use as evaluated for each application of the data set, from the users' perspective, is more relevant to the scope of this case study. The ideal data set would have a one-to-one mapping between the product design's need and the application's need. However this is usually not the case in practice. For instance, the mapping between the available digital ED boundaries (commercial product) and its use for overlaying with Police beat boundary for the Safer Cities Programme Evaluation is not 100%. This gives rise to the research problem this PhD attempts to address.

Although data quality has received a considerable amount of attention in the recent GIS literature, its terminology is still not clearly defined. Terms such as accuracy, error and uncertainty are often used interchangeably and confusingly (Petrick, 1980). Different researchers have used the terms loosely in different contexts. For instance, the terms scale, resolution, accuracy and precision are often mistakenly used interchangeably with regard to spatial data (Goodchild, 1993). Strictly speaking, digital databases do not have a 'scale'. They indirectly represent the scale of the source data (e.g. maps). The digital scale thus does not reflect the accuracy of the data set (Fisher, 1991c). Here I would like to clarify this confusion in terminology, and to define the terms used for the case study. First, one needs to differentiate between accuracy, precision and uncertainty. All three terms are related to data quality issues.

Accuracy can be defined as a measure of the extent to which an estimated value approaches the 'true' value (Burrough, 1986). However, the 'true' value is usually not known. Thus it has to be estimated by means of the best available method. Accuracy is a useful concept for defining the data quality, reflecting the difference between the estimated value, and the 'true' value being estimated (Drummond, 1995). The estimate of the 'true' value may even be quantified using statistical techniques. Such an estimate represents the accuracy of geographical information and indicates the probabilistic nature of the data. In other words, although the concept of accuracy may be qualitative, in practice, its *quantity* can be defined and described by means of statistics. The standard deviation of error in statistics can be defined as

$$SD_{ERROR} = \sqrt{\frac{\sum_{1}^{n}(z - z' \overline{z})^2}{n}} \qquad \textbf{(3.1)}$$

where n is the number of observations

z is the estimated value of a variable from observation

z' the true value of the variable

$\overline{Z}$ is the mean error, the bias.

If one assumes a normal distribution and that there is no bias in the error estimation, i.e., $\overline{Z} = 0$, then the normal standard deviation (SD) would be the same as the root mean square error (RMSE). The extent of accuracy (or inaccuracy) can be inferred by estimating the RMSE:

$$\text{RMSE} = \sqrt{\frac{\sum_{1}^{n}(z-z')^2}{n}} \qquad (3.2)$$

The discrepancies between the values of an entity used in an application and its 'true' value (estimated by a more accurate method as discussed earlier) provide an RMSE value which can be used as an indication of accuracy of the value used. RMSE is a good error indicator as it is regarded as a global error measure covering all sources: field measuring, plotting, paper map production, digitising as well as data representation in the computer (Shi, 1994). RMSE will be used as a basic error indicator throughout this case study.

Precision is defined as the degree of detail used in reporting a measurement (Goodchild, 1995). Accuracy and precision are independent of each other. A number could be accurate but imprecise or it could be very precise but completely inaccurate. Clarke and Clark (1995) refer to *precision* as the degree of detail in the feature that can be resolved or separated into its constituent parts. Repeated measurements provide the input for determining the SD of a particular measurement. In this case SD is defined as

$$\text{SD} = \sqrt{\frac{\sum_{1}^{n}(z-\overline{z})^2}{n}} \qquad (3.3)$$

As in (3.1) except $\overline{Z}$ now is the mean observation.

If there is no gross or systematic error, both its RMSE and the SD of the repeated measurements would have a similar numerical value, $\mathbf{Z'} \approx \overline{Z}$ (Drummond, 1995). Both RMSE as a measure of accuracy (discussed earlier), and SD, as a measure of precision, have been long accepted in the mapping sciences for representing geographical information quality.

Like the term *quality*, the definition of accuracy is relative as it requires identifying a source of higher accuracy for comparison (e.g. larger scale; up-to-dated measurement; ground truth). Some researchers thus prefer the term 'uncertainty' to 'accuracy' (for example, Goodchild, 1995). Burrough (1986) refers *uncertainty* to the awareness of the inaccuracy of the value, as such it is difficult to measure. However Goodchild (1995) defines *uncertainty* as a *measure* of the range of values of an attribute which might result from repeated measurement; different measurements from alternative methods; or interpretation by different observers. This is similar to the definition of *precision* by Clarke and Clark (1995) discussed earlier. One can only infer from these discrepancies of definitions that uncertainty can be a cognitive property from the users' perspective, depending on the application, which may be measured in practice by referring uncertainty to *imprecision*. The available measures of uncertainty and accuracy are related to each other in a antagonistic manner. Uncertainty relates to both accuracy and precision, but accuracy and precision are independent to each other. The precision should reflect the accuracy by using the appropriate fineness of a measurement.

3.2 General issues on errors and policy in the context of the U.K. government

As Chapter two shows, the evaluation of the Safer Cities Programme involves techniques from multiple disciplines. For example, part of the evaluation consists of a large scale survey of respondents' perception. It would also be relevant to review here the errors and uncertainty involved in that survey and to examine the possible interaction between survey and GIS application. A Task Force was set up by the U.K. Government Statistical Service Committee on Methodology [GSS(M)] in December 1996 to categorise types of errors for use within the GSS, and identify high priority topics for future work. The draft definition and classification was made in consultation with Government Social Research heads of profession and Directors of Statistics in March 1997. The number and detail of responses showed a significant level of interest in non-sampling error. The report listed 54 examples of research into non sampling error carried out by all government departments in Britain. Out of which, only one report was on geographical error from Scotland (GSS, 1997). This reflects that most statisticians and social researchers are spatial blind on data quality issues. Nevertheless, it was agreed that non-sampling error work should be given a higher profile; and that priority should be given to processing error, respondent error and instrument error.

3.3 Specific issues: classification of spatial errors

The U.K. GSS(M) Task Force initially agreed to compile a bibliography to assess the extent of current research and practice in measuring non-sampling error. Later, they were reluctant to engage in bringing existing bibliographies together which were extensive. As a result, their review was relatively narrow both in range and scope. The classification proposed was noticeably based on Groves (1989) which was not necessarily a reliable source on the statistics and error issues. For instance many researchers would not accept Groves' (1989) definitions of accuracy as the inverse of RMSE and the precision as the inverse of the standard deviation (as in practice the true values and the inverse functions are usually not known).

Groves (1989) classified errors into two groups: *Error of non-observation* (which include Coverage, Sampling, and Non-response) and *Observational or measurement error* (Interviewer, Respondent, Instrument, Mode). However it did not include processing errors which is extremely important in GIS applications. Furthermore, GSS(M) classification viewed systems errors as a non-measurable entity. On the contrary, as this research will show, errors, including systems error, are quantifiable. From Groves' classification, GSS(M) Task Force added *Processing error* which includes systems, and data handling. This corresponds to the classification developed by Eurostat (GSS, 1997) which consists of four types of errors: frame; non-response; measurement and processing. In practice, many types of error interact, and it is not always possible to distinguish them separately for classification purposes. Although originating from work on social research, many categories of GSS(M) classification also apply to spatial data. Within the GIS tradition, spatial errors are classified broadly into positional error and thematic error (Chrisman, 1989). However, the grain of such broad base classification is not fine enough to understand the processes of GIS errors. The types of error have usually resulted in a complex interaction between the GIS processes and the data types of the geographical information. For instance, Cressie (1993) classified spatial data into three classes: geostatistical data; lattice data; and point pattern data. Geostatistics describe spatial processes indexed over continuous space. Lattice data describe spatial processes indexed over lattices in space. Point data describe spatial point processes. These three classes of data can be grouped into two types: vector and raster data. Vector data are represented by geometrical objects: points, lines and areas (Ehlers *et al*, 1989; Goodchild, 1989; and Goodchild *et al,* 1992). The basic geographical features of the vector data can be represented by means of line segments which form the basis of digitisation (Mark, 1989). These line features can be classified into two types (Shi, 1994). Type I lines represent specific points constructed in the real world (for example, cadastral and political boundaries). Type II lines which represent natural features such as soil and forest have no specific points in the real world. Different types of spatial data tend to associate with different groups of errors.

Although a considerable amount of the GIS literature addresses the error issues, it is still a small proportion of the GIS literature as a whole and it is arguable that all the GIS literature should address error. With a few exceptions which are described in this section, those that addressed the error issue have tended to concentrate on a particular kind of error without giving attention to the classification of the error types.

Burrough (1986) provides an overview of spatial information processing errors and groups 14 basic types of error into three main groups of factors. Group I includes the obvious errors due to factors such as age of data, incomplete areal coverage, map scale, density of observation, relevance, format, accessibility and cost. Group II errors are due to variations which include positional accuracy, accuracy of content, sources of variation in data due to data entry or output faults, observer bias, and natural variation. Group III is the processing error which includes numerical errors in the computer, faults arising through topological analyses, problems associated with map overlay, classification and generalisation problems, methodology, class interval definition, and interpolation. There are also 'unseen errors' and 'natural spatial variation' (Burrough 1986 p132). In practice, these natural sources of variation are augmented by considerable amounts of artificially generated uncertainties added to the data by the GIS process itself.

While Burrough's treatment of error issues are detailed and relatively thorough, it is not a very helpful way of classifying errors. For example the obvious sources of errors are neither obvious (to the novice GIS user) nor helpful. It does not provide a better understanding about the nature of the error. Burrough's text is mostly based on land resources assessment. Some points might not be applicable to the socio-economic context. For example, bias in soil survey and laboratory errors were unlikely to be applicable to the Safer Cities Programme Evaluation. Most of the above data errors are either trivial or not applicable to some applications. The most important types of errors are problems associated with classification, map overlay and interpolation which are the focus for this case study.

Brunsdon and Openshaw (1993) simplified the number of classes of error (14) down into 4: data capture; ageing of data; representation (generalisation and simplification processes); and GIS operations. Group I errors are due to a mix of measurement error and sampling error. Group III errors result from the effect of categorisation and aggregation. Furthermore, many GIS operators add their own errors to data that already contain errors (Group IV errors). This classification scheme is simple, but does not cover all types of errors. Moreover, the mixing between types and groups of error classes does not provide clarity to help understand the error processing.

Other classification approaches can be described in terms of methods rather than data. For example, Kennedy-Smith (1986) has classified methods of assessing accuracy in terms of internal and external testing. The internal testing, usually used in quality control, is a precision assessment based on several independent repeated measurements using the average as the estimate of the 'truth'. External testing assesses the accuracy using external sources, usually with higher standard, as the 'truth' for comparison (Hord & Brooner, 1976). However, internal testing can be considered as no more than measuring the standard deviation (SD) as discussed earlier. Most methods are external testing. Such classification is too broad to be useful. It has left the classification of data accuracy unresolved.

The earliest and most influential effort to identify and standardise different aspects of spatial data quality is the Society of Data Transfer Standards (SDTS) which, in turn, has influenced the National Committee on Digital Cartographic Data Standards (NCDCDS) in the United States. According to the revised NCDCD report, there were five components to be considered as spatial data quality indicators in the standard for digital cartographic data (NCDCDS, 1988). Since then, two additional elements (semantic and temporal) have been added by the ICA Commission, making in total seven elements of the spatial data quality (Guptill and Morrison, 1995):

1. Lineage
2. Positional accuracy
3. Attribute accuracy
4. Completeness
5. Logical consistency
6. Semantic accuracy
7. Temporal information

Guptill and Morrison (1995) have compiled papers from different authors to define and explore the issues of the above seven elements of spatial data quality. It is unfortunate that the interwoven nature of the elements has been presented as separate components in a disjointed fashion. It has been written mainly for data providers rather than from the users' perspectives. Many examples illustrated have been based on the US documentation with direct quotes in many cases (for instance from NASA). There is a danger that the standard might be inappropriate, if applied uncritically to the UK context. Nevertheless the review represents a promising step forward. It provides substantive concepts from which a framework for classification and evaluation may be derived (as exemplified by Veregin and Hargitai, 1995). It is therefore worthwhile to critically review each of the spatial data quality elements in the next section.

3.4 Review of the spatial data quality components

This section re-examines the seven spatial data quality components mentioned above. From the review, a suitable classification framework will be developed for the case study.

1. Lineage. The lineage of a data set is the documentation of its history. It describes different data processing stages from the sources, through data acquisition, compilation methods, conversions, and transformations, to the final products used in an analysis as well as the assumptions and criteria applied at each stage of its use (Clarke and Clark, 1995). Lineage is usually used by the data producers for internal records to ensure that the organisation's standards are being maintained. It provides an estimate of the errors at each stage of the 'production line'. Lineage can thus be used by the data user as a direct assessment of quality.

It can help users assess data quality using their own criteria before they initiate further processing of the data, and decide whether or not the data set is 'fit for use' (also see completeness, Brassel *et al.* 1995). As such, the lineage information should be provided by the supplier as part of the data quality report (though this is not always the case in practice). For efficiency, it is useful to use templates for lineage provision. The one suggested by Clarke and Clark (1995) does not match the list of elements described and is better re-structured as shown in Table 3.1.

Table 3. 1: A lineage framework

Contents

1	Source
1.2	Origin
1.3	Reference fields
1.4	Spatial data characteristics
1.5	Co-ordinate systems
1.6	Map projections
1.7	Corrections and calibrations
2	Pre-processing or Input
2.1	Acquisition (Data collection stage)
2.2	Compilation
2.2.1	Scientific parameter generation stage
2.2.2	Data conversion stage
2.2.2.1	equipment used
2.2.2.2	operator policy
2.2.2.3	digitisation policy
2.2.2.4	source material
2.3	Derivation (Product stage)
3	Transformation and analyses of data [Process]
3.1	Co-ordinate transformation
3.2	Interpolation

Although lineage is an important prerequisite for the users to assess data quality, it is not always provided by the data suppliers (Drummond, 1995). There are disagreements on whether it should be included as part of a data quality model or as metadata (data about data). One of the problems of using lineage for error modelling (say, for this case study) is that it does not provide an index of data quality, but an index of metadata quality (Veregin and Hargitai, 1995). Unlike metadata, lineage is documentation which as a norm (usually in a text format), would not be subjected to further processing (Clarke and Clark, 1995). However, with advanced computer technology such as object oriented programming languages (see Brunsdon, 1995; Khorev *et al*, 1996; Maguire, 1994; Tierney, 1990), there is no reason why a lineage entry facility cannot be 'programmed-in' as part of the GIS environment like meta-data using the template described earlier (e.g. IDRISI). However, this is still not useful enough as it does not provide indices. As a data user, for a one-off application of the data sets, lineage does not form part of the major concern for this case study. It is sufficient to describe the lineage of the data sets used as an example (Appendix 3).

2. *Positional accuracy.* Positional accuracy is defined as the nearness of the values describing the *position* of an entity in a real world in an appropriate co-ordinate system to its 'true' position in that system (Drummond, 1995). (Note the definition is similar to the general definition of accuracy described in Section 2.1 except that a special reference has been made to the term 'position'.) Positional accuracy is one of the concerns of this case study and will be discussed further together with other issues such as attribute accuracy and error propagation.

3. *Attribute accuracy.* Attribute accuracy can be defined similarly to positional accuracy except that the accuracy values in this case are scalars only (while positional accuracy values are both vectors and scalars). It relates to the attribute rather than the co-ordinate system. It follows that its assessment for measures on a continuous scale can also be performed using procedures similar to those used for positional accuracy (NIST, 1994, p. 22). Goodchild's (1995) discussion of attribute accuracy focuses only on well-defined features for which the processes responsible for positional uncertainty are different from those responsible for the uncertainty in attributes. In practice, the positional accuracy and attribute accuracy are not always clearly separable. There is often a link between positional and attribute accuracy such as lengths, areas and the population (within the area). Since the spatial interpolation of incompatible zones would have significant impacts upon the subsequent attribute values, attribute accuracy is one of the major concerns of this case study and will be explored further in later Chapters.

4. *Completeness.* There are a number of definitions of completeness. NCDCDS digital cartographic data standard defines completeness as an attribute describing "the relationship between the objects represented in a data set and the abstract universe of all objects" (Morrison, 1988 p 135). Brassel *et al* (1995) describe completeness as whether "the entity objects within a data set represent *all entity instances of the abstract universe*" as absent or present and to what extent. [Italic added.] Here the term *entity instance* represents the name of a real world phenomenon of a given entity type (SDTS, Barnett and Carlis, 1993); and *entity object* is the digital representation of a real world

phenomenon (SDTS definition). The *abstract universe* is specified through the data capturing rules and usually described within the meta data. The definition does not specify *which* abstract universe the entity instances belong to. The issue of completeness assessment relates to more general issues of *data quality* and *fitness for use*. Both data quality and fitness for use have been discussed earlier in Section 3.1. Completeness can be classified into two kinds: data completeness and model completeness. *Data completeness* (an error of omission and a measurable data quality component) relates to data quality, *model completeness* to fitness of use. Brassel *et al*'s model is confusing as on the one hand they equate model completeness with the fitness of use assessment while on the other they separate the fitness of use from model completeness. The fitness of use is assessed by comparing the abstract universe as specified by the data set to the abstract universe defined by the requirements of the application. However, the *final* fitness of use statement should result from the combination of both data completeness and model completeness.

Users are interested in determining fitness of use; and under what conditions and with which consequences they can use a data set for a specific application. Although data completeness is *application independent*, it generally provides only some information in support for users' decision on the fitness of use. Assessing model completeness is *application dependent* and involves assessing semantic accuracy (see Salgé, 1995). Despite the importance of model completeness, Brassel *et al.* (1995) only focus on data completeness. Data completeness can be further classified into formal completeness and entity object completeness.

Formal completeness specifies whether all information formally required is present; and to what degree the formal structure of a data set is complete. This includes: mandatory meta-information; standard data format; and correct syntax. Formal completeness should be assessed when the data set is assembled.

Entity Object (EO) Completeness specifies to what degree all entity *instances* implicitly or explicitly defined by the data description are present in the data set. The idea of the EO completeness is based on an entity-based conceptual data model (SDTS, 1992 part 2) which has been applied to the GIS context (Shepherd, 1991). An entity has instances (or *entity instances* as Brassel *et al*, 1995, called them) which are represented digitally as objects (*entity objects* or features); and *attributes* (as its properties). An entity instance represents the real world phenomenon. An entity object is a digital representation of the instances. The completeness for object and attributes means to check for the missing entries. *Attribute completeness* expresses degree of omission of information. The *global attribute completeness* specifies attributes which are missing for each object of that data set.

Theoretically, EO completeness is assessed by comparing the data to the data description of an abstract universe (reference frame). This may need expert knowledge for implicit data descriptions (interpretation). In practice, the comparison is reduced to the simple comparison between the data values and the meta-data. Attribute completeness is assessed with an EO x AT (attribute types) table by checking whether the relevant attribute types are specified in the meta data or not. If not, they are totally absent. If they are present, check for the zero values. Similarly, the completeness of entity objects (EO) and attribute types (AT) in combination can be assessed by compiling an EO x AT table with ticks and crosses for the present and missing values respectively. The selection of object entities (of the two cities) for this case study is similar to such completeness assessment (see Chapter 4).

Incomplete attributes may have an impact on attribute accuracy, positional accuracy and logical consistency. Partially complete attribute values (e.g. a missing arc) often lead to problem of logical consistency. The relationship between completeness and other data quality components is interwoven; and is best facilitated by linking all quality components in a comprehensive systems. Veregin and Hargitai (1995) attempt to develop such a system. This will be further explored later in this chapter.

5. Logical consistency. Logical consistency of a data set can be defined as the internal consistency of its structure and attributes defined by a set of logical rules. Kainz (1995) describes various ways to represent spatial data as objects and their relationships using an object oriented approach. Logical consistency tests have been developed as part of data integrity using formal logic: mathematical theories of algebra, graph, topology, and ordered sets (Corbett, 1979; Date, 1985; Meixler and Saalfeld, 1987; White, 1984). These are usually implemented as part of an automatic procedure within a commercial GIS to show spatial errors during the data capturing process (digitisation). The program should identify any topological inconsistencies such as duplicate lines; slivers; undershoots; overshoots; missing centroids; missing nodes; and pseudo nodes. The errors would then be corrected by the human operator using GIS routines such as an on-line editor. Logical consistency is therefore not a significant part of this case study.

6. Semantic accuracy. Semantic accuracy refers to the accuracy in the linguistic meaning of a description. In practice semantic accuracy is assessed by comparing the description of a data set with that of the 'selected model' (Salgé, 1995). The selected model is merely an abstraction of the real world. It represents the users' and system designer's 'perceived reality'. As such, one aspect of evaluating the models performance is its 'ability for abstraction'. The system specification forms an important element of the 'fitness for use' evaluation from both the producer's and user's point of view though each with different emphases in terms of *what* the system does and its required *performance.* The implementation of the systems specification consists of a set of conceptual schema for data specification. This can be used to set the quality parameters for assessing the systems performance. 'Truth in labelling' philosophy in specification is part of semantic accuracy. Semantic accuracy also relates to the assessment of completeness in terms of identifying the attribute accuracy. Defining and designing a GIS for accuracy assessment to warn users of the consequence of semantic error has so far been poorly developed, and is still a subject for further research (Salgé, 1995). This is partly because of the lack of incentive from the producers' point of view to incur extra cost for the users' benefits. It is also partly due to the difficulty in modelling the cognitive aspects of the users' 'perceived reality' in the specification phase. This issue has been addressed elsewhere such as cognitive systems functions specification (see Macleod and Law, 1998). Like completeness assessment, semantic accuracy forms an important integral part of the application development, but it does not fall within the scope of this case study (as here I am concerned with assessing the error processes *retrospectively*, that is, after the completion of an application).

7. *Temporal information.* The problem of handling time in GIS can be viewed as adding a temporal dimension to the general feature-based (object-oriented) spatial data models (see Egenhofer and Frank, 1992; Guptill, 1990; 1994; 1995; Guptill and Stonebraker, 1992; Worboys, 1994a). This requires maintaining and processing a transaction log to keep a database current, that is, maintaining the logical consistency of a database over time using database time stamps (Guptill, 1990; 1994; 1995). This approach is a classic transaction management used in the real-time relational databases. As such the technical aspects of its implementation are well understood. For designing a GIS application, care should be taken to consider the temporal aspects. The information need to be known in advance, that is, before the transaction time so that the database model can be designed to accommodate the values. Otherwise, the change such as inserting an attribute into the existing database can be difficult. Temporal information is part of the data set used in the Safer Cities Programme Evaluation and this case study, *rather than* part of the objectives of the research. This will be discussed where it is appropriate as part of the data description for this case study.

3.5 Evaluation of the spatial data quality components

There is still a long way to go in developing a general model of data quality for geographical data as the models proposed suffer from a number of limitations:

- There is an absence of long term empirical assessment, and the performance of such data quality standards/models is only speculative.
- Some dimensions of data quality (such as consistency and completeness) are undifferentiated over space, time and theme.
- The temporal dimension has significant implications for spatial and thematic data quality. The meaning of location in space is always bound up with location in time (Parkes and Thrift, 1980).
- Resolution is an important component of fitness for use, and yet it has not been explicitly included.
- The model defines data quality as a static attribute of a database. However, data quality characteristics can change as data are transformed in GIS.
- There is a disagreement about whether lineage should be regarded as a component of the data quality model or should it be more accurately regarded as meta-data (Clarke and Clark, 1995; Veregin and Hargitai, 1995).

In view of some of the issues discussed above, Veregin and Hargitai (1995) offer an alternative view of data quality components and attempt to establish a general model of data quality. The seven components have been reduced into three by:

- taking the temporal component out and treating it as a separate dimension
- grouping the three components of accuracy into one dimension: accuracy
- not regarding lineage as a data quality component but a metadata set which act as references.

An extra component - resolution - is added to the components of data quality (Veregin and Hargitai, 1995). The four components are placed against three dimensions: space, time and theme described as probability distribution (from the general model of data quality, Giordano *et al*, 1994). This creates two dimensional arrays of an evaluation matrix. However accuracy is more appropriately conceived as a three-dimensional array, if one considers not only the geographical data dimension (space, time, theme); but also the reference source (such as the lineage) used for the assessment (Figure 3.1).

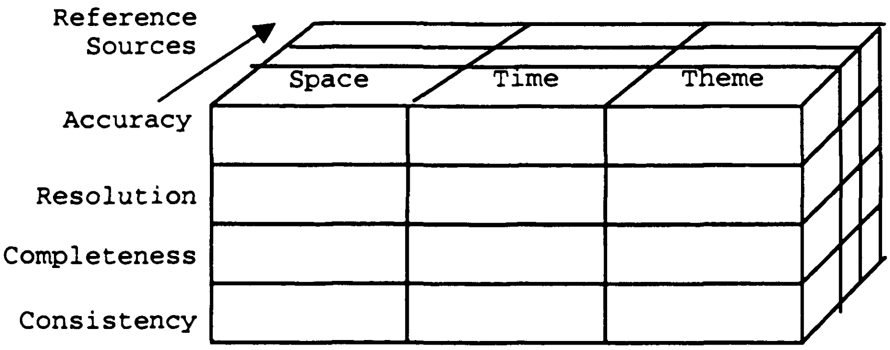


Figure 3. 1: An evaluation matrix, modified from Veregin and Hargitai (1995)

There is a trade off between a comprehensive classification with detailed definitions which might appear too complicated and a more general classification which might not cover all possible sources of errors. The classification should be both comprehensive and flexible enough to be readily applied to other contexts as well as this case study. The classification is a tool to be used for establishing a further framework for error measurement and management (see later Chapters). As this case study concerns the group of errors that can occur in GIS processing, the classification should reflect both the nature of the data types and the processes of the GIS operation. With this framework in mind, the error groups can be classified by first considering the GIS processes, and second, combining the processes with the different natures of error according to the types of data used.

Out of the seven data quality elements, four are directly relevant to the accuracy issue: positional, thematic logical and semantic accuracy. Logical inconsistency may be due to positional and thematic error (Shi, 1994). For example the location of a house at the edge of the city boundary may be wrongly excluded in the digitisation process or misclassified in remote sensing images. Thus the basic error groups can be classified into two: positional and thematic errors according to their data types (This is consistent with Chrisman, 1989). Since the temporal dimension has been described as part of the data sets, the time dimension can be taken out and replaced by the GIS processes which are the concern of this research. By taking positional accuracy as the spatial dimension, thematic (attribute) accuracy as theme, the above error matrix can further be simplified as follows (Figure 3.2):

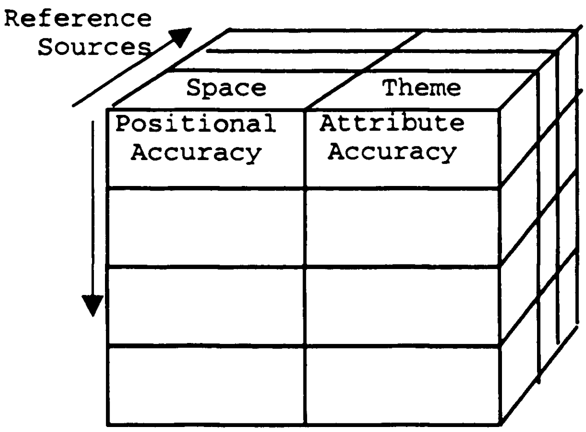


Figure 3. 2: Model of accuracy

The next section examines the GIS processes to further evaluate the model of accuracy within the software engineering context.

3.6 GIS processes and error classification

Figure 3.3 shows the GIS processes which consist of the following logical groups based on the Software Engineering paradigm: input-process-output (Law and Ekblom, 1996). A similar software engineering frame-work applied to GIS error modelling has also been proposed by Shi (1994).

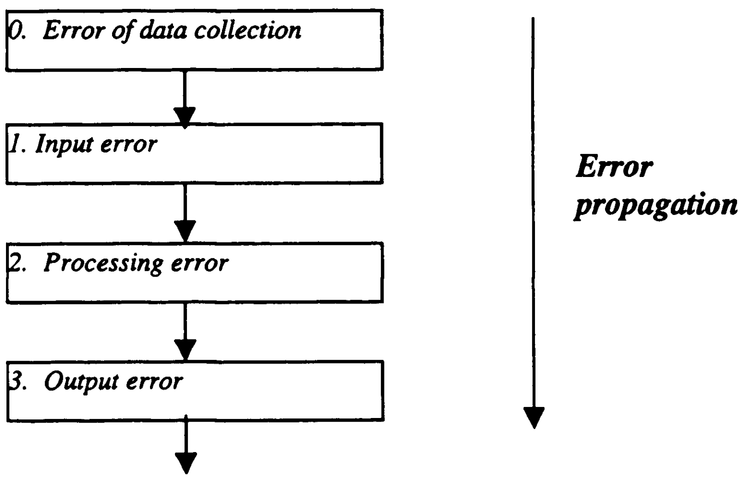


Figure 3. 3: Classification of GIS processing errors

However the above process model was based on the waterfall model derived for systems that could be physically engineered and one would argue that it is no longer adequate (Macleod and Law, 1998). It glosses over the recursive nature of the advanced computer applications where the output from one application very often becomes an input of another application. Further analyses may be carried out within the same application using different software / hardware systems. For example in the case of the Safer Cities Programme Evaluation, the output of the GIS became the input of the multi-level modelling using ML3 software. It follows that the input and output errors should be more appropriately described as transfer errors, and the process of error propagation iterates along different applications or systems as shown in Figure 3.4. The receivers of the output may either be another computer system or human operator. The transfer and process errors include human perceptual and interpretation / cognitive errors respectively. The model is more robust and consistent with the user model of GIS and computer systems (Medyckyj-Scott and Hearnshaw, 1993).

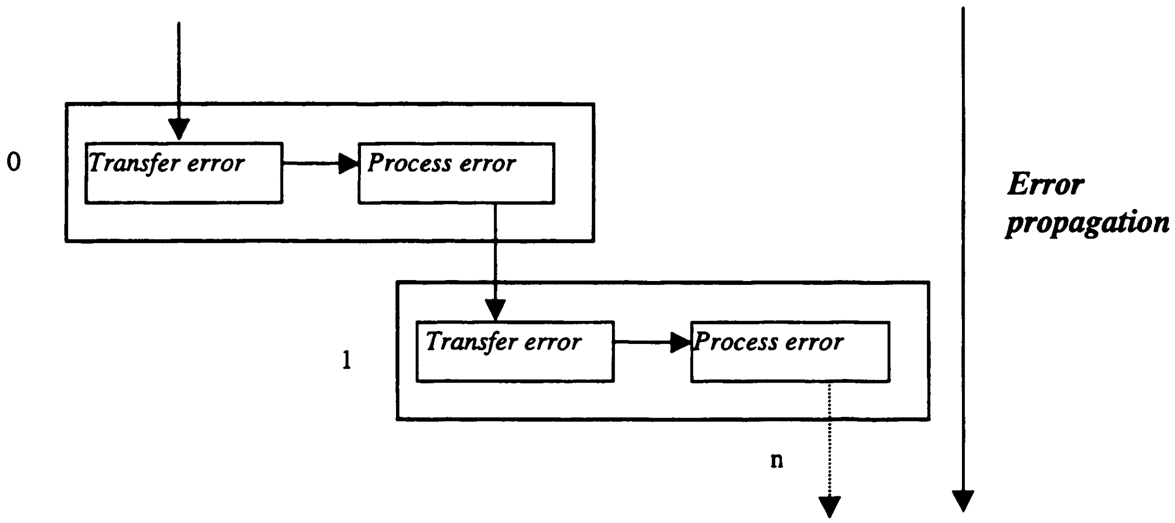


Figure 3. 4: Transfer / Process model

Starting with the initial transfer error (Step 0), the above process can recur indefinitely, say, 1 to n cycles, depending on the application and use of the data sets. This case study is primarily concerned with the GIS process error and the secondary impact upon its further analyses within the application of the Safer Cities Programme Evaluation.

3.7 Proposed classification

Different types of spatial data would produce different kinds of error at each stage of GIS processing. To combine the GIS processes described in the previous section with the nature of the data types discussed in Section 3.2, a complete error classification can be summarised by the following table:

Table 3. 2 Classification of GIS errors

Propagation / Data Type	1. Transfer	2. Process
A. Positional	1A	2A
B. Attribute	1B	2B

As shown in the above table, the transfer and process stages consist of both positional and attribute errors (Group A and Group B errors) as the results of the interaction in GIS processing and data analyses. The coding simply refers to the combinations in the table. For instance, the interpolation of the choropleth maps (Beats and EDs) would introduce both Group 2A and Group 2B errors in addition to Group 1A and Group 1B errors.

3.8 Chapter Summary

This chapter has examined the general sources of error and uncertainty in geographic information by reviewing the previous effort in defining quality for digital data sets. Taking both the users' and developers' perspectives, the review began with a description of the general *Quality* issue and its relationship to *Fitness of use*, and focused specifically on the *data quality* issue. Some key terms such as *accuracy; precision;* and *uncertainty* have been defined. The error indicators including root mean square error (RMSE) and standard deviation (SD) have been formulated. General issues on errors and policy have been discussed first with in the context of the U.K. government; and second within the broader framework of spatial data quality classification from the literature review. The former includes *Error of non-observation; Observational or measurement error;* and *Processing error.* Processing error has been regarded as most important for the investigation of GIS error as it includes systems, and data handling. Seven spatial data quality components have been critically evaluated for the purpose of the scoping study for this research. These include: *lineage; positional accuracy; attribute accuracy; completeness; logical consistency; semantic accuracy; temporal information.* As the result of the review, a general classification system for errors has been formulated which places emphases on *positional* and *thematic* errors and their relationship to different stages of GIS processes: *transfer* and *process* errors. This will provide a framework for this case study. The next chapter will apply this framework to the specific context of the Safer Cities Programme Evaluation.

Chapter Four
Research aim, objectives and scope

This chapter describes the aim, objectives and scope of my research by mapping the general frame-work developed in the previous Chapter upon the specific context of the case study (described in Chapter Two). First a number of research questions were raised in Section 4.1. On the basis of the theoretical importance and relevance to the Safer Cities Programme Evaluation, the aim and objectives of the research are formulated in Section 4.2. Possible scopes of the study were framed by re-examining the steps of the data processing used in the Evaluation. Finally, the process for selecting two cities as a case study for further research is described in Section 4.4.

4.1. Research questions

As shown in Chapters Two and Three have shown, spatial errors are inherent in most GIS applications especially for the complex spatial data processing such as the Safer Cities Programme Evaluation. This raises a number of research questions:

1. Do the errors involved in GIS processing affect the *results* of the Safer Cities Programme evaluation?
2. If so, by how much? In other words, can these errors be quantified?
3. Do the errors involved affect the *conclusion* of the evaluation?
4. How good is the evaluation?

Question three is actually different from questions one and two, though they appear to be similar (Law, 1998). The amount of error may affect the results of the evaluation *quantitatively* but it may not affect its conclusion (that is, the *quality* of the Safer Cities Programme). In other words, *does error matter* (in this application)?

To answer questions two and three, one needs to examine the margins of error of the quantitative estimates of the impact of the Safer Cities Programme. Statistically, question four asks to what extent has the Evaluation managed to minimise both Type I and II errors (failure to accept and reject the null hypothesis respectively i.e. wrongly to conclude that there is a Safer Cities Programme effect or not).

The answers to the above questions also depend, among other things, on how good the data quality is, and how well-founded are the geographical assumptions behind the collection of different data sets and their manipulation (for example, the definition of neighbourhood, assumption of no migrants, and the accuracy of beats and 1991 EDs interpolation). Experience of working on the evaluation over the course of the evaluation programme has revealed many ways in which the collection and manipulation of the data are problematic (for example, collection of the crime statistics, and the fact that police beats change over time). As a result, a number of quite severe trade-offs have had to be confronted (for example, greater continuity of measures over time, or better spatial resolution?). Thus an assessment of the nature and magnitude of accuracy in the evaluation, and of ways to present this uncertainty to end-users (i.e. researchers, administrators, and policy makers), is extremely important not only in retrospect but also to guide the design and conduct of future evaluations. From another perspective - that of GIS as a discipline (if it were)- the measurement accuracy in the data manipulated by GIS is also of fundamental importance (Goodchild and Gopal, 1989).

4.2. Aim and objectives

Judging from the theoretical importance of the spatial accuracy issues and their practical implications upon policy decision making as exemplified by the Safer Cities Programme Evaluation, the primary aim of this research is to *explore the data accuracy issues in GIS processing*.

The objectives of this case study are to:

1. assess the impact of the spatial uncertainty upon the Safer Cities Programme Evaluation.
2. provide a comprehensive quantitative account for the possible errors in the GIS operation for the evaluation.
3. determine whether the errors were significant or not.

In achieving the above aim and objectives, it is necessary to explore the validity of the geographical assumptions made; and to simulate the process of analysis through the error propagation so that more precise error statements can be made about the effect of the Safer Cities Programme. The research will have important implications both to the conclusion of the evaluation as knowledge for policy making within the specific context of the Safer Cities Programme; and for the use of GIS in general.

4.3. Possible Scope of spatial accuracy within the Evaluation of the Safer Cities Programme

As described in Chapter two, the Safer Cities Programme Evaluation involved internal and external comparisons of the recorded crime statistics and survey data sets collected from a number of cities. So the possible scope of the study should be based on the processes of these different data transformations.

Evaluation using recorded crime statistics

The processing of the crime data can be summarised as involving the following steps:

1. Collect monthly crime statistics: number of incidences for each crime type per beat between 1987-1992.
2. Transform them into attribute entities and store them in the database (INFO format).

3. Digitise the police beat boundary maps to generate spatial object entities for beats (named 'BEATGS' in ARC).
4. Join the spatial object entities with attribute entities using beat-ID as an identifier (to form a new BEATGS).
5. Input census data sets from OPCS.
6. Overlay the boundaries of the enumeration districts (ED) with beats to form a new spatial object entities ('EDBEATGS') using the ARC/INFO overlay method.
7. Process the data sets according to the scoping principle (discussed in Chapter two).
8. Compute the action intensity scores based on the scoring principle (2.1).
9. Output the action scores for multi-level modelling.

The above is a classic GIS process and thus could be an ideal context in which to study and explore the spatial accuracy issues. It is also most likely to have an impact, if any, upon the outcome of the Safer Cities Programme Evaluation.

Evaluation using the survey

The steps taken to process the survey data can be briefly described as follows:

1. Conduct the survey before the implementation of the Safer Cities Programme (the so-called BEFORE SURVEY, i.e. before the end of 1989 and hence sampling had to be based on 1981 census).
2. Transform data collected in Step 1 into attribute entities and store them in the database (INFO format).
3. Repeat Step 1 and 2 after the completion of the Safer Cities Programme (AFTER SURVEY, i.e. in 1992 and hence sampling was based on 1991 census).
4. Transform the 1981 ED base BEFORE SURVEY data into 1991 ED base so that both BEFORE and AFTER SURVEY data-sets can be linked with the 1991 ED as a common spatial identifier. (This step was carried out by NWNNL.)
5. Process the data sets according to the scoping principle as before

6. Compute the action intensity scores based on the scoring principle (2.1).
7. Output the action scores for multi-level modelling.

(See Ekblom *et al*, 1996a for details.)

The spatial operations in survey data processing have been far less than those involved with the crime statistics. Step 4 is the only spatial transformation involving spatial uncertainty. Furthermore since there was obviously no action before the initiation of the Safer Cities Programme, the action scores in a BEFORE SURVEY would be zero, and so any spatial errors would have no effect upon the survey action scores. The only effect of the spatial uncertainty upon the survey analysis would be the sampling as the BEFORE and AFTER SURVEY respondents were sampled on the bases of two different geographical units (1981 and 1991 ED). However, survey sampling in social sciences would be beyond the scope of the spatial accuracy in GIS. It was therefore decided to exclude the survey analysis from this case study. Furthermore, since the external comparison was analysed at city level, spatial data were not collected for the external comparison cities. External comparison was therefore excluded from this case study since there is no GIS operation required.

To model the spatial error processing would probably involve first identifying the steps contributing to the accuracy of the spatial object in the GIS; then determining the accuracy of error introduced by each step; and thereafter determining the cumulative effect of these errors. Ideally, it would have been desirable to select all sixteen cities for error modelling. However, considering the cost of the time and effort required, it would not be feasible to do so in practice. While a single case analysis might be most economical, there was a danger that the selected city might have some idiosyncratic properties such that the results might not be generalised. Two cities were therefore chosen. To sum up, <u>this research focuses on the internal comparison of the recorded crime statistics for two Safer Cities as a case study for in depth analyses</u>.

4.4. Selection of two cities for the case study

The selection criteria were based on the data availability and their representativeness of the cities involved in the evaluation. Although the data were eventually collected for all the 14 cities evaluated, it was important that all the data sets were available at the start for the cities selected for this case study. For crime statistics analyses, six years of police recorded crime data were collected (1987-1992). Table 4.1 shows the status of the crime data collection for all the 16 cities and boroughs (Safer Cities) at the beginning of this research. Y indicates a complete data set, and N incomplete. Just on this basis, Birmingham, Islington, Lewisham, Sunderland, Tower Hamlets, Wandsworth, and Wolverhampton would have failed the selection criteria. Birmingham was also considered to be too large to be a typical city for this case study.

Table 4. 1: The status of crime data available for the 16 cities and borough in 1993

Cites	SuperID	ChangeID	1987	1988	1989	1990	1991	1992	CNTRE
Birmingham	16	Y	N	N	N	Y	Y	Y	Y
Bradford	15	Y	Y	Y	Y	Y	Y	Y	Y
Bristol	**25**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**
Coventry	**N**	**N**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**
Hartlepool	N	N	Y	Y	Y	Y	Y	Y	Y
Hull	N	N	Y	Y	Y	Y	Y	Y	Y
Islington	N	N							NA
Lewisham		Y							
Nottingham	5	Y	Y	Y	Y	Y	Y	Y	Y
Rochdale	6	Y	Y	Y	Y	Y	Y	Y	Y
Salford	15	Y	Y	Y	Y	Y	Y	Y	Y*NB
Sunderland	5	Y							Y
Tower Hamlets	N	N							N
Wandsworth	N	N							NA
Wirral	1		Y	Y	Y	Y	Y	Y	Y
Wolverhampton	3	Y	?	Y	Y	Y	Y	?	Y

SuperID - Superbeat; ChangeID - beat change, CNTRE - city centre
Y-complete; N-incomplete; NA - not applicable; NB - see text.

Other issues included the existence of the city centre, and the problems of the beat change. Some areas such as the boroughs did not have city centres. Salford appeared to have more than one city centre. Police beat boundaries changed over time. This problem was resolved by looking back through past maps to uncover 'beat pedigrees', and identify '*superbeats*' - groups of adjacent beats whose common outer boundary remained about the same despite changes within. With this, we arrived at a 'standard beat map' for each city which covered the whole time period. In this way, a full geographical coverage was achieved in 13 of the Safer Cities for which crime data was available, and partial coverage in Birmingham. It was decided that one city selected should have super-beats, while the other one should not.

Based on the representativeness and data availability at the start of the study from the examination of Table 4.1 above, Bristol and Coventry were selected for further investigation on spatial accuracy within the GIS processing. The geographical characteristics of the two cities are shown in Table 4.2.

Table 4. 2: Summary areal statistics of the two Safer Cities of England used in the case study

Cities Km^2	Min. area of EDS	Min. area of beats	Max. area of EDS	Max. area of beats	Mean area of EDS	Mean area of beats
Bristol	.07	0.1	10	9	0.1	1.7
Coventry	.08	0.4	6	11	0.16	2

Cities	No. of EDS	No. of beats	Total. Area Km^2	mean (ED) population	Total population
Bristol	824	58	110	451.91	371020
Coventry	601	46	96	487.76	293141

4.5. Chapter Summary

This chapter has described the aim, objectives and scope of my research. The aim is to explore the data accuracy issues in GIS processing. The objectives are to assess the impact of the spatial uncertainty upon the Safer Cities Programme Evaluation, to provide a comprehensive quantitative account for the possible errors in the GIS operation for the evaluation and thereby determine whether the errors were significant or not. On the basis of the theoretical importance and relevance to the Safer Cities Programme Evaluation, this research focuses on the internal comparison of the recorded crime statistics only by selecting two Safer Cities as a case study. Bristol and Coventry were selected for in depth analyses based on their initial data availability and their representativeness of the cities.

Chapter Five
Methodology

The aim of this chapter is to derive a procedure to model the spatio-thematic errors in the GIS processing. This is achieved by first examining the existing methods from the literature review, identifying the appropriate methodology and then applying it, with modification if necessary, to the spatial uncertainty problems in the Safer Cities Programme Evaluation. Initially, it was intended to structure the literature review around the classification framework developed in Chapter three. However, most methods reviewed in the literature tend to be piece-meal approaches and focus on the particular types of errors introduced by various data transformations.

In general, there is no accepted paradigm for modelling error propagation that explicitly recognises the interdependence between spatial and thematic errors and formal methods of error propagation in a large complex system. Simply, those error models proposed have often been either too general or too simple to be applicable to our context. As Lanter and Veregin (1992) have pointed out, the utility of different dimensions of error is a function of context defined by the requirements of the uses and the classes of geographical data under consideration. However the function is usually non-computable. More basic research on the mechanisms of error propagation is needed. The resultant error model that is useful in practice is context dependent, and is usually tailor-made, and labour-intensive.

Following the above discussion, it is clear that the literature review needs to be organised within a more general and robust approach that can readily be adopted to the context of the Safer Cities Programme Evaluation. Veregin (1989) proposed 5 steps for modelling error in GIS operations and arranged them hierarchically with reference to Maslow's (1954) model of human needs. There is no apparent psychological relationship between Veregin's error modelling and Maslow's hierarchical concept, though it seems sensible to manage error in terms of the following logical sequence. These are: error identification, error detection and measurement, error propagation modelling, and strategies for error reduction. However no appropriate methodology has been suggested for each step as it is likely application specific. Later, Lanter and Veregin (1992) suggested a layer-based error propagation paradigm, within which, geographical features are organised according to a spatial, thematic or temporal scheme (also see Chrisman and Niemann, 1985; Kjerne and Dueker, 1986; Aronson, 1987; Bracken and Webster, 1989). A fundamental characteristic of the framework presented by Lanter and Veregin (1992) is that different error propagation functions can be employed for any particular combination of GIS function and error index. Each error propagation function is determined by the specific error index to be propagated, the GIS transformation function to be employed, and a set of assumptions about the nature of errors in spatial data and their propagation mechanisms. The propagated indices can be evaluated in terms of their utility for judging the quality of GIS derived data products and their appropriateness in decision-making contexts.

The model seems promising to provide a general framework for this research and thus form the structure of this chapter. The proposed model involves the following steps:

1. Identify *error indices* (Section 5.1).
2. Develop *error propagation functions* (Section 5.2).
3. Test the utility of Step 2 by assessing spatial, thematic, and/or temporal accuracy according to the derived geographical data (Section 5.3).

These steps are further expanded in terms of applying them to the context of the Safer Cities Programme Evaluation in each section, drawing upon the classification framework (from Chapter 3) and further literature review of other specific methods where necessary.

5.1 Identifying error Indices

An error index is a unit measurement of accuracy (Lanter and Veregin, 1992). The choice of error indices depends on the types of data, the nature of errors (as indicated in the classification framework in Chapter 3) and the error modelling methods employed. In error modelling of the Safer Cities Programme Evaluation, we need to consider the errors introduced by transferring data (from one system to another as discussed in Chapter 3), and the uncertainty involved when combining different types of data sets as part of the process modelling. The input error of cartographic data is primarily positional though it may involve thematic error as a result of subsequent processing, unlike the uncertainty of classified remote sensing data which is mostly thematic in nature.

The first step of the transfer of the spatial data into a GIS is from a source. In this case study, it is assumed that all input to GIS was carefully checked so that there were no gross error blunders in the input data. However it is unrealistic to assume the input error due to measurement as entirely random in nature. This kind of random input error has been well investigated in surveying science (Mikhail and Ackerman, 1976). The positional error will be examined first.

5.1.1 Group 1A: positional-transfer error

Vector data sets are usually captured by digitising paper maps manually, so are the beat maps for the Safer Cities Programme Evaluation. Errors in digitising are dependent on the mode of operation (stream vs point mode), operators and equipment used. Positional error is inherent in the resolution of the system which indicates the smallest measurement possible by that system. The spatial precision of digitising varies as a result of both the instruments (in this case, a digitizer) with its inherent resolution, and the operators with various pointing skills.

Results of digitising trials have been reported by Maffini *et al* (1989); Boslstad *et al* (1990) and Walsby (1995). While the study by Maffini *et al* (1989) is very limited using only one subject in the experimental trials. Walsby's study is an improvement as she used 12 subjects (9 professional cartographers and 3 CAD experienced users using point mode digitising). The study shows that there is a wide variation in line-following accuracy, techniques and operators' skill. The digitising error is also related to the right handiness of the operators. The digitising task has a characteristic pointing precision. A 'typical' digitizer resolution is 0.02 mm. Operators have an average precision of about 0.1 mm at map scale during manual digitising (Goel, 1992); and about 0.05 mm at photo-scale during stereo-plotting (American Society of Photogrammetry, 1980, p 610-611). This is comparable to Bolstad *et al*'s (1990) finding: mean deviation 0.054 mm and maximum 0.261 mm (also see Maffini *et al*, 1989; Keefer *et al*, 1991).

A co-ordinate Standard Deviation (SD) can thus be estimated to be 0.1 mm at map scale for the digitising step for data derived from a tablet digitizer of stated resolution 0.02 mm when used by a typical operator. However the estimation of SD does not take into account the quality of the information on the map which is affected by the earlier steps of the data collection process. The SD estimate would also change as the data are further processed in the subsequent stages. It does not allow for dynamic modelling in the error. Furthermore, a fixed SD ± 0.12 mm would be unrealistically small to account for the spatial error occurring in the Safer Cities Programme Evaluation, and the unit of measurement (in mm) is not used in the evaluation process.

Methods for more dynamic modelling of positional transfer errors are basically deductive approaches as they tend to focus on individual features. Areal objects are decomposed to three basic elements:

- points,
- lines, and
- areas.

Point features

Point features are the simplest spatial objects. The models of handling the positional uncertainty of points have been most well developed in particular in surveying science. Goodchild (1991) describes a circular normal model of error in point locations which measures the likelihood of the true location of a point by referring to the height of the surface of the bell over that point. Shi (1994) defines a number of error indicators based on the normal distribution model for both circular and linear features. Blakemore (1984) describes a point at the boundary as: definitely in, definitely out, possibly in, possibly out, or ambiguous. The usefulness of such descriptions is not very clear as one still does not know by how much a point is possibly in or out.

There are several problems with point models:

- Variance-covariance matrix is not usually developed for points.
- More measurements of error quality for each point are still required.
- In practice, especially in the context of crime prevention such as the Safer Cities Programme, data sets are not always collected in points.

Line segments

A line segment is a line defined by two end points. The distributions of line segments have been simulated by numerous researchers (Dutton, 1992; Zhang and Tulip, 1990; Caspary and Scheuring, 1992). Models include:

- the epsilon band model, and
- the error band model.

The epsilon band model

The epsilon band is defined as a buffer of constant distance from either side of the line and from its two end points. Chrisman (1982) describes it as the area occupied by rolling a ball along the line (Figure 5.1). The concept of an epsilon band enveloping a line can be traced to Perkal (Perkal, 1956, 1966) and further investigated by Chrisman (1982) and Blakemore (1984).

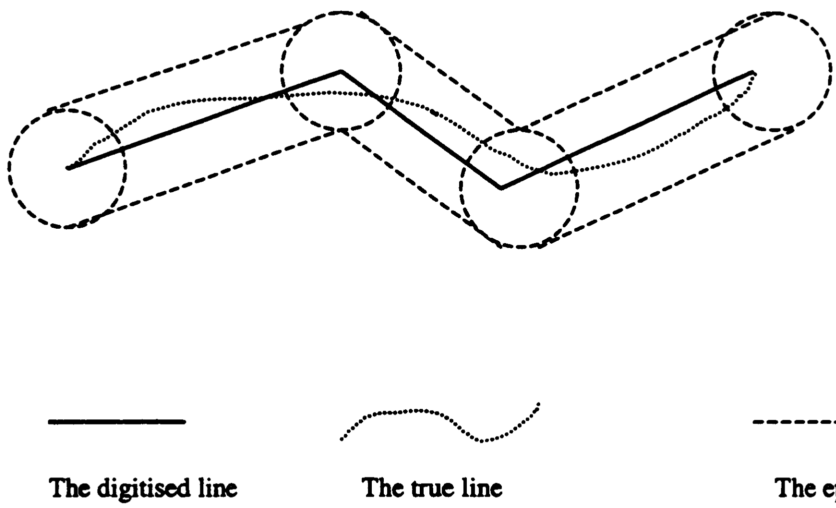


Figure 5. 1: An example of the epsilon Band

The epsilon band is based on cartographic generalisation and probabilities of ground location within and near the defined region (Honeycutt, 1986). As the epsilon band model assumes that the error effects of the digitised lines are random in nature and are independent from the GIS processes. Monte Carlo simulation may be used to perturb the true line to obtain the observed line through the area of the epsilon band. The epsilon band model is not a useful concept for describing errors in the complex GIS processing with a large amount of uncertain data. It has the following set of problematic assumptions:

- The errors are independent from the GIS processes.
- The true line lies within the band.
- Probabilistically, the band is comparable to a standard deviation from the true line if the band width is based on a SD measure but it does not have to be.
- The epsilons on both sides of the true line are equal distance, and thus form a uniform rectangular shape.

A large scale GIS application such as the evaluation of the Safer Cities Programme, would have occasions that the true lines may lie entirely outside the epsilon band. Even if the true line lies within the epsilon band, it is difficult to decide where the upper and lower limits of the band should be. For example, the digitised line may produce a bi-modal distribution around the true line (Honeycutt, 1986). In other words, there should be no upper or lower limit imposed upon the band. However, if there is no upper or lower limit, constructing an epsilon band becomes impossible and meaningless. Even if one can arbitrarily impose the limits of the band boundaries, the model offers no distribution of errors within the epsilon band.

The error band model

The error band model has been offered as an alternative to the epsilon band by Zhang and Tulip (1990), Dutton (1992), Caspary and Scheuring (1992), and Shi (1994). The error band is derived from the random error of a point on a line segment (Caspary and Scheuring, 1992). According to this model, the error at the two end points of a line segment would be largest, and the error at the midpoint smallest. However this may depend on the mode of operation. The error band still has the same problems of the epsilon band described earlier. Furthermore, its assumptions - that co-ordinate errors of end points are independent and have the same variance and covariance - are also erroneous. For instance, the point errors are not independent for lines which are digitised in stream mode. Even for those lines digitised in point mode, there may still consist of large systematic errors. For example, a mis-registration of the digitised points on a map would result in a uniform shift in the location for each point. The relation between the true line and the digitised line should not be modelled as a series of independent position errors at points (Goodchild, 1991). The combination of both systematic and random errors should also be modelled.

Methods for area Objects

The epsilon model can be applied to the 'point-in-polygon' problem by describing the uncertainty of an area object enclosed by the polygon (Blakemore, 1984). However, the model is limited to describing five discrete levels: definitely in, definitely out, possibly in, possibly out, and ambiguous. These have no quantitative indicators. The problem of providing quantitative measures between the 'possibly in/out' region still remains. This can be solved by combining the vector polygon with grid cells (Blakemore, 1984). Nevertheless the model is still limited to describing discrete levels of uncertainty rather than the continuous changes in the probability. We (the users) are more interested in the quantitative statement about the resultant errors, and how it might affect the conclusion of our analyses than in further dividing errors into different finer levels such as 'attributes of points, lines, areas, or cells'.

5.1.2 Group 1B: attribute-transfer error handling methods

Attribute values are attached to points, lines or areas in a spatial database. There are two types of attribute (also known as thematic) data: categorical and continuous. Since the data sets used in the Evaluation of the Safer Cities Programme are categorical, only those methods that handle categorical data are examined.

Categorical data are descriptors of geographical entities (Stutheit, 1990; Rodcay, 1990). The assessment of the uncertainty of categorical data is the domain of classification uncertainty assessment and has been well covered by the remote sensing literature. Although the Safer Cities Programme Evaluation did not make use of the satellite images, the classified images may be regarded as a better standard and used to compare with the GIS processed data sets in the Evaluation. This is to be explored further at a later stage in this study.

At the transfer stage in remote sensing (say, from satellite), one of major sources of attribute error is image classification. There are various sources of classification error:

- uncertainty in the definition of classes and identification;
- measurement error;
- decision making (that a pixel o belongs to a class A: $o \in A$) in interpretation or automatic classification; and
- further data manipulation.

The confidence of the decision ($o \in A$) varies within the area of a polygon with the highest at the centre of a polygon and lowest at the edge with a transition zone in between (Burrough and Heuvelink, 1992). Broadly speaking, the accuracy of the image classification can be assessed by experimental (sampling-based test) or probabilistic (Bayes' theorem) approaches.

Sample-based test method

An experimental approach known as a sampling-based test method assess the classified satellite image by selecting a sample of locations, and then comparing the class assigned to each location to some source of higher accuracy, usually ground truth obtained by direct observation in the field (Congalton and Mead, 1983; Congalton *et al.*, 1983; Story and Congalton, 1986; Rosenfield and Fitzpatrick-Lins, 1986; Congalton, 1991). The results are then expressed in a tabulated form as shown in Table 5.1 known as *error or confusion matrix.*

Table 5. 1: framework for an error matrix

	A B C .... Classified features	A/Total ($C_{A.}$)
A	C_{AA}	Consumer's accuracy
B		(C_{AA} / $C_{A.}$)
C		
:		
Ground truth		
Total ($C_{.A}$)	Producer's accuracy (C_{AA} / $C_{.A}$)	

The diagonal elements of the matrix indicate correct classification. Users can compute their own quality parameters for each class required, such as the percentage correctly measured per class (Chrisman, 1986, Rosenfield, 1986; Greenland *et al.*, 1985; Rosenfield and Fitxpatrick-Lins, 1986) or by means of the same measure but with confidence level (Aronoff, 1985; Hord and Brooner, 1976). A common accuracy standard is 85% (Goodchild, 1995).

However, to express the index of accuracy as the total percentage of correct classification is misleading since a certain number of correct classifications would occur by chance. A better index is the Kappa statistics (*K*) which accounts for correct classification that occurs by chance alone (with value ranging between 1 and 0). Given i row, j column, C_{ij} entries,

$$K = {\sum_{i=1}^{n} C_{ii} - \sum_{i=1}^{n} \frac{C_{i.}C_{.i}}{C_{ij}}} \bigg/ {C_{ij} - \sum_{i=1}^{n} \frac{C_{i.}C_{.i}}{C_{ij}}} \qquad (5.1)$$

The Kappa coefficient can be used as an error index and applied to the entire data sets (Chrisman, 1984; Rosenfield *et al.*, 1986). Alternatively, referring to Table 5.1, within each class the users and producers' accuracy can be used (e.g. Ginevan, 1979; Aronoff, 1982; Story and Congalton, 1986; Hudson and Ramm, 1987).

Producer's accuracy (*PA*) is defined as the probability of a feature (A) being correctly classified.

$$PA = \frac{C_{AA}}{C_{.A}} \tag{5. 2}$$

Consumer's accuracy (*CA*) is defined as the probability of the *classified* feature (A) correctly classified.

$$CA = \frac{C_{AA}}{C_{A.}} \tag{5. 3}$$

One of the criticisms of using the error matrix as an error indicator (say, 0.85 Kappa coefficient or 85% confidence that $o \in A$) is that it does not provide spatial distribution of uncertainty. (See Aronoff, 1989; for more technical details on remote sensing, see Colwell, 1983; and Chrisman, 1984; Rosenfield *et al.*, 1986; Campbell, 1987; and Ehlers *et al*, 1989, for issues on the raster / vector integration).

Bayes' theorem

Bayes' theorem (Rao, 1965) is based on the classic probability theory. Given sets: E, A and B (Figure 5.2):

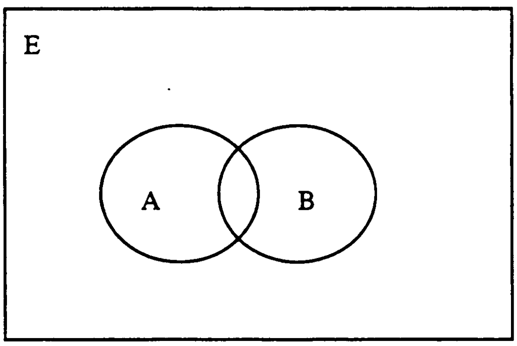


Figure 5. 2: Venn diagram of sets: E, A and B

A measure (there may be many) M on E, is a function, valued on the real numbers or integers, of the subsets of E, such that

if $A, B \subset E$
then $M(A) \geq 0$
$M(\{\}) = 0$

$$M(A \cup B) = M(A) + M(B) - M(A \cap B) \tag{5. 4}$$

The ***probability***, using measure M, ***of A***, given B can be defined as:

$P_M(A|B) = M(A \cap B) / M(B)$

Thus $P_M(A|B)$ depends on:

- M - the measure, there may be more than one
- B - the "conditioning" set
- A - the set whose probability we are calculating.
- and, indirectly, E as $A, B \subset E$ while M is only defined on E.

As a shorthand, we write $P_M(A|E)$ as $P_M(A)$ or P(A). This represents *the* probability of A = P(A). So E, M must be given by the context. It is usual to call $P_M(A|B) = P(A \cap B)/P(B)$, the '*conditional*' probability of A given B, usually written P(A|B). According to the Bayes' theorem:

$P(A|B) = P(A \cap B) / P(B)$ and $P(B|A) = P(A \cap B) / P(A)$

So the '*conditional*' probability of B given A is:

$$P(B|A)P(A) = P(A|B)\,P(B)$$

$$\Rightarrow \quad P(B|A) = P(A|B)\,P(B) / P(A) \tag{5. 5}$$

Bayes' theorem has been used to handle spatial uncertainty (Shi, 1994); and applied in crime pattern analyses (Brunsdon, 1989). In satellite image classification, Hi may be the hypothesis that a given pixel belongs to class h. X is the vector of the density value of the pixel in spectral space. The uncertainty is reflected by the conditional probability $P(H_h|X)$, which indicates the degree to which the class to be correctly assigned. The conditional probability $P(H_h|X)$ is determined by Bayes' theorem:

$$P(H_h|X) = P(X|H_h)\,P(H_h) / P(X) \tag{5. 6}$$

Bayes' model assumes that a *priori* probability $\{P(H_h)\}$ can be assigned correctly, which is often not true in practice. While the theory can accommodate quite complex joint probability, a very large number of probability measures would be required. Furthermore, the model provides neither measure of the quality of the probabilities, nor mechanism for weighting the assigned probabilities as a function of their reliability (Shi, 1994). Given a series of Hypotheses H_i and some evidence E, we can generalise the probability of the hypotheses:

$$P(H_i|E) \propto P(E|H_i)P(H_i)$$

Experiments can be carried out to determine E. If we have the frequency for which H will be true in such experiments, P(H|E) can be calculated. However hypothesis testing using Bayes' theorem is only appropriate if there is a theory to support the hypothesis (Figure 5.3 Grove, 1998).

Theory

$$P(H|E) = P(E|H)\, P(H) / P(E)$$

What is P(H)? Constant

Proportion of possible worlds for which H is true = M(H) / M(E)

Figure 5. 3: Relationship between theory, hypothesis and evidence in Bayes' theorem based on Grove (1998)

Furthermore, very often P(H) cannot be computed. If there is no chance set up then P(H|E) is not defined, we would have to make use of P(E|H) to infer. The *likelihood* of Hypothesis H given the evidence E can be defined as:

$$L(H|E) = P(E|H) \tag{5.7}$$

The likelihood of a hypothesis is *not* a probability. Note the difference:

$P(E|H)$
H fixed
E variable, E_1, E_2,....
and $\sum_i P(E_i \mid H) = 1$

$L(E|H)$
E fixed, H variable, H_1, H_2,....
$\sum_i L(H_i \mid E) = \sum_i P(E \mid H_i) = \text{anything}$

Given $P(H|E)$ does not exist but $P(E|H)$ does, then we can use $L(H|E)$ to choose between $\{H\}$. The problem is that $L(H|E)$ does *not* prove theories (Grove, 1998).

In general, we usually choose the Hypothesis with the Maximum Likelihood. In the maximum likelihood classification, the maximum of $\{L(Hh|X), \text{for } h = 0,1\}$ can be used to produce a binary representation of household and non-household pixels. In other words, a pixel is classified $\{L(Hh|X)>L(Ho|X)$; for all $o \neq h$. $\{L(Hh|X)\}$ is very often mistaken as the probability vector $\{P(Hh|X)\}$ in maximum likelihood classified satellite image (for example, Shi, 1994).

5.1.3 Group 2A and B: positional-process error and attribute-process error

For attribute data, the variables of interest would be their attribute values within the spatial location rather than the vector values of the co-ordinates. Although spatial in nature, attribute errors do not necessarily connect to positional errors in data (Drummond, 1995). These attribute variables are different from the positional variables. The two tend to interrelate to each other. An obvious example is the uncertainty of area estimate which can be computed from the positional uncertainty of the boundary (Chrisman and Yandell, 1988). For example, the estimated error of representing an area on a map, say on a USGS GIRAS digital land cover series map, was found to be 7% (Chrisman, 1982). However, for different classes of variables, their attribute uncertainty may not be reduced at the same rate as the area. It depends on the proportion of inclusions or exclusion of that class within the same area (such as the scoping process in this case study). In this case, knowledge of suitable parameters such as mean inclusion size may help (Goodchild, Sun and Yang, 1992).

Methods of handling error propagation for continuous thematic data are similar to those methods for positional data. These methods are discussed together under the same section. Generally speaking there are two methods for handling attribute- and positional-process errors: 1) Variance propagation, and 2) Monte Carlo method (Drummond, 1995).

Variance propagation

This approach estimates the associated contributing SD identified in each processing step, and applies variance propagation techniques (Mikhail and Ackerman, 1976). The error process can be modelled using the principle of the propagation of a distribution (Mikhail and Ackermann, 1976). The propagation of a distribution can be simplified to the propagation of the mean, variances and covariances (Mikhail and Ackermann, 1976; Shi, 1994). Variance propagation techniques can be used when the process consists of a limited number of steps. When the process is complex involving many steps, the analysis of variance propagation become time consuming and difficult; and may even introduce further errors. Given a functional relationship, the variance analysis can be simplified when the partial derivatives are one ('rule of thumb' estimate).

The variance propagation approach may be suitable for certain applications such as land survey when the mathematical operations performed are relatively standardised, the partial derivatives and the SD can easily be determined (a simple example of the positional accuracy of the telephone pole was given by Drummond, 1995). The mathematical model used to determine the positional accuracy of an area is sufficiently straightforward to ensure that variance propagation can easily be performed (Drummond, 1995).

There are several limitations to using the variance propagation approach:

- It neither considers the relative importance of each measurement in the process nor does it consider the mathematical relationship between the measurements.
- It assumes the unique inverse function exists which it may not.
- It is limited to linear functions. To apply the law of error propagation, a non-linear function has to be 'linearized', for example, using the Taylor expansion (Heuvelink *et al*, 1989). Such linearization would introduce an approximation error. If the function is highly non-linear, the approximation may become unacceptably large (Kuczera, 1988).
- It is usually difficult to define the set of functions describing the transformation in GIS.
- The method requires the density function to be continuously differentiable but it may not be in practice.
- It assumes that the sources of uncertainty in each measurement are independent which is usually not true (for example spatial correlation).
- It is possible only for simple algebraic relationships between input data and calculated results.

In many GIS applications the analysis performed on the data is sufficiently complex to make this analytic approach impossible. In these cases uncertainty would have to be propagated by simulation such as by the Monte Carlo method.

Monte Carlo method

The Monte Carlo technique is a general approach that is not specific to the spatial accuracy issues but can be applied with any method. It can generate distribution of any random variable at an arbitrary level of accuracy. It is easy to implement. The Monte Carlo method is best used when there is insufficient knowledge of the processes operating in a particular situation or it is impossible to develop a process to predict the entire outcomes. Given that some sets of outcomes are known, it is possible to use the statistical summary of those outcomes to determine (at random) new values that conform to the distribution of observed values.

As with the analytic approach, Monte Carlo simulation requires an error model, for example the normal distribution. Rather than work with the parameters of distribution, such as the SD, Monte Carlo simulation uses the error model to generate a sample of possible measurements. While only one measurement may have been taken, the simulated measurements can be interpreted as equally possible but fictitious outcomes of the measurement process. The samples are described as realisations of the error model.

Mathematically, the Monte Carlo computation may be regarded as estimating the value of a multiple integral $\mathbf{R}(E_1, E_2,..., E_N)$ of the sequence of random numbers $E_1, E_2,...$ (Hammersley and Handscomb, 1964). This is an unbiased estimate of the following:

$$\int_0^1 ... \int_0^1 \mathbf{R}(x_1, x_2,..., x_N)\, dx_1, x_2,..., x_N \qquad (5.8)$$

For simplicity, take the one-dimensional integral

$$J = \int_0^1 f(x)\, dx \qquad (5.9)$$

We shall refer to $\bar{f}$ as the crude Monte Carlo estimator of J. In practice, we probably would not know the value of J, so it has to be estimated. If large enough samples are generated from the Gaussian or normal error model, their relative frequencies will form the bell curve. In the Monte Carlo approach, each estimated measurement is then analysed, producing a range of results. These are then analysed to determine the uncertainty in the result, by calculating the standard deviation using the formula (3.2 with mean $\bar{z} = 0$). The factor of n in the denominator of standard error implies that in order to halve the error we must take 4 times as many observations, and so on, so that, in our example, in order to achieve 2 significant figures of accuracy (with a standard error less than 0.005) we should need to make about 3000 observations of values of f. For SD (σ) =

0.05 as acceptable standard practice, 300 observations would be required. So we take n (= 300) random numbers and evaluate the function. With 300 samples of each input measurement, analysis would produce 300 results. We should then announce the result as $j = \bar{f} \pm \sigma$ meaning that $\bar{f}$ is an observation from a distribution whose mean is j and whose standard deviation we estimate at s. Since by Central Limit Theorem we expect that the distribution of $\bar{f}$ is approximately normal (if we have sampled from the normal distribution), we may say with 95% confidence that we are within two standard deviations of the mean. The phrase 'x% confidence' signifies that with the procedure repeatedly applied in the Monte Carlo experiments, x% of results would be correct in the long run. Unless we know the value of the estimated J (in which case there would be no point in carrying out the Monte Carlo work, except for explanatory or calibration purposes discussed later), we cannot say with 100% certainty that whether any particular result is correct or not.

Monte Carlo simulation is a more general and robust technique for propagating uncertainty than mathematical analysis. It can be applied to any kind of propagation problem, no matter how complex, and to any error model (provided only that there exists a method of generating realisation).

According to the literature reviewed, there is a long history of using Monte Carlo methods within the discipline of geography. These are chronologically listed below as examples only and are by no means exhaustive:

- Hagerstrand's (1965) diffusion is based on Monte Carlo simulation.
- Hope (1968) studied migration routes.
- Besag and Diggle (1977) simulated point patterns.
- Openshaw and Taylor (1979), Openshaw (1984), Fotheringham (1989), studied the modifiable area unit problem and the effects of aggregation on area-based demographic statistics.

- Openshaw *et al* (1987) developed the geographical analysis machine in identifying clusters of Leukaemia victims.
- Openshaw *et al* (1991) examined the sensitivity of route selection for nuclear waste transport.
- Fotheringham and Wong (1991) proposed a solution to handle the modifiable area unit problem.
- Fisher (1991a) estimated the effects of soil map errors on land valuation for taxation.
- Fisher (1991b, 1992) established the effects of elevation error on the viewshed for which no process-based formula exists (used the root means squared error reported for a digital elevation model (DEM) to create alternative realisations of the elevation model, and to derive alternative realisations of the viewshed that may be determined mathematically from the DEM).
- Lee *et al* (1992) examined the errors incurred in extracting flood plains.
- Sechrist (1992) implemented Hagerstrand's (1965) diffusion model.
- Caspary and Scheuring (1992) simulated of the probability distribution of a line segment.
- Goodchild, Sun, and Yang (1992) propagated the simulated error into products obtained from the soil map data.
- Englund (1993) investigated the spatial structure of data errors.
- Fisher and Langford (1995) incorporated Monte Carlo technique with Openshaw's (1977) algorithm to generate random zonations of the neighbouring ED in order to evaluate different spatial interpolation methods (more detailed discussion later).
- Brunsdon *et al* (1998) used Monte Carlo significance test for modelling spatial non-stationarity in the geographically weighted regression.

One of the major criticisms of the Monte Carlo method is that it requires a large amount of modelling effort and computing time. In the past improving the efficiency of Monte Carlo methods were of the utmost priority. The relative efficiency (Gain) of two Monte Carlo methods for n_1, n_2 units of computing time with variance $\sigma^2{}_1$ and $\sigma^2{}_2$ respectively can be defined as the product of the labour ratio n_1/n_2 and the variance ratio $\sigma^2{}_1/\sigma^2{}_2$

(5.10). The later depends on the problem and the Monte Carlo methods; the former depends partly on the Monte Carlo method and partly on the computing machinery available.

$$n_1\sigma^2{}_1/\ n_2\sigma^2{}_2 \qquad (5.10)$$

The performance of several Monte Carlo techniques are summarised in the following table based on a standard example f(x) = e^x - 1 / e -1 (Hammersley and Handscomb, 1964):

Table 5. 2: Efficiency Ratio

Methods	Variance x	Labour =	Gain
Hit-or-miss	.34	1/1	0.34
Stratified sampling with 4 equal strata	13	1/1.3	10
Importance sampling g(x) = x	29.9	1/3	10
Control variate	60.4	1/3	30
Antithetic (max.)	2.95×10^6	1/6	460000
Orthonormal	7.2×10^5	1/3	240000

However with the advance of fast super-computer (more than 2^{1000} times faster than the computers in the 1970s), the efficiency is no longer an issue. Many applications described earlier have used the 'hit-or-miss' Monte Carlo method for its simplicity.

Another criticism of the Monte Carlo method is that as it cannot yield an analytical form of the distribution for the variable, a set of new simulation runs has to be performed for each different case (Shi, 1994). However, when spatial errors are difficult to be modelled and their nature depends on different applications, it is argued that the requirement of Monte Carlo method make it more adaptable to different context than the other methods.

5.1.4 The cumulative effect of the Group 1 and Group 2 errors with Group A x Group B interaction

While each approach reviewed above can arguably be used to analyse a specific aspect of uncertainty, the question is which one is the most appropriate solution to the problem of GIS processing in the Safer Cities Programme Evaluation. This consists of all Group 1, 2, A and B errors; their interaction as well as their cumulative effect (error propagation). Methods for a specific group of errors would be of limited use for our context. For instance, the modelling of thematic uncertainty was limited to the classified remote sensing data. In particular, the interpolation of the choropleth maps (Beats and ED) would introduce both positional and thematic errors. The spatial interpolation error is most important in GIS applications as it is most widely used in GIS applications and is much larger than the input error due to measurement (Tempfli, 1980).

Within the context of the Safer Cities Programme, positional accuracy of a polygon (say a Beat in this case) would have an effect upon its attribute value (such as crime rate), and the thematic content of the categorical coverage plays a much larger role than the spatial attribute in determining the final conclusion of the Evaluation. The spatial and thematic errors tend to interact with each other in practice, and it is not meaningful to examine them independently. However, there is a lack of methodology for specifying the interaction among the various error types and models of error propagation with a few exceptions such as the work by Lanter and Veregin (1992) and Shi (1994).

However, methods proposed from the literature that handle both spatial and thematic errors are usually based on simple demonstrative applications and erroneous assumptions. For example, the statistical method based on normal distribution around the error bands developed by Shi (1994) also has similar weaknesses described in other error band models. While theoretically interesting, it is not applicable in practice for at least three reasons. First, the errors are not independently introduced in the GIS processing. In practice, the errors introduced at each stage may not contribute equally to the total error.

They are often wrongly assumed to be equally weighted due to the difficulty in weighing the contributing factors. The error rate may change in each stage in a unpredictable fashion. Second, the true location of the line segment may not be within the buffer zone around the measured line segment. Although assuming that there is no gross error, what if the line is still completely wrong which could happen in spatial interpolation as exemplified in this case study. Third, even if the true location of a line segment is covered by the pre-defined error zone, the error distribution of the line segment need not be normal.

Other strategies include simple methods to estimate the minimum and maximum possible errors, so as to bound the range of error possible in a given application (Veregin, 1989). For example, McAlpine and Cook (1971) have proposed a simple equation to estimate the number of polygons on the composite map:

$$m_c = [\,{}^{n}_{i=1}\Sigma m_i^{1/2}\,]^2 \qquad \textbf{(5.11)}$$

where
m_c is the number of polygons on the composite map
m_i the number of polygons on the individual maps
n the number of data layers.

Thus the possible range of errors can be estimated by comparing the number of polygons in the resultant maps (say, using the overlay method) with the number estimated from the equation (5.4). While the method provides a single statement for the overall error of the composite map, it does not indicate the error in individual polygon. So it could not employ error propagation functions to proceed further in GIS processing. The simple methods such as the equation proposed by McAlpine and Cook (1971) are more appropriate to be used for providing the 'first impression' about the extent of the errors in an application. This may still be useful for analysts to make decisions on whether the extent of errors involved is large enough to warrant further investigation. The framework discussed earlier can thus be revised as follows:

1. Identify the *error index* (RMSE in this case).
2. Perform a quick evaluation of the range of errors using simple methods to check whether further assessment is required (if so, proceed to Step 3; else stop).
3. Develop *error propagation functions* within the data transformation processes using Monte Carlo dasymetric method (to be discussed next).
4. Test the utility of Step 3 by assessing spatio-thematic accuracy in the Evaluation (discussed in Section 5.3).

For categorical data, either a classification error matrix in terms of the proportion of points correctly classified (PCC) or the kappa statistics and users' and producers' accuracy can be employed as the error index (as discussed earlier in Section 5.1.2). However, the PCC index of classification accuracy is a single valued index derived from the classification error matrix. Moreover, the error indices proposed above are usually limited to the raster data type such as land cover, soil, or vegetation; and are inappropriate for numerous data sets used in the Safer Cities Programme. Its value does not permit thematic differentiation of error, i.e. it does not describe how error levels may vary from class to class (Lanter and Veregin, 1992). From the above literature review, using Monte Carlo simulation to generate the root mean squared error as index seems most appropriate for the evaluation of the Safer Cities Programme. (Note that there is no categorical data set used within the scope of this case study.)

5.2 Developing error propagation functions

An *error propagation function* may simply be defined as any unambiguous representation of the mechanisms whereby errors presented in data sources are modified by a particular data transformation function (Lanter and Veregin, 1992). The choice of error propagation function depends on the GIS data transformation function applied and the index used to measure error in data input to the transformation function. This assumes a priori knowledge of input data quality and error propagation mechanisms. There is more than one error propagation function possible. Each represents a different set of assumptions about error propagation.

Appropriate error propagation involves matching the GIS data transformation function with the error measurement index (as discussed earlier). The GIS transformation and error index serve as keys for identifying and selecting the appropriate error propagation function. Each error propagation function is specifically designed to propagate a specific error index through a particular GIS function based on a set of assumptions about error propagation (Lanter and Veregin, 1992). Each of the GIS data transformation functions used in the GIS processing (e.g. UNION, INTERSECT, and RESELECT) induces changes in the spatio-thematic accuracy. Propagation of error indices parallels the propagation of data. The implementation of the paradigm may involve drawing a network of input and output relations between spatial data layers which is similar to a data flow diagram proposed by Martin and McClure (1985), and can be regarded as an application of a 'cartographic model' (Tomlin and Berry, 1979; and Berry, 1987).

Applying Lanter and Veregin's (1992) framework to the context of the Safer Cities Programme Evaluation, the process can be represented by means of the following data flow diagram (Figure 5.4):

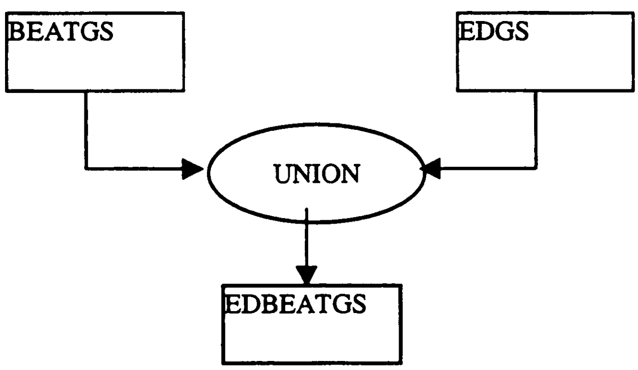


Figure 5. 4: Layer-based error propagation

Household density, Burglary risk and action intensity in the Safer Cities Programme Evaluation are implicit in the registered source layers (EDGS and BEATGS). Data transformations applied in this application link each input to an output map layer (EDBEATGS). The result is a network linking the application's source maps (EDGS and BEATGS) to its products (EDBEATGS). If source errors can be propagated through data transformation functions (UNION), the 'utility' of the GIS processing for the Safer Cities Programme Evaluation would theoretically be assessed.

5.2.1 Data transformation function, map overlay and areal interpolation

While the paradigm proposed by Lanter and Veregins (1992) focuses on the source errors of the data types, one must remember that the data transformation function itself introduces error where none existed before which would require error modelling (also the subject of this research). Different data sets collected are usually based on different spatial units, the so-called zones. For example, crime data are collected based on beats, while census data are based on EDs. Getting the information from one data set, say, census, that covers the same geographical area of another data set, say, beats, can be achieved by overlaying the two maps (beats and EDs) - a simple GIS UNION operation. However, the geographical boundaries used in the census are incompatible with beat boundaries. This is the well-known 'areal interpolation' problem (Flowerdew and Openshaw, 1987; Goodchild and Lam, 1980; Lam, 1983; and Goodchild *et al*, 1993). 'Areal interpolation' is the process of obtaining the information for one zonation of an area (known as 'target zone') by retrieving that information from another zonation of the same area, the so-called 'source zone' (Goodchild and Lam, 1980). The two zonations are very often incompatible. Areal interpolation has thus been regarded as one of the pressing needs in spatial analysis (Fotheringham and Rogerson, 1993).

There are a large number of areal interpolation methods (Goodchild *et al*, 1993; Flowerdew and Green, 1989; 1991; Flowerdew *et al*, 1991; Goodchild and Lam, 1980; Lam, 1983; Langford *et al*, 1991; Tobler, 1979). Under the classification scheme derived in Chapter three, areal interpolation is a GIS process, but would contain both spatial and attribute errors in the resultant map (Group 2 A B). Since the true attribute values in the resultant maps are not known, the errors have to be estimated by comparing the outputs from the 'overlay' method used in the Evaluation with those from a 'better' areal interpolation method. However, no method is truly satisfactory, in the sense that it is impossible to derive perfect results, and thus some error is inevitably introduced. The methods of areal interpolation need to be critically reviewed in the search of the 'better' method. These can be broadly grouped into three: cartographic, regression and surface methods (Fisher and Langford, 1995). Cartographic methods include areal weighting and the dasymetric method.

Areal weighted method

Areal weighted method is perhaps the simplest method of spatial interpolation. It is easy to implement and thus one of the most common methods. The method has been applied widely in various applications (for example, Goodchild and Lam, 1980; Goodchild *et al*, 1993) as well as being a standard function in some GIS packages such as MapInfo.

The value of a variable within the target zone (Z_t) can be calculated according to the following simple equation:

$$Z_t = \Sigma_s Z_s A_{st} / A_s \qquad (5.12)$$

where

Z_s is the value of the variable within the source zone (s)
A_{st} is the area of interaction zone (st)
A_s the area of the source zone (s)

The problem of the areal weighting method is that it wrongly assumes that the density of population within the source zone is uniform. It may only be used to quickly assess the impact of the possible errors so that a decision can be made upon further investigation using more accurate, but labour-intensive methods.

Dasymetric method

The assumption of an even distribution can be refined by using knowledge of the locality to identify smaller areas within zones that have different population densities. Maps providing such knowledge are called dasymetric maps (hence the dasymetric method) and have been used in cartography since Wright (1936). For instance, the areas of residential areas in classified satellite (Landsat) imagery can be identified and used to provide a more realistic mapping of population density (Monmonier and Schnell, 1984 applied the method in Pennsylvania). However it is not until recently that the dasymetric method has been applied to solve the areal interpolation problem (Fisher and Langford, 1995).

Regression method

The regression method computes the population of the target zone as a function of a number of control variables from another zone (and hence known as the control zone). Mathematically, this can be represented as the following general form:

$$P_t = f(x_1, x_2, ..., x_n) \tag{5. 13}$$

where $x_1, x_2, ..., x_n$ are control variables related to zone t.

Langford *et al* (1991) have examined the appropriateness of the areas of different land-cover types as the control variables. The 'shotgun' model uses the areas of different land-cover classes as independent variables to predict the population. The 'focused' model uses only the areas of high-density and low-density residential land use within the source units. The 'simple' regression model uses all residential land as one category.

The appropriate regression model for the areal interpolation may not be linear. It may be a Poisson distribution (Flowerdew, 1988; Flowerdew and Green, 1989; 1991; Flowerdew *et al*, 1991). However for the simple regression model, Langford *et al* (1991) have found little difference to the results between Poisson and linear regression models. Other control variables, such as demographic variables, may also be used. For example Flowerdew *et al* (1991) have used voting outcome, number of voters, car per household, and ethnicity as indicators to predict the population of parliamentary constituencies in Lancashire (as the target zone). Surprisingly, they have found that simple models perform better than complex models, and that the proportion of people living in a situation with more than one person per room provided the best basis for estimation.

Generally speaking control zones can be used to improve the estimates in areal interpolation (Goodchild *et al*, 1993). This is theoretically similar to the dasymetric approach. The power of control zones in improving the areal interpolation would obviously depend on the fineness of the control zones themselves. For example a simple two way split of the control zone (say above or below 400 foot contour in Lancashire by Flowerdew *et al*, 1991) would provide no improvement over the regression method.

Surface methods

The surface methods compute population density of a target zone by integrating the volume under that surface (Tobler, 1979; Bracken, 1994; Bracken and Martin, 1989; Martin, 1989). The problem with the surface methods is that although they provide a dramatic visualisation of the population density distribution, they do not provide a precise statement about the properties of the surface generated (Fisher and Langford, 1995). Although there are numerous studies of areal interpolation, very few authors have evaluated the merit of different methods. Studies are mostly based on a limited number of methods (typically one) and thus are not useful in terms of making a decision on selecting an appropriate method for a particular application (for example, Flowerdew, 1988; Flowerdew and Green, 1989; 1991; Flowerdew *et al*, 1991; Goodchild *et al*, 1993; Langford *et al*, 1991). An exception is the study by Fisher and Langford (1995). They have creatively taken the modifiable area unit (as an opportunity rather than a problem) to provide a baseline for comparison between five different methods (areal weighting, shotgun, focused, simple, and dasymetric) with ED as elemental zone and aggregated unit as target zone covering three districts of Leicestershire. They have also used classified satellite imagery to construct both regression and dasymetric models by incorporating Monte Carlo simulation with Openshaw's (1977) algorithm to generate random zonations of the neighbouring ED. They have found that the method based on dasymetric mapping consistently gave the highest accuracy of those tested, whereas the areal weighting method gave the lowest.

However, Fisher and Langford's study is not without its own limitations. There is a need to examine the effects of this accuracy on the estimation process, and to explore the potential of other data sources. The study has only covered a limited area. There is a need to broaden the area of analysis so that valid estimates of the error can be made over larger areas. The evaluation of the Safer Cities Programme has provided an opportunity for testing the potential of the methods reported by applying them in a real-world, complex situation.

5.2.2 Developing error modelling in areal interpolation for the case study

From the above critical review of the literature, I conclude that to assess the spatio-thematic accuracy of the Safer Cities Programme Evaluation, it is necessary to develop one's own unique approach based on the best practice.

Developing *error propagation functions* is based on the Monte Carlo simulation with dasymetric method as described by Fisher and Langford (1995). Their procedure would need to be modified in order to be applicable to the context of the Safer Cities Programme Evaluation. For instance, since all beats in the selected cities would be subject to error modelling, it would not be necessary to use Openshaw's (1977) algorithm to generate random zonations of the neighbouring ED.

Since data of the household density at ED level were derived from the 1991 Census data, the household density at beat level had to be derived by means of the spatial interpolation using EDs as the source zone (S) and beats as the target zone (T). Unlike Fisher and Langford's method, both the source zones and the target zones in the Safer Cities

Programme were pre-defined, and thus could not be simulated freely. So for ED-beat areal interpolation in our case study, the elementary zones E can be defined as EDs which are aggregated into any Zone A; and the target zones T as beats. The source zones S are the aggregated EDs, i.e. S=A. Using land-use maps of classified satellite (such as Landsat) imagery as a dasymetric map to provide the control zones C, which can be superimposed with T to form a binary representation of populated and unpopulated pixels. Since the population of E is known, the populations of S and T are also known (Figure 5.5).

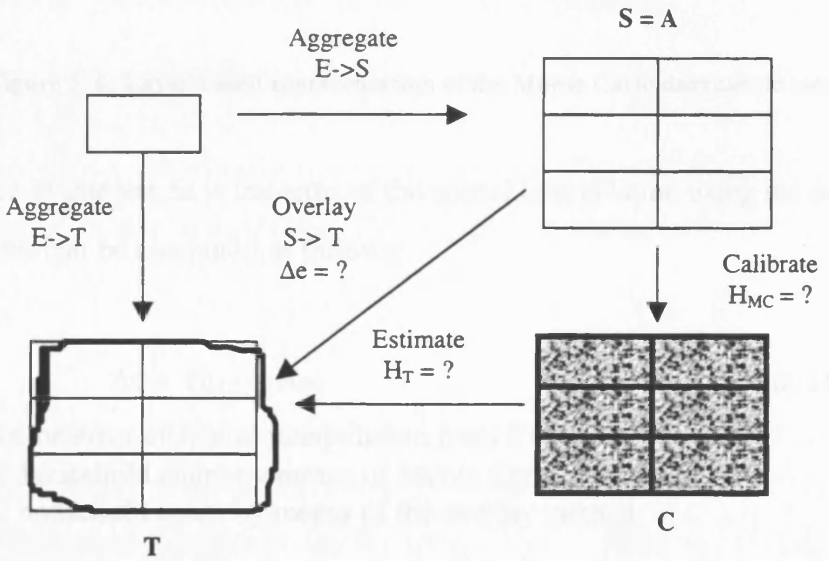


Figure 5. 5 Monte Carlo simulation and dasymetric mapping with satellite imagery to estimate population (based on 1991 census data). E: elementary zones (ED); A aggregated zones; T: target zones (beats); and C control zones (satellite imagery)

The triangle EAT (or SET) effectively defines the spatial interpolation problem of this research; and the triangle CAT (or SCT) offers the solution to that problem. The latter (SCT) can be alternatively viewed in terms of the common map overlay (layer-based representation, Lanter and Veregin, 1992, Shi, 1994) as shown in Figure 5.6.

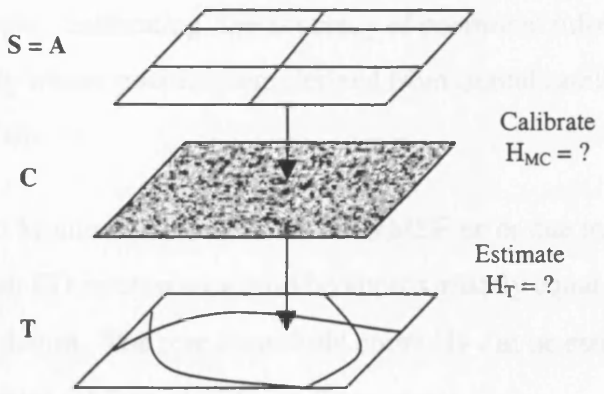


Figure 5. 6: Layer-based representation of the Monte Carlo dasymetric method

The variable of interest Δe is the error of the spatial interpolation using the overlay method. This can be computed as follows:

$$\Delta e = H_{MC} \sim H_{OV} \qquad \textbf{(5. 14)}$$

where Δe is the error of spatial interpolation from S to T
H_{MC} household count by means of Monte Carlo estimation
H_{OV} household count by means of the overlay method

and

$$H_{MC} = H_{OV} \pm \Delta e \qquad \textbf{(5. 15)}$$

The solution requires the following steps:

1. Get the satellite imagery.
2. Calibrate the satellite image using EDs (census) as source zones to estimate household counts.
3. Estimate household density in beats using the calibrated satellite image as a dasymetric map and Monte Carlo simulation.

Step 2 is necessary since the use of remote sensing data such as a Landsat image for error modelling itself also introduces an extra layer of error which needs to be corrected or 'calibrated'. For example, 'calibrating' the accuracy of positional information for specific classes of entity whose positions are derived from digital satellite imagery (see Hord and Brooner, 1976).

Assuming a non-biased Monte Carlo estimator, the RMSE error due to the simulation after the calibration with ED references would be approximately equal to the standard deviation σ of the simulation. The true household count H_T can be estimated by means of the household counts obtained from the Monte Carlo simulation:

$$H_T = H_{MC} \pm \sigma \tag{5. 16}$$

or in terms of the household counts by means of the overlay method (H_{OV}), and its error estimation (Δe) and standard deviation by means of the Monte Carlo method (σ):

$$H_T = H_{OV} \pm \Delta e \pm \sigma \tag{5. 17}$$

The above equation is consistent with the error model associated with the spatial variation proposed by Burgess and Webster (1980 a, b), Burrough (1986), Davis (1986), Heuvelink *et al*, 1989, Journal and Huijbregts (1978), Lodwick *et al* (1990) and Webster (1985). Comparing to their regionalized variable theory, the first term in the right hand side of the equation is equivalent to a function describing the structural component of the spatial (thematic in this case) variable, the second term is a spatial-correlated random component denoting the stochastic, locally-varying spatially-dependent residuals from the first, and the third term is a residual error (white noise) that is spatially uncorrelated having a mean zero.

By defining different zonal areas using Monte Carlo techniques, a precise statement about areal interpolation error can be made. Thus a whole range of error distribution can be worked out in terms of RMSE not only by comparing the population of A and C; but also C and T.

5.3 Testing the methodology

To test the utility of the layer-based paradigm, Lanter and Veregin (1992) have implemented the model using the GEOLINEUS lineage meta-database system to automatically propagate error indices through GIS spatial data transformation functions as an example. However, the hypothetical vector described error indices for a layer is no more than a snap shot of the values in the three data quality dimensions (spatial, thematic, and synoptic / temporal errors). In their detailed application example, Bayes' Theorem has been chosen as the model and the proportion of points correctly classified (PCC) as the error index. The former requires users' supplied error estimates, while the latter is inappropriate to the context of the Safer Cities Programme Evaluation as discussed earlier. The transformation of the PCC index through the application is based on the assumption that errors are uncorrelated across data-layers and are distributed uniformly across classes of the thematic attributes. The assumptions are obviously not true.

To sum up, Lanter and Veregin's (1992) paradigm provides a framework to propagate error through GIS functions, but its implementation says very little about the resultant error and its effect upon the object of users' interest. It does not define what level of error is acceptable in a given situation. Such policy decision is context specific and should be based on a consideration of the significance of the resultant error within a particular decision-making context, in this case, Safer Cities Programme Evaluation. This requires the examination of the effects of imposed perturbations (variation) on the inputs of a geographical analysis on the outputs of that analysis, a process sometime known as geographical sensitivity analysis (Lodwick *et al*, 1990). Other methods such as the geographically weighted regression can be used in the sensitivity analysis (Brunsdon *et al*, 1998). However, since the conclusion of the Safer Cities Programme Evaluation has been derived from the multi-level modelling, which would also be required to be incorporated in the sensitivity analysis of the methodology. This is to be discussed next.

5.3.1 Developing method for testing the methodology: multilevel modelling of geographical data

Multilevel models are usually concerned with the structure of the hierarchical data in the population and thus should be an appropriate tool for modelling of geographical properties such as population density. Statistical theory and examples of multilevel modelling are described by Goldstein (1995); Jones (1992); and Bryk and Raudenbusch (1992). Multi-level modelling techniques have been used to address various statistical problems in criminology such as contextual analyses (Bryk and Raudenbush, 1992; Roundtree *et al*, 1994). It has been applied to a large-scale survey for education (Mortimore *et al.*, 1988; Nuttall *et al.*, 1989). A typical multilevel model in this case would assign individuals (e.g. pupils) to level 1, class to level 2, schools to level 3 and geographical area (authorities) to level 4. Units at one level are recognised as being nested within units at the next higher level. For example, in a household survey, the level-1 units are individual respondents, the level-2 units are households and the level-3 units ED. This hierarchy is described in terms of clusters of level-1 units within each level-2 unit (*clustered population*). Clustering generally causes standard errors of regression coefficients to be underestimated in statistical analyses. If standard errors were underestimated in the evaluation of the Safer Cities Programme, it might be inferred that there was a real impact of preventive action upon the burglary risk when in fact that effect could be ascribed to chance. Correct standard errors would be estimated only if variation at ED and beat level were allowed for in the analysis.

Multilevel modelling provides an efficient way of modelling the influence of socio-economical and geographical characteristics on the outcome of the Safer Cities action. The advantages of multilevel modelling can be summarised as follows (Woodhouse *et al*, 1992):

- Coefficients in a linear model of a process occurring at one level are functions of characteristics of units at another level of a hierarchical system. These coefficients and their variances and covariances can be viewed as variables of interest.
- Coefficients of within-unit relations among variables are generally estimated better than they would be if a single level analysis was conducted for each group.
- The more appropriate model specification resolves the problem of mis-estimated precision inherent in single level analyses of hierarchically structured data.
- Longitudinal data have a nested structure-measurements within individuals. Multilevel analysis permits individuals to have their own 'growth curves' (temporal dimension).

A ordinary linear model assumes entities are completely independent, while in fact, individuals within naturally occurring groups (say, geographical areas) share common features (first law of geography). A multilevel model provides more realistic portrayal of the effects of grouping. This makes it useful for geographical studies (such as spatial correlation). The fidelity is apparent in the estimation of the standard errors of the (fixed) coefficients which are larger than the corresponding values from an ordinary least square analysis as the intra-class correlation among the measurements is taken into account. When there is a considerable between-area variation (say, 25%) it is essential to use multilevel analysis (Woodhouse *et al*, 1992).

Formulating a basic Multilevel model

The basic multi-level model is linear, based on the well-known 'ordinary least square' model. It illustrates on one level, the variation for a single geographical unit, the relationship between different burglary risk and action intensity scores. In the Safer Cities Programme Evaluation, we were interested in the variation between all beats of the city over time in order to make inference about the variation in the underlying burglary risk.

The evaluation is concerned with the relationship between the burglary risk in the individual beat-year y_i and its Safer Cities action intensity scores x_i. For each beat-year i, a linear relationship between these variables can be written as the standard algebraic equation:

$$y_i = \beta_0 + \beta_1 x_i + e_i \qquad (5.18)$$

where subscript i takes values from 1 to n, one for each case (beat-year) in the city. For the ith beat-year, x_i is the action intensity score and y_i the burglary risk respectively. The regression line, regarded as a prediction of y from x, can be expressed as:

$$Y_i = \beta_0 + \beta_1 x_i \qquad (5.19)$$

where Y_i is the burglary risk predicted for the ith beat-year by this particular summary relationship for the beat area. (Note $y_i = Y_i + e_i$)
e_i is the residual as it is that part of the burglary risk y_i which is not predicted by the regression relationship. With only one level, the beat-year variation is simply the variance e_i.

A simple two-level model (between-unit / within-unit version) consists of a random part and fixed part. The random part may either refer to the intercept (Variance components model) or the slope (Random slopes regression) of a linear model.

Variance components model

For the multilevel case, where variation between beats are included, the relationships can be expressed as:

$$Y_{ij} = \beta_{0j} + \beta_1 x_{ij} + u_j \qquad (5.20)$$

where Y is a linear function of x for i beat-year cases as before, but now they are nested in each of the j areas (or beats). (Subscript ij now refers to cases-beat.) β_1 is the regression slope (the impact of x on y) which is assumed to be constant across all beat areas and beat-years. β_{0j} is the intercept which varies across beats but has the same value for the beat-years within a beat.

The random variable $u_j = \beta_{0j} - \beta_0$ is a level-2 residual, representing the departure of the jth beat's actual intercept from the value of β_0 predicted for all beats. (Note u_j is specific to beat j, but is the same for all years in that beat.) The variance in Y is a function of both individual differences in beat-years i and differences among geographical areas j.

Since $y_{ij} = Y_{ij} + e_{ij}$, substituting the value of y_{ij} into equation (5.20), the full model can be expressed as

$$y_{ij} = \beta_{0j} + \beta_1 x_{ij} + u_j + e_{ij} \qquad \textbf{(5. 21)}$$

where

y_{ij} represents	the burglary risk
x_{ij}	action score
β_{0j}	intercept
β_1	coefficients of the action score (tot) (fixed)
u_j and e_{ij}	are the random variables.

The above is known as a variance components model, that is, a multilevel model of a simple type where the only random parameters are the intercept variance at each level.

Random slopes regression

Since both β_{0j} and β_{1j} in general can vary across beats, these coefficients are treated as random variables at level 2. The model allows slopes to vary from beat to beat.

$$y_{ij} = \beta_{0j} + \beta_{1j} x_{ij} + e_{ij} \qquad \textbf{(5. 22)}$$

β_{0j} are the intercepts and β_{1j} slopes which differ across j areas.

Let $\beta_{0j} = \beta_0 + u_j$; $\beta_{1j} = \beta_1 + v_j$, equation (5.22) becomes:

$$y_{ij} = \beta_0 + (\beta_1 + v_j)x_{ij} + u_j + e_{ij} \qquad \textbf{(5. 23)}$$

Re-arranged, we have

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + (v_j x_{ij} + u_j + e_{ij}) \qquad \textbf{(5. 24)}$$

The terms e_{ij} ,u_j and v_j are random variables [error variances] (assumed to have a mean of zero) which represent the sum of all other influences on y_{ij} (i.e. except those of the fitted explanatory variables). Random variables u_j and v_j represent influences on the β_{0j} and β_{1j} not accountable for by x respectively.

The *fixed part* in equation (5.24) contains two explanatory variables (including the arbitrary variable CONS = 1 for the intercept, which will be explained later). Their β coefficients are referred to as fixed coefficients which will be output from an analysis together with estimates of their standard errors. The ratio of a fixed coefficient estimate to its standard error can be referred against the Gaussian or normal distribution.

The set of random terms ($v_j x_{ij} + u_j + e_{ij}$) is referred to as a random part of the model. This involves two explanatory variables: the arbitrary variable *constant* (CONS) attached to u, e and the *variable* x (action score). The random terms can be expressed as *Random parameter indicator matrices:*

The covariance matrix for level 2

$$\Omega_{(2)} = \begin{array}{ccc} & X_0 & X_1 \\ X_0 & (\sigma^2_0 &) \\ X_1 & (\sigma_{01} & \sigma^2_1) \end{array}$$

where X_0 is CONS = 1 and X_1 = x

The covariance matrix for level 1

$$\Omega_{(1)} = \begin{array}{cc} & \begin{array}{cc} X_0 & X_1 \end{array} \\ \begin{array}{c} X_0 \\ X_1 \end{array} & \begin{pmatrix} \sigma^2_e & \\ 0 & 0 \end{pmatrix} \end{array}$$

Multi-level modelling was developed by Goldstein (1995) and Woodhouse *et al* (1992) into a software called ML3 used in the Safer Cities Programme Evaluation. This allows the analysis of three levels. A new version Mlwin is now available on Microsoft Window which allows analysis of n levels (Goldstein *et al*, 1998). Since ML3 was used in the Safer Cities Programme Evaluation, ML3 is also used for this case study.

In ML3, the model $y_{ij} = \beta_{0j} + \beta_1 x_{ij} + u_j + e_{ij}$ can be expressed as

$$y_{ij} = \beta_{0j}\, CONS + \beta_1\, x_{ij} + u_j CONS + e_{ij} CONS \qquad (5.25)$$

ML3 associates every parameter with an explanatory variable. For constants, put CONS = 1. β_{0j} CONS + β_1 x_{ij} are the fixed part; u_jCONS + e_{ij}CONS random part of the model.

Like other ordinary regression models, a typical distribution of the multilevel model has the following assumptions:

- x_{ij} is a fixed, known variable.
- The e_{ij} in group j is independently distributed with an expected mean value of 0 and variance of σ^2_e.
- The level two random terms u_j and v_j have a joint distribution with mean 0 and covariance matrix $\Omega_{(2)}$.
- The level one random term e_{ij} is distributed independently from each of the level two random terms.
- Multivariate normality is assumed for the random terms at each level.

The multilevel analysis allows the Safer Cities Programme evaluation to account for the burglary risk variation y_{ij} in terms of one or more features (which in reality co-exist with the burglary preventive action, for example, merely presence of other Safer Cities action or factors associate with burglary risk such as percentage of household with no car as a measure of socio-economic deprived area). *Between-unit* models for β_{0j} and β_{1j} in terms of such an extra variable z can be written as follows:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}z_j + u_{0j} \tag{5.26}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}z_j + u_{1j} \tag{5.27}$$

For random slopes model, substitute the values of β_{0j} and β_{1j} equations (5.26), (5.27) into (5.22), we have:

$$y_{ij} = (\gamma_{00} + \gamma_{01}z_j + u_{0j}) + (\gamma_{10} + \gamma_{11}z_j + u_{1j})x_{ij} + e_{ij} \tag{5.28}$$

(For variance components model, use equation 5.21 instead of 5.22.)

Re-arranged, this produces a single equation version of the model:

$$y_{ij} = \gamma_{00} + \gamma_{01}z_j + \gamma_{10}x_{ij} + \gamma_{11}z_jx_{ij} + (u_{0j} + u_{1j}x_{ij} + e_{ij}) \tag{5.29}$$

where
y_{ij} = burglary risk
γ_{00} = intercept
$(\gamma_{10} + u_{1j})$ = coefficients of the action score (fixed + random)
x_{ij} = action score
γ_{01} = coefficient of other explanatory variable e.g. presence of action
z_j = other explanatory variable e.g. action presence [1 or 0]
$\gamma_{11}z_jx_{ij}$ = cov = covariates (e.g. census x action scores x action presence)
$(u_{0j} + e_{ij})$ = residuals of levels 2 and 1 respectively.

The product z_jx_{1ij} variable is a between-level interaction. The *fixed part* in equation (5.29) contains three explanatory variables. Their coefficients γs are referred to as fixed coefficients. Among the outputs of an analysis will be the fixed coefficient estimates γ_{00};γ_{01}; γ_{10} and γ_{11} together with estimates of their standard errors. $(u_{0j} + u_{1j}x_{ij} + e_{ij})$ is the random part. (The equation would be similar for variance components except the term would be $u_{ij}x_{ij}$ missing.)

5.4 Chapter Summary

This Chapter has provided a broad literature review and critically evaluated some of the more relevant literature in search of a methodology that can be applied to the context of the Safer Cities Programme Evaluation. Determination of spatio-thematic accuracy involves several processing steps. Errors are inherent within each processing step. The error can be converted to a SD value or RMSE as an error index. A rough estimate of the accuracy can be made using a simple arithmetic 'rule of thumb'. Each processing step may contribute its inherent error in a different way. Variance propagation can be used to determine a 'weight', which affects each of the contributing SD differently. However variance propagation computations are complex even if computer assisted. For the more complex GIS applications, a Monte Carlo approach to error estimation remains to be the most robust method. Based on Lanter and Veregin's (1992) paradigm and Fisher and Langford's (1995) assessment of areal interpolation methods, an innovative methodology has been developed to handle the spatio-thematic accuracy issues in the evaluation of the Safer Cities Programme. The following step by step procedures have been formulated to implement the methodology (in the next chapter):

1. Identify the *error index* (RMSE in this case).
2. Perform a quick evaluation of the range of errors using simple methods (such as area weighting) to check whether further assessment is required (if so, proceed to Step 3; else stop).
3. Develop *error propagation functions* within the data transformation processes using Monte Carlo dasymetric method.
4. Test the utility of Step 3 by assessing spatio-thematic accuracy in the Evaluation using the multilevel modelling technique.

Step 3 requires following sub-steps:

1. Get the satellite imagery
2. Calibrate the satellite image using EDs (census) as source zones to estimate household counts.
3. Estimate household density in beats using the calibrated satellite image as a dasymetric map and Monte Carlo simulation.

Chapter Six
Preliminary estimation of possible sources of errors

The aim of this Chapter is to provide a quick evaluation of the range of errors as the next step of error handling described in the previous chapter. The methods include an arithmetic 'rule of thumb' formula proposed by McAlpine and Cook (1971), and the areal weighted method (Goodchild and Lam, 1980; Goodchild *et al*, 1993). These methods allow checking of whether further assessment is required. First Section 6.1 reviews the possible sources of error in the evaluation based on the classification proposed in Chapter 3. McAlpine and Cook's formula is implemented following this review process as appropriate (Section 6.2). Section 6.3 describes the implementation and the results of the area weighted method, and Section 6.4 is a summary of this chapter.

6.1 Review of the possible sources of errors in GIS processing steps

This section attempts to identify the possible sources of errors (based on the classification proposed in Chapter 3) involved in the process used in the evaluation through a number of transformation function. This also shows how different groups of errors interact in different stages of GIS processing. The GIS transformations are based on ARC/INFO terminology but they can be applied in a broader, more conceptual sense. Getting the spatial data ready for scoping and scoring involves the following steps:

Step 1. Identify the area affected by the Safer Cities schemes, the so-called zone of influence (ZI).

The Safer Cities Co-ordinators were supplied with 1:10,000 OS maps and asked to identify the geographical location where action took place using 91 EDs as the spatial unit. The information was stored in INFO in terms of the following logical schema: Action_locations (<u>scheme-ID</u>, <u>beat-ID</u>, <u>91ED-ID</u>).

Some identification error (Group 1) in this process was likely as the co-ordinators were forced to represent their implicit spatial knowledge of the preventive action explicitly in the form of a map. Some schemes (such as, a city wide mobile bus to give out leaflets on burglary prevention) were difficult to locate spatially.

Step 2. Measure outcome variables using the so-called zone of detection (ZOD).
For crime data analysis, a ZD is simply a beat. To start, the spatial coding of ZI and ZD required digitising both 91EDs (supplied by OPCS as a part of the census data) and beats (mostly digitised by GDC and some in-house). This contained digitising error (Group 1A errors). Even though if one assumed that all input to GIS was carefully checked so that there were no gross errors in the input data, some error caused by operators may still exist as discussed in Chapter 5.

For the surveys, this involved longitudinal random sampling of 7679 residents in the high crime area before and after the implementation of the Safer Cities Programme. The residents were clustered in 406 ED sampling points, each of which contributed to a ZD for scoping purposes. The respondents were aggregated according to their ED locations as the ZD for the scoping purpose. Some of the ZD did overlap with schemes ZI. These were compared with outcome measures in ZDs including those that did not overlap with ZI. The Survey analysis has not been included within the scope of this study, but issues related to it will be discussed in Chapter 12.

Step 3. Detect the zone of overlap (ZO) between ZI *and* ZD
Scoping and scoring also involved overlap in time as well as in space. This should not have caused many errors, as the units of both ZI and ZD are consistent. This should be simply a matter of database management.

The spatial overlay of beats required areal interpolation by overlaying beat maps upon ED boundaries. For this operation, we took a pragmatic approach. The beat coverage was overlaid upon ED coverage to create a new coverage called EDBEATGS using

ARC/INFO. This resulted in a many-to-many relationship between beats and EDs, in other words, one beat may contribute to multiple polygons in the derived coverage with the same ED-ID, and vice versa. For the Safer Cities Programme Evaluation data preparation required a one to many relationship between beats and EDs so that each beat 'contained' a number of unique ED-IDs. Those polygons with ED-IDs occurring more than once were identified using the ARC/INFO FREQUENCY command to count the occurrences of ED-ID in the polygon attribute tables (.PAT files). The sliver polygons with an area less than 10,000 sq. m were ELIMINATEd. Table 6.1 shows a continuous reduction in ED polygons in Bristol and Coventry as a result of the operation. The remaining polygons with a shared ED-ID were forced to merge using the ARC/INFO DISSOLVE command.

Table 6. 1: Reduction in ED polygons in GIS processing

Cities	Number of Polygons with IDs*	Number of beats	Number of EDS	Number of derived polygons	Polygons > 10000 sq. m
Bristol	58	824	1367	1010	996
Coventry	46	601	1243	724	711

* Some polygons were slivers outside the city so that they either lacked beat numbers or IDs.

The above operation may be efficient for our GIS processing purpose, but is hardly satisfactory error handling and is the subject of this research.

6.2 A 'rule of thumb' formula

A quick check can be implemented using the model proposed by McAlpine and Cook (1971) to estimate the number of polygons on the composite map:

$$m_c = [\sum_{i=1}^{n} m_i^{1/2}]^2 \qquad (6.1)$$

where
m_c is the number of polygons on the composite map
m_i the number of polygons on the individual maps
n the number of data layers.

The following table shows the comparison between the estimated and the final overlaid polygons in Bristol and Coventry:

Table 6. 2: The estimated and the final overlaid polygons

	m_1	m_2	$m_{c'}$	m_c	$m_c - m_{c'}$
Bristol	58	824	883	996	113
Coventry	46	601	648	711	63

m_1 number of beats
m_2 number of EDs
$m_{c'}$ estimate of derived polygons using Formulus (6.1)
m_c number of derived polygons using the overlay method.

If the model gives a correct estimate, this implies that some 113 (12.8%) and 63 (9.72%) polygons are wrongly assigned to the beats in Bristol and Coventry respectively. This represents 12% overall errors in practice. The errors of individual areas could be higher.

6.3 Applying areal weighted method

Since at this stage, we are only interested in the change in the surface areas as an indicator of the size of the error due to the overlay method, the equation (5.12) can be simplified (Goodchild and Lam, 1980; Goodchild *et al*, 1993). The weighted factor w can be computed as a simple ratio between the original area (Area 0) and the new area (Area 1) as a result of the overlay operation:

$$w = \frac{A_0}{A_1} \qquad \textbf{(6. 2)}$$

where A_0 is original area
and A_1 new area due to overlay operation
The error rate R can be computed using the following formula:

$$R = \frac{A_1 - A_0}{A_0} \qquad \textbf{(6. 3)}$$

The results of the areal weighted methods for Coventry and Bristol are shown in Tables 6.3 and 6.4 respectively. Figures 6.1 and 6.2 provide a summary of the spatial error in percent for Coventry and Bristol respectively.

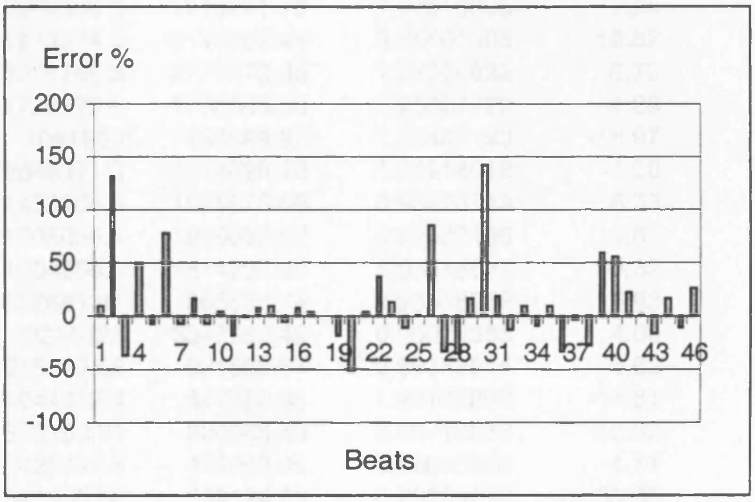


Figure 6. 1: Spatial error of overlay method estimated by area weighted method (Coventry)

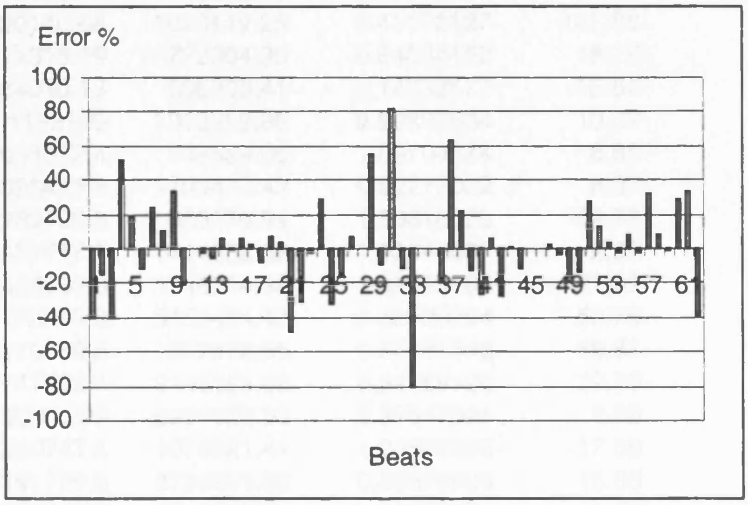


Figure 6. 2: Spatial error of overlay method estimated by area weighted method (Bristol)

Table 6. 3: Results of areal weighted method in Coventry (area in m²)

Beat-ID	Area 0	Area 1	Weight	Error %
2	6049201.5	6618963.60	0.91391974	9.42
3	5764331	13353504.95	0.43167176	131.66
4	10937334	6943106.18	1.57527967	-36.52
5	6183165.5	9314484.13	0.66382265	50.64
6	2337842.3	2147220.38	1.08877613	-8.15
7	1709568	3050761.15	0.56037425	78.45
8	2069306.6	1919241.76	1.07818965	-7.25
9	1875274.3	2190702.41	0.85601508	16.82
10	2976786.3	2776673.35	1.07206932	-6.72
11	1720179.6	1793955.50	0.95887529	4.29
14	1091660	884599.97	1.23407193	-18.97
15	2646371.3	2614629.22	1.01214018	-1.20
16	1432208.3	1525675.08	0.93873743	6.53
17	1708036.4	1856093.32	0.92023196	8.67
18	1704855.4	1614207.52	1.05615627	-5.32
19	622681.81	665837.74	0.93518552	6.93
20	3216325	3347783.49	0.96073268	4.09
21	915899.69	921586.67	0.99382914	0.62
22	1044172.4	847722.35	1.23173867	-18.81
23	673103.81	320968.43	2.09710286	-52.32
24	425497.5	442968.96	0.96055828	4.11
25	541950.06	739124.83	0.73323212	36.38
26	3633323.8	4091189.55	0.88808493	12.60
27	3670629.8	3270820.00	1.12223534	-10.89
28	1447777.4	1486036.59	0.97425420	2.64
29	487856.22	906596.33	0.53811846	85.83
30	463154.59	312251.15	1.48327587	-32.58
31	432136.31	283975.60	1.52173748	-34.29
32	458229.44	533212.48	0.85937494	16.36
33	420140.66	1020449.25	0.41172127	142.88
34	611316.19	722304.30	0.84634162	18.16
35	684040.13	596203.41	1.14732677	-12.84
36	1193729	1313282.86	0.90896564	10.02
37	1034072.4	942599.06	1.09704374	-8.85
38	1098485.8	1190473.42	0.92273022	8.37
39	1289765.8	855176.94	1.50818590	-33.70
40	1457717.8	1414682.62	1.03042037	-2.95
41	2425538.8	1748214.38	1.38743785	-27.92
42	3228783	5159424.47	0.62580294	59.79
43	1270580.5	1979672.65	0.64181343	55.81
44	1747692.4	2145299.52	0.81466126	22.75
45	2242299	2291623.33	0.97847625	2.20
46	1300727.8	1078821.41	1.20569335	-17.06
47	3191789.5	3729601.56	0.85579906	16.85
100	3550448.3	3140518.01	1.13052951	-11.55
101	1512473.5	1897692.28	0.79700672	25.47
Mean	**2097792.59**	**2347824.61**	**0.89350481**	**10.66**

Table 6. 4: Results of area weighted method in Bristol

Beat-ID	Area 0	Area 1	Weight	Error %
1	8087891	4831558.24	1.67397154	-40.26
2	5308205.5	4541126.48	1.16891822	-14.45
3	605837.19	359081.04	1.68718793	-40.73
4	1877857.3	2842237.10	0.66069692	51.36
5	2443735.5	2879946.44	0.84853505	17.85
6	2485665.5	2206208.46	1.12666846	-11.24
7	2036322	2528645.25	0.80530157	24.18
8	5728531	6307044.06	0.90827509	10.10
9	1229332.3	1639125.23	0.74999291	33.33
10	1847501.6	1455934.70	1.26894537	-21.19
11	1528600.1	1521464.96	1.00468965	-0.47
12	1168529.6	1136782.84	1.02792684	-2.72
13	1865524.1	1760116.79	1.05988654	-5.65
14	875925.31	922645.40	0.94936289	5.33
15	768416.94	760720.71	1.01011703	-1.00
16	456065	482409.28	0.94539019	5.78
17	548592.13	562405.75	0.97543834	2.52
18	341469.34	314202.42	1.08678139	-7.99
19	461924.53	494639.55	0.93386089	7.08
20	500596.19	518080.97	0.96625087	3.49
21	456888.56	237654.00	1.92249468	-47.98
22	682793.56	470139.36	1.45232164	-31.14
23	363850.59	356535.06	1.02051840	-2.01
24	319404	410056.03	0.77892771	28.38
25	686003.81	463002.78	1.48164080	-32.51
26	788534.06	669428.93	1.17792050	-15.10
27	593502.13	631551.42	0.93975267	6.41
28	342000.25	342000.25	1	0
29	283934.38	441073.04	0.64373551	55.34
30	642681.06	584790.92	1.09899288	-9.01
31	1055302.1	1917042.21	0.55048454	81.66
32	148533.41	131153.95	1.13251192	-11.70
33	161494.53	32395.61	4.98507418	-79.94
34	121194.41	121194.41	1	0
35	216433.41	216433.41	1	0
36	644246.81	515971.25	1.24860991	-19.91
37	767213.5	1255723.84	0.61097311	63.67
38	841444.06	1021426.80	0.82379282	21.39
39	998221.44	875261.65	1.14048347	-12.32
40	1710821.3	1256144.04	1.36196268	-26.58
41	3496348.5	3679895.29	0.95012174	5.25
42	1108890.1	805348.09	1.37690785	-27.37
43	9226930	9283245.12	0.99393368	0.61
44	881648.13	777748.27	1.13359060	-11.78
45	1116816.3	1116816.30	1	0
46	2673409.3	2673409.30	1.00000000	0.00
47	1933120.3	1988596.40	0.97210289	2.87
48	2010686.5	1933451.55	1.03994667	-3.84
49	1721084	1380850.13	1.24639450	-19.77

50	2907408	2521937.25	1.15284708	-13.26
51	2118766.8	2710590.58	0.78166242	27.93
52	1249670.9	1408983.69	0.88693070	12.75
53	571684.31	589616.69	0.96958637	3.14
54	1221406.8	1186777.64	1.02917915	-2.84
55	2352271.3	2528624.21	0.93025737	7.50
56	1030637.8	1008390.26	1.02206243	-2.16
57	3950992.8	5231379.62	0.75524873	32.41
58	3262265	3539893.12	0.92157161	8.51
60	1905337.4	1872587.00	1.01748939	-1.72
61	3626344	4684174.56	0.77416927	29.17
62	4779629	6361296.40	0.75136084	33.09
63	1764573	1065105.79	1.65671148	-39.64
Mean	**1724208.705**	**1558539.25**	**1.10629790**	**-9.61**

From the above results, the areal errors due to the overlay operation range from -80 to +81% with mean 10% for Bristol. For Coventry, the range is as large as from -52% to +143% with mean error = 11% approximately. The average results are comparable to 12% cartographic errors described earlier. The consistency of the error rates is probably due to the same assumption of homogeneity of spatial distribution used by both methods. There is a large variation within each city. The overall errors could be higher or lower depending on the scoping and scoring processes. The beats that did not have any action would have no error at all in their action scores.

The Root Mean Squared Errors (RMSE) as defined in Chapter 3 can also be computed for the city as a whole. In this case:

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{1}^{n}(Z_{tmi}-Z_{smi})^2}{m}} \qquad \textbf{(6. 4)}$$

where m is the number of target zones in a data layer
Z_{tmi} is the estimated values of the target zone based on the area weighted method
Z_{smi} the estimated value of the target zone without weighting

The RMSE of the beat areas are 0.36 and 0.37 for Bristol and Coventry respectively. However the use of a single RMSE value is rather limited as it can only be applied to the whole city rather than individual beats.

Statistically, it is a common practice to attribute 5% error as due to the chance factor and hence traditionally regarded as 'insignificant'. For an error margin of 10 percent or above, it may have a 'significant' effect upon the data analyses of the Safer Cities Programme Evaluation. This warrants further investigation of:

- improved estimation of the error rates
- RMSE for each beat
- how the above may affect the conclusion of the Safer Cities Programme Evaluation.

These points are the major concern of this research. Furthermore, the use of the areal weighted method cannot be accepted without further investigation due to its erroneous assumption that the population is evenly distributed across the area. Previous research shows that the dasymetric mapping method is a much more accurate method for cross area estimation than the areal weighted method (Fisher and Langford 1995). The modified methodology (as described in the previous chapter) will be implemented next. Nevertheless, establishing the worse case is still worthwhile to set a baseline of errors for future comparison.

6.4 Chapter Summary

This Chapter has rapidly evaluated the range of errors:

- Using an arithmetic 'rule of thumb' formula for overlaying of the categorical data (McAlpine and Cook, 1971), the error due to the overlay operation has been estimated to be 12%.
- Using the areal weighted method, the errors range from -80 to +81% for Bristol, and from -52% to +143% for Coventry. The average errors of 10-11% are comparable to the cartographic errors estimated using McAlpine and Cook's formula (10-13%).

Based on these results, it was decided that further assessment was required which forms the rest of this thesis. The next Chapter describes how a more accurate method for the error estimation can be implemented.

Chapter Seven
Implementation of error modelling

This chapter describes the implementation of the Monte Carlo dasymetric method developed in Chapter 5. This (as described in the earlier chapter) involves the following steps:

1. Get the satellite imagery.
2. Calibrate the satellite image using EDs (census) as source zones to estimate household counts.
3. Estimate household density in beats using the calibrated satellite image as a dasymetric map and Monte Carlo simulation.

The accuracy of the classified Landsat imagery and the Monte Carlo simulation can be assessed and calibrated by comparing the results based on the more accurate ED maps and the household pixel counts respectively. Step 2 Calibrating Landsat TM imagery involves the following sub-steps:

2.1 Rasterize ED boundaries.
2.2 Overlay the rasterized ED with the classified Landsat TM imagery.
2.3 Estimate the household density per ED using the classified imagery.
2.4 Calibrate the Landsat imagery by comparing its pixel counts with ED census counts and estimating the average number of households per pixel (h/p) for each beat.

To estimate the household density for each beat requires the following steps:

3.1 Rasterize beat maps.
3.2 Overlay rasterized beat map with the classified Landsat imagery.
3.3 Perform Monte Carlo simulation to estimate the household counts per beat using h/p obtained from Step 2 as a parameter.

Steps 2.1 - 2.2 and 3.1 - 3.2 are virtually identical. (The only difference is the rasterized maps being either ED or beat.) These steps can be implemented using a raster-based GIS called IDRISI. This is described in Section 7.2. Section 7.3 describes how Steps 2.3 - 2.4 and 3.3 can be implemented together within the programming procedures using Lisp-Stat. The rationale of choosing Lisp-Stat for Monte Carlo and GIS applications is described in Section 7.3.1. Section 7.3.2 describes the testing of the random number generator. The next section describes the process of acquiring the Landsat imagery and transforming it into a usable form for this case study.

7.1 Getting the satellite imagery *Landsat Thematic Mapper*

Five satellites (named Landsat) were first launched between 1972 (Landsat 1) and 1982. The Thematic Mapper (TM) was carried on Landsat 4 and 5, being a Multi-Spectral Scanner system with spectral, radiometric and geometric improvement. It scanned each line from west to east while the satellite was moving its orbit. The TM had seven bands in total, their sensitivity and resolution are shown in Table 7.1 below (MIDAS CSS, 1995).

Table 7. 1: Sensitivity and resolution used on the Landsat-4 and -5 Missions

Band	Sensitivity (μm)	Nominal spectral location	Spatial resolution
1	0.45-.52	Blue	30
2	0.52-.6	Green	30
3	0.6-.69	Red	30
4	0.76-.9	Neat-IR	30
5	1.55-1.75	Mid-IR	30
6	10.4-12.5	Thermal-IR	120
7	2.08-2.35	Mid-IR	30

These Bands can be used to detect land covers such as coastal waters, soil and vegetation, different coniferous and deciduous species (Campbell, 1987; MIDAS 1995).

Importing the imagery data

Each scene was identified by its path and row numbers, corresponding to its longitude and latitude respectively. For example, Bristol was in scene 203/24; and Coventry in 203/23. The size of a scene was about 180 x 160 km^2. This covered a larger area than a city. So the selected cities had to be 'cut out' from the scenes. The 'cut-out' image files were then converted into IDRISI format for further processing. Getting the satellite imagery into a useable form involves a number of logical processing steps:

1. Pre-process.
2. Reduce spectral information.
3. Classify (supervised / unsupervised classification).
   3.1. Select training data.
4. Post-process.

The above sequence is based on Campbell (1987, p 245, Figure 9.1) with some modification. The scheme described by Campbell is an idealised sequence, which would be different in practice (depending on the type of GIS used). For instance no particular order is required for Step 1 and 2. Selecting training data is more appropriate to be regarded as a sub-step within image Classification.

7.1.1 Pre-processing

Satellite imagery is prone to inaccuracy due to 1) the radiometric, and 2) geometric errors (Campbell, 1987). Radiometric errors are caused by atmospheric scatter, mechanical deficiencies in the scanning device and data transcription. The geometric distortion is due to the inherent difference between the imagery and the earth surfaces as both are moving when the imagery is acquired and the projection of the earth curvature to a flat surface in the subsequent mapping. Pre-processing was required to correct those errors (radiometric and geometric pre-processing respectively).

7.1.1.1 radiometric pre-processing

Since physical models of scattering at the level of individual particles and molecules were too complex to apply, a simple radiometric correction was carried out (with destriping if necessary). This was done by examining the reflectance from features of known brightness within the image (e.g. the sea and a larger water body in Bristol). It was assumed that reflection from water was low.

Image enhancement using linear stretch. Normally image enhancement is used for images that are to be interpreted manually for their visual effect and is not essential for further digital analysis. However, since IDRISI only contains pixel values ranging from 0 to 255, a linear stretch may be required for certain bands to convert the original digital values to the new minimum and maximum values.

7.1.1.2 Geometric correction by resampling

Geometric correction is a process known as image registration. It determines the co-ordinates of each pixel. Applying an analytical approach to determine geometry and motion of the sensor is difficult and can only correct some of the geometric errors. Instead, a simpler approach known as resampling was used.

Resampling takes two steps:

1. Record the coordinates of some pixels that can be identified with precision on the ground (known as *ground control points*) in both the image and the target coordinate system (national grid in this instance)
2. Transform the rest of the pixels with reference to the ground control points.

In general, resampling tends to reduce class means of the training data and increase variances of the original data (Kovalick, 1983).

Identifying ground control points

Selection of ground control points is a matter of balance between high precision of the pixel (with high confidence) and large dispersion over the image. The number of good ground control points required is usually small (about 16, Bernstein, 1983). The ground control points for each city were collected by identifying their known features (for example: distinct water bodies and stream junctions, intersection of major highway), and referring their co-ordinates with the 1:10000 OS map. Some 20 ground control points were selected and placed throughout the image, avoiding gaps and clusters.

Transforming the rest of the pixels

By reference to the ground control points, a number of techniques can be employed for transforming the rest of the pixels:

- nearest neighbour;
- bilinear interpolation; and
- Cubic convolution.

The *nearest neighbour* method estimates each value from the nearest point on the reference grid. The method is simple, computational efficient (Kovalick, 1983) and preserves the original value but creates positional errors.

The *bilinear interpolation* method estimates each value in the output image by calculating a weighed average of the four neighbours in the reference image. Each value in the reference image is weighed in proportion to its closeness to the point in the out image. Bilinear interpolation is a more accurate method and has a natural look but original values are lost and variance in the spectral information reduced.

The *cubic convolution* estimates each value in the output matrix by assessing values of a neighbourhood of 16 - 25 pixels in the reference grid. This method is similar to the bilinear interpolation with even more accurate and natural look, but the agitation of the original values is even more drastic. Nevertheless, attempt is made to preserve spectral variance.

For the purpose of this case study, the cubic convolution resampling was used because accuracy is more important than the computational efficiency. Preserving the original values of the pixels was not required for the post-processing. A small amount of positional error at the beat boundaries could be calibrated (discussed later). Three bands of the Landsat imagery of each city were combined using IDRISI to create the rectified image (which was used as the corresponding file for image classification at the later stage).

7.1.2 Spectral information reduction

Spectral information reduction is used to reduce the number of spectral bands, while simultaneously retaining the required information. In order to reduce the cost of analysis, contributions of noise and error, only the more potent bands are selected. The effectiveness of this procedure can be evaluated by computing the optimum coefficients of combinations of the selected set of bands. The transformed value of pixels after the combination (A) can be assumed to be a linear combination of the following form (Campbell, 1987):

$$A = \sum_{1}^{n} C_i X_i \qquad (7.1)$$

where Xi = X_1, X_2, ..., X_n are pixel values in n bands respectively
Ci = C_1, C_2, ..., C_n coefficients (components) of the original values in the respective bands.

The optimal coefficients can be computed by means of principal components analysis (Davis, 1986; Gould, 1967). Band 6 (thermal infrared) was excluded in the principal components analysis as it conveys very little information of interest and has a much larger spatial resolution. The rectified bands were converted into six components. The principal component images were written as integer binary files to account for negative values that might result from the axis rotation. These were converted back to byte binary. Table 7.2 summarised the results of the principal components analysis. (Appendix 7.1 shows the variance and co-variance matrix of the components.)

Table 7. 2: summary of the principal components analysis results.

Bristol	1	2	3	4	5	6
% VAR	79.040	15.240	3.500	1.870	0.029	0.070
Eigenval	1334.040	257.180	59.020	31.540	4.940	1.170
Eigvec.1	-0.027	0.357	0.601	0.029	-0.678	-0.221
2	0.012	-0.217	0.332	-0.060	0.110	0.909
3	0.200	-0.427	0.421	-0.135	0.706	-0.350
4	0.562	0.658	0.490	0.438	0.079	-0.037
5	0.801	-0.409	-0.327	-0.262	-0.121	0.008
6	0.199	-0.206	-0.050	0.952	0.091	0.018

Coventry	1	2	3	4	5	6
Var%	90.860	7.540	1.340	0.120	0.090	0.040
Eigenval	5387.500	447.310	79.700	6.840	5.570	2.580
Eigvec.1	0.493	0.323	-0.534	-0.142	-0.423	-0.410
2	0.221	0.145	0.240	-0.064	-0.423	0.912
3	0.212	0.287	0.003	-0.098	-0.190	0.000
4	0.631	-0.746	0.003	0.181	0.879	0.000
5	0.477	0.307	0.732	-0.379	-0.110	0.000
6	0.193	0.376	0.177	0.889	-0.009	0.000

Eigen values (eigenval) are pixel values in the six components respectively; and the eigen vectors (eignvec) represent the coefficients of the components in relation to the respective transformed channels.

The above data summary shows that component 1 provides maximum information for the single band (which accounts for 91% of the variance) while Component 6 accounts for less than .05% of the variation in the data. From the above analysis, just selecting Components 1, 2 and 3 would have covered more than 95% of the variance. However, it was desirable to use Component 4 to delineate water bodies, and Component 5 to differentiate vegetation from residential areas. Only Component 6 was subsequently discarded. Components 1, 2, 3, 4 and 5 were selected for further image processing. All together they account for more than 99.5% of the total variance.

7.1.3 Classifying Landsat imagery

Having selected an appropriate number of bands and formed a rectified image, each pixel of that image needs to be assigned to a particular class of interest (*information class*) - a process known as *image classification.* This is done by examining the Landsat image and grouping together those pixels that have similar spectral values (*spectral classes*), then assigning them to the known information classes. For this case study only residential areas (value 1) and non-residential areas (0) are required, though many more classes are present in the image (such as water and park areas). The later are regarded as *spectral subclasses* within the information class of non-residential areas.

A simple type of classifier known as a point classifier operates upon each pixel independently according to its spectral value. It is easy to program (and thus provided by numerous image analysis packages) but it does not provide 'image texture' by considering the neighbourhood relationship (*intra-class variation*) between pixels, which usually requires human interpretation. This classification approach requiring human supervision is referred to as *supervised classification,* as oppose to the complete automated process (unsupervised classification).

Unsupervised classification classifies the natural groups from their reflectance patterns within multi-spectral data. A typical unsupervised classification uses cluster analysis to distinguish differences in reflectance values across a set of bands (for example using simple point classifier or AMOEBA classifier with a contiguity constraint, see Bryant, 1979). Since this is mostly an automatic process, no human error would occur and no users' prior knowledge of the area would be required during the classification process (until a later stage for interpretation). However, measures of accuracy based on the unsupervised classification can be misleading as there is a lack of agreement between overall results of discrimination for different discriminators (Kershaw, 1987). Furthermore, as there is usually no one-to-one relationship between spectral classes and the corresponding information classes, the mapping between the two, which is the primary object of image classification, would remain unresolved.

Supervised classification classifies pixels (of unknown identity) in the image by referring to a sample of pixels (which are of known ground cover designated by the users). These user-defined areas of sampled pixels are called *training sites*. Although selecting representative samples of training sites requires detailed knowledge of the area and is relatively time consuming, supervised classification is used for this research as control of the information classes is maintained. The information classes are a simple binary classification in this case study.

Campbell (1987) describes the idealised sequence for conducting supervised classification as follows:

1. Prepare the list of information classes.
2. Select and define training data.
3. Reiterate the above with modifications as necessary to ensure homogeneous training data.
4. Conduct classification.
5. Evaluate the performance of classification.

The last step of the above was covered by calibrating the Landsat image for further processing. Step one was defined as simple residential and non-residential areas as discussed earlier. Steps 2 - 4 were implemented using IDRISI, which took the following steps (see IDRISI User's Guide):

1. Create training sites.
2. Create signature files.
3. Apply a classification procedure to the image bands using the signature created from the training sites.
4. Characterise categories across all the bands to create a signature or spectral response pattern for each information class.
5. Use the signature for each informational class to classify the full image by determining the most likely class for each individual pixel in the image.

Steps 1 and 2 are equivalent to selecting training data. Steps 3 to 5 are to conduct classification using the IDRISI program.

Selecting training data

Selecting the training sites has a far more significant effect upon the classification accuracy than the choice of different classification algorithm (Kershaw and Fuller, 1992; Scholz *et al.*, 1979). For accuracy of creating training sites, shape is not an important factor while size, location, placement and especially uniformity (*homogeneity*) are important (Campbell, 1987). As a rule of thumb, according to the IDRISI User's Guide, the minimal number of pixels = 10 x number of bands for each information class which is somewhat smaller than the 100 pixels recommended by Campbell (1987). Each information class may contain a number of training sites. So a training site could be small. Specific areas of training sites with known features were identified by examining the satellite image and referring to the 1:10000 OS map if necessary. Using IDRISI, each training site was specified by on-screen digitising the known feature as a polygon.

Once signature files have been made from the training sites, several algorithms can be employed to conduct classification (see Campbell, 1987 for various classification techniques). Maximum likelihood is a powerful algorithm and was selected for classifying the Landsat image of the two cities. (Other techniques such as minimum distance and parallelepiped classification were tried with less satisfactory results.)

The basic principle of the maximum likelihood classification has been discussed in Chapter 5. It uses a *prior* knowledge of the probability that any pixel belongs to a given class is expected to occupy over the entire study region. When such conditional probability is absent, equal likelihood is assumed. The classified images for Bristol and Coventry are shown in Figures 7.1 and 7.2 respectively.

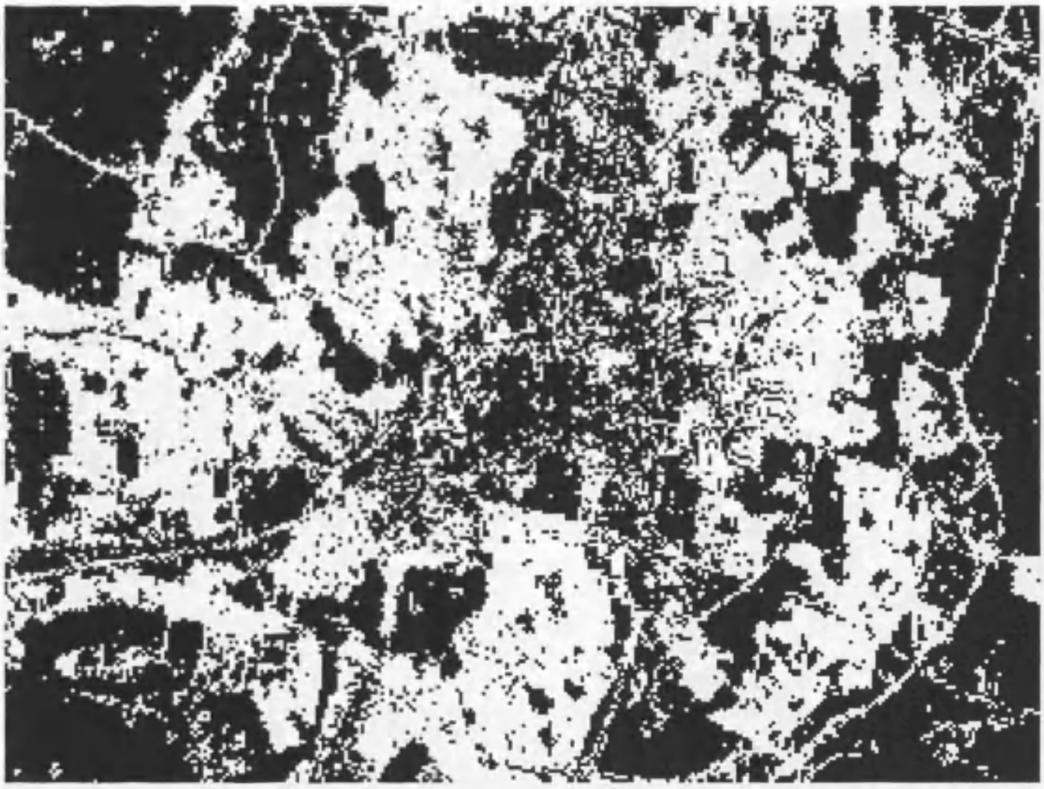


Figure 7. 1:Landsat TM images using Maximum Likelihood Classification (Coventry). Light - residential areas. Dark - non-residential areas.

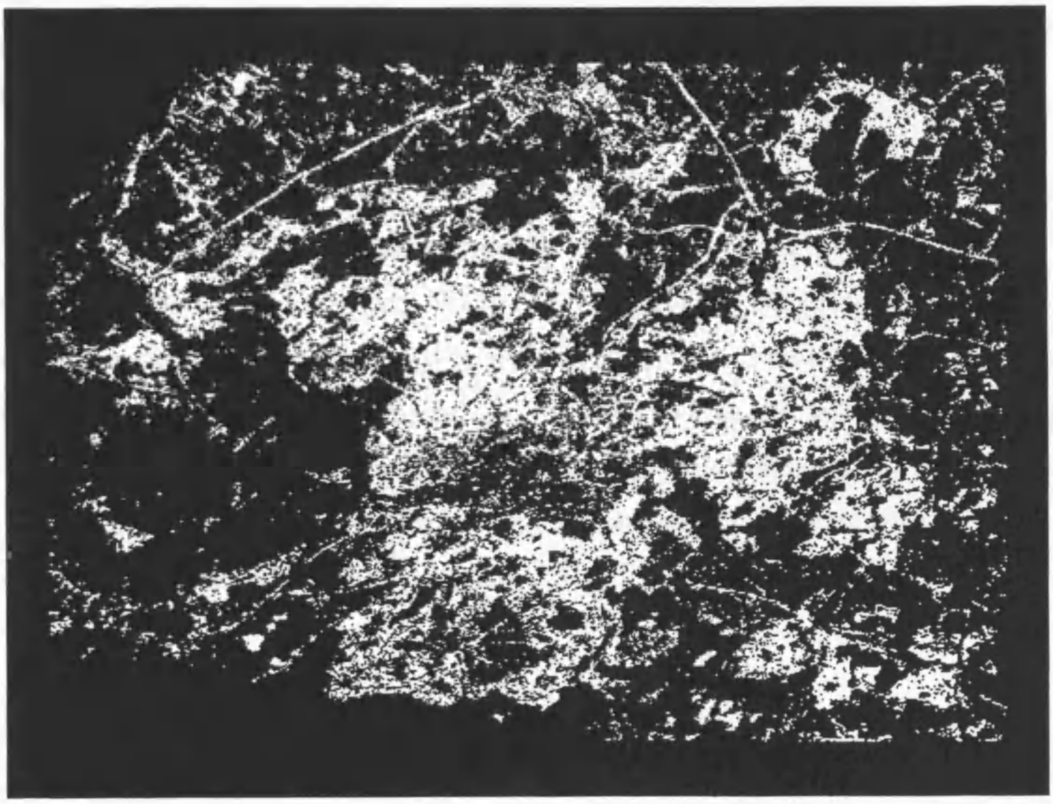


Figure 7. 2: Landsat TM images using Maximum Likelihood Classification (Bristol). Light - residential areas. Dark - non-residential areas.

7.2 Rasterizing ED and beat boundaries, and overlaying with Landsat images

The pre-processing step described in the previous sections would have also introduced errors. The accuracy of the classified Landsat imagery needs to be assessed and calibrated by referring to the more accurate ED maps. To do this, the ED maps (as well as the beat maps, which are the final spatial object of interest for this case study) need to be rasterised and overlayed with the classified Landsat TM images.

The ED and beat boundaries of the two cities were exported from ARC/INFO (export files) and imported into IDRISI as vector files (.vec). The vector files were then rasterized using IDRISI to produce image files (.img). Since the pixel brightness of the rasterized image was represented by numerical values and thus they could be manipulated mathematically (using simple operators such as add, subtract, divide and multiply). The resultant values of the non-residential areas would remain as zero while the residential areas would take on the ID values of ED or beats after the overlay process using the multiply operator:

$$ID \times 0 = 0;$$
$$ID \times 1 = ID.$$

As an example, Figure 7.3 shows a zoom-in section of the resultant overlay Landsat x ED map of Coventry.

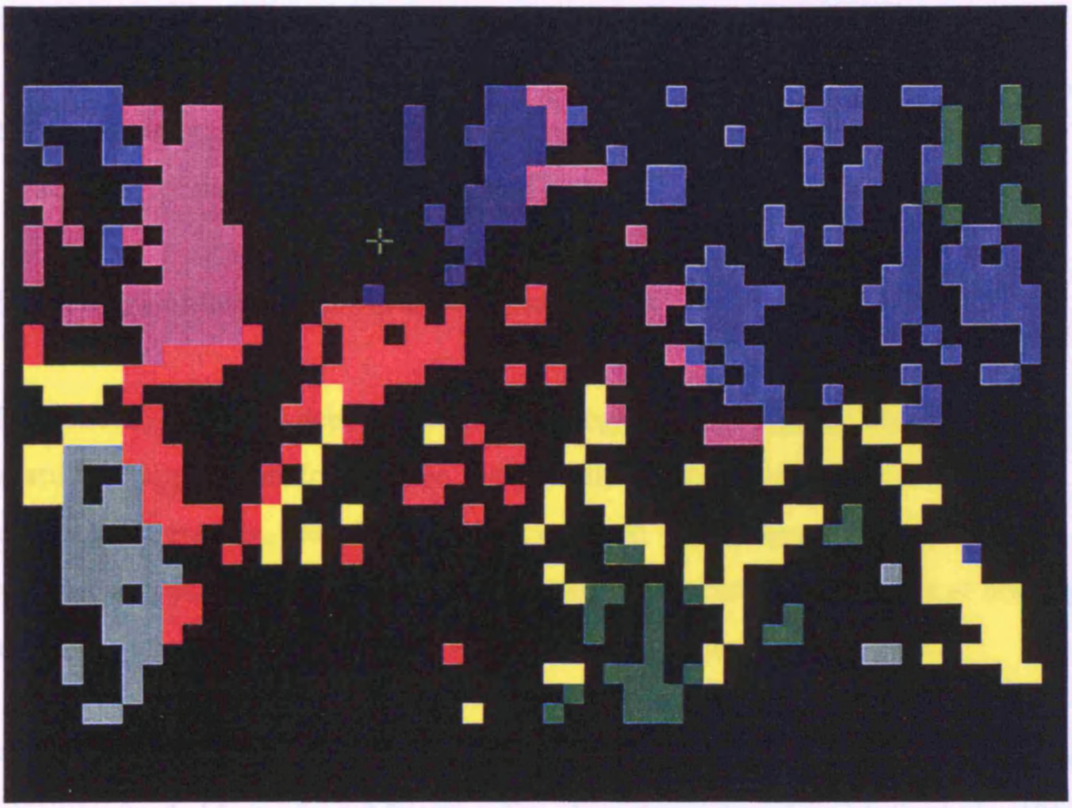


**Figure 7. 3: A section of the resultant overlay map of Coventry (near city centre).
Dark pixel = non-residential area. Colour value = ED-ID**

Note the screen colour may not truly represent that actual value due to the limited colour palette.

The above arrangement made the re-assignment for calibration and Monte Carlo simulation simple using the following conditional rule (see List-Stat implementation next):

If a pixel = 0, ignore

else increment the count by 1 for that ID zone.

7.3 Implementing Monte Carlo simulation using Lisp-stat

Before the Monte Carlo simulation is carried out, one needs to:

1. Select a suitable programming language for implementation.
2. Test the pseudo-random number generator using the selected programming language for the specific hardware used.

7.3.1 Programming and programming languages

Theoretically, there is a debate about which programming language should be used for a particular application, for instance, procedural, declarative, or object oriented approach. In practice, apart from efficiency (which is no longer an important consideration due to the development of fast super-computers), the choice of language is a matter of personal taste, and depends on the programmer's experience.

For this case study, Lisp-Stat is used for implementing Monte Carlo simulation for its robustness and object oriented functionality. List-Stat is based on McCarthy's (1960, 1963) programming language List Processing (Lisp) with improved efficiency and extra functionality to handle statistical data which makes it an ideal candidate for GIS processing (for advantages of using an object oriented approach such as Lisp and Lisp-Stat, see Brunsdon, 1995; 1998; Steele *et al*, 1984; Tierney, 1990; Winston and Horn, 1984). Lisp (or Lisp-Stat) is a pure functional language and consists of the following standard form:

(*function* argument)

In Lisp, the basic element within a list is called an *atom*. An atom may be a number or any symbol. An argument may consist of an atom, a list of atoms, or a list of lists. An empty list () is NIL. Functions can be nested. (See Steele *et al*, 1984; Tierney, 1990; Winston and Horn, 1984 for further technical details.) For example, functions for propagating an index of data error have been incorporated in GEOLINEUS, a lineage information program for GIS (Lanter, 1991), implemented in Lisp and integrated with ARC/INFO.

7.3.2 Generating random numbers using Lisp-Stat

Random number generation is a basic requirement for any Monte Carlo simulation. To generate random numbers with computers, it is necessary for the programmer to write the specialised code modules and insert them directly into the simulation model. The random numbers generated using computers are called pseudorandom numbers as they are not truly random. Psuedorandom numbers are desirable as they usually serve the purpose of estimating the results of an application, and make experiments repeatable so that the reliability of the results can be tested.

Generating pseudorandom numbers using Lisp-Stat functions is simple. A number of distributions are available: uniform; gamma; beta; t; chisq (χ^2); f; binomial; Poisson; and normal. The density, cumulative distribution function, and quartile function of the distribution can also be evaluated (refer to Tierney, 1990). For example, (uniform-rand 100) generates a list of 100 independent random numbers distributed uniformly between 0 and 1. (Random n) accepts a positive integer or floating-point number n and returns a number of the same kind between 0 (inclusive) and n (exclusive) from an approximately uniform distribution. Furthermore the function 'sample' enables the programmer to select a random sample from a list. For example:

```
(sample (iseq 1 100) 10)
```

returns a list of a random sample of size 10 drawn without replacement from the integers 1,2,...100. This is very useful for Monte Carlo simulation and testing of the random number generator as described later.

The random number generators for these distributions are implemented using the Common Lisp random number generator (Steele *et al*, 1984). The Common Lisp random number interface is portable across a variety of systems. The seed of the generator can be saved and restored, and a new seed generated, using the system clock (which is machine dependent). The current value of the random state is held in the global variable *random-state* Within Lisp-Stat environment, simply type *random-state*, the system would return the current value of the random state as a side effect.

```
> *random-state*
```

#$(1 #(2147483562 833502228 1548262346 714760118))

This can be set and saved for the replication of the random numbers. For the purpose of this application, functions 'random', 'sample', and 'uniform-rand' are used to generate uniform random numbers.

7.3.3 Testing pseudo-random number generators using Lisp-Stat

The quality of the random numbers generated is often machine dependent. So it is necessary to test the random number generator before it is used for Monte Carlo simulation. There are many tests developed to determine if a random number sequence {Yn}, (∀n = 1, N) has the desired probabilistic properties (Atkinson, 1980; Dagpunar, 1988; Jennings and Mohan, 1991; Knuth, 1969; MacLaren and Marsaglia, 1965; Sowey, 1972, 1978, 1986; Tauussky and Todd, 1956). For example, Jennings and Mohan (1991) describe twenty different diagnostic tests (in a system called Randalize) to interrogate random number generators. Nevertheless, the tests are not exhaustive, noticeably lattice structure and spectral tests are omitted. It is not necessary to go into details of all the possible tests here. For the purpose of this case study, it is sufficient to determine if the proposed random number generator is appropriate for error modelling according to the following three criteria:

- Visualisation; [See Fisher *et al* (1993), for example, and Knuth's (1969) 3-D view to prevent "mainly in the plane" problem.]
- Period test; and
- Chi-square (χ^2) test.

These are discussed in detail in Appendix 7.2. The random number generator was tested using Lisp-Stat function to generate random numbers on a Digital DECpc Lpx 466d2 microcomputer. If one adopts a criterion such as 1 - p > 0.05 => 'pass' (i.e. the probability that the generator is unbiased > 0.95), then the generator passed the test. Indeed it was found that the value of χ^2 lies between 5% and 95% level (2.17 < 5.92 < 14.07) at d.f.=7 with $p(\chi^2_\alpha|\mu) \approx 0.5$. For the 10,000 random numbers generated, the sequence did not repeat itself. Visually, the numbers appear to be randomly and uniformly distributed.

7.3.4 Calibrating Landsat TM imagery

Since the classified Landsat image contains inaccuracy, we do not know the exact spatial relation between each classified pixel and number of households it represents. The number of households per pixel can be determined (or 'calibrated') by comparing the number of pixels assigned to the residential area with the number of households in the source zone (ED) according to the 1991 Census. This consists of the following steps:

1. Assign pixels to residential / non residential areas within each ED.
2. Define the ED as the source zone from the ED-beat overlayed coverage from ARC/INFO.
3. Get the geographical outcome base (GOB) from the source zone.
4. Compare the results from Step 1 and 3 and work out the ratio of households per pixel for each target zone (beat).

The algorithm of assigning the pixels is as follows (Function *proed_go*):

1. Define ED_ID list (initially nil).
2. Create an empty list with the data structure of a list of value and index lists (Function *create_list0*).
3. Loop.
4. read in a pixel from the image file.
5. Assign the pixel (Function *assignh*).
6. Back to Step 4 till the end of the list.

Assignh applies the following production rule to a list of pixels:

IF CONDITION: p = 0 THEN do nothing
ELSE increment the counter by 1

It means that if a pixel is not a household area (**p** = 0); then do not count, else count the pixel by incrementing the value of the appropriate ED list by 1. This was done by matching the pixel value (which happened to be its ED-ID) and ED-ID in the ED-ID list (Function *match_fill*) and incrementing the counter in that list. The above procedures are listed in PROBED.LSP (Appendix 7. 3).

To run the above program in Lisp-stat, first open the image file, say, Bristol (wbrixed1.img)

```
(setf f (open "wbrixed1.img"))
#<Input-Stream 6:"c:\users\law\gis\ho\wbrixed1.img">
```

then type

```
(proed_go f 825 190920)
```

The function takes the image file (f) a number of EDs (825) and the number of pixels (190920 in the image file) as input arguments, and returns the ED-ID list of the pixel counts as its output. The output ED-ID list was named as **bed_plist** and saved as BEDPLIST.LSP, that is, a list of ED and pixels pairs.

The ED-beat overlayed coverage from ARC/INFO was defined as the source zone.

(def source_zones (read-data-columns "bedbeat.out" 13))

Variables of interest such as ED-ID, Beat-ID, households and etc were retrieved from the geographical outcome base (GOB) of the source zone (Function *get_gob*). Note that the source zone might now have a different number of EDs (compare with Edbeatgs coverage) due to the overlay operation.

The algorithm for getting the attributes for the source zone is listed as follows (Function *pre_beat_go* for Coventry and for *pro_beat_go* Bristol):

1. Initialise all the variable list as nil (Function *init_pre_beat_go*).
2. Define source_zone.
3. Get attributes from the source zone.
4. Get the attribute value for each beat from the many to one relation ship between beats and ED (*get_many_from_one* beatgs-id edgs-id beat-id).
5. Get demographic attributes from GOB (get_gob).
6. The above step is carried out using many to many relationship matching (*get_many_from_many*).

A look up table (LUT) of ID was created for household counts within each beat with a one-to-many relationship from the EDs (*create_lut*). The LUT (**listi = b_hh**) has the following data structure:

$$((ID_1\, h_1)\ (ID_2\, h_2)\ ....\ (ID_n\, h_n))$$

It is a list of lists. Each list within the list has a pair of value: beat-ID and household counts from the census data. A similar LUT can be created for population counts (**listi = bbt_r**).

Residents (r), household counts (hh), household density (r/hh), pixel (p), households per pixel (hh/p), residents per pixel (r/p) for each zone were listed by comparing the values of attributes (such as household counts) and number of pixels for each beat within the source zone (using **bed_plist** and **bbt_r** as data source, Functions *make_pair_list* and *make_list3*)

Function *make_pair_list* creates a beat and ED-ID look-up list for each ED. Function *make_list3* takes lists **bed_plist** and **bbt_r** as its data source, and creates the third list bbt_ped for Bristol (or cbt_ped for Coventry) using a Function called *match_cons*. *Match_cons* constructs a new list by matching the beat- and ED-IDs. Function *make_listv* creates a list of beats and the number of pixels (**bbt_plist**). The desired list can be printed using *my_print* and *print_ratio* functions.

7.3.5 Carry out Monte Carlo simulation

To perform Monte Carlo simulation, first, the classified Landsat TM image was retrieved as a file. The classified Landsat image was defined as the target zone (beat, e.g. bbt for Bristol, and cbt for Coventry).

```
(def bbt (read-data-file "wbbtx01.img"))
```

The image file was then broken up into a number of files (Function *list_batch* in FPRO.LSP) for a stratified random sampling later. Stratified random sampling scheme was selected for this application not only for the technical limitation of Lisp-Stat (maximum 4030 atoms were allowed in a list for random sampling), but also for the theoretical reason such as spatial auto-correlation (see Van Grenderen and Lock, 1977; Rosenfield, 1982; Congalton, 1988; Fukunaga and Hayes, 1989).

Next, run Monte Carlo simulation using Function *Monte_go*. This takes the following parameters as its arguments:

1. A number of simulation runs (n = 300 in this case).
2. A list of files called pn_list (each file contains a list of pixels).
3. A scaling factor (k = 10 in this case).

k represents a constant scaling factor for the sampling size. From a series of test results, a normal distribution was obtained after 300 runs and the optimal value of k was found to be 10 when the RMSE was minimised and yet maintaining a normal distribution of the results (see Appendix 7.4).

So to run the simulation 300 times with scaling factor 10, type:
(monte_go 300 pn_list 10)

The procedure of *Monte_go* consists of the following steps:

1. Iterate: do the following steps for n simulations.
2. Define a list of household count and beat-ID pairs with zero count for each beat.
3. Print message "Please wait... Processing Monte Carlo simulation run".
4. Initialise (*init_fill*).
5. Run Monte Carlo simulation for a list of files (*Monte_file_batch*).
6. Construct a new list to store the results of the simulation.

The list of household count and beat-ID pairs _hid initially was defined as

(def _hID '((0 1) (0 2) (0 3) ..... (0 b)))
where b is the last number of beat-ID

The first atom of each list pair represents the number of household counts. It would be filled with zero values at the beginning of the simulation run (Function *init_fill*).

Monte_file_batch simply runs Monte Carlo simulation using another function *run_monte* for each file within the list. In turn, *run_monte* first initialises the size of random sampling (r) according to the scaling factor k (*init_monte*) and then runs the Monte Carlo function (*monte*) for a list of pixels.

Monte function first uses the standard Lisp-Stat function *sample* to perform random sampling on a list of pixels, and then assigns the pixel to household count using the same *assignh* function as described earlier in calibrating Landsat image (except this time the ID is beat in stead of ED). The experiment can be repeated by initialising the *random-state* in Lisp-Stat.

```
(setq *random-state* #$(1 #(2147483562 833502228 1548262346 714760118)))
```

The output of the simulation is a list of lists called NEW_LIST. Each list with NEW_LIST contains a number of results of the Monte Carlo simulation and the beat-ID as the last atom in the list. So NEW_LIST has the following data structure:

$((h_1\ h_2\ ...., h_{300}\ ID_n)$
$(h_1, h_2\ ...., h_{300}\ ID_{n-1})$
$(h_1\ h_2\ ...., h_{300}\ ID_{n-2})\ ....$
$(h_1\ h_2, ...\ h_{300}\ ID_1))$

The results of the simulation (**new_list**) were saved in the file called bbthid.lsp.

```
> (savevar 'new_list "bbthid")
```

Histograms can be plotted using Lisp-Stat function *histogram* (Figure 7.4):

```
(histogram (car new_list))
#<Object: 228004122, prototype = HISTOGRAM-PROTO>
```

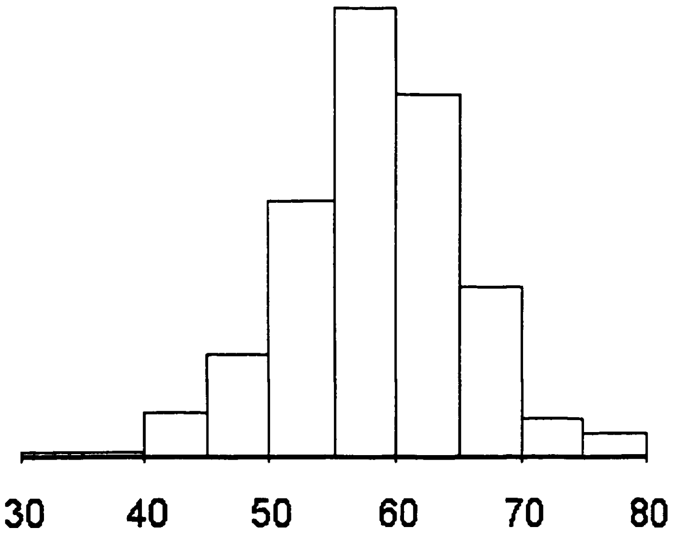


Figure 7. 4: A histogram of 300 runs

Experimentally, it was found that a normal distribution was obtained after 300 runs (Appendix 7.4). This is consistent with the theoretical discussion in Chapter 5. A whole list of histograms can be computed for further analysis (see 040397.Log in Appendix 7.5 and Chapter 8). Appendix 7.3 lists all programs used in the case study and Appendix 7.5 shows the complete log for processing Bristol as an example; with comments after semicolons.

7.3.5 Assessing the accuracy of Monte Carlo simulation

The accuracy of the Monte Carlo simulation itself can be assessed by computing the RMSE using equation (3.2). Here the actual number of pixels (z') was determined by assigning the whole image at one go. The value obtained from the overall assignment for each beat was then compared with those from the simulation (z) (see 121296.log in Appendix 7.5). The results show that the RMSE are almost identical to the SD of the Monte Carlo simulation. The mean RMSE for Bristol is 0% (and for Coventry 1%) indicating that the estimator used is indeed a non-biased estimator. The results can be printed (using *print_tm_mc_results* and POSTBBT.LSP) and read by spreadsheet (such as Excel) for further analysis (next Chapter).

7.4 Chapter Summary

This Chapter has described the processes of implementing the dasymetric method for error modelling using IDRISI for processing Landsat imagery and Lisp-stat for Monte Carlo simulation. As a result, the procedures outlined in the methodology of Chapter 5 can be further expanded and summarised as follows (Table 7.3):

Table 7. 3: Procedures of error modelling

Steps	Procedures
1	Get the satellite imagery
1.1	Pre-process
1.1.1	Resampling
1.2	Extract features
1.3	Classify imagery
1.3.1	Prepare the menu of information classes
1.3.2	Select and define training data
1.3.3	Reiterate Step 1.3.1 and 1.3.2 with modification as necessary to ensure homogeneous training data
1.3.4	Conduct classification
2	Calibrate the satellite image with ED using Monte Carlo simulation
2.1	Rasterize ED boundaries
2.2	Overlay the rasterized ED with the classified Landsat TM imagery
2.3	Estimate the household density per ED
2.4	Calibrate the Landsat imagery
2.4.1	Assign pixels within each ED
2.4.2	Define ED as the source zone from the ED-beat overlayed coverage
2.4.3	Get the geographical outcome base (GOB) from the source zone
2.4.4	Compare the results from Step 1 and 3 and estimate the average number of households per pixel (h/p) for each target zone (beat)
3	Perform Monte Carlo simulation to estimate household density in beats using the calibrated satellite image as a dasymetric map
3.1	Rasterize beat maps
3.2	Overlay rasterized beat map with the Landsat imagery
3.3	Perform Monte Carlo simulation to estimate the household count for each beat using h/p obtained from Step 2
3.3.1	Iterate: do the following steps for n runs
3.3.2	Initialise necessary variables
3.3.3	Run Monte Carlo simulation
3.3.4	Print and store the results of the simulation

The results generated from the above procedures are the subject of further analyses in the next few chapters.

Chapter Eight
Analyses of Results - processing error (I): the impact on household population estimates

This chapter provides a statistical summary of the results from the Monte Carlo simulation (Tables 8.1 and 8.3). The results are analysed quantitatively from their numerical values and qualitatively from their distribution pattern. The object of the analyses is to examine the interrelationship between the geographical factors (such as size and location of an area), spatial statistics (such as household counts in this case) and their error distribution as a result of the spatial interpolation. The results can also be analysed visually by plotting the statistical values onto a series of maps in terms of their spatial distribution. Since each city is different, the results are analysed city by city - first Coventry (Section 8.1) and then Bristol (Section 8.2). Bristol is more complicated not only because it consists of more beats than Coventry, but also geographically it consists of a lot of open spaces such as parks. Furthermore, as discussed in Chapter 3, Bristol consists of super-beats, though this does not affect the analyses at this stage.

8.1 Coventry

The results of the simulation for Coventry are summarised in Table 8.1. It shows the household counts (estimated by means of the overlay method and the Monte Carlo dasymetric method), the standard deviation (SD) of the Monte Carlo Dasymetric method, the error rate (in percentage) of the overlay method as compared with the Monte Carlo dasymetric method and the SD of the percentage error of the Monte Carlo simulation.

Table 8. 1: Household Count and error rate in Coventry

Beat	Overlay	MC (h)	SD	Error %	SD-err %
2	2144	3738	213	-43	0
3	4844	4606	305	5	2
4	3053	2683	193	14	0
5	1933	1769	144	9	0
6	4409	3085	184	43	1
7	1794	2282	188	-21	0

8	5194	4097	264	27	2
9	4320	3837	248	13	0
10	2950	2832	216	4	1
11	3042	2730	203	11	1
14	2505	2224	284	13	3
15	4543	4512	309	1	1
16	1753	1905	197	-8	2
17	2513	2201	282	14	1
18	4241	4065	281	4	1
19	1029	841	179	22	1
20	5819	5578	344	4	0
21	2451	1253	186	96	1
22	2354	1386	154	70	3
23	1707	912	150	87	1
24	829	950	184	-13	1
25	911	1069	158	-15	4
26	2283	3463	238	-34	1
27	2706	2822	205	-4	1
28	3245	2546	156	27	0
29	143	257	84	-44	13
30	402	302	121	33	0
31	1967	1218	269	61	6
32	1229	1421	340	-13	2
33	349	457	130	-24	3
34	1742	1672	278	4	1
35	2020	1427	235	42	1
36	2397	2638	237	-9	0
37	1299	1530	190	-15	1
38	1147	1399	180	-18	0
39	1812	1649	191	10	2
40	2111	1992	157	6	1
41	4548	3618	172	26	0
42	3129	2685	217	17	0
43	1426	1327	114	8	1
44	2923	2569	217	14	1
45	1867	2191	198	-15	1
46	2989	2896	254	3	2
47	3828	3486	217	10	0
100	3396	3949	304	-14	1
101	3297	3016	265	9	2
Total	**116593**	**109085**	**9835**		
Mean	**2535**	**2371**	**214**	**6.88**	**1**

Out of a total of 46 beats, the total number of households estimated by the overlay method was 116593. Using the Monte Carlo dasymetric method, the estimated number of households was 109085 ± 9835. This represents error of 7% ± 1% as a whole. The error ranged from a minimum -1% error (excluding 0 counts) to the maximum error of 96%!

The spatial pattern can be visualised on a series of thematic maps. Figure 8.1 shows the household distribution per beat using the overlay method described in Chapter 6. The city centre can be readily identified by the least populated area (Beat 29, 30, and 33). Naturally, the larger the area, the greater the number of households (Beat 3, and 20) though this is not always the case (Beat 2, 4, 5). Some of the beats consist of a large number of households, and yet their areas are relatively small (Beat 6, 8, 31, 35, 41, 101). These represent the areas of dense household counts - residential areas. The exceptions are near the city centre (Beat 31).

The Monte Carlo dasymetric estimation produces a different spatial pattern (Figure 8.2). The largest areas and the least populated areas such as the city centre can be identified and they are the same areas as identified by the overlay method. There are small changes in the spatial pattern. These changes appear to occur at the medium ranges. The Monte Carlo method seems to have a 'smoothing' effect such that the spatial pattern of the beats become more coherent with their neighbourhood beats (for example, group: Beat 26, 2, 47, 41, group Beat 32, 34, 33, 37, 38, 39, 43).

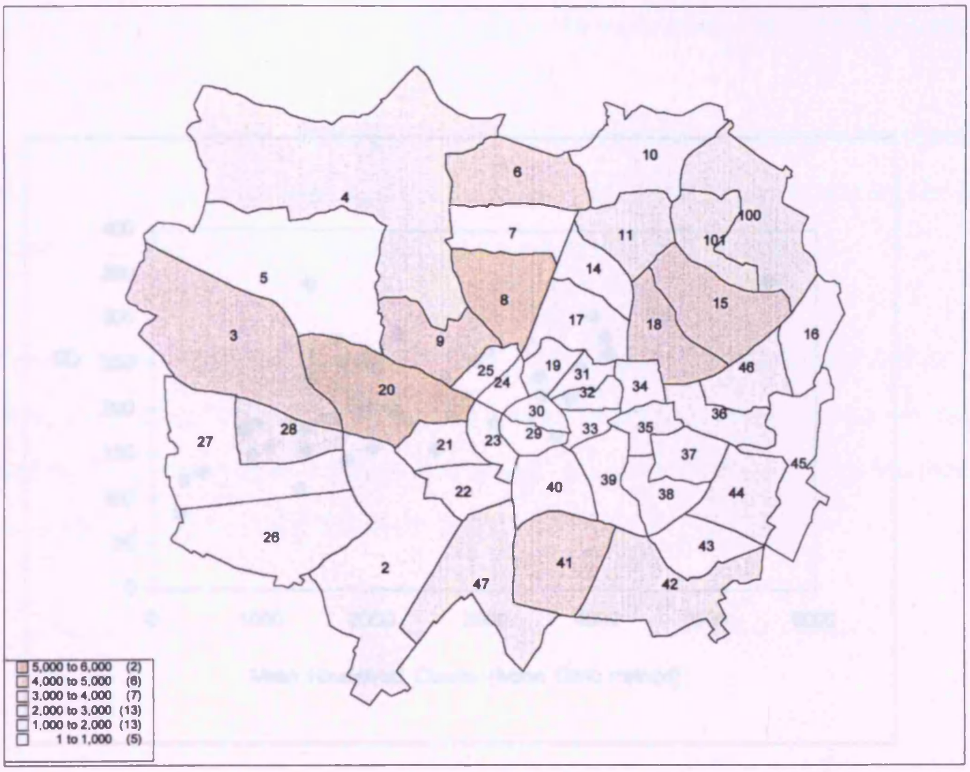


Figure 8. 1: Number of households per beat using overlay method (Coventry)

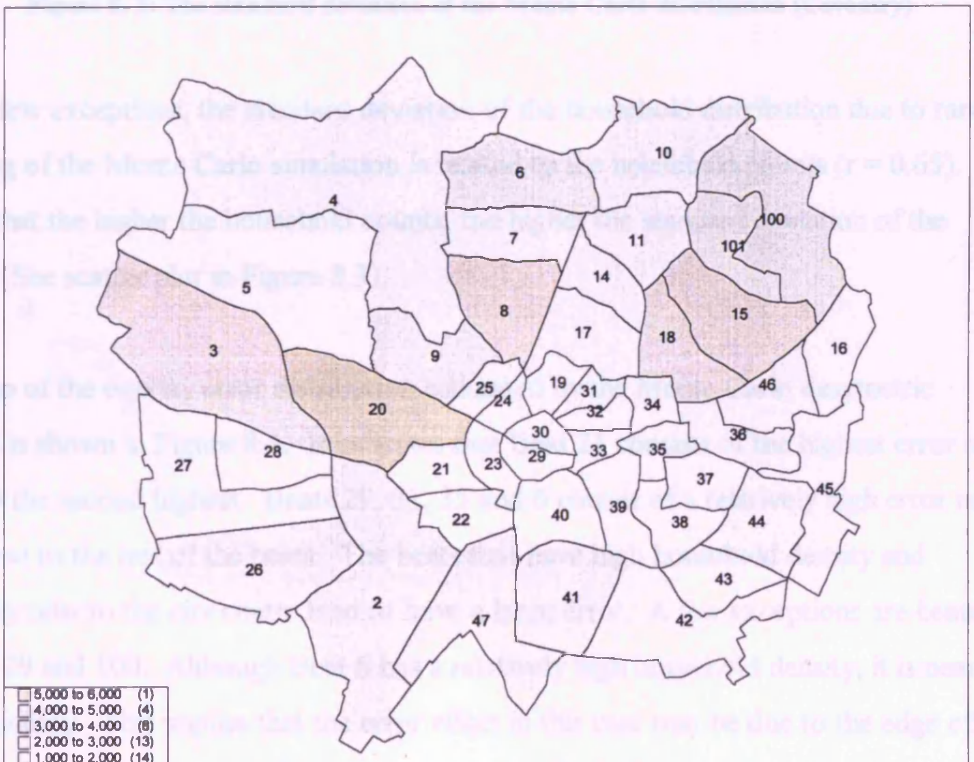


Figure 8. 2: Number of households per beat using Monte Carlo dasymetric method (Coventry)

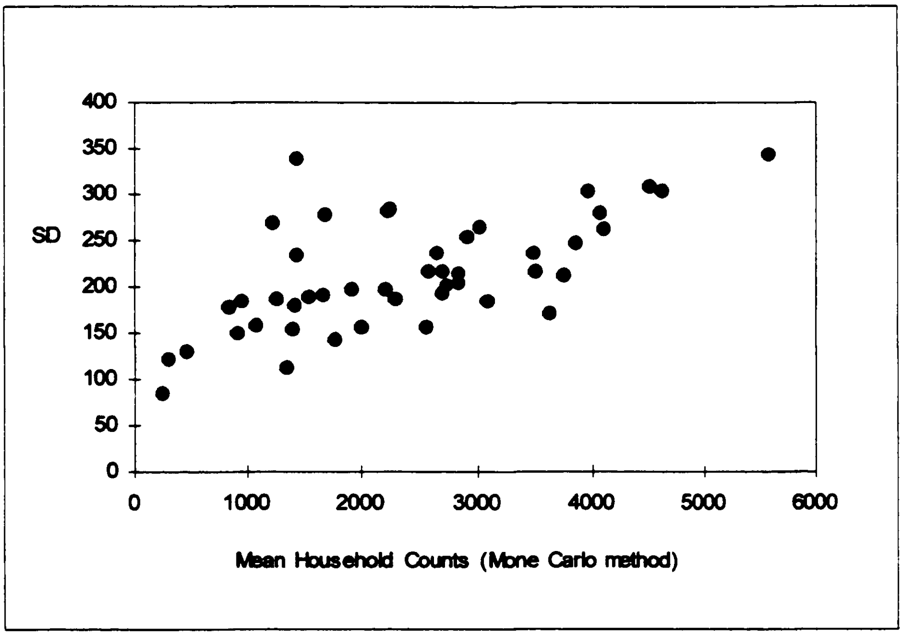


Figure 8. 3: The standard deviation of the Monte Carlo distribution (Coventry)

With a few exceptions, the standard deviation of the household distribution due to random sampling of the Monte Carlo simulation is related to the household counts (r = 0.65). This means that the higher the household counts, the higher the standard deviation of the sample (See scatter plot in Figure 8.3).

The map of the overlay error distribution estimated by the Monte Carlo dasymetric method is shown in Figure 8.4. It indicates that Beat 21 consists of the highest error and Beat 23 the second highest. Beats 29, 31, 35 and 6 consist of a relatively high error rate compared to the rest of the beats. The beats that have high household density and relatively near to the city centre tend to have a large error. A few exceptions are beats 6, 14, 17, 29 and 100. Although Beat 6 has a relatively high household density, it is near the city boundary. This implies that the error effect in this case may be due to the edge effect of the boundary. This edge effect does not necessarily apply to all boundary beats. It depends on whether the boundary shift includes or excludes a high household density area.

For instance, if the boundary shift only happens to the park areas, the change in household counts would be very small (and hence small error).

Beat 29 is a city centre beat. The high Monte Carlo error rate occurring in the sparely populated area would have caused the exceptionally high sampling error during the Monte Carlo runs. This also explains why it has an exceptionally low variance.

For comparison the error rate estimated by the area weighted method (described in Chapter 6) is also mapped (Figure 8.5). The pattern bears no resemblance to the map generated by the Monte Carlo dasymetric method. In general the area weighted method seems to overestimate the error for most beats.

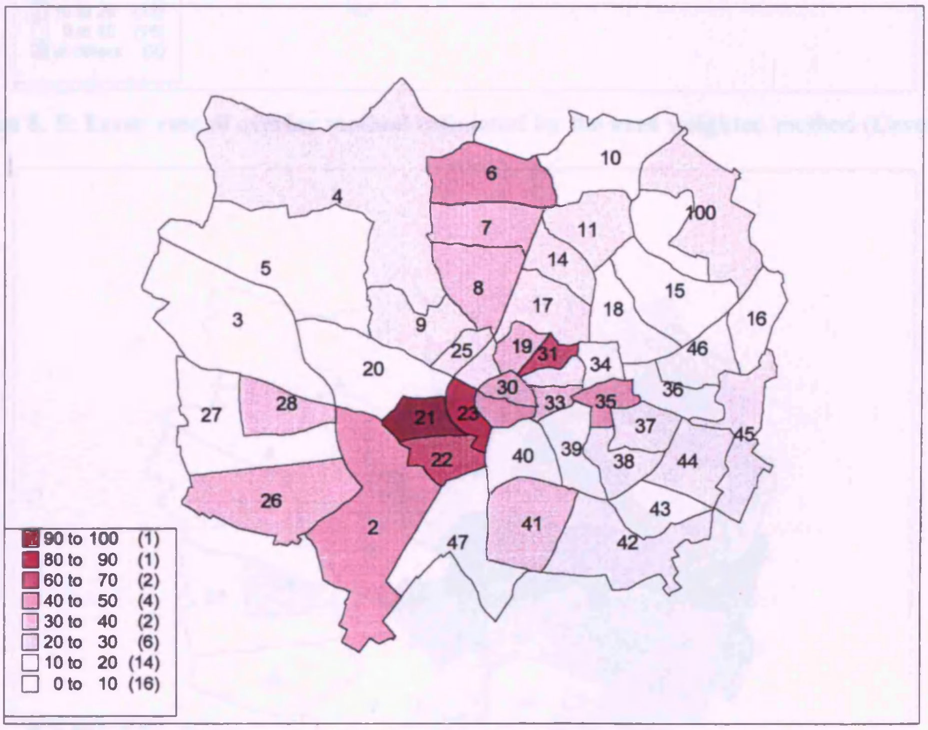


Figure 8. 4: Error rate of overlay method estimated by the Monte Carlo simulation

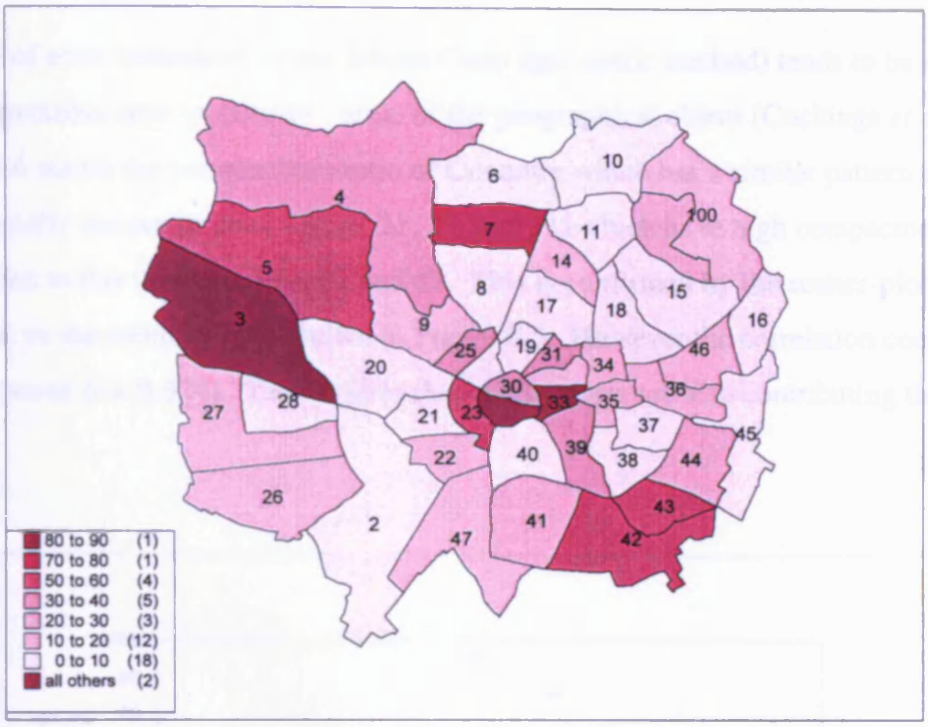


Figure 8. 5: Error rate of overlay method estimated by the area weighted method (Coventry)

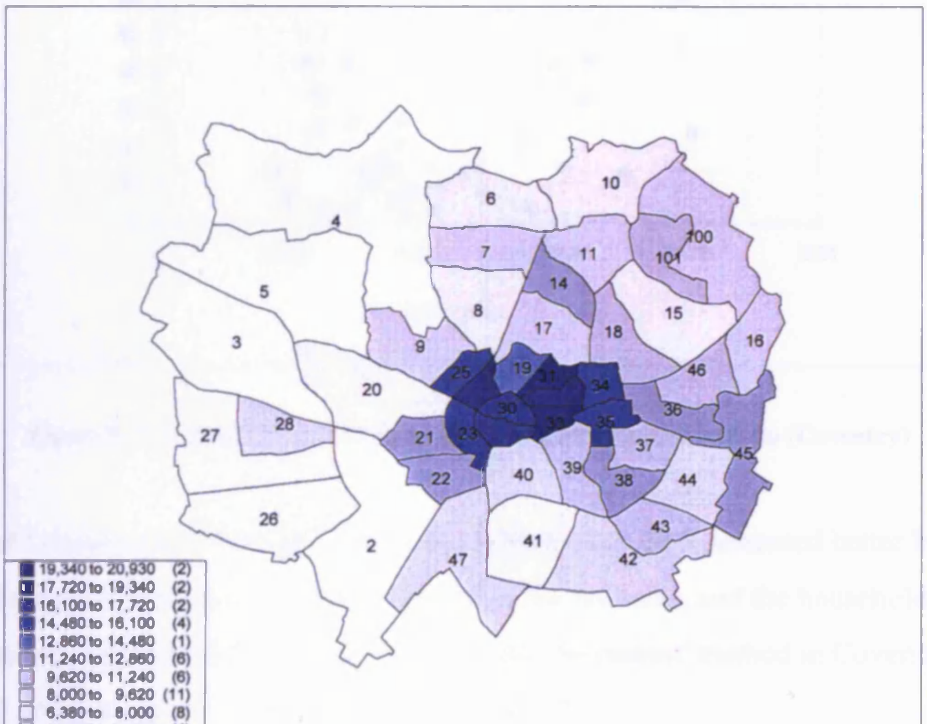


Figure 8. 6: Compactness ratio (Coventry)

The size of error (estimated by the Monte Carlo dasymetric method) tends to be related to the compactness ratio (perimeter : area) of the geographical object (Cockings *et al*, 1997). Figure 8.6 shows the compactness ratio of Coventry which has a similar pattern to Figure 8.4 especially the centre beats (Beats 21, 23, and 31) which have high compactness ratios. Exceptions to this trend are Beats 2 and 47. This is confirmed by the scatter-plot of the error rate vs the compact ratio shown in Figure 8.7. However the correlation coefficient is rather weak (r = 0.324). This implies that other factors are also contributing the error rate.

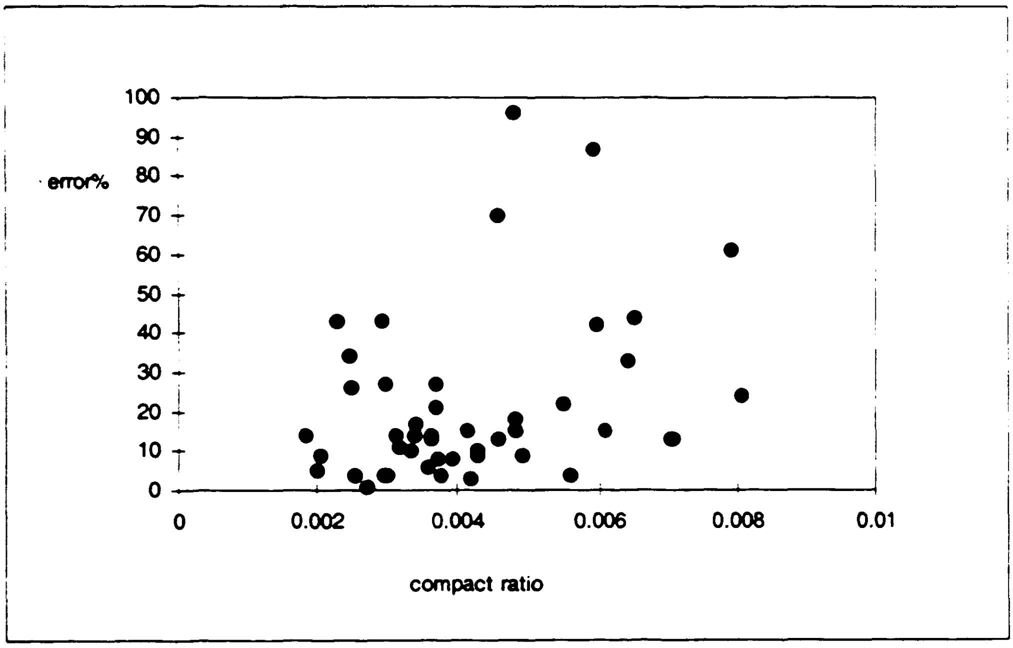


Figure 8. 7: scatter-plot of the error rates vs the compactness ratio (Coventry)

Household density and error rate - household density can be represented better by household counts per unit area. Table 8.2 shows the area size, and the household density per unit area using Arc/INFO and the Monte Carlo dasymetric method in Coventry. This produces different spatial patterns (Figures 8.8 & 8.9).

Table 8. 2: the household density per unit area in Coventry

Beat-ID	Area sq. Km	OvHH/Area	MCHH/Area
2	6.05	354.43	618
3	5.76	840.34	799
4	10.94	279.14	245
5	6.18	312.62	286
6	2.34	1885.93	1319
7	1.71	1049.39	1335
8	2.07	2510.02	1980
9	1.88	2303.66	2046
10	2.98	991	951
11	1.72	1768.42	1587
14	1.09	2294.67	2037
15	2.65	1716.69	1705
16	1.43	1223.98	1330
17	1.71	1471.28	1288
18	1.7	2487.6	2384
19	0.62	1652.53	1350
20	3.22	1809.21	1734
21	0.92	2676.06	1368
22	1.04	2254.42	1327
23	0.67	2536.01	1355
24	0.43	1948.31	2232
25	0.54	1680.97	1973
26	3.63	628.35	953
27	3.67	737.2	769
28	1.45	2241.37	1759
29	0.49	293.12	527
30	0.46	867.96	651
31	0.43	4551.8	2819
32	0.46	2682.06	3100
33	0.42	830.67	1089
34	0.61	2849.59	2734
35	0.68	2953.04	2086
36	1.19	2007.99	2210
37	1.03	1256.2	1480
38	1.1	1044.16	1274
39	1.29	1404.91	1279
40	1.46	1448.15	1367
41	2.43	1875.05	1491
42	3.23	969.1	832
43	1.27	1122.32	1044
44	1.75	1672.49	1470
45	2.24	832.63	977
46	1.3	2297.94	2226
47	3.19	1199.33	1092
100	3.55	956.5	1112
101	1.51	2179.87	1994
Total	**96.49**	**74948.48**	**67584**
Mean	**2.0976**	**1629.315**	**1469.217**

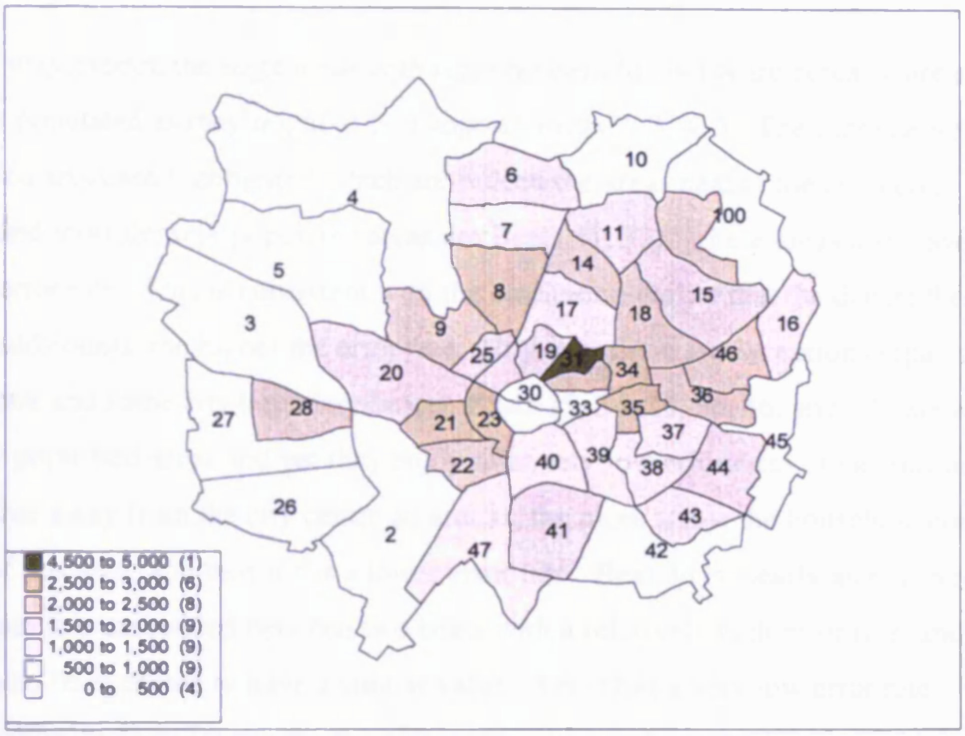


Figure 8. 8: Household density (counts per sq. Km) using the overlay method (Coventry)

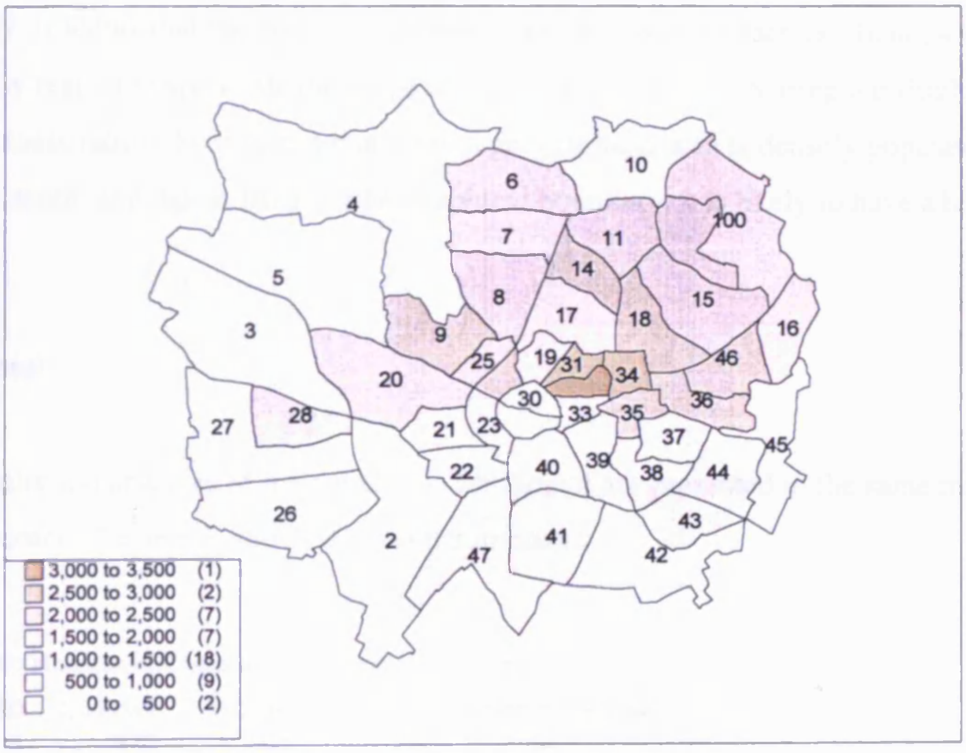


Figure 8. 9: Household density (counts per sq. Km) using the Monte Carlo dasymetric method (Coventry)

As one may expect, the large areas with large household counts are actually not as densely populated as they might at first appear (Beats 2, 3, 4, 5). The more densely populated areas are highlighted which are indeed the areas near to the city centre (Beats 21, 23 and most densely populated areas are Beats 31, 32). These areas also have a higher error rate. This is consistent with the discussion earlier that the denser the household counts, the higher the error rate. However there are exceptions (apart from the city centre and some city boundary beats), Beats 15, 34, 18, 36, 46, and 101 are all densely populated areas and yet they enjoy relatively low error rates. One may argue that the further away from the city centre an area is, the more stable the household counts count of that area, and thus it has a lower error rate. Beat 34 is clearly an exception to that trend. It is embedded between two beats with a relatively high error rate, and so might also be expected to have a similar value. Yet it has a very low error rate. Geographically, they are near to the city centre with relatively high household density. So why is it that the error rate of Beat 34 is more than 10 times lower than that of beat 31 and 35!? It seems that the geometrical shape may be one of the factors. Beat 34 has a relatively regular shape while the shape of Beats 31 and 35 is very irregular (high compactness ratio). So it seems that when a geographical area is densely populated, near the city centre and has an irregular geographical boundary, it is likely to have a high error rate.

8.2 Bristol

The results and analyses of the simulation for Bristol are presented in the same manner as for Coventry. Table 8.3. shows the result summary.

Table 8. 3: Household Count and error rate in Bristol

Beat-ID	Overlay	MC (h)	SD	error %	SD-err %
1	870	563	125	54	0
2	5,830	5,363	148	9	0
3	635	462	46	37	1
4	2,077	2,744	108	-24	0
5	2,288	2,350	123	-3	0
6	2,654	2,333	120	14	1

7	3,680	2,657	126	39	0
8	3,508	4,000	177	-12	-1
9	1,910	2,030	98	-6	0
10	3,297	2,716	100	21	0
11	2,108	2,156	110	-2	-1
12	2,866	2,777	92	3	0
13	2,674	3,963	139	-33	0
14	2,173	2,250	89	-3	1
15	1,772	1,812	88	-2	-1
16	1,438	1,521	70	-5	0
17	1,078	1,159	71	-7	-1
18	1,290	1,145	60	13	1
19	1,439	1,732	71	-17	-2
20	1,538	1,517	52	1	1
21	1,877	1,820	65	3	-1
22	2,519	2,124	88	19	-1
23	1,684	1,622	67	4	1
24	851	1,118	63	-24	-1
25	2,173	2,008	80	8	-1
26	1,478	1,463	71	1	-2
27	2,207	2,169	72	2	2
28	0	0	31	0	-1
29	271	463	40	-41	-2
30	2,352	1,233	61	91	0
31	727	996	49	-27	1
32	0	0	13	0	10
33	164	49	21	233	5
34	0	0	12	0	-5
35	0	0	12	0	-7
36	1,545	1,236	55	25	0
37	923	1,328	34	-31	-1
38	3,014	3,230	75	-7	0
39	1,881	2,933	77	-36	0
40	4,541	3,711	121	22	0
41	6,295	6,655	142	-5	0
42	3,594	2,838	94	27	0
43	10,183	10,568	202	-4	0
44	154	190	48	-19	2
45	3,782	2,646	67	43	-1
46	4,269	5,029	136	-15	-1
47	4,447	4,563	125	-3	0
48	3,980	4,267	142	-7	0
49	2,890	2,441	110	18	1
50	3,215	2,889	139	11	0
51	2,615	3,214	120	-19	0
52	3,813	3,934	92	-3	0
53	1,779	1,880	54	-5	0
54	1,997	1,882	84	6	0
55	1,865	1,994	102	-6	1
56	2,616	2,623	101	0	0
57	2,794	3,288	128	-15	0

	58	4,014	4,669	129	-14	0
	59	261	0	0	0	0
	60	3,683	3,817	96	-4	-1
	61	4,216	4,754	94	-11	-1
	62	6,269	10,600	133	-41	0
	63	3,359	2,357	104	43	-1
Total		**155,422**	**159,851**	**5,562**		
Mean		**2,467**	**2,537**	**88**	**-2.77**	**0**

Out of the total of 63 beats, the total number of households estimated by the overlay method was **155,422**. Using the Monte Carlo dasymetric method, the alternative estimate was **159,851** ± **5,562**. As a whole, this represents the error of -3%. The pattern of the error variation is similar to Coventry, but with a wider range: from a minimum 0% error (excluding those areas with 0 counts) to the maximum error of 233% (Beat 33)! This may be due to the larger household counts in Bristol than Coventry. Notice the extremely large error rate at the city centre beat (33). This does not really matter very much as the actual number of households is very small (164 vs 49).

Similar to Coventry, the spatial pattern of the household and the error distribution can be analysed by examining a series of thematic maps. From Figures 8.10 and 8.11 one can see that Beats 28, 29, 32, and 33 are in the city centre as they consist of relatively few households as well as being at the 'central' part of the city. In contrast to Coventry, the larger areas do not necessarily have the greater number of households (except Beat 1). This is because Bristol consists of a larger number of park areas, and beats in these areas would have relatively fewer households.

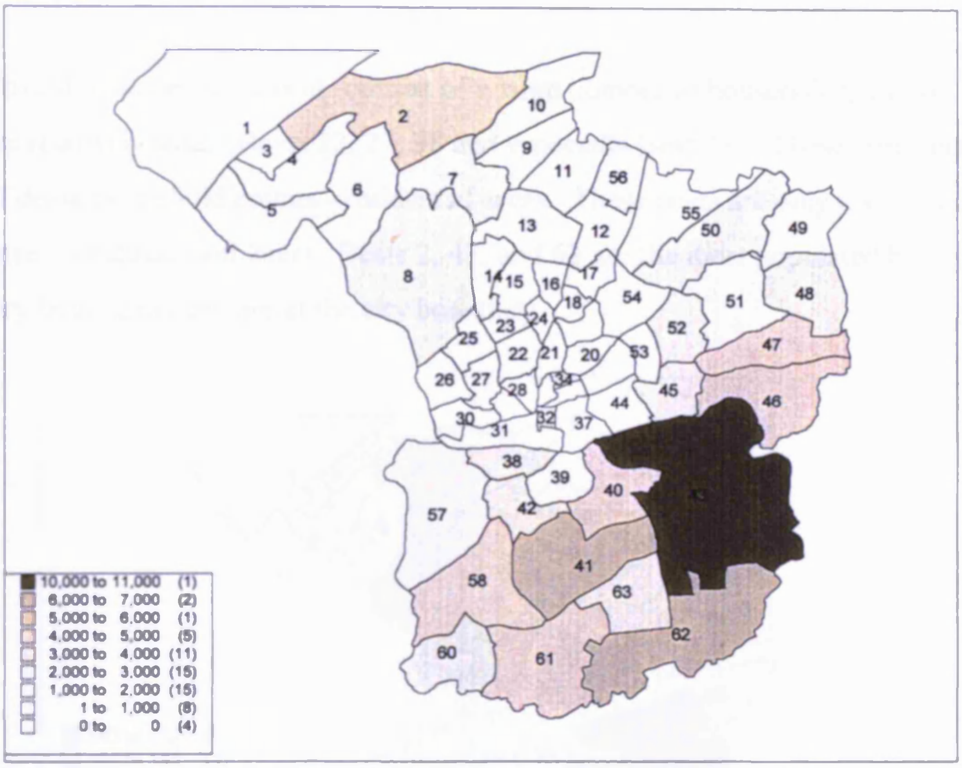


Figure 8. 10: Number of households per beat using overlay method (Bristol)

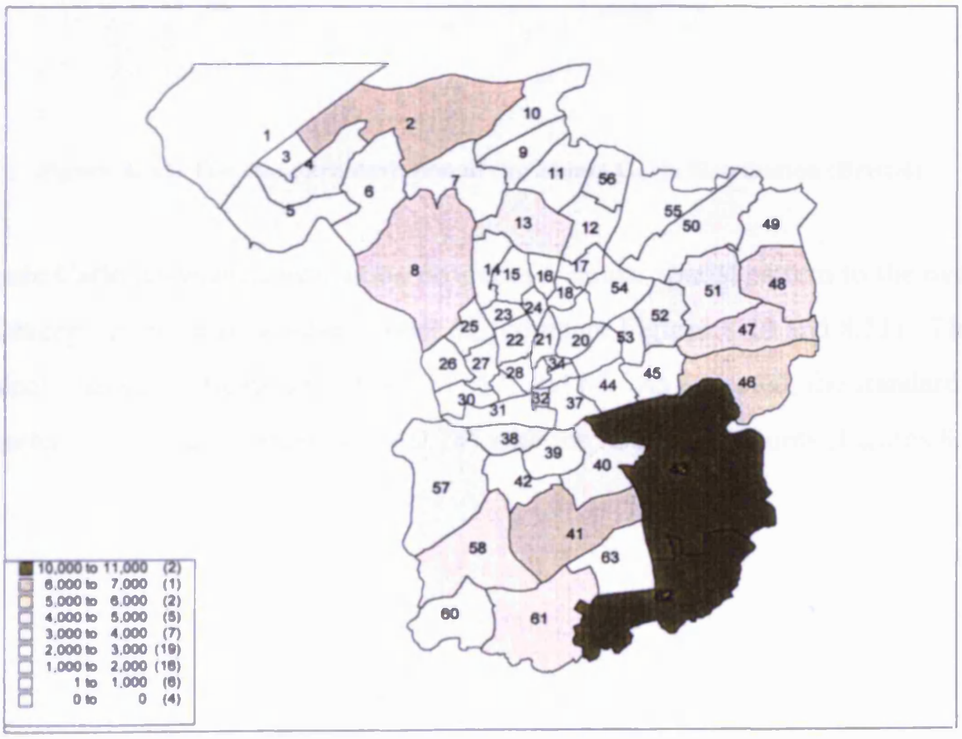


Figure 8. 11: Number of households per beat using Monte Carlo method (Bristol)

As in Coventry, some of the beats consist of a large number of households, and yet their areas are relatively small (Beats 22, 27, 38 and especially Beat 23). These represent the areas of dense household counts - residential areas. These beats are only just outside the city centre (see discussion later). Beats 2, 43, and 62 are the most populated but they all have very large areas and are at the city boundary.

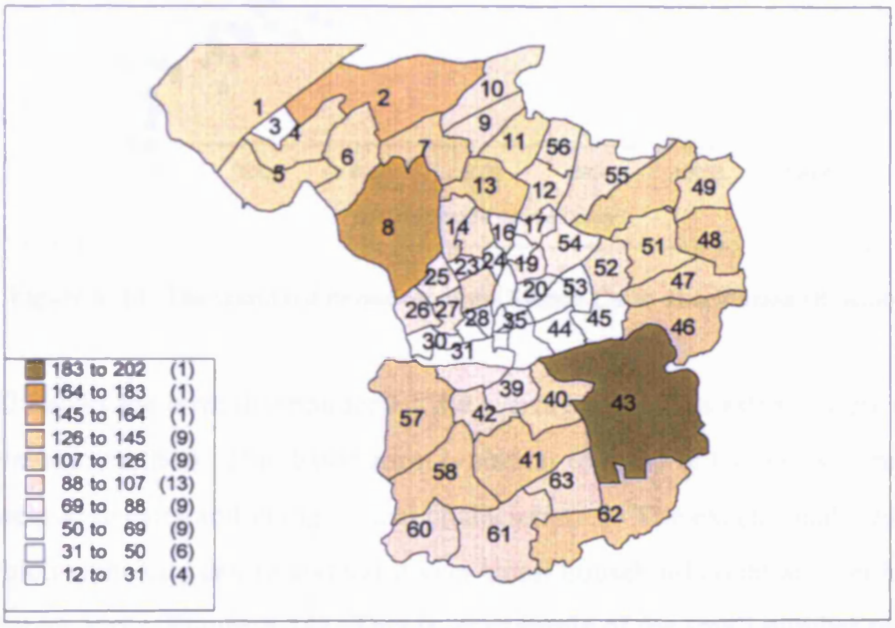


Figure 8. 12: The standard deviation of the Monte Carlo distribution (Bristol)

The Monte Carlo dasymetric estimation produces a similar spatial pattern to the overlay method except in the extreme case in Beat 62 (compare Figures 8.10 and 8.11). There are more minor changes in Beats 38, 39, 42, 51, 57, and 63. As expected, the standard deviation has a positive correlation (r = 0.78) with the household counts (Figures 8.12 and 8.13).

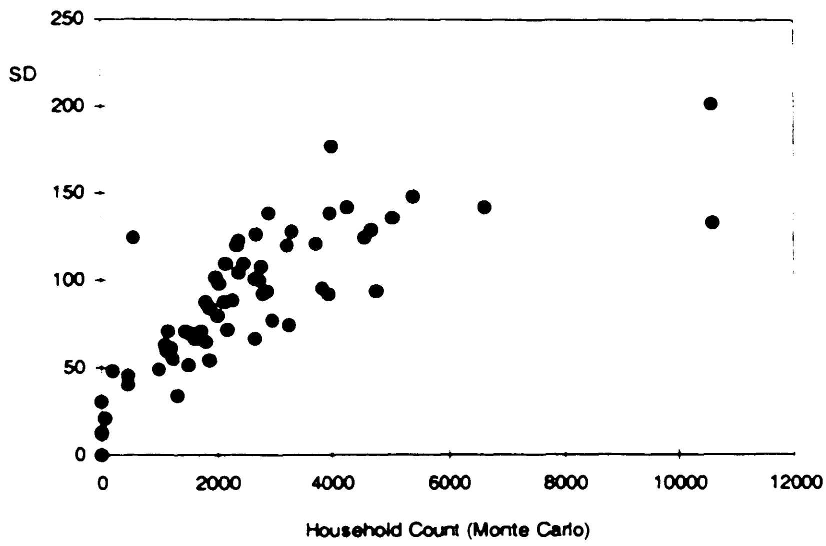


Figure 8. 13: The standard deviation of the Monte Carlo distribution (Bristol)

Figure 8.13 shows the error distribution of the overlay method as estimated by the Monte Carlo dasymetric method. The distribution is similar to those of Coventry in relationship to the household density and geographical characteristics. The exceptionally high error is Beat 33 which is the city centre and has a very small household count and yet has a very high percentage error (Figure 8.14). This is an example of the small number problem as described by Kennedy (1989) where there is a large percentage representation even though the actual value is very small. Beat 30 has the second highest error. It consists of moderate population density and compactness ratio. The only unique geographical feature of this beat is that it is both near the city centre and yet at the city boundary. Some beats are also near or at the city centre, but have no household at all and thus have no error. The unexpected relatively high error of Beat 1 is due to the artificial cut-off boundary of the paper map. Again, for the purpose of comparison, the errors of the overlay method estimated by the area weighted method is shown in Figure 8.15. Except beat 43, the map patterns in Figure 8.15 shows some similarity to Figure 8.14. In general the errors estimated by the area weighted method are higher than those estimated by the Monte Carlo dasymetric method as discussed before.

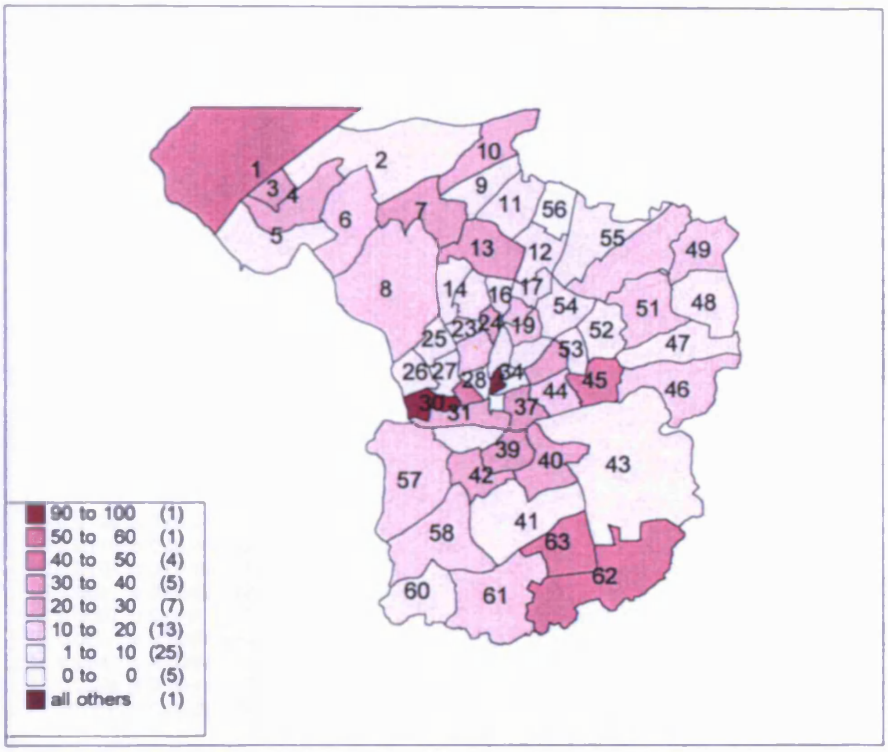


Figure 8. 14: Error rate of overlay method estimated by the Monte Carlo simulation (Bristol)

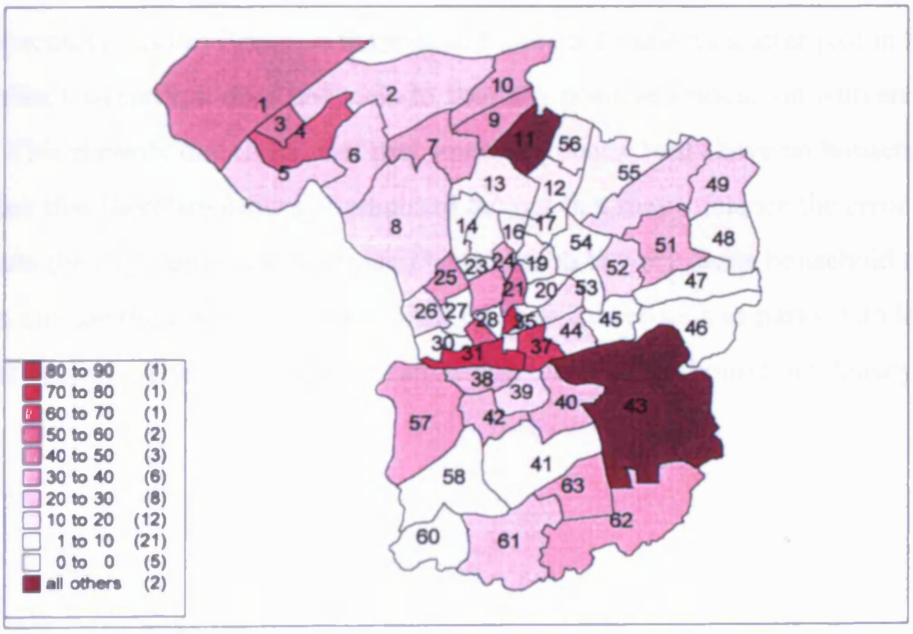


Figure 8. 15: Error rate of overlay method estimated by area weighted method (Bristol)

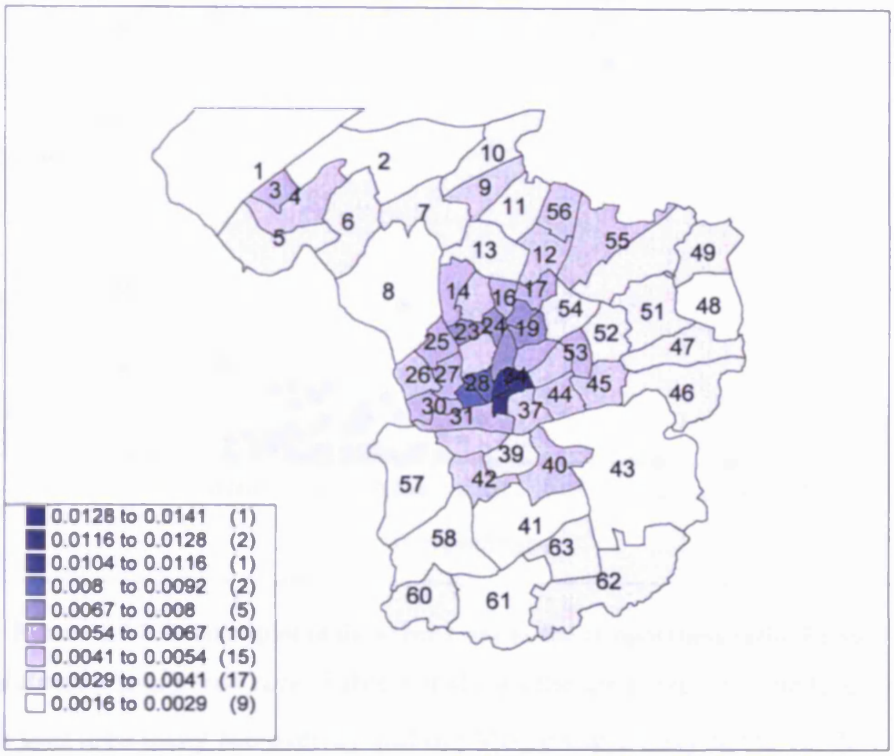


Figure 8. 16: Compact ratio (Bristol)

The compactness ratio of Bristol is mapped in Figure 8.16 and its scatter plot in Figure 8.17. Unlike Coventry, it does not seem to bear any positive association with error rate (r = 0.18). This is partly due to the fact that some city centre beats have no households. It also implies that there are other geographical factors that may influence the error rate. Just outside the city centre, the beats are characterised by very dense household counts. The areas outside these areas consist of a lot of open spaces such as parks with less dense household counts. The effect can be examined in terms of the household density (next).

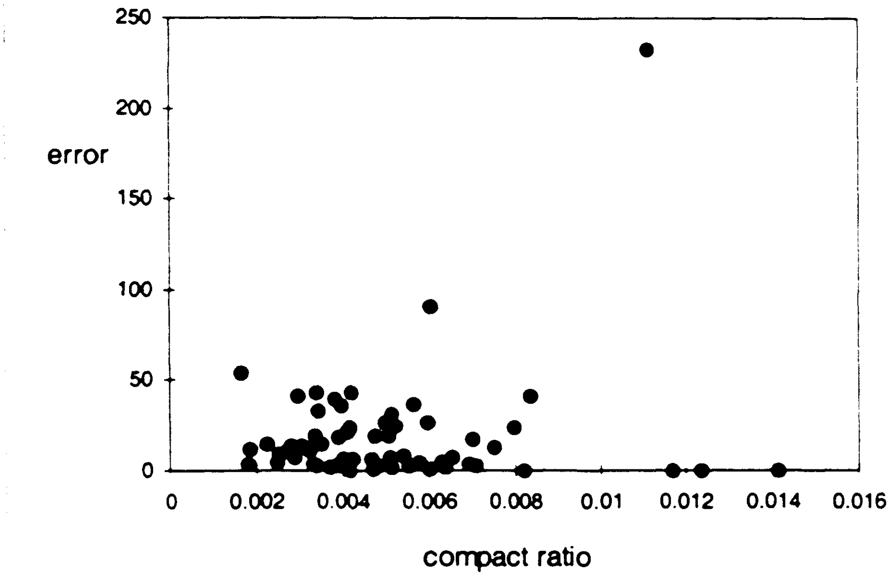


Figure 8. 17: Scatter-plot of the error rates vs the compactness ratio (Bristol)

Household density and error rate. Table 8.4 shows the area size, and the household density per unit area using the overlay and the Monte Carlo dasymetric methods in Bristol. This produces a different spatial pattern (compare Figures 8.18 and 8.19 with 8.10 and 8.11). The effect is similar to Coventry. The beats with high household density have shifted to the areas around and just outside the city centre rather than at the edge of the city (Beat 18, 21, 22, 23, 25, 30, 38, 42, 45, 52, 53). The effect of the near-city-centre area with high household density and irregular geographical boundary can be observed as clearly as in Coventry.

Table 8. 4: The household density per unit area in Bristol

Beat-ID	Area sq. Km	AIHH/Area	MCHH/Area
1	8.09	107.57	69.66
2	5.31	1098.3	1010.3
3	0.61	1048.14	762.88
4	1.88	1106.05	1461.01
5	2.44	936.27	961.78
6	2.49	1067.72	938.75
7	2.04	1807.18	1304.66
8	5.73	612.37	698.29
9	1.23	1553.69	1651.68
10	1.85	1784.57	1470.32
11	1.53	1379.04	1410.42
12	1.17	2452.65	2376.86
13	1.87	1433.38	2124.53
14	0.88	2480.81	2568.69

	15	0.77	2306.04	2357.87
	16	0.46	3153.06	3334.83
	17	0.55	1965.03	2113.03
	18	0.34	3777.79	3352.22
	19	0.46	3115.23	3748.82
	20	0.5	3072.34	3030.05
	21	0.46	4108.22	3984.36
	22	0.68	3689.26	3110.52
	23	0.36	4628.27	4456.66
	24	0.32	2664.34	3499.2
	25	0.69	3167.62	2927.33
	26	0.79	1874.36	1854.87
	27	0.59	3718.61	3654.81
	28	0.34	0	0
	29	0.28	954.45	1630.2
	30	0.64	3659.67	1919.09
	31	1.06	688.9	943.87
	32	0.15	0	0
	33	0.16	1015.51	305.15
	34	0.12	0	0
	35	0.22	0	0
	36	0.64	2398.15	1918.75
	37	0.77	1203.05	1731.23
	38	0.84	3581.94	3838.65
	39	1	1884.35	2938.58
	40	1.71	2654.28	2169.19
	41	3.5	1800.45	1903.46
	42	1.11	3241.08	2559.48
	43	9.23	1103.62	1145.33
	44	0.88	174.67	215.69
	45	1.12	3386.41	2369.04
	46	2.67	1596.84	1881.11
	47	1.93	2300.43	2360.66
	48	2.01	1979.42	2122.02
	49	1.72	1679.17	1418.27
	50	2.91	1105.8	993.53
	51	2.12	1234.21	1516.75
	52	1.25	3051.2	3147.73
	53	0.57	3111.86	3288.83
	54	1.22	1635	1541.04
	55	2.35	792.85	847.67
	56	1.03	2538.23	2544.82
	57	3.95	707.16	832.29
	58	3.26	1230.43	1431.33
	59	0	0	0
	60	1.91	1932.99	2003.57
	61	3.63	1162.6	1311.07
	62	4.78	1311.61	2217.67
	63	1.76	1903.58	1335.48
Total		**106.9**	**117127.82**	**116615.92**
Mean		**1.7**	**1859.17**	**1851.05**

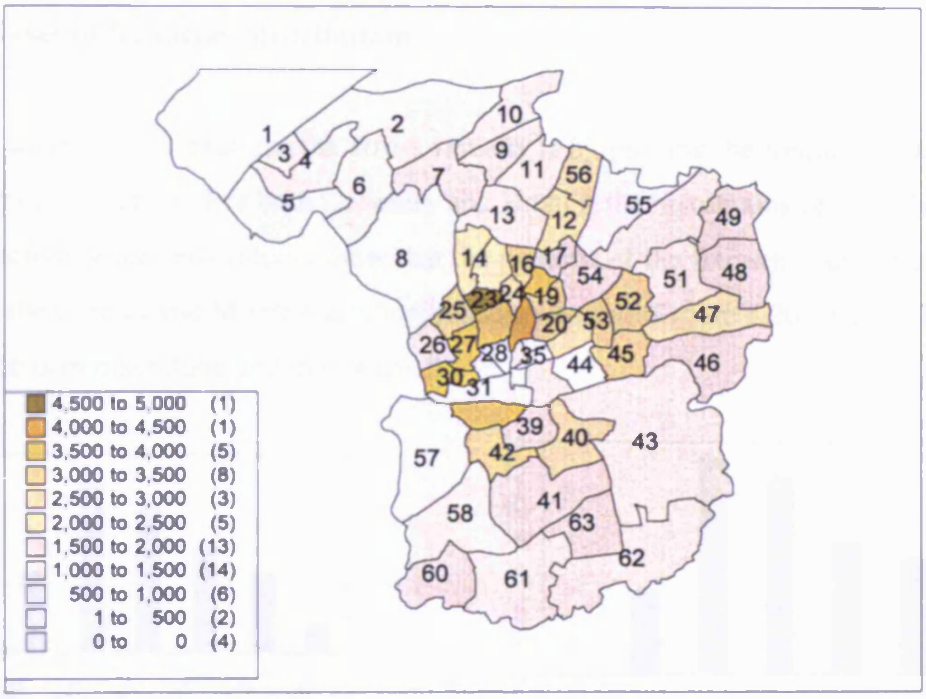


Figure 8. 18: Household density using overlay method (Bristol)

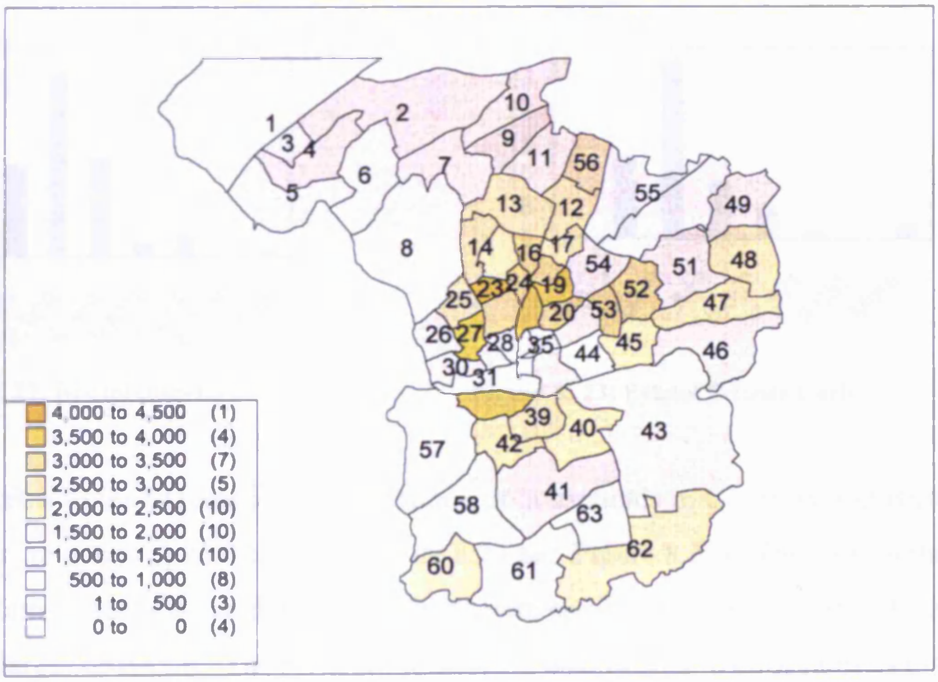


Figure 8. 19: Household density using Monte Carlo Dasymetric method (Bristol)

8.3 Analyses of frequency distribution

Another useful way to examine the errors visually is by plotting the frequency distribution in a form of histogram. For both Coventry and Bristol, the histograms of households over beats (excluding the zero values) show that the patterns of the frequency distribution between the overlay and Monte Carlo methods are similar (Figure 8.20 - 8.23). The only difference is in magnitude and that is small.

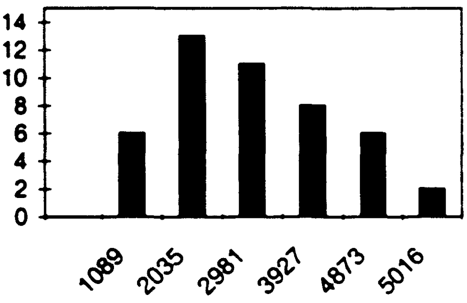


Figure 8. 20 Coventry (overlay)

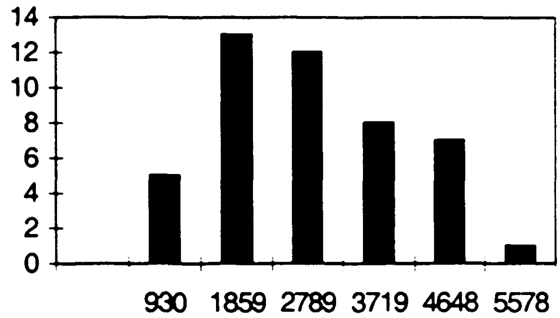


Figure 8. 21 Coventry (Monte Carlo)

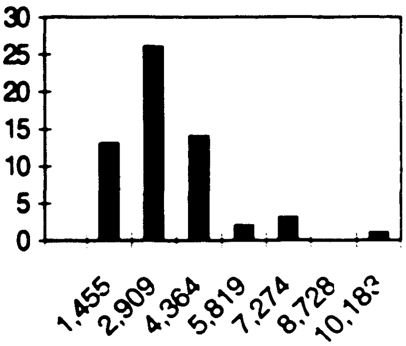


Figure 8. 22: Bristol (overlay)

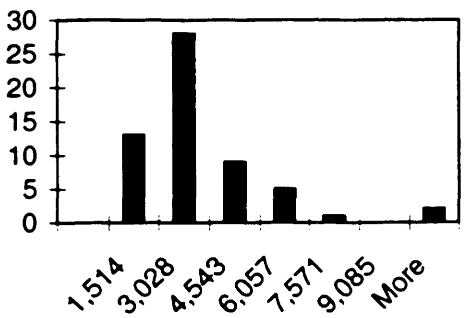
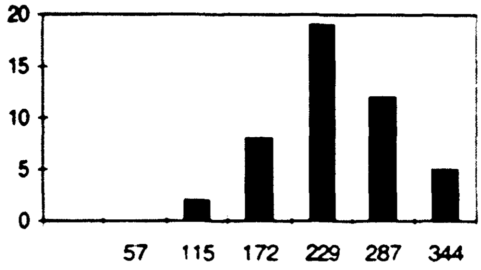


Figure 8. 23: Bristol (Monte Carlo)

The patterns of the SD frequency distribution of households in Coventry and Bristol show a 'mirror' relationship to the above (Figure 8.24 and Figure 8.25). The shift in the distribution pattern may be due to some beats with small household density which might contain larger deviation. The SD distributions are also related to error rates as discussed in the previous sections. Figures 26 and 27 show the error rate (in percentage) of Coventry and Bristol respectively.



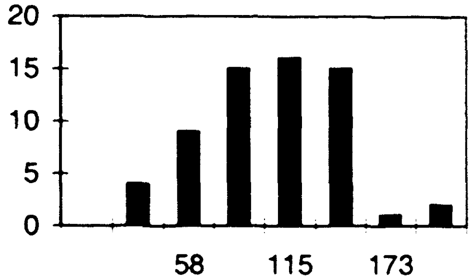
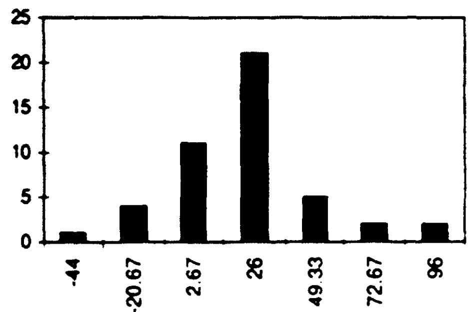


Figure 8. 24: SD (Coventry)

Figure 8. 25: SD (Bristol)

Note that the error distribution for Bristol appears very different in Figure 8.27. This is because the scale has been distorted by one exceptionally high value (Beat 33: 233%!).



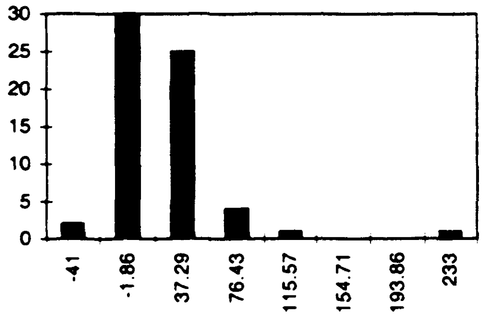
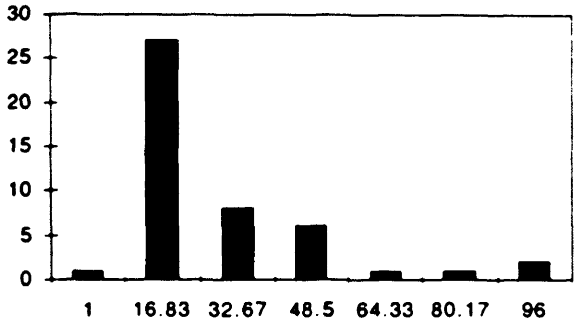


Figure 8. 26: Error distribution in Coventry

Figure 8. 27: Error distribution in Bristol

The mean error appears to be relatively small: 7% for Coventry and -3% for Bristol. This is due to the fact that some of the positive and negative values cancel out each other. However the actual error for individual beats may be much higher. This effect can be shown by plotting the histograms using the absolute values (ignoring the signs) as shown in Figure 8.28 and 8.29.



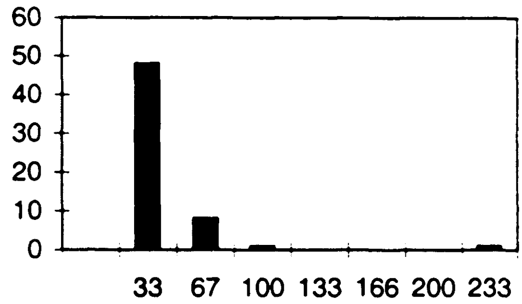


Figure 8. 28: Absolute error rate (Coventry)

Figure 8. 29: Absolute error rate (Bristol)

The above histograms show a lot of the beats could have the absolute error rate as high as 16% for Coventry and 33% for Bristol! It follows that whether the errors are significant or not depends on the beats that are involved in the particular process of data analysis - a topic of further error propagation which will be discussed in greater detail in the next chapter.

8.4 Chapter Summary

Using the Monte Carlo Dasymetric method, more accurate household counts were found for the two cities. These show that the original estimates were an over-estimation of household counts by 7% for Coventry and under-estimation of household counts by -3% for Bristol. Spatial analyses of the error distribution show that **a geographical area would have a higher error if it:**

- **has dense population;**
- **is near the city centre; or**
- **has an irregular geographical boundary.**

Whether or not these errors would significantly affect the Safer Cities Programme evaluation is the subject for further analyses (see next Chapter).

Chapter Nine

Analyses of processing error (II): the impact on Safer Cities Action

This chapter continues with the error analyses in the next phase of the GIS processing. Owing to the erroneous result of spatial interpolation, the output of GIS would also contain errors which are discussed in the previous chapter. How this error affects the further processing of the spatial data as part of error propagation is the subject of this chapter. A tabulation of the results of the spatial data involved in both the transfer and processing errors is provided in Section 9.2. The geographical analyses are shown in Section 9.3 by mapping the results of the overlay and Monte Carlo dasymetric methods for the two cities. The confidence level of error propagation at the initial stage of the data analyses of the Safer Cities Programme Evaluation is discussed in Section 9.4. First however the results of the Safer Cities action will be described which will provide the context for the analyses of the spatial error propagation (next).

9.1.1 Safer Cities action against burglary

This section describes the amount of action present in the Safer Cities Programme evaluation on domestic burglary. As described in Chapter four, domestic burglary was chosen for detailed analyses for Phase 1 reporting because the Safer Cities Programme was most likely to have a measurable impact on burglary (Ekblom *et al.*, 1996a). It follows that to study the propagation of Group 2 errors, I should focus on their impact upon the results of the Safer Cities Programme evaluation on domestic burglary.

The whole evaluation

The local Safer Cities schemes targeted on domestic burglary were identified when they were in scope with our outcome measures (see Chapter 2 for scoping principle). This covered 240 out of a total of 300 schemes. In the *final year of measurement*, 1992, there were 325 action beats. The average action score (intensity) from *Safer Cities* funds was £3 per household. 82 of these beats also had *levered* funds, an average score of nearly £5.50 per household. The *total* score was nearly £4.50. On the basis of the total action present in the *final* year, there were four sets of beats.

Those which

1 *never* had action;

2 ended up in 1992 with *under £5-worth* of action per household (average just under 50p);

3 had between *£5-£13* worth of action (average nearly £8); and

4 had over £13 action (average £34).

The action in each set starts to appear between 1989-90, and reaches the highest level in 1992. Because our scoring calculation system had been set to assume that action, once started, continued to exert its effects for two years, the 1992 scores were in effect cumulative for virtually all action, since the bulk of it started from 1990. There was an association between levered funds and high amounts of Safer Cities action. The 82 action beats with levered funds had an average of just over £7.50 Safer Cities burglary action per household in 1992; the 243 without leverage, just over £1.50. Of the 33 beats in the high (total) action set, 22 had leverage.

The average amount of action across the burglary schemes as a whole was £8,700 (Safer Cities funding). This was above the average Safer Cities fund spent (which was £7,300). The amount varies according to whether there were additional 'levered-in funds' raised from local agencies and institutions, and from other national programmes. According to the data recorded on the Management Information System (see Chapter two), two-thirds of the burglary schemes had no leverage.

For the remaining third of schemes with levered-in funds, the average Safer Cities spend was £11,300 and the average levered supplement £17,800. In other words, the Safer Cites action was actually less than the other non-Safer Cities action. (Safer Cities funds were not used to substitute for funds from other sources. Otherwise, the Safer Cities spending would have been more when levered funds were unavailable.) On average, the levered schemes were geographically smaller than the rest by 53% (19 versus 29 enumeration districts). See Ekblom *et al* (1996a) for further details on the evaluation as a whole.

The two cities

Table 9.1 shows the summary of the total burglary preventive action scores for each of the two cities selected for this case study over the period of the Safer Cities Programme (taking their related errors into account).

Table 9. 1: Total burglary preventive action score (adjusted) per year

Year	1990	1991	1992	Total
Bristol	4.14	27.43	41.36	72.93
Coventry	4.35	37.42	42.07	83.84

The trends of the spending were very similar across the two cities. It is interesting to observe that although Coventry is a smaller city than Bristol in terms of both geographical size and population, the burglary action scores are higher than for Bristol (particularly in 1991).

9.2 Error propagation in the Safer Cities Action data

To analyse the error propagation and its impact upon the Safer Cities Action scores, one can focus only on those beats that had burglary action and compare their action scores obtained from overlay method and Monte Carlo dasymetric method (or Monte Carlo for short). The action scores (S), can be computed from Equation (2.1) which, for beat-based analysis, can be simplified as follows (see Ekblom *et al*, 1994; 1996a for detail).

$$S = A \times N_{ZO} / N_{ZI} / N_{ZD} \qquad \textbf{(9.1)}$$

where S is the action score
N_{ZO} number of household counts in the zone of overlap, N_{ZI} zone of influence, and N_{ZD} zone of detection respectively

For Police Crime data, ZI, ZO, and ZD are all beat-based, after scoping has been completed, the formulas can further be simplified as:

$S = A \times B/B/B$ which simplifies to

$$S = A/B \qquad \textbf{(9.2)}$$

where B is the beat household count.

For corrected scores with corrected beat household counts B' from Monte Carlo simulation, we have

$$S' = A/B'$$

or

$$S'B' = SB$$

$$S' = SB/B' \qquad (9.3)$$

Since we know error rate e:

$$e = (B - B')/B'$$

$$= B/B' - B'/B'$$

$$= B/B' - 1$$

or

$$B/B' = 1 + e \qquad (9.4)$$

Substitute the value of B/B' in terms of e into equation (9.3), S' corrected action score can be computed:

$$S' = S + eS \qquad (9.5)$$

Using the above formula, a new set of action scores for each beat that has action can be calculated.

First those beats that have action are obtained from the scoping process. These are the beats included in Tables 9.2 and 9.3 for Coventry and Bristol respectively, together with the household counts calculated by the **overlay** method, Monte Carlo dasymetric method (**MC**), the standard deviation (**SD**) of the household counts due to the Monte Carlo method (and SD %), and the **error** (in percentage) due to the overlay method as estimated by the Monte Carlo dasymetric method. For a non-biased estimator, the SD is the same as the RMSE of the household estimation by the Monte Carlo simulation (established at the implementation in Chapter 7). The data shown in Tables 9.2 and 9.3 are effectively subsets of the data shown in Tables 8.1 and 8.3. As a result of scoping, smaller sets of beats are left for further processing.

Table 9. 2: Population calculated by overlay and its error in % and standard deviation as estimated by Monte Carlo simulation for those beats that have burglary action in Coventry.
(Beats which had no burglary preventive actions have been excluded)

Beat-ID	Overlay	MC	SD	SD %	error %
8	5194	4096.54	263.63	1.54	26.79
9	4320	3837.19	247.79	0.15	12.58
10	2950	2831.76	216.46	1.09	4.18
15	4543	4512.05	308.71	1.38	0.69
17	2513	2200.6	282.32	1.4	14.2
19	1029	840.77	179.46	0.5	22.39
23	1707	912.12	149.92	0.5	87.15
24	829	949.63	184.12	0.78	-12.7
25	911	1069.21	157.95	3.56	-14.8
26	2283	3462.74	237.62	0.5	-34.07
30	402	301.59	120.92	0.41	33.29
31	1967	1218.1	269.23	5.79	61.48
32	1229	1420.57	340.04	2.04	-13.49
33	349	457.39	129.93	3.45	-23.7
34	1742	1671.63	278.27	0.64	4.21
100	3396	3948.78	303.738	1.18	-14
101	3297	3016.28	265.37	1.7	9.31
Total	**38661**	**36746.95**	**3935.478**	**1.565294**	**5.21**
Mean	**2274.176**	**2161.585**	**231.4987**	**1.57**	**5.21**

Table 9. 3: Population calculated by overlay and its error in % and standard deviation as estimated by Monte Carlo simulation for those beats that have burglary action in Bristol.

Beat-ID	Super-ID	overlay	MC	SD	SD %	error %
36	209	1,545	1,236	39.11	3.16	24.98
53	210	1,779	1,880	38.44	2.04	-5.38
45	218	3,782	2,646	47.17	1.78	42.94
20	219	1,538	1,517	37.13	2.45	1.4
19	220	1,439	1,732	50.37	2.91	-16.9
54	221	1,997	1,882	59.45	3.16	6.1
52	223	3,813	3,934	65.19	1.66	-3.07
41	224	6,295	6,655	100.17	1.51	-5.41
63	225	3,359	2,357	73.8	3.13	42.54
58	227	4,014	4,669	90.9	1.95	-14.04
40	228	4,541	3,711	85.59	2.31	22.36
48	233	3,980	4,267	100.51	2.36	-6.72
47	234	4,447	4,563	88.2	1.93	-2.55
Total		**42,529**	**41,049**	**876.03**	**2.13**	**3.61**
Mean		**3,271**	**3,158**	**67**	**2.13**	**3.61**

The exclusion of some beats has reduced the total error from 7% to 5% for Coventry. This represents the reduction of the total household count from 109,084 ± 9622 to 36,747 ± 3935 in the scoring process (as a result of the scoping). In other words, just under 34% of the households are included in the final analysis.

For Bristol, the total number of households are reduced from 159,853 ± 5,560 to 42,529 ± 41,049. In other words, just over 25% of the households are included in the action score calculation. No superbeat is involved in the burglary preventive actions. In contrast to Coventry, the change in error (%) is small in magnitude (-3% vs +3.61%) but the sign is changed from negative to positive.

Since the action scores represent the amount of money spent per household, the over-estimation of the household density would actually under-estimate the amount of action scores. The results show that this is indeed the case (Tables 9.4 & 9.5).

Table 9. 4: The effect of the error propagation in scoring the cost of preventing burglary on dwellings (BD) in Coventry.

Beat	error %	BD90	Adjusted	BD91	Adjusted	BD92	Adjusted
8	26.79	0	0	0.01	0.01	0.01	0.01
9	12.58	0	0	0.03	0.03	0.03	0.03
10	4.18	0.19	0.2	0.39	0.41	0.53	0.55
15	0.69	0.31	0.31	3.28	3.3	4.81	4.84
17	14.2	0	0	0.1	0.11	1.03	1.18
19	22.39	0	0	0.22	0.27	2.4	2.94
23	87.15	0	0	0.05	0.09	0.05	0.09
24	-12.7	0	0	0.16	0.14	0.16	0.14
25	-14.8	0	0	0.03	0.03	0.03	0.03
26	-34.07	0	0	0	0	0.23	0.15
30	33.29	0	0	0.08	0.11	0.08	0.11
31	61.48	0	0	0.57	0.92	6.27	10.12
32	-13.49	0	0	0.57	0.49	6.27	5.42
33	-23.7	0	0	0.2	0.15	2.25	1.72
34	4.21	0	0	0.11	0.11	1.22	1.27
100	-14	0.88	0.76	1.83	1.57	2.52	2.17
101	9.31	2.59	2.83	26.51	28.98	10.49	11.47
Total		**3.97**	**4.35**	**34.14**	**37.42**	**38.38**	**42.07**
Error %	**5.21**	**-8.74**		**-8.77**		**-8.77**	

Table 9. 5: The effect of the error propagation in scoring the cost of preventing burglary on dwellings (BD) in Bristol.

Beat-ID	error %	BD90	Adjusted	BD91	Adjusted	BD92	Adjusted
36	25	0	0	2.09	2.61	3.28	4.1
53	-5	0	0	10.29	9.74	16.09	15.22
45	43	0	0	0.88	1.26	1.38	1.97
20	1	0	0	0.13	0.13	0.29	0.29
19	-17	0	0	0.09	0.07	0.21	0.17
54	6	0	0	0.64	0.68	1.01	1.07
52	-3	0	0	8.61	8.35	13.53	13.11
41	-5	2.95	2.79	2.95	2.79	2.95	2.79
63	43	0.95	1.35	0.95	1.35	0.95	1.35
58	-14	0	0	0	0	0.61	0.52
40	22	0	0	0	0	0.14	0.17
48	-7	0	0	0.24	0.22	0.24	0.22
47	-3	0	0	0.23	0.22	0.35	0.34
Total		**3.9**	**4.14**	**27.1**	**27.43**	**41.03**	**41.36**
Error %	**3.61**	**-5.8**		**-1.2**		**-0.8**	

For Coventry, over-estimating the household density by 5% (Table 9.2) resulted in the under-estimation of the mean action scores by -9% (Table 9.4). The amount was relatively constant through out the period of the Safer Cities Programme. On the other hand, over-estimating the household density by 3.6% (Table 9.2) resulted in the under-estimation of the mean action scores by only -0.8% (Table 9.4) in Bristol (a very small amount).

9.3 Geographical analyses of error propagation

The thematic maps with appropriate constraints (scoping, scoring and Monte Carlo simulation in this case) can be used as a visualisation tool to explore the implication of the results similar to the error analyses described in the previous chapter. Appendix 9.1 shows the whole range of beat maps of the action scores for burglary prevention for each year of the Safer Cities Programme.

Taking the estimated errors by the Monte Carlo method into account, the more accurate 'estimated' maps of the action scores can be plotted which can be visually compared with the maps generated by the overlay method. Since there are relative small amounts of difference in the spatial pattern shown in the two sets of the maps especially during the first year when there are not much burglary preventive action.

The analysis is focused on the final year when the action scores accumulate to their maximum amounts. For the convenience of cross-referencing, the maps in year 1992 (Appendix 9.1) are included here as shown in Figures 9.1 - 9.4.

Beat 26 in Coventry (near the south-west of the city boundary) has a small amount of action. The amount of the action scores had also been increased for some beats (Beat 19, 31-34 in Coventry: Figure 9.1). The spatial patterns between the two methods are broadly similar. Exceptions are Beat 31 in Coventry and Beat 45 in Bristol (Figures 9.2 and 9.4).

With the Monte Carlo method, Beat 31 in Coventry becomes one class higher, while Beat 45 in Bristol is one class lower though it is less noticeable because the amount of the action intensity is very low (< £1.5). However, both cases indicate that they have a higher error rate. Both beats are near the city centre and have a high compactness ratio (as described in the previous chapter). This shows that such geographical characteristics have an effect upon the error propagation when its attribute values (in this case, action scores) are included in the process (scoping).

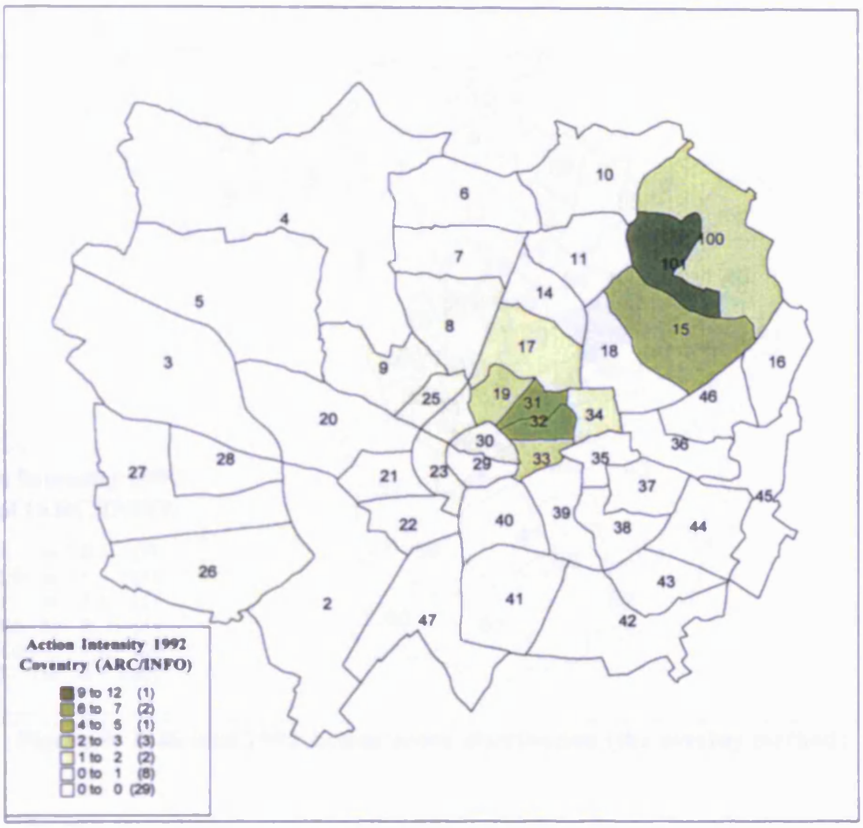


Figure 9. 1: Coventry 1992 Action score distribution (the overlay method)

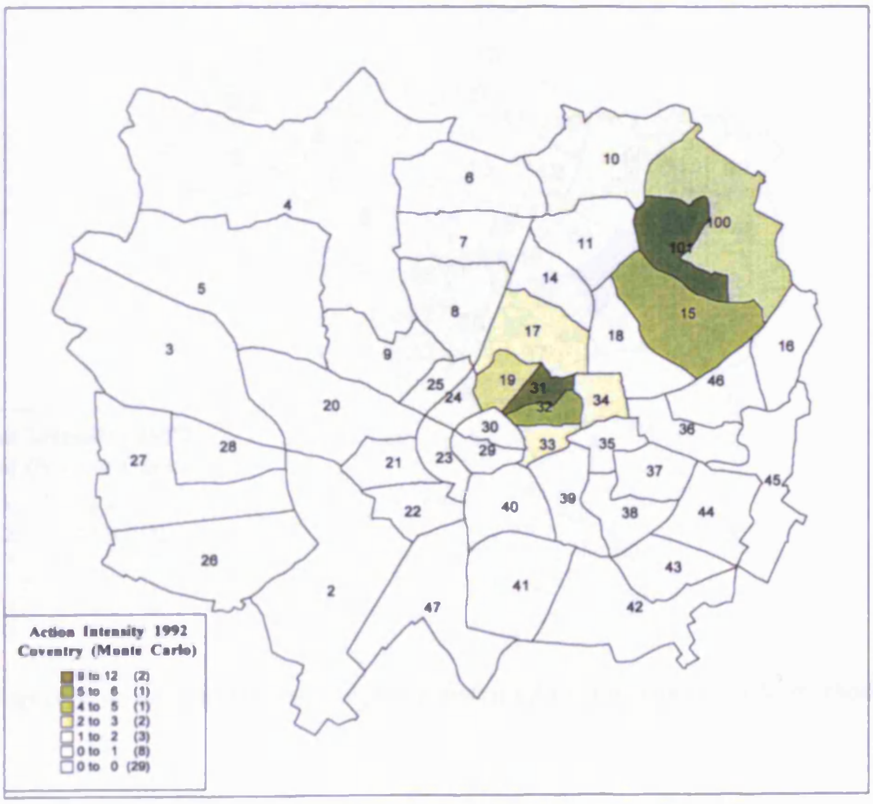


Figure 9. 2: Coventry 1992 Action score distribution (the Monte Carlo method)

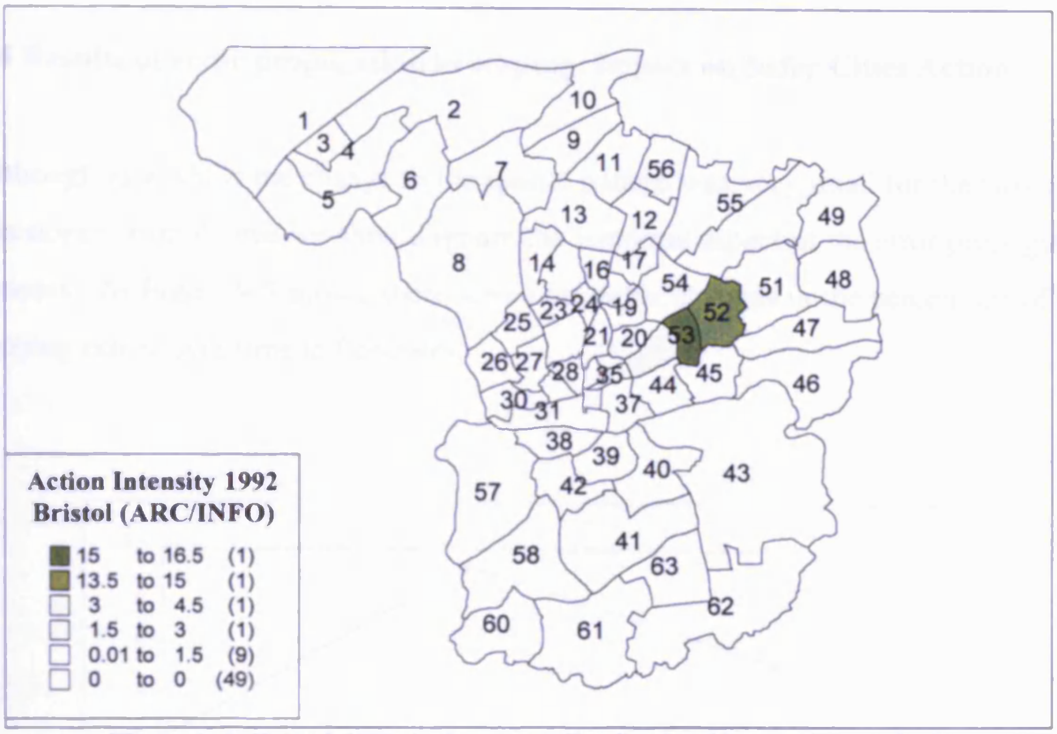


Figure 9. 3: Bristol 1992 Action score distribution (the overlay method)

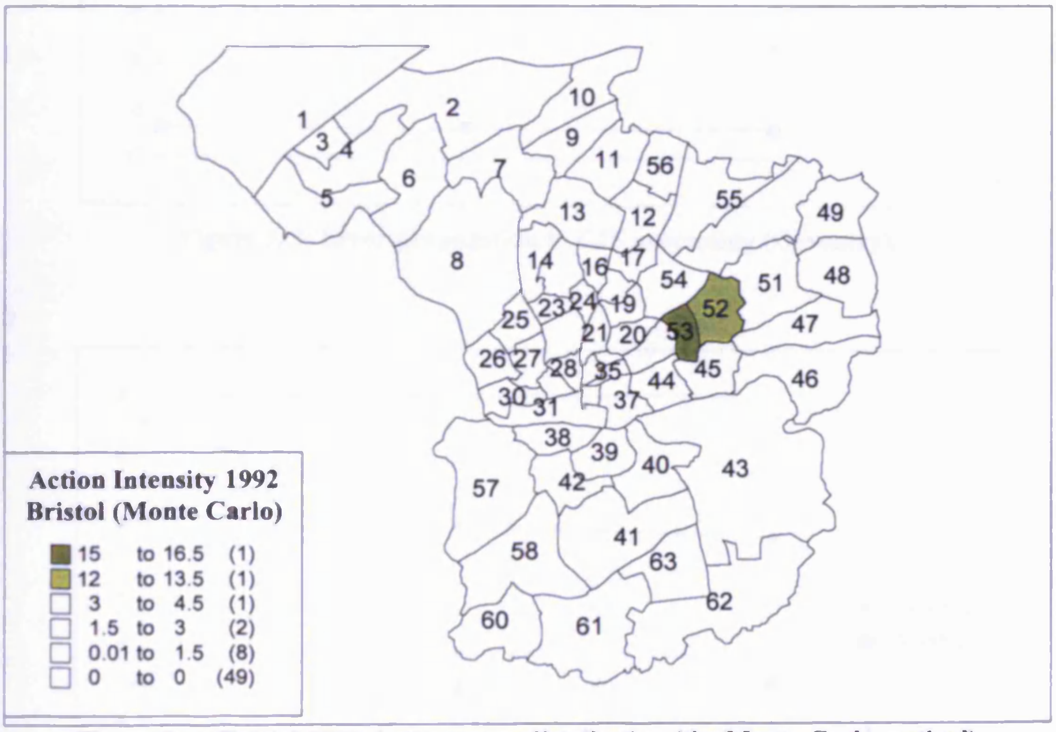


Figure 9. 4: Bristol 1992 Action score distribution (the Monte Carlo method)

9.4 Results of error propagation in scoping: impact on Safer Cities Action

Although as a whole the change in the spatial pattern was very small for the two cities, this does not imply that we should ignore the temporal aspect of the error propagation process. As Figure 9.5 shows, there were noticeable changes in the percentage of the scoping errors over time in Coventry.

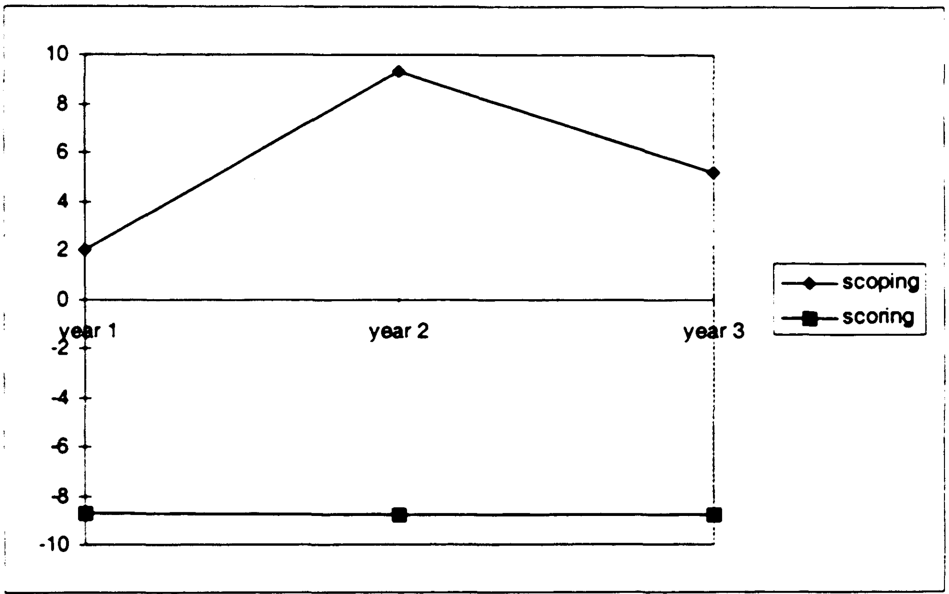


Figure 9. 5: Error propagation in GIS processing (Coventry).

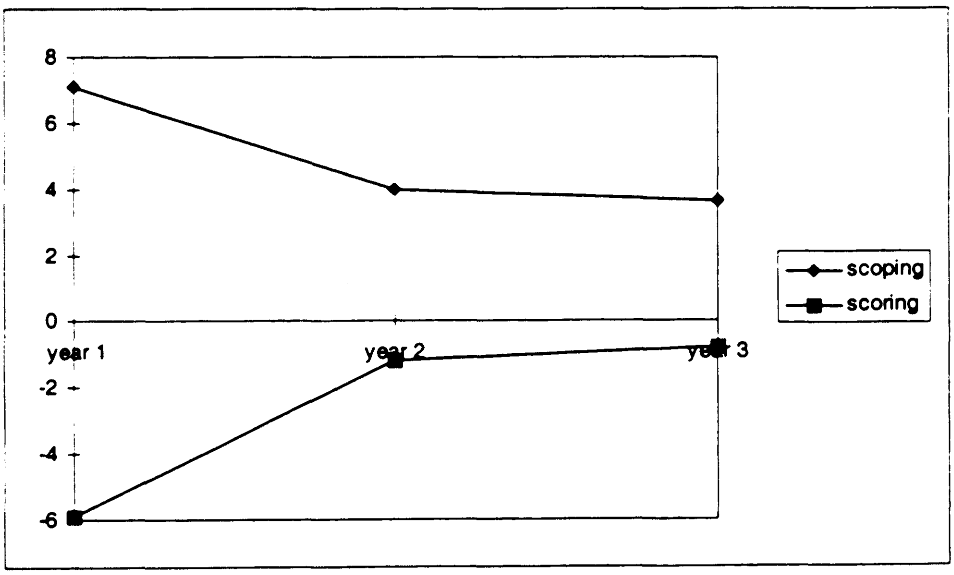


Figure 9. 6: Error propagation in GIS processing (Bristol).

It seems by chance that the irregular nature of the error in household density has been cancelled out by the amount of money spent in Coventry (which was also irregular). However this is not the case for Bristol (Figure 9.6). Despite the irregular action scores, there is a strong negative correlation between scoping and scoring errors for both cities as they were determined by a set of mathematical relationships ($r = -0.9165$ for Coventry & $r = -0.9997$ for Bristol). This implies that even though there is a close relationship between spatial and computational processing (scoping and scoring), the outcome is still unpredictable. This is due to the complex nature of spatial variation and its interaction with the computation.

Having precisely quantified the spatial error of scoping and the propagation error of scoring, the next key question to be answered is: are these errors significant? The meaning of significance depends on a particular statistical context. The degree of significance is analysed at two levels: 1) GIS process error; and 2) transfer error for further statistical analyses. First, in terms of the scoping in GIS processes, one can examine whether the quantity of the household density is significantly deviated from the mean of the household counts. Based on our classification of the pixels used in the simulation, we have independent maximum likelihood values for μ and Δx (Δx is the error process in x). To interpret the quantity of the household counts x with its RMSE Δx (thus the estimated value = $x \pm \Delta x$), one can consider a normal distribution mean μ, and variance Δx^2 (Figure 9.7).

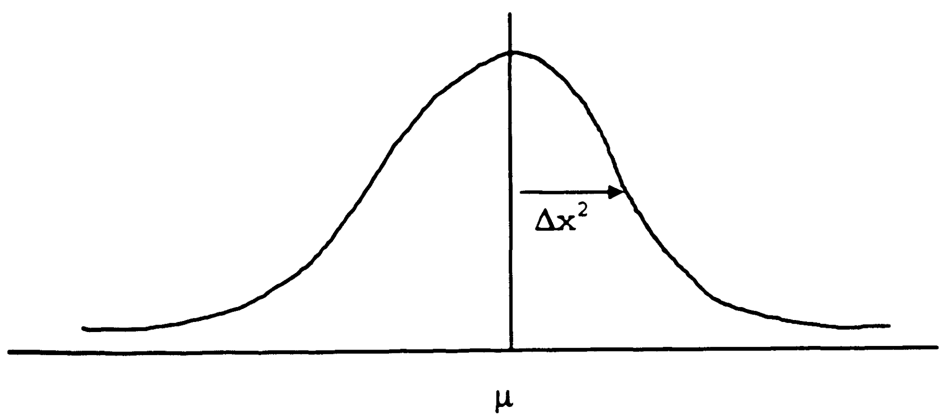


Figure 9. 7: a normal distribution

When Δx = SD, then particular properties follow. Statistically, it is known that 67% of the results would be within $\mu \pm SD$; 95% of the results within $\mu \pm 2SD$; and 99% of the results within $\mu \pm 3SD$. Since we are looking for an effect of the error process (e) on scoping and scoring, it follows that $e = |\mu - y|$ where y is the value estimated by the overlay method. Conservatively, if $e \geq \Delta x$, it can be concluded that the effect of error is significant. Conversely, if $e < \Delta x$, it can be concluded that the results were not significant enough to determine the error effect on the action scores of size $< 2\Delta x$, and that one can assume the error is not important and accept the Null hypothesis ($e = 0$).

For Coventry

Most of the beats in scope are well within one standard deviation of error, that is 67% (for example, figure 9.8). The complete range of the frequency distribution of the Monte Carlo sampling for Coventry in the form of a histogram for the beats in scope is shown in Appendix 9.2. The only exception is Beat 23 (Figure 9.9). This beat is just near the city centre which explains the high error rate as discussed in the previous chapter. Moreover, the action score for this beat is very small (total 18 pence per household over three years). It is unlikely that this would have a significant impact upon the overall conclusion of the evaluation.

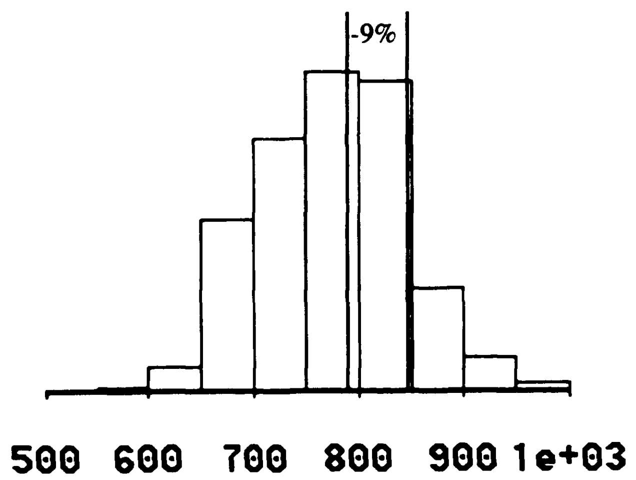


Figure 9. 8: Overlay error deviated from the mean (Coventry Beat 101)

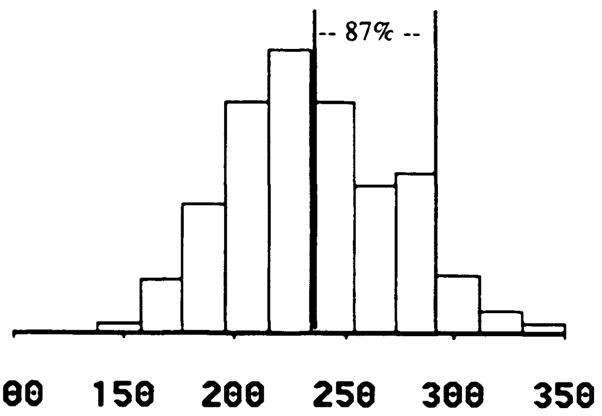


Figure 9. 9: Overlay error deviated from the mean (Coventry Beat 23)

The form of the distribution also deviates from the normal distribution in this case. It looks more like a bimodal distribution. This is associated with a high error rate. For example, Beat 31 is at the borderline of one standard deviation which has a similar bimodal distribution (Figure 9.10). This implies that a number of different household distributions within these beats may exist.

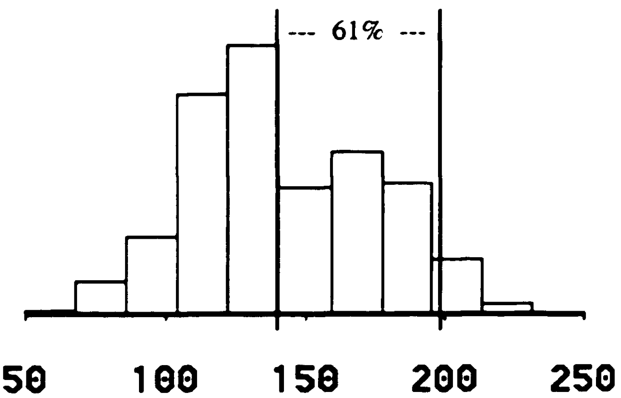


Figure 9. 10: Overlay error deviated from the mean (Coventry Beat 31)

For Bristol

With the exception of Beat 26, all the beats in scope have the error distribution well within one standard deviation (Figure 9.11). The complete range of the frequency distribution of the Monte Carlo sampling for the Bristol beats in scope is shown in Appendix 9.2. Beat 36 (a city centre beat) is excluded from the analyses within this context as the sample household count was not large enough to show the normal distribution. This would not affect our analyses as both the household density and the action scores were very small.

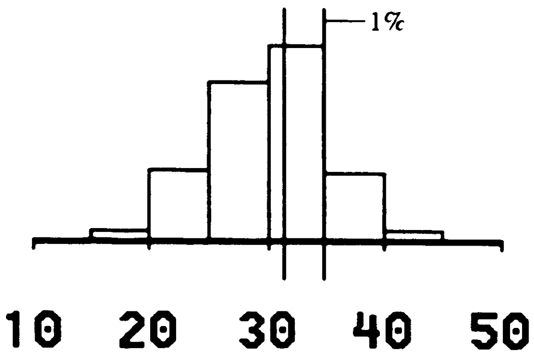


Figure 9. 11 overlay error deviated from the mean (Bristol Beat 19)

The above analyses suggest that most errors in scoping are not significantly different from the estimated values generated from the Monte Carlo experiments. However the statistical results should be used in a theoretical context. Here the importance is the decision criteria. So at the second level of analyses, the word 'significance' is taken to mean: Does the error affect the conclusion of the evaluation? To answer this question, one needs to re-examine the statistical analytical process of the evaluation which, as suggested in Chapter two, is complex, and thus it deserves a separate treatment (Chapter 11). This also relates to the outcome measures of the Safer Cities Programme Evaluation which also required detailed analyses (next chapter).

9.5 Chapter Summary

This chapter has focused on the error propagation of the process error (the so-called *scoping process*) resulting from the spatial interpolation (described in the previous chapter) and examines its the impact upon the Safer Cities action through those beats that had Safer Cities burglary action. The action scores in each beat-year were re-calibrated taking into account the error rate for each beat (the *scoring process*). Since the action scores represent the amount of money spent per household, the over-estimation of the household density would actually under estimate the amount of the action scores. The results show that this is indeed the case. The over-estimation of household density (5% for Coventry and 3.6% for Bristol) has resulted in the under-estimation of the cost of the Safer Cities action by -9% and -0.8% for Coventry and Bristol respectively. The amount of the resultant error of each individual beat depends not only on the initial errors from the GIS spatial interpolation, but also on whether the beat is involved in the further processing (scoping in this case) and the value of the attribute (action score). The analyses so far suggest that the GIS process error has not yet had a significant impact upon the Safer Cities action intensity. Whether the error has a significant impact upon the outcome and the conclusion of the Safer Cities Programme Evaluation requires further analyses.

Chapter Ten
Analyses of processing error (III): the impact on burglary

This chapter examines the impact of spatio-thematic error upon the outcome measures of the Safer Cities Programme Evaluation. Section 10.1 defines the burglary risk as the incidence rate and describes the units of analyses by referring back to the context of the Safer Cities Programme Evaluation. Section 10.2 provides a summary description of the results of error propagation on the burglary risk of the two cities. Section 10.3 discusses the implication of the results upon the crime pattern analysis. Finally, Section 10.4 provides a brief summary of this chapter.

10.1 Outcome data and Units of analyses

As described in Chapter two, the recorded crime data for twelve major offence categories were collected from 14 of the 16 Safer Cities evaluated from 1987 to 1992. Although data for some cities such as Coventry were available in 1993, only data up to 1992 were used for the Evaluation for completeness. The outcome measure is defined as the risk of victimization, which within the scope of this study, is the *burglary risk*. This converts the burglary count into an *incidence rate, which* is defined as *the number of domestic burglaries per 100 households in each beat, in each year* (Ekblom *et al*, 1996a).

Super-beats in Bristol

Some beat boundaries were changed in Bristol during the period of the Safer Cities Programme. The change tended to involve the aggregation of smaller beats adjacent to each other into larger beats, and hence so-called *super-beats*. The super-beat problem was managed by looking back through past maps to uncover 'beat pedigrees'. A 'standard beat map' was used which covered the whole period of the Safer Cities Programme. Super-beat-IDs were used to identify these larger beats. Data were aggregated for the super-beats and processed for data analyses as usual.

For Bristol, the super-beats are Beats 9 and 10 (grouped into Super-beat 211); 15, 16, and 24 (Super-beat 212); 25-30 (Super-beat 222); 38, 39, 42, and 57 (Super-beat 226); 58 and 60 (Super-beat 227); 40, 43, 61, 62 (Super-beat 228); 8 and 14 (Super-beat 229); 11 and 12 (Super-beat 231); 7 and 13 (Super-beat 232); 48-50, and 55 (Super-beat 233); 47 and 51 (Super-beat 234); 22, 23, 28, and 29 (Super-beat 235). The super-beat-IDs are included in Appendices 10.1-10.2 for reference.

10.2 Error propagation in burglary risk

Since most beats had burglary incidence at some point of the Safer Cities Programme, they would all have been affected by the spatial interpolation error. According to the definition, for each beat year, the burglary risk (R):

$$R = (C/B)\ 100 \qquad \textbf{(10.1)}$$

where C is the burglary count per beat-year

B is the beat household count.

Similarly given a new household count (B') estimated by the Monte Carlo dasymetric method, the new burglary risk R':

$$R' = (C/B')100 \qquad \textbf{(10.2)}$$

R' can be computed in terms of R by Dividing (10.2) with (10.2),

$$R'/R = C/B' \ / \ C/B = B/B' \qquad \textbf{(10.3)}$$

From equation (9.4), the ratio between the beat household count by the overlay method and the estimated beat household count by the Monte Carlo dasymetric method (B/B') = 1 + e. Substitute the value of B/B' into equation (10.3), the 'corrected' burglary risk can be calculated:

$$R' = R + eR \qquad \textbf{(10.4)}$$

A new set of burglary risk can be computed using the error rate estimated by the Monte Carlo dasymetric method according to the above formula (in the same way as the action scores described in the previous chapter). Table 10.1 shows the summary of the mean burglary risk per beat for Bristol and Coventry. For a complete listing of the burglary risk computed by the overlay method and the Monte Carlo dasymetric method for Bristol and Coventry, see Appendices 10.1 - 10.4.

Table 10. 1: Mean burglary risk per beat

		1987	1988	1989	1990	1991	1992	mean average
Bristol	overlay	0.032957	0.04402	0.032433	0.044084	0.06668	0.045343	0.04425
	Monte Carlo	0.035486	0.047929	0.036003	0.048015	0.069561	0.047819	0.04747
	mean error %	-7.13	-8.16	-9.92	-8.19	-4.14	-5.18	**-6.77**
Coventry	overlay	0.043823	0.056841	0.051481	0.058354	0.06463	0.081386	0.05942
	Monte Carlo	0.045676	0.059559	0.055223	0.061053	0.067543	0.085546	0.06243
	mean error %	-4.06	-4.56	-6.78	-4.42	-4.31	-4.86	**-4.83**

On average, the above results represent an under-estimation of the burglary risk by -7% for Bristol and -5% for Coventry. Since police recorded crime data are subject to under-reporting (compare with the British Crime Survey, Mayhew *et al*, 1993), a further under-estimation of the burglary risk may have an important impact upon the Evaluation of the Safer Cities Programme. Figures 10.1 and 10.2 show the average burglary risks per beat as they change over time for Bristol and Coventry respectively. In both cases, the Monte Carlo dasymetric method presents the same burglary trends as the overlay method. In contrast to the national crime rate at the time, the burglary risk continued to rise after 1990, and only fell after the completion of the Safer Cities Programme in 1992 for Bristol (and in 1993 for Coventry when the data were included). These patterns show some *prima facie* evidence of a Safer Cities effect. However, the confirmation of such an effect requires statistical modelling to link action with the outcome measures. How the estimated errors affect the conclusion of the Safer Cities Programme Evaluation using multilevel modelling method is the subject of next chapter.

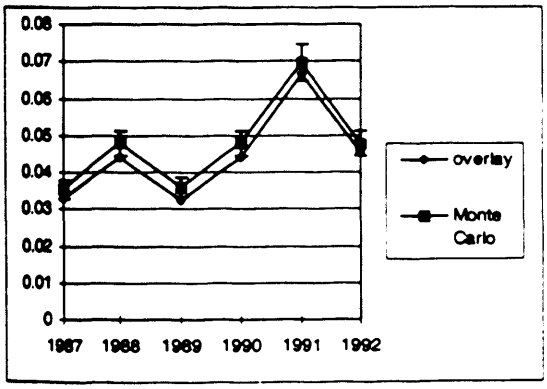


Figure 10. 1: Mean burglary risk (Bristol)

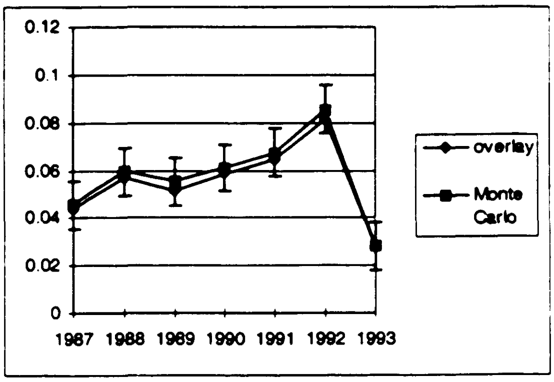


Figure 10. 2: Mean burglary risk (Coventry)

10.3 Geographical analysis of burglary risk - crime pattern analysis

The spatial pattern of the burglary risk for each year can be visualised on a series of thematic maps similar to the error analyses described in the last two chapters. The process of such inspection is also known as crime pattern analysis - an increasing important task in crime prevention as a result of the impact of using GIS (as discussed in Chapter 1). The full sets of the beat maps showing the burglary risk from 1987-1992 for Bristol are shown in Appendices 10.5 and 10.6 (overlay and Monte Carlo methods respectively), and for Coventry in Appendices 10.7 and 10.8 from 1987-1993. Emerging crime patterns can be identified by inspecting these maps over the six year period for the two cities.

For Bristol, Beats 19, 20, and 63 consistently had the highest burglary risk over the six year period. Beat 19 and 20 were densely populated areas just outside the city centre (North East). Beats 32, 34, and 35 had no domestic burglary at all because they were city centre beats and had zero household count. Beats 6, 8, 18, 32, 43, 46, 57, 58, 60, 61 and 62 consistently enjoyed relatively low burglary risk. Most of these beats were country side and park areas and had low population density. Beats 3 and 17 changed from no burglary to relatively high burglary risk over time. This was probably due to a new development of residential housing in these areas.

The thematic maps from the Monte Carlo dasymetric method follow a similar pattern. An exception is beat 33, which is shown up as a hot spot on the Monte Carlo thematic map but not on the overlay map. This is a small beat just next to the city centre (West of Beat 35) and has the highest estimated error (see Chapter 8). An important implication of such an observation is that a less accurate map may miss a potential crime hot spot. This may have a further implication in resource allocation for crime prevention. For easy reference the maps showing the burglary risk in 1989 as estimated by both methods are shown in Figures 10.3 and 10.4.

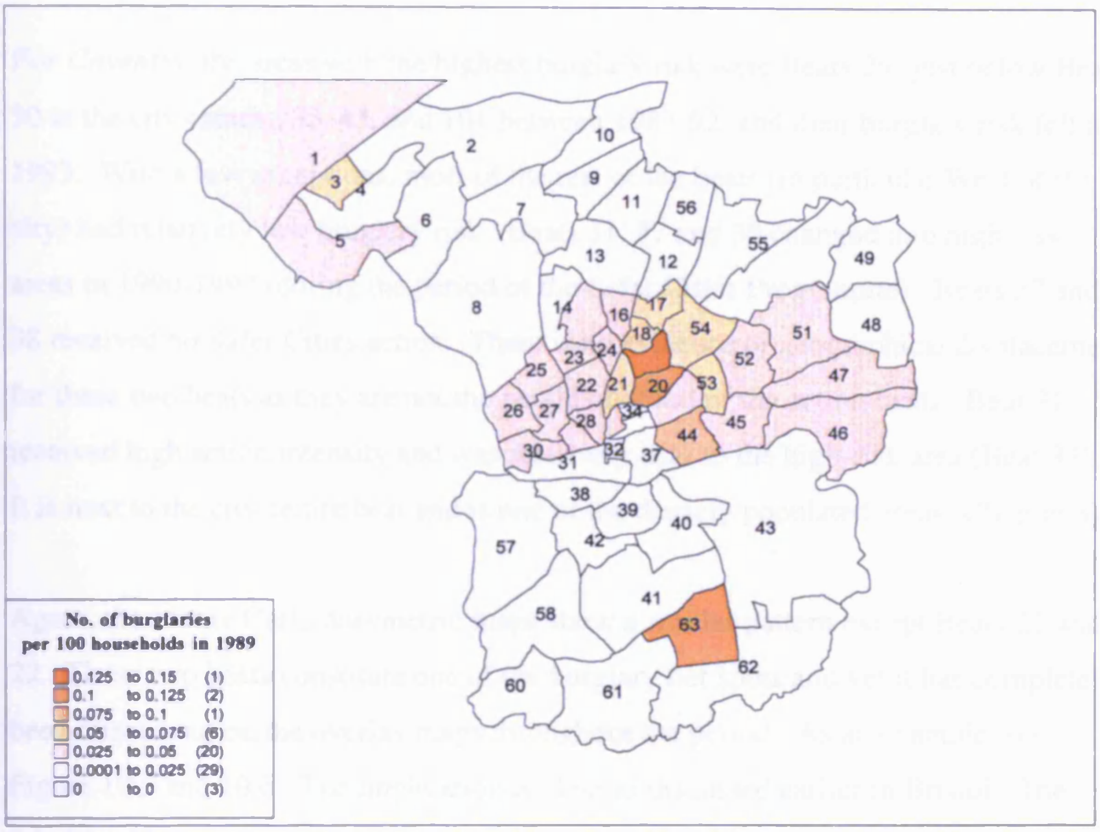


Figure 10. 3: Burglary risk in Bristol 1989 (overlay method)

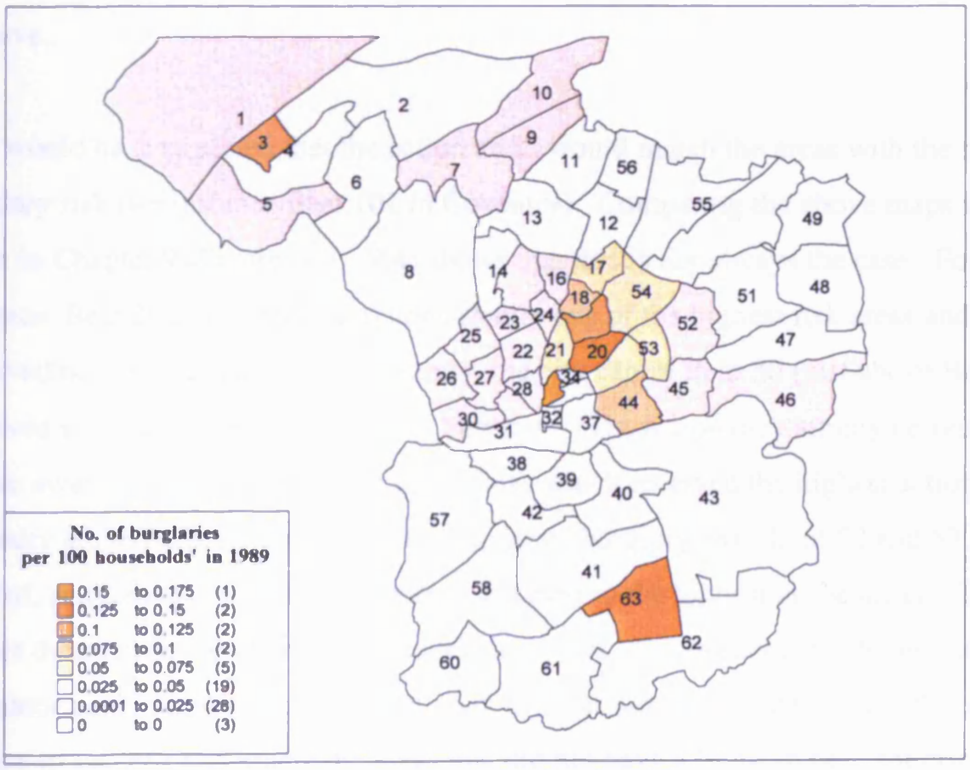


Figure 10. 4: Burglary risk in Bristol 1989 (Monte Carlo dasymetric method)

For Coventry, the areas with the highest burglary risk were Beats 29 (just below Beat 30 at the city centre), 33, 43, and 101 between 1987-92, and their burglary risk fell in 1993. With a few exceptions, most of the rest of the beats (in particular West of the city) had relatively low burglary risk. Beats 31, 37 and 38 changed into high-risk areas in 1990-1992 (during the period of the Safer Cities Programme). Beats 37 and 38 received no Safer Cities action. There is no evidence of geographical displacement for these two beats as they are not the neighbourhood of the action beats. Beat 31 received high action intensity and was relatively near to the high-risk area (Beat 33). It is next to the city centre beat and is one of the densely populated areas (Chapter 8).

Again, the Monte Carlo dasymetric maps show a similar pattern except Beats 21 and 22. These two beats constitute one of the burglary hot spots and yet it has completely been missed out on the overlay maps throughout the period. As an example, see Figure 10.5 and 10.6. The implication is clear as discussed earlier in Bristol. The mis-identification of the hot spots might cause inappropriate allocation of resources such as the Safer Cities Programme fund. The beats that are not perceived as high crime areas would attract less preventive action than the amount that they actually deserve.

One would have expected that the action areas would match the areas with the highest burglary risk (for instance Beat 101 in Coventry). Comparing the above maps with those in Chapter 9 (Figures 9.1 - 9.4) shows that this is not always the case. For instance, Beat 29 in Coventry was consistently one of the highest risk areas and received no action at all. On the contrary, the city centre Beat 30 (just above Beat 29) received some action but had minimal burglary. (The action may simply be raising public awareness at the city centre.) The areas which received the highest action intensity are those beats with relatively moderate burglary risk (Beat 52 and 53 in Bristol, and Beat 31 and 32 in Coventry) just next to the highest crime areas. These reflect the complexity of crime preventive activities. For instance, the Safer Cities co-ordinator might target the areas which they thought would provide them with the best chance to succeed and ignore the areas that did not have a 'hope in hell' for crime reduction. (For further discussion on the process of the Safer Cities Programme see Sutton, 1996).

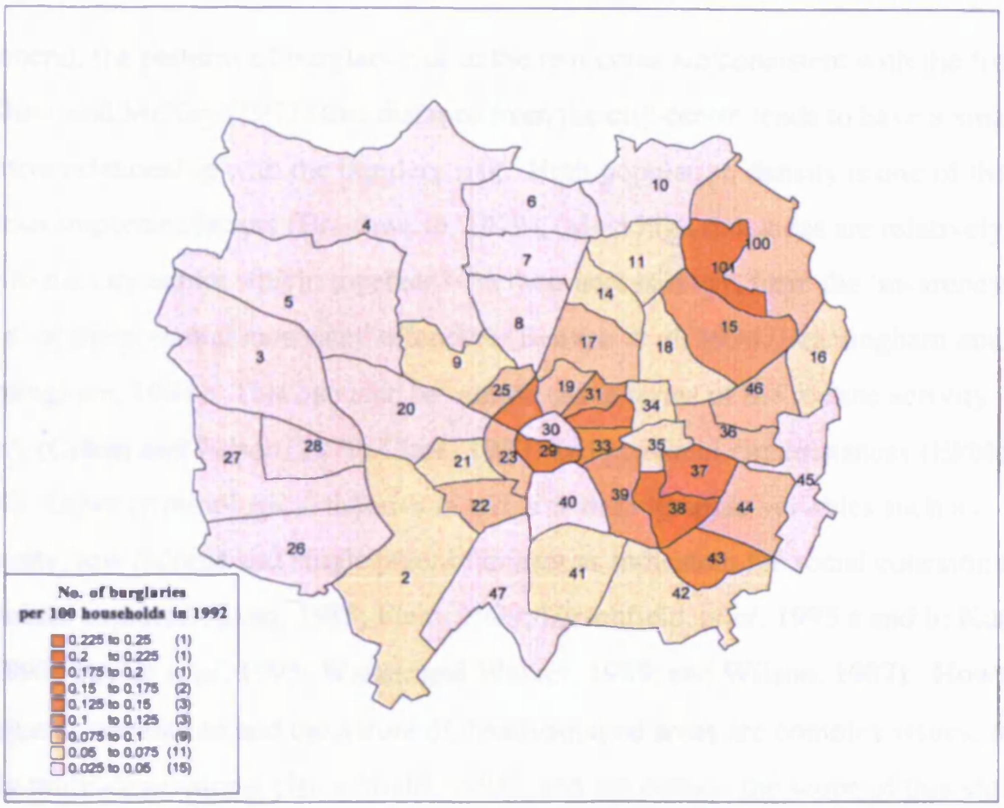


Figure 10. 5: Burglary risk in Coventry 1992 (overlay method)

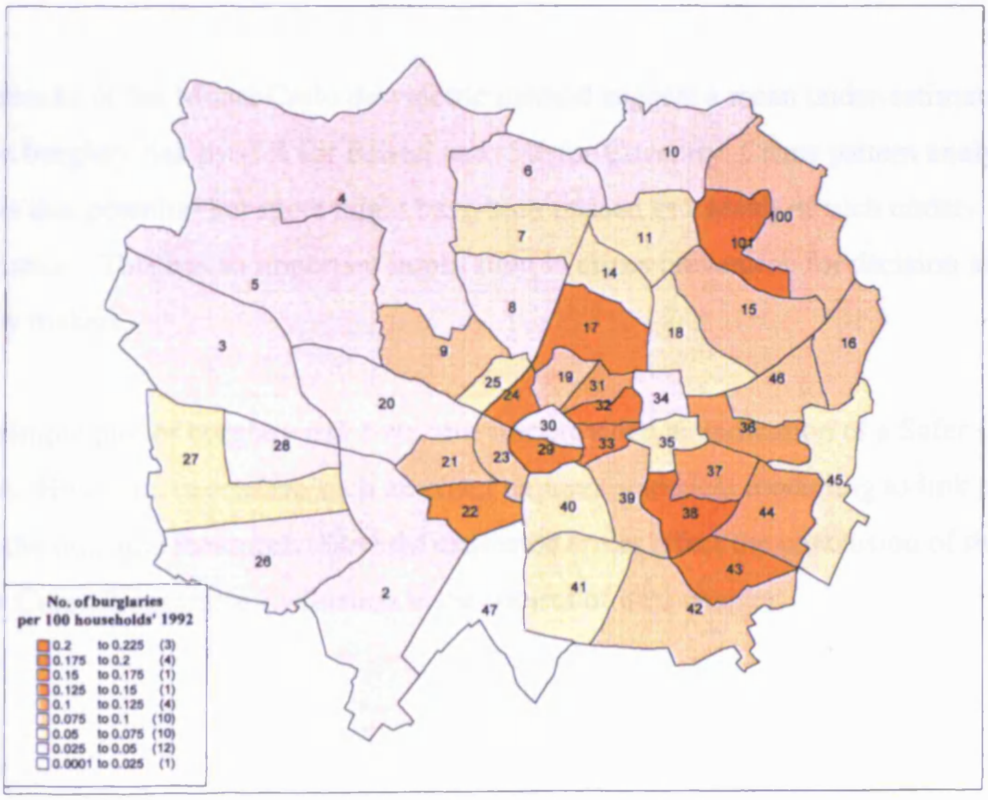


Figure 10. 6: Burglary risk in Coventry 1992 (Monte Carlo dasymetric method)

In general, the patterns of burglary risk in the two cities are consistent with the finding by Shaw and McKay (1972) that distance from the city centre tends to have a small negative relationship with the burglary risk. High population density is one of the obvious important factors (Braithwaite, 1979). Most high risk areas are relatively near to the city centre which, together with their accessibility, form the 'awareness space' of the potential motivated offenders (Beavon *et al*, 1994; Brantingham and Brantingham, 1984). This can also be interpreted in terms of the routine activity theory (Cohen and Felson, 1979; Clark, 1983) and proximal circumstances (Ekblom, 1994). Other criminological theories in terms of demographic variables such as ethnicity, low income and single parent families as indicators for social cohesion may be further explored (Evan, 1989, Field, 1989; Hirschfield, *et al*, 1995 a and b; Kurtz *et al*, 1995; Taylor *et al*, 1995; Walker and Walker, 1989; and Wilson, 1987). However the spatial correlation and the nature of disadvantaged areas are complex issues, most likely multi-dimensional (Hirschfield, 1994), and are outside the scope of this study.

10.4 Chapter Summary

The results of the Monte Carlo dasymetric method suggest a mean under-estimation of the burglary risk by -7% for Bristol and -5% for Coventry. Crime pattern analysis shows that potential hot spots might have been missed as a result of such under-estimation. This has an important implication in crime prevention for decision and policy makers.

The simple plot of burglary risk over time has provided an indication of a Safer Cities effect. However, to confirm such an effect requires statistical modelling to link action with the outcome measures. How the estimated errors affect the conclusion of the Safer Cities Programme Evaluation is the subject of next chapter.

Chapter Eleven
Process Error (IV): Impact upon the Safer Cities Programme Evaluation

As described in the previous chapters, to decide whether the spatial error has had any significant effect upon the data analyses and the subsequent conclusion of the Safer Cities Programme Evaluation, would require re-examining the process of data analyses in the Evaluation using multi-level modelling. This chapter maps the spatial errors upon the further data processing using multi-level modelling in the Evaluation of the Safer Cities Programme. This begins with examining the statistical model used in the Safer Cities Programme Evaluation for explaining variation in risk of burglary incidence as a function of Safer Cities action (Section 11.1). From this general multi-level model, a cut down version of the model is adopted for this case study to examine the impact of the spatial error upon the burglary risk-action intensity relationship (Section 11.2). Section 11.3 describes the initial results of the multi-level modelling. Section 11.4 further assesses how the spatial error influences the Evaluation when the burglary risk data sets were mathematically transformed to fit the statistical assumption of the multi-level modelling. Section 11.5 relates the results back to the context of the Safer Cities Programme Evaluation; and finally, the impact of the action intensity upon burglary risk is re-assessed in the light of this case study by means of significance testing.

11.1 The statistical model used in the Safer Cities Programme Evaluation for explaining variation in risk of burglary incidence

Multilevel models as explained in Chapter 5 (methodology) were used in the analysis of the recorded crime data and the Safer Cities effect. The aim of the data analyses for the Safer Cities Programme evaluation was to explore the link between the intensity (and presence) of Safer Cities action in any one beat in any one year, and the associated domestic burglary incidence risk. The objective of the multi-level modelling for the Evaluation was to measure special effects on burglary risk of the presence of the Safer Cities action, and its intensity whilst taking account of any background trends and differences in area context. The multi-levels consist of: beat-years (Level 1); beats (Level

2); and originally cities (Level 3). See Table 11.1. This was a 'repeated measures' model as there were up to six beat-year observations within each beat. Had the multi-level modelling not been used, the inter-connected spatial temporal relationships of the observations would have caused under-estimation of standard errors in ordinary least squares regression.

Table 11. 1: Recorded crime model parameter

Dependent variable: burglary incidence per household

Hierarchy Level	Unit of analysis	Safer Cities Number	Case Study Number	representation %
3	City	14	2	14
2	Beat/Superbeat	701	91	13
1	Beat-year (6 yrs)	3277	581	18

Two cities represent 13-14% of the total number of cities or beats for the evaluation. Note that the representation increases to 18% in the number of beat-years, and overall this indicates a good representation of the data from the case study. This is due to the nature of the two cities selected for this research (as described in Chapter 4). Not all beats were available for all years. For instance, although Coventry had a complete set of recorded crime data up to 1993, Bristol had some data missing in 1993.

A general multi-level model for the evaluation of the Safer Cities Programme can be expressed according to the *Variance components model* version of the equation (5.29) with the random term $u_{1j}x_{ij}$ missing. (Also see Chapter 5 for detailed methodology). In the original Safer Cities Programme Evaluation, city level was subsequently omitted from the model due to reduction (to zero) in the unexplained between-cities variance of the intercept when all contextual variables from 1991 census and index of local condition were included. The model was therefore simplified to two levels (beat-year and beat only). This made very little difference to the fixed coefficients or their significance in the original evaluation of the Safer Cities Programme. As a result, only the simple variance components model was used. (Random slope regression would have been required to see the variation across cities.) It follows that city level has also been excluded from the model in this case study. Two cities would be too small a number to constitute a level in this case study anyway. Equation (5.29) can thus be simplified as follows:

$$y_{ij} = \gamma_{00} + \gamma_{01} z_j + \gamma_{10} x_{ij} + \gamma_{11} z_j x_{ij} + (u_{0j} + e_{ij}) \qquad (11.1)$$

where

y_{ij} is	the burglary risk
γ_{00}	intercept
γ_{10}	coefficients of the action score (fixed)
x_{ij}	action score
γ_{01}	coefficient of other explanatory variable e.g. presence of action
z_j	other explanatory variable e.g. action presence [1 or 0]
$\gamma_{11} z_j x_{ij}$	covariates (e.g. census x action scores x action presence)
$(u_{0j} + e_{ij})$	residuals of levels 2 and 1 respectively.

11.2 The statistical model used in the case study for explaining variation in risk of burglary incidence and action intensity

The aim of the analysis here is to assess the extent of the impact of the spatial error has upon the conclusion of the Safer Cities Programme Evaluation. To include all the census variables used in the Evaluation for the multi-level analyses is beyond the scope of this case study, though it is feasible to do so. Since a beat consists of a number of Enumeration Districts (ED), the error due to the spatial interpolation would have had very little impact upon the presence of any Safer Cities action (including other actions) in a beat. So the explanatory variables: action presence and all the census variables are removed from the multi-level model. The analysis starts with the simplest version of the variance components model in a form of $y_{ij} = \beta_{0j} + \beta_1 x_{ij} + u_j + e_{ij}$ as described in (5.21).

Explanatory variables

fixed components

A positive coefficient estimate for a fixed component indicates that the variable is associated with an increased risk of burglary victimisation; a negative estimate with a reduced risk.

Random components

The level two random component of the constant is the residual variation (i.e. that which is unexplained by the fixed effects) between beats in the burglary incidence risk. The residual variation for each beat is the average unexplained risk common to all beat-years for that beat. By definition, it does not change over the six-year period of measurement.

The level one random component of the constant is the residual unexplained variance of risk between beat-years, having taken account of the variation between beats.

ML3 associates every parameter with an explanatory variable, the model (5.25) thus represents

$$\mathbf{RISK} = \beta_{0j}\ \mathbf{CONS} + \beta_1\ \mathbf{ACTION} + u_j\mathbf{CONS} + e_{ij}\mathbf{CONS} \qquad \textbf{(11. 2)}$$

(where CONS = 1.)

β_{0j} CONS + β_1 ACTION are the fixed part; u_jCONS + e_{ij}CONS the random part of the model.

So far the spatial error has been analysed on a city by city basis. However, since the multi-level modelling includes all cities in a single model, it is appropriate to combine the data of the two cities in this case study for multi-level modelling in order to compare the results with the Safer Cities Programme as a whole. Following the procedure of the Evaluation to generate estimates of the impact of Safer Cities action, we need to produce a model 1) with action score = 0 [a base line]; and 2) compare with action score = average action intensity.

When the explanatory variable ACTION is excluded in the multi-level model (or action score = 0), equation (5.4) becomes:

$$y_{ij} = \beta_{0j} + u_j + e_{ij} \qquad \textbf{(11. 3)}$$

i.e. (in ML3 terms)

$$RISK = \beta_0\, CONS + u_j\, CONS + e_{ij}\, CONS \qquad (11.4)$$

Specifying the multi-level model according to equations (11.2) & (11.4), a whole range of the results of the burglary risk using the results of the overlay method (RISKov) and of the Monte Carlo Dasymetric method (RISKmc) can be computed for each of the three cases: Coventry; Bristol; and the two cities combined. (See Appendix 11.1 for the detailed log of ML3 implementation.)

11.3 Initial results of multi-level modelling

When the explanatory variable ACTION is excluded in the multi-level model:

For Coventry

RISKov	= 0.05659 ± 0.00532
RISKmc	= 0.05913 ± 0.00504
Estimated error	= RISKov – RISKmc
	= -0.00255 (4.31%)

For Bristol

RISKov	= 0.04435 ± 0.00592
RISKmc	= 0.04179 ± 0.00629
Estimated error	= 0.00256 (6.13%)

When two cities combined

RISKov = 0.05067 ± 0.00402
RISKmc = 0.05201 ± 0.00447
Estimated error = -0.00135 (2.60%)

In all cases so far, the estimated errors due to the overlay method in all cases are well within the range of the standard errors.

When ACTION ($\beta_1\, x_{ij}$) is included in the multi-level model, the equation 11.2 can be worked out from the fixed part of the model, and the residuals and the standard errors from the random part.

For Coventry

RISKov = -0.00064 (± 0.00135) ACTION + 0.05679 ± 0.00537

RISKmc = -0.00074 (± 0.00120) ACTION + 0.05940 ± 0.00512

Since the cut-down model has not taken into account the time trend, external comparison crime series, and all the census variables, one would expect the risk-action relationship would be very small. If any, it might even be a positive correlation, as the preventive Programme tended to take place in the high crime areas. So it is surprising to see a negative coefficient from the ACTION variable in Coventry. Figure 11.1 plots the above equations for both overlay and Monte Carlo methods.

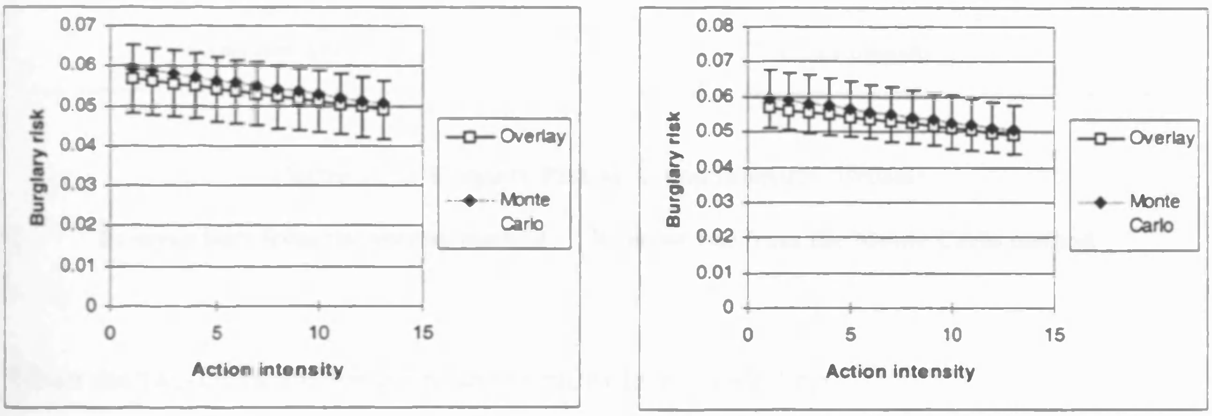


Figure 11. 1: Burglary Risk vs Action Intensity (Coventry)

L: error bars from the overlay method   R: error bar from the Monte Carlo method

The above Figure also illustrates that the equation generated from Monte Carlo data sets falls within the standard errors of the overlay data sets and vice versa.

For Bristol

RISKov = 0.00211 (± 0.00107) ACTION + 0.04364 ± 0.00587

RISKmc = 0.00221 (± 0.00113) ACTION + 0.04104 ± 0.00623

The above equations show a positive correlation between burglary risk and the preventive action. This is as expected and explained earlier, that is, the prevalence of crime and all the census variables have not been accounted for. The important observation is that the outcome from the Monte Carlo data sets is well within the standard error range of the equation from the overlay method. Figure 11.2 shows the above relationship between RISK and ACTION in Bristol.

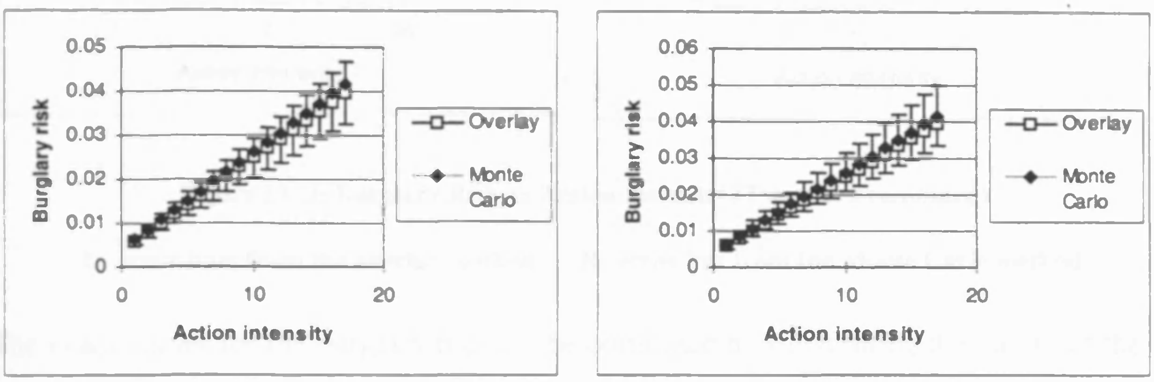


Figure 11. 2: Burglary Risk vs Action Intensity (Bristol)

L: error bars from the overlay method R: error bar from the Monte Carlo method

When the two cities are combined in the multi-level modelling

RISKov = 0.00108 ($\pm$ 0.00084) ACTION + 0.05031 $\pm$ 0.0040

RISKmc = 0.00063 ($\pm$ 0.00088) ACTION + 0.05180 $\pm$ 0.00446

The coefficient of the ACTION shows a very small value in comparison with the overall standard error. This indicates that the effect of the action in relationship to risk is very small. The positive and negative effects of the action in Bristol and Coventry seem to have cancelled each other out when put together. Figure 11.3 derived from the above equations shows this effect. The graph indicates that there is a small interaction between the two methods.

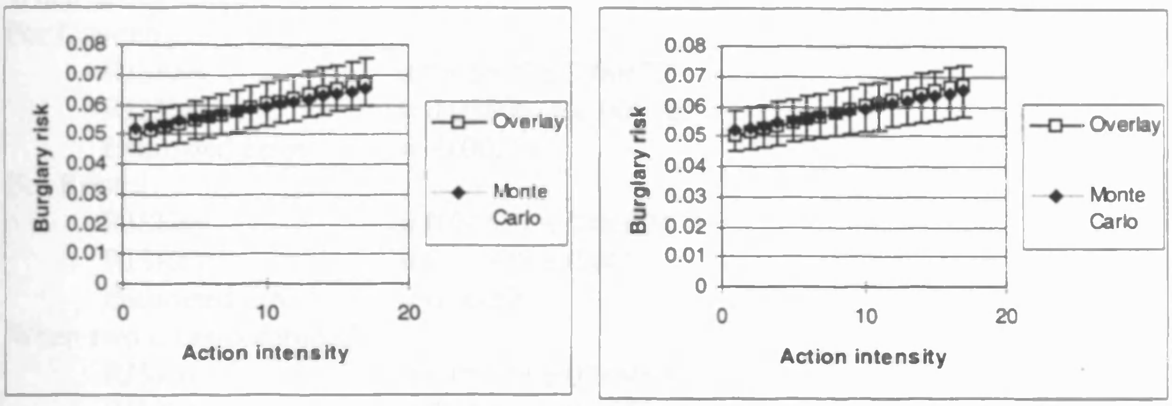


Figure 11. 3: Burglary Risk vs Action Intensity (Two cities combined)

L: error bars from the overlay method R: error bar from the Monte Carlo method

The exact values for the burglary risk can be computed by substituting the values of the action scores into the above equations for each case. The most meaningful values would be when action scores are equal to 0, 1, mean, and maximum. This covers the whole range of values and enables one to examine the impact of the estimated error upon the multi-level model in each case.

When action score = 0
For Coventry

RISKov	= 0.05679 ± 0.00537
RISKmc	= 0.05396 ± 0.00512
Estimated error	= -0.00260

For Bristol

RISKov	= 0.04364 ± 0.00587
RISKmc	= 0.04104 ± 0.00623
Estimated error	= 0.00260

When two cities combined

RISKov	= 0.05031 ± 0.0040
RISKmc	= 0.05180 ± 0.00446
Estimated error	=-0.00149

The above results are more or less the same as when the variable ACTION was excluded from the model as one would have expected when ACTION is equal to zero.

When action score = 1
For Coventry

RISKov	= 0.05615 ± 0.00672
RISKmc	= 0.05866 ± 0.00632
Estimated error	= -0.00251

For Bristol

RISKov	= 0.04575 ± 0.00694
RISKmc	= 0.04325 ± 0.00736
Estimated error	= 0.00250

When two cities combined

RISKov	= 0.05139 ± 0.00483
RISKmc	= 0.05243 ± 0.00533
Estimated error	= -0.00104

When action score = Mean
For Coventry, the mean action score = 2.48

RISKov	= 0.05520 ± 0.00872
RISKmc	= 0.05756 ± 0.00810
Estimated error	= -0.00237

For Bristol, the mean action score = 3.18,

RISKov	= 0.05035 ± 0.00928
RISKmc	= 0.04807 ± 0.00981
Estimated error	= 0.00228

When two cities combined, the mean action score = 2.79

RISKov	= 0.05333 ± 0.00633
RISKmc	= 0.05355 ± 0.00690
Estimated error	= -0.00023

When action score = Maximum
For Coventry, the maximum action score = 11.47

RISKov	= 0.04940 ± 0.02086
RISKmc	= 0.05090 ± 0.01890
Estimated error	= -0.00150

For Bristol, the maximum action score = 15.22

RISKov	= 0.07575 ± 0.02220
RISKmc	= 0.07469 ± 0.02338
Estimated error	= 0.00106

When two cities combined, the maximum action score = 15.22

RISKov	= 0.06675 ± 0.01676
RISKmc	= 0.061350872 ± 0.01779
Estimated error	= 0.00540

The above results show that in all cases, the estimated errors from the Monte Carlo method are well within the range of the standard error of the model generated by the overlay method. The results are summarised in Table 11.2 and their numerical expressions can be described as follows:

<u>Coventry</u>

Zero:	0.00537 > -0.00260 > -0.00537
One:	0.00672 > -0.00251 > -0.00672
Mean:	0.00872 > -0.00237 > -0.00872
Maximum:	0.02086 > -0.00150 > -0.02086

<u>Bristol</u>

zero:	0.00587 > 0.00260 > -0.00587
One:	0.00694 > 0.00250 > -0.00694
Mean:	0.00928 > 0.00228 > - 0.00928
Maximum:	0.02220 > 0.00106 > - 0.02220

<u>Two cities</u>

Zero:	0.0040 > -0.00149 > -0.0040
One:	0.00483 > -0.00104 > -0.00483
Mean:	0.00633 > -0.00023 > -0.00633
Maximum:	0.01676 > -0.00540 > -0.01676

Table 11. 2: Summary of values of Risk at different Action scores (MC - Monte Carlo; OV - Overlay; S - Standard; E - Estimated)

City	Action	Risk(OV)	Risk(MC)	S.ERR(OV)	S.ERR(MC)	E. Error	E. Error %
Coventry	-	0.05659	0.05913	0.00532	0.00504	-0.00255	-4.31
	0	0.05679	0.05396	0.00537	-0.00537	-0.0026	-4.82
	1	0.05615	0.05866	0.00672	-0.00632	-0.00251	-4.28
	2.48	0.0552	0.05756	0.00872	-0.0081	-0.00237	-4.12
	11.47	0.04941	0.0509	0.02086	0.0189	-0.0015	-2.95
Bristol	-	0.04435	0.04179	0.00592	0.00629	0.00255	6.10
	0	0.04364	0.04104	0.00587	0.00623	0.0026	6.34
	1	0.04575	0.04325	0.00694	0.00736	0.0025	5.78
	3.18	0.05035	0.04807	0.00928	0.00981	0.00228	4.74
	15.22	0.07575	0.07469	0.0222	0.02338	0.00106	1.42
Two Cities	-	0.05066	0.05201	0.00402	0.00447	-0.00135	-2.60
	0	0.05031	0.0518	0.00399	0.00446	-0.00149	-2.88
	1	0.05139	0.05243	0.00483	0.00533	-0.00103	-1.96
	2.79	0.05333	0.05355	0.00633	0.0069	-0.00022	-0.41
	15.22	0.06675	0.06135	0.01676	0.01779	0.0054	8.80

This represents the estimated errors (at the mean action scores): –4% for Coventry, 4% for Bristol and 0.4% when the two cities combined. However as action scores increases to the maximum, the estimated error increase to the maximum 8.8% when the two cities combined while the estimated error decreases to –3% and 1.4% for Coventry and Bristol respectively. This indicates that the errors of the two cities tend to interact in the combined model and change in an unpredictable way.

11.4 The transformed risk of burglary incidence used in the Safer Cities Programme Evaluation

In the original Safer Cities Programme Evaluation, the burglary risk (the so-called outcome measure or dependent variable in the multi-level equation) had to be transformed in order to remove negative skew in the frequency distribution. Examination of the histograms of the burglary risk data from the two cities show that this indeed was the case (see Figures 11.4 & 11.5).

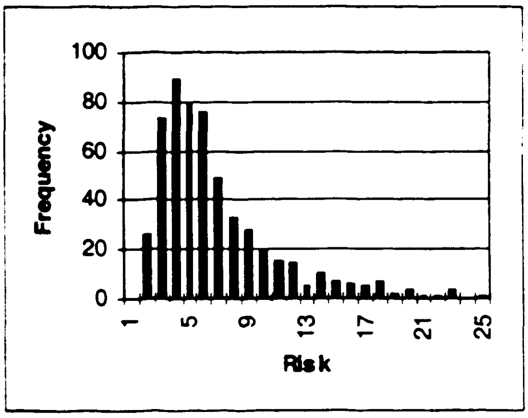


Figure 11. 4: Burglary Risk (two cities)

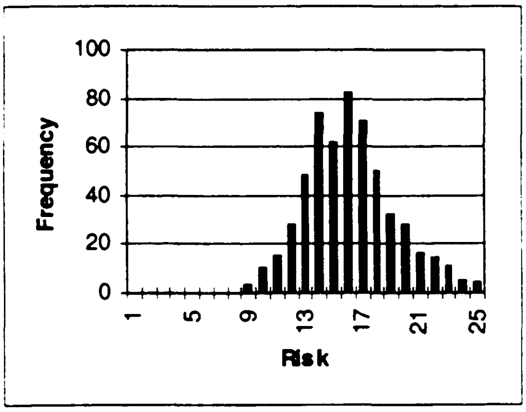


Figure 11. 5: Transformed Burglary Risk (two cities)

In such cases, burglary incidence risk tends to be skewed so that errors associated with the lower risk tend to be larger than the errors associated with the higher risk. So the burglary risks by the overlay method and the Monte Carlo Dasymetric estimation were transformed in the same way as those in the original Safer Cities Programme Evaluation, that is, the coefficient estimate for a fixed component adjusts the predicted transformed risk: {Arcsine [(burglary recorded incidence risk per household) $^{1/4}$]}. (Note normal distribution is a necessary requirement of the data sets used in the multi-level modelling due to its statistical assumptions as outlined in Chapter 5.)

The multi-level model $y_{ij} = \beta_{0j} + \beta_1 x_{ij} + u_j + e_{ij}$ remains the same as before except now y_{ij} represents the transformed burglary risk: [Arcsine ((Burglary rate/Household)$^{1/4}$)].

The whole range of the multi-level modelling results for the transformed risk (RISK') can be computed following the exact procedures as before.

When ACTION is excluded in the multi-level model.

For Coventry

RISK'ov	= 0.49199 ± 0.01320
RISK'mc	= 0.49905 ± 0.01314
Estimated error	= -0.00706

For Bristol,

RISK'ov	= 0.42061 ± 0.03872
RISK'mc	= 0.40999 ± 0.04017
Estimated error	= 0.01062

When two cities combined

RISK'ov	= 0.45953 ± 0.02061
RISK'mc	= 0.45989 ± 0.02198
Estimated error	= -0.00036

When action is included in the multilevel model

For Coventry

RISK'ov = -0.00368 (± 0.00318) ACTION + 0.49319 ± 0.01349

RISK'mc = -0.00365 (± 0.00282) ACTION + 0.50049 ± 0.01348

The transformed results are very similar to the raw data of the burglary risk except the values appear to be accentuated by about eight fold. Figure 11.6 shows that the risk-action relationship has a similar trend as before.

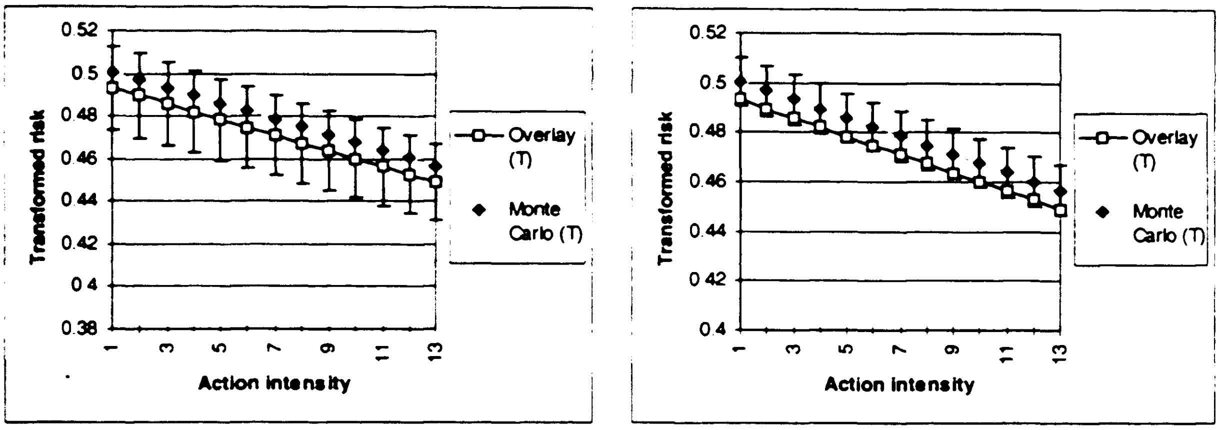


Figure 11. 6: Transformed Burglary Risk vs Action Intensity (Coventry)

L: error bars from the overlay method R: error bar from the Monte Carlo method

Bristol

RISK'ov = 0.00444 (± 0.00267) ACTION + 0.41879 ± 0.03850

RISK'mc = 0.00471 (± 0.00271) ACTION + 0.40803 ± 0.03992

The trend is the same as before except the slope is much 'gentler' in comparison with the above. This is due to the large scale of the residuals (Figure 11.7).

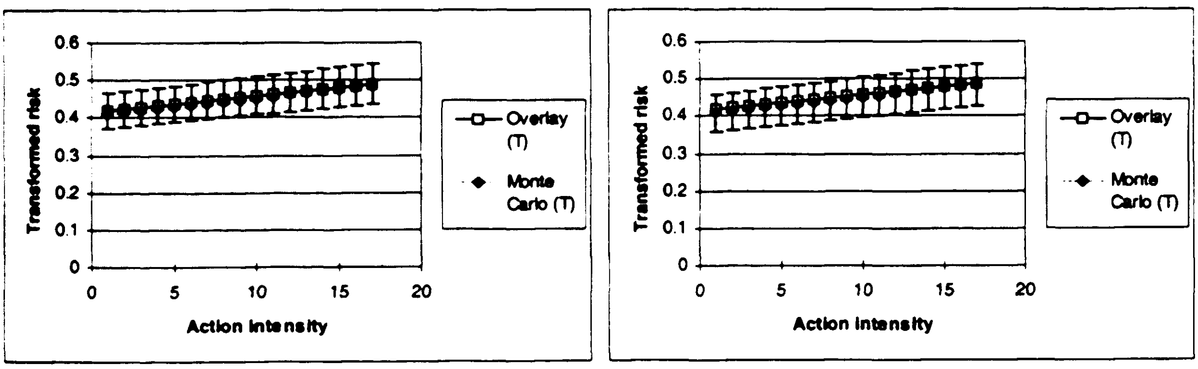


Figure 11. 7: Transformed Burglary Risk vs Action Intensity (Bristol)

L: error bars from the overlay method R: error bar from the Monte Carlo method

When *two cities* are combined in the model
RISK'ov = 0.00091 (± 0.00204) ACTION + 0.45916 ± 0.02059
RISK'mc = -0.00066 (± 0.00217) ACTION + 0.46014 ± 0.02202

Figure 11.8 shows the most interesting contrast between the coefficients of the overlay action and the Monte Carlo action scores. They appear to be in an exact opposite trend though still within the standard error of the overlay method. The graph represents the tail end of the interaction (as indicated before), and here the difference between the two methods has been accentuated by the transformation to such an extent that their coefficients have gone in opposite directions. This implies the coefficients are extremely unstable and the relationship is unlikely to be significant (see significance testing later).

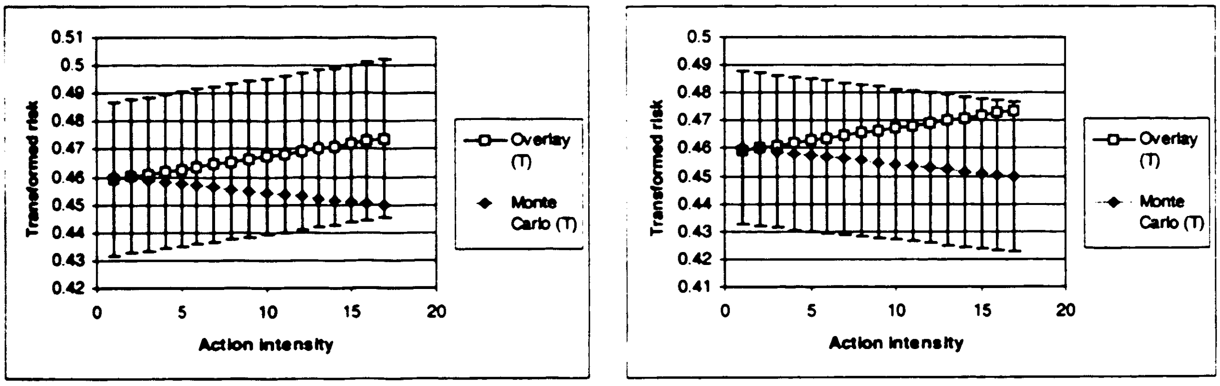


Figure 11. 8: Transformed Burglary Risk vs Action Intensity (Two cities combined)

L: error bars from the overlay method R: error bar from the Monte Carlo method

As in Section 11.2, a whole range of values for the transformed burglary risk can be computed by substituting the values of the action score into the above equations when action scores are equal to 0, 1, mean, and maximum.

When action score = 0
For Coventry

RISK'ov	= 0.49319 ± 0.01349
RISK'mc	= 0.50049 ± 0.01348
Estimated error	= -0.00730

For Bristol
RISK'ov = 0.41879 ± 0.03850
RISK'mc = 0.40803 ± 0.03992
Estimated error = 0.01076
When two cities combined
RISKov = 0.45916 ± 0.02059
RISKmc = 0.46014 ± 0.02202
Estimated error = -0.00098

When action score = 1
For Coventry
RISK'ov = 0.48950 ± 0.01666
RISK'mc = 0.49684 ± 0.0163
Estimated error = -0.00733
For Bristol
RISKov = 0.42323 ± 0.04117
RISKmc = 0.41274 ± 0.04262
Estimated error = 0.01049
When two cities combined
RISK'ov = 0.46007 ± 0.02263
RISK'mc = 0.45949 ± 0.02419
Estimated error = 0.00058

When action score = Mean
For Coventry. The mean action score = 2.48
RISK'ov = 0.48405 ± 0.02137
RISK'mc = 0.49143 ± 0.02047
Estimated error = -0.00738
For Bristol, the mean action score = 3.18
RISK'ov = 0.43291 ± 0.04699
RISK'mc = 0.4230 ± 0.04852
Estimated error = 0.00991

When two cities combined, the mean action score = 2.79
RISK'ov = 0.46169 ± 0.02629
RISK'mc = 0.45831 ± 0.02807
Estimated error = 0.00338

When action score = Maximum
For Coventry. The maximum action score = 11.47
RISK'ov = 0.45095 ± 0.04994
RISK'mc = 0.45862 ± 0.04579
Estimated error = 0.00767

For Bristol, the maximum action score = 15.22

RISK'ov	= 0.48635 ± 0.07911
RISK'mc	= 0.47967 ± 0.08109
Estimated error	= 0.00668

When two cities combined, the maximum action score = 15.22

RISK'ov	= 0.47294 ± 0.05170
RISK'mc	= 0.45017 ± 0.05503
Estimated error	= 0.02278

The above results show that in all cases, as before the RISK has been transformed, the estimated errors from the Monte Carlo method are within the range of the standard errors generated by the overlay method. Again, the summary of the relationship between the estimated errors and the standard errors can be expressed as follows:

<u>Coventry</u>

Zero:	0.01349 > -0.00730 > -0.01349
One:	0.01666 > -0.00733 > -0.01666
Mean:	0.02137 > -0.00738 > -0.02137
Maximum:	0.04993 > -0.00767 > -0.04993

<u>Bristol</u>

Zero:	0.03850 > 0.01076 > -0.03850
One:	0.04117 > 0.01049 > -0.04117
Mean:	0.04699 > 0.00991 > -0.04699
Maximum:	0.07911 > 0.00668 > -0.07911

<u>Two cities</u>

Zero:	0.02059 > 0.00098 > -0.02059
One:	0.02263 > 0.00058 > -0.02263
Mean:	0.02629 > 0.00338 > -0.02629
Maximum:	0.05170 > 0.02278 > -0.05170

The summary of the transformed risk is shown in Table 11.3.

Table 11. 3: Summary of values of the transformed Risk at different Action scores (MC - Monte Carlo; OV - Overlay; S - Standard; E - Estimated)

City	Action	Risk'(OV)	Risk'(MC)	S.ERR(OV)	S.ERR(MC)	E. Error	E. Error %
Coventry -		0.49199	0.49905	0.0132	0.13141	-0.00706	-1.41
	0	0.49319	0.50049	0.01349	0.01348	-0.0073	-1.46
	1	0.48950	0.49684	0.01666	0.0163	-0.00733	-1.48
	2.48	0.48406	0.49143	0.02137	0.02047	-0.00738	-1.50
	11.47	0.45096	0.45862	0.04993	0.04579	-0.00767	-1.67
Bristol -		0.42060	0.40999	0.03872	0.04017	0.01062	2.59
	0	0.41879	0.40803	0.0385	0.03992	0.01076	2.64
	1	0.42323	0.41274	0.04117	0.04262	0.01049	2.54
	3.18	0.43291	0.423	0.04699	0.04852	0.00991	2.34
	15.22	0.48635	0.47967	0.07911	0.08109	0.00668	1.39
Two cities -		0.45953	0.45989	0.02061	0.02198	-0.00036	-0.08
	0	0.45916	0.46014	0.02059	0.02202	-0.00098	-0.21
	1	0.46007	0.45949	0.02263	0.02419	0.00058	0.13
	2.79	0.46169	0.45831	0.02629	0.02807	0.00338	0.74
	15.22	0.47294	0.45017	0.0517	0.05503	0.02278	5.06

This represents the estimated errors (at the mean action scores): –1.5%, 2.3% and 0.7% for Coventry, Bristol, and the two cities combined respectively. For the two cities combined, the overall effect of the underestimation of the action intensity and burglary risk seemed to cancel each other out. However its estimated error increases to the maximum 5% when action scores increase to the maximum.

11.5 Relating the results to the scale of the Safer Cities Programme Evaluation

Unlike the model used in the Safer Cities Programme evaluation, the multi-level model used in this case study has merely attempted to explain the variation in burglary risk in the two Safer Cities beats alone, without using the other indicators as additional explanatory factors. The original model used in the Evaluation included other factors:

At the **beat** level:

- *Geographical factors* such as the size of the beat, the household density, whether it had a city-centre location (since domestic burglary rates are likely to differ in areas comprising mostly shops, offices, transport and entertainment facilities)

- *Social factors* derived directly from the 1991 Census, such as the proportion of the population aged 16-24, the proportion aged 60 and over, and the proportion of households lacking a car. Other factors from the Index of Urban Conditions were also included, such as the overall Index itself, and subsidiary indicators including overcrowding, and children in unacceptable accommodation
- *Measurement factors* which could have introduced bias - whether or not a 'beat' was a superbeat; and whether or not we had obtained burglary data in the beat for all six years ('incomplete' beats may have been areas with special problems or unusual patterns of policing)

At the **beat-year** level:

- *The year* (to indicate the overall trend in burglary)
- *Comparison indicators* for burglary trend, both global and based on city Census-family
- *The amount of* ***other*** *Safer Cities action*, not targeted on burglary, that was present

(Ekblom *et al* 1996, p 49, italic added).

The analysis of the results so far has found very little impact of the estimated error upon the results of the overlay methods used in the evaluation of the Safer Programme. The next question is: How do the results relate to the results of the Safer Cities Programme Evaluation as a whole? To answer such a question, we need to compare the results of the case study with those of the Safer Cities Programme. Table 11.4 provides such a comparison.

Table 11. 4: comparison of the estimates and standard errors (S.ERROR) between the Evaluation and this case study.

Fixed components (CONS)	**ESTIMATE**	**S.ERROR**
Coventry (Overlay)	0.4839	0.01187
Coventry (Monte Carlo)	0.4911	0.01185
Bristol (Overlay)	0.3759	0.02976
Bristol (Monte Carlo)	0.3628	0.03066
Two Cities (Overlay)	0.4310	0.01665
Two cities (Monte Carlo)	0.4284	0.01759
All Safer Cities	-5.5680	1.04500
Fixed components (ACTION)	**ESTIMATE**	**S.ERROR**
Coventry (Overlay)	-0.00368	0.00318
Coventry (Monte Carlo)	-0.00365	0.00282
Bristol (Overlay)	0.00444	0.00267
Bristol (Monte Carlo)	0.00471	0.00271
Two Cities (Overlay)	0.00091	0.00204
Two cities (Monte Carlo)	-0.00066	0.00217
All Safer Cities	-0.01298	0.00689
Random component Level 2	**ESTIMATE**	**S.ERROR**
Coventry (Overlay)	0.00596	0.00133
Coventry (Monte Carlo)	0.00592	0.00134
Bristol (Overlay)	0.03902	0.00837
Bristol (Monte Carlo)	0.04150	0.00890
Two Cities (Overlay)	0.02455	0.00371
Two cities (Monte Carlo)	0.02734	0.00415
All Safer Cities	0.00486	0.00030
Random component Level 1	**ESTIMATE**	**S.ERROR**
Coventry (Overlay)	0.00332	0.00028
Coventry (Monte Carlo)	0.00347	0.00030
Bristol (Overlay)	0.00387	0.00037
Bristol (Monte Carlo)	0.00373	0.00036
Two Cities (Overlay)	0.00361	0.00023
Two cities (Monte Carlo)	0.00440	0.00028
All Safer Cities	0.00338	0.00009

The comparison shows that the intercept of the risk-action equation and its standard error in the Evaluation, is far larger than those in the case study (more than ten times, see Figures 11.9 & 11.10 with both overlay and Monte Carlo methods included respectively for the two cities). The value of the intercept in the Safer Cities Programme Evaluation has a negative value in contrast to the case study. This is because other contextual factors especially the time trend and crime comparison indicators have been taken into account in the Evaluation as explained earlier.

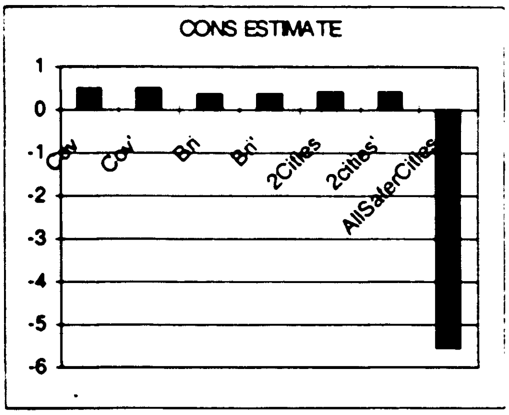


Figure 11. 9: Estimates of the intercept

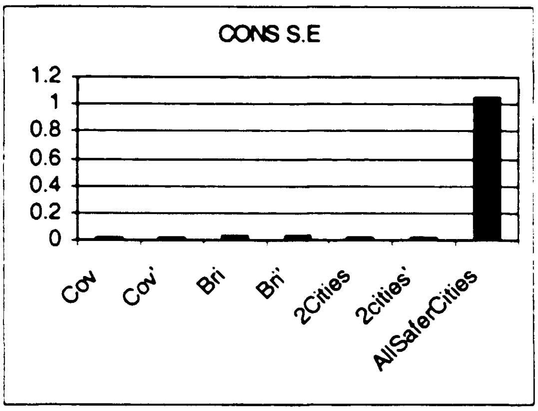


Figure 11. 10: Standard Error of the intercept

In contrast, the coefficients of the action intensity in the case study are comparable to the coefficient of the Safer Cities action intensity. In particular Coventry has the same trend as the Safer Cities Programme as a whole. Bristol appears to have an opposite trend to Coventry. As a result, the effects of the two seem to cancel each other when combined. The values of the coefficient and the standard error of the Safer Cities Programme are larger (but only about twice to three times) than those in the case study. This is expected as the Programme as a whole consists of more cities, more cases and hence has a larger coefficient and a larger standard error (Figures 11.11 & 11.12).

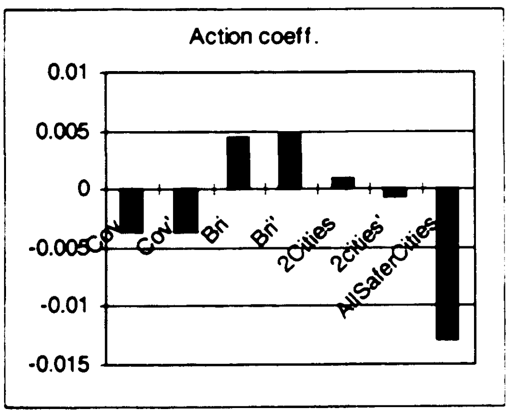


Figure 11. 11: Coefficients of ACTION

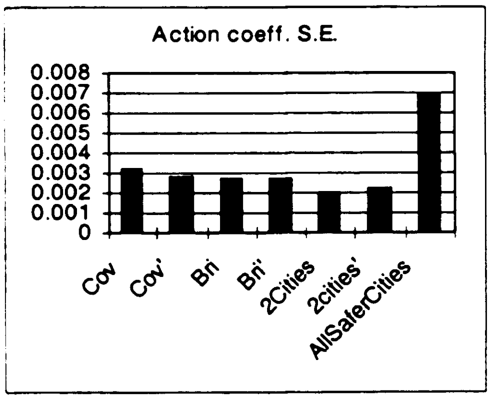


Figure 11. 12: Standard Errors of ACTION

In striking contrast to the fixed components, the level two random components of the Safer Cities Programme consist of a much smaller estimate and standard error than the case study. This indicates that the model used in the Evaluation as a whole is a more 'powerful' test than the model used in the case study as the former left a relatively small amount of unexplained residuals (between beats) and hence, has a smaller standard error (Figures 11.13 & 11.14).

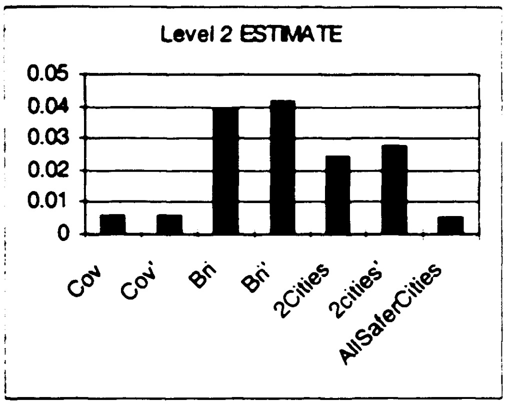


Figure 11. 13: Level 2 estimates

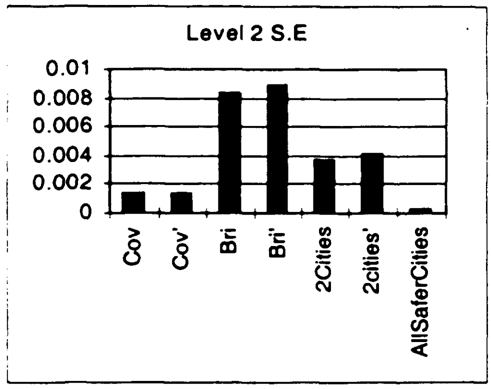


Figure 11. 14: Standard Errors of level 2

Level one estimates of the Safer Cities Programme Evaluation are comparable to those of the case study (Figure 11.15). Nevertheless the standard error of the evaluation as a whole is still consistently smaller than those of the case study (Figure 11.16).

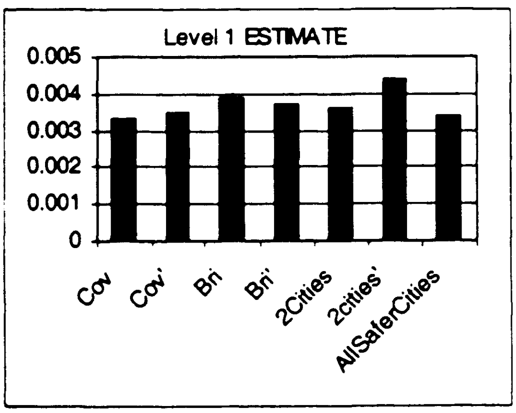


Figure 11. 15: Level 1 estimates

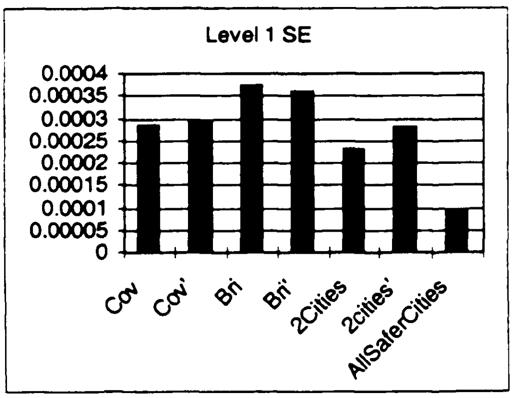


Figure 11. 16: Standard Errors of Level 1

To conclude, the above implies that the impact of the estimated error upon the Safer cities Programme Evaluation is very small. To start, the Evaluation consists of a large negative intercept. To relate the results of the case study to the scale of the evaluation, the multi-level equations of the case study can be scaled up by taking the relative values of the intercepts of the evaluation. This produces Figure 11.17 which shows that the results of the case study are all within the range of the standard error of the evaluation. This is because the evaluation, as a whole, consists of a large standard error in its fixed part of the model, that is, the inferred relationship between the action intensity and burglary risk reduction.

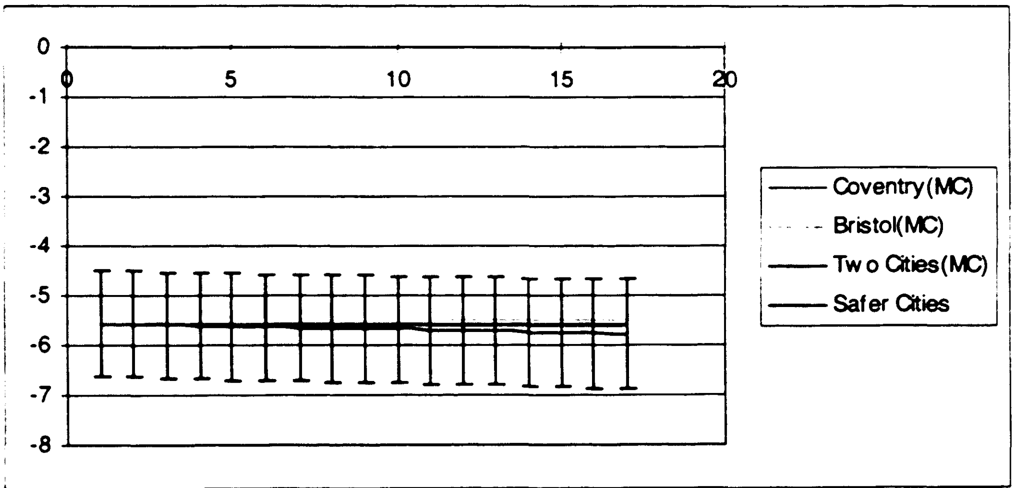


Figure 11. 17: Comparison between the overall Safer Cities effect with the Two Cities effect

Error bars are from the Safer Cities

11.6 Significance testing: Re-examining Safer Cities Programme impact on burglary

The final question to be answered in this research is: how statistically significant are the results of this case study in relation to the evaluation of the Safer Programme Evaluation? We need to examine the significance testing used in the Safer Cities Programme Evaluation. The goodness of fit statistic, or likelihood, in multi-level modelling measures the model that predicts the observed values of burglary risk over all explanatory variables. In the Safer Cities Programme Evaluation model, the mere presence of the action (the so called step effect) on a particular beat-year accounts for one third of the reduction in risk ($p = 0.027$). Other Safer Cities action which was included in the main statistical model, also had a role to play in the burglary risk reduction, but has not been included in this case study.

As described in the previous Chapter, the trend of the burglary risk is somewhat different from the national trend, apart from a small drop in burglary rate in 1988-1989, the burglary rate continued to rise steeply and a deep fall followed shortly after the implementation of Safer Cities Programme (1992-1993). While this trend shows some *prima facie* evidence of a Safer Cities effect, relating the action intensity to the burglary risk for the whole six year period would have masked most of the Safer Cities effect.

Action tended to be located in beats with a higher risk of burglary, so one would expect a positive correlation between burglary risk and action intensity if the time trend and other contextual variables were not taken into account. The effect of the action intensity upon the risk reduction, if any, is expected to be moderately small and to vary from city to city. The effect of the action intensity (given the step effect) found in the Safer Cities Programme as a whole was actually very weak ($p = 0.108$). The evaluation team decided to reject the null hypothesis as a 'borderline' case only on the basis of other related factors, in particular, the significant step effect found in the mere presence of the action ($p = 0.027$). The step and marginal-intensity effects together were found to be jointly significant at $p = 0.01$. Had there been no other contextual evidence, the null hypothesis would have been accepted in the normal academic practice.

Table 11.5 shows the comparison of the results of the significance testing in from this case study in comparison with the Safer Cities Programme Evaluation for all cities. These are calculated by first calculating the likelihood of each model (for example, Coventry, -792.508 with Monte Carlo dasymetric method and –790.895 when the explanatory variable ACTION is removed from the model). Under the null hypothesis of zero parameter values (variance), the difference of likelihoods (1.613) follows a chi-square distribution with degree of freedom equal to the number of parameters removed, in this case 1. The fit statistics are calculated as -2[log (likelihood ratio)]. (see Appendix 11.2 for the full significance testing).

Table 11. 5: Significance testing

Significant testing	P	df
Coventry (Overlay)	0.25440	1
Coventry (Monte Carlo)	0.20407	1
Bristol (Overlay)	0.09701	1
Bristol (Monte Carlo)	0.08270	1
Two Cities (Overlay)	0.66292	1
Two cities (Monte Carlo)	0.76418	1
All Safer Cities	0.10800	1

The causal relationship between risk and action intensity found in the case study can be concluded to be insignificant in every case, though the values for the individual city are comparable to the conclusion of the Safer Cities Programme evaluation.

11.5 Conclusion of the Chapter

This case study has found that the errors due to the spatial interpolation estimated by the Monte Carlo method are well within the standard errors generated by the overlay method in the multi-level modelling (mean: 0.02136536 > -0.00738 > -0.02136536 for Coventry; 0.04698854 > 0.009906 > -0.04698854 Bristol; 0.02629 > 0.00338 > -0.02629 when two cities combined. This represents the estimated errors (at the mean action scores): –1.5%, 2.3% and 0.7% for Coventry, Bristol, and the two cities combined respectively. Multi-level modelling has shown that the risk-action relationship between the overlay and Monte Carlo Dasymetric methods tend to interact in the multi-level modelling such that the overall effect of the underestimation of the action intensity and burglary risk seemed to cancel each other out. However as action scores increases to the maximum, the estimated error increase to the maximum 5% when the two cities combined. Taking the spatial error into account (i.e. using Monte Carlo Dasymetric method) the difference between the two method (if any) tended to enhance the action intensity effect, though such an effect was too small to be significant ($p = 0.2$ for Coventry; 0.08 for Bristol; and 0.76 when two cities combined; c.f. $p = 0.108$ for all Safer Cities; with d.f. = 1 in all cases). **It is concluded that no significant impact due to spatial error upon the conclusion of the Safer Cities Programme Evaluation has been found.**

The analyses also show that the Safer Cities effect varies across different cities as exemplified by Bristol and Coventry. While it is justified for the Evaluation to combine all the cities in the multi-level model to assess the impact of the Safer Cities Programme as a whole (as required by the Treasury), analysing the effect city by city would have enabled the policy makers to assess the effectiveness of the Programme in certain cities. This would also help researchers to unpack the mechanism of the preventive processes and raise further research questions such as: why is it that some cities are more successful in implementing crime prevention initiatives than others?

Chapter Twelve
Conclusion and Discussion

This chapter provides an overall summary of this research and examines the implications of the results and discusses the recommendation for the future research and development. This research has explored the spatio-thematic accuracy issues involved in the evaluation of the £30 million Safer Cities Programme carried out by the Home Office RDS between 1988 and 1995. Domestic burglary on dwellings was chosen for the impact evaluation. Two cities (Bristol and Coventry) were selected for in depth analyses. Based on a broad literature review and in particular, Lanter and Veregin's (1992) paradigm and Fisher and Langford's (1995) assessment of areal interpolation methods, an innovative methodology has been implemented to handle the spatio-thematic accuracy issues. The results have a number of important implications to:

- The evaluation of the Safer Cities Programme (Section 12.1).
- The crime pattern analyses and the evaluation of the future crime preventive action (Section 12.2).
- Spatial data quality assessment (Section 12.3).

Finally, recommendations for future research and development are discussed in Section 12.4 (which are summarised in Section 12.5).

12.1 Implications to the evaluation of the Safer Cities Programme

The initial estimation using the areal weighted method has suggested that the average errors of the overlay method is in the range 10-11% for the two cities which are comparable to the cartographic errors estimated using McAlpine and Cook's formula. The results of using the more accurate Monte Carlo dasymetric method show that the overlay method used in the evaluation, after the scoping process, has over-estimated the household counts by 3.6% and 5% for Bristol and Coventry respectively, much less than those estimated by the areal weighted method. As a result, the Safer Cities

Programme evaluation has underestimated the action intensity by -0.8 and -9% and the burglary risk by –7% and –5% (for Bristol and Coventry respectively). Multi-level modelling has shown that the mean errors due to the spatial interpolation estimated by the Monte Carlo dasymetric method are –1.5%, 2.3% and 0.7% for Coventry, Bristol, and the two cities combined respectively. However as action scores increase to the maximum, the estimated error increases to the maximum (5% when the two cities are combined). In all cases, these are well within the standard error generated by the overlay method. This supports the general practice in statistics that when the error is less than 5%, the impact is unlikely to be significant.

Taking the spatial and thematic errors into account (i.e. using Monte Carlo dasymetric method) the difference between the two methods is too small to have a significant impact upon the conclusion of the Safer Cities Programme evaluation ($p = 0.2$ for Coventry; 0.08 for Bristol; and 0.76 when the two cities are combined; c.f. $p = 0.108$ for all Safer Cities; with d.f. = 1 in all cases). <u>It is concluded that no significant impact due to spatial and thematic errors upon the conclusion of the Safer Cities Programme evaluation has been found.</u>

This conclusion is encouraging to both the evaluation team and the policy makers. The research finding suggests that the methodology used by the Safer Cities Programme evaluation is robust enough to cope with the spatio-thematic error. From now on, we can use the evaluation strategy developed from the Safer Cities Programme with confidence, and apply it to the future evaluation of crime prevention initiatives and GIS application. If necessary, the methodology developed in this research can be used to test the validity of the future application as an extra-quality assurance.

There are at least three reasons why no significant impact are detected in the evaluation as a whole (at the 'global' aggregated scale) while the estimated error varies widely across different beats (at the 'local' disaggregated scale). First, as the number of data sets used in a GIS analysis increases, the accuracy of the result decreases due to the aggregation effects (see Veregin, 1989). This finding confirms this general conjecture. Second, the multi-level modelling shows that the risk-action relationship between the overlay and Monte Carlo dasymetric methods tend to interact in the multi-level modelling and that the overall effect of the underestimation of the action intensity and burglary risk seemed to cancel each other out when the two cities are combined.

Third, because this research has only focused on the spatio-thematic error in GIS processing, any other types of errors would have been excluded from the scope of this case study. For example, as said in Chapter 5 (5.1), it is *assumed* that there are no gross error blunders in all the input data. However, such an assumption may be unrealistic for this application with a very large amount of data sets. Some errors might have occurred in the early stages of data capture such as data collection and input. For instance, on a closer re-examination of Figure 2.7 (top left hand corner of the map), it seems likely that there might be some gross error such as mis-registration in the digitizing stage. This may account for the large error margin within the evaluation of the Safer Cities Programme. So we must not be too optimistic about the scale of error within the evaluation just because this case study shows that the aspects of GIS processing (spatial interpolation in particular) have not affected the conclusion significantly.

12.2 Implications to crime pattern analyses and the evaluation of the future crime preventive action

The crime pattern analysis in this case study shows that potential hot spots might have been missed as a result of such under-estimation. This has an important implication for decision and policy makers in terms of crime prevention and resource allocation. If more accurate high-risk areas were identified, decision-makers would be able to target those areas with more appropriate resources and crime preventive strategies. The observation also implies that a spatial unit as large as a beat with aggregated data is susceptible to the spatio-thematic error and therefore is not appropriate for crime pattern analysis. Alternative approaches with smaller spatial units such as postcodes and enumeration districts would give a more accurate spatial pattern (for example, Brunsdon 1989; Ratcliffe and McCullagh, 1998; Bowers and Hirschfield, 1999).

The analyses also show that the Safer Cities effect varies across different cities as exemplified by Bristol and Coventry. Coventry appears to have a greater action effect upon the burglary risk reduction than Bristol. It may be generalised that the Safer Cities schemes of some Safer Cities may be more 'successful' than the others. Combining the data sets from all the Safer Cities in the analysis tend to dilute the effect of the more 'successful' Safer Cities schemes. With such hindsight, it would have been better to analyse each city individually as exemplified by this case study. This would have enabled the evaluation team to 'comb' the more 'successful' Safer Cities schemes and investigate further the mechanism behind the success (as well as the failure) of Safer Cities schemes (rather than the Programme as a whole). While it is justified for the evaluation to combine all the cities in the multi-level model to assess the impact of the Safer Cities Programme as a whole (as required by the Treasury), analysing the effect city by city would have enabled the policy makers to assess the effectiveness of the Programme in certain cities. This would also help researchers to unpack the mechanism of the preventive processes and raise further research questions such as: why is it that some cities are more successful in implementing crime prevention initiatives than others?

12.3 Implications to the data quality assessment in GIS processing

Spatial analyses of the error distribution show that a geographical area would have a higher error when it has:

- dense population;
- nearness to the city centre; or
- an irregular geographical boundary.

Further techniques need to be developed to assess the spatial structure of errors in practical applications.

Development of spatial data accuracy assessment procedures

Since there are no formal agreed methods for individual applications as it depends on the context, there is a need to develop a standard of best practice as exemplified by this research. The following step by step procedures have been formulated to assess the spatio-thematic accuracy of the spatial interpolation:

1. Identify the *error index*.
2. Perform a quick evaluation of the range of errors using simple methods (such as area weighting) to check whether further assessment is required (if so, proceed to Step 3; else stop).
3. Develop *error propagation functions* within the data transformation processes using Monte Carlo dasymetric method.

3.1. Get the satellite imagery.

3.2. Calibrate the satellite image using the source zones to estimate the attribute values of interests.

3.3. Estimate the attribute values in the target zones using the calibrated satellite image as a dasymetric map and Monte Carlo simulation.

4. Test the utility of Step 2 by assessing spatio-thematic accuracy in the GIS application.

12.4 Recommendations for future research and development

Within the scope of this research, the case study has left no stone unturned. It has followed every important step of the GIS processes thoroughly right until the end of the conclusion. The scope has excluded the 'survey' sampling which is within the domain of social sciences. The findings of this case study can be generalised to be applied to those in the survey in the evaluation of the Safer Cities Programme as the results of the evaluation using official crime data are consistent with victimisation data from survey. Previous research has also provided a similar consistency between official records and survey data (Blumstein *et al*, 1991; McDowall and Laffin, 1992). Nevertheless, if GIS is to have a wider impact, it should have a functionality to handle spatial sampling. While sampling approaches have been well developed in both geography and social sciences, how to use GIS to facilitate sampling is still under developed. Future research should address how to handle small number sampling over different geographical and temporal units.

Within the discipline of quantitative criminology, it is not yet standard practice to use GIS as a research tool. There is no agreed 'ecologically valid' measure of the geographical context of potential crime victimisation. Methods and research strategies of measuring crime are still being carried out without the benefits of incorporating GIS (for example, see Holzman and Piper, 1998 on measuring crime in public housing). Maltz (1988) discusses the utility of graphical methods in visualising crime data such as homicide, again, without reference to the use of GIS. If GIS is to have a significant contribution to the field and other disciplines, it would have to be incorporated as an integral part of the methodological framework. Future development should continue to try to integrate GIS with other systems, in particular, visualisation, simulation and statistical tools. While such tools exist, they create a system of loose coupling at present (for example, ARC/INFO S-plus, MapInfo/SAS). This still requires a tremendous amount of work from the users to 'connect' up the system as a whole for a particular kind of application.

While much is known of the motivation of crime such as burglaries (Bennett and Wright, 1984, Field, 1990), relatively little is known about the validity of the criminological theories within the geographical context such as social cohesion (Kurtz *et al*, 1995, Hirschfield, *et al*, 1995, and Taylor *et al*, 1995) routine activity theories, (Cohen and Felson, 1979; Clark, 1983), proximal circumstances (Ekblom 1994) and evolutionary struggles (Ekblom, 1999). Further research using GIS is required to explore the ecological validity of these theories.

Within the discipline of GIS, the experience from this research suggests that there is still a long way to the ideal of pressing a single button to show the error quality of any application (Openshaw, 1989). It would help the users if the function to assess data accuracy is integrated as part of the standard GIS. With a few exceptions, GIS developers are reluctant to add error-handling functions to their products. IDRISI includes tools for describing uncertainty in metadata, and propagating its effects in GIS operations such as overlay. GRASS provides access to tools for Monte Carlo simulation, and the spatial structure of uncertainty. As assessing the geographical data quality is context specific, each application has to be assessed according to an acceptable set of procedures as exemplified by this case study. At best we can only provide guidelines and standards of best practice. Recommendation to "give the customer no more, and no less, quality that what he [or she] needs" (ISO-9000) is not helpful as it does not specify the exact quantity of the required quality. The information should include accuracy in a form of user report or meta-data stored in the GIS. In UK, meta-data will be available on 2001 Census definition, concepts, output classifications, geography, data quality and coverage (ONS, 1999). The information about the data quality in the input data sets would serve as a starting point for the data accuracy assessment. Further data processing in an application may then be assessed using the procedures outlined in Section 12.3 or some other appropriate procedures. A few important areas are discussed in greater detail in the following sub-sections. Other recommendations are included in the summary in Section 12.5.

Spatial structure. As indicated in Section 12.3, spatial structure of error is very important as it has a direct impact on the outcome of GIS applications. As Goodchild (1995) pointed out, very little is known about the spatial structure of errors and it receives little attention in implemented or proposed standards. Many spatial data standards are limited to requirements for positional accuracy alone (e.g. Federal Information Processing Standard 173, Morrison, 1992). Estimates of the final attribute values (such as the action intensity in this case study) depend not only on the original values of the attributes, but also on the spatial structure of inclusions in polygons (in this case beats included in the scoping process). Although some insight has been gained from this research about the spatial structures of uncertainty in geographic data, more research is needed to investigate how these structures govern the outcome of an application in a finer geographical unit.

Units of analyses. The phenomena of the Modifiable Units implies that only spatial analysis methods that exhibit invariance of conclusions under alternative spatial partitioning should be used (Tobler, 1989). It would be useful to find the optimal units of analyses if they do exist.

Paradigm shift. A great improvement is being made in the Census 2001 by adopting some of the recommendations made by Openshaw and Rao (1994; 1995) and by separating the output geography from the collection geography using postcodes as building bricks (Martin, 1998; ONS, 1999). However this process is not free of accuracy issues. First, certain inaccurate placements of address points would result in inaccurate output area boundaries, and hence inaccurate counts in those areas. Second, some postcodes are split into two or more separate parts, adding the cost of creating output areas. Even if we could find the optimal units of analyses, the Modifiable Units Problem would still exist (Openshaw and Taylor, 1979). For the problems of the Modifiable Units and the spatial interpolation completely disappear, one may argue that we have to abandon the concept of an areal unit (a zone) all together. This calls for a paradigm shift. For instance, one can use raster data rather than vector data for GIS application. A key question is how to develop a probability surface for geographical data.

In the past vector data were sometimes regarded (somewhat wrongly) as intelligent data because many methods for vector data handling had been well developed (as the literature review of this research shows) in comparison to those employed in raster data. However, since the early attempt as exemplified by Tobler's (1979) work on the smooth pycnophylactic surfaces, recent research has been developed to close this gap. For example, Arbia *et al* (1998) analyse how both spatial and thematic errors interact with the source map geography through propagation in raster GIS as a result of overlay operations (though in a much smaller scale than this research and using different methodology and error index). Brunsdon (1995) has developed an adaptive kernel algorithm to estimate the probability for point data. This can be readily applied to our context of crime prevention, say, to estimate a "risk surface" of household burglaries in various parts of the study area (Brunsdon, 1995). Other geographical phenomena can be modelled in a similar way in terms of their relative likelihoods. The approach is consistent with the recent attempts within the discipline of statistics to develop a unified approach for nonparametric smoothing (for example, the optimal bandwidth selection and confidence interval construction in local likelihood estimation by Fan *et al*, 1998). Statistical techniques using surface models for handling spatial data are now well developed (Cressie, 1993). More research and practical examples of this kind of application in the future would enable closer linkage between statistical analysis and the use of GIS.

12.5 Summary of recommendations

While many spatial interpolation methods and their accuracy assessment have been developed, more work is needed on the integration of statistical methods for spatial data quality measurement. To summarise, the recommendations for the future research and development are listed as follows (in no particular order):

- Develop system functions to help users to assess data accuracy as part of the GIS standard.
- Develop meta-database include data quality as part of its attribute.
- Develop spatial sampling strategy for other disciplines such as social research.
- Integrate GIS with other systems such as statistical analytical system.
- Use GIS as a tool for testing criminological theories standardised error reporting.
- Provide a range of evaluation methods as part of the GIS functionality.
- Develop further techniques to assess the spatial structure of errors in practical applications.
- Investigate the optimal unit of analyses.
- Investigate alternative approaches for spatial representation such as probability surfaces.
- Implement standard of best practice.

If the above research and development flourish, more accurate information on data quality is made available to the GIS users, and more accurate conclusion of GIS applications can be made. This would also help GIS to be accepted by the users within other disciplines such as statistics and social sciences. This calls for a multi-disciplinary research using GIS as a tool and based on geography as a common ground, and translating the standard of best practice into policy. Some development on this field is already under way. For example, on completion of writing up this thesis, I have written the recommendations for the adoption of Geographical Information Charter Standard Statement by the Home Office to the Permanent Under Secretary of State via the Director of the RDS (see Appendix 12.1). Other development in UK includes National Geo-Data Format. Policy formulation of the spatial data quality standards may lead to the implementation of the best practice for providing geographical information. Hopefully this would reduce the burden on a user's part for quality assessment in the future.

Appendix 1

Home Office's statement of purpose

Following the General Election in 1997, the Home Secretary agreed a new statement of purpose and aims for the Home Office. These are set out below (Home Office Annual Report 1998, p 4)

Statement of Purpose

To build a safe, just and tolerant society in which the rights and responsibilities of individuals, families and communities are properly balanced and the protection and security of the public are maintained.

AIMS

1. Reduction in crime, particularly youth crime, and in the fear of crime; and the maintenance of public safety and good order.
2. Delivery of justice through effective and efficient investigation, prosecution, trial and sentencing, and through support for victims.
3. Prevention of terrorism, reduction in other organised and international crime and protection against threats to national security.
4. Effective execution of the sentences of the courts so as to reduce re-offending and protect the public.
5. Helping to build, under a modernised constitution, a fair and prosperous society, in which everyone has a stake, and in which the rights and responsibilities of individuals, families and communities are properly balanced.
6. Regulation of entry to and settlement in the UK in the interests of social stability and economic growth and facilitation of travel by UK citizens.
7. Reduction in the incidence of fire and related death, injury and damage and ensuring the safety of the public through civil protection.

(Also available at http://intranet/intranet/internet/howeb/webwork/ho_funct.htm)

Appendix 2

Entity-Attribute Tables in the INFO database

This documentation provides a complete listing of all the entities used in the Evaluation of the Safer Cities Programme, and represents the *storage schema* in the relational database (Everest, 1986). Broadly speaking, these can be grouped into three types:

1. Spatial (Section A2.1);
2. Thematic (Section A2.2); and
3. Relational (Section A2.3).

Relational data sets are the geographical linkages to provide a common key to all entities. The spatial data sets consist of geographical entities used in the Evaluation. These are described in the ARC/INFO format. They consist of AAT, BND, PAT and TIC tables. These are listed in Section A2.1.

The thematic data sets consist of:

1. Action data – MIS.DAT (Section A2.2.1)
2. Outcome data – Crime statistics (Section A2.2.2)
3. Demographic (Section A2.2.3)

All the above entities (spatial, thematic, and relational) are stored in separate directories (city by city) in the INFO database using ARC/INFO. The first three letters of the entities identify City name (for example 'bri' for Bristol and 'cov' for Coventry). The description is in standard INFO format: attribute name (ITEM), input width (WDTH), output width (OPUT), data type (TYP), number of decimal (N.DEC), and ALTERNATE NAME if any. Comments are added in each section wherever necessary to explain the semantic meaning of the attributes. The following listing shows the data sets within ARC/INFO on a VAXstation 4000.60 named Yeats:

```
Yeats> arc
Copyright (C) 1994 Environmental Systems Research Institute, Inc.
        All Rights Reserved Worldwide.
ARC Version 6.1.3  (April 15, 1994)

This software is provided with RESTRICTED AND LIMITED RIGHTS.  Use,
duplication, or disclosure by the Government is subject to the
restrictions as set forth in FAR 52.227-14 (JUN 1987) Alternate III
```

```
Arc: w D$YEATS2:[HO.CITIES2.BRI]
Arc: info
 INFO CALL EXCHANGE
 26-APR-1999 13:11
 INFO 9.23D 1/4/91 52.74-63*
 COPYRIGHT 1983 HENCO SOFTWARE, INC.
 PROPRIETARY TO HENCO SOFTWARE, INC.
 ENTER USER NAME>arc

 ENTER COMMAND >dir
 TYPE NAME                   INTERNAL NAME  NO. RECS LENGTH EXTERNL
  DF  BEATGS.TIC               ARC000DAT       4    12   XX
  DF  BEATGS.BND               ARC001DAT       1    16   XX
  DF  BEATGS.PAT               ARC002DAT      63    50   XX
  DF  BEATGS.AAT               ARC003DAT     219    28   XX
  DF  BRI-GOB.DAT              ARC007DAT     822   467
  DF  BRI-MIS.DAT              ARC010DAT     298   726
  DF  BRI-LOC.DAT              ARC011DAT    2645    35
  DF  SUPER-CRIMES.DAT         ARC013DAT    2652    87
  DF  CRIMES.DAT               ARC023DAT    4631    67
  DF  BRI-EDBEATGS.TIC         ARC031DAT       4    12   XX
  DF  BRI-EDBEATGS.BND         ARC033DAT       1    16   XX
  DF  BRI-EDBEATGS.PAT         ARC034DAT     850    60   XX
  DF  EDGS.PAT                 ARC042DAT     825    42   XX
  DF  SUPERBTS1.DAT            ARC054DAT     135    62
```

A2.1 Spatial data sets

BEATGS and EDGS are the boundaries of beats and ED respectively. The two combined using ARC/INFO UNION command to form a new entity called BRI-EDBEATGS. The superbeat entity (SUPERBTS1.DAT) is used to keep a record of the dates of change in beat boundaries (START-DATE, LAST-DATE).

```
ENTER COMMAND >select BEATGS.TIC
   4 RECORD(S) SELECTED

 ENTER COMMAND >item
 DATAFILE NAME: BEATGS.TIC                    4/26/1999
   3 ITEMS: STARTING IN POSITION   1
 COL  ITEM NAME      WDTH OPUT TYP N.DEC  ALTERNATE NAME
   1  IDTIC            4    5   B    0
   5  XTIC             4   12   F    3
   9  YTIC             4   12   F    3

 ENTER COMMAND >list
 $RECNO  IDTIC      XTIC         YTIC
    1      1  350,000.000  166,000.000
    2      2  365,000.000  166,000.000
    3      3  350,000.000  181,000.000
```

```
  4     4 365,000.000 181,000.000
ENTER COMMAND >sel BEATGS.BND
   1 RECORD(S) SELECTED

ENTER COMMAND >item
DATAFILE NAME: BEATGS.BND                          4/26/1999
  4 ITEMS: STARTING IN POSITION   1
COL  ITEM NAME          WDTH OPUT TYP N.DEC  ALTERNATE NAME
  1  XMIN                 4  12  F   3
  5  YMIN                 4  12  F   3
  9  XMAX                 4  12  F   3
 13  YMAX                 4  12  F   3

ENTER COMMAND >list
$RECNO          XMIN         YMIN        XMAX        YMAX
   1   350,344.500 166,642.000 364,671.500 180,001.000


ENTER COMMAND >SEL BEATGS.AAT
   219 RECORD(S) SELECTED

ENTER COMMAND >IT
DATAFILE NAME: BEATGS.AAT                           5/ 8/1999
  7 ITEMS: STARTING IN POSITION   1
COL  ITEM NAME          WDTH OPUT TYP N.DEC  ALTERNATE NAME
  1  FNODE#               4   5  B   0
  5  TNODE#               4   5  B   0
  9  LPOLY#               4   5  B   0
 13  RPOLY#               4   5  B   0
 17  LENGTH               4  12  F   3
 21  BEATGS#              4   5  B   0
 25  BEATGS-ID            4   5  B   0

ENTER COMMAND >SEL BEATGS.PAT
   63 RECORD(S) SELECTED

ENTER COMMAND >IT
DATAFILE NAME: BEATGS.PAT                           5/ 8/1999
  8 ITEMS: STARTING IN POSITION   1
COL  ITEM NAME          WDTH OPUT TYP N.DEC  ALTERNATE NAME
  1  AREA                 4  12  F   3
  5  PERIMETER            4  12  F   3
  9  BEATGS#              4   5  B   0
 13  BEATGS-ID            4   5  B   0
 17  BEAT-ID             10  10  C   -
 27  OLD-BEAT-ID         10  10  C   -
 37  BEAT-NUMBER          4   5  B   0
 41  SUPERBEAT           10  10  C   -


ENTER COMMAND >SEL EDGS.AAT
  2346 RECORD(S) SELECTED

ENTER COMMAND >IT
DATAFILE NAME: EDGS.AAT                             5/ 8/1999
  8 ITEMS: STARTING IN POSITION   1
COL  ITEM NAME          WDTH OPUT TYP N.DEC  ALTERNATE NAME
  1  FNODE#               4   5  B   0
```

```
  5 TNODE#          4  5 B  0
  9 LPOLY#          4  5 B  0
 13 RPOLY#          4  5 B  0
 17 LENGTH          4  12 F  3
 21 EDGS#           4  5 B  0
 25 EDGS-ID         4  5 B  0
 29 F_CODE          10  10 C  -


ENTER COMMAND >SEL EDGS.BND
   1 RECORD(S) SELECTED

ENTER COMMAND >IT
DATAFILE NAME: EDGS.BND                       5/ 8/1999
  4 ITEMS: STARTING IN POSITION  1
COL  ITEM NAME      WDTH OPUT TYP N.DEC  ALTERNATE NAME
  1 XMIN            4  12 F  3
  5 YMIN            4  12 F  3
  9 XMAX            4  12 F  3
 13 YMAX            4  12 F  3

ENTER COMMAND >LI
$RECNO        XMIN        YMIN        XMAX        YMAX
   1   350,396.500  166,642.000  364,674.000  183,044.000

ENTER COMMAND >SEL EDGS.TIC
   4 RECORD(S) SELECTED

ENTER COMMAND >IT
DATAFILE NAME: EDGS.TIC                       5/ 8/1999
  3 ITEMS: STARTING IN POSITION  1
COL  ITEM NAME      WDTH OPUT TYP N.DEC  ALTERNATE NAME
  1 IDTIC           4  5 B  0
  5 XTIC            4  12 F  3
  9 YTIC            4  12 F  3

ENTER COMMAND >LI
$RECNO  IDTIC      XTIC        YTIC
   1     1  84,000.000   5,000.000
   2     2 656,000.000   5,000.000
   3     3  84,000.000 660,000.000
   4     4 656,000.000 660,000.000

ENTER COMMAND >SEL EDGS.PAT
  825 RECORD(S) SELECTED


ENTER COMMAND >IT
DATAFILE NAME: EDGS.PAT                       5/ 8/1999
  9 ITEMS: STARTING IN POSITION  1
COL  ITEM NAME      WDTH OPUT TYP N.DEC  ALTERNATE NAME
  1 AREA            4  12 F  3
  5 PERIMETER       4  12 F  3
  9 EDGS#           4  5 B  0
 13 EDGS-ID         4  5 B  0
 17 ED-ID           10  10 C  -
 27 ED-NUMBER       4  5 B  0
 31 ZONE            4  4 I  -
```

```
 35 ZA1              4   4 I  -
 39 ZA2              4   4 I  -
   ** REDEFINED ITEMS **
 17 WARD             4   4 C  -


ENTER COMMAND >SEL BRI-EDBEATGS.TIC
    4 RECORD(S) SELECTED

ENTER COMMAND >LI
$RECNO  IDTIC      XTIC        YTIC
   1      1  84,000.000   5,000.000
   2      2 656,000.000   5,000.000
   3      3  84,000.000 660,000.000
   4      4 656,000.000 660,000.000

ENTER COMMAND >IT
DATAFILE NAME: BRI-EDBEATGS.TIC                          5/ 8/1999
  3 ITEMS: STARTING IN POSITION   1
COL  ITEM NAME        WDTH OPUT TYP N.DEC  ALTERNATE NAME
  1 IDTIC              4   5 B   0
  5 XTIC               4  12 F   3
  9 YTIC               4  12 F   3


ENTER COMMAND >SEL BRI-EDBEATGS.BND
    1 RECORD(S) SELECTED

ENTER COMMAND >IT
DATAFILE NAME: BRI-EDBEATGS.BND                          5/ 8/1999
  4 ITEMS: STARTING IN POSITION   1
COL  ITEM NAME        WDTH OPUT TYP N.DEC  ALTERNATE NAME
  1 XMIN               4  12 F   3
  5 YMIN               4  12 F   3
  9 XMAX               4  12 F   3
 13 YMAX               4  12 F   3

ENTER COMMAND >LI
$RECNO         XMIN         YMIN        XMAX         YMAX
   1    350,344.500  166,642.000  364,674.000  183,044.000


ENTER COMMAND >SEL BRI-EDBEATGS.PAT
  850 RECORD(S) SELECTED

ENTER COMMAND >IT
DATAFILE NAME: BRI-EDBEATGS.PAT                          5/ 8/1999
  9 ITEMS: STARTING IN POSITION   1
COL  ITEM NAME        WDTH OPUT TYP N.DEC  ALTERNATE NAME
  1 AREA                      4  12 F   3
  5 PERIMETER                 4  12 F   3
  9 BRI-EDBEATGS#             4   5 B   0
 13 BRI-EDBEATGS-ID           4   5 B   0
 17 ED-ID                    10  10 C   -
 27 BEAT-ID                  10  10 C   -
 37 BEAT-NUMBER               4   5 B   0
 41 SUPERBEAT                10  10 C   -
 51 OLD-BEAT-ID              10  10 C   -
```

```
DATAFILE NAME: SUPERBTS1.DAT                      5/15/1995
  4 ITEMS: STARTING IN POSITION   1
 COL  ITEM NAME        WDTH OPUT TYP N.DEC  ALTERNATE NAME
   1  SUPER-ID          36   36  C    -
  37  BEAT-ID           10   10  C    -
  47  START-DATE         8   10  D    -
  55  LAST-DATE          8   10  D    -
```

A2.2 Thematic data sets

Both the action (MIS.DAT) and the demographic data sets share some of the Census variables called Geographical Outcome Base (GOB). These are described first.

Geographical Outcome Base (GOB)

In the MIS.DAT entity these are demographic totals derived from 1991 Census data for the entire scheme coverage. In the GOB.DAT entity they relate solely to the ED in question. The names of the attributes (or called items in INFO) are the same and refer to the same Census variable for both MIS.DAT and GOB.DAT entities. These are listed as follows:

TOTAL RESIDENTS,GOB1
TOTAL-ADULT,GOB2
TOTAL-FEMALE,GOB3
CHILDREN LESS THAN 10,GOB4
CHILDREN LESS THAN 10 FEMALE,GOB5
CHILDREN LESS THAN 10 BLACK,GOB6
CHILDREN LESS THAN 10 ASIAN, GOB7
CHILDREN LESS THAN 10 FEMALE ASIAN, GOB9
YOUTH11-17,GOB10
YOUTH11-17 FEMALE,GOB11
YOUTH11-17 BLACK,GOB12
YOUTH11-17 ASIAN,GOB13
ELDERLY,GOB35
ELDERLY-FEMALE,GOB36
BLACK,GOB14
BLACK-ADULT,GOB15
BLACK-ADULT-ELDERLY,GOB16
BLACK-ADULT-FEMALE,GOB17
ASIAN,GOB19
ASIAN-ADULT,GOB20
ASIAN-ADULT-FEMALE,GOB22
CHINESE,GOB23
CHINESE-ADULT,GOB24
CHINESE-ADULT-FEMALE,GOB25
BLACK-ASIAN-CHINESE,GOB30
BLACK-ASIAN-CHINESE-ADULT,GOB31
BLACK-ASIAN,GOB32

CHILD-YOUTH,GOB34
TOTAL-HOUSEHOLDS,GOB50
BLACK-HEAD OF HOUSEHOLD,GOB51
ASIAN-HEAD OF HOUSEHOLD,GOB52
CHILD-YOUTH IN HOUSEHOLD, GOB53
FEMALE-HEAD OF HOUSEHOLD,GOB56
ELDERLY-HEAD OF HOUSEHOLD,GOB57
ELDERLY-BLACK HEAD OF HOUSEHOLD,GOB58
ELDERLY ASIAN-HEAD OF HOUSEHOLD,GOB59
ELDERLY-FEMALE-HEAD OF HOUSEHOLD,GOB60
NO CAR - HOUSEHOLD, GOB61

A2.2.1 Action data from the Management Information System (MIS)

This entity is formed by combining a number of different data sources. AREA represents the scheme coverage in m^2. AREA-ID locates the size of the target area: 1 City centre, 2 Citywide, 3 Local area. A scheme may cover 1 to 3 crime types CRIME1-ID, CRIME2-ID, and CRIME3-ID. Each attribute may have one of the following values: 1 represents Violence against the person, 2 Sexual offences, 3 Domestic Violence, 4 Burglary, 5 Robbery, 6 Theft from the person, 7 Theft from shops, 8 Theft from vehicles, 9 Theft of motor vehicles, 10 Handling stolen goods, 11 Criminal Damage, 12 Drug related, 13 Alcohol related, 14 Vandalism, 15 Graffiti & Litter, 16 Racially-Motivated Offences & Harassment, 17 Public Disorder / Rowdiness, 18 Fear of crime, and 19 Other.

CITY-ID identifies the name of the city: HAC (value 1), SOU (2), HAR (3), MID (4), MAN (5), NOR (6), WIG (7), OLD (8), LEE (9), BIR (10), BRI (11), COV (12), HUL (13), LEW (14), ROC (15), SAL (16), SUN (17), TOW (18), WAN (19), WIR (20), BRA (21), HAR (22), ISL (23), NOT (24), and WOL (25).

BODY-ID identifies the organisation that involved in the Safer Cities Programme: 1 unknown, 2 Business, 3 Charity, 4 Local Authority, 5 Other, 6 Police, 7 Private Sector, 8 Probation, 9 Voluntary Organisation.

A scheme may target for more than one objectives of the Safer Cities Programme (OBJEC1-ID & OBJEC2-ID): 1 Crime Reduction, 2 Crime Reduction, 3 Improve Economic Activity, and 4 Improve Community Life (5 represents a missing value).

OFFEND1-ID, OFFEND2-ID, and OFFEND3-ID describe the schemes that target offenders and have the following values: 1 represents Whole community, 2 Youth (11-17), 3 Victims, 4 Female, 5 Male, 6 Black, 7 White, 8 Asian, 9 Chinese, 10 Drug /Alcohol Abusers, 11 Unemployed, 12 Known cautioned / convicted offenders, and 13 Other.

OUTCOME-ID (the outcome of the scheme operation) has the following values:

- 1 - Abandoned early - Implementation incomplete
- 2 - Abandoned early - Implementation complete but with problems
- 3 - Action ceased at planned time
- 4 - Scheme still operational for at least another
- 5 - Other body takes over funding and /or direct
- 6 - Missing

PHYSICAL1 describes the physical characteristics of the target area:

- 1 - Car Parks
- 2 - Dwellings
- 3 - Other
- 4 - Shopping Precincts/Centres/Streets
- 5 - Shops/Stores
- 6 - Industrial Estates
- 7 - Public Buildings
- 8 - Other leisure sites
- 9 - Factories/Warehouses
- 10 - Offices
- 11 - Pubs/Bars/Clubs
- 12 - Fast food outlets

PREVENTIV1 & PREVENTIV2 describe the nature of the preventive action:

- 1 - Target hardening - locks
- 2 - Social action
- 3 - Research and evaluation
- 4 - Education and training
- 5 - Publicity and campaigns
- 6 - Conferences and seminars
- 7 - Homewatch and CPO related
- 8 - Graffiti & Litter removal
- 9 - Victim Support
- 10 - Lighting
- 11 - Surveillance
- 12 - CCTV
- 13 - Environmental design
- 14 - Community development
- 15 - Improvement of leisure facilities for youth

- 16 - Potential offender -oriented
- 17 - Other

STATUS-ID describes the STATUS of the scheme:

- 1 - Planned
- 2 - Current
- 3 - Completed
- 4 - Cancelled
- 5 - Aborted
- 6 - Missing

TIME-ID describes the duration of the scheme:

- 1 - Immediate / very short term
- 2 - Medium term
- 3 - Longer term
- 4 - Very long term
- 5 - Missing

The types of victim supported by the Safer Cities action (VICTIM1, VICTIM2, VICTIM3) are:

- 1 - Whole community
- 2 - Children (10 & under)
- 3 - Youth (11-17)
- 4 - Elderly
- 5 - Female
- 6 - Black
- 7 - White
- 8 - Asian
- 9 - Chinese
- 10 - Unemployed
- 11 - Residents - Housing problems
- 12 - Single parent
- 13 - Victims of Crime / Fear
- 14 - Other

DATAFILE NAME: BRI-MIS.DAT 4/30/1999

108 ITEMS: STARTING IN POSITION 1

COL	ITEM NAME	WDTH	OPUT	TYP	N.DEC	ALTERNATE NAME
1	SCHEME-ID	3	3	I	-	
4	PROJECT	3	3	C	-	
7	OBJEC1-ID	1	1	I	-	
8	OBJEC2-ID	1	1	I	-	
9	STATUS-ID	1	1	I	-	
10	AREA-ID	1	1	I	-	
11	NEIGH-SCOPE	2	2	I	-	
13	SCOPE	2	2	I	-	
15	LOCAL	1	1	I	-	
16	OFFEND1-ID	2	2	I	-	
18	OFFEND2-ID	2	2	I	-	
20	OFFEND3-ID	2	2	I	-	

22	VICTIM1-ID	2	2	I	-
24	VICTIM2-ID	3	3	I	-
27	VICTIM3-ID	3	3	I	-
30	CRIME1-ID	2	2	I	-
32	CRIME2-ID	2	2	I	-
34	CRIME3-ID	2	2	I	-
36	PREVEN1-ID	2	2	I	-
38	PREVEN2-ID	2	2	I	-
40	PHYS-ID	2	2	I	-
42	TIME-ID	1	1	I	-
43	OUTCOME-ID	1	1	I	-
44	COMPNENT-ID	1	1	I	-
45	SOURCE	30	30	C	-
75	VALUE_1	9	9	N	2
84	VALUE_2	9	9	N	2
93	SSURV_FILTER	1	1	I	-
94	START_DATE	8	10	D	-
102	START_CODE	4	4	I	-
106	COMPLETION_DATE	8	10	D	-
114	COMPLETION_CODE	4	4	I	-
118	COST_SC	9	9	N	2
127	AFT-SEPT90	9	9	N	2
136	LAST-12	9	9	N	2
145	OLD_COST_SC	9	9	N	2
154	COST_TOT	9	9	N	2
163	GRANT87-88	9	9	N	2
172	GRANT88-89	9	9	N	2
181	GRANT89-90	9	9	N	2
190	GRANT90-91	9	9	N	2
199	GRANT91-92	9	9	N	2
208	GRANT92-93	9	9	N	2
217	GRANT93-94	9	9	N	2
226	GRANT94-95	9	9	N	2
235	BUDGET87-88	9	9	N	2
244	BUDGET88-89	9	9	N	2
253	BUDGET89-90	9	9	N	2
262	BUDGET90-91	9	9	N	2
271	BUDGET91-92	9	9	N	2
280	BUDGET92-93	9	9	N	2
289	BUDGET93-94	9	9	N	2
298	BUDGET94-95	9	9	N	2
307	BEF-SEP92	9	9	N	2
316	BEF-DEC92	9	9	N	2
325	BEF-JUN92	9	9	N	2
334	LEVERED	9	9	N	2
343	TOTAL_PAID	9	9	N	2
352	ACT_CAPITAL	9	9	N	2
361	ACT_REV	9	9	N	2
370	TILLEY	1	1	I	-
371	GOB-ID	2	2	I	-
373	OLD-GOB-ID	2	2	I	-
375	OLD-AREA-ID	1	1	I	-
376	GOB1	9	9	N	2
385	GOB2	9	9	N	2
394	GOB3	9	9	N	2
403	GOB4	9	9	N	2
412	GOB5	9	9	N	2
421	GOB6	9	9	N	2

430	GOB7	9	9 N	2
439	GOB9	9	9 N	2
448	GOB10	9	9 N	2
457	GOB11	9	9 N	2
466	GOB12	9	9 N	2
475	GOB13	9	9 N	2
484	GOB35	9	9 N	2
493	GOB36	9	9 N	2
502	GOB14	9	9 N	2
511	GOB15	9	9 N	2
520	GOB16	9	9 N	2
529	GOB17	9	9 N	2
538	GOB19	9	9 N	2
547	GOB20	9	9 N	2
556	GOB22	9	9 N	2
565	GOB23	9	9 N	2
574	GOB24	9	9 N	2
583	GOB25	9	9 N	2
592	GOB30	9	9 N	2
601	GOB31	9	9 N	2
610	GOB32	9	9 N	2
619	GOB34	9	9 N	2
628	GOB50	9	9 N	2
637	GOB51	9	9 N	2
646	GOB52	9	9 N	2
655	GOB53	9	9 N	2
664	GOB56	9	9 N	2
673	GOB57	9	9 N	2
682	GOB58	9	9 N	2
691	GOB59	9	9 N	2
700	GOB60	9	9 N	2
709	GOB61	9	9 N	2
718	AREA	4	12 F	3
722	MECH	1	1 I	-
723	REVCAP	1	1 I	-
724	BODY-ID	1	1 I	-
725	APPROVE_REFUSE	1	1 C	-
726	LINK	1	1 I	-

The mechanism of methods of crime prevention (MECH) have values: 0 NIL, 1 Offender, 2 Situational. Revenue / Capital split (REVCAP) has values: 4 (No funding) 1 (Revenue only) 2 (Capital only) 3 (Mix). The funding body (BODY-ID) has values: 1 NOT GIVEN, 2 Business, 3 Charity, 4 Local Authority, 5 Other, 6 Police, 7 Private Sector, 8 Probation, 9 Voluntary Organisation. APPROVE_REFUSE has a binary value: Either A for approve or W for waiting. LINK implies that the scheme has links with others (this variable was not used in Scoping and Scoring).

A2.2.2 Outcome data

The twelve crime types as described in the main text are stored in CRIMES.DAT. This also includes DATE of the offence and SUPERBEAT identifier.

```
DATAFILE NAME: CRIMES.DAT                                5/15/1995
 16 ITEMS: STARTING IN POSITION   1
COL  ITEM NAME        WDTH OPUT TYP N.DEC  ALTERNATE NAME
  1  BEAT-ID           10   10  C    -
 11  VIOLENCE           3    3  I    -
 14  SEXUAL             3    3  I    -
 17  BURGLARY-D         3    3  I    -
 20  THEFT-IN-CAR       3    3  I    -
 23  DAMAGE             3    3  I    -
 26  OTHER              3    3  I    -
 29  DATE               8   10  D    -
 37  BURGLARY-O         3    3  I    -
 40  ROBBERY            3    3  I    -
 43  THEFT-P            3    3  I    -
 46  THEFT-S            3    3  I    -
 49  THEFT-OF-CAR       3    3  I    -
 52  THEFT-O            3    3  I    -
 55  FRAUD              3    3  I    -
 58  SUPERBEAT         10   10  C    -
```

SUPER-CRIMES.DAT is the product of frequency operation on the above (ie remove duplicates). Crime ratios are derived from this and GOB attributes.

```
DATAFILE NAME: SUPER-CRIMES.DAT                          5/15/1995
 19 ITEMS: STARTING IN POSITION   1
COL  ITEM NAME        WDTH OPUT TYP N.DEC  ALTERNATE NAME
  1  CASE#              4    5  B    0
  5  FREQUENCY          4    5  B    0
  9  SUPERBEAT         10   10  C    -
 19  DATE               8   10  D    -
 27  VIOLENCE           4   10  B    0
 31  SEXUAL             4   10  B    0
 35  BURGLARY-D         4   10  B    0
 39  BURGLARY-O         4   10  B    0
 43  ROBBERY            4   10  B    0
 47  THEFT-IN-CAR       4   10  B    0
 51  THEFT-OF-CAR       4   10  B    0
 55  THEFT-P            4   10  B    0
 59  THEFT-S            4   10  B    0
 63  THEFT-O            4   10  B    0
 67  DAMAGE             4   10  B    0
 71  FRAUD              4   10  B    0
 75  OTHER              4   10  B    0
 79  BEAT-NUMBER        4    5  B    0
 83  BEAT-YEAR          5    5  C    -
```

A2.2.3 Demographic data

Demographic data sets (derived from the 1991 Census using the C91 software) include links to the survey. Although the ED-ID in the INFO database is defined with 10-character width, the input values are 6 characters (Indexed). ED-NUMBER is a number assigned from the survey with default value = 0. AREA was taken from EDGS.PAT area as digitised by GDC in m^2. DOMAIN is a geographical area defined by MORI and updated using Target Action Area data (Values: 1 - Target action ED; 2 – Citywide; 3 - External comparison, 0 - Not part of survey).

If a city is a Safer City, SC is set to 1. If it is a London borough, PROV is set to 1. CENTRE is set to 1 if this ED is at the city centre. TAA is set to 1 if this ED is a Target Action Area. ZONE is a group of surveyed EDs. ZA1 and ZA2 are the adjacent zones. When the area consists of more than one adjacent zone, the EDs in or around the ZONE are identified by the attribute called GROUP.

The deprivation index (INDX) and its related variables were supplied by the Department of Environment (now Department of Environment, Transport and the Regions). The variables included unemployment (UNEMPL), over-crowding (OV-CROWD), lack of amenities (LACKAM), children in unacceptable accommodation (CHUNACC), low earning households (LOERNH), and households with no car (NOCAR).

Arc: items bir-gob.dat

1	ED-ID	10	10	C	
11	ED-NUMBER	4	5	B	
15	AREA	4	12	F	3
19	CITY-ID	2	2	I	
21	DOMAIN	3	3	I	
24	GOB1	9	9	N	2
33	GOB2	9	9	N	2
42	GOB3	9	9	N	2
51	GOB4	9	9	N	2
60	GOB5	9	9	N	2
69	GOB6	9	9	N	2
78	GOB7	9	9	N	2
87	GOB9	9	9	N	2
96	GOB10	9	9	N	2
105	GOB11	9	9	N	2
114	GOB12	9	9	N	2
123	GOB13	9	9	N	2
132	GOB35	9	9	N	2
141	GOB36	9	9	N	2
150	GOB14	9	9	N	2

159 GOB15	9	9	N	2	
168 GOB16	9	9	N	2	
177 GOB17	9	9	N	2	
186 GOB19	9	9	N	2	
195 GOB20	9	9	N	2	
204 GOB22	9	9	N	2	
213 GOB23	9	9	N	2	
222 GOB24	9	9	N	2	
231 GOB25	9	9	N	2	
240 GOB30	9	9	N	2	
249 GOB31	9	9	N	2	
258 GOB32	9	9	N	2	
267 GOB34	9	9	N	2	
276 GOB50	9	9	N	2	
285 GOB51	9	9	N	2	
294 GOB52	9	9	N	2	
303 GOB53	9	9	N	2	
312 GOB56	9	9	N	2	
321 GOB57	9	9	N	2	
330 GOB58	9	9	N	2	
339 GOB59	9	9	N	2	
348 GOB60	9	9	N	2	
357 GOB61	9	9	N	2	
366 SC	1	1	I	-	
367 PROV	1	1	I	-	
368 CENTRE	1	1	I	-	
369 TAA	2	2	I	-	
371 OLD-DOMAIN	1	1	I	-	[Original DOMAIN before updating by use of TAA]
372 ZD [Not used]	3	3	I	-	
375 SUPERBEAT	10	10	C	-	
385 BEAT-NUMBER	4	5	B	0	
389 ZONE	4	4	I	-	
393 ZA1	4	4	I	-	
397 ZA2	4	4	I	-	
401 GROUP	4	4	I	-	
405 INDX	9	9	N	2	
414 UNEMPL	9	9	N	2	
423 OV-CROWD	9	9	N	2	
432 LACKAM	9	9	N	2	
441 CHUNACC	9	9	N	2	
450 LOERNH	9	9	N	2	
459 NOCAR	9	9	N	2	
** REDEFINED ITEMS **					
1 WARD	4	4	C	-	

A2.3 Relational data: Geographical linkage

This entity provides a common key (composite) to link all data sets as described in Chapter 2.

Arc: items bir-loc.dat
1 SCHEME-ID
4 ED-ID
14 ED-NUMBER from survey
18 BEAT-NUMBER from crime data via edbeatgs
22 SUPERBEAT where appropriate
32 ZONE defined visually - groups of surveyed eds in this city
** REDEFINED ITEMS **
4 WARD

Appendix 3
Lineage

An example of lineage of data sets - beats, Enumeration Districts (ED), and their combined coverage (EDbeatgs) - used in the Evaluation of the Safer Cities Programme based on the modification of Clarke and Clark's (1995) framework.

Contents

1. Source

1.2 Origin: Beat maps were digitised by GDC Limited. ED boundaries from OPCS

1.3 Reference fields: Beat-ID; ED-ID

1.4 Spatial data characteristics: polygons

1.5 Co-ordinate systems: British Coordinate systems

1.6 Map projections: British Ordinance Survey (OS) Grid

1.7 Corrections and calibrations

2. Pre-processing or Input

2.1 Acquisition (Data collection stage): The 1: 50,000maps were from OS, and beat maps from police via the Safer Cities Ordinators. ED boundaries were purchased from OPCS

2.2 Compilation

- 2.2.1 Scientific parameter generation stage
- 2.2.2 Data conversion stage
  - 2.2.2.1 equipment used: digitizer, VAXstation 4000.60
  - 2.2.2.2 operator policy: GDC, OPCS
  - 2.2.2.3 digitisation policy: minimum 0.5 mm accuracy
  - 2.2.2.4 source material: beat boundaries were drawn on OS maps

2.3 Derivation (Product stage): boundary files (.e00) were generated and transferred into ARC/INFO

3. Transformation and analyses of data [Process]

3.1 Co-ordinate transformation: British Coordinate systems

3.2 Interpolation: beat and ED combined to form a single coverage (EDbeatgs)

Appendix 7.1
Variance and Co-Variance matrix

Table A7.1 shows the variance and co-variance matrix of the principal components for the two cities, converted from the rectified Landsat Image. Band 6 (thermal infrared) was excluded from the principal components analysis. So component 6 represents Band 7.

Table A7.1: The variance and co-variance matrix of the principal components

Coventry

VAR/COVAR	1	2	3	4	5	7
1	1394.45	624.99	607.05	1571.76	1284.52	554.54
2	624.99	283.22	276.85	700.53	583.24	254.39
3	607.05	276.85	282.26	610.16	563.75	263.92
4	1571.76	700.53	610.16	2382.7	1524.1	524.83
5	1284.52	583.24	563.75	1524.1	1316.63	552.42
7	554.54	254.39	263.92	524.83	552.42	270.25

COR MATRIX	1	2	3	4	5	7
1	1	0.995	0.968	0.862	0.948	0.903
2	0.995	1	0.979	0.853	0.955	0.920
3	0.968	0.979	1	0.744	0.925	0.956
4	0.862	0.853	0.744	1	0.860	0.654
5	0.948	0.955	0.925	0.860	1	0.926
7	0.903	0.920	0.956	0.654	0.926	1

Bristol

VAR/COVAR	1	2	3	4	5	6
1	57.93	31.49	51.05	-64.69	-3.21	11.03
2	31.49	20.91	33.15	-17.66	29.2	11.62
3	51.05	33.15	60.12	-45.56	58.21	23.31
4	-64.69	-17.66	-45.56	546.22	522.33	114.89
5	-3.21	29.2	58.21	522.33	909.65	227.79
6	11.03	11.62	23.31	114.89	227.79	92.65

COR MATRIX	1	2	3	4	5	6
1	1	0.905	0.862	-0.0364	-0.014	0.151
2	0.905	1	0.932	-0.165	0.212	0.264
3	0.862	0.93	1	-0.25	0.248	0.311
4	-0.364	-0.165	0.25	1	0.741	0.511
5	-0.014	0.212	0.248	0.741	1	0.785
6	0.15	0.264	0.311	0.511	0.785	1

Appendix 7.2
Testing random numbers

There are many tests developed to determine if a random number sequence {Yn}, (∀n = 1, N) has the desired probabilistic properties (Atkinson, 1980; Dagpunar, 1988; Jennings & Mohan, 1991; Knuth, 1969; MacLaren & Marsaglia, 1965; Sowey, 1972, 1978, 1986; Tauussky and Todd, 1956). For the purpose of this case study, the following three criteria are used.

A7.2.1 Visualisation

The simplest way to see if {Yn}, (∀n = 1,N) is distributed randomly is to examine the sequence visually using 2-D and 3-D scatterplots (Fisher et al, 1993). A 2-D scatterplot is easily implemented using Lisp-Stat 'plot-points' function (Figure A7.2.1 & A7.2.2).

```
For N= 100
> (plot-points (uniform-rand 100) (uniform-rand 100))
#<Object: 527391322, prototype = SCATTERPLOT-PROTO>
```

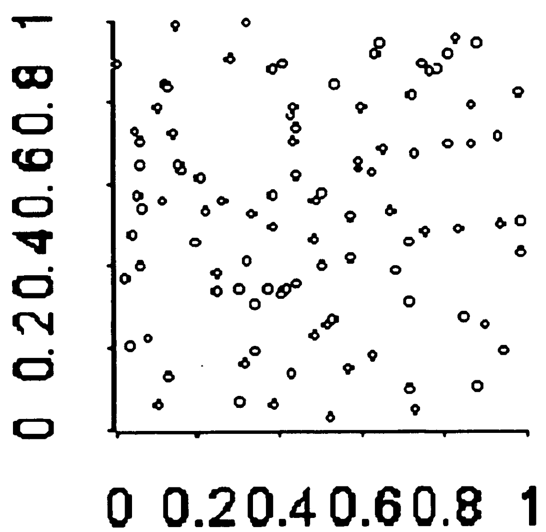


Figure A7.2.1: scatterplot of 100 uniform random numbers

For N=1000

```
> (plot-points (uniform-rand 1000) (uniform-rand 1000))
#<Object: 463452266, prototype = SCATTERPLOT-PROTO>
>
```

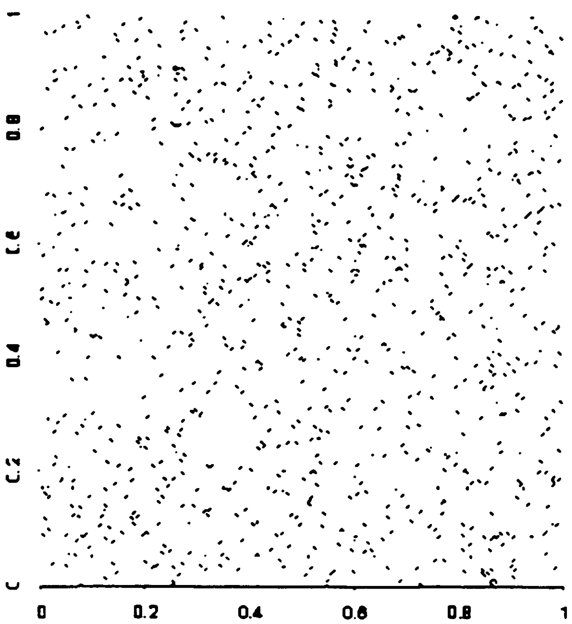


Figure A7.2.2: scatterplot of 1000 uniform random numbers

To prevent the possibility that the random numbers may stay "mainly in the plane", a 3-D view is necessary (Knuth, 1969). For a 3-D scatterplot, we can only observe a 2-D projection of the plot on a computer screen. One way to recover some of the 3-D depth perception is to rotate the points around the same axis. This is achieved using Lisp-Stat 'spin-plot' function to construct a rotatable 3-D plot (Figure A7.2.3 & A7.2.4).

For N=1000

```
(spin-plot (list (uniform-rand 1000) (uniform-rand 1000) (uniform-rand 1000)))
#<Object: 475524522, prototype = SPIN-PROTO>
```

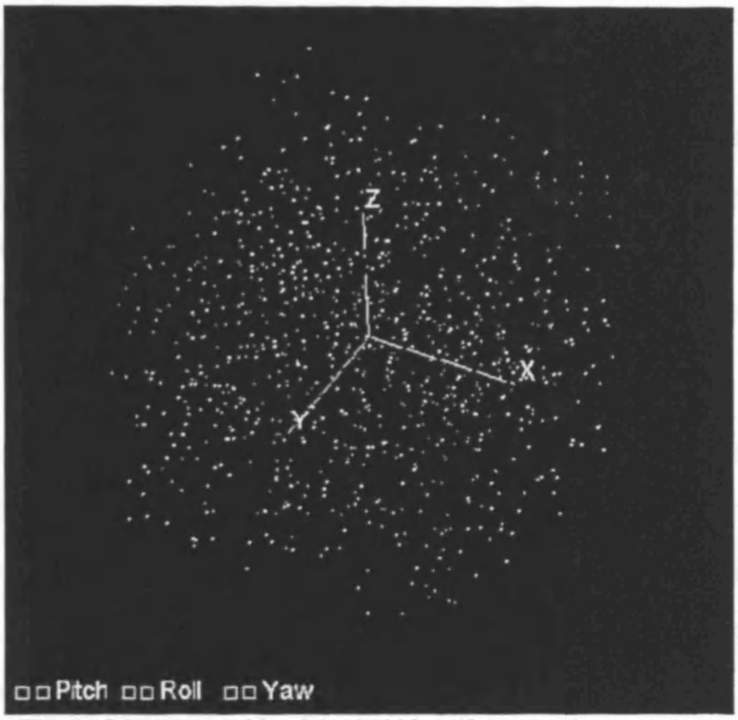


Figure 5.3: A rotatable plot of 1000 uniform random numbers

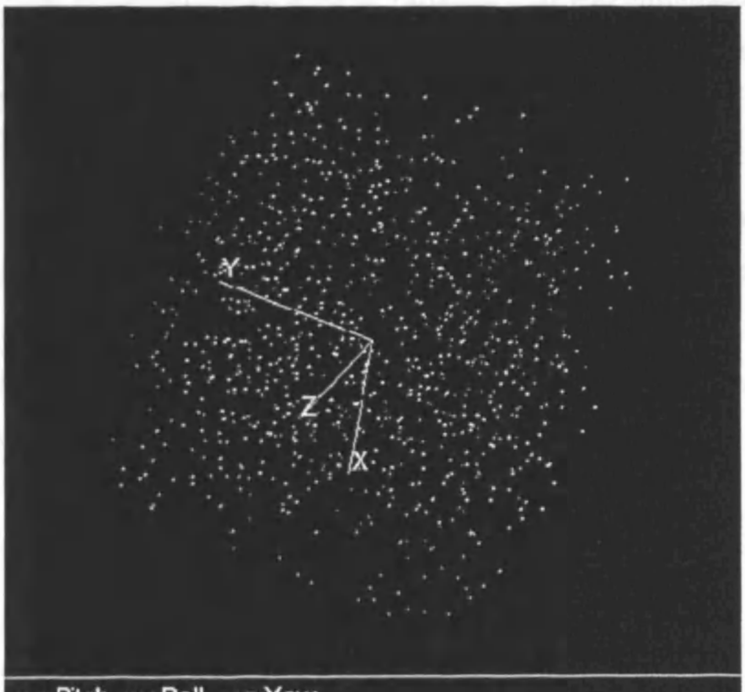


Figure A7.2.4: A second view of 1000 uniform random numbers

A7.2.2 Period test

The Period test is a predicate designed to test if the sequence of numbers begins to repeat (Jennings and Mohan, 1991). This is determined by matching the first 7 numbers in the repeating sequence.

$Yi = Ynp+i; \forall i = 1,7$

The test matches more than one number because there is a small probability that any one number may repeat randomly within a sequence. The probability that 10 random numbers [0,9] will appear in order in any one sequence is $(0.1)^k$, where k is the number of the repeating numbers in the sequence. There is, however, essentially no chance $[(0.1)^k \rightarrow 0]$ that a sequence of 7 numbers will randomly repeat.

In terms of Lisp-Stat implementation, a set of variables first was initialised: Count7 as a loop counter to hold the number of the first 7 iteration; first7 as a list of the first seven numbers; Next7 a list Next 7 numbers (see Function *init*).

We then define the function period_test with argument of number of iterations (cycle) defined by the user. When the function period_test is run, it will try to match the first 7 numbers of the sequence. If they match, the function will print the 'cyclic_property_occurs' message, else it return nil (see Function *period_test* Appendix 7.2.5). For example, when cycle = 10000:

```
> (period_test 10000)
NIL
>
; This took about 20s for 10000 numbers
```

A7.2.3 Chi-square (χ^2) test

Chi-square (χ^2) analysis, first introduced by Karl Pearson, is a more generalised procedure (Cochran, 1954). χ^2 is originally intended to test if observed outcomes are sufficiently random to imply that they come from unbiased samples. If {Yn} is a sequence of random variables, with mean :, following the Gaussian PDF such that

$$Z(\chi) = 2\P^{-1/2} \exp(-\chi^2/2) \qquad \textbf{(A7.1)}$$

then the Chi-square (χ^2) PDF can be defined as

$$\chi^2 = \frac{\sum_{i=1}^{n} (Y_i - \mu_i)^2}{\mu_i} \qquad \textbf{(A7.2)}$$

with

$$PDF = [2^{n/2}\,\Gamma(n/2)]^{-1} \chi^{2(n/2-1)} \exp(-\chi^2/2) \qquad \textbf{(A7.3)}$$

where n is the number of degrees of freedom;
$\Gamma(\ )$ the Gamma function.

The above definition can easily be defined as a Lisp-Stat function (see Function *chi_square*). This definition of χ^2 gives a positive, semi-infinite random variable $\chi^2 \in [0,\infty]$ that can only take on its lower bound when $Yi = 0$, $\forall i = 1,n$

The χ^2 variate has

- a mean $\mu = n$;
- a standard deviation of 2μ;
- a central tendency of χ^2_{ct}, $[p(\chi^2_{ct} | \mu) = 0.5]$; and
- is usually skewed to the right.
- In general, $\chi^2_{ct} \neq \mu$ but $\chi^2_{ct} \to \mu$ as $n \to \infty$.

For any sequence of numbers, one can compute the probability 1 - p of the resulting χ^2 value falling at least the observed distance away from the mean. As 1 - p becomes small, it implies that the observed sequence Yi is likely to represent a biased set of random outcomes. The test can also be repeated to determine the probability that a sequence of random numbers is unbiased and sufficiently random, and the generator can be selected or rejected on this basis.

For Lisp-Stat implementation to run the test, a function 'run_test' taking a list of numbers n as its argument, is defined in Function *run_test* (Appendix 7.2.5). This represents the following steps:

1. Initialise variables (init).
2. Create a list of random number r_list with n numbers (list_rand n n).
3. Random sample from the list with defined constraint (my_sample r_list).
4. Set r_list = sample_list (setq r_list sample_list).
5. Run χ^2 with d.f. = length of list - 1 (chisq_test (- (length r_list) 1))

The above steps consists of the following sub-functions:
Init: Initially, r_list and sample_list are nil, and χ^2 (x_sq)= 0.
List_rand; and
Chisq_test: Compute χ^2, print out the values of χ^2 at 5%, the level of sample, 95%, and the χ^2 probability respectively.

For all tests, n must be large enough to yield stable results. The minimum value of n should be at least 5 as suggested by Knuth (1969). For this application, first, 10,000 random numbers are generated by the function 'list_rand'. the numbers Y_i are then sampled which satisfy

$$n/j - 1 \leq Zi < n/j \qquad (j = 1,2,\text{---}, 1000)$$

This gives a set of 8 numbers in the run-test (which still satisfies the requirement of n > 5 but makes the χ^2 test feasible and reduces the run time enormously). The function then proceeds to compute the statistic for the sample on the assumption that the Zi's are independently and uniformly distributed. Since the only automatically satisfied condition is the total entries of 10,000, the χ^2 distribution for 7 degree of freedom is computed to find bounds between which χ^2 should lie. Ideally the value of χ^2 should lie between 5% and 95% level with $p({\chi^2}_{ct} | \mu) \approx 0.5$. Indeed, these results were found from the test.

A7.2.4 Conclusion of the random number testing

The random number generator was tested using Lisp-Stat function to generate random numbers on a Digital DECpc Lpx 466d2 microcomputer. If one adopts a criterion such as 1 - p > 0.05 => 'pass' (i.e. the probability that the generator is unbiased > 0.95), then the generator passed the test. Indeed it was found that the value of χ^2 lies between 5% and 95% level (2.17 < 5.92 < 14.07) at d.f.=7 with $p(\chi^2_\alpha \mid \mu) \approx 0.5$. For the 10,000 random numbers generated, the sequence did not repeat itself. Visually, the numbers appear to be randomly and uniformly distributed.

A7.2.5 Implementation of the testing procedure in XLISP-STAT

The specification of Lisp-Stat is implemented using a system called XLISP-STAT based on the XLISP (dialect of Lisp) with the window system (Tierney, 1990). The following test functions are listed within the XLISP-STAT environment.

```
XLISP-PLUS version 2.1g
Portions Copyright (c) 1988, by David Betz.
Modified by Thomas Almy and others.
XLISP-STAT 2.1 Release 3.45 (Beta).
Copyright (c) 1989-1994, by Luke Tierney.
Initialization may take a moment.

; Comments are after a semicolon.
; '>' is the XLISP-STAT prompt.
; The output from LISP-STAT is in block capital.

; First initialize the variables
; count7 is the count of the first seven numbers called first7
; next7 is the next seven numbers.

> (defun init ()
 (setq count7 1)
(setq first7 nil)
(setq next7 first7)
 (print count7)
(print first7)
(print next7))
INIT

> (defun period_test (cycle)
  (do ((count 0 (+ count 1)))
      ((equal count cycle) nil)
```

```
(setq *random-state* #$(1 #(2147483562 1342567740 491571990 1296487812)))
(setq z (random cycle))
  (cond ((< count 7) (setq first7 (cons z first7)) (setq next7 first7) )
       ((< count7 8) (setq next7 (cons z (butlast next7))) (setq count7 (+ count7 1)) )
       (T T))
   (cond ((> count7 7) (setq count7 1))
      (T T))
  (cond ((< count 7) nil)
         ((equal next7 first7) (print 'cyclic_property_occurs))
       (T T))))
PERIOD_TEST

> (defun chi_square (list_no)
    (do ((x list_no (cdr x)))
        ((null x))
    (setq x_sq (+ x_sq (/ (sq (- (car x) (mean list_no))) (mean list_no))))))
CHI_SQUARE

; where sq (square) can be defined as:

> (defun sq (x)
   (* x x))
SQ
> (defun run_test (n)
   (init)
   (list_rand n n)
   (my_sample r_list)
   (setq r_list sample_list)
    (chisq_test (- (length r_list) 1)))
RUN_TEST

> (defun init ()
   (setq sample_list nil)
   (setq r_list nil)
   (setq x_sq 0))
INIT
>
> (defun list_rand (range cycle)
  (let ((r_list NIL)))
   (dotimes (count cycle)
(setq *random-state* #$(1 #(2147483562 1955864722 1011176338 1202795178)))
(setq r_list (cons (random range) r_list)) ))
LIST_RAND
> (defun my_sample (my_list)
   (do ((x my_list (cdr x)))
        ((null x) )
  (cond ((< (car x) 10) (setq sample_list (cons (car x) sample_list)))
         (T nil))))
MY_SAMPLE
> (defun chisq_test (df)
   (print (chisq-quant .05 df))
   (print (chi_square r_list))
   (print (chisq-quant .95 df))
       (chisq-cdf x_sq df))
CHISQ_TEST
```

```
>
> (run_test 10000)

2.1673499091965045
5.923076923076923
14.067140449703 25
0.45124048214395046
>
```

; Random Numbers

; recursion
; A recursion to generate random numbers [ranged to the number of cycles; initial count 0].

```
> (defun generate_rand (count cycle)
   (cond ((equal count cycle) nil)
         (T (setq z (random cycle)) (print z) (generate_rand (+ count 1) cycle))))
GENERATE_RAND
> (generate_rand 0 10)

8
5
2
7
9
9
4
1
2
2
NIL
>

> (generate_rand 0 74)
NIL
> (generate_rand 0 75)
error: system stack overflow
>

(generate_rand 0 10)
NIL
> (generate_rand 1 75)
NIL
```

; limit to 75 cycles

;To initialise the random state

```
(setq rst (make-random-state T))
#$(1 #(2147483562 0 12285 31861))
```

; which changes each time as a side effect from RANDOM.

```
> (random 5)
4
```

```
> (random 5)
0

;:
;:
 (print rst)

#$(1 #(2147483562 1342567740 491571990 1296487812))
#$(1 #(2147483562 1342567740 491571990 1296487812))

> (setq *random-state* #$(1 #(2147483562 1342567740 491571990 1296487812)))
#$(1 #(2147483562 1342567740 491571990 1296487812))

> (defun generate_rand (count cycle)
   (cond ((equal count cycle) nil)
         (T (setq *random-state* #$(1 #(2147483562 1342567740 491571990 1296487812))) (setq z
(random cycle)) (print z) (generate_rand (+ count 1) cycle))))
GENERATE_RAND

; Need to be defined again each time to ensure the same random number - i.e. to start with the same seed.

> (defun generate_rand (count cycle)
   (cond ((equal count cycle) nil)
         (T (setq *random-state* #$(1 #(2147483562 1342567740 491571990 1296487812))) (setq z
(random cycle)) (print z) (generate_rand (+ count 1) cycle))))
GENERATE_RAND
>  (generate_rand 0 5)

3
0
2
0
4
NIL
> (defun generate_rand (count cycle)
   (cond ((equal count cycle) nil)
         (T (setq *random-state* #$(1 #(2147483562 1342567740 491571990 1296487812))) (setq z
(random cycle)) (print z) (generate_rand (+ count 1) cycle))))
GENERATE_RAND
> (generate_rand 0 5)

3
0
2
0
4
NIL
>
```

; Iteration

```
> (defun generate_rand (cycle)
   (do ((count 0 (+ count 1)))
       ((equal count cycle) nil)
(setq *random-state* #$(1 #(2147483562 1342567740 491571990 1296487812)))
```

```
(setq z (random cycle))
(print z)))
GENERATE_RAND

> (generate_rand 5)

3
0
2
0
4
NIL
> (defun generate_rand (cycle)
   (do ((count 0 (+ count 1)))
        ((equal count cycle) nil)
(setq *random-state* #$(1 #(2147483562 1342567740 491571990 1296487812)))
(setq z (random cycle))
(print z)))
GENERATE_RAND
> (generate_rand 5)

3
0
2
0
4
NIL
>
```

;No problem with the stack overflow limit, note the PRINT is removed for obvious reason.

```
> (generate_rand 77)
NIL
> (generate_rand 77)
NIL
> (generate_rand 777)
NIL
> (generate_rand 10000)
NIL
>
```

; Test cycliability

```
> (defun generate_rand (cycle)
   (do ((count 0 (+ count 1)))
        ((equal count cycle) nil)
(setq *random-state* #$(1 #(2147483562 1342567740 491571990 1296487812)))
(setq z (random cycle))
(print z)))
GENERATE_RAND

> (defun init ()
 (setq count7 1)
(setq first7 nil)
(setq next7 first7)
```

```
 (print count7)
(print first7)
(print next7))
INIT

> (defun test7 (z cycle)
  (do ((count 0 (+ count 1)))
      ((equal count cycle) nil)
   (cond ((< count 7) (setq first7 (cons z first7))(print 'first7) (print first7))
        ((< count7 8) (setq next7 (cons z next7)) (setq count7 (+ count7 1)) (print 'next7) (print next7))
        ((> count7 7) (setq count7 1) (setq next7 nil))
        (T T))
   (cond ((equal next7 first7) (print 'cyclic_property_occurs))
   (T T))))
TEST7
> (init)

0
NIL
NIL
NIL
> (test7 1 23)

FIRST7
(1)
FIRST7
(1 1)
FIRST7
(1 1 1)
FIRST7
(1 1 1 1)
FIRST7
(1 1 1 1 1)
FIRST7
(1 1 1 1 1 1)
FIRST7
(1 1 1 1 1 1 1)
NEXT7
(1)
NEXT7
(1 1)
NEXT7
(1 1 1)
NEXT7
(1 1 1 1)
NEXT7
(1 1 1 1 1)
NEXT7
(1 1 1 1 1 1)
NEXT7
(1 1 1 1 1 1 1)
CYCLIC_PROPERTY_OCCURS
NEXT7
(1 1 1 1 1 1 1 1)
NEXT7
```

```
(1)
NEXT7
(1 1)
NEXT7
(1 1 1)
NEXT7
(1 1 1 1)
NEXT7
(1 1 1 1 1)
NEXT7
(1 1 1 1 1 1)
NEXT7
(1 1 1 1 1 1 1)
CYCLIC_PROPERTY_OCCURS
NIL
>

> (defun generate_rand (cycle)
 (do ((count 0 (+ count 1)))
     ((equal count cycle) nil)
(setq *random-state* #$(1 #(2147483562 1342567740 491571990 1296487812)))
(setq z (random cycle))
   (cond ((< count 7) (setq first7 (cons z first7)) (setq next7 first7) )
       ((< count7 8) (setq next7 (cons z (butlast next7))) (setq count7 (+ count7 1)) )
       (T T))
    (cond ((> count7 7) (setq count7 1))
       (T T))
   (cond ((< count 7) nil)
         ((equal next7 first7) (print 'cyclic_property_occurs))
       (T T))))
GENERATE_RAND

> (generate_rand 26)
NIL
> (init)

1
NIL
NIL
NIL
> (generate_rand 10000)
NIL
>
;Took about 20s for 10000 numbers
;testing the first 28 random numbers.

> (defun generate_rand (cycle)
 (do ((count 0 (+ count 1)))
     ((equal count cycle) nil)
(setq *random-state* #$(1 #(2147483562 1342567740 491571990 1296487812)))
(setq z (random cycle))
   (print z)
   (cond ((< count 7) (setq first7 (cons z first7)) (setq next7 first7) (print 'first7) (print first7) )
       ((< count7 8) (setq next7 (cons z (butlast next7))) (setq count7 (+ count7 1))(print 'next7) (print
next7) )
```

```
      (T T))
   (cond ((> count7 7) (setq count7 1))
      (T T))
  (cond ((< count 7) nil)
        ((equal next7 first7) (print 'cyclic_property_occurs))
      (T T))))
GENERATE_RAND
```

; need to re-'defun' generaye_rand containing the *random-state* to re-run the function.

```
> (init)
1
NIL
NIL
NIL
> (generate_rand 26)

18
FIRST7
(18)
2
FIRST7
(2 18)
14
FIRST7
(14 2 18)
1
FIRST7
(1 14 2 18)
21
FIRST7
(21 1 14 2 18)
25
FIRST7
(25 21 1 14 2 18)
7
FIRST7
(7 25 21 1 14 2 18)
15
NEXT7
(15 7 25 21 1 14 2)
25
NEXT7
(25 15 7 25 21 1 14)
10
NEXT7
(10 25 15 7 25 21 1)
0
NEXT7
(0 10 25 15 7 25 21)
20
NEXT7
(20 0 10 25 15 7 25)
23
NEXT7
```

```
(23 20 0 10 25 15 7)
3
NEXT7
(3 23 20 0 10 25 15)
20
NEXT7
(20 3 23 20 0 10 25)
17
NEXT7
(17 20 3 23 20 0 10)
4
NEXT7
(4 17 20 3 23 20 0)
6
NEXT7
(6 4 17 20 3 23 20)
5
NEXT7
(5 6 4 17 20 3 23)
10
NEXT7
(10 5 6 4 17 20 3)
18
NEXT7
(18 10 5 6 4 17 20)
14
NEXT7
(14 18 10 5 6 4 17)
1
NEXT7
(1 14 18 10 5 6 4)
4
NEXT7
(4 1 14 18 10 5 6)
6
NEXT7
(6 4 1 14 18 10 5)
5
NEXT7
(5 6 4 1 14 18 10)
NIL
>
; Similarly, for uniform random number, replace random with rand-uniform

 (defun generate_rand_uniform (cycle)
   (do ((count 0 (+ count 1)))
       ((equal count cycle) nil)
(setq *random-state* #$(1 #(2147483562 685461157 730303945 44842788)))
(setq z (uniform-rand 1))
(print z)))
GENERATE_RAND_UNIFORM
> (generate_rand_uniform 5)

(0.020437848631456576)
(0.9623325209866916)
```

```
(0.899333927969133)
(0.32823186502131513)
(0.9508440423670703)
NIL
> (defun generate_rand_uniform (cycle)
   (do ((count 0 (+ count 1)))
       ((equal count cycle) nil)
(setq *random-state* #$(1 #(2147483562 685461157 730303945 44842788)))
(setq z (uniform-rand 1))
(print z)))
GENERATE_RAND_UNIFORM
> (generate_rand_uniform 5)

(0.020437848631456576)
(0.9623325209866916)
(0.899333927969133)
(0.32823186502131513)
(0.9508440423670703)
NIL
>

 (defun generate_rand_uniform (cycle)
   (do ((count 0 (+ count 1)))
       ((equal count cycle) nil)
(setq *random-state* #$(1 #(2147483562 685461157 730303945 44842788)))
(setq z (car (uniform-rand 1)))
(print z)))
GENERATE_RAND_UNIFORM
> (generate_rand_uniform 18)

0.020437848631456576
0.9623325209866916
0.899333927969133
0.32823186502131513
0.9508440423670703
0.2604499287564407
6.512096409337872E-4
0.023054245374461 25
0.5281240473818933
0.9383118980176751
0.8346784901397056
0.3661913937473534
0.6783505047816244
0.28081100846503465
0.602064884337613
0.8777487159573908
0.5707873709000508
0.9317943179808009
NIL
>

> (defun generate_rand_uniform (cycle)
  (do ((count 1 (+ count 1)))
      ((equal count cycle) nil)
(setq *random-state* #$(1 #(2147483562 685461157 730303945 44842788)))
```

```
(setq z (car(uniform-rand 1)))
  (print z)
  (cond ((< count 8) (setq first7 (cons z first7)) (setq next7 first7) (print 'first7) (print first7) )
      ((< count7 8) (setq next7 (cons z (butlast next7))) (setq count7 (+ count7 1))(print 'next7) (print
next7) )
      (T T))
   (cond ((> count7 7) (setq count7 1))
     (T T))
  (cond ((< count 8) nil)
        ((equal next7 first7) (print 'cyclic_property_occurs))
     (T T))))
GENERATE_RAND_UNIFORM
> (init)

1
NIL
NIL
NIL
> (generate_rand_uniform 18)

0.020437848631456576
FIRST7
(0.020437848631456576)
0.9623325209866916
FIRST7
(0.9623325209866916 0.020437848631456576)
0.899333927969133
FIRST7
(0.899333927969133 0.9623325209866916 0.020437848631456576)
0.32823186502131513
FIRST7
(0.32823186502131513 0.899333927969133 0.9623325209866916 0.020437848631456576)
0.9508440423670703
FIRST7
(0.9508440423670703 0.32823186502131513 0.899333927969133 0.9623325209866916
0.020437848631456576)
0.2604499287564407
FIRST7
(0.2604499287564407 0.9508440423670703 0.32823186502131513 0.899333927969133
0.9623325209866916 0.020437848631456576)
6.512096409337872E-4
FIRST7
(6.512096409337872E-4 0.2604499287564407 0.9508440423670703 0.32823186502131513
0.899333927969133 0.9623325209866916 0.020437848631456576)
0.02305424537446125
NEXT7
(0.02305424537446125 6.512096409337872E-4 0.2604499287564407 0.9508440423670703
0.32823186502131513 0.899333927969133 0.9623325209866916)
0.5281240473818933
NEXT7
(0.5281240473818933 0.02305424537446125 6.512096409337872E-4 0.2604499287564407
0.9508440423670703 0.32823186502131513 0.899333927969133)
0.9383118980176751
NEXT7
```

```
(0.9383118980176751 0.5281240473818933 0.02305424537446125 6.512096409337872E-4
0.2604499287564407 0.9508440423670703 0.3282318650213151 3)
0.8346784901397056
NEXT7
(0.8346784901397056 0.9383118980176751 0.5281240473818933 0.02305424537446125
6.512096409337872E-4 0.2604499287564407 0.9508440423670703)
0.3661913937473534
NEXT7
(0.3661913937473534 0.8346784901397056 0.9383118980176751 0.5281240473818933
0.02305424537446125 6.512096409337872E-4 0.2604499287564407)
0.6783505047816244
NEXT7
(0.6783505047816244 0.3661913937473534 0.8346784901397056 0.9383118980176751
0.5281240473818933 0.02305424537446125 6.512096409337872E-4)
0.2808110084650346 5
NEXT7
(0.2808110084650346 5 0.6783505047816244 0.3661913937473534 0.8346784901397056
0.9383118980176751 0.5281240473818933 0.02305424537446125)
0.602064884337613
NEXT7
(0.602064884337613 0.2808110084650346 5 0.6783505047816244 0.3661913937473534
0.8346784901397056 0.9383118980176751 0.5281240473818933)
0.8777487159573908
NEXT7
(0.8777487159573908 0.602064884337613 0.2808110084650346 5 0.6783505047816244
0.3661913937473534 0.8346784901397056 0.9383118980176751)
0.5707873709000508
NEXT7
(0.5707873709000508 0.8777487159573908 0.602064884337613 0.2808110084650346 5
0.6783505047816244 0.3661913937473534 0.8346784901397056)
NIL
>
> (defun generate_rand_uniform (cycle)
  (do ((count 1 (+ count 1)))
      ((equal count cycle) nil)
(setq *random-state* #$(1 #(2147483562 685461157 730303945 44842788)))
(setq z (car(uniform-rand 1)))
   (cond ((< count 8) (setq first7 (cons z first7)) (setq next7 first7) )
        ((< count7 8) (setq next7 (cons z (butlast next7))) (setq count7 (+ count7 1)) )
        (T T))
    (cond ((> count7 7) (setq count7 1))
        (T T))
   (cond ((< count 8) nil)
          ((equal next7 first7) (print 'cyclic_property_occurs))
        (T T))))
GENERATE_RAND_UNIFORM
> (init)

1
NIL
NIL
NIL
> (generate_rand_uniform 10000)
NIL
; took about 20s.
```

```
(defun test7 (z cycle)
 (do ((count 0 (+ count 1)))
     ((equal count cycle) nil)
   (cond ((< count 7) (setq first7 (cons z first7))(print 'first7) (print first7))
       ((< count7 8) (setq next7 (cons z next7)) (setq count7 (+ count7 1)) (print 'next7) (print next7))
       ((> count7 7) (setq count7 1) (setq next7 nil))
       (T T))
   (cond ((equal next7 first7) (print 'cyclic_property_occurs))
   (T T))))
```

; Testing the program
; Just to show the test does work if cyclic property occurs

```
> (defun init ()
 (setq count 0)
 (setq count7 1)
(setq first7 nil)
(setq next7 nil)
 (print count7)
(print first7)
(print next7))
INIT

>  (defun test7 (sys_list)
 (do ((z sys_list (cdr z)))
     ((null z) )
   (setq count (+ count 1))
   (cond ((< count 8) (setq first7 (cons (car z) first7)) (setq next7 first7)(print 'first7) (print first7 ))
       ((< count7 8) (setq next7 (cons (car z) (butlast next7))) (setq count7 (+ count7 1)) (print 'next7) (print
next7))
       (T T))
    (cond ((> count7 7) (setq count7 1))
      (T T))
   (cond ((< count 8) nil)
         ((equal next7 first7) (print 'cyclic_property_occurs))
       (T T))))
TEST7
> (init)
1
NIL
NIL
NIL
> (test7 (repeat (iseq 10) 2))

FIRST7
(0)
FIRST7
(1 0)
FIRST7
(2 1 0)
FIRST7
(3 2 1 0)
FIRST7
(4 3 2 1 0)
```

```
FIRST7
(5 4 3 2 1 0)
FIRST7
(6 5 4 3 2 1 0)
NEXT7
(7 6 5 4 3 2 1)
NEXT7
(8 7 6 5 4 3 2)
NEXT7
(9 8 7 6 5 4 3)
NEXT7
(0 9 8 7 6 5 4)
NEXT7
(1 0 9 8 7 6 5)
NEXT7
(2 1 0 9 8 7 6)
NEXT7
(3 2 1 0 9 8 7)
NEXT7
(4 3 2 1 0 9 8)
NEXT7
(5 4 3 2 1 0 9)
NEXT7
(6 5 4 3 2 1 0)
CYCLIC_PROPERTY_OCCURS
NEXT7
(7 6 5 4 3 2 1)
NEXT7
(8 7 6 5 4 3 2)
NEXT7
(9 8 7 6 5 4 3)
NIL
>
```

Appendix 7.3
Program listing

The Lisp-Stat functions are stored in the following files (in bold) and directories (in front of the file names):

- C:\USERS\LAW\GIS\HO\DATA**proBED.lsp**
- C:\USERS\LAW\GIS\HO**PROBBT.LSP**
- C:\HO\PHD**FPRO.LSP**
- C:\USERS\LAW\GIS\HO\DATA**MONTEK.LSP**
- C:\USERS\LAW\GIS\HO\DATA**MCBH.lsp**
- C:\USERS\LAW\GIS\HO\DATA**POSTBBT.LSP**

The files are directly loaded into the XLISP-STAT environment based on the XLISP (dialect of Lisp) with the window system (Tierney, 1990). **ProBED.lsp** contains the functions for counting pixels (p) of households in each ED (first index in the id_list).

The number of households per ED is retrieved from the geographical outcome base (GOB) using the functions specified in **PROBBT.LSP** (get_values_from_keys bedgs-id bed_pid where a list of pairs: number of households & identifier; and bed_pid is a list of pairs: number of pixels & identifier).

Lists of pixels are read into files using functions in **FPRO.LSP**. Monte Carlo simulation is performed using functions in **MONTEK.LSP**. **MCBH.lsp** contains file-handling functions to process a list of files using the *Monte* Carlo function developed in **MONTEK.LSP**. The final output functions are specified in **POSTBBT.LSP** - computing the number of households per pixel (r); RMSE; error %; and printing the results. Their uses are discussed in the main text (also see comments after the semicolon in the listing and Appendix 7.5 which shows the complete log for processing).

```
; C:\USERS\LAW\GIS\HO\DATA\proBED.lsp

; Initially, ed_id list is empty

(def ed_id nil)

(defun init_proed ()
 (def ed_id nil)
)

; Create the data structure: a list of (value index) lists.
; Initially, the value of each index is 0.

(defun create_list0 (n)
    (dotimes (i n)
      (def ed_id (cons (list 0 i) ed_id))))

; Function to assign pixel to household.

(defun assignh (p id_list)
 (cond ((= p 0) NIL)
         (T (match_fill id_list p))))

; If the pixel (p) matches the index (ED-ID) in the list, increment the value (first index) of the zone

(defun match_fill (id_list p)
     (dolist (index id_list)
      (cond ((= p (car (last index))) (def (first index) (+ 1 (first index))))
             (T NIL))))

; Process ED, here goes the following steps:
; 1. Define an empty ed_id list
; 2. Create the data structure
; 3. Iteratively read in a pixel from the image file and
; 4. assign it.

(defun proed_go (in_image_file n np)
(def ed_id nil)
 (create_list0 (+ 1 n))
 (def ed_id (cdr (reverse ed_id)))
(dotimes (i np)
(def p (read in_image_file nil))
 (assignh p ed_id)
))
```

; C:\USERS\LAW\GIS\HO\PROBBT.LSP

```
; Initialise all the variable used

(def i 0)
(def found nil)
(def found1 nil)
(def true_flag nil)
(def found_attributes nil)
(def value nil)
(def found_value nil)

; Function to initialise a list of beats

(defun init_pre_beat_go ()
  (def i 0)
(def found nil)
(def found1 nil)
(def true_flag nil)
(def found_attributes nil)
(def value nil)
(def found_value nil)
 )

; Set counter i, start from zero

(setq i 0)

; Get geographical outcome base (GOB) from the INFO entity.

(defun get_gob (entity)
     (do ((my_list entity (cdr my_list)))
      ((null my_list) )
       (def attribute (car my_list))
        (def i (+ i 1))
         (cond ((null entity) nil)
             ((= i 1) (def ed-id attribute))
             ((= i 2) (def edgs-id attribute))
             ((= i 3) (def beatgs-id attribute))
             ((= i 4) (def beat-id attribute))
              ((= i 5) (def superbt-id attribute))
               ((= i 6) (def ward-id attribute))
               ((= i 7) (def residents attribute))
                ((= i 8) (def households attribute))
                (T  (def attribute (cdr entity))))))

; Get the the values of the attributes from many_entity for one_entity in a many-to-one relation

(defun get_many_from_one (one_entity many_entity key)
     (dotimes (i (length one_entity))
        (def one_attribute (car one_entity))
        (def many_attribute (car many_entity))
        (cond ((eq one_attribute key) (def found (cons many_attribute found)))
                (T nil))
         (def one_entity (cdr one_entity))
```

```
       (def many_entity (cdr many_entity))
      ))

; Function to match the key with the attribute of the relation entity
; if match, define true_flag True.

 (defun match_keys_p (attribute key_entity)
     (dolist (key key_entity true_flag)
        (cond ((equal attribute key) (def true_flag T))
              (T nil))
        ))

; Get the the values of the attributes from the entity for the key_entity in a many-to-many relation

 (defun get_many_from_many (keys key_entity entity)
     (dotimes (i (length entity))
       (def key_attribute (car key_entity))
       (def attribute (car entity))
       (cond ((match_keys_p key_attribute keys) (def found_attributes (cons attribute
found_attributes)) (def  true_flag nil))
             (T nil))
        (def key_entity (cdr key_entity))
        (def entity (cdr entity))
      ))

(defun get_values_from_keys (keys key_entity)
     (dolist (value_key key_entity)
       (def key (second value_key))
       (def value (car value_key))
       (cond ((match_keys_p key keys) (def found_value (cons value found_value)) (def
true_flag nil))
             (T nil))
      ))

; Procedure to get the values of interest from ED zone for beats

(defun pre_beat_go (beat-id)
 (init_pre_beat_go)
(def beateds (read-data-columns "cedbeat.dat" 4))
(get_attributes beateds)
(get_many_from_one beatgs-id edgs-id beat-id)
(def bedgs-id found)
(setq found1 found)
(def i 0)
(def cgobs (read-data-columns "cgob.dat" 8))
(get_gob cgobs)
(get_many_from_many found ed-id residents)
(setq bbt_residents found_attributes)
(def found_attributes nil)
(get_many_from_many found ed-id households)
(setq bbt_hh found_attributes)
(def true_flag nil)
(get_values_from_keys bedgs-id bed_pid))

(defun pro_beat_go (beat-id)
```

```
 (init_pre_beat_go)
(get_many_from_one beatgs-id edgs-id beat-id)
(def bedgs-id found)
(setq found1 found)
(get_many_from_many found ed-id residents)
(setq bbt_residents found_attributes)
(def found_attributes nil)
(get_many_from_many found ed-id households)
(setq bbt_hh found_attributes)
(def true_flag nil)
(get_values_from_keys bedgs-id bed_pid))
```

; C:\HO\PHD**FPRO.LSP**

```
; Read in a list of numbers and save it a file

(defun list_batch (file_list my_file list_length)
   (dolist (file_name file_list)
    (def p_list nil)
    (read_file_into_list my_file list_length)
     (print "Please wait...")
     (Princ "Saving this list into file ")
     (prin1 file_name)
     (save_my_var 'p_list file_name)))

; Read a file of number into a list

(defun read_file_into_list (my_file list_length)
   (dotimes (i list_length)
    (def p (read my_file nil))
    (setq p_list (cons p p_list))))

; Save my variable as my file

 (defun save_my_var (my_var my_file)
    (savevar my_var my_file))

; Initially, the list of numbers (pixels) is empty

(def p_list nil)
```

; C:\USERS\LAW\GIS\HO\DATA**MONTEK.LSP**

```
; Initially a list of counts is empty

(setq c_list nil)

; Get and set the current value of the random state from the global variable *random-state*

(setq *random-state* #$(1 #(2147483562 833502228 1548262346 714760118)))
```

```
; The batch function to initialise the count, the list of counts and the *random-state*

(defun init_monte_TM_batch ()
   (setq c 0)
   (setq c_list nil)
(setq *random-state* #$(1 #(2147483562 833502228 1548262346 714760118)))
)

; Run Monte Carlo function over a zone list
; count = count x scaling factor k

 (defun run_monte (k zone)
     (monte zone)
     (def c (* c k)))

; Initiase the Monte Carlo function: the count start from zero, the list of count empty, *random-state*
; and sample size r = size of zone / scaling factor

(defun init_monte (k zone)
   (setq c 0)
   (setq c_list nil)
(setq *random-state* #$(1 #(2147483562 833502228 1548262346 714760118)))
   (setq r (round (/ (length zone) k))))

; Monte Carlo function gets a random sample from a zone
; and assign the list of numbers from the random sample

(defun Monte (zone)
 (setq zlist (my_sample zone r))
 (assign zlist))

; Define the random sample from my list as my sample with n numbers

(defun my_sample (my_list n)
   (sample my_list n))

; Assign pixel numbers
; and increment the counter c if it is not zero

(defun assign (my_list)
    (dolist (p my_list)
    (cond ((= p 0) NIL)
          (T (setq c (+ 1 c))))))

; Compute error_% = (observed – actual )/ actual x 100%

(defun error_% (observed actual)
     (* (/ (- observed actual) actual) 100))

; Function to compute the total squared deviation from the mean

(defun sq_d (list_no mean_x)
     (do ((x list_no (cdr x)))
          ((null x) d_sq)
     (setq d_sq (+ d_sq (^ (- (car x) mean_x) 2)))))
```

```
; Compute the RMSE

(defun rmse (observed_list actual)
      (sqrt (/ (sq_d observed_list actual) (length observed_list))))

; Print the results of length, mean, error% of the Monte Carlo simulation, RMSE, SD, SD-1
; and save the list counts into a file called c_list

(defun print_results (actual save_file)
  (print 'length= ) (prin1 (length c_list))
  (print 'mean= ) (prin1 (mean c_list))
  (print 'error_MC%= ) (prin1 (error_% (mean c_list) actual))
  (setq d_sq 0)
  (print 'RMSE= ) (prin1 (rmse c_list actual))
  (setq d_sq 0)
  (print 'SD= ) (prin1 (rmse c_list (mean c_list)))
  (print ' SD-1=) (prin1 (standard-deviation c_list))
  (savevar 'c_list save_file)
)

; Batch processing of the Monte Carlo simulation
; Run Monte Carlo function over a zone
; and construct a list of counts c_list.

(defun monte_batch (n zone k)
    (dotimes (i n)
     (setq c 0)
     (run_monte k zone)
     (setq c_list (cons c c_list))))

; Batch processing of the Monte Carlo function for a list of files
; with scaling factor k as the passing parameter

(defun Monte_file_batch (file_list k)
    (dolist (file_name file_list)
      (load file_name)
       (setq r (round (/ (length p_list) k)))
      (monte p_list)))

; Process the wholeTM image file

(defun monte_TM_batch (n zone k)
    (dotimes (i n)
     (setq c 0)
      (print "Please wait...")
       (Princ "Processing Monte Carlo simulation run ")
       (prin1 i)
     (monte_file_batch zone k)
     (setq c_list (cons (* k c) c_list))))

; Define the base-line value from assigning the whole image for comparison

(defun base_value (image)
   (def zone (read-data-file image))
   (print 'length_of_zone_)
```

```
   (prin1 (length zone))
   (setq c 0)
   (assign zone)
   (setq c0 c)
   (print c0))
```

; Monte Carlo processing in a batch of 300 runs with scaling factor 10 over an image with an output file

```
(defun mc_go_go (image out_file)
   (base_value image)
   (monte_go 300 zone c0 10 out_file))
```

; Monte Carlo processing in a batch of n runs with scaling factor k over a zone with an output file

```
(defun monte_go (n zone actual_c k out_file)
(init_monte k zone)
(monte_batch n zone k)
(print_results actual_c out_file))
```

; C:\USERS\LAW\GIS\HO\DATA**MCBH.lsp**

; Define the data structure for a list of (count Beat-ID) called _hID
; initial count has a zero value

```
(def _hID '((0 1) (0 2) (0 3) (0 4) (0 5) (0 6) (0 7) (0 8) (0 9) (0 10) (0 11) (0 12)
(0 13) (0 14) (0 15) (0 16) (0 17) (0 18) (0 19) (0 20) (0 21) (0 22) (0 23) (0 24) (0 25) (0 26)
(0 27) (0 28) (0 29) (0 30) (0 31) (0 32) (0 33) (0 34) (0 35) (0 36) (0 37) (0 38) (0 39) (0 40)
(0 41) (0 42) (0 43) (0 44) (0 45) (0 46) (0 47) (0 48) (0 49) (0 50) (0 51) (0 52) (0 53) (0 54)
(0 55) (0 56) (0 57) (0 58) (0 59) (0 60) (0 61) (0 62) (0 63)))
```

; Batch processing of the Monte Carlo function for file handling

```
(defun monte_go (n file_list k)
    (dotimes (i n)
       (def _hID '((0 1) (0 2) (0 3) (0 4) (0 5) (0 6) (0 7) (0 8) (0 9) (0 10) (0 11) (0 12)
(0 13) (0 14) (0 15) (0 16) (0 17) (0 18) (0 19) (0 20) (0 21) (0 22) (0 23) (0 24) (0 25) (0 26)
(0 27) (0 28) (0 29) (0 30) (0 31) (0 32) (0 33) (0 34) (0 35) (0 36) (0 37) (0 38) (0 39) (0 40)
(0 41) (0 42) (0 43) (0 44) (0 45) (0 46) (0 47) (0 48) (0 49) (0 50) (0 51) (0 52) (0 53) (0 54)
(0 55) (0 56) (0 57) (0 58) (0 59) (0 60) (0 61) (0 62) (0 63)))
      (print "Please wait...")
      (Princ "Processing Monte Carlo simulation run ")
      (prin1 i)
      (init_fill _hid)
      (Monte_file_batch file_list _hid k)
      (def new_list nil)
       (cons_value_into_list _hid1 _hid)
      (def _hid1 new_list)
))
```

; Run Monte Carlo function n times over a zone with initial vid_list and output list1 as a new list

```
(defun monte1_go (n zone vid_list list1)
    (dotimes (i n)
      (init_fill vid_list)
      (Monte zone vid_list)
```

```
    (def list1 new_list)
    (def new_list nil)
     (cons_value_into_list list1 vid_list)))

; Initially, fill the index list with zero values

(defun init_fill (id_list)
    (dolist (index id_list)
       (def (first index) 0)))

; If an index match ID of list1, construct the new list with its value

(defun match_cons (list1 i)
   (dolist (v list1)
    (cond ((= (car (last i)) (car (last v))) (def new_list (cons (cons (car i) v) new_list)))
          (T NIL))))

; Construct the new list with the index and values from list1 and list2 respectively

(defun cons_value_into_list (list1 list2)
  (dolist (h list2)
    (match_cons list1 h)))

; Monte Carlo processing in a batch of files with scaling factor k

 (defun Monte_file_batch (file_list id_list k)
   (dolist (file_name file_list)
    (load file_name)
    (run_monte p_list id_list k)))

(defun run_monte (zone id_list k)
   (init_monte k zone)
   (monte zone id_list))

(defun init_monte (n zone)
  (setq r (round (/ (length zone) n))))

(defun Monte (zone id_list)
 (setq zlist (sample zone r))
 (assignh zlist id_list))

; Assign my list of numbers by matching the ID from id_list

(defun assignh (my_list id_list)
  (dolist (p my_list)
  (cond ((= p 0) NIL)
        (T (match_fill id_list p)))))

; if the number match  with ID, increment the count which is the first index in the list

(defun match_fill (id_list p)
    (dolist (index id_list)
    (cond ((= p (car (last index))) (def (first index) (+ 1 (first index))))
          (T NIL))))
```

```
; a list file names

(def pn_list '("bbt1f" "bbt2f" "bbt3f" "bbt4f" "bbt5f" "bbt6f" "bbt7f" "bbt8f" "bbt9f" "bbt10f"
"bbt11f" "bbt12f" "bbt13f" "bbt14f" "bbt15f" "bbt16f" "bbt17f" "bbt18f" "bbt19f" "bbt20f"
"bbt21f" "bbt22f" "bbt23f" "bbt24f" "bbt25f" "bbt26f" "bbt27f" "bbt28f" "bbt29f" "bbt30f"
"bbt31f" "bbt32f" "bbt33f" "bbt34f" "bbt35f" "bbt36f" "bbt37f" "bbt38f" "bbt39f"))

(setq *random-state* #$(1 #(2147483562 833502228 1548262346 714760118)))

; Define the data structure for a list of beat-ID

(def _hid1 '((1) (2) (3) (4) (5) (6) (7) (8) (9) (10) (11) (12)
(13) (14) (15) (16) (17) (18) (19) (20) (21) (22) (23) (24) (25) (26)
(27) (28) (29) (30) (31) (32) (33) (34) (35) (36) (37) (38) (39) (40)
(41) (42) (43) (44) (45) (46) (47) (48) (49) (50) (51) (52) (53) (54)
(55) (56) (57) (58) (59) (60) (61) (62) (63) ))

(defun monte_go1 ()
(Monte_file_batch pn_list _hid 10)
(def new_list _hid)
(monte_go 1 pn_list 10))


; C:\USERS\LAW\GIS\HO\DATA\POSTBBT.LSP

; Get attributes of interest from the entity

(defun get_attributes (entity)
    (do ((my_list entity (cdr my_list)))
     ((null my_list) )
      (def attribute (car my_list))
       (def i (+ i 1))
        (cond ((null entity) nil)
            ((= i 1) (def beat-id attribute))
            ((= i 2) (def residents attribute))
            ((= i 3) (def households attribute))
            ((= i 4) (def r/h attribute))
            ((= i 5) (def pixels attribute))
            ((= i 6) (def h/p attribute))
             ((= i 7) (def r/p attribute))
            (T  (def attribute (cdr entity))))))

; Define the source zone as
; the data file with the attributes of interest, their calibrated values such as households per pixel

(def source_zones (read-data-columns "bbt_rhp.dat" 7))

(setq i 0)

(get_attributes SOURCE_ZONES)

(defun get_bmc (entity)
    (do ((my_list entity (cdr my_list)))
     ((null my_list) )
```

```
    (def attribute (car my_list))
     (def i (+ i 1))
      (cond ((null entity) nil)
          ((= i 1) (def beat-id attribute))
          ((= i 2) (def c0 attribute))
          ((= i 3) (def c1 attribute))
          ((= i 4) (def mc%error attribute))
          ((= i 5) (def mc_rmsc attribute))
          ((= i 6) (def sd attribute))
          ((= i 7) (def sd-1 attribute))
          (T  (def attribute (cdr entity))))))

(setq i 0)

; Define target zone as
; the file containing Monte Carlo results

(def target_zone (read-data-columns "bbtmc.dat" 7))

(get_bmc target_zone)

, calibrating factor r = households / pixel
; household count c = pixel count x households / pixel

(defun scale_c (plist rlist)
   (dolist (p plist)
    (def r (car rlist))
     (def c (* p r))
     (def clist (cons c clist))
     (def rlist (cdr rlist))))

(def clist nil)

(scale_c c1 h/p)

(def clist (reverse clist))

(def hlist clist)
(def clist nil)

(scale_c mc%error h/p)
(def clist (reverse clist))
(def mc%errl clist)

(def clist nil)
(scale_c mc_rmsc h/p)
(def clist (reverse clist))
(def mcsdl clist)

(defun print_btmc%err (slist mclist mcsdlist mc%errlist)
   (dotimes (i (length mclist))
     (print (+ 1 i))
      (prin1 (car slist))
      (prin1 "  ")
      (prin1 (round_2d (car mclist))) (princ " ±") (prin1 (round_2d (car mcsdlist)))
```

```
    (prin1 " ")
    (cond ((= 0 (car mclist)) nil)
        (T (prin1 (round_2d (error_% (car slist) (car mclist))))))
    (princ " ±") (prin1 (round_2d (car mc%errlist)))
    (def slist (cdr slist))
    (def mclist (cdr mclist))
    (def mcsdlist (cdr mcsdlist))
    (def mc%errlist (cdr mc%errlist))))

(print_btmc%err households hlist mcsdl mc%errl)

(defun get_bbtdat (entity)
    (do ((my_list entity (cdr my_list)))
     ((null my_list) )
      (def attribute (car my_list))
       (def i (+ i 1))
        (cond ((null entity) nil)
           ((= i 1) (def beat-no attribute))
           ((= i 2) (def beatgs-id attribute))
           ((= i 3) (def beat-id attribute))
           ((= i 4) (def residents attribute))
           ((= i 5) (def households attribute))
           (T  (def attribute (cdr entity))))))

; Get the counts from the overlay method in ARC/INFO geographical outcome base

(def arcinfo_h (read-data-columns "bbthgob.dat" 1))

(def arcinfo_h (car arcinfo_h))

; Print the estimated error % of the overlay method from arcinfo_h list of counts
; and a list of households called h_list as well as a list of SDs and % errors of the Monte Carlo method

(print_btmc%err arcinfo_h hlist mcsdl mc%errl)

(def target_zone (read-data-columns "bbth2d.lsp" 6))

(defun get_attributes (entity)
    (do ((my_list entity (cdr my_list)))
     ((null my_list) )
      (def attribute (car my_list))
       (def i (+ i 1))
        (cond ((null entity) nil)
           ((= i 1) (def record-id attribute))
           ((= i 3) (def mcbth attribute))
           (T  (def attribute (cdr entity))))))

(def i 0)

(get_attributes target_zone)

(def target_zone (read-data-columns "bbtmc.dat" 7))

(def i 0)
```

```
(get_bmc target_zone)
(defun round_2d (n) (/ (round (* n 100)) 100))

; Print results

(defun print_mcbt (idlist mclist)
  (dotimes (i (length mclist))
    (print (car idlist)) (princ ",")
     (prin1 (car mclist))
     (def idlist (cdr idlist))
     (def mclist (cdr mclist))))

(print_mcbt beat-id mcbth)
```

Appendix 7.4
Test runs for Monte Carlo simulation

This Appendix presents the log (Section A7.4.1) and the results (Section A7.4.2) of the test runs of the Monte Carlo experiments. An Enumeration District (ED) is regarded as an elementary zone. A list of EDs was randomly selected. A number of EDs adjacent to the selected EDs were then aggregated into the target zone using the ARC/INFO *Reselect Adjacent* command. The Monte Carlo simulation was carried out using these randomly aggregated target zones. From a series of the experiments, a normal distribution was obtained after 300 runs and the optimal value of k was found to be 10 when the RMSE was minimised and yet maintaining a normal distribution of the results.

A7.4.1 Random Selection of elementary zone for aggregation of target zone

XLISP-PLUS version 2.1g
Portions Copyright (c) 1988, by David Betz.
Modified by Thomas Almy and others.
XLISP-STAT 2.1 Release 3.45 (Beta).
Copyright (c) 1989-1994, by Luke Tierney.
Initialization may take a moment.

; Comments are after a semicolon.

; Select a list of EDs for random aggregation

; Technical note: Lisp-Stat could only process up to 4912 elements in a list. For example, for Bristol Landsat image, there were 201521 pixels containing 371020 population. In IDRISI, this can be stored as a single column ASCII file of integers. We can use Read-data-columns 1 function to define the city as a list.

```
 (setq Bristol (read-data-columns "outbri.img" 1))

> (length bristol)
1
> (length (cdr bristol))
0
> (length (car bristol))
4912

> (def Bristol (read-data-columns "outbri.img" 1))
BRISTOL
> (length (car bristol))
4912
>
```

; Thus the maximum limit for a list 4912. So a file processing function needd to be used...

; To select and aggregate a number of source zones, theoretically the number of source zones should be a fair representation of the whole city and approximately equal to the number of target zones (62?). This gives approximately 10-12 elementary units for each zone. At first, it was thought that this can be achieved by first randomly generate an Edgs-id. A list of its neighbourhood Eds can be built up by generating randomly a number with a random range of 13.

```
> (defun rand_agg (selected)
    (cond ((> (length agg_list) mean_n) agg_list)
          (T (setq agg_list (cons selected agg_list)) (rand_agg (+ selected (random mean_n))))))
RAND_AGG
> (defun init_agg (n)
   (setq mean_n 13)
   (setq selected (random n))
   (setq agg_list nil))
INIT_AGG
> (init_agg 824)
NIL
> selected
718
> (rand_agg selected)
(795 791 790 780 771 771 770 759 750 738 736 726 722 718)
>
```

; This gives a list of 13 EDs:
(795 791 790 780 771 770 759 750 738 736 726 722 718)

; However, this did not work because neighbourhood EDs did not necessarily follow the same order as the EDgs-ids. So a random number as a starting point to select an ED and then use the selected ED as a basis to create an aggregated zone using ARC/INFO Reselect Adjacent command.

; For 824 EDs, a simple random function can generate a number for selection.

```
> (random 824)
327

> (setq agg_list '(404 403 499 325 326 498 327 328 329 330))
(404 403 499 325 326 498 327 328 329 330)

> (sort-data agg_list)
(325 326 327 328 329 330 403 404 498 499)
>
```

; Note they are not in sequential order.

; A list of ED-IDs as a basis for aggregation can easily be produced (see Rand_list function)

```
> (def r_list nil)
R_LIST
> r_list
NIL
> (defun rand_list (n)
(setq *random-state* #$(1 #(2147483562 833502228 1548262346 714760118)))
```

```
(dotimes (i n)
 (setq r_list (cons (random 824) r_list))))
RAND_LIST
> (rand_list 10)
NIL
> r_list
(20 247 758 517 602 810 17 536 762 790)
>

; An aggregated zone (zone1) can be saved as a variable in Lisp-Stat using Savevar function.

> (savevar 'zone1 "zone1")
(ZONE1)
```

A7.4.2 Test results of the Monte Carlo experiment

The following table shows the target zone, household counts (Po) based on ED, household pixel counts based on Landsat (TM), household estimation based on the Monte Carlo (MC) method, its the error %, RMSE, SD, and SD-1 (for sampling).

Table A7.1 Test results of the Monte Carlo experiment

zone	Eds	Po (ED)	Po (TM)	%error(TM)	Po (MC)	%error(MC)	RMSE	SD	SD-1
1	10	3341	3919	-17.30	3874.53	-15.97	590.99	254.19	254.44
2	8	3636	3562	2.0	3916.15	-7.70	331.52	177.26	177.44
3	9	4706	4490	4.59	4764.38	-1.24	200.70	192.02	192.22
4	13	5544.0	5482	1.12	5143.52	+7.22	453.30	212.36	212.57
5	8	2856.0	3350	-17.30	3303.76	-15.68	527.06	278.03	278.31
6	6	3186.0	3096	2.82	3389.52	-6.39	268.47	175.08	175.25
7	7	3835.0	3742	2.44	3474.77	9.39	399.26	172.17	172.35
8	11	5338.0	5515	-3.32	5852.06	-9.63	558.32	217.87	218.09
9	10	6297	5302	15.81	4826.04	23.36	1519.46	380.85	381.23
10	14	7814.0	7295	6.63	7033.94	9.98	855.16	350.43	350.78

zone	Eds	Po (ED)	Po (TM)	%error(TM)		Po (MC)	%error(MC)	RMSE	SD	SD-1
1	10	3341	3919	-17.30	250	2175.79	34.88	1174.25	145.43	145.72
					500	2179.41	34.77	1170.29	142.51	142.65
2	8	3636	3562	2.0	250	4647.28	-27.81	1034.59	218.38	218.82
					500	4645.99	-27.77	1032.44	214.15	214.36
3	9	4706	4490	4.59	250	6214.49	-32.05	1531.36	263.65	264.18
					500	6204.86	-31.85	1523.04	270.33	270.60
4	13	5544.0	5482	1.12	250	4122.61	25.63	1430.85	164.25	164.58
					500	4125.61	25.58	1427.81	163.75	163.91
5	8	2856.0	3350	-17.30	250	1485.22	47.99	1376.43	124.68	124.93
					500	1494.42	47.67	1367.43	126.46	126.58
6	6	3186.0	3096	2.82	250	3782.01	-18.71	624.17	185.36	185.73
					500	3773.18	-18.43	619.39	197.14	197.34
7	7	3835.0	3742	2.44	250	4909.72	-28.02	1093.62	202.41	202.82
					500	4895.27	-27.64	1082.58	218.61	218.83

8	11	5338.0	5515	-3.32	250	6960.63	-30.40	1643.81	263.04	263.57
					500	6963.62	-30.45	1646.31	260.21	260.47
9	10	6297	5302	15.81	250	2267.124	64.00	4033.82	178.24	178.60
					500	2266.61	64.00	4034.18	174.75	174.93
10	14	7814.0	7295	6.63	250	4761.888	39.06	3060.01	219.78	220.22
					250k	4773.98	38.90	4997.46	3966.47	3974.43
				(2 x 250)	500	4736.74	39.38	3085.71	228.20	228.42
					2x250k	4617.52	40.91	4902.32	3716.90	3720.62
					500	4765.07	39.02	6192.15	5389.50	5394.90
					250	4934.23	36.85	3121.41	1204.22	1206.63
			4917.0	37.07	1					
					Runs					
T	822	371490.	373123	-0.44	101	370132.30	0.365	2171.79	1695.09	1703.54
					200	370033.03	0.390	2356.17	1851.70	1856.35
					254	370061.81	0.385	2333.55	1845.46	1849.10

zone	%error(TM)	Po (ED)	runs	k	Po (MC)	%error(MC)	RMSE	SD	SD-1
10	6.63	7814.0	500	+500	7033.94	9.98	855.16	350.43	350.78
			500	500	7070.48	9.52	8031.32	7996.83	8004.84
			250	250	7083.69	9.35	5930.65	5885.52	5897.33
			500	250	6851.55	12.32	5598.54	5515.20	5520.72
			250	100	7237.67	7.38	3482.62	3434.59	3441.49
			500	100	7211.11	7.72	3414.17	3360.52	3363.89
			250	50	7163.32	8.33	2730.45	2651.79	2657.10
			500	50	7219.05	7.61	2575.30	2505.63	2508.14
			100	25	7120.18	8.88	1900.09	1768.88	1777.79
			250	25	7321.18	6.31	1853.55	1786.84	1790.42
			500	25	7315.21	6.38	1806.50	1736.28	1738.02
			250	10	7235.49	7.40	1229.97	1085.42	1087.60
			500	10	7213.46	7.68	1249.05	1095.21	1096.31
			250	5	7246.13	7.23	870.73	660.06	661.39
			500	5	7218.39	7.62	897.38	671.22	671.89
			250	2	7269.26	6.97	644.32	344.10	344.79
			500	2	7280.39	6.83	638.57	350.78	351.14
9	15.81	6297	500	+500	4826.04	23.36	1519.46	380.85	381.23
			500	500	4946.61	21.44	9344.94	9246.85	9256.12
			250	500	4392.06	30.25	9015.11	8811.56	8829.23
			250	100	5696.30	9.54	4217.45	4174.45	4182.82
			500	100	5798.35	7.92	4321.65	4292.79	4297.09
			250	50	5518.94	12.35	3011.43	2909.18	2915.02
			500	50	5532.25	12.14	3016.26	2917.70	2920.62
			250	25	5685.31	9.71	2068.85	1976.36	1980.32
			500	25	5436.87	13.66	2174.26	1996.89	1998.89
			250	20	5337.91	15.23	1939.53	1685.80	1689.18
			500	20	5356.54	14.93	1970.81	1731.95	1733.69
			250	10	5330.84	15.34	1552.80	1215.61	1218.05
			500	10	5283.386	16.10	1559.30	1184.91	1186.09
			50	20	5341.54	15.17	2171.51	1950.01	1969.81
			100	20	5394.73	14.33	2019.97	1807.26	1816.37
			150	20	5409.50	14.09	1976.68	1766.25	1772.16
			200	20	5354.76	14.96	1963.41	1722.55	1726.87

			250	20	5337.91	15.23	1939.53	1685.80	1689.18
			300	20	5365.12	14.80	1954.24	1717.75	1720.62
			350	20	5363.00	14.83	1951.97	1714.01	1716.46
			400	20	5310.41	15.66	1987.48	1725.32	1727.47
			450	20	5324.70	15.44	1970.14	1713.51	1715.42
			500	20	5356.54	14.93	1970.81	1731.95	1733.69
			10	10	5057.5	19.68	1653.77	1094.80	1154.02
			20	10	5345.95	15.10	1515.42	1179.84	1210.49
			50	10	5496.78	12.70	1583.28	1366.18	1380.05
			100	10	5518.93	12.35	1576.47	1371.09	1378.00
			200	10	5354.80	14.96	1574.20	1261.09	1264.26
			250	10	5330.84	15.34	1552.80	1215.61	1218.05
			300	10	5296.38	15.89	1570.25	1210.14	1212.16
			400	10	5273.85	16.25	1579.47	1203.28	1204.79
			500	10	5283.386	16.10	1559.30	1184.91	1186.09

zone	%error(TM)	Po (ED)	runs	k	Po (MC)	%error(MC)	RMSE	SD	SD-1
3	4.59	4706	500	+500	4764.38	-1.24	200.70	192.02	192.22
			500	500	4566.02	2.97	4297.46	4295.18	4299.48
			250	10	4479.30	4.82	680.34	641.46	642.75
			300	**10**	**4486.61**	**4.66**	**680.80**	**644.49**	**645.57**
			500	10	4478.41	4.84	643.31	601.71	602.32
4	1.12	5544.0	500	+500	5143.52	+7.22	453.30	212.36	212.57
		->	**300**	**10**	**5555.12**	**-0.20**	**720.36**	**720.28**	**721.48**
			500	10	5531.066	0.23	685.64	685.52	686.20

(Base on the above results, number of runs 300 and scaling factor 10 were selected for Monte Carlo simulation.)

Appendix 7.5

Processing the Monte Carlo dasymetric method in XLISP-STAT

This Appendix shows the complete log for the implementation of the Monte Carlo dasymetric method for Bristol in the XLISP-STAT environment. Coventry was processed first, initially, beat by beat until the file-handling algorithm was developed (see Appendix 7.3). The processing represents the following steps:

1. Estimate the household density per ED using the classified imagery.
2. Calibrate the Landsat imagery by comparing its pixel counts with ED census counts and estimating the average number of households per pixel for each beat (h/p).
3. Perform Monte Carlo simulation to estimate the household counts per beat using h/p obtained from Step 2 as a parameter.

```
XLISP-PLUS version 2.1g
Portions Copyright (c) 1988, by David Betz.
Modified by Thomas Almy and others.
XLISP-STAT 2.1 Release 3.45 (Beta).
Copyright (c) 1989-1994, by Luke Tierney.
Initialization may take a moment.

; Comments are after a semicolon.

; Log files of processing error model for Bristol
; for Coventry, change first letter b into c, e.g. bbt becomes cbt

; 270599.log (Redo 071196)
; Assign pixels within each ED

> ; loading "C:\USERS\LAW\GIS\HO\PROBED.LSP"
 (setf f (open "wbrixed1.img"))
#<Input-Stream 5:"c:\users\law\gis\ho\wbrixed1.img">
> (proed_go f 825 190920)
NIL
> (length ed_id)
825
> (first ed_id)
(1061 1)
> (second ed_id)
(115 2)
> (last ed_id)
((1 825))
> (third ed_id)
(103 3)
> ed_id
```

((1061 1) (115 2) (103 3) (93 4) (64 5) (47 6) (39 7) (84 8) (62 9) (78 10) (69 11) (66 12) (65 13) (63 14) (94 15) (51 16) (57 17) (41 18) (25 19) (34 20) (61 21) (74 22) (38 23) (43 24) (85 25) (79 26) (46 27) (53 28) (56 29) (90 30) (53 31) (64 32) (26 33) (58 34) (47 35) (71 36) (2 37) (3 38) (1 39) (50 40) (30 41) (17 42) (0 43) (35 44) (52 45) (74 46) (60 47) (26 48) (210 49) (82 50) (95 51) (191 52) (0 53) (65 54) (43 55) (45 56) (51 57) (55 58) (53 59) (81 60) (52 61) (129 62) (75 63) (69 64) (27 65) (69 66) (65 67) (101 68) (34 69) (60 70) (44 71) (63 72) (91 73) (79 74) (71 75) (40 76) (65 77) (76 78) (54 79) (65 80) (55 81) (18 82) (82 83) (31 84) (24 85) (66 86) (81 87) (20 88) (91 89) (111 90) (35 91) (62 92) (62 93) (16 94) (32 95) (31 96) (55 97) (70 98) (44 99) (53 100) (38 101) (45 102) (40 103) (70 104) (53 105) (34 106) (76 107) (68 108) (64 109) (16 110) (49 111) (37 112) (156 113) (21 114) (109 115) (82 116) (55 117) (24 118) (29 119) (53 120) (56 121) (61 122) (32 123) (36 124) (43 125) (36 126) (83 127) (84 128) (40 129) (69 130) (49 131) (26 132) (87 133) (113 134) (51 135) (68 136) (112 137) (217 138) (43 139) (40 140) (77 141) (130 142) (50 143) (183 144) (132 145) (81 146) (26 147) (29 148) (39 149) (48 150) (24 151) (33 152) (98 153) (66 154) (60 155) (83 156) (54 157) (85 158) (71 159) (83 160) (64 161) (95 162) (62 163) (29 164) (29 165) (36 166) (29 167) (40 168) (42 169) (57 170) (28 171) (44 172) (43 173) (45 174) (87 175) (59 176) (94 177) (65 178) (45 179) (51 180) (14 181) (39 182) (47 183) (26 184) (62 185) (48 186) (80 187) (47 188) (25 189) (63 190) (47 191) (39 192) (19 193) (29 194) (38 195) (50 196) (36 197) (26 198) (118 199) (29 200) (20 201) (77 202) (90 203) (24 204) (46 205) (51 206) (45 207) (39 208) (65 209) (58 210) (57 211) (53 212) (4 213) (29 214) (39 215) (42 216) (11 217) (39 218) (29 219) (48 220) (31 221) (78 222) (57 223) (35 224) (15 225) (37 226) (28 227) (67 228) (64 229) (114 230) (53 231) (28 232) (32 233) (101 234) (80 235) (67 236) (120 237) (79 238) (62 239) (88 240) (83 241) (36 242) (28 243) (90 244) (17 245) (19 246) (43 247) (33 248) (33 249) (15 250) (37 251) (43 252) (27 253) (18 254) (34 255) (31 256) (30 257) (49 258) (5 259) (22 260) (22 261) (14 262) (20 263) (67 264) (72 265) (62 266) (83 267) (36 268) (3 269) (73 270) (41 271) (49 272) (60 273) (34 274) (138 275) (27 276) (161 277) (63 278) (27 279) (41 280) (10 281) (36 282) (33 283) (74 284) (21 285) (0 286) (0 287) (8 288) (11 289) (3 290) (25 291) (29 292) (26 293) (15 294) (4 295) (45 296) (25 297) (24 298) (11 299) (161 300) (45 301) (190 302) (14 303) (26 304) (9 305) (34 306) (18 307) (6 308) (27 309) (39 310) (50 311) (30 312) (27 313) (31 314) (31 315) (16 316) (44 317) (52 318) (22 319) (46 320) (12 321) (45 322) (27 323) (26 324) (5 325) (11 326) (31 327) (23 328) (43 329) (6 330) (13 331) (10 332) (17 333) (10 334) (19 335) (9 336) (7 337) (156 338) (17 339) (39 340) (2 341) (14 342) (7 343) (41 344) (45 345) (20 346) (7 347) (31 348) (42 349) (27 350) (41 351) (76 352) (93 353) (28 354) (28 355) (42 356) (36 357) (20 358) (32 359) (23 360) (33 361) (19 362) (12 363) (28 364) (10 365) (20 366) (48 367) (29 368) (23 369) (21 370) (15 371) (23 372) (56 373) (45 374) (30 375) (33 376) (28 377) (38 378) (25 379) (14 380) (8 381) (21 382) (27 383) (8 384) (12 385) (62 386) (45 387) (25 388) (27 389) (38 390) (27 391) (28 392) (6 393) (9 394) (23 395) (22 396) (33 397) (24 398) (30 399) (21 400) (35 401) (25 402) (56 403) (32 404) (25 405) (39 406) (41 407) (7 408) (2 409) (4 410) (0 411) (15 412) (79 413) (77 414) (56 415) (15 416) (32 417) (48 418) (29 419) (20 420) (30 421) (25 422) (4 423) (14 424) (7 425) (42 426) (121 427) (100 428) (30 429) (56 430) (74 431) (36 432) (26 433) (55 434) (21 435) (47 436) (25 437) (37 438) (27 439) (88 440) (66 441) (14 442) (14 443) (21 444) (28 445) (16 446) (8 447) (12 448) (27 449) (31 450) (34 451) (19 452) (23 453) (39 454) (23 455) (26 456) (8 457) (36 458) (20 459) (8 460) (26 461) (14 462) (14 463) (22 464) (25 465) (23 466) (29 467) (36 468) (22 469) (39 470) (41 471) (12 472) (18 473) (15 474) (12 475) (18 476) (21 477) (25 478) (23 479) (22 480) (22 481) (21 482) (18 483) (30 484) (36 485) (40 486) (29 487) (24 488) (24 489) (28 490) (10 491) (25 492) (6 493) (27 494) (22 495) (25 496) (15 497) (4 498) (11 499) (6 500) (21 501) (20 502) (24 503) (11 504) (7 505) (35 506) (3 507) (46 508) (52 509) (19 510) (25 511) (34 512) (50 513) (18 514) (22 515) (14 516) (22 517) (18 518) (26 519) (33 520) (15 521) (11 522) (4 523) (5 524) (41 525) (17 526) (21 527) (51 528) (56 529) (3 530) (48 531) (32 532) (23 533) (49 534) (74 535) (19 536) (65 537) (75 538) (50 539) (72 540) (13 541) (49 542) (17 543) (44 544) (7 545) (5 546) (9 547) (12 548) (11 549) (20 550) (74 551) (29 552) (40 553) (44 554) (65 555) (53 556) (51 557) (52 558) (41 559) (54 560) (56 561) (32 562) (52 563) (36 564) (1 565) (0 566) (33 567) (6 568) (43 569) (66 570) (61 571) (78 572) (62 573) (56 574) (62 575) (35 576) (29 577) (34 578) (3 579) (25 580) (15 581) (20 582) (17 583) (10 584) (26 585) (0 586) (42 587) (34 588) (12 589) (6 590) (11 591) (20 592) (12 593) (23 594) (26 595) (37 596) (15 597) (28 598) (12 599) (31 600) (8 601) (7 602) (94 603) (60 604) (29 605) (24 606) (18 607) (94 608) (24 609) (38 610) (36 611) (16 612) (20 613) (22 614) (13 615) (38 616) (83 617) (46 618) (75 619) (46 620) (62 621) (15 622) (43 623) (58 624) (9 625) (68 626) (27 627) (42 628) (27 629) (48 630) (21 631) (55 632) (27 633) (31 634) (97 635) (50 636) (51 637) (57 638) (37 639) (26 640) (40 641) (31 642) (18 643) (30 644) (19 645) (19 646) (15 647) (11 648) (26 649) (9 650) (23 651) (15 652) (11 653) (21 654) (32 655) (20 656) (7 657) (18 658) (35 659) (72 660) (55 661) (54 662) (55 663) (98 664) (60 665) (22 666) (59 667) (48 668) (6 669) (9 670) (42 671) (3 672)

(12 673) (1 674) (27 675) (27 676) (30 677) (36 678) (36 679) (38 680) (43 681) (76 682) (58 683) (41 684) (30 685) (35 686) (97 687) (44 688) (67 689) (24 690) (28 691) (17 692) (17 693) (31 694) (31 695) (54 696) (29 697) (37 698) (34 699) (51 700) (61 701) (37 702) (13 703) (5 704) (62 705) (140 706) (74 707) (74 708) (107 709) (73 710) (168 711) (39 712) (24 713) (130 714) (35 715) (40 716) (61 717) (60 718) (52 719) (42 720) (62 721) (53 722) (66 723) (58 724) (47 725) (23 726) (70 727) (94 728) (57 729) (74 730) (50 731) (47 732) (54 733) (49 734) (40 735) (57 736) (40 737) (49 738) (57 739) (103 740) (42 741) (45 742) (50 743) (16 744) (52 745) (51 746) (68 747) (37 748) (57 749) (51 750) (58 751) (36 752) (77 753) (57 754) (53 755) (55 756) (25 757) (57 758) (48 759) (3 760) (28 761) (44 762) (24 763) (54 764) (57 765) (8 766) (68 767) (131 768) (42 769) (124 770) (23 771) (25 772) (12 773) (18 774) (56 775) (30 776) (44 777) (10 778) (51 779) (27 780) (39 781) (26 782) (59 783) (12 784) (16 785) (8 786) (18 787) (9 788) (110 789) (22 790) (25 791) (22 792) (45 793) (1 794) (0 795) (0 796) (0 797) (25 798) (0 799) (0 800) (6 801) (0 802) (0 803) (16 804) (0 805) (56 806) (54 807) (44 808) (23 809) (2 810) (5 811) (3 812) (45 813) (53 814) (33 815) (40 816) (11 817) (0 818) (20 819) (5 820) (0 821) (0 822) (30 823) (9 824) (1 825))
>

; The above was named as bed_plist and saved as BEDPLIST.LSP, that is, a list of ED and pixels pairs.

```
; 261196.log
> ; loading "C:\USERS\LAW\GIS\HO\BEDPLIST.LSP"
(length bed_plist)
825
> (def source_zones (read-data-columns "bedbeat.out" 13))
SOURCE_ZONES
; there were 13 items from ARC/INFO but only the first 8 are of interest for this case study.
> (defun get_gob (entity)
     (do ((my_list entity (cdr my_list)))
      ((null my_list) )
       (def attribute (car my_list))
        (def i (+ i 1))
         (cond ((null entity) nil)
            ((= i 1) (def ed-id attribute))
            ((= i 2) (def edgs-id attribute))
            ((= i 3) (def beatgs-id attribute))
            ((= i 4) (def beat-id attribute))
             ((= i 5) (def superbt-id attribute))
              ((= i 6) (def ward-id attribute))
              ((= i 7) (def residents attribute))
               ((= i 8) (def households attribute))
               (T  (def attribute (cdr entity))))))
GET_GOB

> (setq i 0)
0
> (get_gob source_zones)
NIL
> (length households)
821
> (length edgs-id)
821
; source zone has 4 EDs missing due to the overlay method.

> ; loading "C:\USERS\LAW\GIS\HO\PROBBT.LSP"
 (init_pre_beat_go)
FOUND_VALUE
> FOUND_VALUE
NIL
```

; 081296.log

; Creat LUT of id : household counts with 1 to many relationship

```
 (defun create_lut (list1 list2)
    (dolist (i list1)
      (def lut (cons (list i (car list2)) lut))
      (def list2 (cdr list2))))

(def lut nil)
(create_LUT beatgs-id households)

> (defun create_index_list (n)
   (def listi nil)
   (dotimes (i n)
    (def j (+ 1 i))
     (def listi (cons (list j 0) listi))))
CREATE_INDEX_LIST
> (create_index_list 63)
NIL
> (length listi)
63
> (first listi)
(63 0)
> (last listi)
((1 0))
>
(defun get_value_from_lut (index_list data_list)
   (dolist (i index_list)
    (match_fill_value_list i data_list)))

> (defun match_fill_value_list  (index_value data_list)
    (dolist (record data_list)
     (cond ((= (car index_value) (car record)) (def (second index_value) (+ (second index_value)
(second record))))
        (T nil))))
MATCH_FILL_VALUE_LIST
> (first listi)
(1 0)
> (last listi)
((63 0))
> (get_value_from_lut listi lut)
NIL
> (first listi)
(1 870.0)
> (last listi)
((63 3359.0))
> (second listi)
(2 5830.0)
> listi
((1 870.0) (2 5830.0) (3 635.0) (4 2077.0) (5 2288.0) (6 2654.0) (7 3680.0) (8 3508.0) (9 1910.0) (10
3297.0) (11 2108.0) (12 2866.0) (13 2674.0) (14 2173.0) (15 1772.0) (16 1438.0) (17 1078.0) (18
1290.0) (19 1439.0) (20 1538.0) (21 1877.0) (22 2519.0) (23 1684.0) (24 851.0) (25 2173.0) (26
1478.0) (27 2207.0) (28 0) (29 271.0) (30 2352.0) (31 727.0) (32 0) (33 164.0) (34 0) (35 0) (36 1545.0)
(37 923.0) (38 3014.0) (39 1881.0) (40 4541.0) (41 6295.0) (42 3594.0) (43 10183.0) (44 154.0) (45
3782.0) (46 4269.0) (47 4447.0) (48 3980.0) (49 2890.0) (50 3215.0) (51 2615.0) (52 3813.0) (53
```

```
1779.0) (54 1997.0) (55 1865.0) (56 2616.0) (57 2794.0) (58 4014.0) (59 261.0) (60 3683.0) (61
4216.0) (62 6269.0) (63 3359.0))
>
 (setq b_hh listi)

b_hh
((1 870.0) (2 5830.0) (3 635.0) (4 2077.0) (5 2288.0) (6 2654.0) (7 3680.0) (8 3508.0) (9 1910.0) (10
3297.0) (11 2108.0) (12 2866.0) (13 2674.0) (14 2173.0) (15 1772.0) (16 1438.0) (17 1078.0) (18
1290.0) (19 1439.0) (20 1538.0) (21 1877.0) (22 2519.0) (23 1684.0) (24 851.0) (25 2173.0) (26
1478.0) (27 2207.0) (28 0) (29 271.0) (30 2352.0) (31 727.0) (32 0) (33 164.0) (34 0) (35 0) (36 1545.0)
(37 923.0) (38 3014.0) (39 1881.0) (40 4541.0) (41 6295.0) (42 3594.0) (43 10183.0) (44 154.0) (45
3782.0) (46 4269.0) (47 4447.0) (48 3980.0) (49 2890.0) (50 3215.0) (51 2615.0) (52 3813.0) (53
1779.0) (54 1997.0) (55 1865.0) (56 2616.0) (57 2794.0) (58 4014.0) (59 261.0) (60 3683.0) (61
4216.0) (62 6269.0) (63 3359.0))
>
> (savevar 'b_hh "bbt_hh")

; Similarly for residents

> (create_index_list 63)
NIL
> listi
((63 0) (62 0) (61 0) (60 0) (59 0) (58 0) (57 0) (56 0) (55 0) (54 0) (53 0) (52 0) (51 0) (50 0) (49 0) (48
0) (47 0) (46 0) (45 0) (44 0) (43 0) (42 0) (41 0) (40 0) (39 0) (38 0) (37 0) (36 0) (35 0) (34 0) (33 0)
(32 0) (31 0) (30 0) (29 0) (28 0) (27 0) (26 0) (25 0) (24 0) (23 0) (22 0) (21 0) (20 0) (19 0) (18 0) (17
0) (16 0) (15 0) (14 0) (13 0) (12 0) (11 0) (10 0) (9 0) (8 0) (7 0) (6 0) (5 0) (4 0) (3 0) (2 0) (1 0))
> (def lut nil)
LUT
> lut
NIL
> (create_LUT beatgs-id residents)
NIL
> (first lut)
(61 603.0)
> (last lut)
((59 722.0))
>

> (get_value_from_lut listi lut)
NIL
> (savevar 'b_hh "bbt_hh")
(B_HH)
> (first listi)
(63 8640.0)
> (last listi)
((1 2239.0))
> listi
((63 8640.0) (62 16967.0) (61 10993.0) (60 9562.0) (59 722.0) (58 10169.0) (57 6598.0) (56 7076.0)
(55 4565.0) (54 4672.0) (53 3906.0) (52 9039.0) (51 6146.0) (50 7455.0) (49 6522.0) (48 10193.0) (47
10481.0) (46 10132.0) (45 7835.0) (44 313.0) (43 25358.0) (42 8191.0) (41 17679.0) (40 10554.0) (39
4466.0) (38 6535.0) (37 1505.0) (36 3364.0) (35 0) (34 0) (33 338.0) (32 0) (31 1423.0) (30 4715.0) (29
479.0) (28 0) (27 4053.0) (26 3008.0) (25 4269.0) (24 1910.0) (23 3291.0) (22 5227.0) (21 3568.0) (20
3375.0) (19 3382.0) (18 2987.0) (17 2671.0) (16 3449.0) (15 4320.0) (14 5020.0) (13 6291.0) (12
6832.0) (11 4814.0) (10 8265.0) (9 5256.0) (8 8827.0) (7 8684.0) (6 6813.0) (5 5741.0) (4 4845.0) (3
1543.0) (2 14146.0) (1 2239.0))
>

> (setq bbt_r listi)
```

```
> (savevar 'bbt_r "bbt_r")
(BBT_R)
>

; To make zone  residents hh r/hh p  hh/p  r/p table...

> (defun make_pair_list (list1 list2)
    (def pair_list nil)
    (dolist (i list1)
     (def pair_list (cons (list i (car list2)) pair_list))
       (def list2 (cdr list2))))
MAKE_PAIR_LIST
> (make_pair_list beatgs-id edgs-id)
NIL
> (length pair_list)
821
> (first pair_list)
(61 803)
> (last pair_list)
((59 797))

> ; loading "C:\HO\PHD\BEDPLIST.LSP"
(length bed_plist)
825
>
> (create_index_list 63)
NIL
> listi
((63 0) (62 0) (61 0) (60 0) (59 0) (58 0) (57 0) (56 0) (55 0) (54 0) (53 0) (52 0) (51 0) (50 0) (49 0) (48
0) (47 0) (46 0) (45 0) (44 0) (43 0) (42 0) (41 0) (40 0) (39 0) (38 0) (37 0) (36 0) (35 0) (34 0) (33 0)
(32 0) (31 0) (30 0) (29 0) (28 0) (27 0) (26 0) (25 0) (24 0) (23 0) (22 0) (21 0) (20 0) (19 0) (18 0) (17
0) (16 0) (15 0) (14 0) (13 0) (12 0) (11 0) (10 0) (9 0) (8 0) (7 0) (6 0) (5 0) (4 0) (3 0) (2 0) (1 0))

> (defun make_list3 (list1 list2)
   (def list3 nil)
   (dolist (i list1)
     (match_cons i list2)))
MAKE_LIST3
> (defun match_cons (id_pair list2)
     (dolist (i list2)
     (cond ((= (second id_pair) (second i)) (def list3 (cons (cons (first id_pair) i) list3)))
            (T nil))))
MATCH_CONS
>
> (make_list3 pair_list bed_plist)

> (last pair_list)
((59 797))
> (last list3)
((61 0 803))
> (tenth list3)
(2 41 18)
> (def bbt_ped list3)
BBT_PED
> (length bbt_ped)
821
> (savevar 'bbt_ped "bbt_ped")
```

```
(BBT_PED)
>

> (defun make_listv (n list1)
    (def listv nil)
    (dotimes (i n)
     (def j (+ 1 i))
     (def v 0)
      (dolist (k list1)
     (cond ((= j (car k)) (def v (+ v (second k))))
         (T nil)))
      (def listv (cons (list j v) listv))))
MAKE_LISTV
> (make_listv 63 bbt_ped)
NIL
> (length listv)
63
> (first listv)
(63 859)
> (second listv)
(62 576)
> (last listv)
((1 1346))
> (first bbt_ped)
(59 0 797)
> listv
((63 859) (62 576) (61 492) (60 489) (59 0) (58 863) (57 849) (56 587) (55 544) (54 459) (53 174) (52
452) (51 649) (50 1186) (49 865) (48 1129) (47 1002) (46 974) (45 421) (44 92) (43 2463) (42 599) (41
1184) (40 1033) (39 241) (38 312) (37 49) (36 191) (35 0) (34 0) (33 73) (32 0) (31 106) (30 468) (29
50) (28 0) (27 296) (26 230) (25 399) (24 171) (23 276) (22 487) (21 251) (20 161) (19 252) (18 251)
(17 263) (16 311) (15 471) (14 451) (13 744) (12 520) (11 694) (10 684) (9 465) (8 1533) (7 1343) (6
847) (5 751) (4 488) (3 168) (2 1555) (1 1346))
>
> (def bbt_plist listv)
BBT_PLIST
> (savevar 'bbt_plist "bbt_plist")
(BBT_PLIST)
>
> bbt_r
((63 8640.0) (62 16967.0) (61 10993.0) (60 9562.0) (59 722.0) (58 10169.0) (57 6598.0) (56 7076.0)
(55 4565.0) (54 4672.0) (53 3906.0) (52 9039.0) (51 6146.0) (50 7455.0) (49 6522.0) (48 10193.0) (47
10481.0) (46 10132.0) (45 7835.0) (44 313.0) (43 25358.0) (42 8191.0) (41 17679.0) (40 10554.0) (39
4466.0) (38 6535.0) (37 1505.0) (36 3364.0) (35 0) (34 0) (33 338.0) (32 0) (31 1423.0) (30 4715.0) (29
479.0) (28 0) (27 4053.0) (26 3008.0) (25 4269.0) (24 1910.0) (23 3291.0) (22 5227.0) (21 3568.0) (20
3375.0) (19 3382.0) (18 2987.0) (17 2671.0) (16 3449.0) (15 4320.0) (14 5020.0) (13 6291.0) (12
6832.0) (11 4814.0) (10 8265.0) (9 5256.0) (8 8827.0) (7 8684.0) (6 6813.0) (5 5741.0) (4 4845.0) (3
1543.0) (2 14146.0) (1 2239.0))
> bbt_hh
Error: The variable BBT_HH is unbound.
> b_hh
((1 870.0) (2 5830.0) (3 635.0) (4 2077.0) (5 2288.0) (6 2654.0) (7 3680.0) (8 3508.0) (9 1910.0) (10
3297.0) (11 2108.0) (12 2866.0) (13 2674.0) (14 2173.0) (15 1772.0) (16 1438.0) (17 1078.0) (18
1290.0) (19 1439.0) (20 1538.0) (21 1877.0) (22 2519.0) (23 1684.0) (24 851.0) (25 2173.0) (26
1478.0) (27 2207.0) (28 0) (29 271.0) (30 2352.0) (31 727.0) (32 0) (33 164.0) (34 0) (35 0) (36 1545.0)
(37 923.0) (38 3014.0) (39 1881.0) (40 4541.0) (41 6295.0) (42 3594.0) (43 10183.0) (44 154.0) (45
3782.0) (46 4269.0) (47 4447.0) (48 3980.0) (49 2890.0) (50 3215.0) (51 2615.0) (52 3813.0) (53
1779.0) (54 1997.0) (55 1865.0) (56 2616.0) (57 2794.0) (58 4014.0) (59 261.0) (60 3683.0) (61
4216.0) (62 6269.0) (63 3359.0))
```

```
> (def bbt_hh b_hh)
BBT_HH
> bbt_plist
((63 859) (62 576) (61 492) (60 489) (59 0) (58 863) (57 849) (56 587) (55 544) (54 459) (53 174) (52
452) (51 649) (50 1186) (49 865) (48 1129) (47 1002) (46 974) (45 421) (44 92) (43 2463) (42 599) (41
1184) (40 1033) (39 241) (38 312) (37 49) (36 191) (35 0) (34 0) (33 73) (32 0) (31 106) (30 468) (29
50) (28 0) (27 296) (26 230) (25 399) (24 171) (23 276) (22 487) (21 251) (20 161) (19 252) (18 251)
(17 263) (16 311) (15 471) (14 451) (13 744) (12 520) (11 694) (10 684) (9 465) (8 1533) (7 1343) (6
847) (5 751) (4 488) (3 168) (2 1555) (1 1346))
> (def bbt_r (reverse bbt_r))
BBT_R
> (def bbt_plist (reverse bbt_plist))
BBT_PLIST

> (defun my_21 (v)
    (second (car v)))

> (defun my_print (list1 list2 list3)
   (dotimes (i (length list1))
    (def j (+ 1 i))
    (print j) (princ ",")
    (prin1 (my_21 list1)) (princ ",")
    (prin1 (my_21 list2)) (princ ",")
    (print_ratio list1 list2)
    (prin1 (my_21 list3)) (princ ",")
    (print_ratio list2 list3)
    (print_ratio list1 list3)
    (def list1 (cdr list1))
    (def list2 (cdr list2))
    (def list3 (cdr list3))))
MY_PRINT
>
> (defun print_ratio (list1 list2)
    (cond ((= (my_21 list2) 0) (prin1 0))
      (T (prin1 (/ (my_21 list1) (my_21 list2))) (princ " , "))))
PRINT_RATIO

> (my_print bbt_r bbt_hh bbt_plist)

1 ,2239.0,870.0,2.57 , 1346,0.65 , 1.66 ,
2 ,14146.0,5830.0,2.43 , 1555,3.75 , 9.1 ,
3 ,1543.0,635.0,2.43 , 168,3.78 , 9.18 ,
4 ,4845.0,2077.0,2.33 , 488,4.26 , 9.93 ,
5 ,5741.0,2288.0,2.51 , 751,3.05 , 7.64 ,
6 ,6813.0,2654.0,2.57 , 847,3.13 , 8.04 ,
7 ,8684.0,3680.0,2.36 , 1343,2.74 , 6.47 ,
8 ,8827.0,3508.0,2.52 , 1533,2.29 , 5.76 ,
9 ,5256.0,1910.0,2.75 , 465,4.11 , 11.3 ,
10 ,8265.0,3297.0,2.51 , 684,4.82 , 12.08 ,
11 ,4814.0,2108.0,2.28 , 694,3.04 , 6.94 ,
12 ,6832.0,2866.0,2.38 , 520,5.51 , 13.14 ,
13 ,6291.0,2674.0,2.35 , 744,3.59 , 8.46,
14 ,5020.0,2173.0,2.31 , 451,4.82 , 11.13 ,
15 ,4320.0,1772.0,2.44 , 471,3.76 , 9.17 ,
16 ,3449.0,1438.0,2.4 , 311,4.62 , 11.09 ,
17 ,2671.0,1078.0,2.48 , 263,4.1 , 10.16 ,
18 ,2987.0,1290.0,2.32 , 251,5.14 , 11.9 ,
19 ,3382.0,1439.0,2.35 , 252,5.71 , 13.42 ,
```

```
20 ,3375.0,1538.0,2.19 , 161,9.55 , 20.96 ,
21 ,3568.0,1877.0,1.9 , 251,7.48 , 14.22 ,
22 ,5227.0,2519.0,2.08 , 487,5.17 , 10.73 ,
23 ,3291.0,1684.0,1.95 , 276,6.1 , 11.92 ,
24 ,1910.0,851.0,2.24 , 171,4.98 , 11.17 ,
25 ,4269.0,2173.0,1.96 , 399,5.45 , 10.7 ,
26 ,3008.0,1478.0,2.04 , 230,6.43 , 13.08 ,
27 ,4053.0,2207.0,1.84 , 296,7.46 , 13.69 ,
28 ,0,0,00,00
29 ,479.0,271.0,1.77 , 50,5.42 , 9.58 ,
30 ,4715.0,2352.0,2 , 468,5.03 , 10.07 ,
31 ,1423.0,727.0,1.96 , 106,6.86 , 13.42 ,
32 ,0,0,00,00
33 ,338.0,164.0,2.06 , 73,2.25 , 4.63 ,
34 ,0,0,00,00
35 ,0,0,00,00
36 ,3364.0,1545.0,2.18 , 191,8.09 , 17.61 ,
37 ,1505.0,923.0,1.63 , 49,18.84 , 30.71 ,
38 ,6535.0,3014.0,2.17 , 312,9.66 , 20.95 ,
39 ,4466.0,1881.0,2.37 , 241,7.8 , 18.53 ,
40 ,10554.0,4541.0,2.32 , 1033,4.4 , 10.22 ,
41 ,17679.0,6295.0,2.81 , 1184,5.32 , 14.93 ,
42 ,8191.0,3594.0,2.28 , 599,6 , 13.67 ,
43 ,25358.0,10183.0,2.49 , 2463,4.13 , 10.3 ,
44 ,313.0,154.0,2.03 , 92,1.67 , 3.4 ,
45 ,7835.0,3782.0,2.07 , 421,8.98 , 18.61 ,
46 ,10132.0,4269.0,2.37 , 974,4.38 , 10.4 ,
47 ,10481.0,4447.0,2.36 , 1002,4.44 , 10.46 ,
48 ,10193.0,3980.0,2.56 , 1129,3.53 , 9.03 ,
49 ,6522.0,2890.0,2.26 , 865,3.34 , 7.54 ,
50 ,7455.0,3215.0,2.32 , 1186,2.71 , 6.29 ,
51 ,6146.0,2615.0,2.35 , 649,4.03 , 9.47 ,
52 ,9039.0,3813.0,2.37 , 452,8.44 , 20 ,
53 ,3906.0,1779.0,2.2 , 174,10.22 , 22.45 ,
54 ,4672.0,1997.0,2.34 , 459,4.35 , 10.18 ,
55 ,4565.0,1865.0,2.45 , 544,3.43 , 8.39 ,
56 ,7076.0,2616.0,2.7 , 587,4.46 , 12.05 ,
57 ,6598.0,2794.0,2.36 , 849,3.29 , 7.77 ,
58 ,10169.0,4014.0,2.53 , 863,4.65 , 11.78 ,
59 ,722.0,261.0,2.77 , 0,00
60 ,9562.0,3683.0,2.6 , 489,7.53 , 19.55 ,
61 ,10993.0,4216.0,2.61 , 492,8.57 , 22.34 ,
62 ,16967.0,6269.0,2.71 , 576,10.88 , 29.46 ,
63 ,8640.0,3359.0,2.57 , 859,3.91 , 10.06 ,
NIL
>
; Calibration of Landsat image completed
```

; 091296.log

```
> (def bbt_r '((63 8640.0) (62 16967.0) (61 10993.0) (60 9562.0) (59 722.0) (58 10169.0) (57 6598.0)
(56 7076.0) (55 4565.0) (54 4672.0) (53 3906.0) (52 9039.0) (51 6146.0) (50 7455.0) (49 6522.0) (48
10193.0) (47 10481.0) (46 10132.0) (45 7835.0) (44 313.0) (43 25358.0) (42 8191.0) (41 17679.0) (40
10554.0) (39 4466.0) (38 6535.0) (37 1505.0) (36 3364.0) (35 0) (34 0) (33 338.0) (32 0) (31 1423.0)
(30 4715.0) (29 479.0) (28 0) (27 4053.0) (26 3008.0) (25 4269.0) (24 1910.0) (23 3291.0) (22 5227.0)
(21 3568.0) (20 3375.0) (19 3382.0) (18 2987.0) (17 2671.0) (16 3449.0) (15 4320.0) (14 5020.0) (13
6291.0) (12 6832.0) (11 4814.0) (10 8265.0) (9 5256.0) (8 8827.0) (7 8684.0) (6 6813.0) (5 5741.0) (4
4845.0) (3 1543.0) (2 14146.0) (1 2239.0)))
```

```
BBT_R
> (length bbt_r)
63
> (def bbt_hh '((1 870.0) (2 5830.0) (3 635.0) (4 2077.0) (5 2288.0) (6 2654.0) (7 3680.0) (8 3508.0) (9
1910.0) (10 3297.0) (11 2108.0) (12 2866.0) (13 2674.0) (14 2173.0) (15 1772.0) (16 1438.0) (17
1078.0) (18 1290.0) (19 1439.0) (20 1538.0) (21 1877.0) (22 2519.0) (23 1684.0) (24 851.0) (25
2173.0) (26 1478.0) (27 2207.0) (28 0) (29 271.0) (30 2352.0) (31 727.0) (32 0) (33 164.0) (34 0) (35 0)
(36 1545.0) (37 923.0) (38 3014.0) (39 1881.0) (40 4541.0) (41 6295.0) (42 3594.0) (43 10183.0) (44
154.0) (45 3782.0) (46 4269.0) (47 4447.0) (48 3980.0) (49 2890.0) (50 3215.0) (51 2615.0) (52
3813.0) (53 1779.0) (54 1997.0) (55 1865.0) (56 2616.0) (57 2794.0) (58 4014.0) (59 261.0) (60
3683.0) (61 4216.0) (62 6269.0) (63 3359.0)))
BBT_HH
> (length bbt_hh)
63
> (def bbt_plist '((63 859) (62 576) (61 492) (60 489) (59 0) (58 863) (57 849) (56 587) (55 544) (54
459) (53 174) (52 452) (51 649) (50 1186) (49 865) (48 1129) (47 1002) (46 974) (45 421) (44 92) (43
2463) (42 599) (41 1184) (40 1033) (39 241) (38 312) (37 49) (36 191) (35 0) (34 0) (33 73) (32 0) (31
106) (30 468) (29 50) (28 0) (27 296) (26 230) (25 399) (24 171) (23 276) (22 487) (21 251) (20 161)
(19 252) (18 251) (17 263) (16 311) (15 471) (14 451) (13 744) (12 520) (11 694) (10 684) (9 465) (8
1533) (7 1343) (6 847) (5 751) (4 488) (3 168) (2 1555) (1 1346)))
BBT_PLIST
> (length bbt_plist)
63
```

; 121296.log from probtf.log

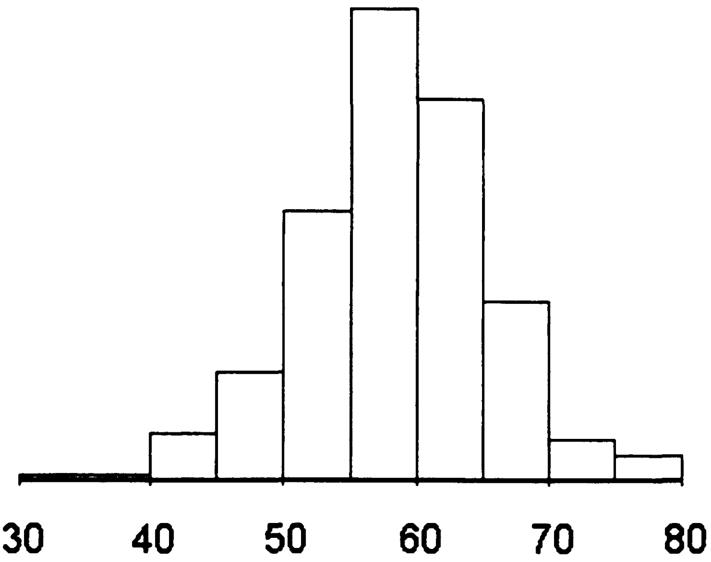
; Carry out Monte Carlo simulation Bristol Beat

```
> ; loading "C:\HO\PHD\FPRO.LSP"
(setf f (open "wbbtx01.img"))
#<Input-Stream 5:"c:\idrisi\data\wbbtx01.img">
> (list_batch pn_list f 4030)

"Please wait..." Saving this list into file "bbt1f"
"Please wait..." Saving this list into file "bbt2f"
:
:
:
"Please wait..." Saving this list into file "bbt39f"

> (close f)
T
> ; loading "C:\HO\PHD\MONTEK.LSP"
```

; also load MCBH.LSP

```
> (def bbt (read-data-file "wbbtx01.img"))
BBT
> (length bbt)
160460
>
> (max bbt)
56
>
:
 (monte_go 300 pn_list 10)

> (savevar 'new_list "bbthid")
```

```
(NEW_LIST)
> (histogram (car new_list))
#<Object: 228004122, prototype = HISTOGRAM-PROTO>
```



; A normal distribution was obtained after 300 runs.

; **Statistical note.** Since here one is sampling from a list of numbers with only two possible outcomes: 0 or 1, it is equivalent to the so-called *independent Bernoulli trials* (e.g. the coin-tossing experiment). The corresponding probability density function (pdf) is:

$$P\{x=0\}= p$$

and

$$P\{x=1\} = q = 1 - p$$

where $0 \leq q \leq 1$

Assume the case of n numbers in the list (Bernoulli trials from the pixels of the Landsat image) each with an equal and independent chance (P=1/n) where n is specified *in advanced* from the stratified random sample. The probability of a particular combination of outcomes with k zero values and (n-k) counts is

$$p^k q^{n-k}, \qquad 0 \leq k \leq n$$

(For k = 1, i.e. the probability that the first number is zero and the remaining numbers are non-zero is pq^{n-1})

In this case all the distinct combinations having k zeros (regardless of their order of occurrence in the list) are:

$$\binom{n}{k}\frac{n!}{k!(n-k)!} \qquad \textbf{(A7.5. 1)}$$

According to the addition law of probability,

$$p\{x=k\} = \binom{n}{k} p^k q^{n-k} \qquad \textbf{(A7.5. 2)}$$

k = 0, 1, 2, ..., n

This is the binomial distribution with parameters n and p. It satisfies the conditions for pdf, since

$P\{x=k\} \geq 0, \qquad \forall k = 0, 1, 2, ..., n$

$$\sum_{k=0}^{n} P\{x=k\} = \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} \qquad \textbf{(A7.5. 3)}$$

$$= (p+q)^n = 1$$

So theoretically, the distribution from the stratified random sampling should be binomial. However given a fixed p, $n \rightarrow \infty$ infinity (or very large, in this case n= 300), the binomial distribution approximates the normal distribution (Taha, 1982):

$$\sum_{k=a}^{b} \binom{n}{k} p^k q^{n-k} \longrightarrow \frac{1}{\sqrt{2\pi}} \int_{(a-\mu-1/2)}^{(b-\mu+1/2)} e^{-y^2/2} dy \qquad \textbf{(A7.5. 4)}$$

where $\mu = np$ and $\sigma = \sqrt{npq}$

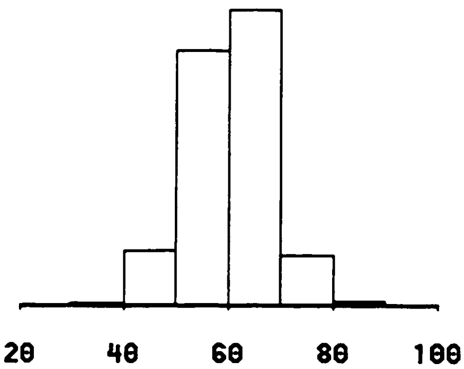
As $n \rightarrow \infty$, p and $\mu \rightarrow 0$; and
Since $P = pq = 1/n$, $\sigma = 1$. This satisfies the conditions for the standard normal pdf:

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}, \quad -\infty < z < \infty \qquad \textbf{(A7.5. 5)}$$

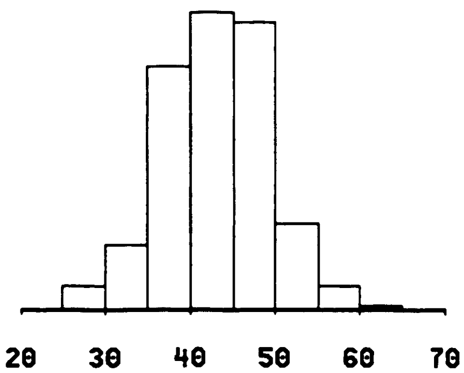
with its parameters given by $\mu = 0$ and $\sigma = 1$. The corresponding cumulative density function is given by:

$$\phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz \qquad \textbf{(A7.5. 6)}$$

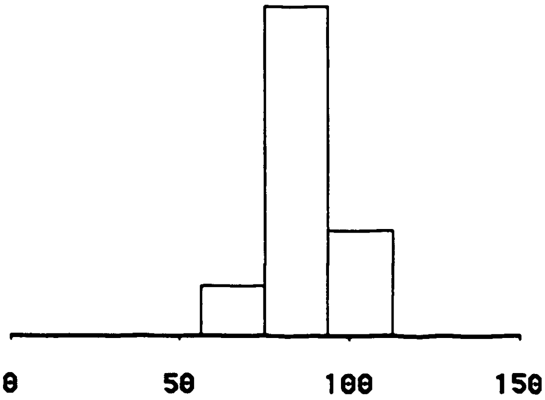
```
; Re-do 10/12/1996
> ; loading "C:\HO\PHD\MONTEK.LSP"

; also load MCBH.LSP

> (monte_go1)
(monte_go1)
(monte_go 298 pn_list 10)

> (length new_list)
63
> (length (car new_list))
301
> (savevar 'new_list "bbthid")
(NEW_LIST)
```




```
> (histogram (car new_list))
#<Object: 234329114, prototype = HISTOGRAM-PROTO>
>
```




```
> (histogram (tenth new_list))
#<Object: 233301690, prototype = HISTOGRAM-PROTO>
>
```




```
> (histogram (last new_list))
#<Object: 233795738, prototype = HISTOGRAM-PROTO>
>
```

; 121296.log
; Assessing Monte Carlo accuracy

```
(def _hID '((0 1) (0 2) (0 3) (0 4) (0 5) (0 6) (0 7) (0 8) (0 9) (0 10) (0 11) (0 12)
(0 13) (0 14) (0 15) (0 16) (0 17) (0 18) (0 19) (0 20) (0 21) (0 22) (0 23) (0 24) (0 25) (0 26)
(0 27) (0 28) (0 29) (0 30) (0 31) (0 32) (0 33) (0 34) (0 35) (0 36) (0 37) (0 38) (0 39) (0 40)
(0 41) (0 42) (0 43) (0 44) (0 45) (0 46) (0 47) (0 48) (0 49) (0 50) (0 51) (0 52) (0 53) (0 54)
(0 55) (0 56) (0 57) (0 58) (0 59) (0 60) (0 61) (0 62) (0 63)))
_HID
> (defun init_fill (id_list)
    (dolist (index id_list)
      (def (first index) 0)))
INIT_FILL
> (defun match_cons (list1 i)
   (dolist (v list1)
    (cond ((= (car (last i)) (car (last v))) (def new_list (cons (cons (car i) v) new_list)))
          (T NIL))))
MATCH_CONS
> (defun cons_value_into_list (list1 list2)
   (dolist (h list2)
     (match_cons list1 h)))
CONS_VALUE_INTO_LIST
> (defun assignh (my_list id_list)
   (dolist (p my_list)
   (cond ((= p 0) NIL)
         (T (match_fill id_list p)))))
ASSIGNH
> (defun match_fill (id_list p)
    (dolist (index id_list)
     (cond ((= p (car (last index))) (def (first index) (+ 1 (first index))))
           (T NIL))))
MATCH_FILL
> (def pn_list '("bbt1f" "bbt2f" "bbt3f" "bbt4f" "bbt5f" "bbt6f" "bbt7f" "bbt8f" "bbt9f" "bbt10f"
"bbt11f" "bbt12f" "bbt13f" "bbt14f" "bbt15f" "bbt16f" "bbt17f" "bbt18f" "bbt19f" "bbt20f"
"bbt21f" "bbt22f" "bbt23f" "bbt24f" "bbt25f" "bbt26f" "bbt27f" "bbt28f" "bbt29f" "bbt30f"
"bbt31f" "bbt32f" "bbt33f" "bbt34f" "bbt35f" "bbt36f" "bbt37f" "bbt38f" "bbt39f"))
PN_LIST
> (def _hid1 '((1) (2) (3) (4) (5) (6) (7) (8) (9) (10) (11) (12)
(13) (14) (15) (16) (17) (18) (19) (20) (21) (22) (23) (24) (25) (26)
(27) (28) (29) (30) (31) (32) (33) (34) (35) (36) (37) (38) (39) (40)
(41) (42) (43) (44) (45) (46) (47) (48) (49) (50) (51) (52) (53) (54)
```

```
(55) (56) (57) (58) (59) (60) (61) (62) (63) ))
_HID1
> (defun assignh_file_batch (file_list id_list)
    (dolist (file_name file_list)
      (load file_name)
      (assignh p_list id_list)))
ASSIGNH_FILE_BATCH
> ; loading "C:\USERS\LAW\GIS\HO\DATA\OK.LSP"
; a simple flag to indicate the programs run in the correct data directory
HO_IS_OK

(assignh_file_batch pn_list _hID)
; loading "bbt1f.lsp"
; loading "bbt2f.lsp"
; loading "bbt3f.lsp"
; loading "bbt4f.lsp"
; loading "bbt5f.lsp"
; loading "bbt6f.lsp"
; loading "bbt7f.lsp"
; loading "bbt8f.lsp"
; loading "bbt9f.lsp"
; loading "bbt10f.lsp"
; loading "bbt11f.lsp"
; loading "bbt12f.lsp"
; loading "bbt13f.lsp"
; loading "bbt14f.lsp"
; loading "bbt15f.lsp"
; loading "bbt16f.lsp"
; loading "bbt17f.lsp"
; loading "bbt18f.lsp"
; loading "bbt19f.lsp"
; loading "bbt20f.lsp"
; loading "bbt21f.lsp"
; loading "bbt22f.lsp"
; loading "bbt23f.lsp"
; loading "bbt24f.lsp"
; loading "bbt25f.lsp"
; loading "bbt26f.lsp"
; loading "bbt27f.lsp"
; loading "bbt28f.lsp"
; loading "bbt29f.lsp"
; loading "bbt30f.lsp"
; loading "bbt31f.lsp"
; loading "bbt32f.lsp"
; loading "bbt33f.lsp"
; loading "bbt34f.lsp"
; loading "bbt35f.lsp"
; loading "bbt36f.lsp"
; loading "bbt37f.lsp"
; loading "bbt38f.lsp"
; loading "bbt39f.lsp"
NIL
> (length new_list)
Error: The variable NEW_LIST is unbound.
> (length _hid)
63
> (first _hid)
(868 1)
> _hid
```

```
((868 1) (1435 2) (124 3) (643 4) (769 5) (753 6) (974 7) (1736 8) (494 9) (561 10) (704 11) (503 12)
(1108 13) (471 14) (479 15) (330 16) (280 17) (224 18) (296 19) (161 20) (242 21) (405 22) (269 23)
(223 24) (365 25) (222 26) (297 27) (66 28) (84 29) (246 30) (147 31) (10 32) (23 33) (8 34) (8 35)
(153 36) (70 37) (336 38) (376 39) (843 40) (1251 41) (475 42) (2562 43) (116 44) (292 45) (1139 46)
(1029 47) (1204 48) (736 49) (1068 50) (796 51) (467 52) (184 53) (432 54) (587 55) (588 56) (996
57) (1004 58) (0 59) (501 60) (548 61) (970 62) (598 63))
> (def bbt_hid _hid)
BBT_HID
> bbt_hid
((868 1) (1435 2) (124 3) (643 4) (769 5) (753 6) (974 7) (1736 8) (494 9) (561 10) (704 11) (503 12)
(1108 13) (471 14) (479 15) (330 16) (280 17) (224 18) (296 19) (161 20) (242 21) (405 22) (269 23)
(223 24) (365 25) (222 26) (297 27) (66 28) (84 29) (246 30) (147 31) (10 32) (23 33) (8 34) (8 35)
(153 36) (70 37) (336 38) (376 39) (843 40) (1251 41) (475 42) (2562 43) (116 44) (292 45) (1139 46)
(1029 47) (1204 48) (736 49) (1068 50) (796 51) (467 52) (184 53) (432 54) (587 55) (588 56) (996
57) (1004 58) (0 59) (501 60) (548 61) (970 62) (598 63))
> (savevar 'bbt_hid "bbt_hid")
(BBT_HID)
> (defun error_% (observed actual)
    (* (/ (- actual observed) actual) 100))
ERROR_%

; Note the above I accidentally define the error as actual – observed
; it should be observed – actual i.e. (- observed actual)
; so the final printed results of the error % had to be multiply by –1.
; This was done in Excel spreadsheet and reported in the main text.

> (rmse 1 2)
Error: The function RMSE is unbound.
>
(defun sq_d (list_no mean_x)
    (do ((x list_no (cdr x)))
        ((null x) d_sq)
    (setq d_sq (+ d_sq (^ (- (car x) mean_x) 2)))))
SQ_D
>
(defun rmse (observed_list actual)
     (sqrt (/ (sq_d observed_list actual) (length observed_list))))
RMSE
> (rmse 1 2)
Error: The variable D_SQ is unbound.
Happened in: #<FSubr-DO: #1d57393a>

> ; loading "C:\USERS\LAW\GIS\HO\DATA\BBTHID.LSP"
(length (car new_list))
301
> (last (car new_list))
(63)
> (last (last bbt_hid))
((598 63))
> (last (car new_list))
(63)
> (last (car bbt_hid))
(1)
> (def bbt_hid (reverse bbt_hid))
BBT_HID
> (last (car bbt_hid))
(63)
> bbt_hid
((598 63) (970 62) (548 61) (501 60) (0 59) (1004 58) (996 57) (588 56) (587 55) (432 54) (184 53)
(467 52) (796 51) (1068 50) (736 49) (1204 48) (1029 47) (1139 46) (292 45) (116 44) (2562 43) (475
```

```
42) (1251 41) (843 40) (376 39) (336 38) (70 37) (153 36) (8 35) (8 34) (23 33) (10 32) (147 31) (246
30) (84 29) (66 28) (297 27) (222 26) (365 25) (223 24) (269 23) (405 22) (242 21) (161 20) (296 19)
(224 18) (280 17) (330 16) (479 15) (471 14) (1108 13) (503 12) (704 11) (561 10) (494 9) (1736 8)
(974 7) (753 6) (769 5) (643 4) (124 3) (1435 2) (868 1))
> (car (car bbt_hid))
598
> (cdr (reverse (car new_list)))
(61 68 46 66 76 62 63 68 50 68 65 53 58 63 51 59 70 50 51 59 71 66 60 51 66 55 64 56 52 49 60 55 62
48 72 63 61 65 53 62 69 64 60 57 64 54 72 48 57 59 51 66 57 63 54 57 56 58 78 62 62 57 51 64 51 61
53 74 50 68 68 71 67 78 57 62 64 59 68 51 65 59 46 62 71 46 69 60 55 47 56 67 64 76 58 51 65 59 69
65 62 66 57 67 66 62 54 68 57 56 64 75 53 64 52 61 61 48 49 56 62 50 62 42 72 53 56 67 61 65 68 66
68 63 43 57 60 52 53 44 63 61 50 59 62 63 53 52 59 59 60 59 63 55 74 56 61 51 49 52 65 65 65 65 64
60 71 52 56 64 54 57 51 83 61 50 60 53 59 60 45 65 57 62 62 34 57 68 77 50 56 66 57 59 45 50 68 60
70 58 66 71 67 59 54 54 54 64 52 61 61 65 59 67 68 60 54 69 59 56 69 64 77 68 67 53 59 60 58 52 59
72 59 66 63 55 61 68 55 63 59 65 61 60 54 57 56 60 64 63 68 57 74 67 65 55 64 66 55 58 55 59 58 59
58 59 56 57 69 61 52 63 71 70 69 63 55 57 67 59 59 73 71 59 64 55 56 58 58 53 58 60 73 61 49 45 62
58 80 62)
> (length (cdr (reverse (car new_list))))
300
> (rmse (cdr (reverse (car new_list))) (/ (car (car bbt_hid)) 10))
Error: The variable D_SQ is unbound.
Happened in: #<FSubr-DO: #1d57393a>
> (setq d_sq 0)
0
> (rmse 1 2)
Error: bad argument type - 1
Happened in: #<Subr-CAR: #1d574d3a>
> (rmse (cdr (reverse (car new_list))) (/ (car (car bbt_hid)) 10))
7.37956638292522
> (* 10 (rmse (cdr (reverse (car new_list))) (/ (car (car bbt_hid)) 10)))
104.36282863165415
> (setq d_sq 0)
0
> (* 10 (rmse (cdr (reverse (car new_list))) (/ (car (car bbt_hid)) 10)))
73.79566382925221
> (standard-deviation (cdr (reverse (car new_list))))
7.376889248408003
> (rmse (cdr (reverse (car new_list))) (mean (cdr (reverse (car new_list)))))
10.425694221489511
> (setq d_sq 0)
0
> (rmse (cdr (reverse (car new_list))) (mean (cdr (reverse (car new_list)))))
7.364584170202682
> d_sq
16271.129999999976
> (defun round_2d (n) (/ (round (* n 100)) 100))
ROUND_2D
> (DEF BBT_HID (QUOTE ((868 1) (1435 2) (124 3) (643 4) (769 5) (753 6) (974 7) (1736 8) (494 9)
(561 10) (704 11) (503 12) (1108 13) (471 14) (479 15) (330 16) (280 17) (224 18) (296 19) (161 20)
(242 21) (405 22) (269 23) (223 24) (365 25) (222 26) (297 27) (66 28) (84 29) (246 30) (147 31) (10
32) (23 33) (8 34) (8 35) (153 36) (70 37) (336 38) (376 39) (843 40) (1251 41) (475 42) (2562 43)
(116 44) (292 45) (1139 46) (1029 47) (1204 48) (736 49) (1068 50) (796 51) (467 52) (184 53) (432
54) (587 55) (588 56) (996 57) (1004 58) (0 59) (501 60) (548 61) (970 62) (598 63))))
BBT_HID
> bbt_hid
((868 1) (1435 2) (124 3) (643 4) (769 5) (753 6) (974 7) (1736 8) (494 9) (561 10) (704 11) (503 12)
(1108 13) (471 14) (479 15) (330 16) (280 17) (224 18) (296 19) (161 20) (242 21) (405 22) (269 23)
(223 24) (365 25) (222 26) (297 27) (66 28) (84 29) (246 30) (147 31) (10 32) (23 33) (8 34) (8 35)
(153 36) (70 37) (336 38) (376 39) (843 40) (1251 41) (475 42) (2562 43) (116 44) (292 45) (1139 46)
```

```
(1029 47) (1204 48) (736 49) (1068 50) (796 51) (467 52) (184 53) (432 54) (587 55) (588 56) (996
57) (1004 58) (0 59) (501 60) (548 61) (970 62) (598 63))
> (length bbt_hid)
63
> (def new_list (reverse new_list))
NEW_LIST
> (last (car new_list))
(1)
> (defin print_tm_mc_results (list_mc list_tm)
   (dolist (i list_tm)
    (print (second i))
     (prin1 (car i))
     (setq d_sq 0)
     (def rmse (* 10 (rmse (cdr (reverse (car list_mc))) (/ (car i) 10))))
     (def sd (* 10 (rmse (cdr (reverse (car list_mc))) (mean (cdr (reverse (car list_mc)))))))
     (def sd (standard-deviation (* 10 (cdr (reverse (car list_mc))))))
     (prin1 (round_2d %_err))
     (prin1 (round_2d rmse))
     (prin1 (round_2d sd))
     (prin1 (round_2d sd-1))
     (def list_mc (cdr list_mc))))
Error: The function DEFIN is unbound.
> (defun print_tm_mc_results (list_mc list_tm)
   (dolist (i list_tm)
    (print (second i))
     (prin1 (car i))
     (setq d_sq 0)
     (def rmse (* 10 (rmse (cdr (reverse (car list_mc))) (/ (car i) 10))))
     (def sd (* 10 (rmse (cdr (reverse (car list_mc))) (mean (cdr (reverse (car list_mc)))))))
     (def sd (standard-deviation (* 10 (cdr (reverse (car list_mc))))))
     (prin1 (round_2d %_err))
     (prin1 (round_2d rmse))
     (prin1 (round_2d sd))
     (prin1 (round_2d sd-1))
     (def list_mc (cdr list_mc))))
PRINT_TM_MC_RESULTS
> sd
Error: The variable SD is unbound.
> sd-1
Error: The variable SD-1 is unbound.
> d_sq
16271.129999999976
> (setq d_sq 0)
0
> (print_tm_mc_results new_list bbt_hid)

1 868
Error: The variable %_ERR is unbound.
Happened in: #<Closure-PRINT_TM_MC_RESULTS: #164771da>
> (defun print_tm_mc_results (list_mc list_tm)
   (dolist (i list_tm)
    (print (second i))
     (prin1 (car i)) (princ " , ")
     (setq d_sq 0)
     (setq rmse 0)
     (setq sd 0)
     (setq sd-1 0)
     (setq %_err 0)
   (cond ((= 0 (car i)) (setq rmse 0)
                  (setq sd 0)
```

```
                (setq sd-1 0)
                (setq %_err 0))
    (T
     (def rmse (* 10 (rmse (cdr (reverse (car list_mc))) (/ (car i) 10))))
     (def sd (* 10 (rmse (cdr (reverse (car list_mc))) (mean (cdr (reverse (car list_mc)))))))
     (def sd-1 (* 10 (standard-deviation (cdr (reverse (car list_mc))))))
     (def %_err (error_% (* 10 (mean (cdr (reverse (car list_mc))))) (car i))))
     (def mc_mean (* 10 (mean (cdr (reverse (car list_mc))))))
     (prin1 (round_2d mc_mean))(princ " , ")
     (prin1 (round_2d %_err)) (princ " , ")
     (prin1 (round_2d rmse)) (princ " , ")
     (prin1 (round_2d sd)) (princ " , ")
     (prin1 (round_2d sd-1))
     (def list_mc (cdr list_mc))))
PRINT_TM_MC_RESULTS
> (print_tm_mc_results new_list bbt_hid)

1 868 , 866.73 , 0.15 , 88.71 , 125.45 , 88.85
2 1435 , 1430.1 , 0.34 , 104.42 , 147.6 , 104.48
3 124 , 122.27 , 1.4 , 32.39 , 45.78 , 32.4
4 643 , 644.03 , -0.16 , 76.13 , 107.66 , 76.25
5 769 , 770.6 , -0.21 , 86.83 , 122.79 , 86.96
6 753 , 745.5 , 1 , 85.29 , 120.38 , 85.10
7 974 , 969.6 , 0.45 , 89.09 , 125.91 , 89.13
8 1736 , 1746.8 , -0.62 , 125.08 , 176.56 , 124.82
9 494 , 494.03 , -0.01 , 69.39 , 98.13 , 69.5
10 561 , 563.57 , -0.46 , 71.08 , 100.49 , 71.16
11 704 , 709.2 , -0.74 , 77.87 , 110 , 77.82
12 503 , 504.07 , -0.21 , 65.32 , 92.37 , 65.42
13 1108 , 1104 , 0.36 , 98.38 , 139.07 , 98.46
14 471 , 466.8 , 0.89 , 62.67 , 88.53 , 62.63
15 479 , 481.87 , -0.6 , 62.46 , 88.29 , 62.5
16 330 , 329.2 , 0.24 , 49.33 , 69.76 , 49.4
17 280 , 282.73 , -0.98 , 49.95 , 70.58 , 49.96
18 224 , 222.7 , 0.58 , 42.29 , 59.79 , 42.34
19 296 , 303.27 , -2.45 , 50.37 , 70.87 , 49.93
20 161 , 158.83 , 1.35 , 37.13 , 52.47 , 37.13
21 242 , 243.37 , -0.56 , 45.62 , 64.5 , 45.67
22 405 , 410.8 , -1.43 , 62.27 , 87.87 , 62.1
23 269 , 265.83 , 1.18 , 47.32 , 66.84 , 47.29
24 223 , 224.43 , -0.64 , 44.83 , 63.38 , 44.88
25 365 , 368.47 , -0.95 , 56.84 , 80.3 , 56.82
26 222 , 227.47 , -2.46 , 50.11 , 70.66 , 49.9
27 297 , 290.77 , 2.1 , 51.26 , 72.22 , 50.96
28 66 , 66.57 , -0.86 , 22.19 , 31.37 , 22.22
29 84 , 85.40 , -1.67 , 28 , 39.58 , 28.01
30 246 , 245.2 , 0.33 , 43 , 60.8 , 43.06
31 147 , 145.2 , 1.22 , 34.78 , 49.15 , 34.79
32 10 , 9 , 10 , 8.94 , 12.61 , 8.9
33 23 , 21.9 , 4.78 , 14.83 , 20.94 , 14.81
34 8 , 8.43 , -5.42 , 8.76 , 12.38 , 8.76
35 8 , 8.53 , -6.67 , 8.45 , 11.94 , 8.45
36 153 , 152.8 , 0.13 , 39.11 , 55.31 , 39.17
37 70 , 70.5 , -0.71 , 23.94 , 33.85 , 23.97
38 336 , 334.37 , 0.49 , 53.33 , 75.41 , 53.4
39 376 , 376.07 , -0.02 , 54.37 , 76.90 , 54.46
40 843 , 843.43 , -0.05 , 85.60 , 121.06 , 85.75
41 1251 , 1250.97 , 0 , 100.17 , 141.66 , 100.34
42 475 , 473.03 , 0.41 , 66.77 , 94.41 , 66.85
43 2562 , 2558.8 , 0.12 , 142.83 , 201.97 , 143.04
```

44 116 , 113.87 , 1.84 , 33.73 , 47.65 , 33.72
45 292 , 294.63 , -0.9 , 47.17 , 66.66 , 47.18
46 1139 , 1148.17 , -0.8 , 96.36 , 135.97 , 96.09
47 1029 , 1027.8 , 0.12 , 88.2 , 124.72 , 88.34
48 1204 , 1208.7 , -0.39 , 100.51 , 142.07 , 100.57
49 736 , 730.83 , 0.7 , 78.22 , 110.49 , 78.18
50 1068 , 1065.9 , 0.2 , 98.08 , 138.69 , 98.22
51 796 , 797.43 , -0.18 , 84.58 , 119.6 , 84.71
52 467 , 466.07 , 0.2 , 65.19 , 92.18 , 65.29
53 184 , 183.97 , 0.02 , 38.44 , 54.36 , 38.5
54 432 , 432.7 , -0.16 , 59.45 , 84.08 , 59.55
55 587 , 581.33 , 0.97 , 72.27 , 102.05 , 72.17
56 588 , 588.07 , -0.01 , 71.12 , 100.58 , 71.24
57 996 , 999.5 , -0.35 , 90.56 , 128.03 , 90.65
58 1004 , 1004.17 , -0.02 , 90.90 , 128.55 , 91.05
59 0 , 0 , 0 , 0 , 0 , 0
60 501 , 506.97 , -1.19 , 67.7 , 95.55 , 67.55
61 548 , 554.77 , -1.23 , 66.88 , 94.34 , 66.65
62 970 , 974.23 , -0.44 , 93.75 , 132.52 , 93.81
63 598 , 602.7 , -0.79 , 73.8 , 104.26 , 73.77
NIL
>

; 131296.log

> ; loading "C:\USERS\LAW\GIS\HO\DATA\POSTBBT.LSP"

1 870.0" "563.37 ±57.66" "-54.43 ±0.1
2 5830.0" "5362.88 ±391.58" "-8.71 ±1.28
3 635.0" "462.18 ±122.43" "-37.39 ±5.29
4 2077.0" "2743.57 ±324.31" "24.3 ±-0.68
5 2288.0" "2350.33 ±264.83" "2.65 ±-0.64
6 2654.0" "2333.42 ±266.96" "-13.74 ±3.13
7 3680.0" "2656.7 ±244.11" "-38.52 ±1.23
8 3508.0" "4000.17 ±286.43" "12.3 ±-1.42
9 1910.0" "2030.46 ±285.19" "5.93 ±-0.04
10 3297.0" "2716.41 ±342.61" "-21.37 ±-2.22
11 2108.0" "2155.97 ±236.72" "2.22 ±-2.25
12 2866.0" "2777.43 ±359.86" "-3.19 ±-1.16
13 2674.0" "3963.36 ±353.18" "32.53 ±1.29
14 2173.0" "2249.98 ±302.07" "3.42 ±4.29
15 1772.0" "1811.83 ±234.85" "2.2 ±-2.26
16 1438.0" "1520.9 ±227.9" "5.45 ±1.11
17 1078.0" "1159.19 ±204.8" "7 ±-4.02
18 1290.0" "1144.68 ±217.37" "-12.7 ±2.98
19 1439.0" "1731.67 ±287.61" "16.9 ±-13.99
20 1538.0" "1516.83 ±354.59" "-1.4 ±12.89
21 1877.0" "1820.41 ±341.24" "-3.11 ±-4.19
22 2519.0" "2123.84 ±321.94" "-18.61 ±-7.39
23 1684.0" "1621.56 ±288.65" "-3.85 ±7.2
24 851.0" "1117.66 ±223.25" "23.86 ±-3.19
25 2173.0" "2008.16 ±309.78" "-8.21 ±-5.18
26 1478.0" "1462.63 ±322.21" "-1.05 ±-15.82
27 2207.0" "2169.14 ±382.4" "-1.75 ±15.67
28 0" "0 ±0" " ±0
29 271.0" "462.87 ±151.76" "41.45 ±-9.05
30 2352.0" "1233.36 ±216.29" "-90.7 ±1.66
31 727.0" "996.07 ±238.59" "27.01 ±8.37
32 0" "0 ±0" " ±0
33 164.0" "49.28 ±33.37" "-232.83 ±10.76

```
34 0" "0 ±0" " ±0
35 0" "0 ±0" " ±0
36 1545.0" "1236.15 ±316.4" "-24.98 ±1.05
37 923.0" "1328.22 ±451.03" "30.51 ±-13.38
38 3014.0" "3230.01 ±515.17" "6.69 ±4.73
39 1881.0" "2933.35 ±424.09" "35.88 ±-0.16
40 4541.0" "3711.09 ±376.6" "-22.36 ±-0.22
41 6295.0" "6655.16 ±532.9" "5.41 ±0
42 3594.0" "2838.18 ±400.62" "-26.63 ±2.46
43 10183.0" "10567.84 ±589.89" "3.64 ±0.5
44 154.0" "190.16 ±56.33" "19.02 ±3.07
45 3782.0" "2645.78 ±423.59" "-42.94 ±-8.08
46 4269.0" "5028.98 ±422.06" "15.11 ±-3.5
47 4447.0" "4563.43 ±391.61" "2.55 ±0.53
48 3980.0" "4266.71 ±354.8" "6.72 ±-1.38
49 2890.0" "2440.97 ±261.25" "-18.4 ±2.34
50 3215.0" "2888.59 ±265.8" "-11.3 ±0.54
51 2615.0" "3213.64 ±340.86" "18.63 ±-0.73
52 3813.0" "3933.63 ±550.2" "3.07 ±1.69
53 1779.0" "1880.17 ±392.86" "5.38 ±0.2
54 1997.0" "1882.24 ±258.61" "-6.1 ±-0.7
55 1865.0" "1993.96 ±247.89" "6.47 ±3.33
56 2616.0" "2622.79 ±317.2" "0.26 ±-0.04
57 2794.0" "3288.36 ±297.94" "15.03 ±-1.15
58 4014.0" "4669.39 ±422.69" "14.04 ±-0.09
59 261.0" "0 ±0" " ±0
60 3683.0" "3817.48 ±509.78" "3.52 ±-8.96
61 4216.0" "4754.38 ±573.16" "11.32 ±-10.54
62 6269.0" "10599.62 ±1020" "40.86 ±-4.79
63 3359.0" "2356.56 ±288.56" "-42.54 ±-3.09
>
```

; 040397.Log
; Printing histograms
; For Bristol

```
> (def bbthid (reverse new_list))
BBTHID
> (last (car bbthid))
(1)
> (def list1 (car bbthid))
LIST1
(def list1 (reverse list1))
(def list1 (cdr list1))
(histogram list1)

> (defun list_histogram (alists)
   (dolist (list1 alists)
      (def list1 (reverse list1))
      (def list1 (cdr list1))
      (histogram list1)))

LIST_HISTOGRAM
>
```

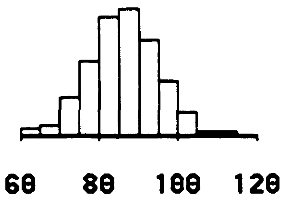
; Similarly for Coventry...

; see file: lsplog.doc

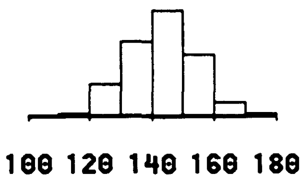
; file: lsplog.doc
; Frequency distribution of the Monte Carlo sampling

; The following two sets of histograms show the frequency distribution of household-pixels based on the Monte Carlo simulation for Bristol and Coventry respectively

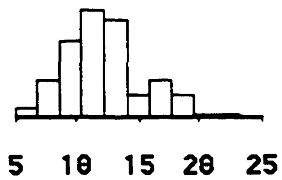
; Bristol
; The frequency distribution of household-pixels in each beat: Beat-ID from 1 to 63 (except Beat 59 which contain no household pixel):



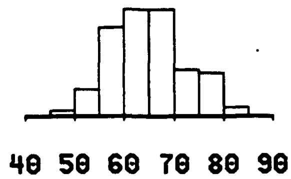
Beat 1



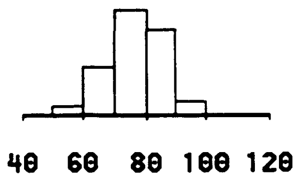
Beat 2



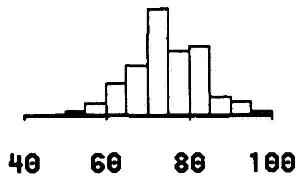
Beat 3



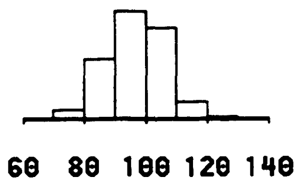
Beat 4



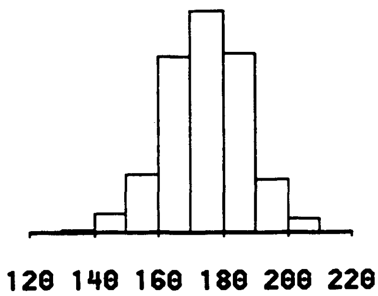
Beat 5



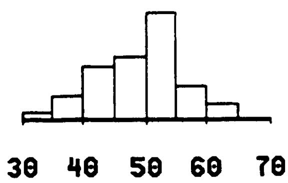
Beat 6



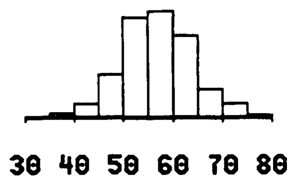
Beat 7



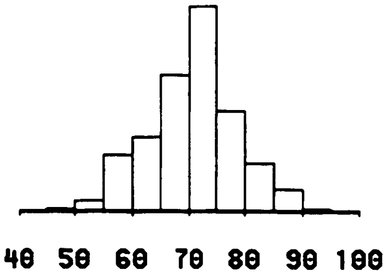
Beat 8



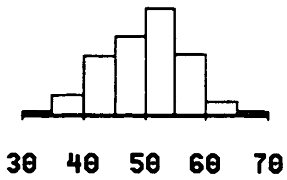
Beat 9



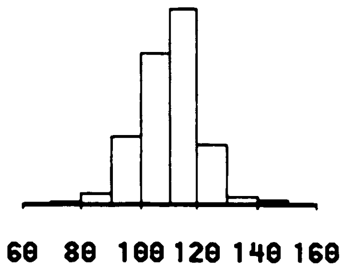
Beat 10



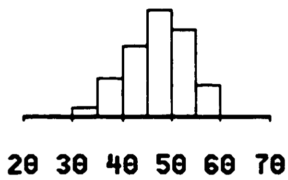
Beat 11



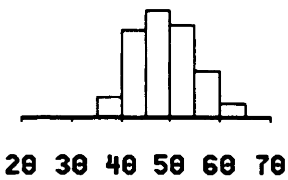
Beat 12



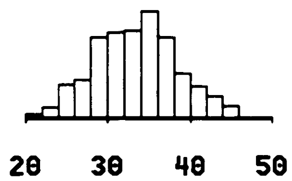
Beat 13



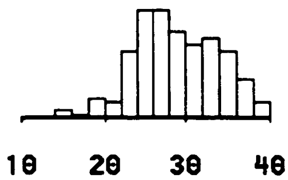
Beat 14



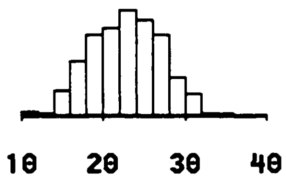
Beat 15



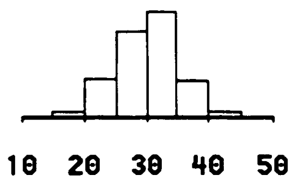
Beat 16



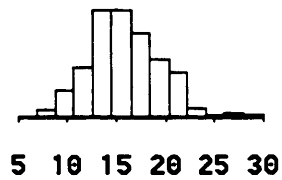
Beat 17



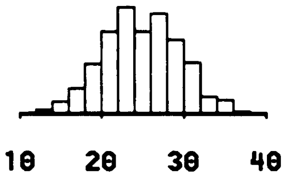
Beat 18



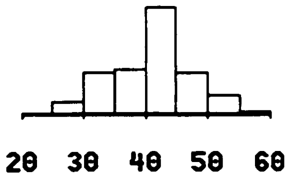
Beat 19



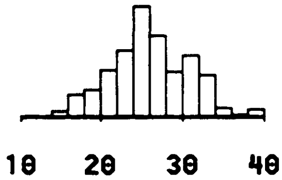
Beat 20



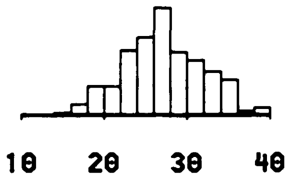
Beat 21



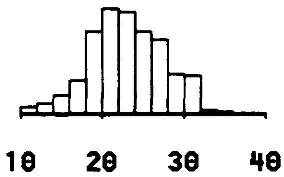
Beat 22



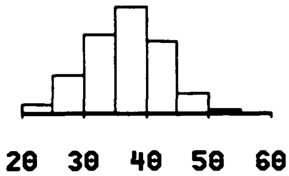
Beat 23



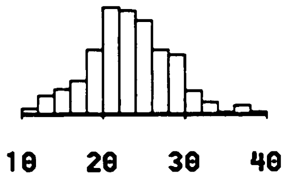
Beat 24



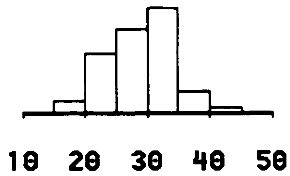
Beat 2 5



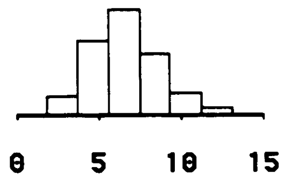
Beat 26



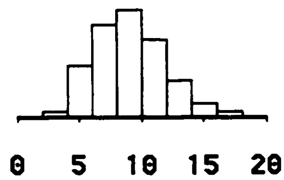
Beat 27



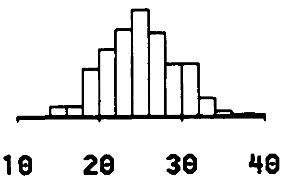
Beat 28



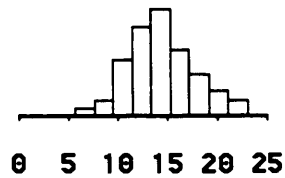
Beat 29



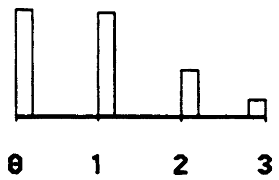
Beat 30



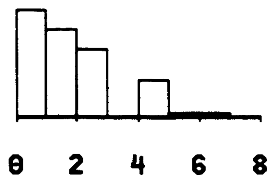
3 Beat 1



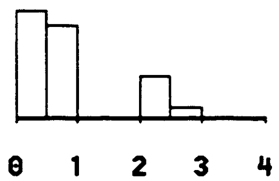
Beat 32



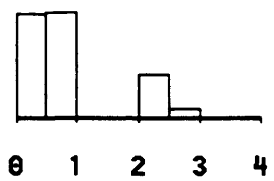
Beat 33



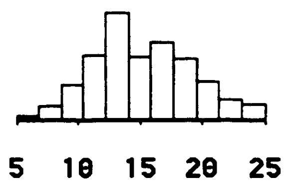
Beat 34



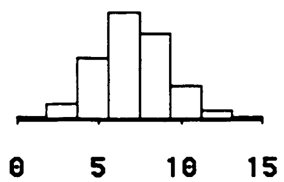
Beat 35



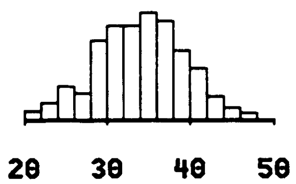
Beat 36



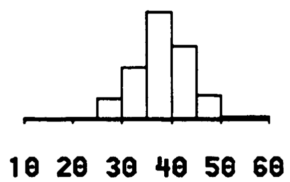
Beat 37



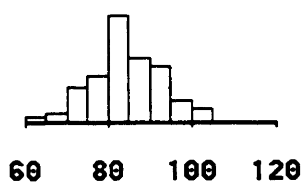
Beat 38



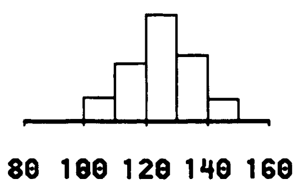
Beat 39



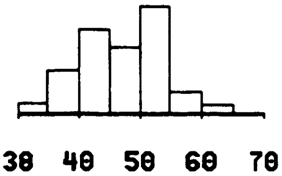
Beat 40



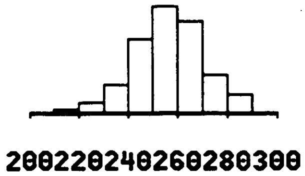
Beat 41



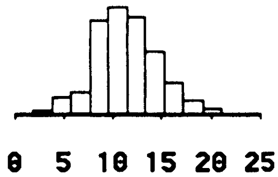
Beat 42



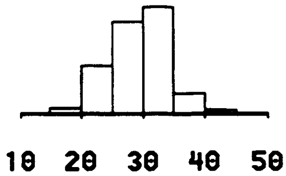
Beat 43



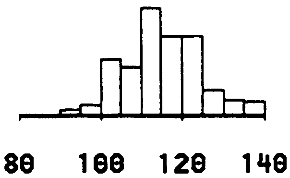
Beat 44



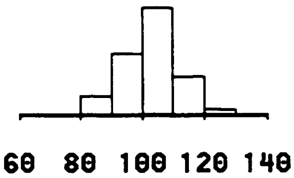
Beat 45



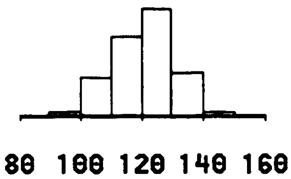
Beat 46



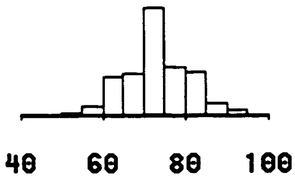
Beat 47



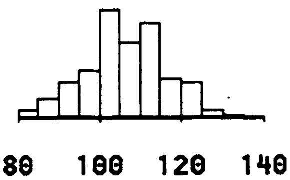
Beat 48



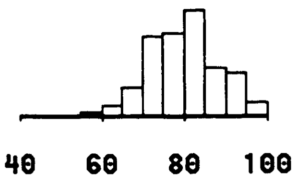
Beat 49



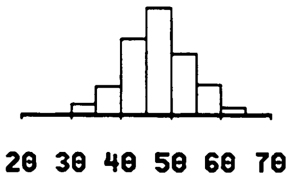
Beat 50



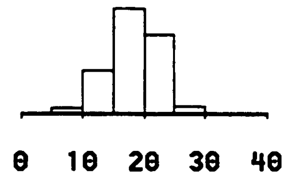
Beat 51



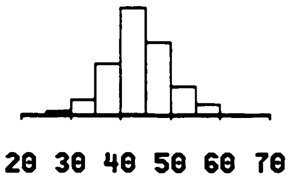
Beat 52



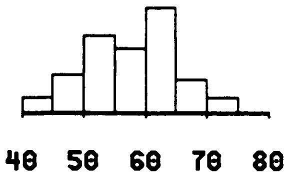
Beat 53



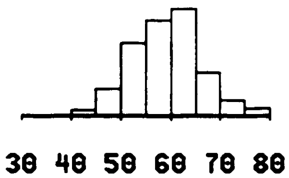
Beat 54



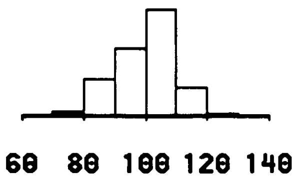
Beat 55



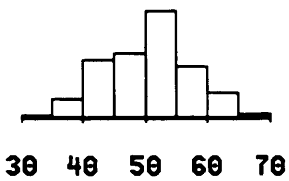
Beat 56



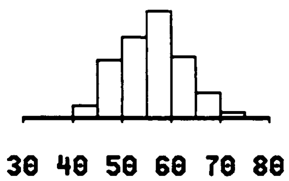
Beat 57



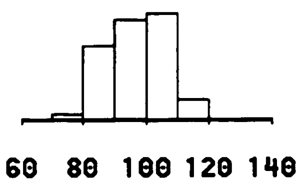
Beat 58



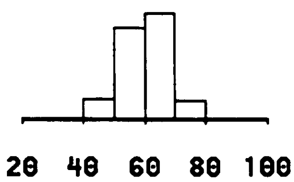
Beat 60



Beat 61

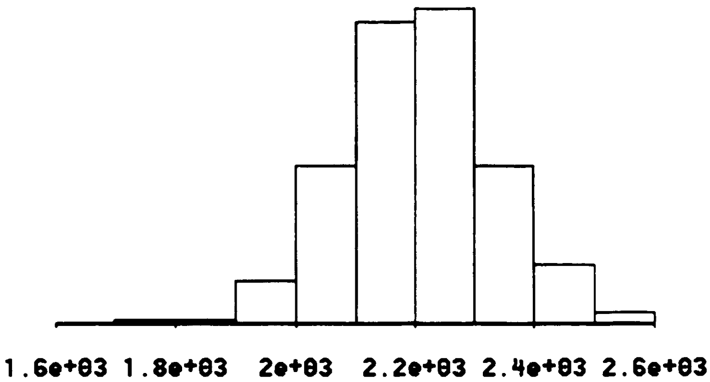


Beat 62

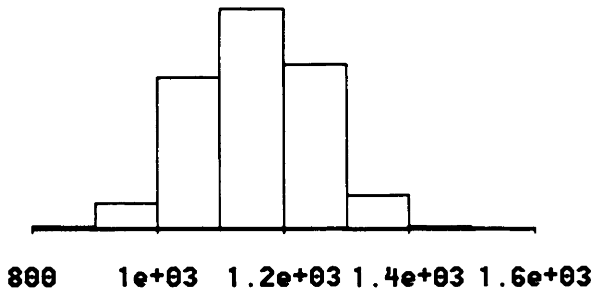


Beat 63

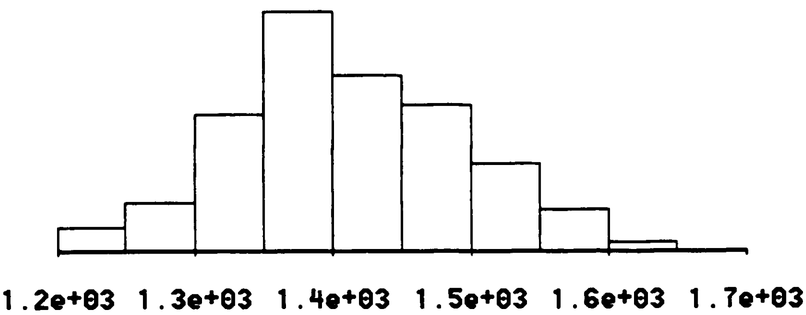
; Coventry
; The frequency distribution of household-pixels in each beat (except Beat 1 which is outside the city boundary):



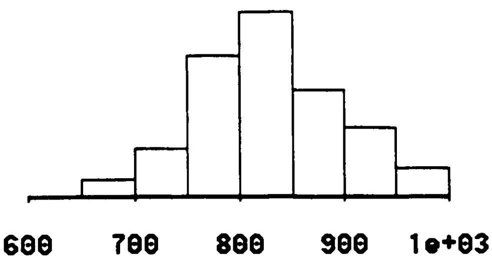
Beat 2



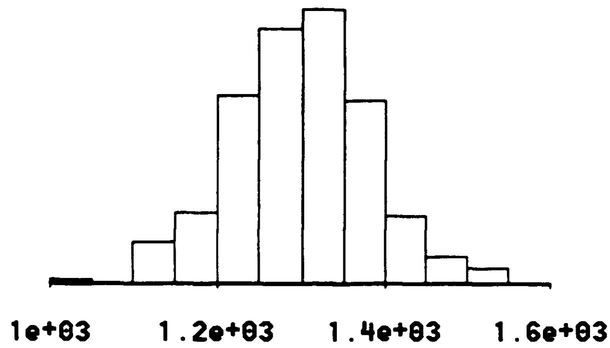
Beat 5



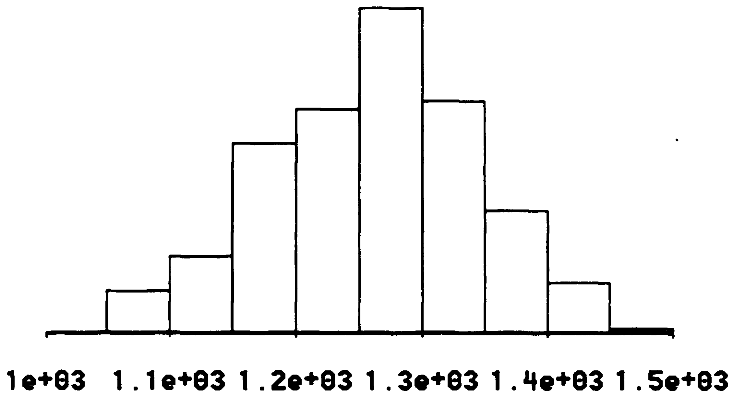
Beat 6



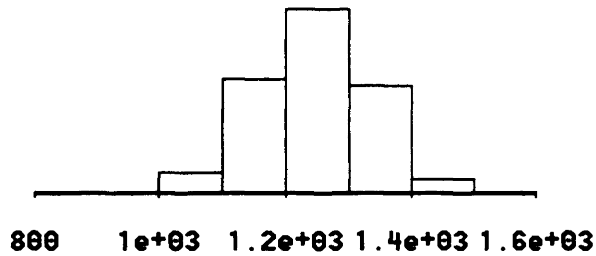
Beat 7



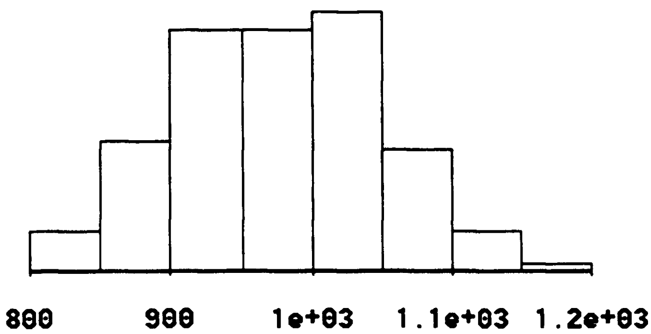
Beat 8



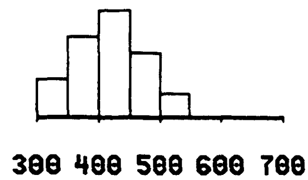
Beat 9



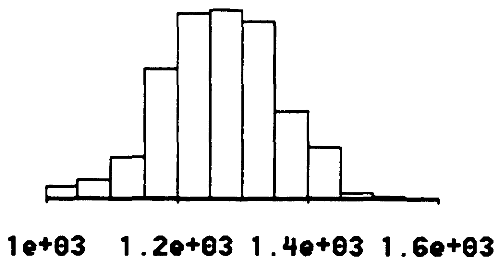
Beat 10



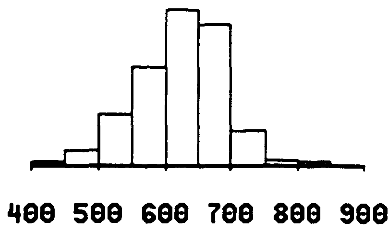
Beat 11



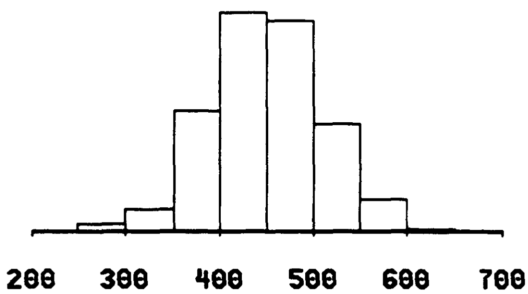
Beat 14



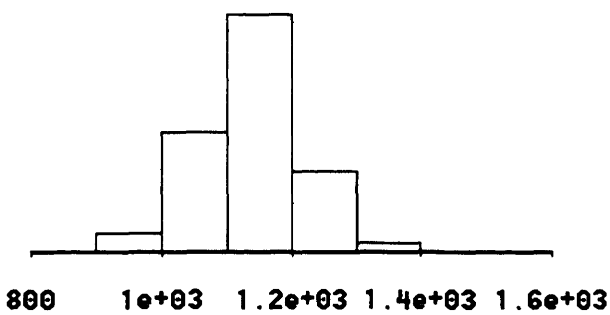
Beat 15



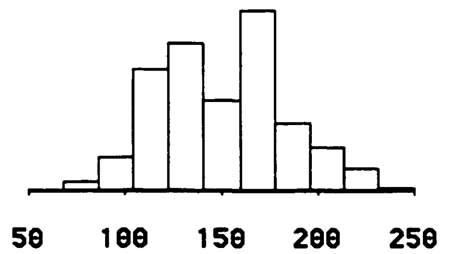
Beat 16



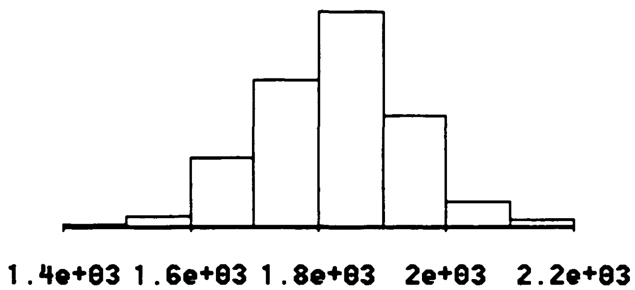
Beat 17



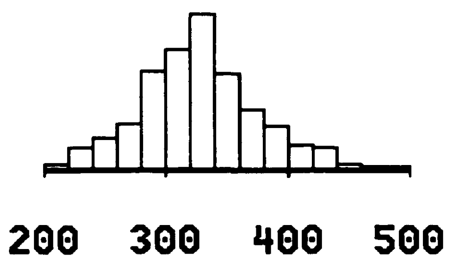
Beat 18



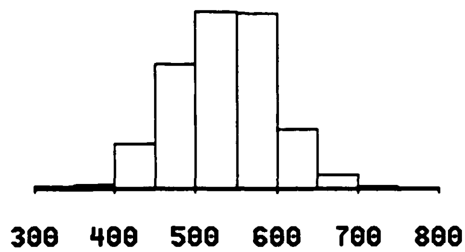
Beat 19



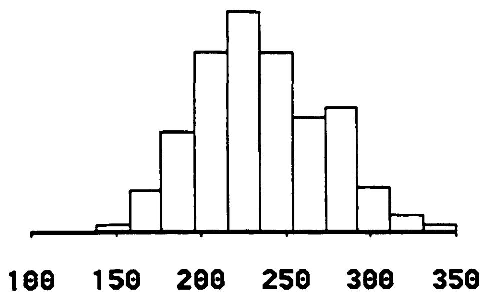
Beat 20



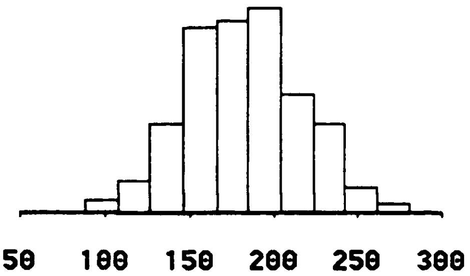
Beat 21



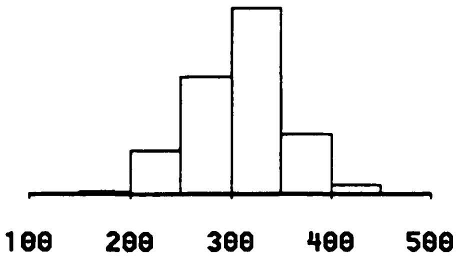
Beat 22



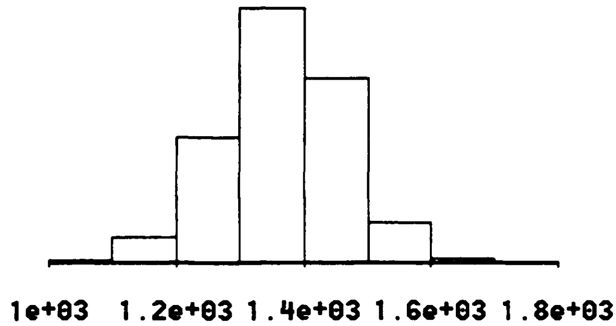
Beat 23



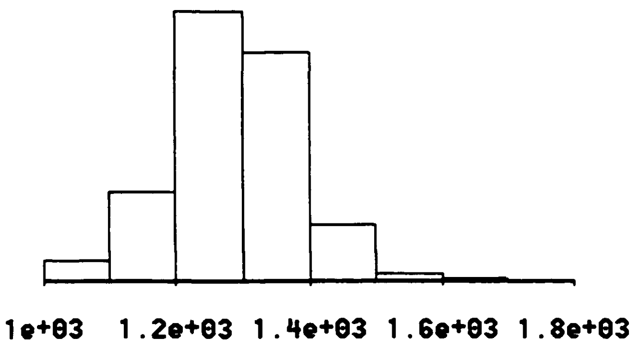
Beat 24



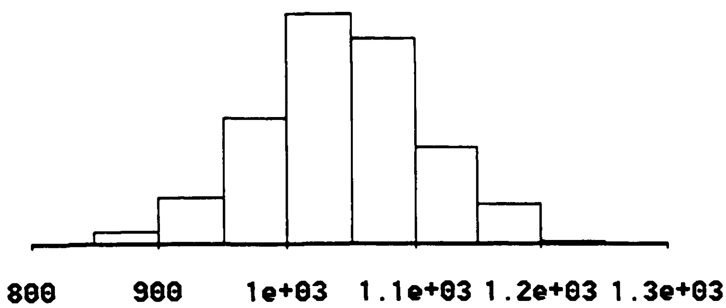
Beat 25



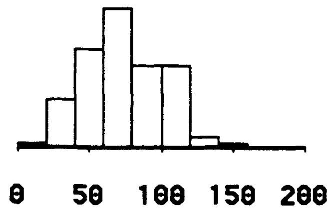
Beat 26



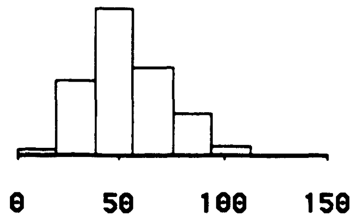
Beat 27



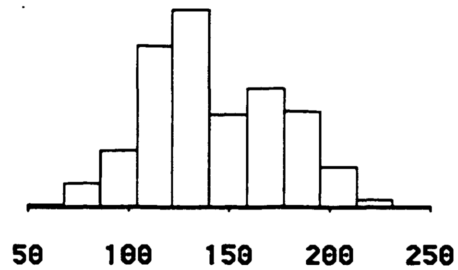
Beat 28



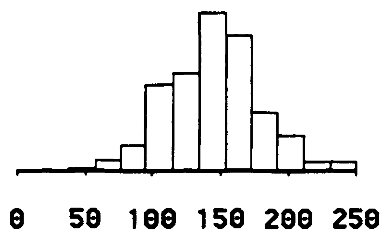
Beat 29



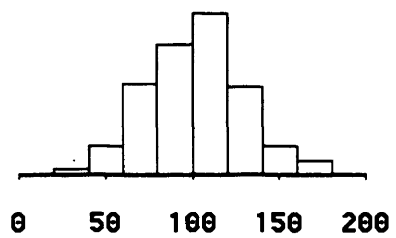
Beat 30



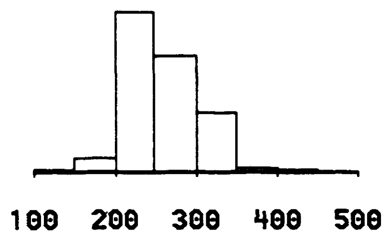
Beat 31



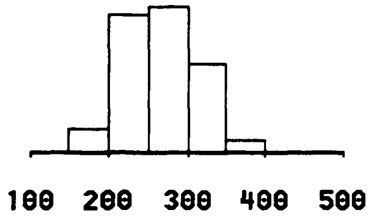
Beat 32



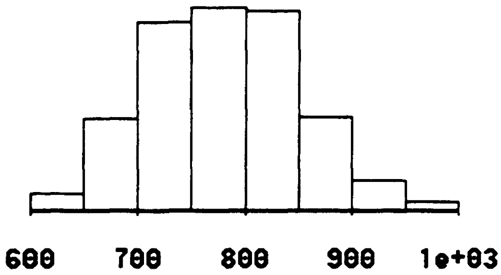
Beat 33



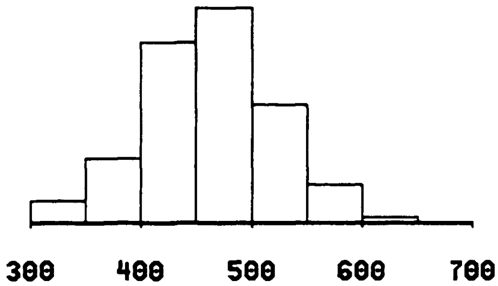
Beat 34



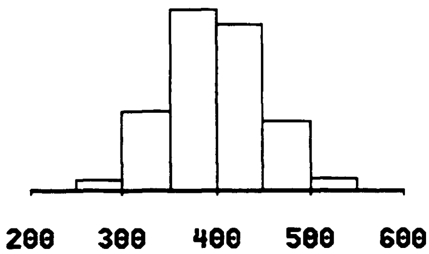
Beat 35



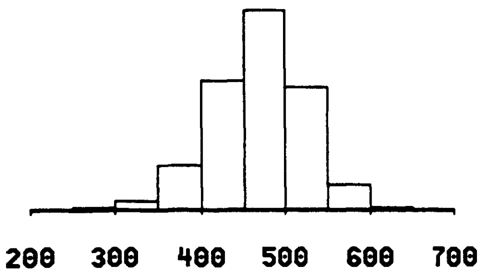
Beat 36



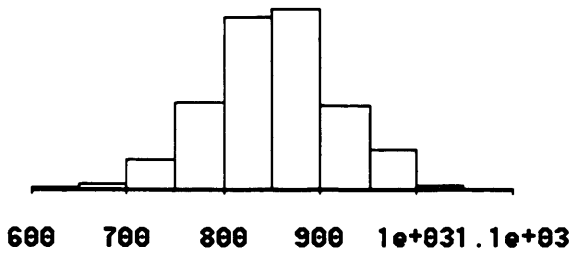
Beat 37



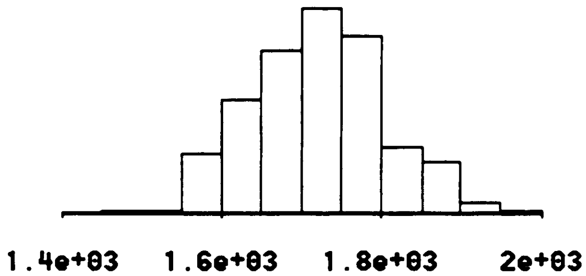
Beat 38



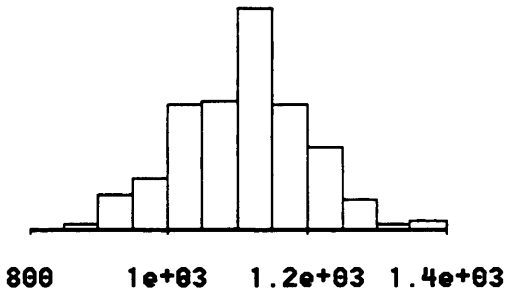
Beat 39



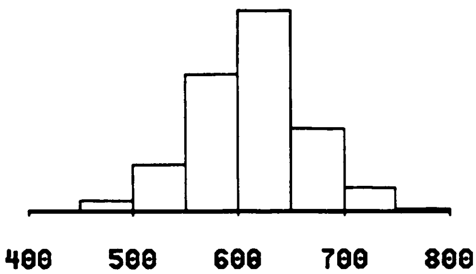
Beat 40



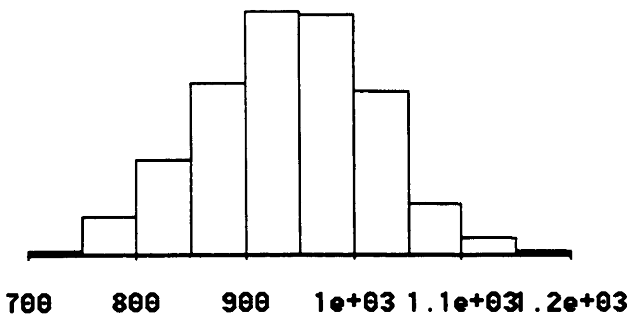
Beat 41



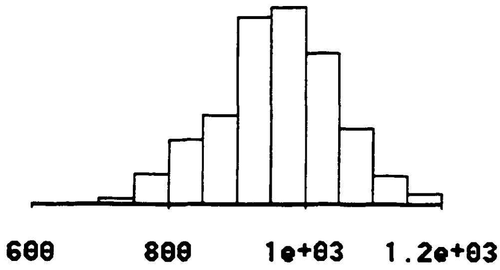
Beat 42



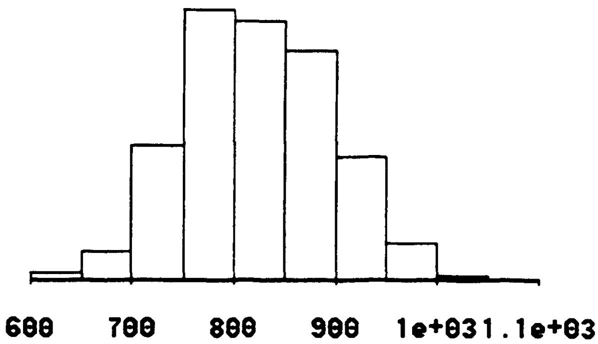
Beat 43



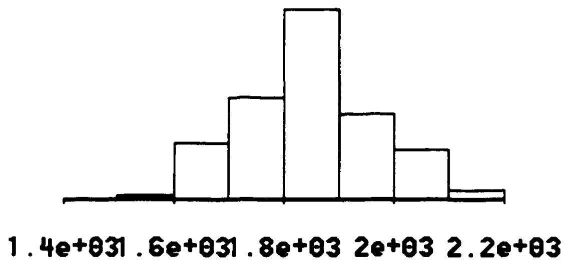
Beat 44



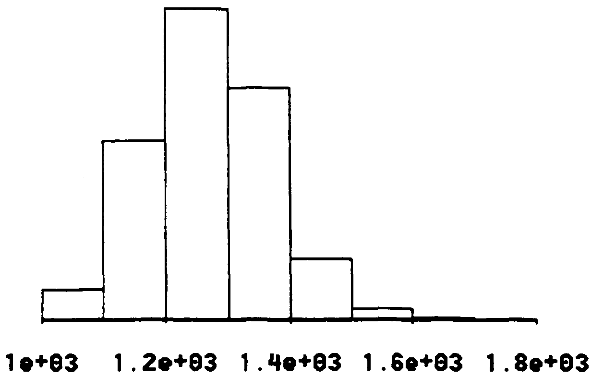
Beat 45



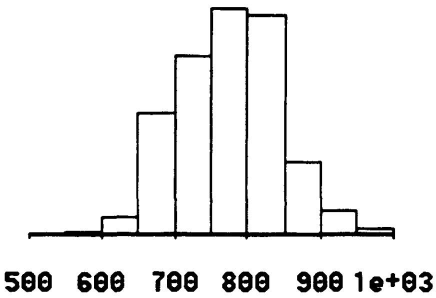
Beat 46



Beat 47



Beat 100



Beat 101

Appendix 9.1

Thematic maps of the Safer Cities action intensity

This appendix shows the whole range of beat maps of the action intensity for burglary prevention for each year of the Safer Cities Programme. The action intensity score is defined as the average amount of funds acting on each household over a given year (Ekblom et al, 1996).

The overlay method and the Monte Carlo method were implemented using ARC/INFO and XLISP-STAT respectively. MapInfo (version 4) was used to produce the lay out for printing the final output maps. See the main text for the explanation and interpretation of the maps.

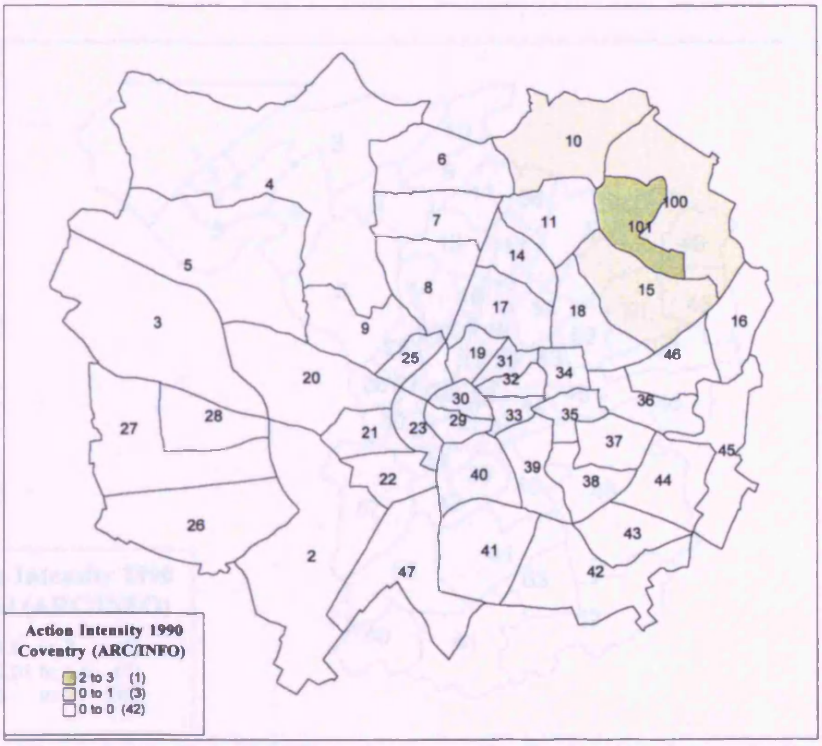


Figure A9. 1: Coventry 1990 Action score distribution (overlay method)

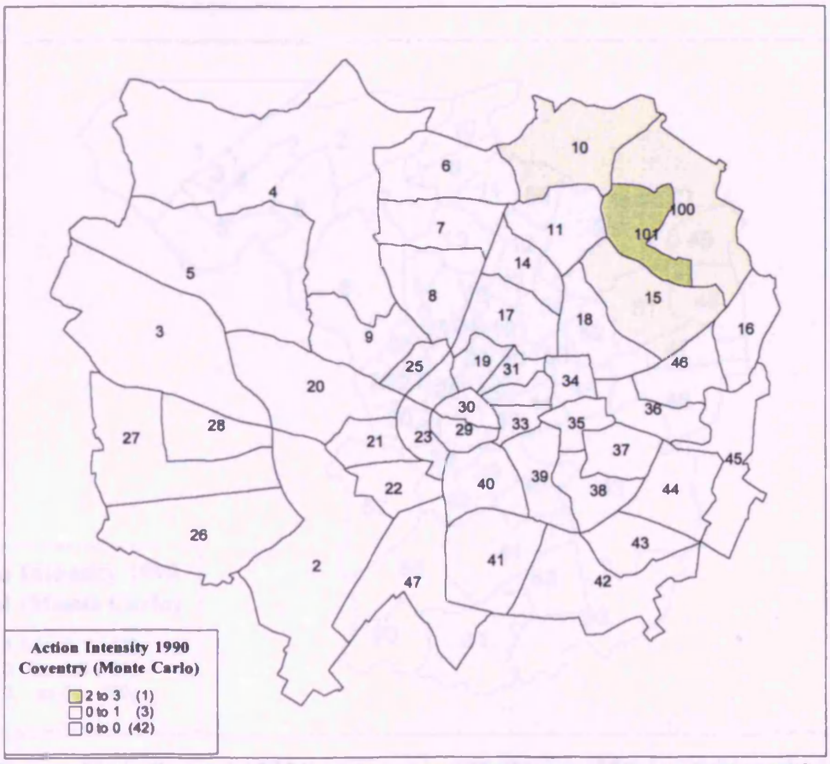


Figure A9. 2: Coventry 1990 Action score distribution (Monte Carlo method)

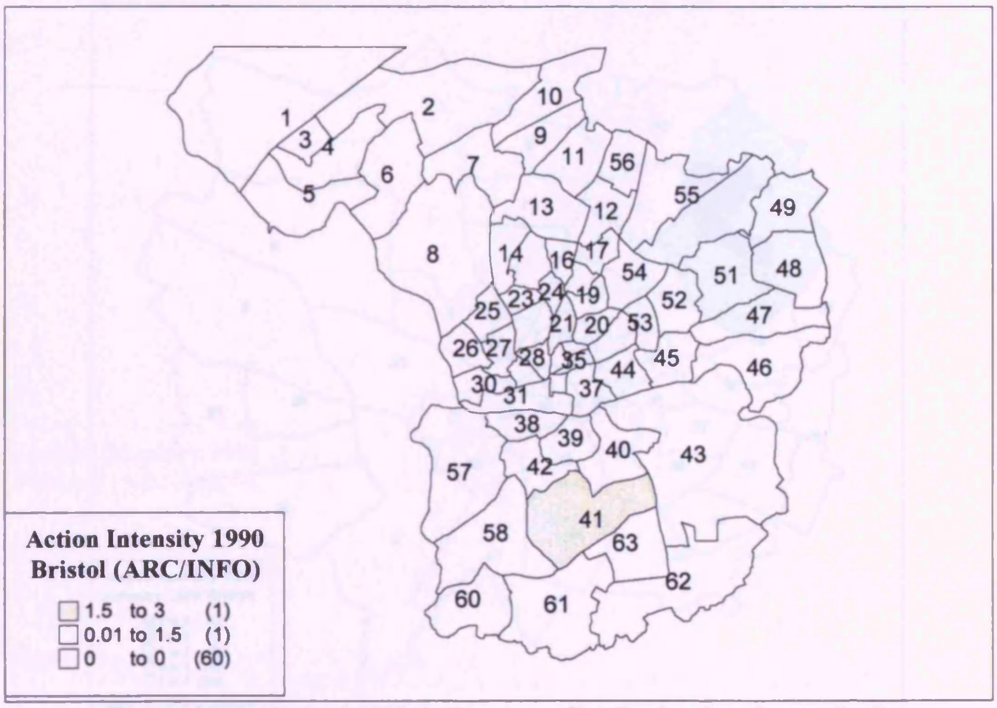


Figure A9. 3: Bristol 1990 Action score distribution (overlay method)

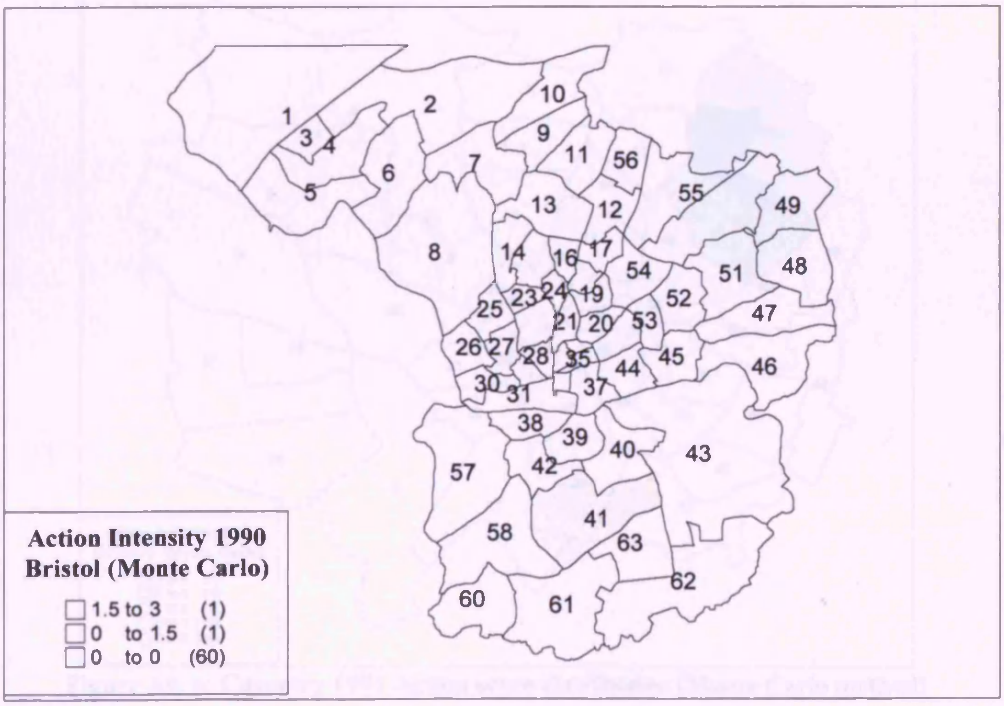


Figure A9. 4: Bristol 1990 Action score distribution (Monte Carlo method)

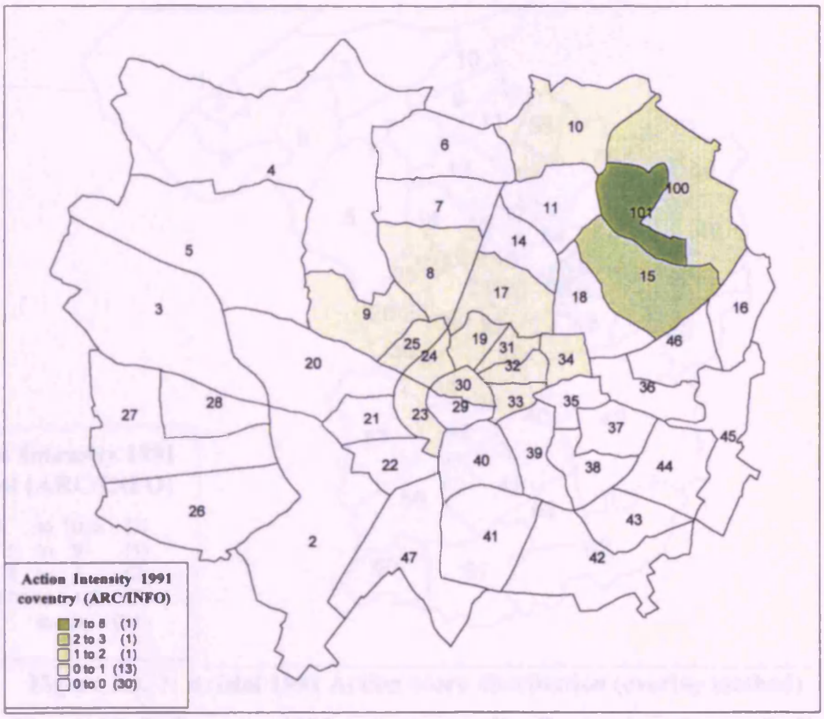


Figure A9. 5: Coventry 1991 Action score distribution (overlay method)

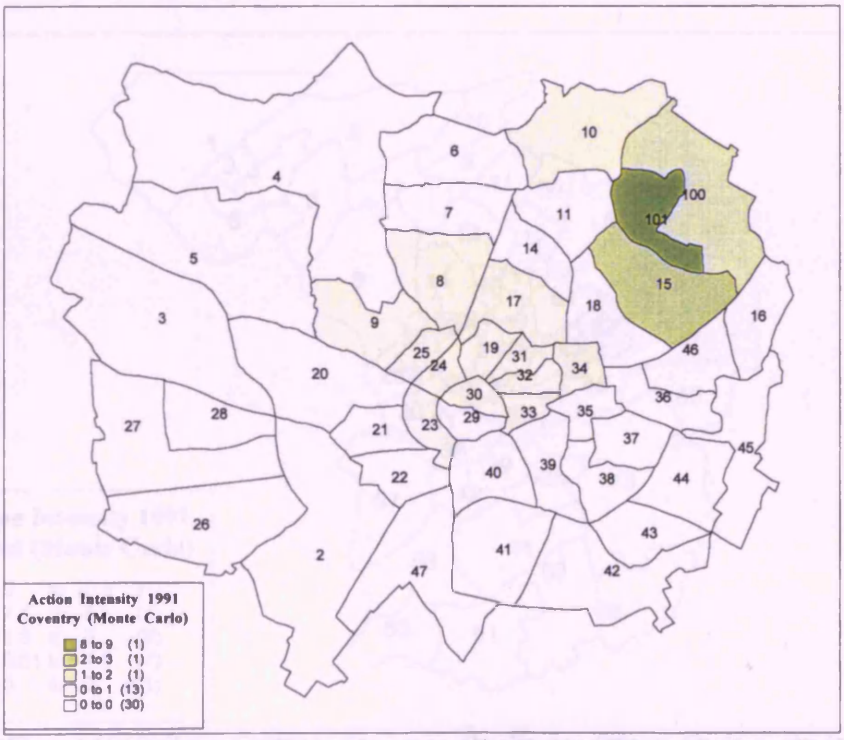


Figure A9. 6: Coventry 1991 Action score distribution (Monte Carlo method)

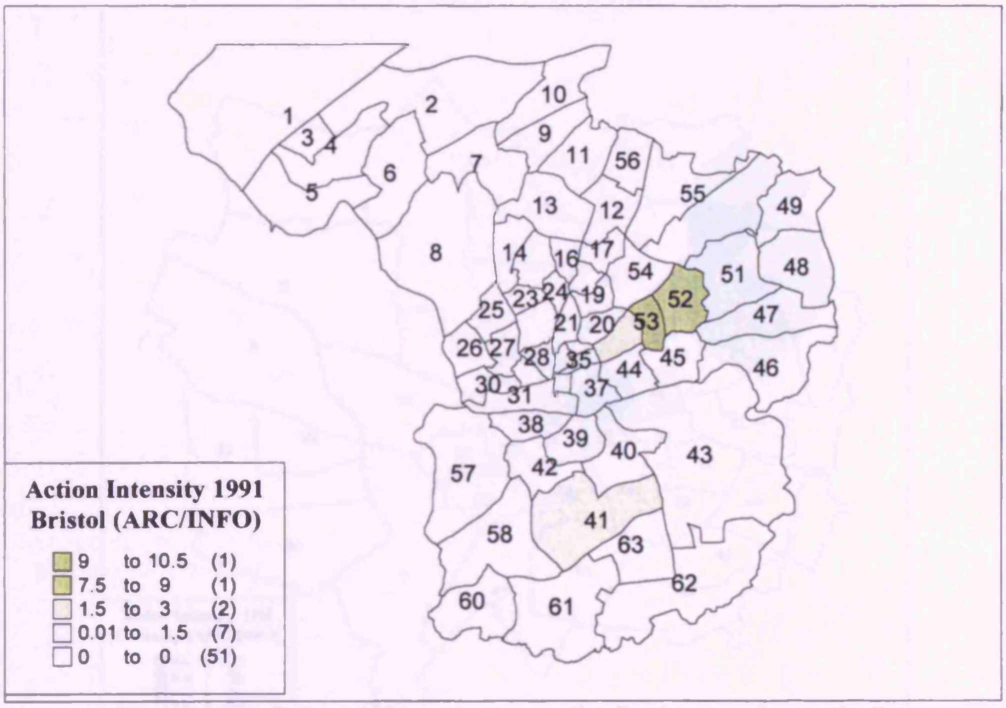


Figure A9. 7: Bristol 1991 Action score distribution (overlay method)

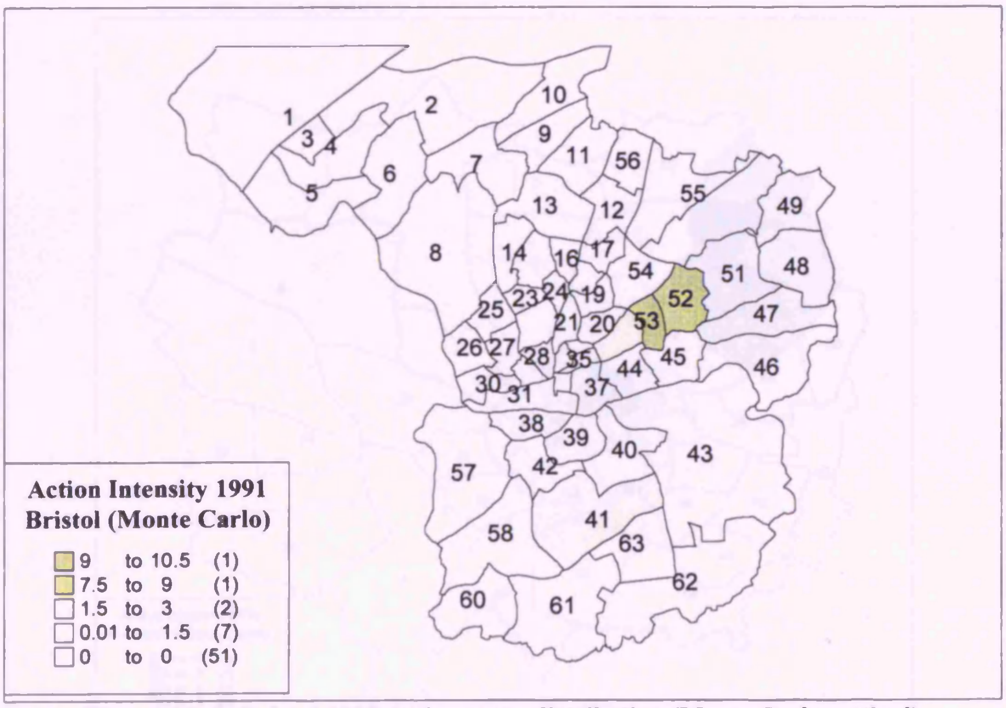


Figure A9. 8: Bristol 1991 Action score distribution (Monte Carlo method)

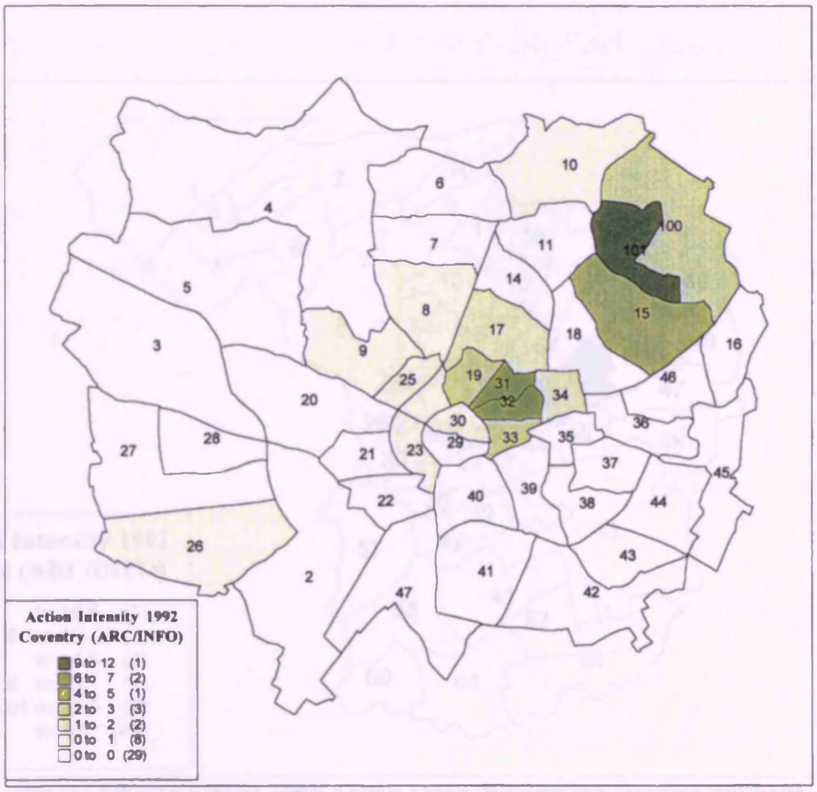


Figure A9. 9: Coventry 1992 Action score distribution (overlay method)

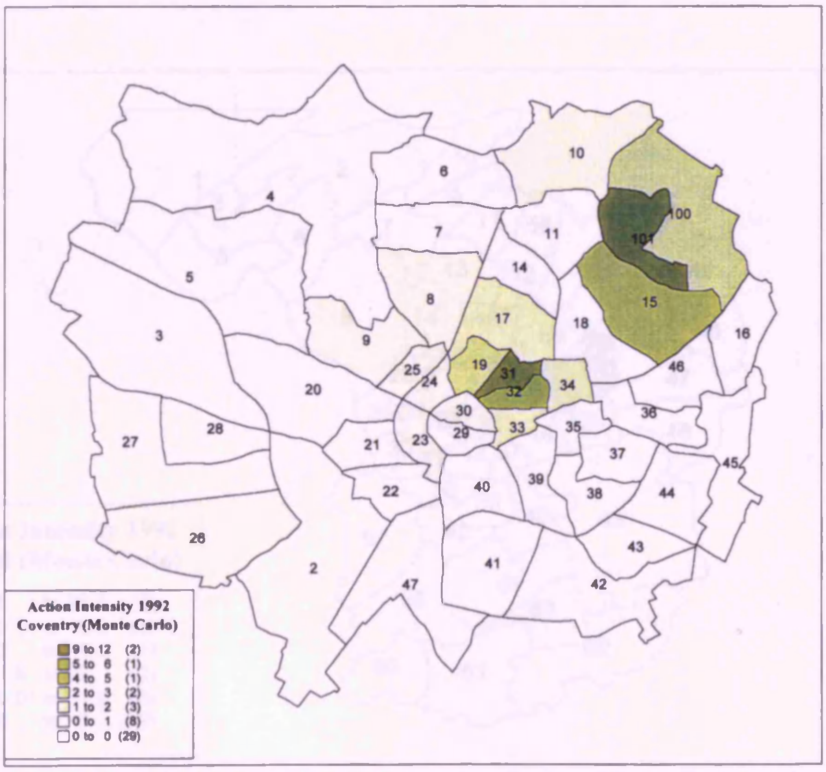


Figure A9. 10: Coventry 1992 Action score distribution (Monte Carlo method)

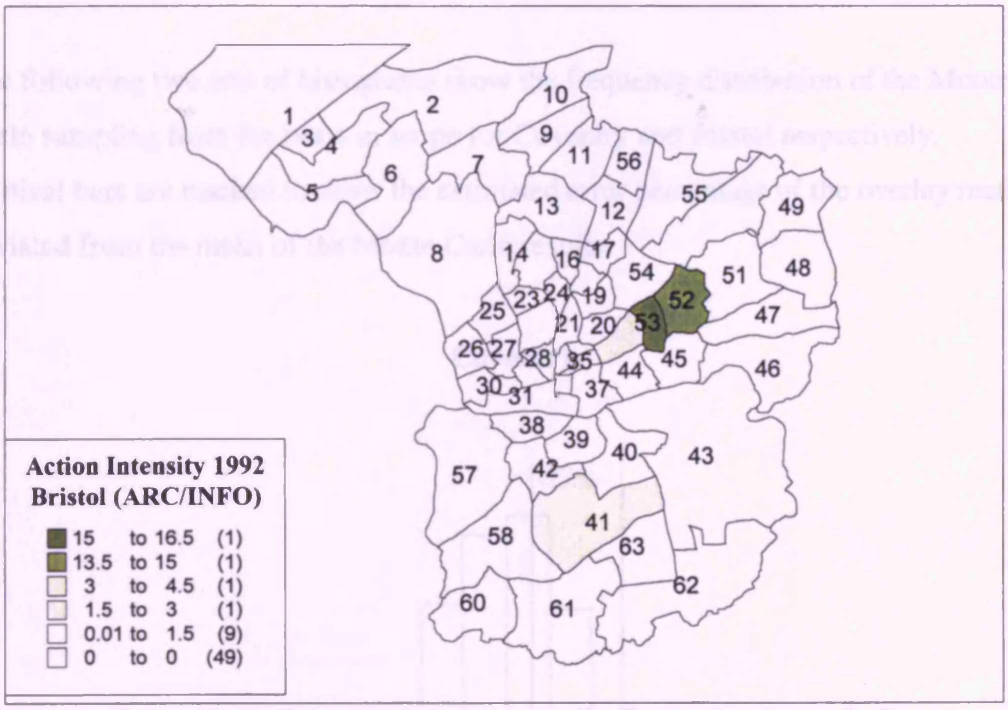


Figure A9. 11: Bristol 1992 Action score distribution (overlay method)

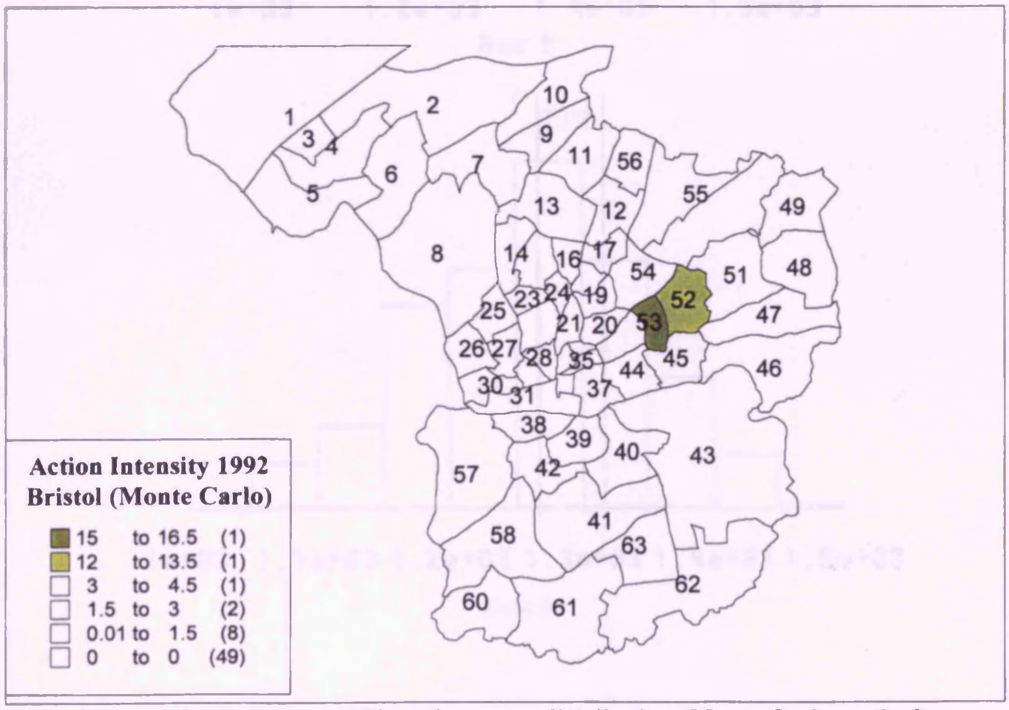


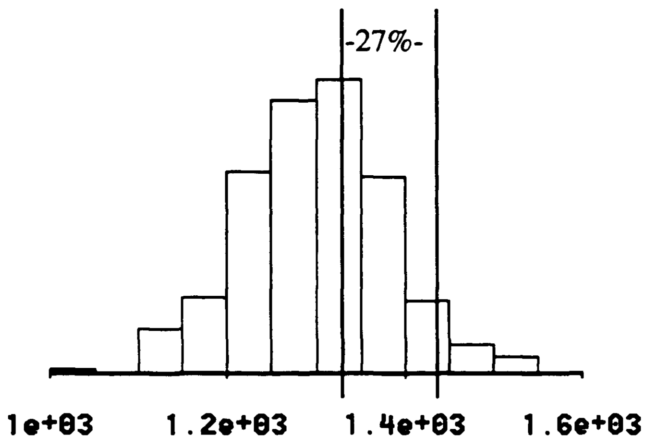
Figure A9. 12: Bristol 1992 Action score distribution (Monte Carlo method)

Appendix 9.2

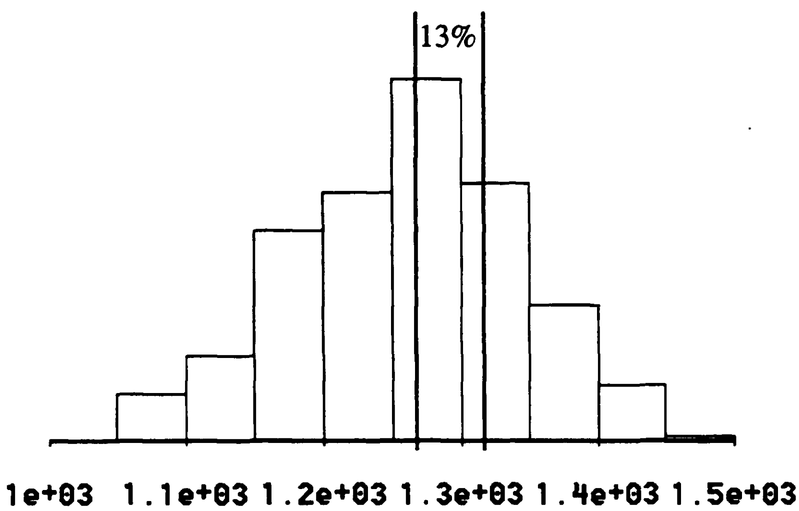
Frequency distribution of the Monte Carlo sampling

The following two sets of histograms show the frequency distribution of the Monte Carlo sampling from the beats in scope for Coventry and Bristol respectively. Vertical bars are marked to show the estimated error percentage of the overlay method deviated from the mean of the Monte Carlo results

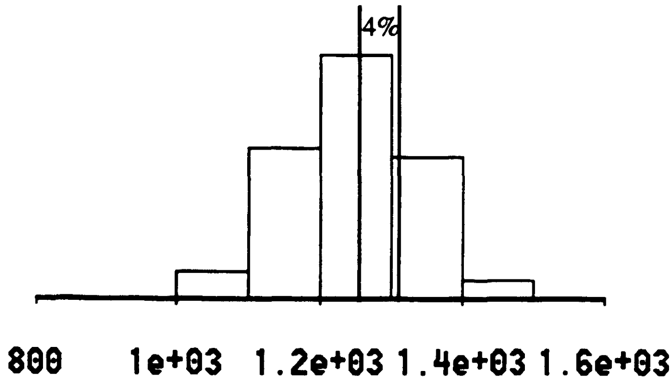
Coventry



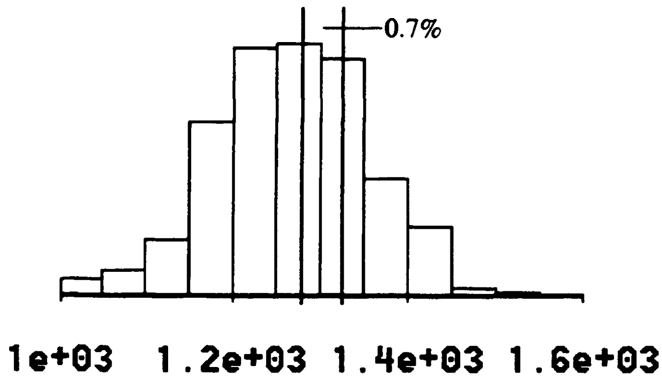
Beat 8



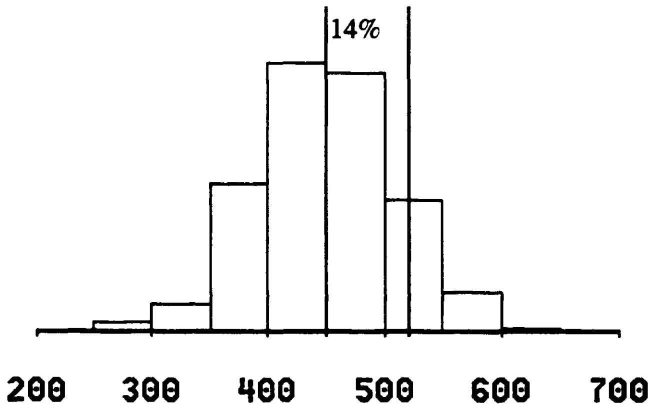
Beat 9



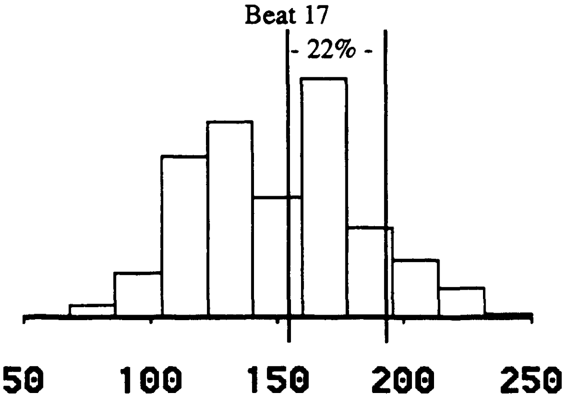
Beat 10



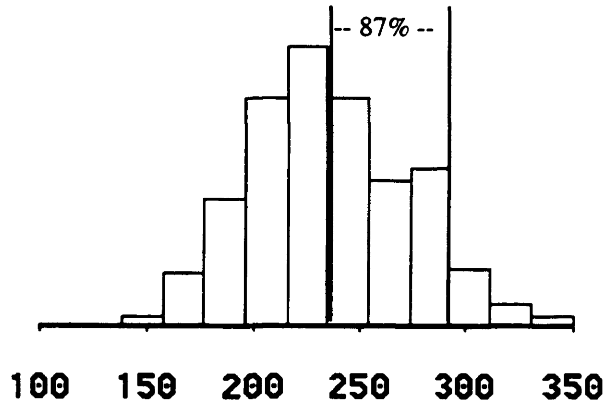
Beat 15



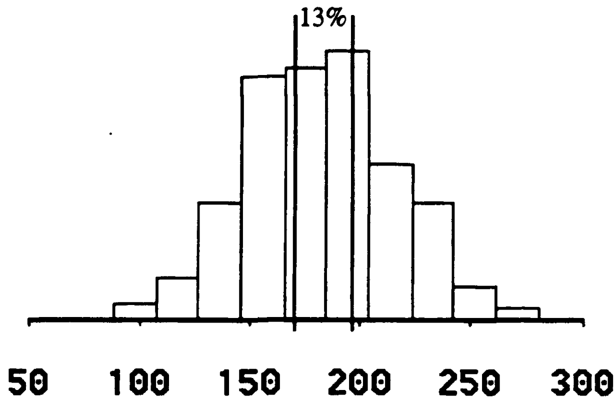
Beat 17



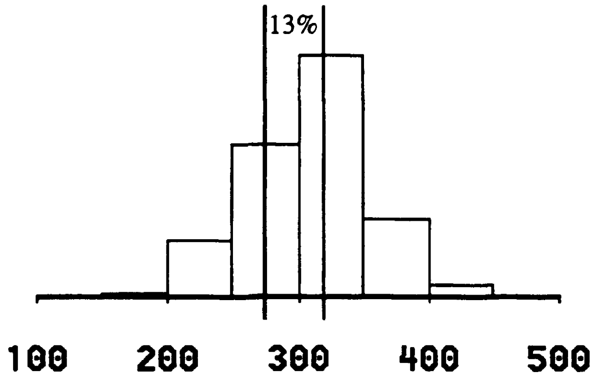
Beat 19



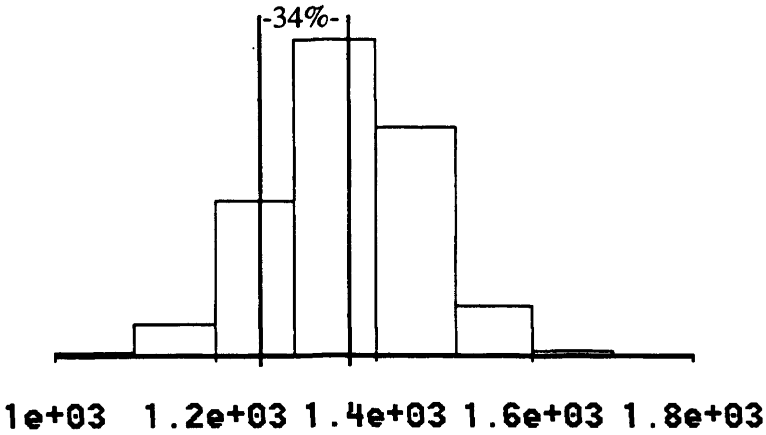
Beat 23



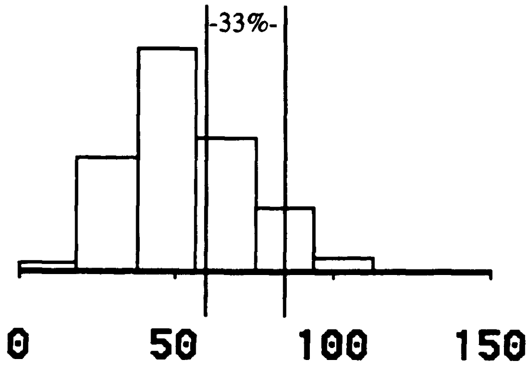
Beat 24



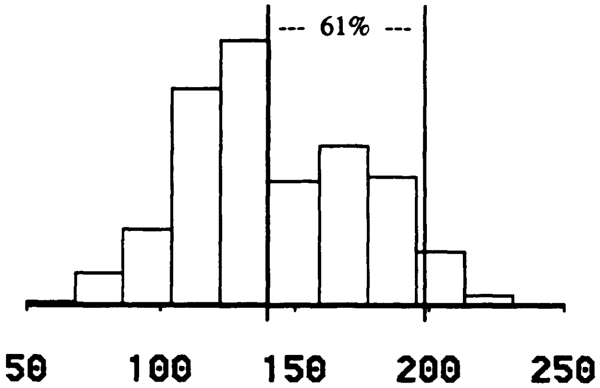
Beat 25



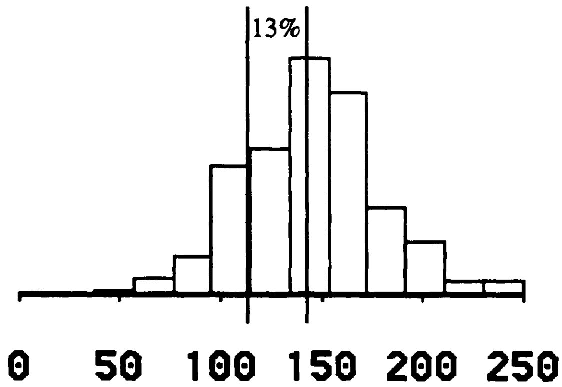
Beat 26



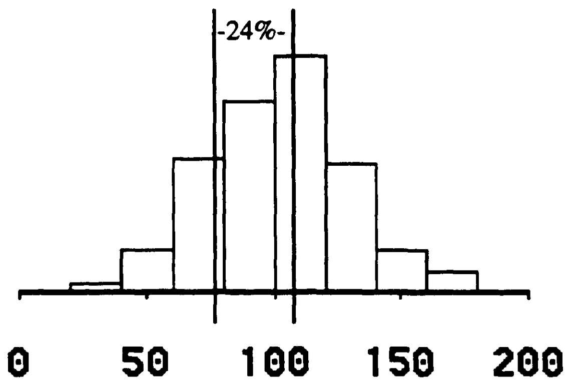
Beat 30



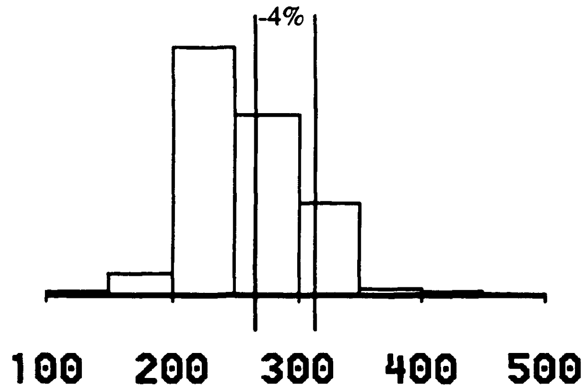
Beat 31



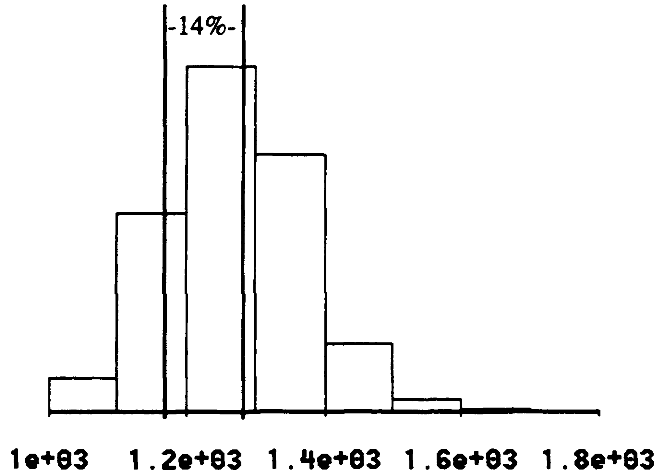
Beat 32



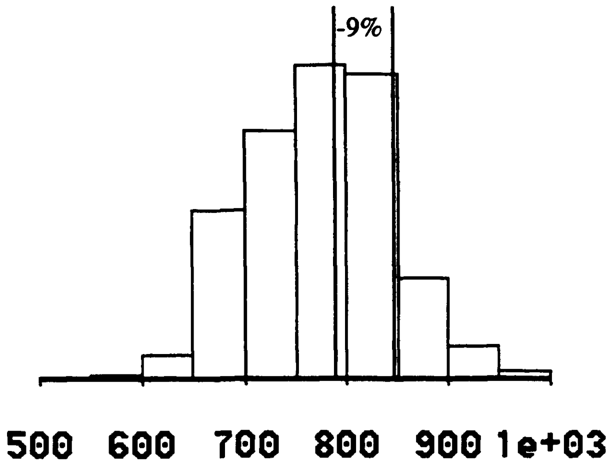
Beat 33



Beat 34

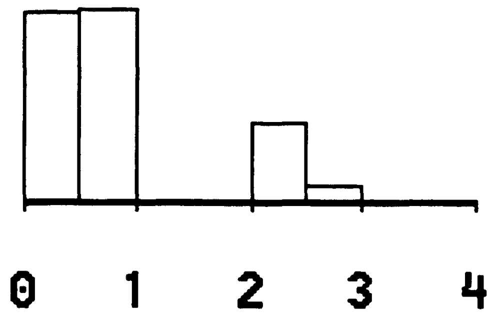


Beat 100

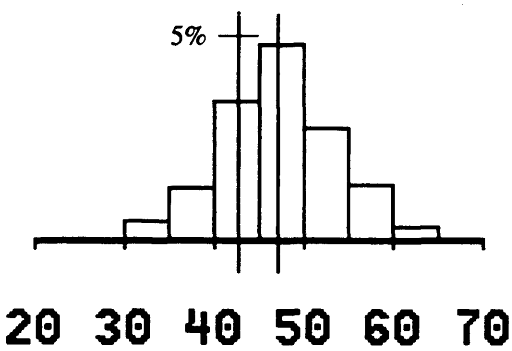


Beat 101

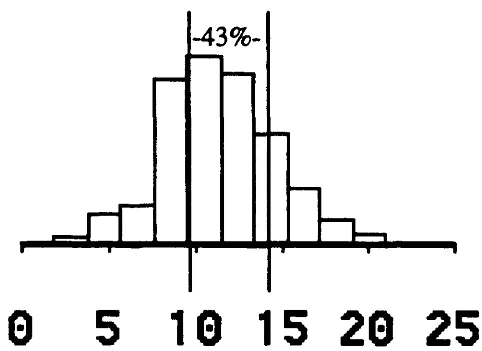
Bristol



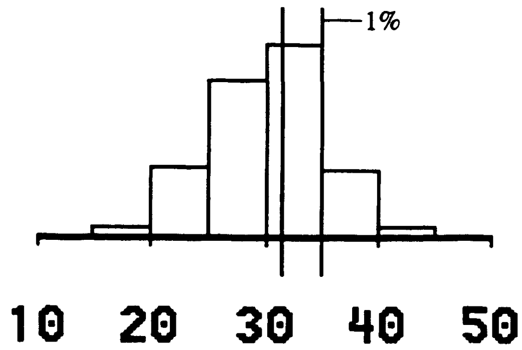
Beat 36



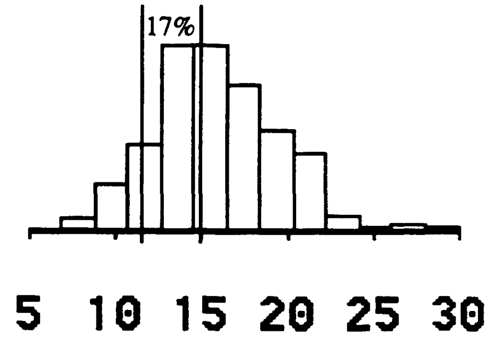
Beat 53



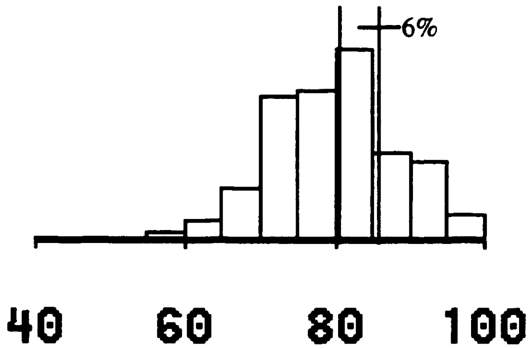
Beat 45



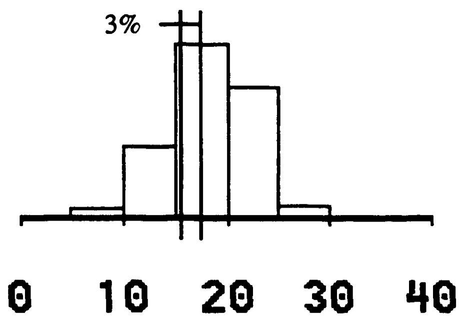
Beat 19



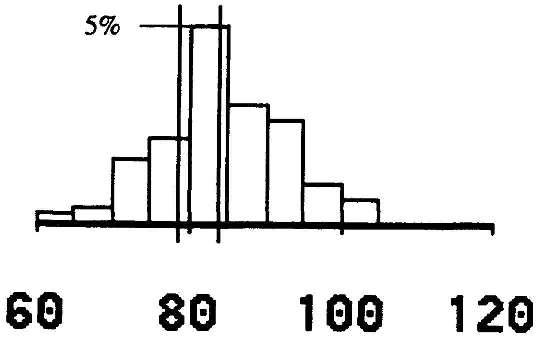
Beat 20



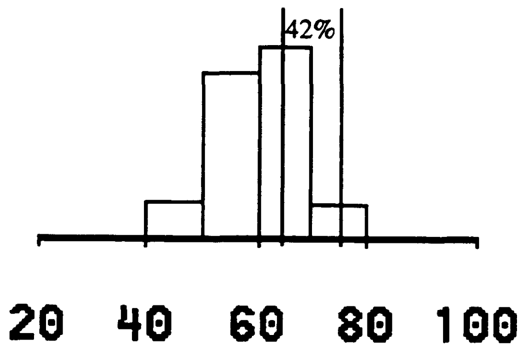
Beat 52



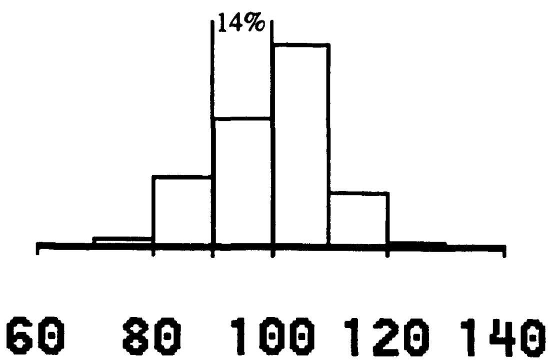
Beat 54



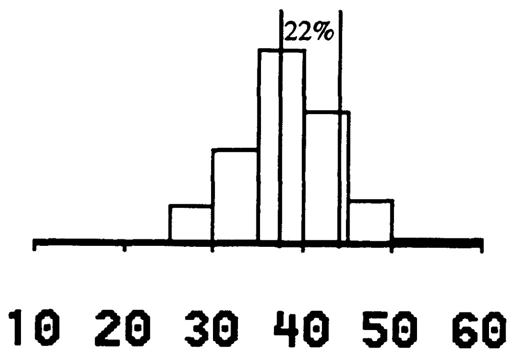
Beat 41



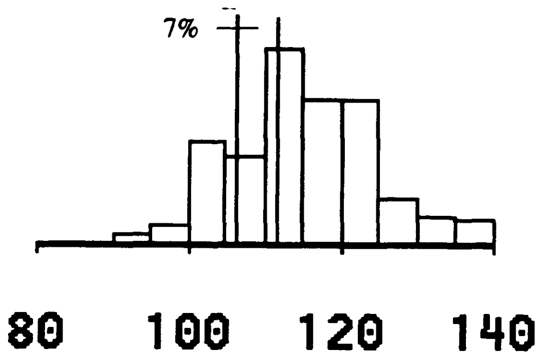
Beat 63



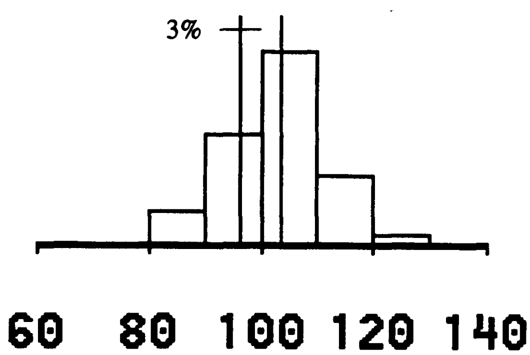
Beat 58



Beat 40



Beat 47



Beat 48

Appendix 10.1

Burglary risk in Bristol (overlay method)

Table A10.1 shows the burglary risk computed by the overlay method for Bristol. The burglary risk is defined as *the number of domestic burglaries per 100 households in each beat, in each year.*

Table A10.1: Burglary risk in Bristol (overlay method)

Beat-ID	super-id	1987	1988	1989	1980	1991	1992
32	206	0	0	0	0	0	0
33	207	0.024421	0.024421	0.048821	0.036621	0.018321	0.0061
44	208	0.058421	0.168821	0.097421	0.084421	0.162321	0.0325
36	209	0.024621	0.028521	0.042721	0.038821	0.051121	0.0252
53	210	0.073621	0.126521	0.069721	0.106821	0.128721	0.0433
9	211	0.030221	0.026521	0.024621	0.059721	0.0691	0.073
10	211	0.030221	0.026521	0.024621	0.059721	0.0691	0.073
15	212	0.035212	0.051212	0.041621	0.068212	0.108121	0.0926
16	212	0.035212	0.051212	0.041621	0.068212	0.108121	0.0926
24	212	0.035212	0.051212	0.041621	0.068212	0.108121	0.0926
34	213	0	0	0	0	0	0
35	214	0	0	0	0	0	0
31	215	0.019322	0.012422	0.012422	0.033215	0.042622	0.0261
37	216	0.013216	0.015222	0.024922	0.030322	0.026216	0.0498
21	217	0.045322	0.0778	0.058122	0.084222	0.111322	0.0991
45	218	0.025422	0.022722	0.028322	0.033322	0.050222	0.0701
20	219	0.100822	0.171219	0.124822	0.109222	0.135222	0.1131
19	220	0.143822	0.186922	0.134822	0.111222	0.223122	0.2168
54	221	0.039122	0.076122	0.057622	0.054122	0.079622	0.1292
25	222	0.0705	0.041722	0.025122	0.036222	0.051922	0.0544
26	222	0.0705	0.041722	0.025122	0.036222	0.051922	0.0544
27	222	0.0705	0.041722	0.025122	0.036222	0.051922	0.0544
30	222	0.0705	0.041722	0.025122	0.036222	0.051922	0.0544
52	223	0.050922	0.052222	0.034622	0.060122	0.090222	0.123
41	224	0.020922	0.016122	0.014422	0.047522	0.054322	0.0787
63	225	0.107225	0.130423	0.102323	0.124723	0.156323	0.1747
38	226	0.022223	0.022823	0.017523	0.025123	0.031623	0.0199
39	226	0.022223	0.022823	0.017523	0.025123	0.031623	0.0199
42	226	0.022223	0.022823	0.017523	0.025123	0.031623	0.0199
57	226	0.022223	0.022823	0.017523	0.025123	0.031623	0.0199
58	227	0.019227	0.022123	0.020823	0.020323	0.019227	0.0152
60	227	0.019227	0.022123	0.020823	0.020323	0.019227	0.0152
40	228	0.011823	0.010523	0.0079	0.011323	0.013123	0.0135
43	228	0.011823	0.010523	0.0079	0.011323	0.013123	0.0135
61	228	0.011823	0.010523	0.0079	0.011323	0.013123	0.0135
62	228	0.011823	0.010523	0.0079	0.011323	0.013123	0.0135
8	229	0	0.016229	0.019229	0.029623	0.0729	0.044923
14	229	0	0.016229	0.019229	0.029623	0.0729	0.044923

2	230	0.0184	0.025423	0.016523	0.022523	0.051723	0.053423
11	231	0.0273	0.037223	0.023923	0.037423	0.075223	0.039232
12	231	0.0273	0.037223	0.023923	0.037423	0.075223	0.039232
7	232	0.0057	0.022523	0.019223	0.042723	0.082523	0.057423
13	232	0.0228	0.036123	0.021823	0.036523	0.048523	0.014123
48	233	0.0228	0.036123	0.021823	0.036523	0.048523	0.014123
49	233	0.0228	0.036123	0.021823	0.036523	0.048523	0.014123
50	233	0.0228	0.036123	0.021823	0.036523	0.048523	0.014123
55	233	0.0228	0.036123	0.021823	0.036523	0.048523	0.014123
47	234	0.0215	0.032234	0.025523	0.032723	0.063923	0.012924
51	234	0.0215	0.032234	0.025523	0.032723	0.063923	0.012924
22	235	0.072	0.061224	0.030224	0.054524	0.08235	0.064824
23	235	0.072	0.061224	0.030224	0.054524	0.08235	0.064824
28	235	0.072	0.061224	0.030224	0.054524	0.08235	0.064824
29	235	0.072	0.061224	0.030224	0.054524	0.08235	0.064824
1	239	0.046	0.020724	0.026424	0.026424	0.050624	0.012624
5	240	0.007	0.015324	0.025824	0.013524	0.036724	0.0087
4	241	0.0462	0.040424	0.022124	0.018824	0.040424	0.007724
3	242	0	0.1024	0.074242	0.127624	0.146524	0.037824
46	244	0	0.019224	0.0305	0.021324	0.041724	0.010325
56	246	0.050525	0.059325	0.0245	0.044246	0.089125	0.032125
18	247	0	0.076247	0.067425	0.1031	0.172925	0.038825
17	248	0	0.079825	0.061225	0.085325	0.1744	0.030625
6	249	0	0.010225	0.006425	0.017249	0.033925	0.0045
Mean		**0.032957**	**0.04402**	**0.032433**	**0.044084**	**0.06668**	**0.045343**

Appendix 10.2

Burglary risk in Bristol (Monte Carlo dasymetric method)

Table A10.2 shows the burglary risk computed by the overlay method for Bristol.

Table A10.2: Burglary risk in Bristol (Monte Carlo dasymetric method)

Beat-ID	super-id	1987'	1988'	1989'	1980'	1991'	1992'
32	206	0	0	0	0	0	0
33	207	0.081279	0.081279	0.16249	0.121885	0.060977	0.020303
44	208	0.047309	0.136711	0.078891	0.068364	0.131447	0.026319
36	209	0.030771	0.035645	0.053393	0.048518	0.063891	0.031495
53	210	0.06966	0.119714	0.06597	0.101074	0.121796	0.04097
9	211	0.032554	0.028569	0.026522	0.064332	0.074435	0.078636
10	211	0.032554	0.028569	0.026522	0.064332	0.074435	0.078636
15	212	0.031514	0.045833	0.03725	0.061047	0.096765	0.082874
16	212	0.031514	0.045833	0.03725	0.061047	0.096765	0.082874
24	212	0.031514	0.045833	0.03725	0.061047	0.096765	0.082874
34	213	0	0	0	0	0	0
35	214	0	0	0	0	0	0
31	215	0.014103	0.009066	0.009066	0.024244	0.031109	0.01905
37	216	0.009184	0.010577	0.017318	0.02107	0.018217	0.034606
21	217	0.046731	0.08022	0.059929	0.086841	0.114784	0.102182
45	218	0.036338	0.032479	0.040483	0.04763	0.071787	0.100201
20	219	0.102233	0.173616	0.126569	0.110751	0.137115	0.114683
19	220	0.119516	0.155332	0.112037	0.092425	0.185414	0.180161
54	221	0.041509	0.080766	0.061137	0.057424	0.084479	0.137081
25	222	0.088426	0.052331	0.03151	0.045433	0.065125	0.068233
26	222	0.088426	0.052331	0.03151	0.045433	0.065125	0.068233
27	222	0.088426	0.052331	0.03151	0.045433	0.065125	0.068233
30	222	0.088426	0.052331	0.03151	0.045433	0.065125	0.068233
52	223	0.049359	0.050619	0.033559	0.058277	0.087452	0.119224
41	224	0.01979	0.01525	0.013642	0.044951	0.051384	0.074442
63	225	0.152839	0.185904	0.14585	0.177779	0.222822	0.249017
38	226	0.020502	0.021056	0.016166	0.023177	0.029174	0.018359
39	226	0.020502	0.021056	0.016166	0.023177	0.029174	0.018359
42	226	0.020502	0.021056	0.016166	0.023177	0.029174	0.018359
57	226	0.020502	0.021056	0.016166	0.023177	0.029174	0.018359
58	227	0.017539	0.02018	0.018994	0.018538	0.017539	0.013865
60	227	0.017539	0.02018	0.018994	0.018538	0.017539	0.013865
40	228	0.010834	0.02018	0.018994	0.018538	0.017539	0.013865
43	228	0.010834	0.02018	0.018994	0.018538	0.017539	0.013865
61	228	0.010834	0.02018	0.018994	0.018538	0.017539	0.013865
62	228	0.010834	0.02018	0.018994	0.018538	0.017539	0.013865
8	229	0	0.014953	0.017718	0.027295	0.06717	0.041392
14	229	0	0.014953	0.017718	0.027295	0.06717	0.041392
2	230	0.020003	0.083896	0.017962	0.024485	0.056228	0.058076
11	231	0.028118	0.038338	0.02464	0.038544	0.077476	0.040407
12	231	0.028118	0.038338	0.02464	0.038544	0.077476	0.040407
7	232	0.007896	0.031199	0.026628	0.05918	0.114311	0.079543
13	232	0.022069	0.034966	0.021124	0.035353	0.046969	0.013671

48	233	0.022069	0.034966	0.021124	0.035353	0.046969	0.013671
49	233	0.022069	0.034966	0.021124	0.035353	0.046969	0.013671
50	233	0.022069	0.034966	0.021124	0.035353	0.046969	0.013671
55	233	0.022069	0.034966	0.021124	0.035353	0.046969	0.013671
47	234	0.019223	0.034966	0.021124	0.035353	0.046969	0.013671
51	234	0.019223	0.034966	0.021124	0.035353	0.046969	0.013671
22	235	0.07739	0.065807	0.032486	0.058605	0.088515	0.069677
23	235	0.07739	0.065807	0.032486	0.058605	0.088515	0.069677
28	235	0.07739	0.065807	0.032486	0.058605	0.088515	0.069677
29	235	0.07739	0.065807	0.032486	0.058605	0.088515	0.069677
1	239	0.071038	0.032004	0.040806	0.040806	0.078178	0.019495
5	240	0.006815	0.014918	0.02514	0.013166	0.035751	0.008469
4	241	0.034973	0.030601	0.016748	0.01425	0.030601	0.005847
3	242	0	0.140687	0.102001	0.175343	0.20131	0.051967
46	244	0	0.01632	0.025891	0.018102	0.03542	0.008765
56	246	0.050393	0.05917	0.024436	0.044131	0.088893	0.032041
18	247	0	0.08593	0.075988	0.116194	0.194886	0.043756
17	248	0	0.074237	0.056939	0.079352	0.162192	0.028481
6	249	0	0.01163	0.007308	0.019619	0.038586	0.005118
Mean		**0.035486**	**0.047929**	**0.036003**	**0.048015**	**0.069561**	**0.047819**

Appendix 10.3

Burglary risk in Coventry (overlay method)

Table A10.1 shows the burglary risk computed by the overlay method for Coventry. The burglary risk is defined as *the number of domestic burglaries per 100 households in each beat, in each year.*

Table A10.3: Burglary risk in Coventry (overlay method)

Beat-ID	1987	1988	1989	1980	1991	1992	1993
2	0.021529	0.040629	0.018229	0.028929	0.055529	0.063429	0.0168
3	0.014728	0.010528	0.010128	0.019284	0.025428	0.026228	0.0068
4	0.011128	0.031828	0.027228	0.023628	0.034128	0.031128	0.0052
5	0.010928	0.018628	0.012928	0.017128	0.028528	0.025328	0.0041
6	0.033328	0.047628	0.026328	0.033128	0.028128	0.035628	0.0166
7	0.035128	0.060828	0.055228	0.044628	0.031228	0.040128	0.0262
8	0.026428	0.038528	0.052278	0.033928	0.038128	0.041278	0.0227
9	0.011628	0.038228	0.041928	0.042628	0.034528	0.056328	0.0132
10	0.022727	0.042272	0.045827	0.056327	0.038627	0.044427	0.0227
11	0.040827	0.064827	0.047327	0.054227	0.061527	0.051927	0.024
14	0.045927	0.102627	0.061527	0.063527	0.063527	0.063127	0.0176
15	0.053728	0.071328	0.078428	0.070728	0.064128	0.104328	0.0352
16	0.034228	0.042828	0.038228	0.047328	0.042228	0.073628	0.0434
17	0.056927	0.091527	0.060127	0.069227	0.066527	0.0995	0.0279
18	0.045727	0.062227	0.042227	0.059727	0.07427	0.064627	0.025
19	0.0632	0.064126	0.117626	0.102256	0.101126	0.09901	0.0253
20	0.015629	0.023287	0.015129	0.021729	0.032529	0.035929	0.01
21	0.042429	0.046129	0.055929	0.059629	0.055929	0.073829	0.0159
22	0.044229	0.052329	0.043829	0.048429	0.037829	0.082829	0.0174
23	0.038127	0.047527	0.047527	0.028127	0.043927	0.075627	0.0217
24	0.043425	0.06765	0.076254	0.054325	0.0627	0.144825	0.0338
25	0.038426	0.027426	0.039526	0.032926	0.034255	0.075726	0.0165
26	0.025429	0.029329	0.02129	0.023729	0.02829	0.029329	0.0066
27	0.017429	0.019229	0.0096	0.028529	0.039229	0.042129	0.0067
28	0.017929	0.016329	0.010829	0.021629	0.026829	0.031129	0.0049
29	0.153825	0.174825	0.0909	0.188825	0.216825	0.195825	0.1119
30	0.017425	0.024925	0.054725	0.034825	0.044825	0.049825	0.0149
31	0.050326	0.082426	0.088	0.086426	0.090526	0.100226	0.0178
32	0.047226	0.105258	0.122126	0.093626	0.100926	0.129426	0.0399
33	0.157626	0.174826	0.143326	0.108926	0.088826	0.151926	0.0516
34	0.063126	0.065426	0.049926	0.055126	0.068326	0.065426	0.0253
35	0.041126	0.053261	0.042126	0.037626	0.065326	0.070326	0.0158
36	0.047626	0.045126	0.035526	0.036726	0.057226	0.081826	0.0246
37	0.050826	0.079326	0.076226	0.137263	0.137263	0.127826	0.0408
38	0.057526	0.063626	0.091526	0.132526	0.150826	0.246726	0.0898
39	0.059627	0.061327	0.059627	0.072827	0.0778	0.112265	0.0386
40	0.018927	0.022727	0.027527	0.042627	0.049727	0.049727	0.027
41	0.011929	0.009729	0.014529	0.035829	0.049529	0.060929	0.0182
42	0.025229	0.038729	0.024929	0.045429	0.044729	0.061429	0.0332

	43	0.10733	0.11433	0.08423	0.11013	0.15853	0.17043	0.0645
	44	0.03423	0.04296	0.03423	0.065296	0.0753	0.09893	0.051
	45	0.03297	0.03323	0.02683	0.038297	0.05033	0.04453	0.023
	46	0.038528	0.045528	0.032128	0.051228	0.050928	0.079276	0.0368
	47	0.007629	0.010729	0.013129	0.0099	0.016529	0.036829	0.0078
	100	0.044227	0.047727	0.055427	0.048927	0.040327	0.083327	0.0321
	101	0.137427	0.160827	0.145627	0.166227	0.189273	0.215327	0.0516
Mean		**0.043823**	**0.056841**	**0.051481**	**0.058354**	**0.06463**	**0.081386**	**0.027878**

Appendix 10.4

Burglary risk in Coventry (Monte Carlo dasymetric method)

Table A10.1 shows the burglary risk computed by the Monte Carlo dasymetric method for Coventry.

Table A10.4: Burglary risk in Coventry (Monte Carlo dasymetric method)

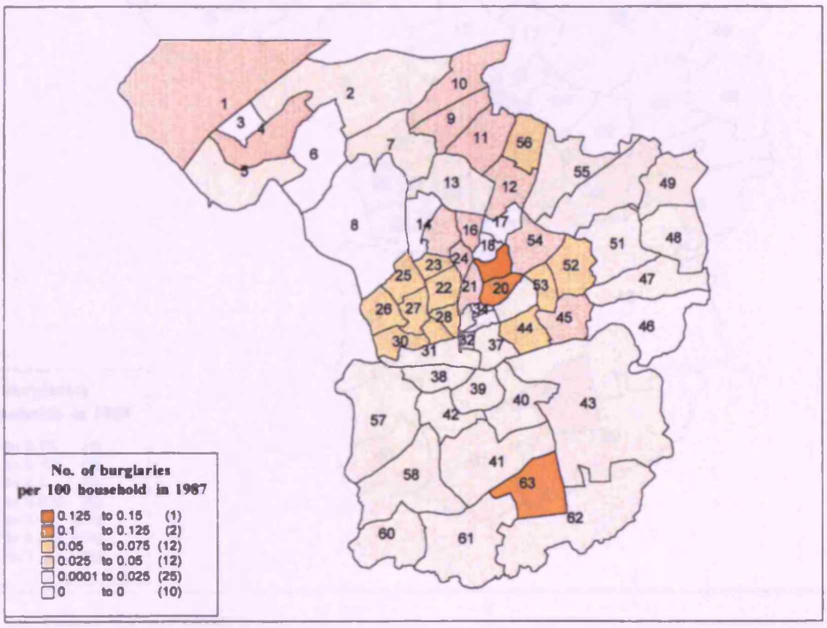
Beat-ID	1987'	1988'	1989'	1980'	1991'	1992'	1993'
2	0.012347	0.023301	0.010454	0.016591	0.031846	0.036377	0.009635
3	0.015488	0.011072	0.010651	0.020279	0.026741	0.027582	0.007151
4	0.012663	0.036217	0.030983	0.026887	0.038834	0.035421	0.005917
5	0.011941	0.020355	0.014127	0.018716	0.031173	0.027676	0.00448
6	0.047639	0.068079	0.037633	0.047353	0.040206	0.050927	0.023728
7	0.027614	0.047817	0.043415	0.035082	0.024548	0.031545	0.020596
8	0.033508	0.048849	0.066283	0.043017	0.048342	0.052336	0.028781
9	0.013091	0.043037	0.047203	0.047991	0.038872	0.063414	0.014861
10	0.023677	0.044039	0.047743	0.058682	0.040242	0.046284	0.023649
11	0.045494	0.072237	0.052737	0.060425	0.06856	0.057862	0.026743
14	0.051732	0.115599	0.069304	0.071557	0.071557	0.071106	0.019825
15	0.054098	0.07182	0.078969	0.071216	0.06457	0.105047	0.035443
16	0.0315	0.039414	0.035181	0.043556	0.038862	0.06776	0.039941
17	0.06501	0.104524	0.068665	0.079057	0.075974	0.113629	0.031862
18	0.047707	0.064921	0.044055	0.062313	0.077486	0.067425	0.026083
19	0.07735	0.078483	0.143962	0.125151	0.123768	0.121178	0.030965
20	0.016304	0.024293	0.015782	0.022667	0.033934	0.037481	0.010432
21	0.082999	0.090237	0.109408	0.116646	0.109408	0.144424	0.031104
22	0.075118	0.088875	0.074439	0.082252	0.064249	0.140677	0.029552
23	0.071354	0.088946	0.088946	0.052639	0.082209	0.141535	0.040612
24	0.03791	0.059058	0.06657	0.047426	0.054737	0.126433	0.029507
25	0.032739	0.023367	0.033676	0.028053	0.029185	0.064518	0.014058
26	0.016765	0.019337	0.014036	0.015645	0.018652	0.019337	0.004351
27	0.016714	0.01844	0.009206	0.027359	0.03762	0.040401	0.006425
28	0.02285	0.020811	0.013801	0.027566	0.034193	0.039673	0.006245
29	0.085527	0.097203	0.05054	0.104987	0.120555	0.108879	0.062216
30	0.023226	0.033223	0.072943	0.046419	0.059748	0.066412	0.01986
31	0.081266	0.133101	0.142102	0.13956	0.146181	0.161844	0.028743
32	0.040855	0.091059	0.105651	0.080996	0.087311	0.111966	0.034517
33	0.120269	0.133392	0.109358	0.08311	0.067774	0.115919	0.039371
34	0.065784	0.06818	0.052028	0.057447	0.071203	0.06818	0.026365
35	0.058218	0.075396	0.059634	0.053264	0.092476	0.099554	0.022366
36	0.043268	0.040997	0.032276	0.033366	0.05199	0.074339	0.022349
37	0.043146	0.06734	0.064709	0.116523	0.116523	0.108512	0.034635
38	0.047154	0.052155	0.075024	0.108632	0.123632	0.202242	0.073609
39	0.065512	0.067379	0.065512	0.080014	0.085479	0.123346	0.04241
40	0.020053	0.024079	0.029164	0.045163	0.052685	0.052685	0.028607
41	0.014998	0.012232	0.018266	0.045045	0.062268	0.0766	0.022881
42	0.029397	0.045127	0.029048	0.052934	0.052119	0.071578	0.038685
43	0.115379	0.122904	0.090547	0.118389	0.170419	0.183212	0.069338
44	0.038953	0.048888	0.038953	0.074307	0.085691	0.112582	0.058038

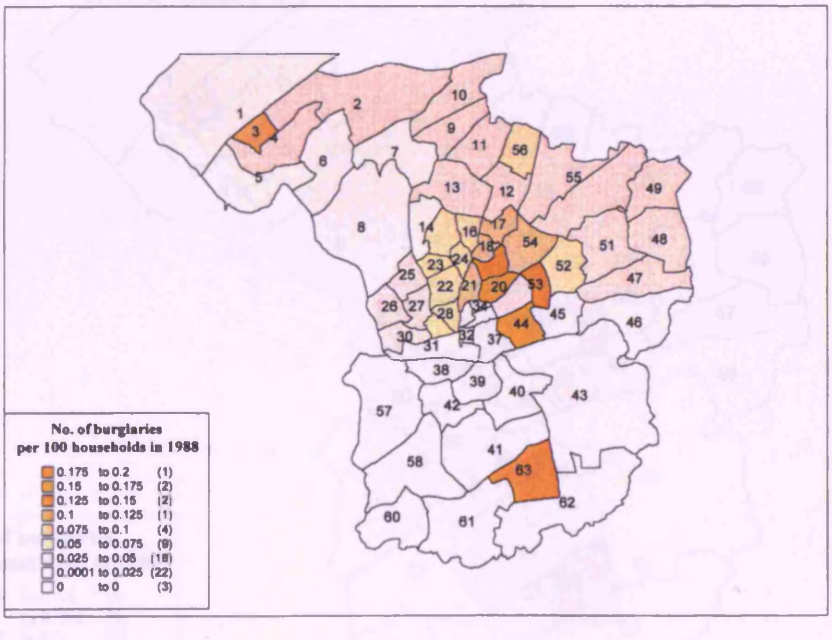
	45	0.02809	0.028312	0.022859	0.032629	0.042881	0.037939	0.019596
	46	0.039764	0.046989	0.033159	0.052872	0.052562	0.081821	0.037981
	47	0.008378	0.011782	0.014417	0.010871	0.018151	0.040442	0.008565
	100	0.038036	0.041046	0.047668	0.042078	0.034682	0.071662	0.027606
	101	0.150222	0.1758	0.159185	0.181703	0.206894	0.235374	0.056404
Mean		**0.045676**	**0.059559**	**0.055223**	**0.061053**	**0.067543**	**0.085546**	**0.028176**

Appendix 10.5

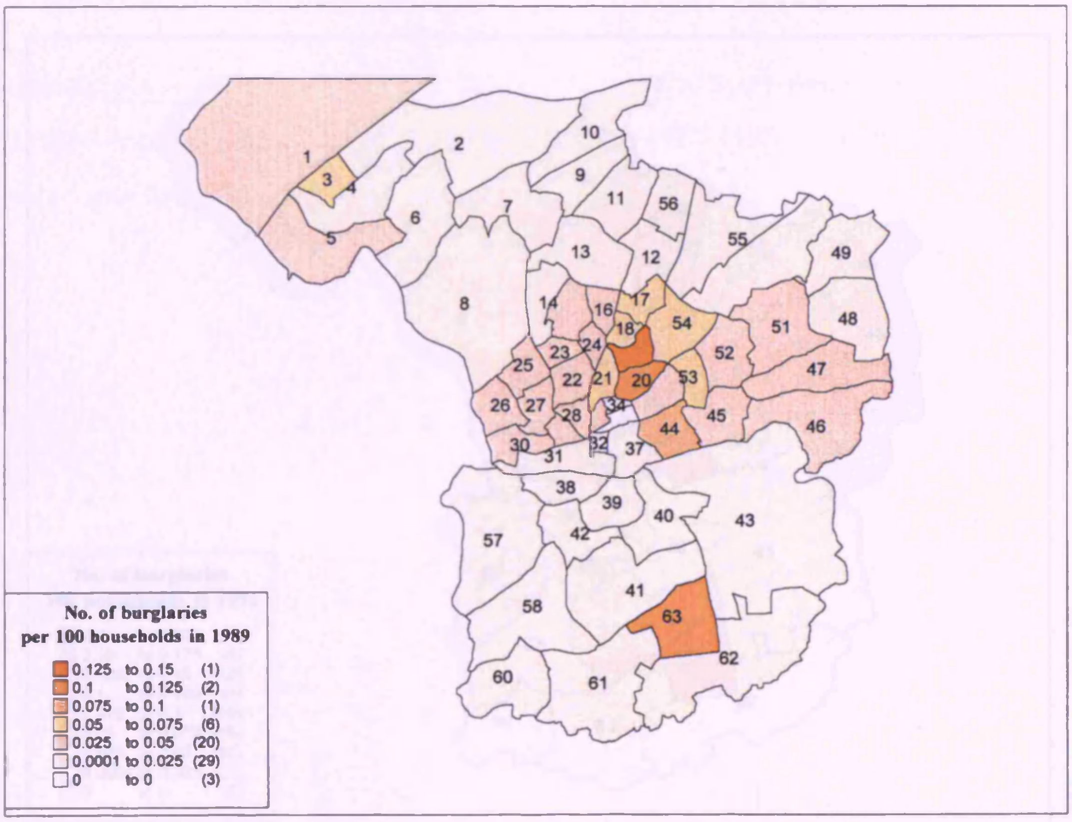
Thematic maps of Burglary Risk in Bristol (Overlay method)

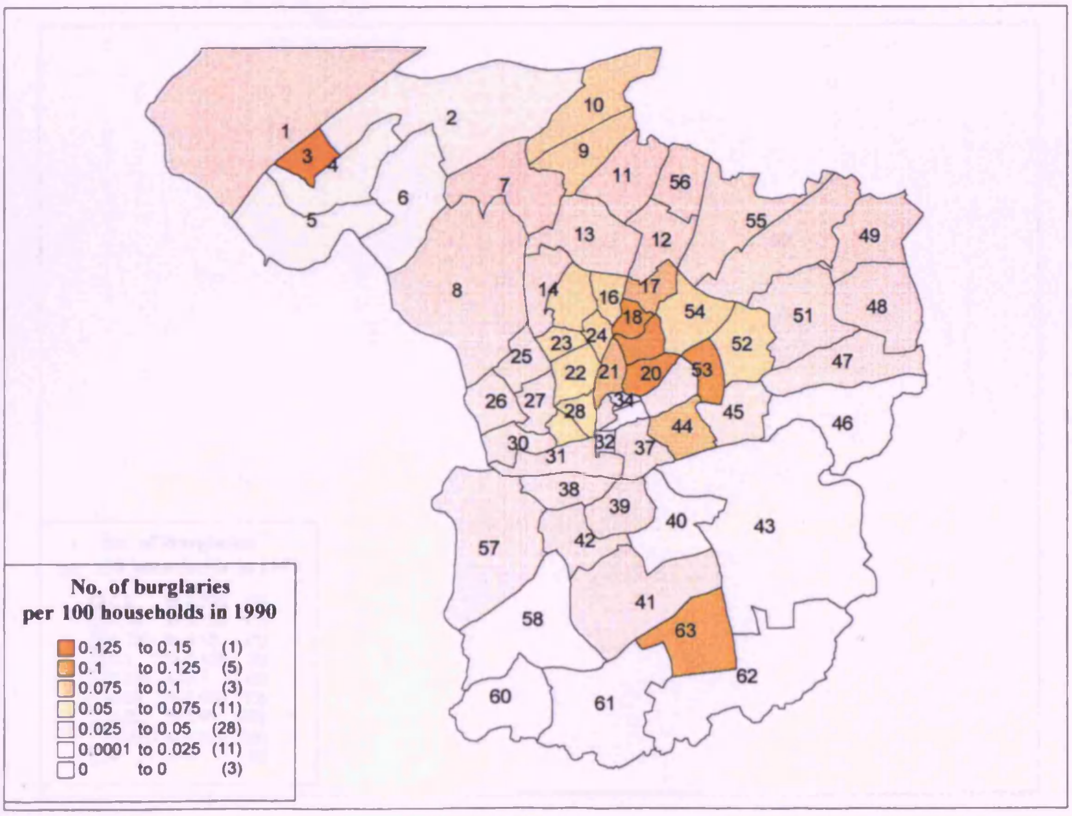
The following beat maps show the burglary risk from 1987-1992 for Bristol using the overlay method.



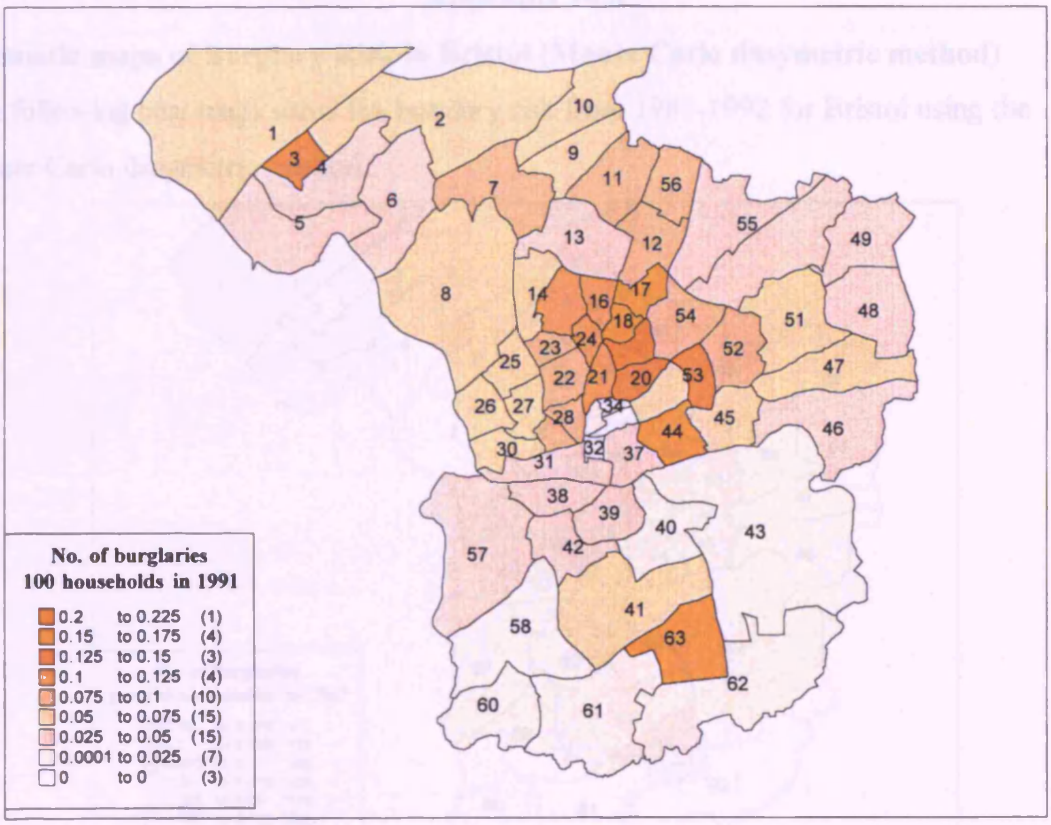


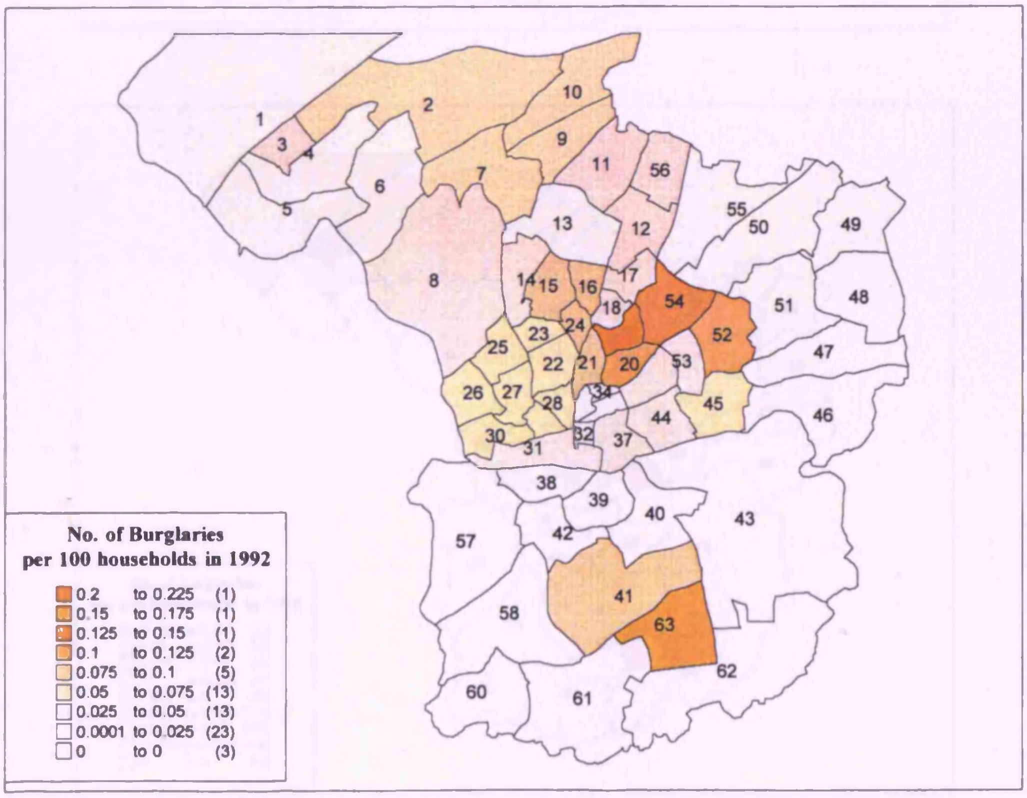
Appendix 10.5 Thematic maps of Burglary Risk in Bristol (Overlay method)





Appendix 10.5 Thematic maps of Burglary Risk in Bristol (Overlay method)

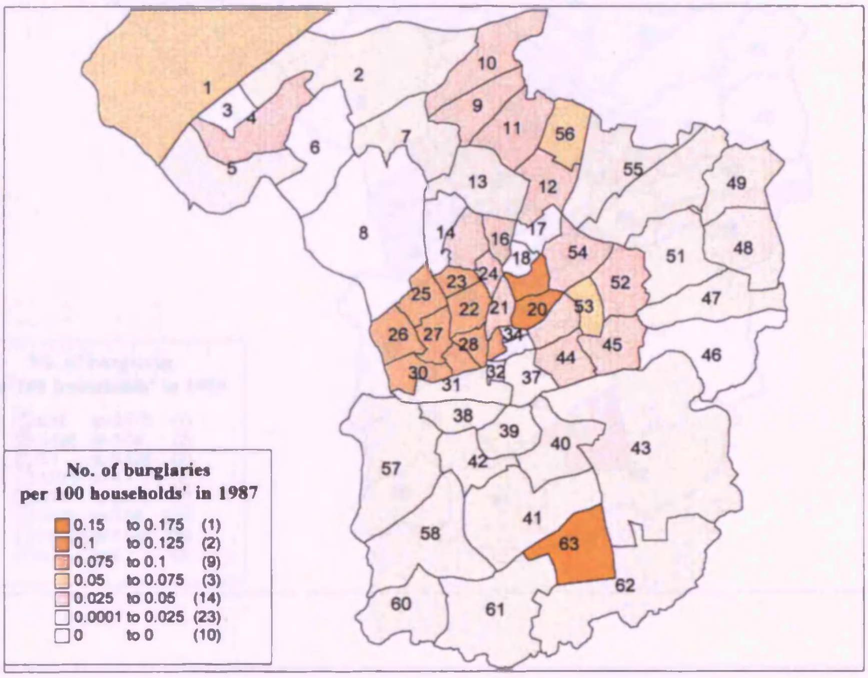


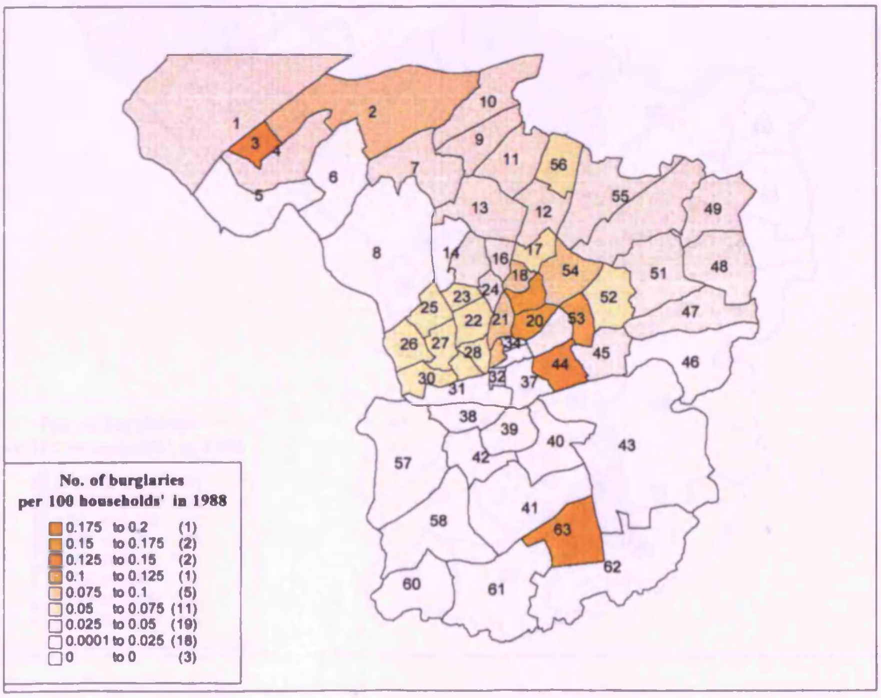


Appendix 10.6

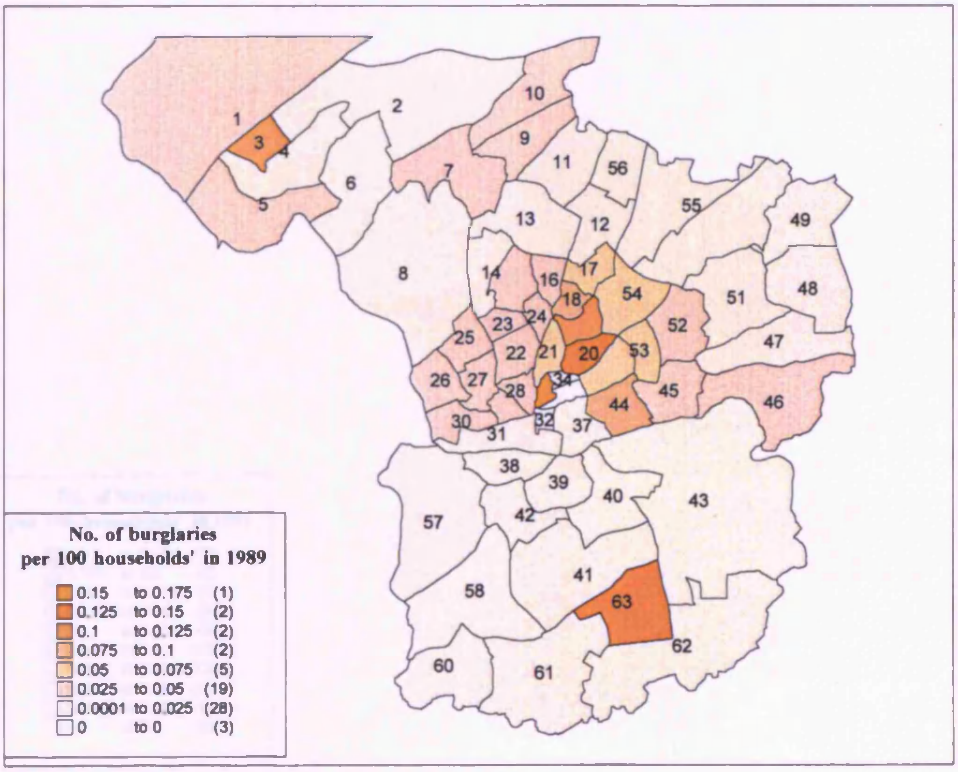
Thematic maps of Burglary Risk in Bristol (Monte Carlo dasymetric method)

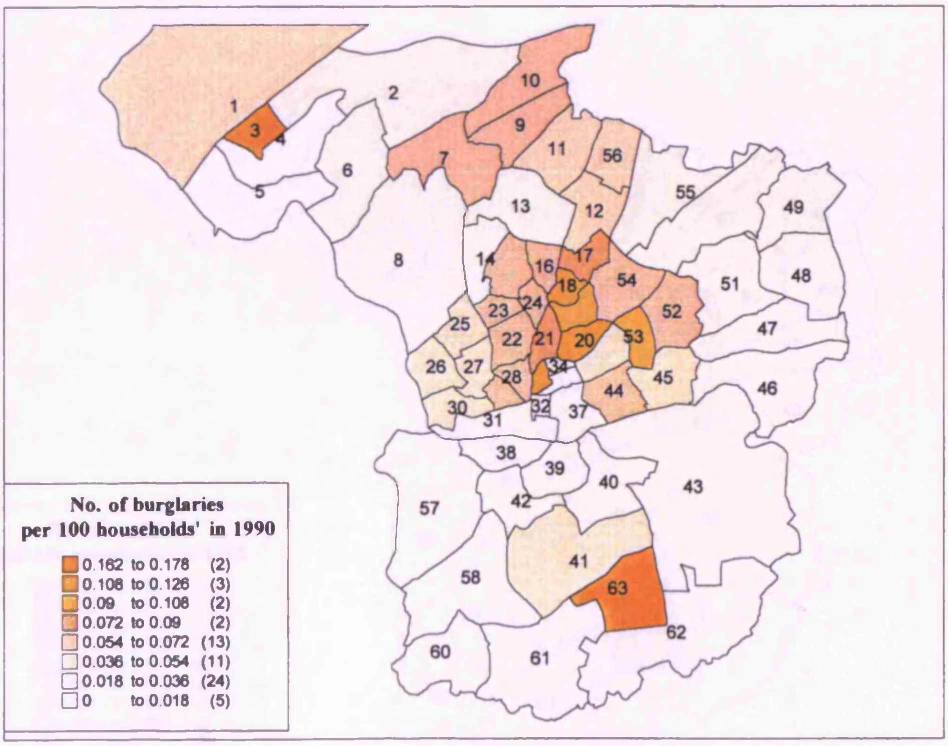
The following beat maps show the burglary risk from 1987-1992 for Bristol using the Monte Carlo dasymetric method.



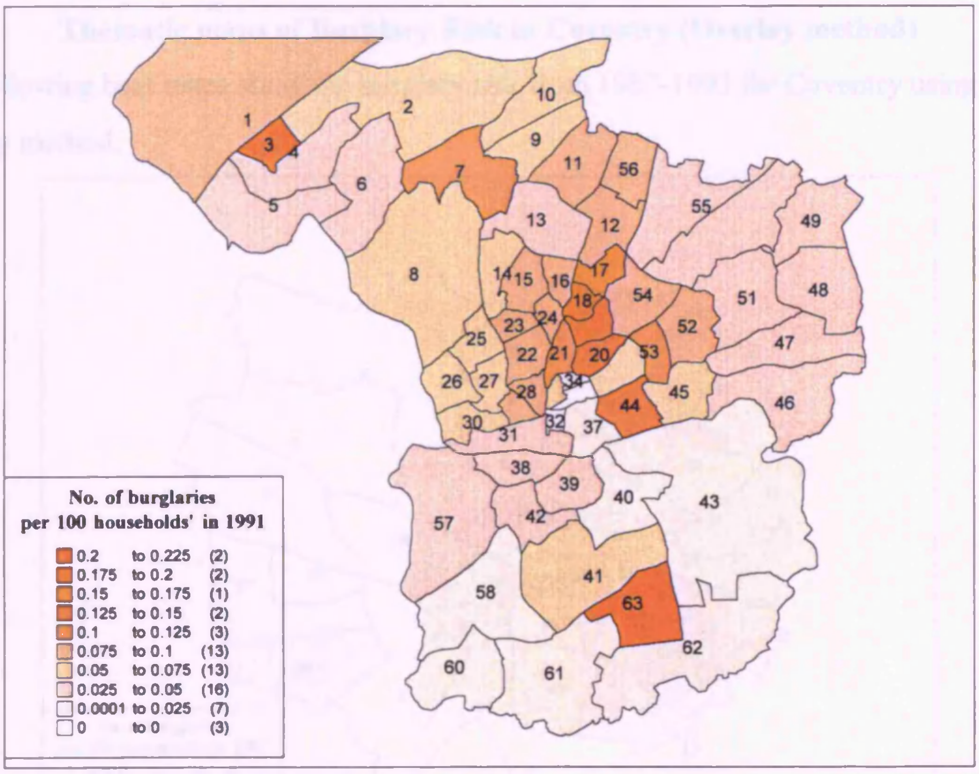


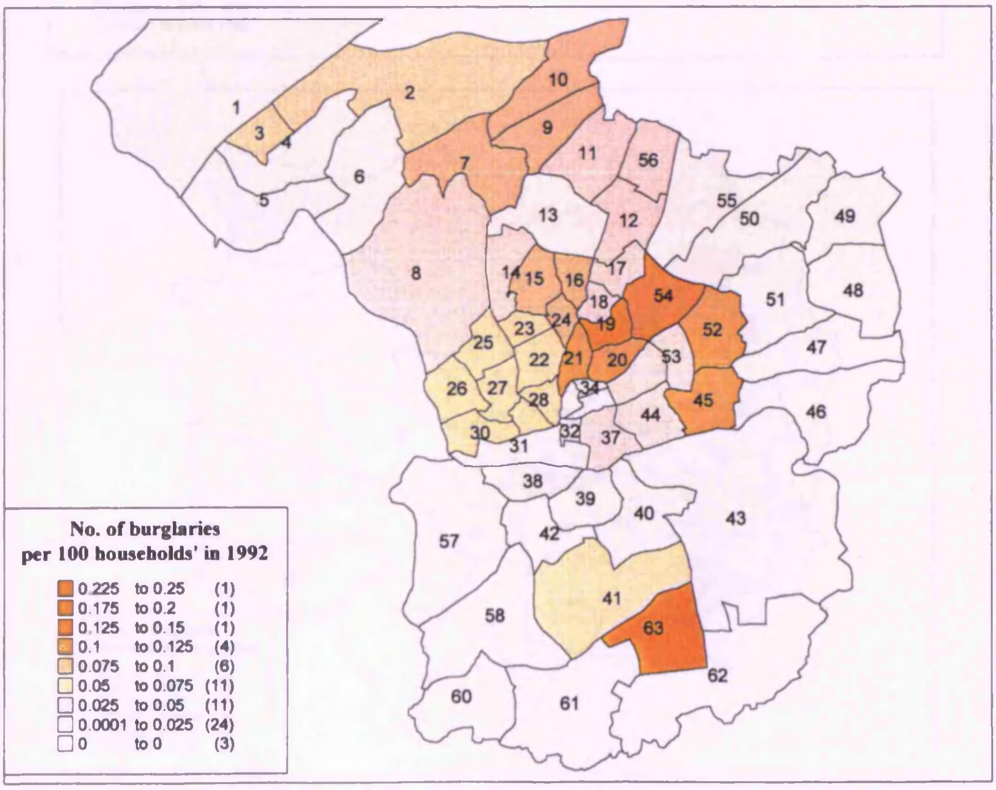
Appendix 10.6 Thematic maps of Burglary Risk in Bristol (Monte Carlo dasymetric method)





Appendix 10.6 Thematic maps of Burglary Risk in Bristol (Monte Carlo dasymetric method)

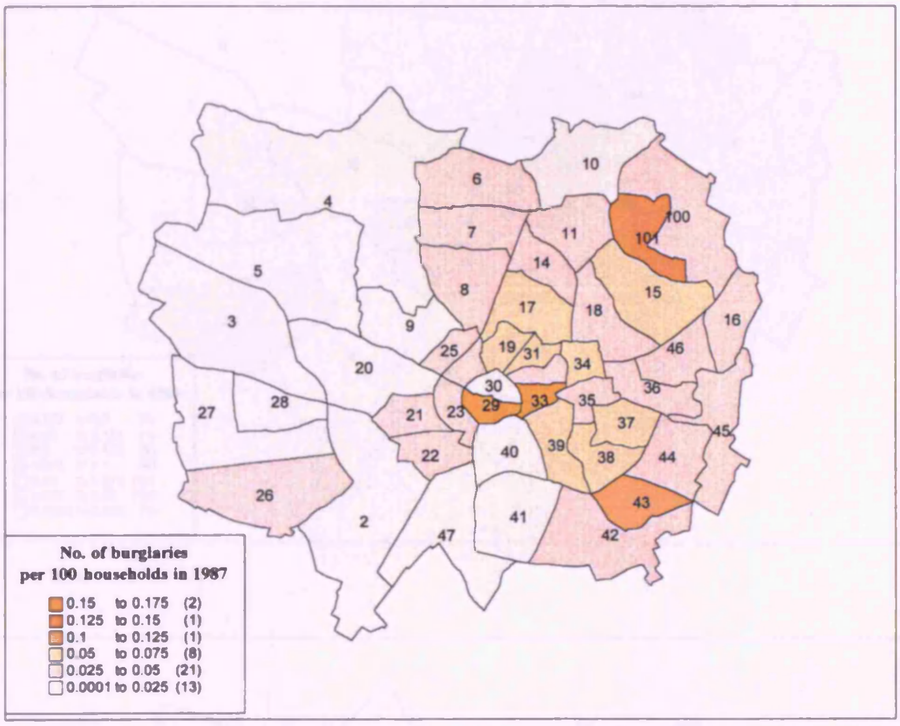


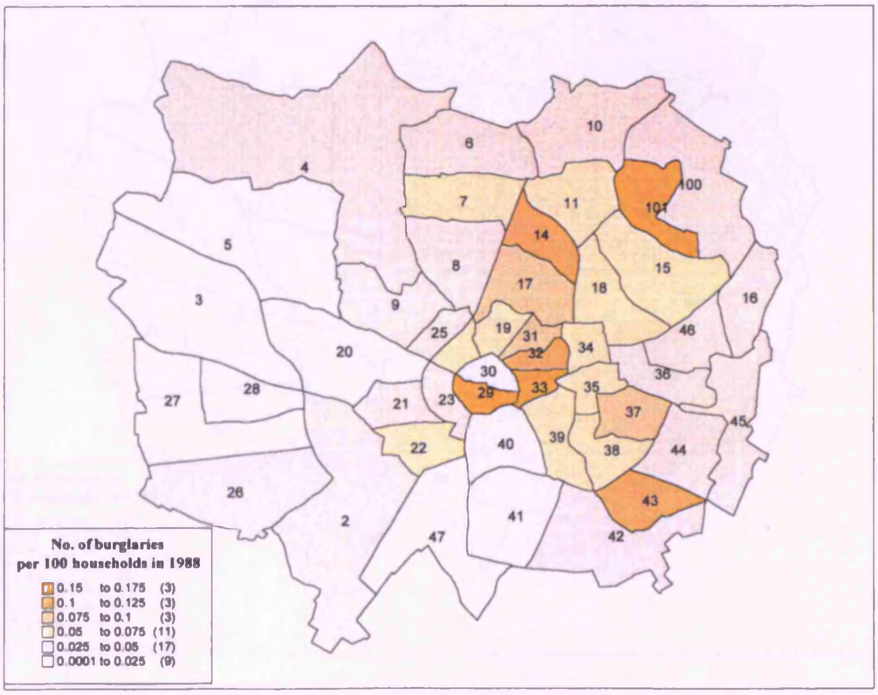


Appendix 10.7

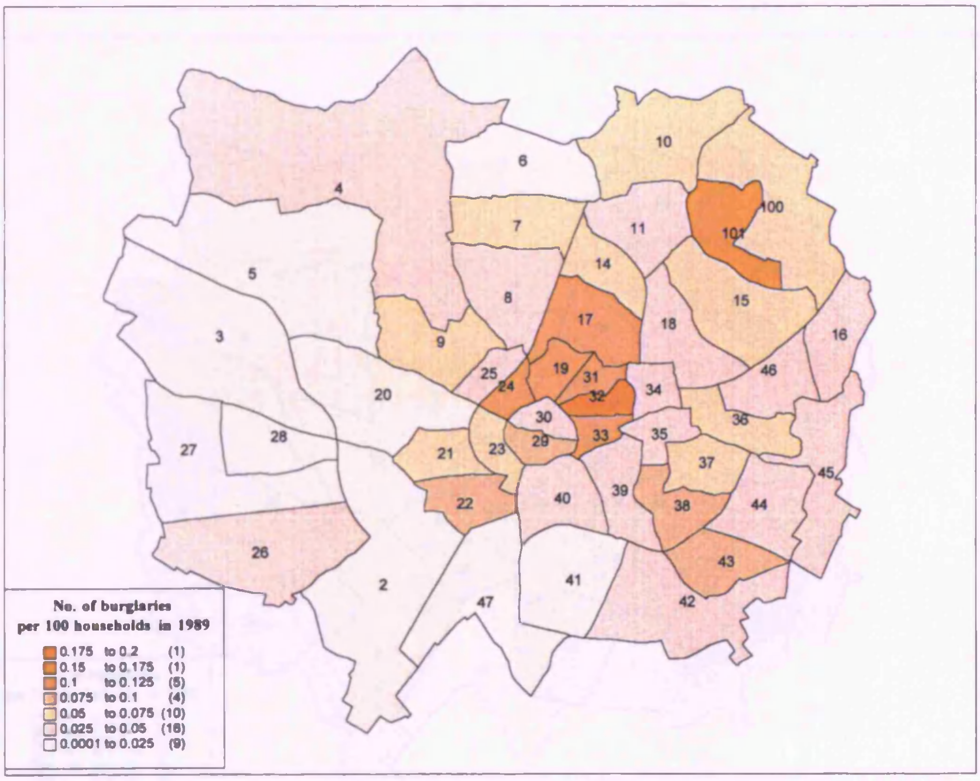
Thematic maps of Burglary Risk in Coventry (Overlay method)

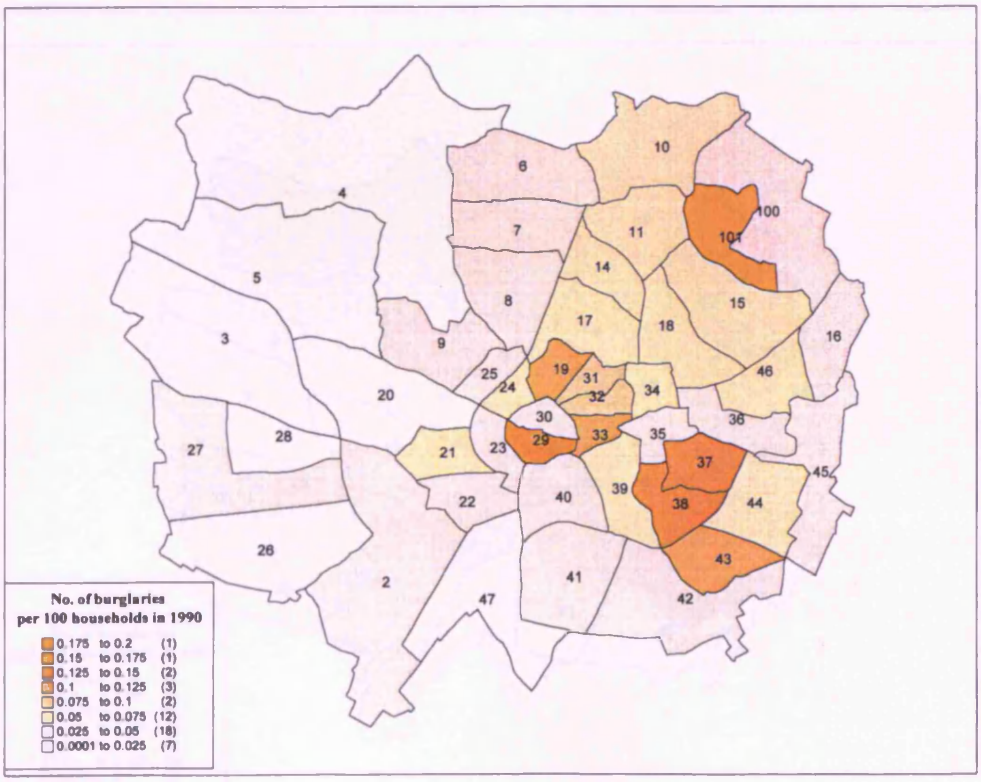
The following beat maps show the burglary risk from 1987-1993 for Coventry using the overlay method.



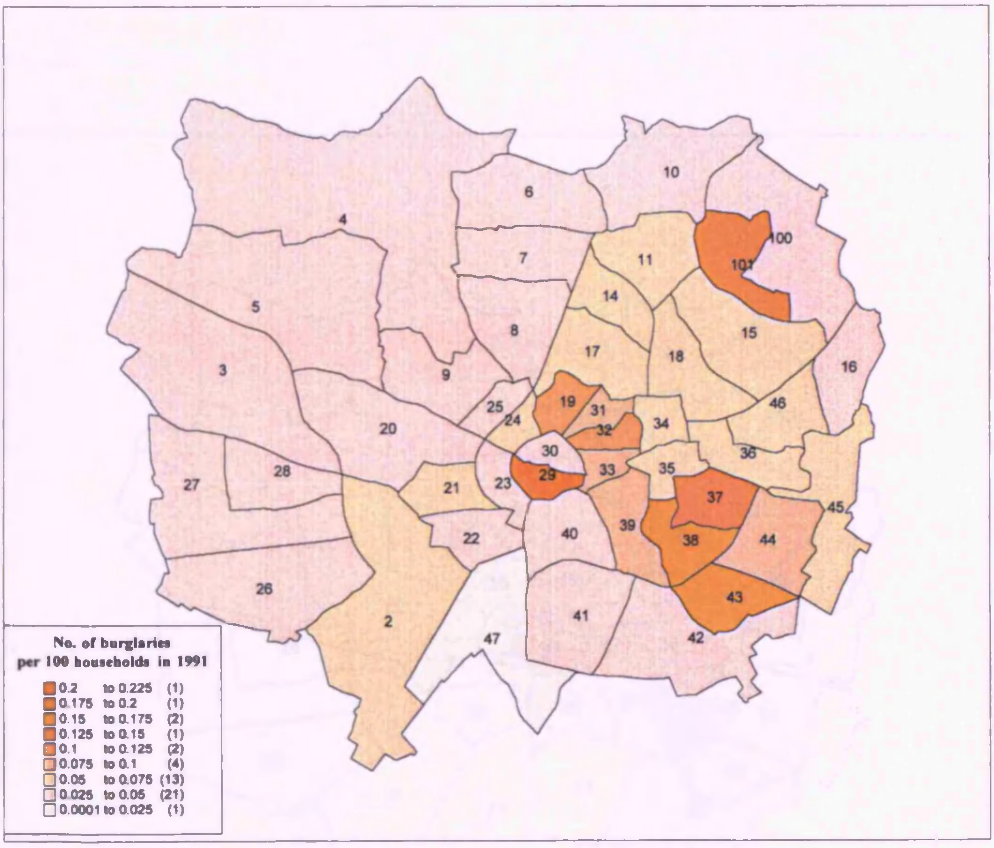


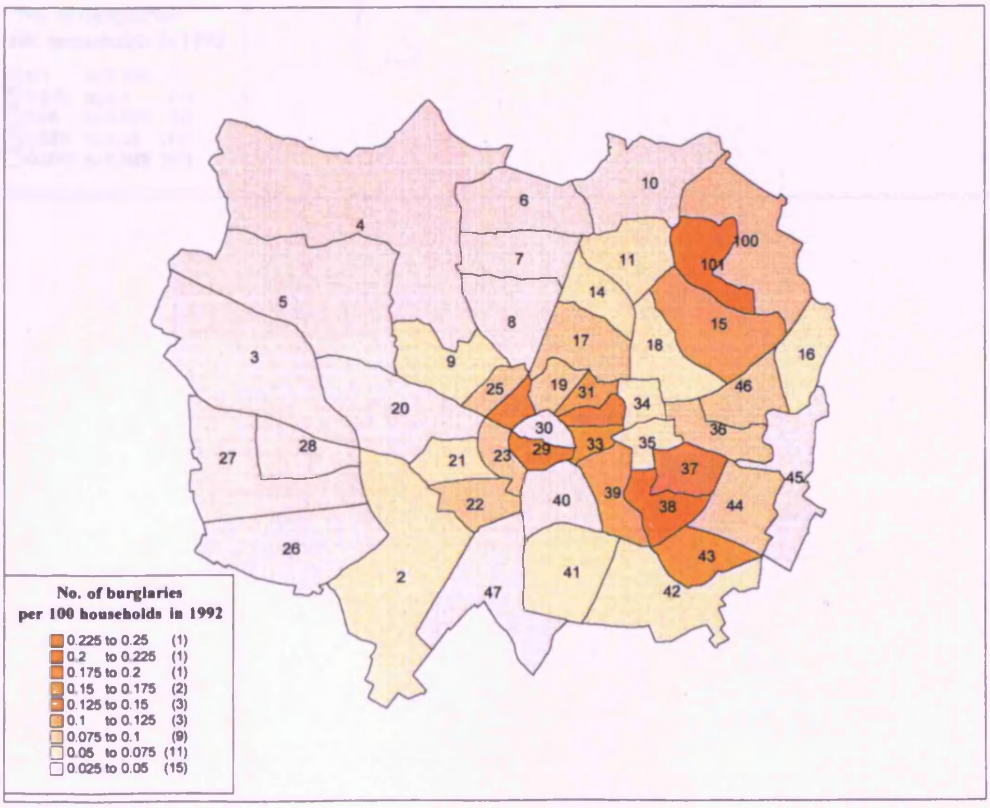
Appendix 10.7 Thematic maps of Burglary risk in Coventry (Overlay method)

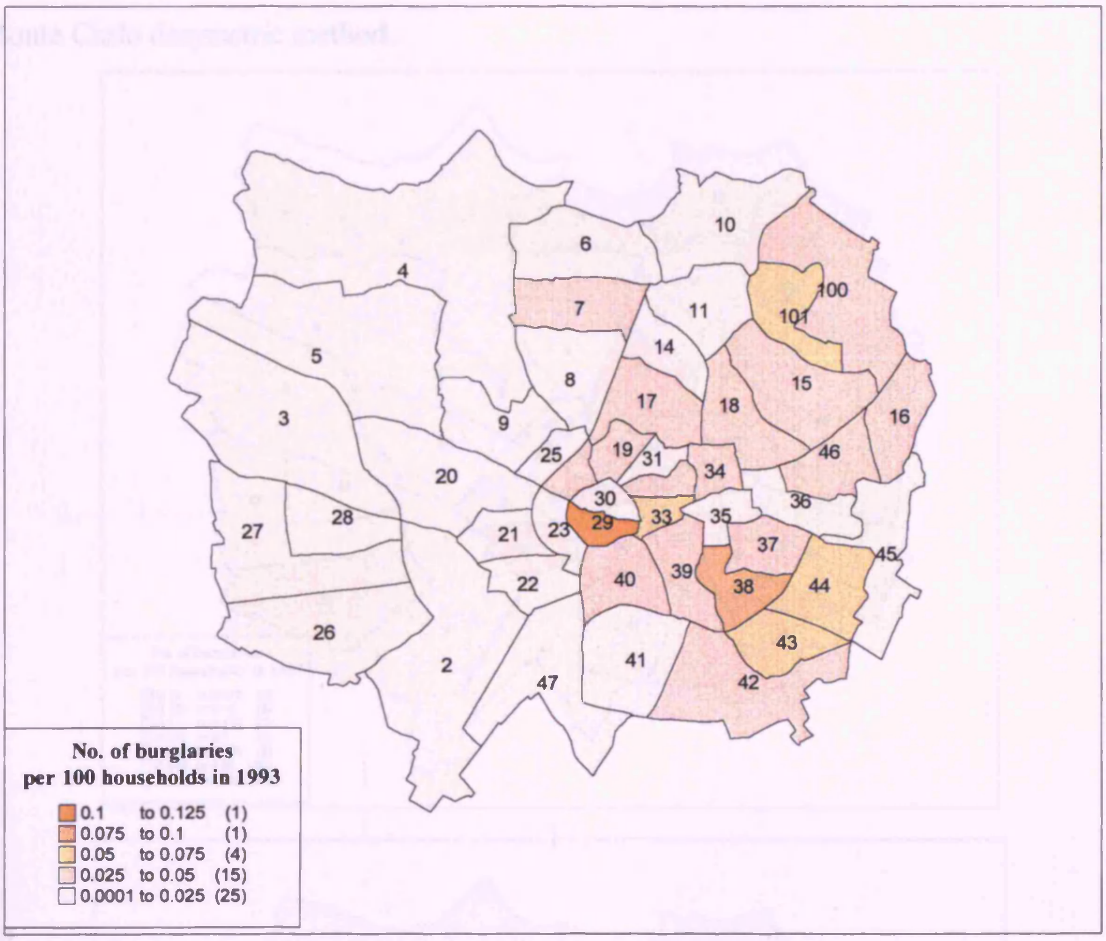




Appendix 10.7 Thematic maps of Burglary risk in Coventry (Overlay method)





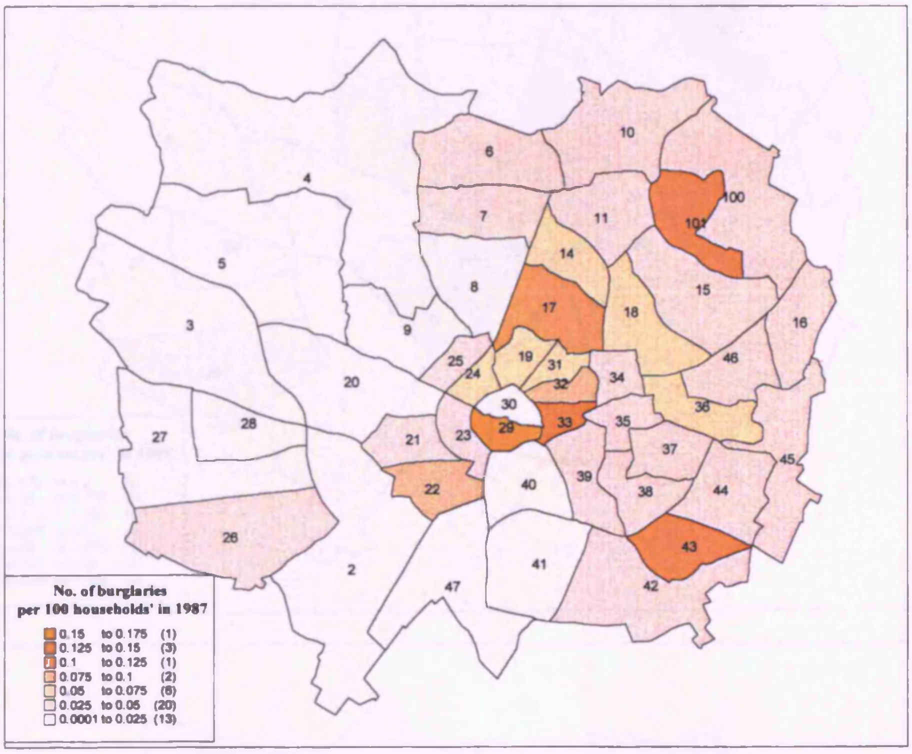


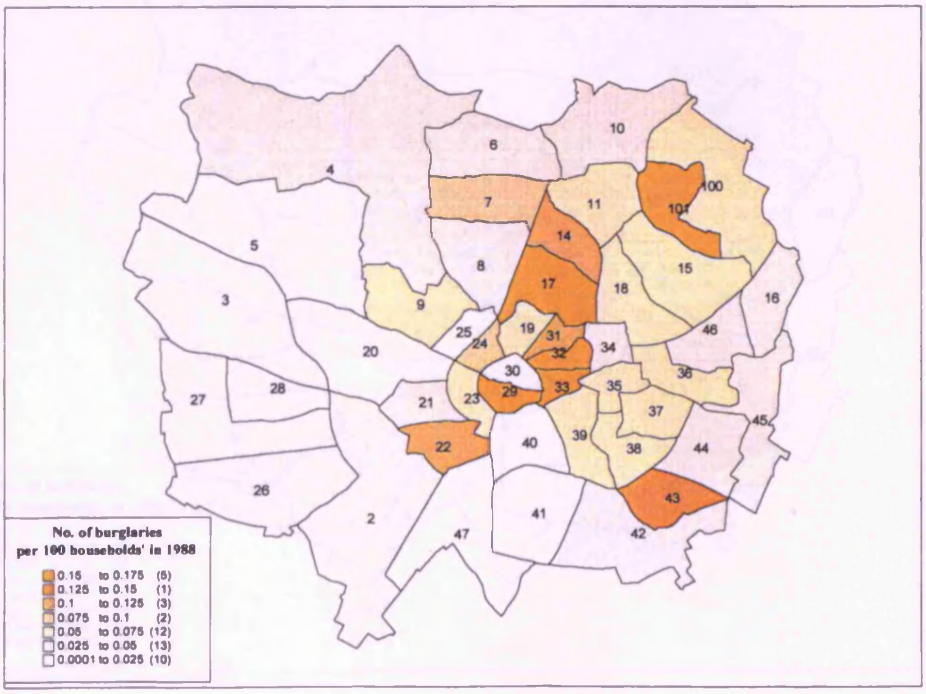
10
6
4
100
7
11
101
14
5
8
15
17
18
3
9
16
46
25
19
31
34
20
30
36
28
29
33
35
27
21
23
37
45
39
40
38
44
22
43
26
41
2
47
42
No. of burglaries
per 100 households in 1993
0.1 to 0.125 (1)
0.075 to 0.1 (1)
0.05 to 0.075 (4)
0.025 to 0.05 (15)
0.0001 to 0.025 (25)

Appendix 10.8

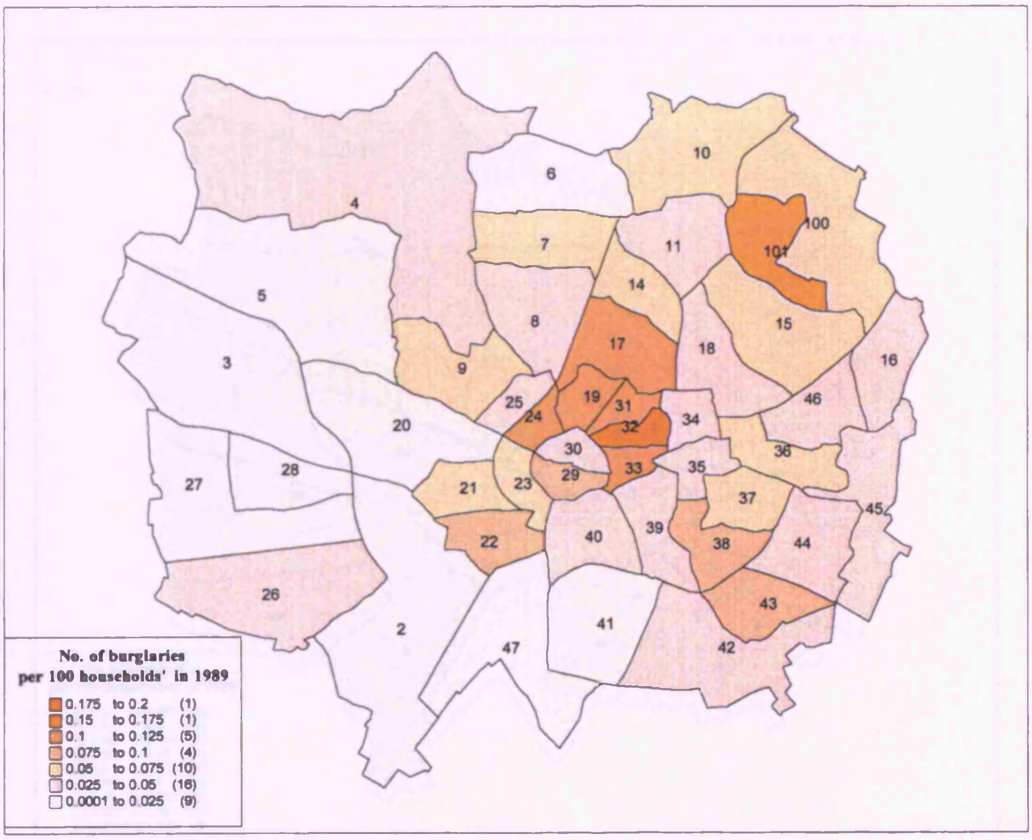
Thematic maps of Burglary Risk in Coventry (Monte Carlo dasymetric method)

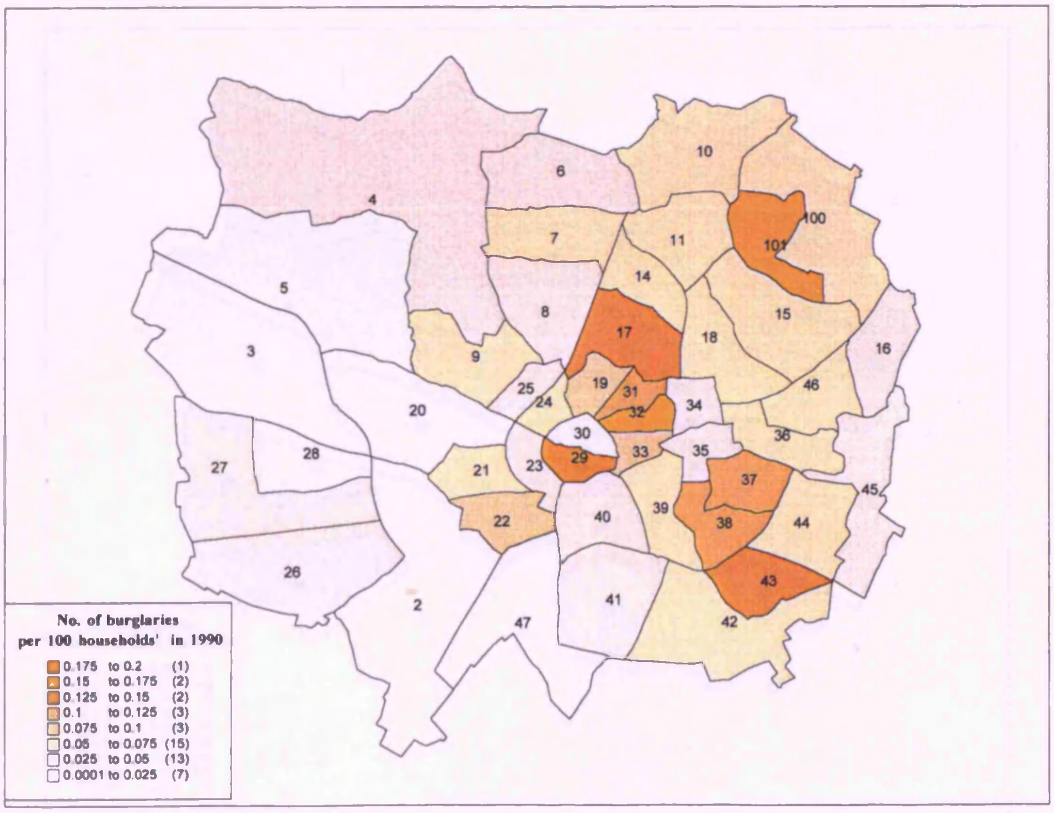
The following beat maps show the burglary risk from 1987-1993 for Coventry using the Monte Carlo dasymetric method.



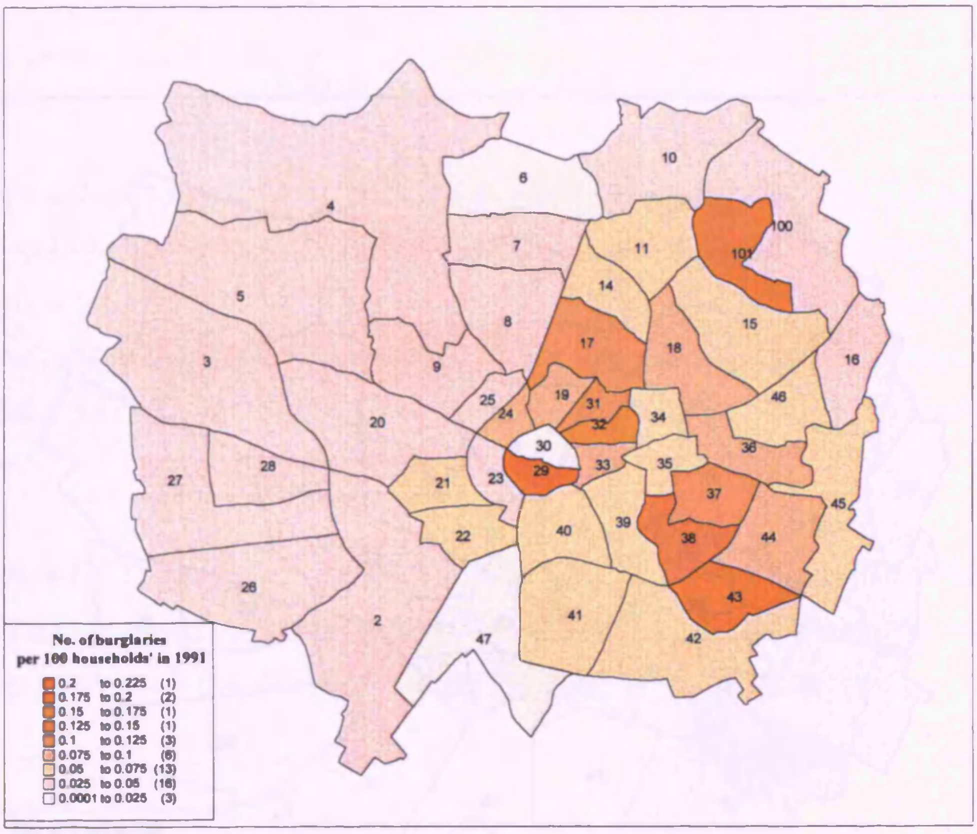


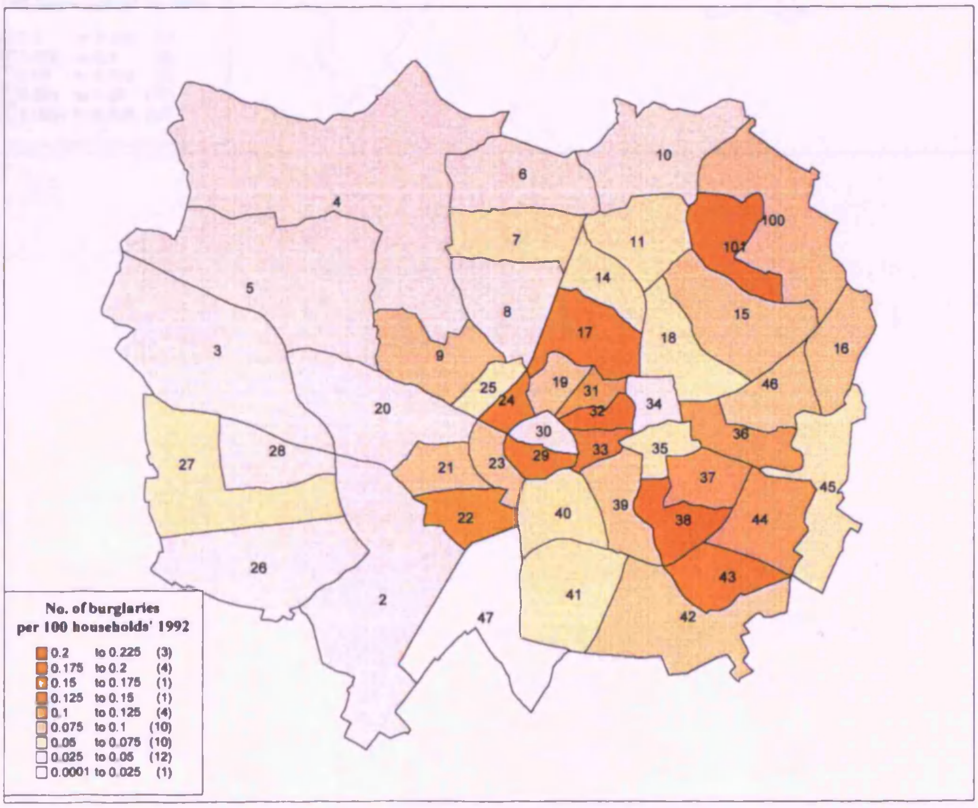
Appendix 10.8 Thematic maps of Burglary risk in Coventry (Monte Carlo dasymetric method)



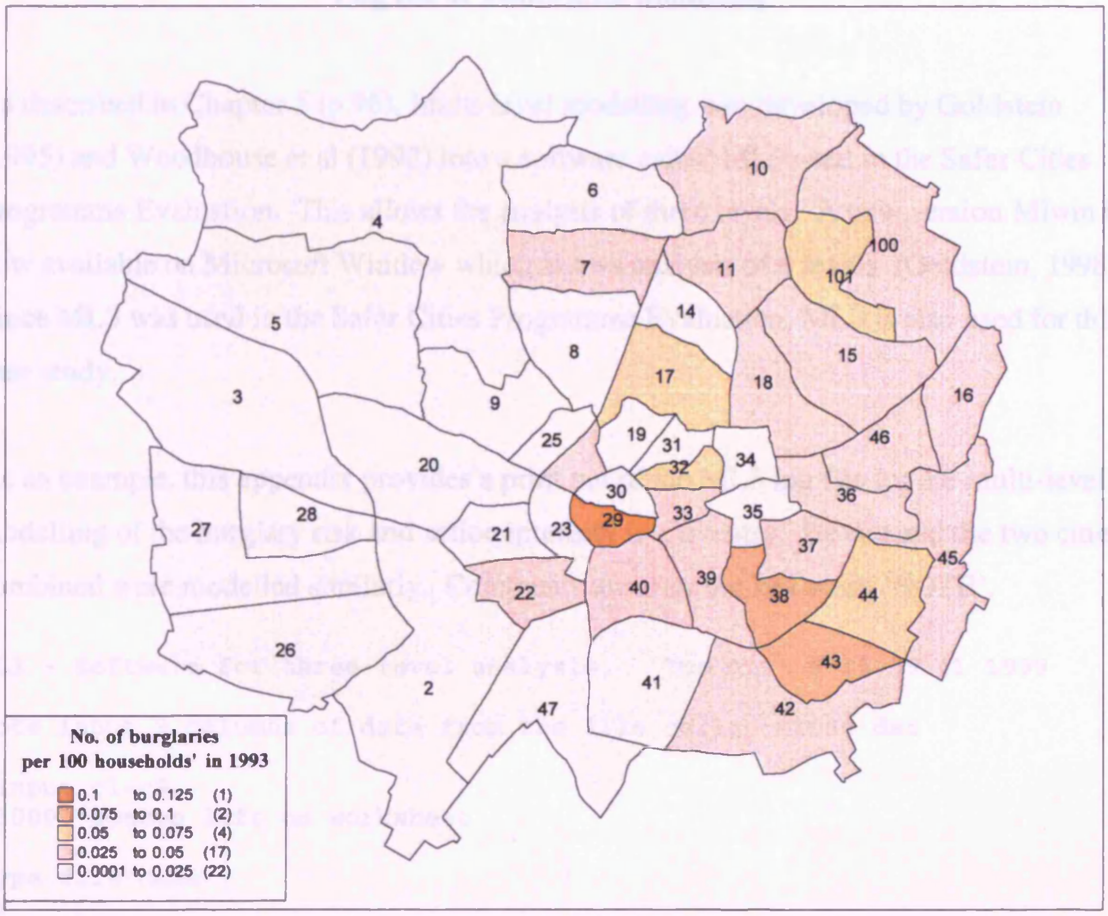


Appendix 10.8 Thematic maps of Burglary risk in Coventry (Monte Carlo dasymetric method)





Appendix 10.8 Thematic maps of Burglary risk in Coventry (Monte Carlo dasymetric method)



Appendix 11.1
Log file of Multi-level modelling

As described in Chapter 5 (p 96), Multi-level modelling was developed by Goldstein (1995) and Woodhouse et al (1992) into a software called ML3 used in the Safer Cities Programme Evaluation. This allows the analysis of three levels. A new version Mlwin is now available on Microsoft Window which allows analysis of n levels (Goldstein, 1998). Since ML3 was used in the Safer Cities Programme Evaluation, ML3 is also used for this case study.

As an example, this appendix provides a print out of the ML3 log file for the multi-level modelling of the burglary risk and action intensity in Coventry. Bristol and the two cities combined were modelled similarly. Comments are after the key word 'NOTE'.

```
ML3 - Software for three-level analysis.   Tue Apr  6 15:39:41 1999

Note input 9 columns of data from the file called mlc99.dat

dinput c1-c9
250000 spaces left on worksheet

Type file name
->
mlc99.dat
1      1      0.021529     0.39309636   0.012347     0.339846367 0     0
       1
1      2      0.040629     0.465602705  0.023301     0.401392135 0     0
       1
1      3      0.018229     0.376258863  0.010454     0.325473449 0     0
       1
1      4      0.028929     0.425102277  0.016591     0.3670844   0     0
       1
1      5      0.055529     0.50685922   0.031846     0.436134225 0     0
       1
1      6      0.063429     0.525733671  0.036377     0.451953425 0     0
       1
1      7      0.0168       0.368289946  0.009635     0.318667924 0     0
       1
2      8      0.014728     0.355827497  0.015488     0.360536018 0     0
       1
2      9      0.010528     0.326069146  0.011072     0.330358147 0     0
       1
2      10     0.010128     0.322812327  0.010651     0.327052641 0     0
       1
2      11     0.019284     0.381861373  0.020279     0.386949283 0     0
       1
```

2	12 1	0.025428	0.410782034	0.026741	0.416305659	0	0
2	13 1	0.026228	0.414170689	0.027582	0.419744393	0	0
2	14 1	0.0068	0.291262909	0.007151	0.295060971	0	0
3	15 1	0.011128	0.330790971	0.012663	0.342088078	0	0
3	16 1	0.031828	0.436068354	0.036217	0.451418795	0	0
3	17 1	0.027228	0.418306015	0.030983	0.43294648	0	0
3	18 1	0.023628	0.402873931	0.026887	0.416908017	0	0
3	19 1	0.034128	0.444283646	0.038834	0.459966716	0	0
3	20 1	0.031128	0.433486364	0.035421	0.448734453	0	0
3	21 1	0.0052	0.271871807	0.005917	0.281032953	0	0
4	22 1	0.010928	0.32923788	0.011941	0.336904833	0	0
4	23 1	0.018628	0.378404156	0.020355	0.387330569	0	0
4	24 1	0.012928	0.343937217	0.014127	0.351979389	0	0
4	25 1	0.017128	0.370160533	0.018716	0.3788729	0	0
4	26 1	0.028528	0.423525818	0.031173	0.433653558	0	0
4	27 1	0.025328	0.410353211	0.027676	0.420124161	0	0
4	28 1	0.0041	0.255825319	0.00448	0.261690375	0	0
5	29 1	0.033328	0.441471023	0.047639	0.486106428	0	0
5	30 1	0.047628	0.486075923	0.068079	0.536118379	0	0
5	31 1	0.026328	0.414589146	0.037633	0.456094875	0	0
5	32 1	0.033128	0.440760559	0.047353	0.485311747	0	0
5	33 1	0.028128	0.421937777	0.040206	0.464290223	0	0
5	34 1	0.035628	0.449436499	0.050927	0.495017994	0	0
5	35 1	0.0166	0.367136536	0.023728	0.403324192	0	0
6	36 1	0.035128	0.447735866	0.027614	0.419873777	0	0
6	37 1	0.060828	0.519702477	0.047817	0.486599382	0	0
6	38 1	0.055228	0.506105408	0.043415	0.474020982	0	0
6	39 1	0.044628	0.47757078	0.035082	0.447578568	0	0

6	40 1	0.031228	0.433857678	0.024548	0.406966853	0	0
6	41 1	0.040128	0.46404717	0.031545	0.435029299	0	0
6	42 1	0.0262	0.41405332	0.020596	0.38853302	0	0
7	43 1	0.026428	0.41500649	0.033508	0.442107914	0	0
7	44 1	0.038528	0.458988142	0.048849	0.489433102	0	0
7	45 1	0.052278	0.49856664	0.066283	0.53216495	0	0
7	46 1	0.033928	0.443584815	0.043017	0.472841484	0	0
7	47 1	0.038128	0.457700851	0.048342	0.488046093	0.01	0.01
7	48 1	0.041278	0.467598294	0.052336	0.498717592	0.01	0.01
7	49 1	0.0227	0.398630021	0.028781	0.424522231	0.01	0.01
8	50 1	0.011628	0.334587598	0.013091	0.345061132	0	0
8	51 1	0.038228	0.458023543	0.043037	0.472900933	0	0
8	52 1	0.041928	0.469575491	0.047203	0.48489365	0	0
8	53 1	0.042628	0.471681416	0.047991	0.487080054	0	0
8	54 1	0.034528	0.445672837	0.038872	0.460087867	0.03	0.03
8	55 1	0.056328	0.508846959	0.063414	0.525699366	0.03	0.03
8	56 1	0.0132	0.345807114	0.014861	0.356663832	0.03	0.03
9	57 1	0.022727	0.39875521	0.023677	0.403094726	0	0
9	58 1	0.042272	0.470613389	0.044039	0.475855444	0	0
9	59 1	0.045827	0.481015416	0.047743	0.486394598	0	0
9	60 1	0.056327	0.508844484	0.058682	0.51459526	0.19	0.2
9	61 1	0.038627	0.459305323	0.040242	0.464402292	0.39	0.41
9	62 1	0.044427	0.476987158	0.046284	0.482312157	0.53	0.55
9	63 1	0.0227	0.398630021	0.023649	0.402968597	0.53	0.55
10	64 1	0.040827	0.466213834	0.045494	0.480064962	0	0
10	65 1	0.064827	0.52890747	0.072237	0.545013649	0	0
10	66 1	0.047327	0.485239341	0.052737	0.499758162	0	0
10	67 1	0.054227	0.503578465	0.060425	0.518752686	0	0

10	68	0.061527	0.521339961	0.06856	0.537165428	0	0	1
10	69	0.051927	0.497650697	0.057862	0.51261074	0	0	1
10	70	0.024	0.404542178	0.026743	0.416313926	0	0	1
11	71	0.045927	0.481299917	0.051732	0.497140029	0	0	1
11	72	0.102627	0.601644128	0.115599	0.622531868	0	0	1
11	73	0.061527	0.521339961	0.069304	0.538775455	0	0	1
11	74	0.063527	0.525957663	0.071557	0.543582441	0	0	1
11	75	0.063527	0.525957663	0.071557	0.543582441	0	0	1
11	76	0.063127	0.525041956	0.071106	0.542628266	0	0	1
11	77	0.0176	0.372808085	0.019825	0.384650248	0	0	1
12	78	0.053728	0.502307026	0.054098	0.503250535	0	0	1
12	79	0.071328	0.543098446	0.07182	0.544137039	0	0	1
12	80	0.078428	0.5576545	0.078969	0.558727643	0	0	1
12	81	0.070728	0.54182546	0.071216	0.542861359	0.31	0.31	1
12	82	0.064128	0.527326324	0.06457	0.528327457	2.52	2.54	1
12	83	0.104328	0.604474202	0.105047	0.605661698	4.81	4.84	1
12	84	0.0352	0.447981786	0.035443	0.448809202	4.81	4.84	1
13	85	0.034228	0.444631998	0.0315	0.434863482	0	0	1
13	86	0.042828	0.472278753	0.039414	0.461807058	0	0	1
13	87	0.038228	0.458023543	0.035181	0.447916924	0	0	1
13	88	0.047328	0.485242127	0.043556	0.474437069	0	0	1
13	89	0.042228	0.470480958	0.038862	0.460055993	0	0	1
13	90	0.073628	0.547913932	0.06776	0.535421272	0	0	1
13	91	0.0434	0.473976663	0.039941	0.46346314	0	0	1
14	92	0.056927	0.510324745	0.06501	0.529319545	0	0	1
14	93	0.091527	0.582401569	0.104524	0.604798425	0	0	1
14	94	0.060127	0.518047632	0.068665	0.537393347	0	0	1
14	95	0.069227	0.538609358	0.079057	0.558901749	0	0	1
14	96	0.066527	0.532706187	0.075974	0.552724977	0.1	0.11	1

14	97	0.0995	0.596363093	0.113629	0.619457652	1.03	1.18	1
14	98	0.0279	0.421025515	0.031862	0.436192755	1.03	1.18	1
15	99	0.045727	0.48073049	0.047707	0.486294895	0	0	1
15	100	0.062227	0.522967389	0.064921	0.529119233	0	0	1
15	101	0.042227	0.470477947	0.044055	0.475902247	0	0	1
15	102	0.059727	0.517097568	0.062313	0.523166486	0	0	1
15	103	0.07427	0.549240387	0.077486	0.555774328	0	0	1
15	104	0.064627	0.52845623	0.067425	0.534686857	0	0	1
15	105	0.025	0.408938265	0.026083	0.413561932	0	0	1
16	106	0.0632	0.52520936	0.07735	0.555501649	0	0	1
16	107	0.064126	0.527321783	0.078483	0.557763819	0	0	1
16	108	0.117626	0.625661248	0.143962	0.663621033	0	0	1
16	109	0.102256	0.601022939	0.125151	0.636995741	0	0	1
16	110	0.101126	0.599122078	0.123768	0.634944719	0.22	0.27	1
16	111	0.09901	0.595526061	0.121178	0.631065459	2.4	2.94	1
16	112	0.0253	0.410232927	0.030965	0.432879337	2.4	2.94	1
17	113	0.015629	0.361391311	0.016304	0.365411164	0	0	1
17	114	0.023287	0.401328368	0.024293	0.405843374	0	0	1
17	115	0.015129	0.358332951	0.015782	0.362313183	0	0	1
17	116	0.021729	0.394056285	0.022667	0.398476871	0	0	1
17	117	0.032529	0.438614739	0.033934	0.443605821	0	0	1
17	118	0.035929	0.450452331	0.037481	0.455598804	0	0	1
17	119	0.01	0.321750554	0.010432	0.325295764	0	0	1
18	120	0.042429	0.471085157	0.082999	0.566574855	0	0	1
18	121	0.046129	0.481873325	0.090237	0.580070356	0	0	1
18	122	0.055929	0.507856721	0.109408	0.612756977	0	0	1
18	123	0.059629	0.516864153	0.116646	0.62415249	0	0	1
18	124	0.055929	0.507856721	0.109408	0.612756977	0	0	1
18	125	0.073829	0.548330037	0.144424	0.664247766	0	0	1

18	126 1	0.0159	0.36301982	0.031104	0.433397125	0	0
19	127 1	0.044229	0.476410482	0.075118	0.55098098	0	0
19	128 1	0.052329	0.498699379	0.088875	0.577585667	0	0
19	129 1	0.043829	0.475240073	0.074439	0.54958831	0	0
19	130 1	0.048429	0.4882848	0.082252	0.565139307	0	0
19	131 1	0.037829	0.456732508	0.064249	0.527600844	0	0
19	132 1	0.082829	0.566248897	0.140677	0.659129635	0	0
19	133 1	0.0174	0.37169249	0.029552	0.427521477	0	0
20	134 1	0.038127	0.457697621	0.071354	0.543153449	0	0
20	135 1	0.047527	0.485795612	0.088946	0.577715795	0	0
20	136 1	0.047527	0.485795612	0.088946	0.577715795	0	0
20	137 1	0.028127	0.421933788	0.052639	0.499504353	0	0
20	138 1	0.043927	0.475527496	0.082209	0.565056414	0.05	0.09
20	139 1	0.075627	0.552019566	0.141535	0.660308738	0.05	0.09
20	140 1	0.0217	0.393917484	0.040612	0.465550139	0.05	0.09
21	141 1	0.043425	0.474050522	0.03791	0.456995355	0	0
21	142 1	0.06765	0.535180386	0.059058	0.515499023	0	0
21	143 1	0.076254	0.553292648	0.06657	0.532801432	0	0
21	144 1	0.054325	0.503827239	0.047426	0.485514895	0	0
21	145 1	0.0627	0.524060164	0.054737	0.504869809	0.16	0.14
21	146 1	0.144825	0.66479078	0.126433	0.638884602	0.16	0.14
21	147 1	0.0338	0.443136062	0.029507	0.427347936	0.16	0.14
22	148 1	0.038426	0.458660761	0.032739	0.439370131	0	0
22	149 1	0.027426	0.419112135	0.023367	0.401692388	0	0
22	150 1	0.039526	0.462160283	0.033676	0.442700207	0	0
22	151 1	0.032926	0.440039961	0.028053	0.421638264	0	0
22	152 1	0.034255	0.444725932	0.029185	0.426100725	0.03	0.03
22	153 1	0.075726	0.552221037	0.064518	0.528209913	0.03	0.03

22	154 1	0.0165	0.366556113	0.014058	0.351530137	0.03	0.03
23	155 1	0.025429	0.410786316	0.016765	0.368088809	0	0
23	156 1	0.029329	0.426659665	0.019337	0.382137024	0	0
23	157 1	0.02129	0.391940944	0.014036	0.351386566	0	0
23	158 1	0.023729	0.403328688	0.015645	0.361488017	0	0
23	159 1	0.02829	0.422582825	0.018652	0.378532151	0	0
23	160 1	0.029329	0.426659665	0.019337	0.382137024	0.23	0.15
23	161 1	0.0066	0.289034584	0.004351	0.259741673	0.23	0.15
24	162 1	0.017429	0.371854816	0.016714	0.367795186	0	0
24	163 1	0.019229	0.381574751	0.01844	0.377397453	0	0
24	164 1	0.0096	0.318367922	0.009206	0.314934911	0	0
24	165 1	0.028529	0.423529768	0.027359	0.418839815	0	0
24	166 1	0.039229	0.461222089	0.03762	0.456052502	0	0
24	167 1	0.042129	0.470182643	0.040401	0.464896441	0	0
24	168 1	0.0067	0.290154797	0.006425	0.287044116	0	0
25	169 1	0.017929	0.374623769	0.02285	0.399324186	0	0
25	170 1	0.016329	0.365557748	0.020811	0.389597365	0	0
25	171 1	0.010829	0.32846151	0.013801	0.349842805	0	0
25	172 1	0.021629	0.393577106	0.027566	0.419679662	0	0
25	173 1	0.026829	0.416668999	0.034193	0.444510155	0	0
25	174 1	0.031129	0.433490081	0.039673	0.462622849	0	0
25	175 1	0.0049	0.267763327	0.006245	0.284955151	0	0
26	176 1	0.153825	0.676750475	0.085527	0.5713718	0	0
26	177 1	0.174825	0.703148515	0.097203	0.59241633	0	0
26	178 1	0.0909	0.581271145	0.05054	0.493989746	0	0
26	179 1	0.188825	0.719749695	0.104987	0.605562799	0	0
26	180 1	0.216825	0.751015787	0.120555	0.630124736	0	0
26	181 1	0.195825	0.727791991	0.108879	0.611905879	0	0

26	182	0.1119	0.61673205	0.062216	0.52294191	0	0	1
27	183	0.017425	0.371832437	0.023226	0.40105021	0	0	1
27	184	0.024925	0.408612895	0.033223	0.441098399	0	0	1
27	185	0.054725	0.504839517	0.072943	0.546490228	0	0	1
27	186	0.034825	0.446697114	0.046419	0.48269355	0	0	1
27	187	0.044825	0.478141047	0.059748	0.517147552	0.08	0.11	1
27	188	0.049825	0.49207589	0.066412	0.53245126	0.08	0.11	1
27	189	0.0149	0.356908057	0.01986	0.384828806	0.08	0.11	1
28	190	0.050326	0.493418862	0.081266	0.563231433	0	0	1
28	191	0.082426	0.565474445	0.133101	0.648525183	0	0	1
28	192	0.088	0.575976453	0.142102	0.661085586	0	0	1
28	193	0.086426	0.573055537	0.13956	0.657588119	0	0	1
28	194	0.090526	0.580594471	0.146181	0.666620394	0.57	0.92	1
28	195	0.100226	0.59759847	0.161844	0.687063881	6.27	10.12	1
28	196	0.0178	0.373914693	0.028743	0.424372964	6.27	10.12	1
29	197	0.047226	0.484957817	0.040855	0.466300092	0	0	1
29	198	0.105258	0.606009212	0.091059	0.58155828	0	0	1
29	199	0.122126	0.632491226	0.105651	0.606655309	0	0	1
29	200	0.093626	0.586150134	0.080996	0.562706371	0	0	1
29	201	0.100926	0.598784243	0.087311	0.574702039	0.57	0.49	1
29	202	0.129426	0.64324927	0.111966	0.616836573	6.27	5.42	1
29	203	0.0399	0.46333484	0.034517	0.445634784	6.27	5.42	1
30	204	0.157626	0.681677285	0.120269	0.629691875	0	0	1
30	205	0.174826	0.703149727	0.133392	0.648939149	0	0	1
30	206	0.143326	0.662756289	0.109358	0.612676644	0	0	1
30	207	0.108926	0.611981601	0.08311	0.566787453	0	0	1
30	208	0.088826	0.577495821	0.067774	0.535451911	0.2	0.15	1
30	209	0.151926	0.674262171	0.115919	0.62302815	2.25	1.72	1
30	210	0.0516	0.496793607	0.039371	0.461671261	2.25	1.72	1

31	2111	0.063126	0.525039662	0.065784	0.531053947	0	0
31	2121	0.065426	0.530253424	0.06818	0.536338643	0	0
31	2131	0.049926	0.492347363	0.052028	0.497914686	0	0
31	2141	0.055126	0.505849336	0.057447	0.511599184	0	0
31	2151	0.068326	0.536656662	0.071203	0.542833824	0.11	0.11
31	2161	0.065426	0.530253424	0.06818	0.536338643	1.22	1.27
31	2171	0.0253	0.410232927	0.026365	0.414743693	1.22	1.27
32	2181	0.041126	0.467132856	0.058218	0.513474613	0	0
32	2191	0.053261	0.501109885	0.075396	0.551548793	0	0
32	2201	0.042126	0.470173595	0.059634	0.516876068	0	0
32	2211	0.037626	0.45607206	0.053264	0.501117598	0	0
32	2221	0.065326	0.530029294	0.092476	0.584103119	0	0
32	2231	0.070326	0.54096857	0.099554	0.596455177	0	0
32	2241	0.0158	0.362421219	0.022366	0.397072637	0	0
33	2251	0.047626	0.486070376	0.043268	0.473586203	0	0
33	2261	0.045126	0.479009073	0.040997	0.46673692	0	0
33	2271	0.035526	0.449090917	0.032276	0.437700156	0	0
33	2281	0.036726	0.453113975	0.033366	0.441605676	0	0
33	2291	0.057226	0.511058498	0.05199	0.497815404	0	0
33	2301	0.081826	0.564316846	0.074339	0.549382501	0	0
33	2311	0.0246	0.407194946	0.022349	0.396992931	0	0
34	2321	0.050826	0.494750152	0.043146	0.473224598	0	0
34	2331	0.079326	0.559433175	0.06734	0.534500129	0	0
34	2341	0.076226	0.553235943	0.064709	0.52864135	0	0
34	2351	0.137263	0.654394668	0.116523	0.623962571	0	0
34	2361	0.137263	0.654394668	0.116523	0.623962571	0	0
34	2371	0.127826	0.640923765	0.108512	0.611313895	0	0
34	2381	0.0408	0.466130619	0.034635	0.446042554	0	0

35	239	0.057526	0.511792122	0.047154	0.484756876	0	0	1
35	240	0.063626	0.526183707	0.052155	0.498246143	0	0	1
35	241	0.091526	0.58239977	0.075024	0.550788674	0	0	1
35	242	0.132526	0.647705592	0.108632	0.611507597	0	0	1
35	243	0.150826	0.672812413	0.123632	0.634742269	0	0	1
35	244	0.246726	0.78211335	0.202242	0.735026242	0	0	1
35	245	0.0898	0.579275784	0.073609	0.54787456	0	0	1
36	246	0.059627	0.516859387	0.065512	0.530445993	0	0	1
36	247	0.061327	0.520872713	0.067379	0.534585824	0	0	1
36	248	0.059627	0.516859387	0.065512	0.530445993	0	0	1
36	249	0.072827	0.546248265	0.080014	0.560787038	0	0	1
36	250	0.0778	0.556402702	0.085479	0.57128158	0	0	1
36	251	0.112265	0.617309613	0.123346	0.634316082	0	0	1
36	252	0.0386	0.459218875	0.04241	0.471028128	0	0	1
37	253	0.018927	0.379990477	0.020053	0.385809441	0	0	1
37	254	0.022727	0.39875521	0.024079	0.404894109	0	0	1
37	255	0.027527	0.419521768	0.029164	0.426019052	0	0	1
37	256	0.042627	0.471678425	0.045163	0.479115501	0	0	1
37	257	0.049727	0.491812123	0.052685	0.499623527	0	0	1
37	258	0.049727	0.491812123	0.052685	0.499623527	0	0	1
37	259	0.027	0.417372652	0.028607	0.423837612	0	0	1
38	260	0.011929	0.336816804	0.014998	0.35751974	0	0	1
38	261	0.009729	0.319469766	0.012232	0.339020322	0	0	1
38	262	0.014529	0.354565989	0.018266	0.376459196	0	0	1
38	263	0.035829	0.4501155	0.045045	0.478775874	0	0	1
38	264	0.049529	0.491278128	0.062268	0.52306233	0	0	1
38	265	0.060929	0.519939855	0.0766	0.553992246	0	0	1
38	266	0.0182	0.376101643	0.022881	0.399467246	0	0	1
39	267	0.025229	0.409927503	0.029397	0.426922943	0	0	1

39	268	0.038729	0.459631531	0.045127	0.47901195	0	0	1
39	269	0.024929	0.408630265	0.029048	0.425567164	0	0	1
39	270	0.045429	0.479878884	0.052934	0.500267407	0	0	1
39	271	0.044729	0.477863364	0.052119	0.498152242	0	0	1
39	272	0.061429	0.521111136	0.071578	0.543626774	0	0	1
39	273	0.0332	0.441016668	0.038685	0.459490887	0	0	1
40	274	0.10733	0.609398706	0.115379	0.622190178	0	0	1
40	275	0.11433	0.620555346	0.122904	0.633656234	0	0	1
40	276	0.08423	0.568922357	0.090547	0.580632514	0	0	1
40	277	0.11013	0.613914441	0.118389	0.62683054	0	0	1
40	278	0.15853	0.682838759	0.170419	0.697768613	0	0	1
40	279	0.17043	0.697782143	0.183212	0.713180034	0	0	1
40	280	0.0645	0.528169211	0.069338	0.538848758	0	0	1
41	281	0.03423	0.444638958	0.038953	0.460345839	0	0	1
41	282	0.04296	0.47267195	0.048888	0.48953939	0	0	1
41	283	0.03423	0.444638958	0.038953	0.460345839	0	0	1
41	284	0.065296	0.529962011	0.074307	0.549316604	0	0	1
41	285	0.0753	0.551352869	0.085691	0.571679806	0	0	1
41	286	0.09893	0.595389153	0.112582	0.617810273	0	0	1
41	287	0.051	0.495211359	0.058038	0.513038268	0	0	1
42	288	0.03297	0.440197184	0.02809	0.421786094	0	0	1
42	289	0.03323	0.441123266	0.028312	0.422670224	0	0	1
42	290	0.02683	0.416673123	0.022859	0.399365734	0	0	1
42	291	0.038297	0.458245861	0.032629	0.438974871	0	0	1
42	292	0.05033	0.493429548	0.042881	0.472436727	0	0	1
42	293	0.04453	0.477286453	0.037939	0.457089366	0	0	1
42	294	0.023	0.400015144	0.019596	0.383476417	0	0	1
43	295	0.038528	0.458988142	0.039764	0.462908608	0	0	1
43	296	0.045528	0.480162222	0.046989	0.484295597	0	0	1

43	297	0.032128	0.437162831	0.033159	0.440870876	0	0	1
43	298	0.051228	0.49581409	0.052872	0.500107275	0	0	1
43	299	0.050928	0.495020644	0.052562	0.499304707	0	0	1
43	300	0.079276	0.559334486	0.081821	0.564307176	0	0	1
43	301	0.0368	0.453359071	0.037981	0.457225433	0	0	1
44	302	0.007629	0.300021171	0.008378	0.307358087	0	0	1
44	303	0.010729	0.327672084	0.011782	0.335733242	0	0	1
44	304	0.013129	0.345321703	0.014417	0.353850547	0	0	1
44	305	0.0099	0.320914194	0.010871	0.328791501	0	0	1
44	306	0.016529	0.366724693	0.018151	0.375835591	0	0	1
44	307	0.036829	0.45345503	0.040442	0.465023646	0	0	1
44	308	0.0078	0.301740778	0.008565	0.309115159	0	0	1
45	309	0.044227	0.476404648	0.038036	0.457403459	0	0	1
45	310	0.047727	0.486350292	0.041046	0.466887415	0	0	1
45	311	0.055427	0.506604084	0.047668	0.486186826	0	0	1
45	312	0.048927	0.48964562	0.042078	0.470028778	0.88	0.76	1
45	313	0.040327	0.464666626	0.034682	0.446204702	1.83	1.57	1
45	314	0.083327	0.567202541	0.071662	0.543804019	2.52	2.17	1
45	315	0.0321	0.437060981	0.027606	0.419841441	2.52	2.17	1
46	316	0.137427	0.654623734	0.150222	0.672013705	0	0	1
46	317	0.160827	0.685772684	0.1758	0.704328527	0	0	1
46	318	0.145627	0.665874123	0.159185	0.683677896	0	0	1
46	319	0.166227	0.692575358	0.181703	0.71139462	2.59	2.83	1
46	320	0.189273	0.720269314	0.206894	0.740192623	7.9	8.64	1
46	321	0.215327	0.749400477	0.235374	0.770549565	10.49	11.47	1
46	322	0.0516	0.496793607	0.056404	0.509035041	10.49	11.47	1

```
note Give each column-variable a sensible name
note riskov and riskmc are the transformed burglary risks from overlay
method and the Monte Carlo method respectively, risktov and risktmc are
```

```
the burglary risks from overlay method and the Monte Carlo method
respectively.
Note actov and actmc are the action scores from overlay method and the
Monte Carlo method respectively.

name c1 'beat-no' c2 'case' c3 'riskov' c4 'risktov' c5 'riskmc'
name c6 'risktmc' c7 'actov' c8 'actmc' c9 'cons'
Note look at the initial setting of the multi-level model

sett
EXPLanatory variables in
FPARameters
FMEAns
RMEAns
RESPonse variable in
IDENtifying codes for level 1:          level 2:          level 3:
RESEtting covariances level 1: ON       level 2: ON       level 3: ON
MAXIterations   5     TOLErance   2     METHod is IGLS    BATCh is OFF


LEVEL 3 RANDOM PARAMETER MATRIX unspecified
LEVEL 2 RANDOM PARAMETER MATRIX unspecified
LEVEL 1 RANDOM PARAMETER MATRIX unspecified

Note Check to see the variables are all ok

names

      Name               n  missing     min             max

    1 BEAT-NO          322      0        1.0000          46.000
    2 CASE             322      0        1.0000          322.00
    3 RISKOV           322      0     0.0041000         0.24673
    4 RISKTOV          322      0       0.25583         0.78211
    5 RISKMC           322      0     0.0043510         0.23537
    6 RISKTMC          322      0       0.25974         0.77055
    7 ACTOV            322      0       0.00000          10.490
    8 ACTMC            322      0       0.00000          11.470
    9 CONS             322      0        1.0000          1.0000
   10 C10                0
   11 C11                0
   12 C12                0
   13 C13                0
   14 C14                0
   15 C15                0
   16 C16                0
   17 C17                0
   18 C18                0
   19 C19                0
   20 C20                0
q

Note above quit for no more printing
Note print the variables to their values read in ok

print 'beat-no'-'riskmc'
```

N =	BEAT-NO 322	CASE 322	RISKOV 322	RISKTOV 322	RISKMC 322
1	1.0000	1.0000	0.021529	0.39310	0.012347
2	1.0000	2.0000	0.040629	0.46560	0.023301
3	1.0000	3.0000	0.018229	0.37626	0.010454
4	1.0000	4.0000	0.028929	0.42510	0.016591
5	1.0000	5.0000	0.055529	0.50686	0.031846
6	1.0000	6.0000	0.063429	0.52573	0.036377
7	1.0000	7.0000	0.016800	0.36829	0.0096350
8	2.0000	8.0000	0.014728	0.35583	0.015488
9	2.0000	9.0000	0.010528	0.32607	0.011072
10	2.0000	10.000	0.010128	0.32281	0.010651
11	2.0000	11.000	0.019284	0.38186	0.020279
12	2.0000	12.000	0.025428	0.41078	0.026741
13	2.0000	13.000	0.026228	0.41417	0.027582
14	2.0000	14.000	0.0068000	0.29126	0.0071510
15	3.0000	15.000	0.011128	0.33079	0.012663
16	3.0000	16.000	0.031828	0.43607	0.036217
17	3.0000	17.000	0.027228	0.41831	0.030983
18	3.0000	18.000	0.023628	0.40287	0.026887
19	3.0000	19.000	0.034128	0.44428	0.038834
20	3.0000	20.000	0.031128	0.43349	0.035421

q
print 'riskmc'-'cons'

N =	RISKMC 322	RISKTMC 322	ACTOV 322	ACTMC 322	CONS 322
1	0.012347	0.33985	0.00000	0.00000	1.0000
2	0.023301	0.40139	0.00000	0.00000	1.0000
3	0.010454	0.32547	0.00000	0.00000	1.0000
4	0.016591	0.36708	0.00000	0.00000	1.0000
5	0.031846	0.43613	0.00000	0.00000	1.0000
6	0.036377	0.45195	0.00000	0.00000	1.0000
7	0.0096350	0.31867	0.00000	0.00000	1.0000
8	0.015488	0.36054	0.00000	0.00000	1.0000
9	0.011072	0.33036	0.00000	0.00000	1.0000
10	0.010651	0.32705	0.00000	0.00000	1.0000
11	0.020279	0.38695	0.00000	0.00000	1.0000
12	0.026741	0.41631	0.00000	0.00000	1.0000
13	0.027582	0.41974	0.00000	0.00000	1.0000
14	0.0071510	0.29506	0.00000	0.00000	1.0000
15	0.012663	0.34209	0.00000	0.00000	1.0000
16	0.036217	0.45142	0.00000	0.00000	1.0000
17	0.030983	0.43295	0.00000	0.00000	1.0000
18	0.026887	0.41691	0.00000	0.00000	1.0000
19	0.038834	0.45997	0.00000	0.00000	1.0000
20	0.035421	0.44873	0.00000	0.00000	1.0000

q

Note initial model

sett
EXPLanatory variables in
FPARameters
FMEAns
RMEAns
RESPonse variable in

```
IDENtifying codes for level 1:          level 2:          level 3:
RESEtting covariances level 1: ON       level 2: ON       level 3: ON
MAXIterations   5     TOLErance   2     METHod is IGLS    BATCh is OFF


LEVEL 3 RANDOM PARAMETER MATRIX unspecified
LEVEL 2 RANDOM PARAMETER MATRIX unspecified
LEVEL 1 RANDOM PARAMETER MATRIX unspecified

Note first define the explanatory variable as CONS to set the base
line, and response variable as burglary risk (overlay method.  Identify
the number of cases (beat-years) as Level 1, and Beat-no as Level 2.
Set variance at Level 1 & 2 CONS as their variable names so that their
coefficients become the values of the variances.

expl 'cons'
resp 'riskov'
identify 1 'case' 2 'beat-no'
setv 1 'cons'
setv 2 'cons'

Note the new setting
sett
EXPLanatory variables in       CONS
FPARameters                    CONS
FMEAns
RMEAns
RESPonse variable in           RISKOV
IDENtifying codes for level 1: CASE     level 2: BEAT-NO  level 3:
RESEtting covariances level 1: ON       level 2: ON       level 3: ON
MAXIterations   5     TOLErance   2     METHod is IGLS    BATCh is OFF


LEVEL 3 RANDOM PARAMETER MATRIX unspecified
LEVEL 2 RANDOM PARAMETER MATRIX
         CONS
CONS     1
LEVEL 1 RANDOM PARAMETER MATRIX
         CONS
CONS     1

Note start running the model

start


Iteration number 1 in progress
Iteration number 1 completed
Convergence not achieved

Note so turn on the BATCH mode (a toggle command)
batch

BATCh mode is ON
next
```

```
Iteration number 2 in progress
Iteration number 2 completed
Convergence achieved

Note here goes the results of the fixed components
fixe


  PARAMETER                 ESTIMATE     S. ERROR     PREV. ESTIMATE
CONS                         0.05491     0.005022            0.05491


Note here goes the results of the random components
rand

                              LEVEL 3

  PARAMETER                 ESTIMATE     S. ERROR     PREV. ESTIMATE
NCONV

                              LEVEL 2

  PARAMETER                 ESTIMATE     S. ERROR     PREV. ESTIMATE
NCONV
CONS      /CONS             0.001074     0.000242           0.001074
1

                              LEVEL 1

  PARAMETER                 ESTIMATE     S. ERROR     PREV. ESTIMATE
NCONV
CONS      /CONS            0.0006024   0.00005128          0.0006024
1

Note look at the setting again
sett
EXPLanatory variables in       CONS
FPARameters                    CONS
FMEAns
RMEAns
RESPonse variable in           RISKOV
IDENtifying codes for level 1: CASE     level 2: BEAT-NO  level 3:
RESEtting covariances level 1: ON       level 2: ON       level 3: ON
MAXIterations   5     TOLErance   2     METHod is IGLS    BATCh is ON


LEVEL 3 RANDOM PARAMETER MATRIX unspecified
LEVEL 2 RANDOM PARAMETER MATRIX
        CONS
CONS    1
LEVEL 1 RANDOM PARAMETER MATRIX
        CONS
CONS    1

Note do the similar routine as above for the Monte Carlo method
resp 'riskmc'
sett
```

```
EXPLanatory variables in       CONS
FPARameters                    CONS
FMEAns
RMEAns
RESPonse variable in           RISKMC
IDENtifying codes for level 1: CASE     level 2: BEAT-NO  level 3:
RESEtting covariances level 1: ON       level 2: ON       level 3: ON
MAXIterations   5     TOLErance   2     METHod is IGLS    BATCh is ON


LEVEL 3 RANDOM PARAMETER MATRIX unspecified
LEVEL 2 RANDOM PARAMETER MATRIX
        CONS
CONS     1
LEVEL 1 RANDOM PARAMETER MATRIX
        CONS
CONS     1
start


Iteration number 1 in progress
Iteration number 1 completed

Iteration number 2 in progress
Iteration number 2 completed
Convergence achieved
fixe


  PARAMETER             ESTIMATE     S. ERROR     PREV. ESTIMATE
CONS                     0.05754     0.004768            0.05754
rand

                             LEVEL 3

  PARAMETER             ESTIMATE     S. ERROR     PREV. ESTIMATE
NCONV

                             LEVEL 2

  PARAMETER             ESTIMATE     S. ERROR     PREV. ESTIMATE
NCONV
CONS      /CONS        0.0009548    0.0002182          0.0009548
1

                             LEVEL 1

  PARAMETER             ESTIMATE     S. ERROR     PREV. ESTIMATE
NCONV
CONS      /CONS         0.000637   0.00005422           0.000637
1
sett
EXPLanatory variables in       CONS
FPARameters                    CONS
FMEAns
RMEAns
RESPonse variable in           RISKMC
```

```
IDENtifying codes for level 1: CASE     level 2: BEAT-NO  level 3:
RESEtting covariances level 1: ON       level 2: ON       level 3: ON
MAXIterations    5     TOLErance    2     METHod is IGLS    BATCh is ON


LEVEL 3 RANDOM PARAMETER MATRIX unspecified
LEVEL 2 RANDOM PARAMETER MATRIX
         CONS
CONS     1
LEVEL 1 RANDOM PARAMETER MATRIX
         CONS
CONS     1

Note do the similar routine as above for the transformed risk

resp 'risktov'

sett
EXPLanatory variables in        CONS
FPARameters                     CONS
FMEAns
RMEAns
RESPonse variable in            RISKTOV
IDENtifying codes for level 1: CASE     level 2: BEAT-NO  level 3:
RESEtting covariances level 1: ON       level 2: ON       level 3: ON
MAXIterations    5     TOLErance    2     METHod is IGLS    BATCh is ON


LEVEL 3 RANDOM PARAMETER MATRIX unspecified
LEVEL 2 RANDOM PARAMETER MATRIX
         CONS
CONS     1
LEVEL 1 RANDOM PARAMETER MATRIX
         CONS
CONS     1
start


Iteration number 1 in progress
Iteration number 1 completed

Iteration number 2 in progress
Iteration number 2 completed
Convergence achieved
fixe


  PARAMETER               ESTIMATE     S. ERROR     PREV. ESTIMATE
CONS                        0.4829      0.01162             0.4829
rand

                              LEVEL 3

  PARAMETER               ESTIMATE     S. ERROR     PREV. ESTIMATE
NCONV

                              LEVEL 2
```

```
  PARAMETER              ESTIMATE     S. ERROR     PREV. ESTIMATE
NCONV
CONS      /CONS          0.005732     0.001296          0.005732
1

                                LEVEL 1

  PARAMETER              ESTIMATE     S. ERROR     PREV. ESTIMATE
NCONV
CONS      /CONS           0.00336     0.000286           0.00336
1
sett
EXPLanatory variables in       CONS
FPARameters                    CONS
FMEAns
RMEAns
RESPonse variable in           RISKTOV
IDENtifying codes for level 1: CASE     level 2: BEAT-NO  level 3:
RESEtting covariances level 1: ON       level 2: ON       level 3: ON
MAXIterations   5     TOLErance   2     METHod is IGLS    BATCh is ON


LEVEL 3 RANDOM PARAMETER MATRIX unspecified
LEVEL 2 RANDOM PARAMETER MATRIX
         CONS
CONS     1
LEVEL 1 RANDOM PARAMETER MATRIX
         CONS
CONS     1

Note do the similar routine as above for the transformed risk with the
Monte Carlo method

resp 'risktmc'
sett
EXPLanatory variables in       CONS
FPARameters                    CONS
FMEAns
RMEAns
RESPonse variable in           RISKTMC
IDENtifying codes for level 1: CASE     level 2: BEAT-NO  level 3:
RESEtting covariances level 1: ON       level 2: ON       level 3: ON
MAXIterations   5     TOLErance   2     METHod is IGLS    BATCh is ON


LEVEL 3 RANDOM PARAMETER MATRIX unspecified
LEVEL 2 RANDOM PARAMETER MATRIX
         CONS
CONS     1
LEVEL 1 RANDOM PARAMETER MATRIX
         CONS
CONS     1
start


Iteration number 1 in progress
```

```
Iteration number 1 completed

Iteration number 2 in progress
Iteration number 2 completed
Convergence achieved
fixe


  PARAMETER               ESTIMATE      S. ERROR     PREV. ESTIMATE
CONS                        0.4899       0.01156             0.4899
rand

                                LEVEL 3

  PARAMETER               ESTIMATE      S. ERROR     PREV. ESTIMATE
NCONV

                                LEVEL 2

  PARAMETER               ESTIMATE      S. ERROR     PREV. ESTIMATE
NCONV
CONS      /CONS           0.005644      0.001282           0.005644
1

                                LEVEL 1

  PARAMETER               ESTIMATE      S. ERROR     PREV. ESTIMATE
NCONV
CONS      /CONS            0.00351     0.0002988            0.00351
1
sett
EXPLanatory variables in       CONS
FPARameters                    CONS
FMEAns
RMEAns
RESPonse variable in           RISKTMC
IDENtifying codes for level 1: CASE     level 2: BEAT-NO  level 3:
RESEtting covariances level 1: ON       level 2: ON       level 3: ON
MAXIterations    5     TOLErance    2     METHod is IGLS    BATCh is ON


LEVEL 3 RANDOM PARAMETER MATRIX unspecified
LEVEL 2 RANDOM PARAMETER MATRIX
         CONS
CONS     1
LEVEL 1 RANDOM PARAMETER MATRIX
         CONS
CONS     1

Note do the similar routine as before using action scores from the
Monte Carlo method

expl 'actmc'
sett
EXPLanatory variables in       CONS     ACTMC
FPARameters                    CONS     ACTMC
FMEAns
```

```
RMEAns
RESPonse variable in           RISKTMC
IDENtifying codes for level 1: CASE     level 2: BEAT-NO  level 3:
RESEtting covariances level 1: ON       level 2: ON       level 3: ON
MAXIterations   5     TOLErance   2     METHod is IGLS    BATCh is ON


LEVEL 3 RANDOM PARAMETER MATRIX unspecified
LEVEL 2 RANDOM PARAMETER MATRIX
         CONS
CONS     1
LEVEL 1 RANDOM PARAMETER MATRIX
         CONS
CONS     1
start


Iteration number 1 in progress
Iteration number 1 completed

Iteration number 2 in progress
Iteration number 2 completed

Iteration number 3 in progress
Iteration number 3 completed

Iteration number 4 in progress
Iteration number 4 completed
Convergence achieved
fixe


  PARAMETER                ESTIMATE      S. ERROR      PREV. ESTIMATE
CONS                         0.4911       0.01185              0.4911
ACTMC                      -0.00365      0.002817           -0.003649
rand

                                    LEVEL 3

  PARAMETER                ESTIMATE      S. ERROR      PREV. ESTIMATE
NCONV

                                    LEVEL 2

  PARAMETER                ESTIMATE      S. ERROR      PREV. ESTIMATE
NCONV
CONS      /CONS            0.005921      0.001338            0.005919
1

                                    LEVEL 1

  PARAMETER                ESTIMATE      S. ERROR      PREV. ESTIMATE
NCONV
CONS      /CONS            0.003465      0.000295            0.003465
2
sett
EXPLanatory variables in        CONS     ACTMC
```

```
FPARameters                   CONS      ACTMC
FMEAns
RMEAns
RESPonse variable in          RISKTMC
IDENtifying codes for level 1: CASE     level 2: BEAT-NO  level 3:
RESEtting covariances level 1: ON       level 2: ON       level 3: ON
MAXIterations   5     TOLErance   2     METHod is IGLS    BATCh is ON


LEVEL 3 RANDOM PARAMETER MATRIX unspecified
LEVEL 2 RANDOM PARAMETER MATRIX
         CONS
CONS     1
LEVEL 1 RANDOM PARAMETER MATRIX
         CONS
CONS     1
resp 'riskmc'
sett
EXPLanatory variables in      CONS      ACTMC
FPARameters                   CONS      ACTMC
FMEAns
RMEAns
RESPonse variable in          RISKMC
IDENtifying codes for level 1: CASE     level 2: BEAT-NO  level 3:
RESEtting covariances level 1: ON       level 2: ON       level 3: ON
MAXIterations   5     TOLErance   2     METHod is IGLS    BATCh is ON


LEVEL 3 RANDOM PARAMETER MATRIX unspecified
LEVEL 2 RANDOM PARAMETER MATRIX
         CONS
CONS     1
LEVEL 1 RANDOM PARAMETER MATRIX
         CONS
CONS     1
start


Iteration number 1 in progress
Iteration number 1 completed

Iteration number 2 in progress
Iteration number 2 completed

Iteration number 3 in progress
Iteration number 3 completed

Iteration number 4 in progress
Iteration number 4 completed
Convergence achieved
fixe


  PARAMETER              ESTIMATE      S. ERROR      PREV. ESTIMATE
CONS                      0.05778      0.004844            0.05778
ACTMC                  -0.0007402      0.001202         -0.0007389
rand
```

```
                              LEVEL 3

  PARAMETER                ESTIMATE      S. ERROR      PREV. ESTIMATE
NCONV

                              LEVEL 2

  PARAMETER                ESTIMATE      S. ERROR      PREV. ESTIMATE
NCONV
CONS     /CONS             0.000982     0.0002236            0.000981
1

                              LEVEL 1

  PARAMETER                ESTIMATE      S. ERROR      PREV. ESTIMATE
NCONV
CONS     /CONS            0.0006335    0.00005393           0.0006336
2
sett
EXPLanatory variables in       CONS     ACTMC
FPARameters                    CONS     ACTMC
FMEAns
RMEAns
RESPonse variable in           RISKMC
IDENtifying codes for level 1: CASE     level 2: BEAT-NO  level 3:
RESEtting covariances level 1: ON       level 2: ON       level 3: ON
MAXIterations   5     TOLErance   2     METHod is IGLS    BATCh is ON


LEVEL 3 RANDOM PARAMETER MATRIX unspecified
LEVEL 2 RANDOM PARAMETER MATRIX
         CONS
CONS     1
LEVEL 1 RANDOM PARAMETER MATRIX
         CONS
CONS     1
expl 'actov'
resp 'riskov'
sett
EXPLanatory variables in       CONS     ACTMC     ACTOV
FPARameters                    CONS     ACTMC     ACTOV
FMEAns
RMEAns
RESPonse variable in           RISKOV
IDENtifying codes for level 1: CASE     level 2: BEAT-NO  level 3:
RESEtting covariances level 1: ON       level 2: ON       level 3: ON
MAXIterations   5     TOLErance   2     METHod is IGLS    BATCh is ON


LEVEL 3 RANDOM PARAMETER MATRIX unspecified
LEVEL 2 RANDOM PARAMETER MATRIX
         CONS
CONS     1
LEVEL 1 RANDOM PARAMETER MATRIX
         CONS
CONS     1
```

```
expl 'actmc'
sett
EXPLanatory variables in       CONS     ACTOV
FPARameters                    CONS     ACTOV
FMEAns
RMEAns
RESPonse variable in           RISKOV
IDENtifying codes for level 1: CASE     level 2: BEAT-NO  level 3:
RESEtting covariances level 1: ON       level 2: ON       level 3: ON
MAXIterations   5     TOLErance   2     METHod is IGLS    BATCh is ON


LEVEL 3 RANDOM PARAMETER MATRIX unspecified
LEVEL 2 RANDOM PARAMETER MATRIX
         CONS
CONS     1
LEVEL 1 RANDOM PARAMETER MATRIX
         CONS
CONS     1
start


Iteration number 1 in progress
Iteration number 1 completed

Iteration number 2 in progress
Iteration number 2 completed

Iteration number 3 in progress
Iteration number 3 completed

Iteration number 4 in progress
Iteration number 4 completed
Convergence achieved
fixe


  PARAMETER              ESTIMATE     S. ERROR     PREV. ESTIMATE
CONS                       0.0551     0.005076             0.0551
ACTOV                  -0.0006442     0.001351          -0.000644
rand

                              LEVEL 3

  PARAMETER              ESTIMATE     S. ERROR     PREV. ESTIMATE
NCONV

                              LEVEL 2

  PARAMETER              ESTIMATE     S. ERROR     PREV. ESTIMATE
NCONV
CONS      /CONS          0.001092    0.0002457           0.001092
2

                              LEVEL 1
```

```
  PARAMETER             ESTIMATE     S. ERROR     PREV. ESTIMATE
NCONV
CONS     /CONS         0.0006004   0.00005111         0.0006004
2
sett
EXPLanatory variables in       CONS     ACTOV
FPARameters                    CONS     ACTOV
FMEAns
RMEAns
RESPonse variable in           RISKOV
IDENtifying codes for level 1: CASE     level 2: BEAT-NO  level 3:
RESEtting covariances level 1: ON       level 2: ON       level 3: ON
MAXIterations   5     TOLErance   2     METHod is IGLS    BATCh is ON


LEVEL 3 RANDOM PARAMETER MATRIX unspecified
LEVEL 2 RANDOM PARAMETER MATRIX
         CONS
CONS     1
LEVEL 1 RANDOM PARAMETER MATRIX
         CONS
CONS     1

Note do the similar routine as above for the overlay method (vs
transformed risk)

resp 'risktov'
sett
EXPLanatory variables in       CONS     ACTOV
FPARameters                    CONS     ACTOV
FMEAns
RMEAns
RESPonse variable in           RISKTOV
IDENtifying codes for level 1: CASE     level 2: BEAT-NO  level 3:
RESEtting covariances level 1: ON       level 2: ON       level 3: ON
MAXIterations   5     TOLErance   2     METHod is IGLS    BATCh is ON


LEVEL 3 RANDOM PARAMETER MATRIX unspecified
LEVEL 2 RANDOM PARAMETER MATRIX
         CONS
CONS     1
LEVEL 1 RANDOM PARAMETER MATRIX
         CONS
CONS     1
start


Iteration number 1 in progress
Iteration number 1 completed

Iteration number 2 in progress
Iteration number 2 completed

Iteration number 3 in progress
Iteration number 3 completed
Convergence achieved
```

```
fixe

  PARAMETER                ESTIMATE     S. ERROR     PREV. ESTIMATE
CONS                         0.4839      0.01187             0.4839
ACTOV                     -0.003682     0.003178          -0.003662
rand

                                LEVEL 3

  PARAMETER                ESTIMATE     S. ERROR     PREV. ESTIMATE
NCONV

                                LEVEL 2

  PARAMETER                ESTIMATE     S. ERROR     PREV. ESTIMATE
NCONV
CONS      /CONS            0.005961     0.001333           0.005912
1

                                LEVEL 1

  PARAMETER                ESTIMATE     S. ERROR     PREV. ESTIMATE
NCONV
CONS      /CONS            0.003324    0.0002834           0.003329
1
sett
EXPLanatory variables in        CONS      ACTOV
FPARameters                     CONS      ACTOV
FMEAns
RMEAns
RESPonse variable in            RISKTOV
IDENtifying codes for level 1: CASE      level 2: BEAT-NO  level 3:
RESEtting covariances level 1: ON        level 2: ON       level 3: ON
MAXIterations   5     TOLErance   2      METHod is IGLS    BATCh is ON


LEVEL 3 RANDOM PARAMETER MATRIX unspecified
LEVEL 2 RANDOM PARAMETER MATRIX
         CONS
CONS     1
LEVEL 1 RANDOM PARAMETER MATRIX
         CONS
CONS     1

Note save the model for the future use

save mlc99.ws
247092 spaces left on worksheet

Note logoff (toggle command) and type stop to exit ML3
logo
```

Appendix 11.2

Log file of significance testing in ML3

This appendix provides a print out of the ML3 log file for the significance testing of the burglary risk and action intensity in Coventry as an example. The results of the multi-level modelling from Bristol and the two cities combined were tested similarly. Comments are after the key word 'NOTE'.

```
ML3 - Software for three-level analysis.  Tue Apr 20 14:05:54 1999

Note retrieve the saved model

retr mlc99.ws

note look at the setting of the model
sett
EXPLanatory variables in      CONS    ACTMC
FPARameters                   CONS    ACTMC
FMEAns
RMEAns
RESPonse variable in          RISKTMC
IDENtifying codes for level 1: CASE   level 2: BEAT-NO  level 3:
RESEtting covariances level 1: ON     level 2: ON      level 3: ON
MAXIterations  5    TOLErance  2    METHod is IGLS    BATCh is ON


LEVEL 3 RANDOM PARAMETER MATRIX unspecified
LEVEL 2 RANDOM PARAMETER MATRIX
      CONS
CONS    1
LEVEL 1 RANDOM PARAMETER MATRIX
      CONS
CONS    1

Note re-run the model
start


Iteration number 1 in progress
Iteration number 1 completed

Iteration number 2 in progress
Iteration number 2 completed

Iteration number 3 in progress
Iteration number 3 completed

Iteration number 4 in progress
Iteration number 4 completed
```

Convergence achieved
fixe

```
 PARAMETER        ESTIMATE    S. ERROR    PREV. ESTIMATE
CONS             0.4911     0.01185        0.4911
ACTMC            -0.00365   0.002817       -0.003649
rand

                 LEVEL 3

 PARAMETER        ESTIMATE    S. ERROR    PREV. ESTIMATE    NCONV

                 LEVEL 2

 PARAMETER        ESTIMATE    S. ERROR    PREV. ESTIMATE    NCONV
CONS    /CONS     0.005921    0.001338       0.005919       1

                 LEVEL 1

 PARAMETER        ESTIMATE    S. ERROR    PREV. ESTIMATE    NCONV
CONS    /CONS     0.003465    0.000295       0.003465       2
```

Note compute the likelihood of the model (the fit statistics are calculated as -2[log (likelihood ratio)].

likelihood

-2*log(lh) is   -792.508

Note look at the setting again
sett

```
EXPLanatory variables in     CONS    ACTMC
FPARameters                  CONS    ACTMC
FMEAns
RMEAns
RESPonse variable in         RISKTMC
IDENtifying codes for level 1: CASE    level 2: BEAT-NO  level 3:
RESEtting covariances level 1: ON      level 2: ON       level 3: ON
MAXIterations  5    TOLErance  2    METHod is IGLS    BATCh is ON


LEVEL 3 RANDOM PARAMETER MATRIX unspecified
LEVEL 2 RANDOM PARAMETER MATRIX
      CONS
CONS    1
LEVEL 1 RANDOM PARAMETER MATRIX
      CONS
CONS    1
```

Note remove the explanary variable - action score (Monte Carlo method) - from the model
expl 'actmc'
sett

```
EXPLanatory variables in     CONS
FPARameters                  CONS
FMEAns
```

```
RMEAns
RESPonse variable in        RISKTMC
IDENtifying codes for level 1: CASE    level 2: BEAT-NO  level 3:
RESEtting covariances level 1: ON     level 2: ON     level 3: ON
MAXIterations  5    TOLErance  2    METHod is IGLS   BATCh is ON


LEVEL 3 RANDOM PARAMETER MATRIX unspecified
LEVEL 2 RANDOM PARAMETER MATRIX
      CONS
CONS    1
LEVEL 1 RANDOM PARAMETER MATRIX
      CONS
CONS    1
start

Iteration number 1 in progress
Iteration number 1 completed

Iteration number 2 in progress
Iteration number 2 completed
Convergence achieved
fixe


 PARAMETER        ESTIMATE    S. ERROR    PREV. ESTIMATE
CONS            0.4899    0.01156        0.4899
rand

                LEVEL 3

 PARAMETER        ESTIMATE    S. ERROR    PREV. ESTIMATE    NCONV

                LEVEL 2

 PARAMETER        ESTIMATE    S. ERROR    PREV. ESTIMATE    NCONV
CONS    /CONS      0.005644    0.001282       0.005644       1

                LEVEL 1

 PARAMETER        ESTIMATE    S. ERROR    PREV. ESTIMATE    NCONV
CONS    /CONS       0.00351   0.0002988        0.00351       1

Note compute the likelihood with the action variable removed
likelihood

-2*log(lh) is    -790.895

Note compare the probability of the above two models, the difference is 1.613 with 1 degree of freedom, so
compute the probability (a chi-square distribution)
cprob 1.613 with 1 df

 0.20407

logoff
```

Appendix 12.1
Geographical Information Charter Standard Statement
Draft adoption by the Home Office

- *1. Consult users when preparing applications or drafting legislation for the collection of data*

Where appropriate, the Home Office will consult its main customers and data users, usually in the development phase of the data.

- *2. Provide information about what data are available*

The Home Office produces a variety of publications to provide customers with a broad understanding of the work and services. Usually, data are documented in Technical Reports. Other methods of increasing awareness will be investigated and adopted where appropriate.

- *3. Provide clear statements on the price of data;*

At present most of the data sets are available free of charge for appropriate organisations (such as research institutions) on request. This will be subject to review in the future, and it may vary across different agencies. It is envisaged that pricing of data will also vary depending on the proposed use intended by the purchaser and will be negotiated on an individual basis.

- *4. Make data available, unless there are specific reasons for not doing so, in which case those reasons should be explained;*

The Home Office will seek to make data and statistics readily available unless there are reasons to withhold them. Specific reasons will be explained if data are withheld (for example, due to confidentiality constraints, commercial consideration etc.). Some data sets would be too sensitive to release (in their entirety) outside the Home Office although aggregate data will be available through publications etc.

- *5. Ensure that data adhere to British, European and international standards and classifications, unless there are specific reasons for not doing so, in which case those reasons should be explained;*

At present there are no British, or international standards which relate to spatial data held by the Home Office. However the Home Office would assess emerging standards and where appropriate will consider adoption of a relevant standard when it becomes established. It will review its adoption of these standards when they are published and clearly explain its position to users in the event of any not being adopted.

- *6. Deliver data in standard digital, or other, formats wherever possible;*

The Home Office produces statistics in forms convenient to users, and aims to adopt common data exchange standards and meet their needs wherever possible.

- *7. Supply accompanying documentation with data to enable users to judge the fitness for their purpose;*

Home Office produces a variety of publications to provide customers with a broad understanding of the work and services. Usually, data are documented in Technical Reports to enable users to judge the fitness for their purpose.

- *8. Consult users before the destruction of any data set, subject to the prior guidance of the appropriate national record office;*

The Home Office has no plans to destroy any of the data sets. However, the Home Office would consult users before the destruction of any data sets. Certain data sets, such as the British Crime Survey, are also deposited in the ESRC data archive, which is available for secondary analysis to bona fide academic researchers.

- *9. Publish a contact point to deal with enquiries;*

The Home Office operates a central enquiry desk at each office to answer telephone or written enquires. The Home Office's published reports include the Information & Publication Group contact point for further enquiries.

- *10. Provide information about how users can complain if they are not satisfied with the service they receive.*

The Home Office publishes a contact point for details of its complaint procedures.

Bibliography

American Society of Photogrammetry (Committee for Specifications and Standards, Professional Practice Division, 1980, 1985) Accuracy specification for large scale line maps. *Photogrammetric Engineering & Remote Sensing*. Vol. **51**, pp. 195-199.

Arbia, G., Griffith, D. & Haining, R. (1998) Error propagation in raster GIS: overlay operation. *International Journal of Geographical Information Science.* Vol. **12**. No. 2, pp. 145-167.

Aronoff, S. (1982) Classification accuracy: a user approach. *Photogrammetric Engineering & Remote Sensing*. Vol. **48**. No. 8 pp. 1299-1307.

Aronoff, S. (1985) The minimum accuracy as an index of classification accuracy. *Photogrammetric Engineering & Remote Sensing.* Vol. **51**. No. 1, pp. 1687-1694.

Aronoff, S. (1989) *Geographical Information Systems: A Mangement Perspective.* Canada: WDL.

Aronson, P. (1987) Attribute handling for Geographical Information Systems. *Proceedings of the Eighth International Symposium on Computer-Assisted Cartography*, pp. 246-355.

Atkinson, A. C. (1980) Test of pseudo-random numbers. *Applied Statistics.* Vol. **29**. No. 2, pp. 164-171.

Barnett, L. & Carlis, J. C. (1993) Feature information support and the SDTS conceptual data model: clarification and extension. In *Proceedings of the AUTOCARTO 11*. Minneopolis, pp. 132-144.

Beavon, D., Brantingham, P. & Brantingham, P. (1994) The influences of street network on the patterning of property offences. In R. V. Clark (1994 Eds) *Crime Prevention Studies*. Vol. **II**. Monsey, New York: Criminal Justice Press.

Bennett, T. & Wright, R. (1984) *Burglars on Burglary: Prevention and the Offender*. Hampshire, England: Bower.

Bernstein,R. (1983 Eds) Image geometry and rectification. In R. N. Colwell, (Ed.) *Manual of Remote Sensing*. 2nd edition, pp. 873-922. Falls Church, VA: American Society of Photogrammetry.

Berry, J. K. (1987) Fundamental operations in computer-assisted map analysis. *International Journal of Geographical Information Systems*. Vol. **1**. No. 2, pp. 119-136.

Berry, B. J. L. & Baker, A. M. (1968) Geographical Sampling. In B. J. L. Berry & D. F. Marble (Eds) *Spatial analysis: a Reader. Statistical Geography*, pp. 91-100. Englewood Cliffs, NJ: Prentice Hall

Besag J & Diggle, P. J. (1977) Simple Monte Carlo tests for spatial patterns. *Applied Statistics*. Vol. **26**, p. 327.

Blakemore, M. (1984) Generalization and error in spatial databases. *Cartographica*. Vol. **21**, pp. 131-139.

Blumstein, A., Cohen, J., Rosenfeld, R. (1991) Trend and deviation in crime rates: a comparison of UCR and NCVS data for burglary and Robbery. *Criminology*. Vol. **29**, pp. 237-248.

Bolstad, P. V., Gessler, P. & Lillesand,. T. M. (1990) Positional uncertainty in maually digitised maps. *International Journal of Geographical Information Systems*. Vol. **4**, No. 4, 399-412.

Bowes, K. & Hirschfield (1999) Exploring links between crime and disadvantage in north-west England: an analysis using geographical information systems. *International Journal of Geographical Information Science*. Vol. **13**. No. 2, pp. 159-184.

Bracken, I. (1994) A Surface model approach to the representation of population-related social indicators. In S. Fotheringham & P. Rogerson (Eds) *Spatial Analysis and GIS*, pp. 247-260. London: Taylor & Francis.

Bracken, I. & Martin, D. (1989) The generation of spatial population distributions from census centroid data. *Environment & Planning A*. Vol. **21**, pp. 537-543.

Bracken, I. & Webster, C. (1989)Toward a typology of Geographical Information Systems. *International Journal of GIS*. Vol. **3**. No. 2, pp. 137-152.

Braithwaite, J. (1979) *Inequality, Crime and Public Policy*. Routlege & Kegan Paul: London.

Brantingham, P. & Brantingham, P. (1984) *Patterns in Crime*. New York: MacMillian.

Brantingham, P. J. & Brantingham, P. L. (1991) *Environmental Criminology*. Waveland, III.

Brassel, K. Bucher, F. Stephan, E. & Vckovski, A. (1995) Completeness. In S. C. Guptill & J. L. Morrison (1995 Eds) *Elements of Spatial Data Quality*, pp. 81-108. Oxford: Elsevier Science.

Brown, A. & van Elzakker, C. P. J. M. (1993) Kleurgebruik bij de weergave van de kwaliteit van nominale gebeidsinformatie. *Kartografisch Tijdschroft* **XIX** (4) pp. 59-73.

Brunsdon, C. (1989) *Spatial Analysis Techniques Applied to the Local Crime Pattern Analysis*, PhD thesis. Newcastle University.

Brunsdon C. (1995) Developing an exploratory spatial analysis system in Xlisp-Stat. Paper presented at *GISRUK'95*, Newcastle University.

Brunsdon, C. (1995) Estimating probability surfaces for geographical point data: an adaptive kernel algorithm. *Computer & Geoscience.* Vol. **21**, No. 7, pp. 877-894.

Brunsdon, C. (1998) Exploring spatial data and local indicators of spatial association with XLIS-STAT. *The Statistician.* Vol. **47** (3), pp. 471-484.

Brunsdon, C., Fotheringham, S. & Charlton, M. (1998) Geographically weighted regression - modelling spatial non-stationarity. *The Statistician.* Vol. **47** (3), pp. 431-443.

Brunsdon, C. & Openshaw, S. (1993) Simulating the effect of error in GIS. In P. M. Mather (Ed) *Geographical Information Handling – Research and Applications*, pp. 48-61. Chichester: Wiley.

Bryk, A. S. & Raudenbush, S. (1992) *Hierarchical Linear Models - Applications and Data Analysis.* Newbury Park, CA: Sage.

Burgess, T. M. & Webster, R. (1980a) Optimal interpolation and isarithmic mapping I. The semivariogram and punctual kriging. *Journal of Soil Science.* Vol. **31**, p. 315.

Burgess, T. M. & Webster, R. (1980b) Optimal interpolation and isarithmic mapping II. Block kriging. *Journal of Soil Science.* Vol. **31**, p. 333.

Burrough, P. A. (1986) *Principles of Geographical Information Systems for Land Resources Assessment.* Oxford: Clarendon Press.

Burrough, P. A. & Heuvelink, G. (1992) The sensitivity of Boolean and continuous logical modelling to uncertainty data. *Proceedings of EGIS'92*, pp. 1032-1039.

Bursik, R. Jr. (1988) Social disorganization and theories of crime and delinquency, problem and prospects. *Criminology*. Vol. **26**: pp. 529-51.

Buttenfield, B. P. (1993) Representing data quality. *Cartographica*. Vol. **30**, pp. 1-7.

Campbell, J. B. (1987) *Introduction to Remote Sensing*. New York: Guilford.

Carter, J. R. (1989) On defining the geographical Information System. In W. J. Ripple (Ed) *Fundamentals of Geographical Systems: A Compendium*, pp. 3-7. Falls Church, Virginia: ASPRS/ACSM.

Caspary, W. & Scheuring, R. (1992) Error-band as measures of geometrical accuracy. *Proceedings of EGIS'92*, pp. 226-233.

Chrisman, N. R. (1982) A theory of cartographic error and its measurement in digital database. *Proceedings of Auto-Carto 5*.

Chrisman, N. R. (1984) The role of quality information in the long term functioning of a Geographical Information System. *Cartographica*. Vol. **21**. No. 2 & 3.

Chrisman, N. R. (1986) Alternative for specifying quality standards for digital cartographic data . In H. Moellering (Ed) *National Committee for Cartographic Data Standards* (NCCDS). Columbus, Ohio: ACSM.

Chrisman, N. R., (1989) Modelling error in overlaid categorical maps. In Goodchild, M. F. & Gopal, S. (Eds) *Accuracy of Spatial Databases*, pp. 21-34. London: Taylor & Francis.

Chrisman, N. R. & Niemann, B. (1985) Alternative routes to a mutipurpose cadastre: merging institutional and technical reasoning. *Preceedings of the Seventh International Symposium on Automated Cartography*, pp. 84-93.

Chrisman, N. R. & Yandell, B. (1988) A model for the variance in area. *Surveying & Mapping*. Vol. **48**, pp. 241-246.

Clarke, G. C. & Clark, D. M. (1995) Lineage. In S. C. Guptill & J. L. Morrison (1995 Eds) *Elements of Spatial Data Quality*, pp. 13-30. Oxford: Elsevier Science.

Clark, R. V. (1983) Situational crime prevention: its theoretical basis and practical scope. In M. Tonry & N. Morris (1994 Eds) *Crime and Justice: An Annual Review of Research.* Vol. **4**, pp. 225-256. Chicago: University of Chicago Press.

Colwell, R. N. (1983) *Manual of Remote Sensing*. Second Edition. Falls church, Virgina: American Society of Photogrammetry.

Congalton, R. G. (1988) A comparison of sampling schemes used in generating error matrices for assessing the accuracy of maps generated from remotely sensed data. *Photogrammetric Engineering & Remote Sensing*. Vol. **54**. No. 5, pp. 593-600.

Congalton, R. G. (1991) A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment*. Vol. **37**, pp. 35-46.

Congalton, R. G. & Mead, R. A. (1983) A quantitiative method to test for consistency and correctness in photointerpretation. *Photogrammetric Engineering & Remote Sensing*. Vol. **49**, pp. 69-74.

Congalton, R. G., Oderwald, R. G. & Mead, R. A. (1983) Assessing Landsat classification accuracy using discrete multivariate analysis statistical techniques. *Photogrammetric Engineering & Remote Sensing*. Vol. **49**, pp. 1671-1678.

Cochran, W.G. (1954) Some methods for strengthening the common X^2 test. *Biometrics*. Vol. **10**. No. 4, pp. 417-451.

Cohen, L. E. & Felson, M. (1979) Social change and crime rate trends: a routine activity approach. *American Sociological Review*. Vol. **44**, pp. 588-608.

Cressie, N. A. C. (1991) *Statistics for Spatial Data*. New York: Wiley.

Cressie, N. A. C. (1993) *Statistics for Spatial Data* (revised). New York: Wiley.

Corbett, J. P. (1979) *Topological Principles in Cartography*. Technical paper – U. S. Bureau of the Census, p. 48.

Craig, J. (1985). *A 1981 Socio-Economic Classification of Local and Health Authorities of Great Britain*. OPCS Studies of Medical and Population Subject 48. London: HMSO.

Dagpunar, J. (1988) *Principles of Random Variate Generation*. Oxford: Clarendon Press.

Davis, J. C. (1986) *Statistics and Data Analysis in Geology*. Second Edition. New York: John Wiley & Sons.

Date, C. J. (1985) *An Introduction to Database Systems*. Vol. **II**. Reading, MA: Addison-Wesley.

de Man, E. (1988) Establishing a Geographical Information System in relation to its use: a process of strategic choice. *International Journal of Geographical Information Systems*. Vol. **2**, pp. 245 - 161.

Devery, C. (1992) *Mapping Crime in Local Government Areas: Assault and Break and Enter in Waverley*. NSW Bureau of Crime Statistics and Research.

Department of Environment (1995) 1991 *Deprivation Index: A Review of Approaches and a Matrix of Results*. London: HMSO.

Drummond, J. E. (1995) Positional accuracy. In S. C. Guptill & J. L. Morrison (1995 Eds) *Elements of Spatial Data Quality*, pp. 31-58. Oxford: Elsevier Science.

Drummond, J. E. & Ramlal, B. (1992) An uncertainty sub-system applied to a Dutch land reallocation project. *EGIS Proceedings*. Munich.

Bryant, J. (1979) On the clustering of multidimensional pictorial data. *Pattern Recognition*. Vol. **11**, pp. 115-125.

Dutton, G. (1992) Handling positional uncertainty via hierarchical tessellation. In M. F. Goodchild & S. Gopal, (1989 Eds) *The Accuracy of Spatial Databases*. London: Taylor & Francis.

Ehlers, M., Edwards, G. & Bedard, Y. (1989) Integration of remote sensing with Geographical Information Systems: a necessary evolution. *Photogrammetric Engineering & Remote Sensing*. Vol. **55**. No. 11, pp. 1619-1627.

Egenhofer, M. J. & Frank, A. U. (1992) Object-oriented modeling for GIS. *URISA Journal*. Vol. **4**, pp. 3-19.

Ekblom, P. (1990) Evaluation crime prevention: the management of uncertainty. In Kemp, C. (Ed) *Current issues in Criminological Research*. Bristol: Bristol Centre for Criminal Justice.

Ekblom, P. (1992) The Safer Cities Programme impact evaluation: problems and progress. *Studies on Crime and Crime Prevention*, 1, pp. 35-51. Scandinavian University Press.

Ekblom, P. (1994) *Proximal Circumstances: A Theory-Based Classification of Crime Prevention.* Crime Prevention Studies, 2. London: Home Office.

Ekblom, P. (1999) Can we make crime prevention adaptive by learning from other evolutionary struggles? *Studies on Crime & Crime Prevention*, pp. 27-51. Scandinavian University Press.

Ekblom, P. Howes, D. & Law, H. C. (1994) Scoping, scoring and modelling: linking measures of crime preventive action to measures of outcome in a large, multi-site. evaluation using GIS and multilevel modelling. *Proceedings of the GISRUK* 1994, pp. 123-132.

Ekblom, P., Law, H. C. & Sutton, M. (1996a) *Safer Cities and Residential Burglary.* Home Office Research Study 164. London: HMSO.

Ekblom, P., Law, H. C. & Sutton, M. (1996b) *Domestic burglary schemes in the Safer Cities Programme.* Home Office RSD Research Findings No. 42, London: Home Office.

Ekblom, P., Sutton, M & Wiggins, R. (1993) *Scoping, scoring and modelling: linking measures of crime preventive action to measures of outcome in a large, multi-site evaluation.* Paper presented to *Royal Statistical Society*, London.

Ekblom, P. & Pease, K. (1995) Evaluating crime prevention. In M. Tonry & D. Farrington (Eds.) Building a Safer Society: strategic approach to crime prevention. *Crime & Justice: A review of research.* Vol. **19**, pp. 585-662. London & Chicago: University of Chicago Press.

Englund, E. (1993) Spatial simulation: environmental applications. In M. F. Goodchild, B. O. Parks & L. T. Steyaert. (Ed) *Environmental Modeling with GIS.* New York: Oxford University Press.

Evan, D. J. (1989) Geographical analyses of residential burglary. In D. J. Evans & D. T. Herbert (Ed) *The Geography of Crime*. London: Routledge.

Everest, G. (1986) *Database Management - Objectives, System Functions and Administration.* New York: McGraw-Hill.

Fan, J. Farmen, M. & Gijbels, I. (1998) Local likelihood estimation and interference. *Journal of the Royal Statistical Society, Series B. Statistical Methodology.* Vol. **60**. Part 3, pp. 591-608.

Farrington, D. P., Sampson, R. J. & Wikstrom, P-O. (1993 Eds) *Integrating Individual and Ecological Aspects of Crime.* Stockholm: National Council for Crime Prevention.

Field, F. (1989) *Losing Out: The Emergence of Britain's Underclass.* Oxford: Basil Blackwell.

Field, S. (1990) *Trends in Crime and their Interpretation: A study in Post-war England and Wales*. Home Office Research Study 119. London: HMSO.

Fisher, P. F. (1991a) Modelling soil map-unit inclusions by Monte Carlo simulation. *International Journal of Geographical Information Systems*. Vol. **5**, pp. 193-208.

Fisher, P. F. (1991b) First Experiments in Viewshed Uncertainty: the accuracy of the viewable area. *Photogrammetric Engineering & Remote Sensing*. Vol. **57**, pp. 1321-1327.

Fisher, P. F. (1991c) Data sources and data problems. In J. Maguire, M. Goodchild & D. Rhind (Eds) *Geographical Information Systems: Principles and Applications.* Vol. **2**, pp. 217-231. Essex: Longman.

Fisher, P. F. (1992) First Experiments in Viewshed Uncertainty: Simulating the Fuzzy Viewshed. *Photogrammetric Engineering & Remote Sensing*. Vol. **58**, pp. 345-352.

Fisher, P., Dykes, J. & Wood, J. (1993) Map design & visualization. *The Cartographic Journal.* Vol. **30**. Dec.

Fisher, P.F. & Langford, M. (1995) Modelling the errors in areal interpolation between zonal systems by Monte Carlo simulation. *Environmental & Planning Section A.* Vol. **27**, pp. 211-224.

Flowerdew, R. (1988) *Statistical Method for Areal Interpolation: Predicting Count Data from a Binary Variable*. RR-15. NRRL. University of Lancaster & University of New Castle Upon Tyne.

Flowerdew, R. & Green, M. (1989) *Statistical methods for inference between incompatible zonal systems.* In M. F. Goodchild & S. Gopal (Eds) *The Accuracy of Spatial Databases*, pp. 239-247. London: Taylor & Francis.

Flowerdew, R. & Green, M. (1991) Data integration: statistical methods for transferring data between zonal systems. In I. Masser & M. Blakemore (Eds) *Handling Geographical Information: Methodology & Potential Applications*, pp. 38-54. London: Longman.

Flowerdew, Green, M. & Kehris, E. (1991) Using areal interpolation methods in Geographical Information Systems. *Regional Science*. Vol. **70**, pp. 303-315.

Flowerdew, R. & Openshaw, S. (1987) *A Review of the Problems of Transferring Data from One Set of Areal Units to Another Incompatible Set.* Research Report. NRRL, Lancaster & Newcastle, England.

Fotheringham, A. S. (1989) Scale-independent spatial analysis. in the accuracy of spatial databases. In M. F. Goodchild & S. Gopal (Eds) *The Accuracy of Spatial Databases*, pp. 221-228. London: Taylor & Francis.

Fotheringham, A. S. & Rogerson, P. A. (1993) GIS and spatial analytical problems. *International Journal of Geographical Information Systems*. Vol. **7**, pp. 3-19.

Fotheringham, A. S. & Wong, D. W. S. (1991) The modifiable areal unit problem in multivariate statistical analysis. *Environment & Planning A*. **23**, pp. 1025-1044.

Fukunaga, K. & Hayes, R. R. (1989) Effects of sample size in classifier design. *Transactions on Pattern Analysis and Machine Intelligence*. IEEE. Vol. PAMI-**11**. No. 8.

Giordano, A. Veregin, H. Borak, E. & Lanter, D. (1994) A conceptual model of GIS-based spatial analysis. *Unpublished manuscript*. Department of Geography, Kent State University, Kent, Ohio.

Ginevan, M. E. (1979) Testing land-use map accuracy: another look. *Photogrammetric Engineering & Remote Sensing*. Vol. **45**. No. 10, pp. 1371-1377.

Goel, S. P. (1992) *Building a national digital database through digitization of 1:250,000 maps: an evaluation of alternatives*. M.Sc. thesis. International Institute of Aerospace Survey and Earth Science (ITC). P. O. Box 6, 7500 AA Enschede. The Netherlands.

Goldstein, H. (1990) *Problem-Oriented Policing*. New York: McGraw-Hill.

Goldstein, H. (1995) *Multilevel Statistical Models*. London: Edward Arnold.

Goldstein, H., Rasbash, J. Plewis, I., Draper, D., Browne, W., Yang, M., Woodhouse, G. & Healy, M. (1998) *A User's Guide to Mlwin*. Multilevel Models Project, Institute of Education, University of London.

Goodchild, M. F. (1989) Modeling error in object and fields. In M. F. Goodchild & S. Gopal (Ed) *The Accuracy of Spatial Databases*, pp. 107-114. London: Taylor & Francis.

Goodchild, M. F. (1990) Spatial Information Science. *Proceedings of the Spatial Data Handling Symposium*, pp. 3-14. International Geographical Union: Ohio.

Goodchild, M. F. (1991) Spatial Information Science. *Proceedings of the Spatial Data Handling Symposium*, pp. 3-4. International Geographical Union: Ohio.

Goodchild, M. (1993) Data models and data quality: problems and prospects. In M. Goodchild, B. Parks & L. Steyaert (Ed) *Environmental Modelling with GIS*, pp. 94-103.New York: Oxford University Press.

Goodchild, M. F. (1995) Attribute accuracy. In S. C. Guptill & J. L. Morrison (1995 Eds) *Elements of Spatial Data Quality*, pp. 31-58. Oxford: Elsevier Science.

Goodchild, M. F. Anselin, L. & Deichmann, U. (1993) A framework for the areal interpolation of socioeconomic data. *Environment & Planning A*. Vol. **25**, pp. 383-397.

Goodchild, M. F. & Gopal, S. (1989 Eds) *The Accuracy of Spatial Databases*. London: Taylor & Francis.

Goodchild, M. F. & Lam, N. S-N. (1980) Areal interpolation: a variant of the tradition spatial problem. *Geoprocessing*. Vol. **1**, pp. 297-312.

Goodchild, M. F. Sun, G. O. & Yang, S. (1992) Development and test of an error model for cartographical data. *International Journal of Geographical Information Systems*. Vol. **6**. No. 2, pp. 87-104.

Gould, P. (1967) On the geographical interpretation of eigenvalues. *Transactions. Institute of British Geographers*. Vol. **42**, pp. 53-86.

Greenland, A., Socher, R. M. & Thompson, M. R. (1985) Statistical evaluation of accuracy for digital cartographic database. *Proceedings of Auto-Carto 7*. Washington: ASP-ACSM.

Grady, R. B. (1993) Practical results from measuring software quality. *Communications of the ACM*. Vol. **36**, pp. 62-68.

Grove, P. (1998) Three talks in Statistics. *Unpublished manuscripts*. London: Home Office.

Groves R. (1989) *Survey Errors and Survey Costs*. Chichester: Wiley.

GSS (1997) *Report of the Task Force on Non Sampling Error* (October). London: Government Statistical Service Method Committee.

Guptill, S. C. (1990) An enhanced digital line graph design. *U.S. Geographical Survey Circular 1048*. Reston, VA: U.S. Geographical Survey.

Guptill, S. C. (1994) Synchronization of discrete geospatial databases. *Proceedings of the 6th International Conference on Spatiasl Data Handling*. Edinburgh, Vol. **2**, pp. 945-956.

Guptill, S. C. (1995) Temporal information. In S. C. Guptill & J. L. Morrison (1995 Eds) *Elements of Spatial Data Quality*, pp. 153-165. Oxford: Elsevier Science.

Guptill, S. C. & Morrison, J. L. (1995 Eds) *Elements of Spatial Data Quality*. Oxford: Elsevier Science.

Guptill, S. C. & Morrison, J. L. (1995) Looking ahead. In S. C. Guptill & J. L. Morrison (1995 Eds) *Elements of Spatial Data Quality*, pp. 189-202. Oxford: Elsevier Science.

Guptill, S. C. & Stonebraker, M. (1992) The Sequoia 2000 approach to managing large spatial object databases. *Proceedings of the 5th Internation Spatial Dasta Handling Symposium*. Charleston, SC, Vol. **2**, pp. 642-651.

Hagerstrand, T. (1965) A Monte Carlo approach to diffusion. *European Journal of Sociology*. Vol. **6**, pp. 43-67.

Hammersley, J. M. & Handscomb, D. C. (1964) *Monte Carlo Methods*. London: Chapman & Hall.

Heuvelink, G. B. M., Burrough, P. & Stein, A. (1989) Propagation of error in spatial modelling with GIS. *International Journal of Geographical Information Systems*. Vol. **3**. No. 4, pp. 303-322.

Hirschfield, A. (1994) Using the 1991 population Census to study deprivation. *Planning Practice and Research*. Vol. **9**. No. 1, pp. 43-54.

Hirschfield, A. & Bowers, K. (1997) The development of a social, demographic and land use profiler for areas of high crime. *British Journal Criminology*. Vol. **37**. No. 1.

Hirschfield, A. Bowers, K. & Brown, P. (1995a) Exploring relations between crime and disadvantage on Merseyside. *European Journal on Criminal Policy and Research*. Vol. **3**. No. 3, pp. 93-112.

Hirschfield, A. Brown, P. & Todd, P. (1995b) GIS and the analysis of spatially-referenced crime data: experiences in Merseyside, UK. *International Journal of Geographical Information Science*. Vol. **9**. No. 2, pp. 191-210.

Holzman, H. R. & Piper, L. (1998) Measuring crime in public housing: methodological issues and research strategies. *Journal of Quantitative Criminology*. Vol. **4**. No. 4, pp. 331-352.

Home Office (1998) *Home Office Annual Report 1998*, Cm 3908, ISBN: 0101390823. London: The Stationery Office.

Honeycutt, D. M. (1986) Epsilon bands based on probability. *Unpublished Manuscript*.

Hope, A. C. A. (1968) A simplified Monte Carlo significance test procedure. *Journal of the Royal Statistical Society, Series B*. Vol. **30**, pp. 582-598.

Hord, R. M. & Brooner, W. (1976) Land use map accuracy criteria. *Photogrammetric engineering & Remote Sensing*. Vol. **42**. No. 5.

Houghton, G. R. & Berry, G. (1989) *Microcomputers - An Aid to Crime Analysis*. SC/84 111/169/1. London: Home Office SRDB.

Hudson, W. D. & Ramm, C. W. (1987) Correct formulation of the Kappa coefficient of agreement. *Photogrammetric Engineering & Remote Sensing*. Vol. **53**. No. 4, pp. 421-422.

ISO 8402 (1986) *Quality – vocabulary*. International Standard Organization.

Jennings, A. A. & Mohan, S. (1991) Testing random number generators for microcomputer applications of Monte Carlo simulation. *Environmental Software*, Vol. **6**. No. 4, pp.176-193.

Johnson, S. D., Bowers, K. & Hirschfield, A. (1997) New Insights into the spatial and temporal distribution of repeat victimisation. *British Journal of Criminology*.

Jones, K. (1992) Multi-level modeling. In A. Westlake, R. Banks, C. Payne & T. Orchard (Ed) *Survey and Statistical Computing*. New York: North-Holland.

Journel, A. G. & Huijbregts, Ch. J. (1978) *Mining Geostatistics*. New York: Academic Press.

Junger-Tas, J. (1993) Policy evaluation research in criminal justice. *Studies on Crime and Crime Prevention*. Vol. **2**, pp. 7-20. Stockholm, Sweden: National Council for Crime Prevention.

Kainz, W. (1995) Logical consistency. In S. C. Guptill & J. L. Morrison (1995 Eds) *Elements of Spatial Data Quality*, pp. 109-137. Oxford: Elsevier Science.

Keefer, B. J., Smith, J. L. & Gregoire, T. G. (1991) Modelling and evaluating the effects of stream digitising errors on map variables. *Photogrammetric Engineering and Remote Sensing*. Vol. **57**. No. 7, pp. 957-63.

Kennedy, S. (1989) Small number problem and the accuracy of spatial databases. In M. F. Goodchild & S. Gopal (Eds) *The Accuracy of Spatial Databases*. London: Taylor & Francis.

Kennedy-Smith, G. M. (1986) Data quality - a management philosophy. *Preceeedings of Autocarto 1986*. London.

Kershaw, C. D. (1987) Discrimination problems for satellite images. *International Journal of Remote Sensing* Vol. **8**. No. 9, pp. 1377-1383.

Kershaw, C.D. & Fuller, R M. (1992). Evaluation of Land Use Discrimination Procedures for Satellite Imagery in Lowland Britain: Statistical Considerations and Problems. *International Journal of Remote Sensing* Vol. **13**. No. 16, pp. 3085-3104.

Khorev, A. G. Govorov, M. O. & Kasianova, E. L. (1996) Representation on multi-detailed data in object-oriented Gist. In M. Rumor, R. McMillam & H. F. L. Ottens (Eds) *From Research to Application through Cooperation, the Second Joint European Conference & Exhibition on Geographical Information.* Vol. **1**, pp. 226-229. Oxford: IOS.

Kjerne, D. & Dueker, K. J. (1986) Modeling cadatral relationships using an object-oriented language. *Proceedings of the Second International Symposium on Spatial Data Handling*, pp. 142-157.

Knuth, D. E. (1969) *The Art of computer programming.* Second Addition, Vol **2**. Massachusetts: Addison-Wesley.

Kottman, C. A. & Buttenfield, B. P. (1994) Standards for spatial data use: similes improve our understanding. *Cartography and Geographic Information Systems.* Vol. **21**, pp. 141-144.

Kovalick, W. M. (1983) *The Effects of Selected Preprocessing Procedures Upon the Accuracy of a Landsat - Dervised Classification of a Forested Wetland.* M.S. thesis. Blacksburg: Virginia Polytechnic Institute.

Kuczera, G. (1988) On validity of first order prediction limits for conceptual hydrological models. *Journal of Hydrology.* Vol. **103**, pp. 229-247.

Kurtz, E. Koon, B. & Taylor, R. B. (1995) Impacts of nonresidential landuse and physical deterioration on resident-based control and calls for police service. Paper presented at *the Annual Meeting of the Academy of Criminal Justice Sciences.* Boston.

Lam, N. S-N. (1983) Spatial interpolation interpolation methods: a review. *American Cartographer.* Vol. **10**, pp. 129-149.

Langbein, L. I. & Lichtman, A. J. (1978) *Ecological Inference*. Beverly Hills: Sage.

Langford M, Maguire, D. J. & Unwin, D. J. (1991) The areal interpolation problem: estimating population using remote sensing in a GIS framework. In I. Masser & Blakemore (Eds) *Handling Geographical Information: Methodology & Potential Application*, pp. 55-77. London: Longman.

Langford M., Fisher, P. F. & Troughear, D. (1993) Comparative accuracy measurements of the cross areal interpolation of population in *Proceedings of EGIS'93*, pp. 663-674. EGIS Foundation: Utrecht.

Lanter, D. P. (1991) *Lineage in GIS: The problem and a Solution*. Technical paper 90-6. Santa Barbara, California: National Center for Geographical Information and Analysis.

Lanter, D. P. & Veregin, H. (1992) A research paradigm for propagating error in layer-based GIS. *Photogrammetric Engineering & Remote Sensing*. Vol. **58**, pp. 825-833.

Law, H. C. (1998) Error: the neglected area, but does it matter? *The AGI Conference Proceedings at GIS'98 – profiting from collaboration.* Chapter 8.1. London: AGI & Miller Freeman.

Law, H. C. & Ekblom, P. (1994a) Application of GIS: from knowledge to data representation & data storage. *ARC/INFO European User Conference'94*. Paris.

Law, H. C. & Ekblom, P. (1994b) Application of GIS: Evaluation of Safer Programme - from knowledge to data representation and transformation. *AGI'94. Conference Proceedings*. Chapter 17.2, pp. 1-12. London: AGI.

Law, H. C. & Ekblom, P. (1996) Describing the nature of Safer Cities Action using GIS. In M. Rumor, R. McMillam & H. F. L. Ottens (eds) *Geographical Information - From Research to Application.* Second Joint European & Exhibition on Geographical Information. Proceedings, Vol. **2**, pp. 1007-1016. Oxford: IOS Press.

Law, H. C. & Fisher, P. F. (1995) Error modelling in areal interpolation: a case study on the data precision of the Safer Cities Evaluation. *Proceedings of the GIS Research UK.* Paper presented at GISRUK'95, Leicester University.

Laycock, G. K. & Tilley, N. (1995) Implementing crime prevention. In M. Tonry & D. Farrington (Eds.) Building a Safer Society: Strategic approach to crime prevention. *Crime & Justice: A review of research.* Vol. **19**, pp. 535-584. London & Chicago: University of Chicago Press.

Lee, J., Snyder, P. K. & Fisher, P. F. (1992) Modeling the effect of data errors on feature extraction from digital elevation models. *Photogrammetric Engineering & Remote Sensing*. Vol. **57**, pp. 1321-1327.

Liddle, M. & Bottoms, A. E. (1992) *Implementing Circular 8/84: A retrospective assessment of the 5 Towns Initiate*. Unpublished report to Home Office, available from Institute of Criminology, Cambridge.

Lodwick, W. A., Monson, W. & Svoboda, L. (1990) Attribute error and sensitivity analysis of map operations in geographical informations systems: suitability analysis. *International Journal of Geographical Information Systems.* Vol. **4**. No. 4, pp. 413-428.

Lunetta, R. S. Congalton, R. G., Fenstermarker, L. K., Jensen, J. R., McGwire, K. C. & Tinney, L. R. (1991) Remote sensing and geographical information data integration: error sources and research issues. *Photogrammetric Engineering & Remote Sensing.* Vol. **57**. No. 6, pp. 677-687.

MacLaren, M.D. & Marsaglia, G. (1965) Uniform random number generators. *Journal of the Association for Computing Machinery.* Vol. **12**. No.1, pp. 83-89.

MacLeod, I. S. & Law, H. C. (1998) System Cognitive Functions: The Next Step. Paper presented at the Second InternationalConference of Engineering Psychology and Cognitive Ergonomics, Oxford. In D. Harris (Ed in press) *Engineering Psychology and Cognitive Ergonomics.* Aldershot: Ashgate.

Maffini, G. Arno, M. & Bitterlich, W. (1989) Observations and comments on the generation and treatment error in digital GIS data. In Goodchild, M. F. & Gopal, S. (Eds) *The Accuracy of Spatial Databases*, pp. 55-67. London: Taylor & Francis.

Maguire, D. J. (1991) An overview and definition of GIS. In D. J. Maguire, M. F. Goodchild & D. W. Rhind (Eds) *Geographical Information Systems: Principles and Applications.* Vol. **1**, pp. 9-20. Essex: Longman.

Maguire D. (1994) Object oriented GIS. Paper presented at the 9th ARC/INFO European User Conference. Paris.

Maltz, M. D. (1988) Visualizing Homicide: a research note. *Journal of Quantitative Criminology.* Vol. **4**. No. 4, pp. 397- 410.

Maltz, M. D., Gordon, A. C. & Friedman, W. (1991) *Mapping Crime in its Community Setting - Event Geography Analysis.* Springer-Verlag.

Mark, D. & Csillag, F. (1989) The nature of boundaries on 'area-class' map. *Cartographica.* Vol. **26**. No. 1, pp. 65-77.

Martin, D. (1998) Optimizing census geography: the separartion of collection and output geographies. *International Journal of Geographical Information Science.* Vol. **12**. No. 7, pp. 673-686.

Martin, J. & McClure, C. (1985) *Diagramming Techniques for Analysts and Programmers*. Englewood Cliffs, Inc.: Prentice Hall.

Maslow, A. H. (1954) *Motivation and Personality*. New York: Harper & Row.

Martin, D. (1989) Mapping population data from zone centroid locations. *Transactions of the Institute of British Geographers NS*. Vol. **14**, pp. 90-97.

Mayhew, P., Maung, N. A. & Mirlees-Black, C. (1993) *The 1992 British Crime Survey*. Home Office Research Study 132. London: HMSO.

McAlpine, J. R. & Cook, B. G. (1971) Data reliability from map overlay. *Proceedings of Australian & New Zealand Association for Advancement of Science, 43rd Congress.*

McCarthy, J. (1960) Recursive Functions of symbolic expression and their computation by machine: Part 1. *Communication of the ACM*. Vol. **3**. No. 4, pp. 184-195.

McCarthy, J. (1963) A basis for a mathematical theory of computation. In P. Bradford & D. Hirshber (Ed) *Computer Programming and Formal Systems*. Amsterdam: North Holland.

McDowall, D. & Laffin, C. (1992) Comparing the UCR and NCVS over time. *Criminology*. Vol. **30**, pp. 125-133.

McHarg, L. L. (1969) *Design with Nature*. New York: Doubleday.

Medyckyj-Scott D. & Hearnshaw H. M. (1993 Eds) *Human Factors in Geographical Information Systems*. London & Florida: Belhaven Press.

Meixler, D. & Saalfeld, A. (1987) Polygonization and topological editing at the bureau of the census. In N. R. Chrisman (Ed) *Proceedings Auto-Carto 8*, pp. 731-738. American Society for Photogrammetry and Remote Sensing and American Congress on Surveying and Mapping.

MIDAS (1995) *MIDAS CSS 618.* Manchester Information Datasets and Associated Services. Manchester University.

Mikhail, E. M. & Ackerman, F. (1976) *Observations and Least Square.* Dun-Donnelley, New York: IEPA.

Moellering, H. (1984a) *Digital cartographic data standards*, Columbus: Ohio.

Moellering, H. (1992 Eds) *Spatial Database Transfer Standards: Current International Status*. Oxford: Elsevier.

Mortimore, P., Sammons, P., Stoll, L., Lewis, D. & Ecob, R. (1988) *School Matters.* Open books.

Monmonier, M. & Schnell, G. (1984) Land use & land cover data the maping density. *International Yearbook of Cartography,* pp. 115-121.

Morrison, J. L. (1988) The proposed stanard for digital cartographic data. *The American Cartographer*. Vol. **15**, pp. 129-135.

Morrison, J. L. (1992) Implementing the spatial data transfer standard - introduction. *Cartography and Geographic Information Systems*. Vol. **19**. p. 277.

NCDCDS (1988) National Committee on Digital Cartographic Data Standards. In Brassel, K. Bucher, F. Stephan, E. & Vckovski, A. (1995) Completeness. In S. C. Guptill & J. L. Morrison (1995 Eds) *Elements of spatial data quality*, pp. 81-108. Oxford: Elsevier Science.

NIST, (1994). In Drummond, J. (1995) Positional accuracy. In S. C. Guptill & J. L. Morrison (1995 Eds) *Elements of spatial data quality*, pp. 31-58. Oxford: Elsevier Science.

Nutall, D., Goldstein, H., Prosser, R. & Rasbash, J. (1989) Differential school effectiveness. *International Journal of Education Research.* Vol. **13**, pp. 769-776.

ONS (1999) *Census 2000 - Fill in your Future, Consultation* (Spring). London: Office for National Statistics.

Openshaw, S. (1977) Algorithm 3: a procedure to generate pseudo-random aggregations of N zones into M zones, where M is less than N. *Environmental & Planning, Series A*, pp. 169-184.

Openshaw, S. (1984) *The Modifiable Areal Unit Problem.* CATMOG 38. Norwich: Geobooks.

Openshaw, S. (1989) Learning to live with errors in spatial databases. In M. F. Goodchild & S. Gopal, (Eds) *The Accuracy of Spatial Databases*, London: Taylor & Francis.

Openshaw, S. (1991) Developing appreciate spatial analysis methods for GIS. In D. J. Maguire, M. F. Goodchild & D. W. Rhind (Ed) *Geographical Information Systems: Principles and Applications.* Essex: Longman.

Openshaw, S. (1993) GIS crime and GIS criminality. *Environmental Planning, Series A.* Vol. **26**, pp. 451-458.

Openshaw S. & Rao, L. (1994) *Re-engineering 1991 census geography: serial and parallel algorithms for unconstrained zone design.* A507265011 report. ESRC Census Initiative Award.

Openshaw S. & Rao, L. (1995) Algorithms for re-engineering 1991 census geography. *Environmental and Planning, Series A*. Vol. **27**, pp. 425-446.

Openshaw, S. & Taylor, P. J. (1979) A million or so correlation coefficients in three experiment on the modifiable area unit problem. In N. Wrigely (Ed) *Statistical Application in the Spatial Science*, pp. 127-144. London: Pion.

Openshaw, S. Charlton, M., Wymer, C. & Craft, A. (1987) A mark 1 geographical analysis machine for the automated analysis of point data sets. *International Journal of Geographical Information Systems*, pp. 335-358.

Openshaw, S., Chartlton, M. & Craver, S. (1991) Error propagation: a Monte Carlo simulation. In I. Masser & M. Blackemore (Eds) *Handling Geographical Information: Methodology and Potential Applications*. Essex: Longman.

Park, R. E. (1936) Human ecology. *American Journal of Sociology*. Vol.**42**, pp. 1-15.

Parkes, D. N. & Thrift, N. J. (1980). *Times, Spaces & Places: A Chronogeographic Perspective*. New York: John Wiley.

Perkal, J. (1956) On Epsilon Length. *Bulletin de l'Academie Polonaise des Sciences*. Vol. **4**, pp. 399-403.

Perkal, J. (1966) *On the Length of Empirical Curves*. Discussion Paper 10. Ann Arbor MI: Michigan Inter-University Community of Mathematical Cartographers.

Petrick, K. (1980) Concepts of accuracy methods and results of determination. In *Proceedings of the second European Seminar of the Technical Committee on Standardization in Quality Control*, pp. 189. Paris.

Polder, W. (1992) Crime prevention in the Netherlands: Pilot projects evaluated. *Dutch Penal Law and Policy* 7. The Hague, Netherlands: Research and Documentation Centre.

Ralphs, M. P. (1993) In Brassel, K. Bucher, F. Stephan, E. & Vckovski, A. (1995) Completeness. In S. C. Guptill & J. L. Morrison (1995 Eds) *Elements of spatial data quality*, pp. 81-108. Oxford: Elsevier Science.

Rao, C. R. (1965) *Linear Statistical Inference and its Applications.* New York: Wiley.

Rodcay, G. (1990) Tucson plants with GIS. *GIS World.* Vol. **3**, pp. 51-53.

Roundtree, P., Land, K. C. & Mirthe, T. (1994) Macro-micro integration in the study of victimization: A hierarchical logistic model analysis across Seattle neighbourhoods. *Criminology.* Vol. **32**. No. 3, pp. 387-414.

Rosenfield, G. H. (1982) Sample design for estimating change in land use and land cover. *Photogrammetric Engineering & Remote Sensing.* Vol. **48**. No. 5, pp. 793-801.

Rosenfield, G. H. (1986) Analysis of thematic map classification error matrices. *Photogrammetric Engineering & Remote Sensing.* Vol. **52**. No. 5, pp. 681-686.

Rosenfield, G. H. & Fitzpatrick-Lins, K. (1986) A coefficient of agreement as a measure of thematic classification accuracy. *Photogrammetric Engineering and Remote Sensing.* Vol. **52**, pp. 223-227.

Rossmo, D. K. & Fisher, D. K. (1993) *Problem-Oriented Policing: A Cooperative Approach in Mount Pleasant.* Vancouver, Burnaby, BC.

Ratcliffe, J. & McCullagh, M. (1998) Aoristic crime analysis. *International Journal of Geographical Information Science.* Vol. **12**. No. 7, pp. 751-764.

Rhind, D. W. (1911) Counting the people: the role of GIS. In D. J. Maguire, M. F. Goodchild & D. W. Rhind (Eds) *Geographical Information Systems: Principles and Applications.* Volume 2, pp. 127-137. Essex: Longman.

Rich, T. F. (1995) The use of computerized mapping in crime control and prevention programs. *National Institute of Justice - Research in Action.* U.S. Department of Justice, National Institute of Justice.

Salgé, F. (1995) Semantic accuracy. In S. C. Guptill & J. L. Morrison (1995 Eds) *Elements of Spatial Data Quality*, pp. 139-151. Oxford: Elsevier Science.

Sampson, R. J. & Groves, W. B. (1989) Community structure and crime: testing socio-disorganization theory. *American Journal of Sociology.* Vol. **94**, pp. 774-802.

Scholz, D., Fuhs, D. & Hixson, M. (1979) An evaluation of several different classification schemes, their parameters & performance. *Proceedings of The Thirteenth International Symposium on Remote Sensing of the Environment*, pp. 1143-1149. Ann Arbor: University of Michigan.

Sechrist, R. P. (1992) Simulation of the spatial diffusion process. *Computers & Geosciences.* Vol. **18**, pp. 965-974.

Shaw, C. & McKay, H. (1972) *Juvenile Delinquency and Urban Areas.* Chicago: University of Chicago Press.

Shepherd, I. D. H. (1991) Information integration and GIS. In D. J. Maguire, M. F. Goodchild & Rhind, D. W. (Ed) *Geograohical Information Systems: Principles and Applications*, pp. 337-359. Essex: Longman.

Shi, W. (1994) *Modelling Positional and Thematic Uncertainties in Integration of Remote Sensing and Geographical Information Systems.* Publication 22. Enschede, Netherlands: ITC.

Sowey, E.R. (1972) A chronology and classified bibliography on random number generation and testing. *International Statistical Review.* Vol. **40**. No. 3, pp355-371.

Sowey, E.R. (1978) A second classified bibliography on random number generation and testing. *International Statistical Review*. Vol. **46**, pp. 89-102.

Sowey, E.R. (1986) A third classified bibliography on random number generation and testing. *Royal Statistical Society*. Vol. **149.** No. 1, pp. 83-107.

SDTS (1992) *Spatial data transfer standard* (SDTS). U. S. Department of Commerce.

Spelman, W. G. & Eck, J. E. (1987) *Problem-Oriented Policing*. National Institute of Justice Publication. Washington, DC: U. S. Government Printing Office.

Star, J. L. Estes, J. E. & Davis, F. (1991) *Improved Integration of Remote Sensing and Geographical Information Systems: A background to NCGIA*. No. **6**, pp. 643-645.

Steele, Jr. G. L., Fahlman, S. E., Gabriel, R. P., Moon, D. A. & Weinreb, D. L. (1984) *Common Lisp*. Bedford, Massachusetts: Digital Press.

Story, M. & Congalton, R. G. (1986) Accuracy assessment: a user's perspective. *Photogrammetric Engineering & Remote Sensing*. Vol. **52.** No. 3, pp. 397-399.

Sträng, D. (1996) *Response Times for Interior Fire-Fighting In Residential Housing*. Swedish Rescue Services Agency, Räddnings Verket.

Stutheit, J. (1990 Eds) *The 1990 GIS Source book*. GIS World. Colorado, USA: Fort Collins.

Sutton, M. (1996) *Implementing Crime Prevention Schemes in a Multi-Agency Setting: Aspects of Process in the Safer Cities Programme*. Home Office Research Study 160. London: HMSO.

Taha, H. A. (1982) *Operations Research - An Introduction*. Third Edition. Review of Probability Theory, pp. 382-385. New York: MacMillan.

Tempfli, K. (1980) Spectral analysis of terrain relief for the accuracy estimation of digital terrain models. *ITC Journal,* Vol. **3**, pp. 478-510.

Tauussky, O. & Todd, J. (1956) Generation and testing of pseudo-random numbers. In Ed. H. A. Meyer (Ed) *Symposium on Monte Carlo methods*, pp. 15-28. New York: Wiley.

Taylor, D.R.F. (1995) Forward in S. C. Guptill & J. L. Morrison (Ed) *Elements of Spatial Data Quality*, p. ix. Oxford: Elsevier Science.

Taylor, R. B. Koons, B. A. Kurtz, E. M., Greene, J. R. & Perkins, D. D. (1995) Street blocks with more nonresidential land use have more physical deterioration: evidence from Baltimore and Philadelphia. Research Note. *Urban Affairs Review.* Vol. **31**, pp. 120-136.

Tierney, L. (1990) *Lisp-Stat an Object-Oriented Environment for Statistical Computing and Dynamic Graphics.* New York: John Wiley & Sons.

Tilley, N. (1992) *Safer Cities and Community Safety Strategies.* Crime Prevention Unit Paper 38. London: Home Office.

Tilley, N. (1993a) *Understanding Car Parks, Crime and CCTV: Evaluation Lessons from Safer Cities.* Home Office Crime Prevention Unit Paper 42. London: Home Office.

Tilley, N. (1993b) Crime prevention and the Safer Cities story. *Howard Journal of Criminal Justice.* Vol. **32**, pp. 32-57.

Tilley, N. & Webb, J. (1994) *Burglary Reduction: Findings from Safer Cities Schemes.* Police Reserach Group, Crime Prevention Series, 60. London: Home Office.

Tobler, W. (1979) Smooth pycnophylactic interpolation for geographical regions. *Journal of the American Statistical Association*. Vol. **74**, pp. 519-530.

Tomlin, C. D. (1991) Cartographic Modelling. In D. J. Maguire, M. F. Goodchild & D. W. Rhind (Ed) *Geographical Information Systems: Principles and Applications*. Essex: Longman.

Tomlin, C. D. & Berry, J. K. (1979) A mathematical structure for cartographic modelling in environmental analysis. *Proceedings of the American Congress on Surveying and Mapping*, pp. 269-283.

Van der Wel, F. J. M. & Hootsman, R. M. (1993) Visualisaties van kwaliteits-informatie in een GIS-omgeving. *Kartografisch Tijdschrift*. Vol. **XIX**. No. 4, pp. 49-58.

Van Elzakker, C. P. J. M., Ramlal, B. & Drummond, J. E. (1992) The visualisation of GIS generated information quality. ISPRS XVII International Congress Washington DC. USA Aug. 1992. *International Archives of Photogrammetry and Remote Sensing Part B4. Commission*. Vol. **IV**, pp. 608-615.

Van Genderen, J. L. & Lock, B. F. (1977) Testing land-use map accuracy. *Photogrammetric Engineering & Remote Sensing*. Vol. **43**, pp. 1135-1137.

Veregin, H. (1989) Error modelling for the map overlay operation. In M. F. Goodchild & S. Gopal (Eds) *The Accuracy of Spatial Databases*. London: Taylor & Francis.

Veregin, H. & Hargitai, P. (1995) An evaluation matrix for geographical data quality. In S. C. Guptill & J. L. Morrison (1995 Eds) *Elements of Spatial Data Quality*, pp. 167-188. Oxford: Elsevier Science.

Walker, A. & Walker, C. (1989 Eds) *The Growing Divide: A Social Audit 1979-1987*. London: CPAG.

Walsby, J. C. (1995) The cause and effect of manual digitizing on error creation in data input to GIS. In P. Fisher (Ed) *Innovations in GIS 2*. London: Taylor & Francis.

Webster, R. (1985) Quantitative spatial analysis of soil in the field. *Advances in Soil Science*. Vol. **3**. New York: Spring-Verlag.

White, M. S. (1984) Technical requirements and standards for a multipurpose geographic data system. *The American Cartographer*. Vol. **11**, pp. 15-26.

Wilson, W. J. (1987) *The Truly Disadvantaged*. University of Chicago Press.

Winston, P. H. & Horn, B. K. P. (1984) *Lisp*. Second Edition. London: Addison-Wesley.

Woodhouse, G., Rasbash, J. Goldstein, H. & Yang, M. (1992) *A guide to ML3 for New Users*. Institute of Education, University of London.

Worboys, M. F. (1994a). Object-oriented approaches to geo-referenced information. *International Journal of Geographical Information Systems*. Vol. **8**, pp. 385-399.

Wright, J. K. (1936) A method of mapping densities of population with Cape Cod as an example. *Geographical Review*. Vol. **26**, pp. 103-110.

Youell, J. (1993) *Assessment and Monitoring of Safer Cities Schemes: A Guidance Manual for Project Staff*. London: Home Office.

Zhang, G. Y. & Tulip, J. (1990) An algorithm for the avoidance of sliver polygons and clusters of points in spatial overlay. *Proceedings of 4th Spatial Data Handling*, pp. 141-150.