Developing a Combined Risk Model for the Prediction of Temporally Clustered Offences

Emily Jane Sheard

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

The University of Leeds
School of Geography

April, 2020

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

Acknowledgements

I would like to thank my supervisors, Prof Mark Birkin, Prof Nick Malleson, and Dr Daniel Birks, for their input and support throughout the course of my research, and also for giving me the opportunity to pursue my own ideas and interests. Thank you also to my Research Support Group, Prof Alison Heppenstall, Prof Lex Comber, and Prof Alex Hirschfield, for their constructive feedback and debate, and especially to Alex who took time out of his retirement to support me. I would also like to acknowledge the Leeds Institute for Data Analytics, who provided me with relevant training opportunities, the Economic and Social Research Council (ESRC) for funding my research, and West Yorkshire Police for contributing the crime data on which this thesis is based.

Thank you also to my friends and colleagues in the Consumer Data Research Centre, and especially to Dr Stephen Clark who (patiently) introduced me to the world of R, to Nick Malleson for setting me on the path to Python coding, to Eusebio for his permanently positive outlook, to Michelle for her wise words and advice, and to Verity, Robin, Eleri, and Kylie for their support and encouragement when the going got tough! Thank you also to Prof Stillwell for giving me the opportunity to join you all in Montpellier, and to Rachel for being an excellent travel buddy.

Thank you to Mum and Dad for their continued support and encouragement (and willingness to embrace all things coding), and to Joshua for the occasional brotherly ‘how’s it going?’ texts.

And finally, but by no means least, to Faisal for always believing in me – I couldn’t have done it without you.

This work was supported by the Economic and Social Research Council [grant numbers ES/J500215/1, ES/L011891/1].

Abstract

Having the means to estimate when and where future offences are likely to occur is of immense value to crime practitioners and partner agencies, hence a number of crime models have been developed to this end. Some of these models, including ProMap and SEPP, incorporate elements of a criminological theory known as 'repeat/ near-repeat victimisation'. This theory is based on the idea that offenders typically operate within the confines of residentially-anchored routine activity spaces, thus rendering them well-placed to return to previously targeted locations over time. Therefore, associated offences are likely to present as spatially-clustered, yet temporally extended crime series, which provides a window of opportunity for operational intervention. Although 'repeat/ near-repeat victimisation' theory has informed the modelling of some high-level crime recording categories, including residential burglary, this thesis presents empirical evidence that failure to disaggregate beyond official crime classifications risks neglecting heterogeneity of offence characteristics within these. A potential implication of this is that the spatio-temporal parameters on which some prevailing crime modelling techniques are based might not apply to all offences, meaning that any related decision-making could be misinformed.

Table of Contents

List of Tables

List of Figures

Chapter 1 Introduction to the Research

1.1 Introduction

This thesis will argue that failure to disaggregate crime data beyond official Home Office (HO) recording classifications runs the risk of overlooking heterogeneity of offence characteristics. Given that spatio-temporal specificity is key to the success of both problem-oriented and intelligence-led policing, if predictive crime models are calibrated only with reference to the dominant parameters of high-level summary offence categories, such as 'Burglary Dwelling', as opposed to unofficial sub-categories of these, for example, 'Car Key' burglary, then it is possible that any related decision-making, including when and where to deploy resources, could be misinformed. A key theory underpinning a number of crime modelling techniques is 'repeat/ near-repeat victimisation', which is based on the idea that offenders typically operate within the bounds of residentially-anchored routine activity spaces, thus making it relatively easy for them to return to the locations of their most recent offences. The implication of this is that associated offences are likely to be spatially-clustered within neighbourhoods, and for an extended time period, which presents an opportunity for intervention work on the part of police and partner agencies. Although repeat/ near-repeat victimisation has been used to inform the development of residential burglary models at the HO offence classification level, this thesis will challenge the prevailing narrative within the literature that 'all residential burglary events are created equal'.

1.2 Research Rationale

The overarching rationale for the current research is that there will be heterogeneity of offence characteristics within the Home Office's 'Burglary Dwelling' (residential burglary) classification, meaning that some crime modelling techniques might not be as effective for every offence that is recorded under this category. Given that conventional repeat/ near-repeat victimisation (RV/ N-RV) behaviours are analogous with Optimal Foraging Theory (e.g. see Krebs and Davies, 1987 in Johnson and Bowers, 2004, p.242), most residential burglary offenders are likely to target properties within the vicinity of their own neighbourhoods, thus generating spatially clustered, but temporally longitudinal hot spots, these being especially conducive to intervention activity. However, it is hypothesised by the author of the current work that when the target (stealable) property type for a sub-category of residential burglary is most prevalent beyond the confines of offenders' residential areas/ routine activity spaces, then the associated spatio-temporal parameters will be different to those for other types of residential burglary. This is likely to render crime modelling techniques that are based on conventional foraging behaviours, such as the heuristic geographical buffering of recent offences, less effective. The main reasons for this are that: (i) longer journey-to-crime distances are likely to be indicative of more spatially extensive criminal activity spaces, (ii) offences are likely to present as temporally clustered crime sprees because offenders will offset the additional effort required to access the target property by committing more than one offence during a single trip, and (iii) in order to mitigate the risks inherent to return visits to unfamiliar areas, offenders will probably choose a different target area on their next trip. Therefore, prospective buffers drawn around recent burglary events in non-high offender rate areas and based on the dominant spatio-temporal parameters for all Burglary Dwelling offences are unlikely to capture many, if any, related offences in the immediate future.

It was fundamental to the testing of the research rationale that a sub-category of residential burglary could be identified that was expected to have longer associated journey-to-crime (JTC) distances than most other residential burglary offences. Having reviewed the relatively small literature on the characteristics of high offender rate/ high offender count areas, it was inferred that increased offender mobility would be associated with a target property type that was unlikely to be prevalent in such areas, e.g. expensive cars. Prior to commencing the research, the author was already aware of an unofficial crime type termed 'Car Key burglary', whereby offenders commit burglaries in order to obtain the keys for vehicles that they subsequently steal. Given that West Yorkshire Police had also recently made available a spatially and temporally disaggregated residential burglary data set for the purposes of academic research, it was decided to pursue the hypothesis that '**Car Key**' burglary offences would produce a different RV/ N-RV pattern to '**Regular**' (all other residential) burglary offences (see Figure 1.1 below). If shown to be true, this could then potentially call into question the relevance of some existing crime modelling techniques to the offence type, especially where the associated results are used to inform operational police activity, as well as the appropriateness of recording Car Key burglaries under the same HO crime classification as other residential burglaries, as per the assertion of Chapman et al. (2012, p.943). Thus, the research findings could have notable implications for policy and practice.

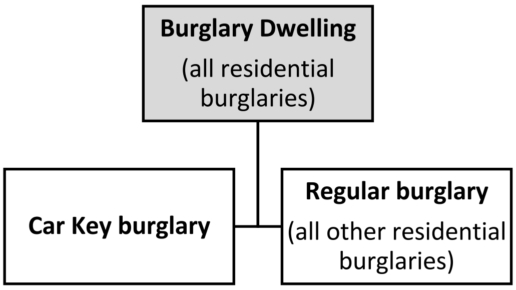

Burglary Dwelling
(all residential burglaries)
Car Key burglary
Regular burglary
(all other residential burglaries)

Figure 1.1 - Residential burglary types in the current research

Since Car Key burglars were expected to travel farther from home base to offence than Regular burglars, and also that they might work in offender groups, especially if part of an organised criminal network, it was hypothesised that they would offset the additional effort required to access the target property type by committing more than one theft during a single trip, i.e. a crime spree. Recognising the risks that are inherent to offending in unfamiliar areas, it was also anticipated that Car Key burglars would only return to the general vicinity of their most recent offences, at least in the short-term. This was as opposed to Regular burglars, who, based on the existing RV/ N-RV literature, were thought more likely to return to the exact location of a recent offence, or nearby properties, and over an extended time period, e.g. weeks. Taking all of this together, it was postulated that Car Key burglary hot spots would be far more transient in nature with subsequent offences following an initial event in quick succession – "temporally clustered".

From a practical perspective, a potential impact of highly mobile offending is that the spatial parameters on which some existing crime models are based might need be far more extensive in order to capture a large proportion of future events. For example, prospective geographical buffers would probably need to be much larger for Car Key burglary offences than for Regular burglary offences, although rational choice would still likely impose an average maximum JTC for the majority of Car Key burglars, i.e. the distance at which additional travel becomes unprofitable in terms of the time and effort required, financial costs incurred, and potential risk. In light of this, it was decided to develop a combined risk model for the prediction of temporally clustered offences, with prospective buffers placed around recent Car Key burglary offences to provide a general indication as to the likely spatial extent of offenders' criminal activities – Carden observed a median JTC distance of 4.9 km for Car Key burglars (2012, p.74). The areal extents of these 'dynamic' risk surfaces were then filtered using a 'static' risk heterogeneity surface created from relevant predictor variables, including an area type juxtaposition measure.

1.3 Aims and Objectives

1.3.1 Overall Aim

To develop a combined risk model for the prediction of temporally clustered offences.

1.3.2 Objectives

The following objectives have been set to facilitate the realisation of the overall research aim:

1. Conduct an initial review of the literature to inform the identification of a sub-category of an official Home Office crime recording classification where the spatio-temporal signature of the offences is expected to differ to that of others in the same classification.

2. Perform a detailed review of the literature to seek evidence in support of the hypothesis that the spatio-temporal signatures of Car Key burglary and Regular burglary will differ, and that prevailing crime modelling methods do not take account of this in predictions.

3. Devise a method by which to differentiate Car Key burglary offences from Regular burglary offences in a police-recorded crime data set, and then use this to create appropriate crime samples for the current research analysis, including household rates.

4. Use the findings of the detailed literature review to identify appropriate independent variables for analysis with the crime rates, e.g. area socio-demographic characteristics, and then use the most relevant of these predictors to derive the static risk surface, employing an approach that reflects the statistical attributes of the dependent variable.

5. Undertake repeat/ near-repeat victimisation analysis for the two burglary types to understand if the associated spatio-temporal patterns differ, and then use the Car Key burglary results to establish distance and time parameters for the dynamic risk surfaces.

6. Develop and evaluate dynamic and combined risk models for the prediction of temporally clustered offences, focusing on the capacity of the combined risk model to maintain previously observed hit rates whilst reducing the areal extent of identified risk.

7. Review the key findings against the original research rationale, aims and objectives, and existing crime models, identify any limitations of the methods employed, and suggest some potential areas for future work, including automation of the modelling approach.

1.4 Thesis Structure

Table 1.1 below sets out the thesis structure, including the main purpose of each chapter.

Chapter	**Chapter title**	**Main purpose**
Chapter 1	Introduction to the Research	To justify the core research rationale and to set out the overall aim and objectives.
Chapter 2	Literature Review: Applying Crime Theory to the Prediction of Residential Burglary	To present evidence that the spatio-temporal signatures of Car Key burglary and Regular burglary are likely to differ, and that prevailing crime modelling techniques do not take account of this.
Chapter 3	Data Selection and Exploratory Analysis	To provide an overview of the key data sets employed in the research, together with any inherent limitations, and to give an initial empirical indication that the research rationale is likely to be valid.
Chapter 4	Measuring the Impact of Area Accessibility on Burglary Rates	To assess the impact of offender mobility parameters on observed crime patterns, focusing specifically on the effects of area type juxtapositions and average area accessibility (street closeness centrality).
Chapter 5	Developing the Static and Dynamic Risk Surfaces	To identify the most appropriate input variables for the static risk layer and to devise a suitable method for deriving this, as well as identifying spatio-temporal parameters for the dynamic risk layer.
Chapter 6	Developing and Evaluating a Combined Risk Model	To evaluate the performance of the dynamic and combined risk surfaces based on associated daily hit rates and size of predicted risk areas, and to compare with existing crime models.
Chapter 7	Using Machine Learning in an End-to-End Burglary Model	To illustrate how text classification could be used to improve the crime samples selection and to outline a proposal for an end-to-end burglary estimation model that incorporates machine learning.
Chapter 8	Discussion and Conclusions	To review the key findings of the current research against the original aims and objectives, to identify any limitations of the methods employed, and to suggest potential areas for future development.

Table 1.1 - Thesis structure, including the main purpose of each chapter

Chapter 2 Literature Review: Applying Crime Theory to the Prediction of Residential Burglary

2.1 Introduction

Given the time and effort that will be invested in the development of the combined risk model, it is imperative to locate evidence within the related literature to support the research rationale, namely that the spatio-temporal signatures of Car Key burglary and Regular burglary differ, thus potentially reducing the efficacy of contemporary predictive crime modelling techniques. Notably, a preliminary review of the literature found no explicit reference to the possible role of repeat victimisation/ near-repeat victimisation (RV/ N-RV) in determining the spatio-temporal distribution of Car Key burglary offences, which of course renders the current research novel. However, this means that assumptions underpinning the research rationale, including that Car Key burglaries will typically present as either lone events or temporally clustered crime sprees due to desirable vehicles being more prevalent in less deprived/ low offender rate areas, will be based entirely on the triangulation of findings from existing residential burglary-related studies.

Recognising the extensive research that exists on variations in the spatial distribution of crime, generally considered to have begun in the 19th century with André-Michel Guerry and Adolphe Quetelet's macro-level analyses of French judicial[1] and population data (Quetelet, 1842; Chainey and Ratcliffe, 2005; Friendly, 2007; Wortley and Mazerolle, 2008; Andresen, 2014), the first part of this Review will be structured around the following question; 'Why here, why now?'. Rather than providing a detailed history of environmental criminology and the current "GIS school" (Chainey and Ratcliffe, 2005, p.85), the response will instead integrate relevant ideas from different points in the academic timeline to explain why crime in general, and more specifically residential burglary, usually only occurs for a limited number of "space-time loci" (Brantingham and Brantingham, 2013, p.537). If this were not the case, i.e. if crime was randomly distributed in both space and time, then there would be little value in developing a predictive crime model.

[1] Compte Ǵ eń eral de l'Administration de la Justice Criminelle en France (Friendly, 2007, p.3).

Having outlined the different behavioural and situational mechanisms that are thought to influence spatio-temporal distributions of crime, section 2.3 and section 2.4 of this chapter will describe in more detail the key attributes of the two residential burglary sub-categories on which the research is based, including a review of the prevailing characteristics of those neighbourhoods and households that are disproportionately victimised for each crime type. Regular burglary will be considered first so that it can be used as a benchmark for the Car Key burglary literature, as well as for any empirical findings that result from the current work, including those presented in Chapter 3 of the thesis; 'Data Selection and Exploratory Analysis'.

Section 2.5 of the Review will provide a summary of contemporaneous current crime modelling techniques but focusing primarily on those that have been applied to residential burglary data, including heuristic geographical buffering of recent offences (e.g. see Fielding and Jones, 2012). This will be followed by a detailed examination of two modelling approaches that combine flag account (risk heterogeneity) and boost account (event-based risk) in the associated predictions, these being 'ProMap' (Johnson et al., 2009) and 'PROVE' software (Ratcliffe et al, 2016; 2016b). The final section of this chapter will discuss how the key findings of the Literature Review can be applied to the current work.

2.2 Why Here, Why Now?

Crime has historically been studied at different geographical scales, although the longitudinal trend has been one of spatial disaggregation. Brantingham and Brantingham (1991 in Wortley and Mazerolle, 2008, p.3) define these scales as "macro" (large), "meso" (middle), and "micro" (small), a terminology that is now commonplace within the environmental criminological literature. Each term can be linked to a specific point in the evolution of spatial crime analysis, for example, Guerry mapped crime type distributions for départements in France (macro-level) (Friendly, 2007; Wortley and Mazerolle, 2008), members of the Chicago School, which was active during the early to mid-20th century, considered zones within a city (meso-level) (Chainey and Ratcliffe, 2005; Wortley and Mazerolle, 2008), and modern day environmental criminologists seek to understand why crime occurs at specific places (micro-level) (Eck and Weisburd, 1995). Macro-level analysis is useful for identifying general crime trends, for example, Guerry observed that the highest property crime rates were found in wealthy, industrialised, and more educated départements to the north of France, leading him to suppose that opportunity, as opposed to poverty, as had been previously thought, influenced the spatial distribution of offences – a finding that is still very much relevant today (Friendly, 2007; Wortley and Mazerolle, 2008). However, the problem with undertaking analyses at such a spatially aggregate level is that local heterogeneity can be masked, an issue that is intrinsic to both the Modifiable Areal Unit Problem (MAUP) and the ecological fallacy, as will be discussed in more detail in Chapter 3.

The remainder of this section will therefore consider how meso-level phenomena, together with the characteristics of 'small' places, such as streets and individual properties, can determine spatio-temporal distributions of crime. Before proceeding, it is pertinent to situate the three spatial scales – macro, meso, micro – in the context of the West Yorkshire study area, hence the inclusion of Figure 2.1 below. The Metropolitan county and local authority districts (LADs) have been assigned to the macro scale category because, given their large areal extents, any analysis at these levels is unlikely to offer many useful insights for the identification of meso/ micro risk. Moving next to the meso scale, Lower Layer Super Output Areas (LSOAs) are thought by the author to sit well here because these official geographies are expected to most accurately reflect the spatial extent of aggregate processes, such as social disorganisation (Shaw and McKay, 1942) and collective efficacy (Sampson et al., 1997), that can influence individual crime distributions. Output Areas (OAs) have also been assigned to the meso scale category, although, for very small instances of these, there might be some cross-over with the micro category. Finally, streets and individual properties have been assigned to the micro category, as per Eck and Weisburd (1995).

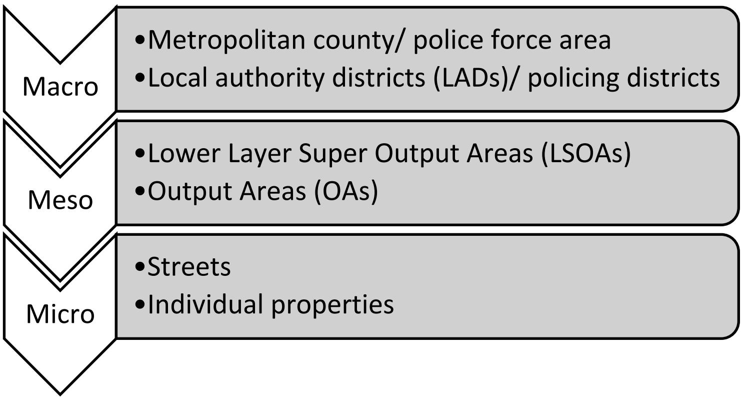

Macro
•Metropolitan county/ police force area
•Local authority districts (LADs)/ policing districts
Meso
•Lower Layer Super Output Areas (LSOAs)
•Output Areas (OAs)
Micro
•Streets
•Individual properties

Figure 2.1 - Example spatial disaggregations in the current study

Risk in the current study will be conceptualised at the meso, or neighbourhood, scale (LSOAs), however, it is important to note that area level crime rates generally result from a combination of both meso and micro factors, or, to put it another way, ecological and environmental factors. To overly simplify, in Figure 2.2 below, LSOA (A) has a high offender rate, making it particularly vulnerable to burglaries, however, only certain houses in LSOA (A) will be attractive to burglars.

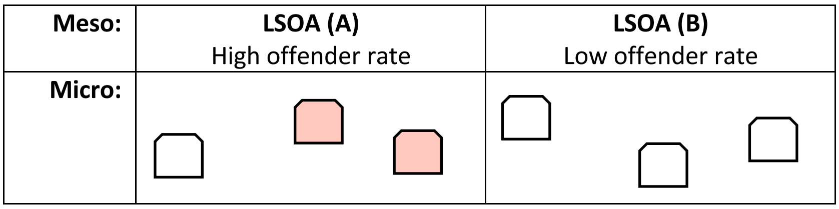


Figure 2.2 - Meso and micro scale influences on within-area crime rates

2.2.1 Meso Scale Factors

There are a number of meso scale factors that can influence the spatial distribution of crime, the majority of which can be explained with reference to the work of 20th century Chicago School social ecologists, including, amongst others, Ernest Burgess, Clifford Shaw, and Henry McKay. The socio-ecological approach considers how groups of individuals with similar characteristics inhabit particular niches within the urban environment, being subject to the same processes as are found in the natural world, including competition for land and resources (Brown, 2002). 1925 saw the publication of Burgess' zonal model in which he used concentric circles to chart the five key stages in a city's development (Andresen, 2014). Burgess noted that the 'zone in transition', adjacent to the central business district, experienced the highest population mobility, was under pressure from the invasion of business and light industry (Chainey and Ratcliffe, 2005, pp.82-83), and characterised by "juvenile delinquency, boys' gangs, crime, poverty, wife desertion, divorce, abandoned infants, [and] vice" (Burgess, 1925 in Chainey and Ratcliffe, 2005, p.83). In 1929, and building on Burgess' earlier work, Shaw and McKay published the results of their analysis into male delinquency rates for concentric zones within Chicago, finding these to be highest near the centre, where, notably, social disorganisation was also greatest (Sanders, 1943). Shaw and McKay's subsequent analysis of juvenile delinquency in other US cities uncovered remarkably similar patterns, that is, rates were highest near the centre but decreased, almost uniformly, the closer to the periphery (Sanders, 1943). The authors also observed longitudinal stability of delinquency rates, even when the ethnic makeup of areas changed, leading them to conclude that community-level mechanisms, as opposed to the traits of individuals, caused delinquency (Shaw and McKay, 1942; Sanders, 1943; Sampson and Groves, 1989; Andresen, 2014). Observing that social disorganisation theory had never been tested directly, Sampson and Groves (1989) took the three structural factors from Shaw and McKay's original model (low economic status, ethnic heterogeneity, and residential mobility), added two new structural variables, 'family disruption' and 'urbanisation', and incorporated three intervening factors; 'sparse local friendship networks', 'unsupervised teenage peer groups', and 'low organisational participation' (p.783). The authors were interested to see how the structural factors influenced the intervening factors, and how all of these influenced crime and delinquency. To explain, structural factors, such as family disruption, are not necessarily direct causes of crime themselves, rather these can affect a community's ability to control the conditions that can lead to criminality and offending, such as unsupervised teenage peer groups. Sampson and Groves identified empirical evidence in support of social disorganisation theory, for example, 27% of the total effects of family disruption on total victimisation was found to be exerted indirectly via the 'unsupervised teenage peer groups' intervening factor (1989, p.790). Closely related to social disorganisation is collective efficacy, defined by Sampson et al. (1997, p.918) as: "social cohesion among neighbours combined with their willingness to intervene on behalf of the common good". Sampson et al. (1997) analysed 343 neighbourhood clusters in Chicago to understand which of a number of 'personal-level' (e.g. socioeconomic status) and 'neighbourhood-level' (e.g. concentrated disadvantage) variables were the most strongly associated with collective efficacy. A collective efficacy measure was created based on individuals' responses to a community survey that asked questions around "informal social control" and "social cohesion and trust" (Sampson et al., 1997, pp.919-920). The authors included all of the variables in a regression model and found that high 'socioeconomic status', 'home ownership', and 'age' were linked to increased collective efficacy, whereas high mobility ("number of moves in past 5 years") was negatively associated (Sampson et al., 1997, p.921). Of the three neighbourhood-level variables, 'concentrated disadvantage' and 'immigrant concentration' were negatively associated with collective efficacy, whereas 'residential stability' was positively associated (Sampson et al., 1997, p.921). Further, the neighbourhood-level variables explained 70% of the variation in collective efficacy between neighbourhoods (Sampson et al., 1997, p.921, p.923). These findings make theoretical sense, for example, neighbourhoods experiencing constant population churn are unlikely to be conducive to feelings of collective responsibility and a vested interest in the realisation of shared goals. It is also interesting to note that, in a similar vein to Sampson and Groves (1989), Sampson et al. (1997) observed mediating effects between collective efficacy and two of their structural variables, namely 'residential stability' and 'disadvantage', and numerous measures of violence (Sampson et al., 1997, p.923). The concept of mediating variables is illustrated in Figure 2.3 below (NB some independent variable effects will directly influence the dependent variable).

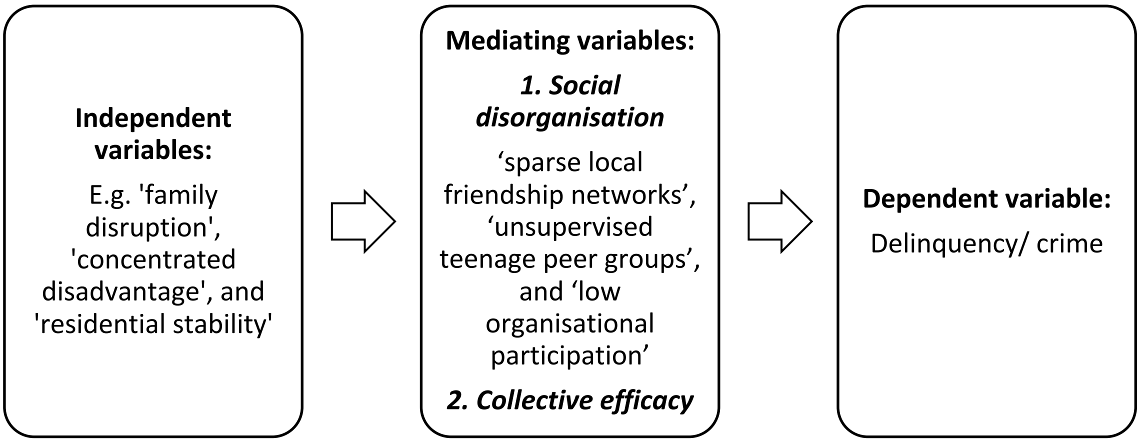

Independent variables:
E.g. 'family disruption', 'concentrated disadvantage', and 'residential stability'
Mediating variables:
1. Social disorganisation
'sparse local friendship networks', 'unsupervised teenage peer groups', and 'low organisational participation'
2. Collective efficacy
Dependent variable:
Delinquency/ crime

Figure 2.3 - Mediating community-level variables (based on and adapted from Sampson and Groves,1989, p.783).

2.2.2 Impact of Socio-Ecological Factors on the Current Work

Now to relate the findings of the Chicago School and Sampson et al. to the current research. It is expected that those LSOAs that are more socially disorganised and that also have lower levels of collective efficacy will be most vulnerable to Regular burglary because these factors are likely to facilitate crime and criminality. Conversely, given the target property type for Car Key burglary, i.e. desirable vehicles, these offences are expected to be most prevalent in more affluent/ less deprived LSOAs which, in turn, are likely to be more socially organised and have higher levels of collective efficacy. For example, Sampson et al. (1997, p.919) suggest that the financial aspect of home ownership renders collective action mutually beneficial for residents, from which we can infer that LSOAs with a high proportion of owner-occupied properties (proxy: more affluent/ less deprived areas) will be better equipped, and more willing, to act in the interests of the 'neighbourhood'. A closely related point to consider here is how community control mechanisms, such as collective efficacy, might cause Car Key burglars to modify their spatio-temporal behaviours. For example, if the target areas for Car Key burglary are socially organised, have high levels of collective efficacy, and the populations are fairly stable, then it is likely that, following an initial crime event, residents will be empowered to take collective action, such as by increasing neighbourhood surveillance. Further to the socio-ecological findings outlined in the previous section, the additional, albeit limited, literature on the characteristics of high offender rate/ high offender count areas also indicates that Car Key burglars are unlikely to reside in the same neighbourhood types as desirable vehicles. For example, Boggs (1965, p.904) identified a strong positive correlation between residential day burglary rates and burglary offender rates for 128 census tracts in St Louis, US; Bottoms and Wiles (1986, p.105) state that "if we are interested in residential areas with high offender rates in the British context we should concentrate very strongly on the privately rented and the local authority housing sectors" (acknowledging some deviations from this); and Malleson (2010, p.79) presented strong empirical evidence that more burglars were likely to live in the more deprived areas of Leeds, West Yorkshire, including a Pearson's correlation coefficient of 0.56 between IMD Score and LSOA offender count. Finally, if Car Key burglars do not typically reside in the same areas as the target property type, then they will probably have to travel farther from home base to offence than Regular burglars (to be discussed in detail in Chapter 4). Taking all of this together, Car Key burglaries are expected to be temporally clustered in nature, presenting as either lone events, or small sprees, within spatially transient target areas. This is based on the idea that offenders will mitigate the risks inherent to unfamiliar areas by avoiding recently victimised locations in the short to medium-term, and also that they might commit more than one offence during a single trip due to the additional financial costs and effort associated with multiple visits.

2.2.3 Micro Scale Factors

For a successful crime to occur there generally needs to be "convergence in space and time of likely offenders, suitable targets and the absence of capable guardians against crime" (Cohen and Felson, 1979, p.588). This quote is particularly relevant here because it can be used to explain how crime opportunity arises from a combination of both meso and micro-level factors. Considering first the "likely offenders" dimension, meso-level influences appear to locate offenders in the same areas over time (Shaw and McKay, 1942), hence these areas are likely to be particularly vulnerable to neighbourhood-based crime types, such as residential burglary. However, the "convergence in space and time" aspect must occur at a specific 'micro' location, for example, a street corner or an individual property. Similarly, an offender might be attracted to an area at the meso-level because it is known to contain lots of "suitable targets" (Sampson and Wooldredge, 1987; Brantingham and Brantingham, 1995), or because it is characterised by low levels of guardianship, e.g. student areas, whereas the decision to target a specific location is likely to be determined by micro-level characteristics; these might include, amongst other things, ease of access and egress, the presence of stealable property, or the absence of security.

2.2.4 Environmental Criminological Theories and Metatheory

The diagram in Figure 2.4 below depicts the spatial components of a metatheory within environmental criminology known as 'Crime Pattern Theory' (Andresen, 2014). Crime Pattern Theory (1993) combines aspects of three other theories, namely 'Routine Activity Theory' (1979), the 'Geometric Theory of Crime' (1981), and 'Rational Choice Theory' (1985) (Andresen, 2014, pp.29-30), to explain the conditions in which "likely offenders" might come to converge with "suitable targets" in the absence of "capable guardians" (Cohen and Felson, 1979, p.588). Taking Routine Activity Theory first, Cohen and Felson (1979, p.588) noted that crime rates in the United States increased between 1960 and 1975, despite an improvement in social and economic conditions, i.e. those things that would theoretically be expected to reduce crime. The authors attributed this contradictory trend to a post-World War II change in the structure of routine activities, defining these as being the recurrent activities that are necessary for either a population or an individual to meet their basic needs, including food, shelter, employment, education and leisure (Cohen and Felson, 1979, pp.593-594). For example, the rate of married females participating in the work force increased between 1960 and 1970 (USBC, 1975 in Cohen and Felson, 1979, p.598), which meant that more homes were likely to be left empty during the day, thus increasing the probability that an offender would come across a "suitable target" in the absence of a "capable guardian".

Considering next the Geometric Theory of Crime, this relates to the physical constructs of the spaces in which individuals enact their routine activities (Andresen, 2014). Although offenders are perhaps less likely to be engaged in employment, they will probably have other routine activities that they perform on a regular basis, such as shopping for food, visiting friends, or even meeting a drug dealer/ fence. Brantingham and Brantingham (1993) term the locations of these routine activities 'nodes', e.g. houses and shops, the routes between them 'paths', e.g. footpaths and roads, and the areas in which they occur 'routine activity spaces'. Importantly, whenever individuals – both offenders and non-offenders – enter their routine activity spaces, they will increase their knowledge of surrounding areas, a concept known in environmental criminology as 'awareness spaces' (Brantingham and Brantingham, 1993). From the offender perspective, detailed awareness spaces could facilitate such things as the identification of potential targets and potential escape routes.

To now apply Routine Activity Theory and the Geometric Theory of Crime using Figure 2.4 below, assuming that a "likely offender" lives at the centre of the grey buffer, as this offender travels along the paths and between the nodes of their routine activities they will develop an awareness space, and where this awareness space intersects with "suitable targets", but in the absence of a "capable guardian", the offender might then decide to commit an offence. Note that the convergence of an offender and a target could either be due to the offender actively searching for a victim within their awareness space, or simply because they happened upon one whilst conducting their routine activities. This idea leads nicely on to the third theory, Rational Choice.

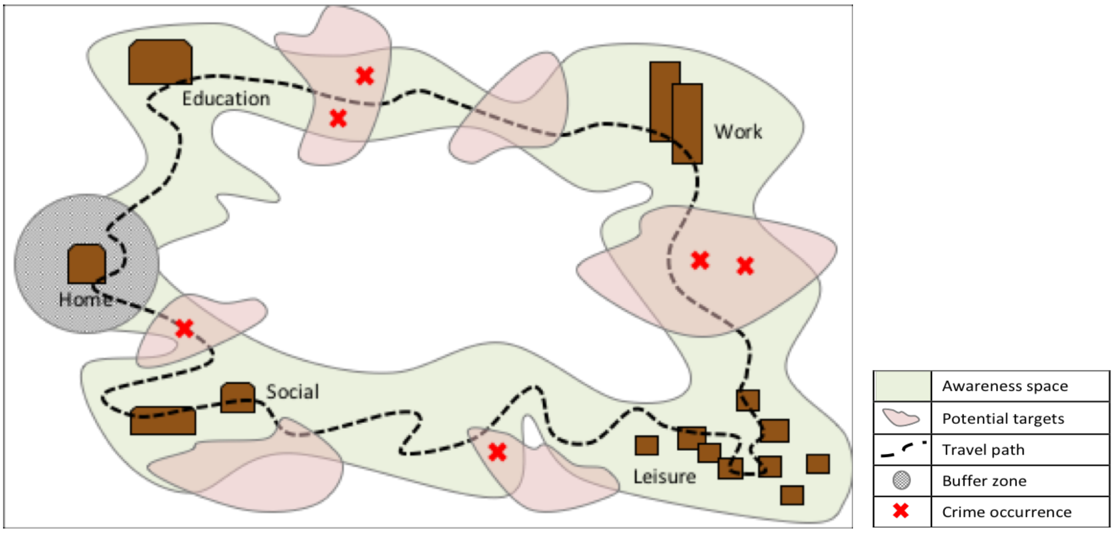


Figure 2.4 - Diagram to illustrate the spatial components of crime pattern theory (adapted from Brantingham and Brantingham, 1993, p.10)

As a concept, Rational Choice can either be involvement-centred or crime-centred (Cornish and Clarke, 2008), however, only the latter will be considered here because it pertains to the crime event itself, as opposed to the decision to engage in criminal activity generally. The theory views offenders as reasoning individuals who commit crime for a specific purpose, e.g. to acquire money for drugs, but who also weigh up the potential costs, e.g. the risk of apprehension, and benefits, e.g. financial gain, of this. For pre-planned searches especially, rational choice is likely to be evident at different spatial scales, i.e. hierarchical in nature (Bernasco, 2006). For example, Addis et al. (2019) reported the findings of interviews with incarcerated steal-to-order burglars, one of whom stated that “Would target nice areas, Harrogate, York, Leeds, Weetwood, Pudsey, all over. Drive about and see” (Addis et al., 2019, p.465). This indicates that Car Key burglars will probably select a general target area based on the assumed characteristics of the area, e.g. affluent so likely to contain more desirable vehicles, and then perform additional spatial filtering of potential targets within this, most likely at the micro-level, and considering such factors as makes and models of vehicles on driveways, presence/ absence of security, etc. Recalling Figure 2.4, even if an offender converges with a potential target as a result of their routine activities, i.e. no pre-planning at the meso-level, then rational choice is still likely to be evident in the associated decision-making process, for example, choice of MO with a view risk minimisation. It is also worth noting here that acquisitive offenders are unlikely to offend on their own doorstep, as indicated by the buffer zone in Figure 2.4, the reason for this being increased risk of detection (Brantingham and Brantingham, 1981b in Andresen, 2014, pp.53-54).

Also closely associated with Rational Choice Theory is journey-to-crime (JTC), which refers to the distance travelled between an offender's home base/ anchor-point and a potential target. Analogous with Optimal Foraging Theory (e.g. see: Krebs and Davies, 1987 in Johnson and Bowers, 2004, p.242), offender mobility patterns, with the exception of the buffer zone already mentioned, are usually subject to an inverse decay function (e.g. see: Bernasco and Luykx, 2003). This is because longer crime trips generally incur greater risks, costs, and effort for an offender. The influence of JTC on repeat/ near-repeat victimisation patterns will be discussed in the next section of this Review, and also in Chapter 4 (Measuring the Impact of Area Accessibility on Burglary Rates).

2.2.5 Repeat Victimisation and Near Repeat Victimisation

Central to the current research are the concepts of 'repeat victimisation' and 'near-repeat victimisation'. These relate to the idea that some individuals/ properties experience a greater risk of victimisation relative to others. In the context of residential burglary, repeat victimisation (RV) refers to offences committed at the same property and near-repeat victimisation (N-RV) refers to offences committed at neighbouring properties of a previously targeted property.

Research conducted over at least the last 30 years (e.g. see: Polvi et al., 1991; Johnson and Bowers, 2004) has identified that a previous residential burglary offence is a good indicator of future burglary risk, i.e. that a property is likely to be targeted again (Bernasco et al., 2015, p120). For example, Polvi et al. (1990 in Polvi et al., 1991, p.412) analysed residential burglaries in the City of Saskatoon, Canada, and found that the chance of a repeat burglary within one year of an initial event was approximately four times the expected rate for independent events. However, when burglaries within one month of an initial event were considered, the rate was found to be over twelve times that expected, but less than twice for offences committed six months apart (Polvi et al., 1990 in Polvi et al., 1991, p.412). Further to this, half of the burglaries committed within one month of an initial event occurred within seven days (Polvi et al., 1990 in Polvi et al., 1991, p.412). Of particular note, is that the authors observed clear evidence of longitudinal decay, that is, the risk of re-victimisation decreased over time. Over a decade later, Johnson and Bowers (2004) analysed residential burglaries in the county of Merseyside, UK, to understand if the spatio-temporal influence of victimisation extended beyond the immediate location of an initial event, i.e. was there evidence of communicability of risk? They used statistical techniques from epidemiology to test if the Merseyside offences were clustered in both space and time (Johnson and Bowers, 2004, pp.245-248), subsequently identifying that:

> …a residential burglary flags the risk of further residential burglaries in the near future (1-2 months) and in close proximity (up to 300-400 metres) to the victimized home.
>
> Johnson and Bowers, 2004, p.237.

These findings, referred to as the 'near repeat' phenomenon, are particularly useful in the context of crime prevention, with operational examples provided by the Trafford Burglary Model, GMP, 2010 (Fielding and Jones, 2012) and Project Optimal, Safer Leeds Partnership/ WYP, 2012 (Addis, 2013).

Possible Explanations for Repeat Victimisation and Near-Repeat Victimisation

There are two main theories as to the causal mechanisms of repeat burglary victimisation. The first, 'boost', is based on the idea that an initial burglary increases the risk of another offence, whereas the second, 'flag', relates to enduring risk heterogeneity across space (Tseloni and Pease, 2003). It should be noted, however, that there is a clear bias within the related burglary literature towards the boost explanation, which appears to be supported by offenders' accounts of their criminal decision-making, coupled with the fact that repeat burglary victimisation usually occurs quickly following an initial event (Johnson and Bowers, 2004, pp.238-239).

Boost

As explained by rational choice theory, offenders will typically seek to balance potential rewards and risks when selecting future targets, so they are likely to look to locations where they have been successful previously (Bernasco et al., 2015, p.121). Having committed one burglary at a property, an offender should be familiar with its layout, as well as being aware of any desirable items that were left behind the first time around, or which are likely to have been replaced through insurance (Johnson and Bowers, 2004, p.239; Bowers and Johnson, 2005, p.69). A second factor to consider in the context of boost repeat victimisation is the dissemination of intelligence between criminal associates, i.e. one offender informs another that they are aware of a suitable target property and that person then goes on to commit a burglary there.

Further to the above, and adding additional weight to the boost account theory, research indicates that the majority of repeat and near repeat burglary offences are committed by the same offender, particularly if the time between the initial event and the follow-up is short (Everson and Pease, 2001 in Bernasco et al., 2015, p.120). Bernasco et al. (2008 in Bernasco et al., 2015, p.123) analysed detected burglaries committed in the Greater The Hague Area (the Netherlands) and found that 95% of direct repeats committed within two weeks of an initial offence were attributable to the same offender. For near repeat offences, defined in the Netherlands case study as those committed less that 400m apart, 64% of burglaries that occurred within two weeks of an initial event were detected to the same offender.

Similarly, Bernasco et al. (2015) analysed detected residential burglary offences committed in the West Midlands police force area from January 2007 to January 2012 and identified a statistically significant relationship between recent and previous offence location choices of offenders. Using a discrete spatial crime location choice method, a number of variables were tested to assess the impact of each of these on the likelihood of a burglar targeting a particular LSOA (Lower Super Output Area). The highest positive odds (16.6) related to an offender having committed a burglary in the same LSOA within the last two days, as compared to those LSOAs not targeted by the offender over the last two years. The highest negative odds (0.58) related to the distance (km) from the offender's place of residence to the LSOA, a finding which pertains to optimal forager theory.

Flag

The flag account theory, also termed risk heterogeneity, relates to the idea that some properties experience a heightened risk of burglary victimisation, relative to other locations, and that this does not fluctuate over time, in the absence of intervention (Johnson and Bowers, 2004, p.239). This approach suggests that a visible indicator of vulnerability, such as poor security or a means of covert entry, flags a property to would-be offenders. An initial event taken in this context reflects enduring risk as opposed to the transient risk that is associated with the boost mechanism. It is reasonable to infer, therefore, that flag offences are likely to be committed by different offenders over an extended time period and with no tangible links between past and future events.

Repeat Burglary and Relative Deprivation

Although the concepts of repeat and near repeat burglary victimisation are widely acknowledged within the burglary literature, the spatio-temporal distribution of these varies according to the socio-demographic characteristics of an area. Research indicates that repeat burglary offences are more prevalent in deprived neighbourhoods, probably due to the fact that offenders typically reside in such areas and are therefore better placed to return to properties that they have targeted previously, especially if these are located within the catchment of their daily routine activities (Bowers and Johnson, 2005; Cohen and Felson, 1979; Hirschfield and Bowers, 1998 in Bowers and Johnson, 2005, pp.69; Johnson and Bowers, 2004).

Near-Repeat Burglary and Relative Affluence

Conversely, affluent areas are more susceptible to near repeat burglary offences, or 'spates' of victimisation, whereby crime events are clustered in both space and time (Johnson and Bowers, 2004, p.243). Given that less deprived areas are unlikely to fall within an offender's routine activity space, the benefits of any burglaries committed in these areas need to be accrued quickly in order to justify the additional time and effort that is required on the part of the offender. It is also likely that the concept of defensible space (Chainey and Ratcliffe, 2005, pp.94-95) plays a role in this type of offending behaviour, with return visits to non-frequented areas potentially increasing the risk of apprehension. The relationship between near repeats and affluent areas is particularly relevant given the current research topic of car key burglaries.

Pre-Victimisation and Risk Heterogeneity

Although the theories of repeat and near repeat burglary victimisation can provide useful insights for persons undertaking prospective hot spot mapping, they are less powerful when it comes to identifying pre-victimised properties that do not sit within the immediate vicinity of an ongoing crime series, but which might be targeted in the future (Johnson and Bowers, 2004). It is perhaps here, under this second point of consideration, that flag account should be viewed as useful tool for crime prediction. Although repeat and near repeat burglary dwelling offences are effectively mini crime hot spots, given that they cluster in space and time, the spatio-temporal decay that is inherent to these renders them particularly transient in nature. As such, a crime practitioner would probably want to know where and when the next initial event is likely to occur. To this end, data should be acquired which is able to unmask the spatial distribution of risk heterogeneity.

2.3 Regular Burglary as a Benchmark

2.3.1 Legal Definition of Burglary

Under Section 9(1,2) of the Theft Act 1968, a burglary is committed when a person enters a building, or part thereof, as a trespasser, and then:

- steals (or attempts to steal) any item therein;
- inflicts (or attempts to inflict) grievous bodily harm (GBH) on any person therein; or
- enters a building as a trespasser and with <u>intent</u> to steal any item therein, inflict GBH on any person therein, or unlawfully damage the building or anything therein.

(Bedfordshire Police, 2019; Home Office, 2015, p.4).

2.3.2 Household and Area Characteristics for Residential Burglary

Using data from the British Crime Survey (BCS), Budd (2001) identified the types of households that were at high risk of domestic burglary in 1999. The top three characteristics were (i) “single parent household”, (ii) “head of household 16-24”, and (iii) “head of household student” (Budd, 2001, p.2), which perhaps alludes to the concept or guardianship, or the potential lack thereof. Similarly, the ONS (2017d) also identified variables (i) and (ii) as risk factors for domestic burglary, together with households in urban areas, and Malleson (2010, p.72) uncovered a Pearson’s correlation of 0.64 between SOA burglary rates and his “students (full-time)” variable. Other household risk factors identified by Budd (2001, p.2) were: “head of household economically inactive”, “head of household unemployed”, “privately rented”, “household income less than £5,000”, “rented from social landlord”, “flat”, and “terraced property”. Looking at the area level, Budd (2001, p.3) found that those in ACORN classifications: “council flats, very high unemployment, singles”, “multi-occupied terraces, multi-ethnic areas”, “council areas, high unemployment, lone parents”, “academic centres, students and young professionals”, “council flats, greatest hardship, many lone parents”, “furnished flats and bedsits, younger single people”, and “council areas, residents with health problems” experienced the highest domestic burglary risks (based on data from the 1996-2000 BCS). Taking these findings together with those of Boggs (1965), Bottoms and Wiles (1986), and Malleson (2010), as discussed earlier in the chapter, it would be reasonable to infer that the highest ‘Regular’ burglary rates in West Yorkshire are likely to present in the more deprived LSOAs and OAs and/or high student rate areas, and within close proximity of the urban centres. It is also worth mentioning here that studies have shown that household affluence exerts a positive influence on crime victimisation versus the negative influence of area affluence (Tseloni et al., 2002), that is, offenders are more likely to target wealthier households in poorer areas (e.g. see Trickett et al., 1995). This is likely to be attributable, at least in part, to the modifying effects of rational choice and the principle of least effort on offender mobility patterns – these are also likely to be important factors in determining Car Key burglary locations, i.e. the highest rates are anticipated to occur in more affluent areas that are closest to high offender rate areas.

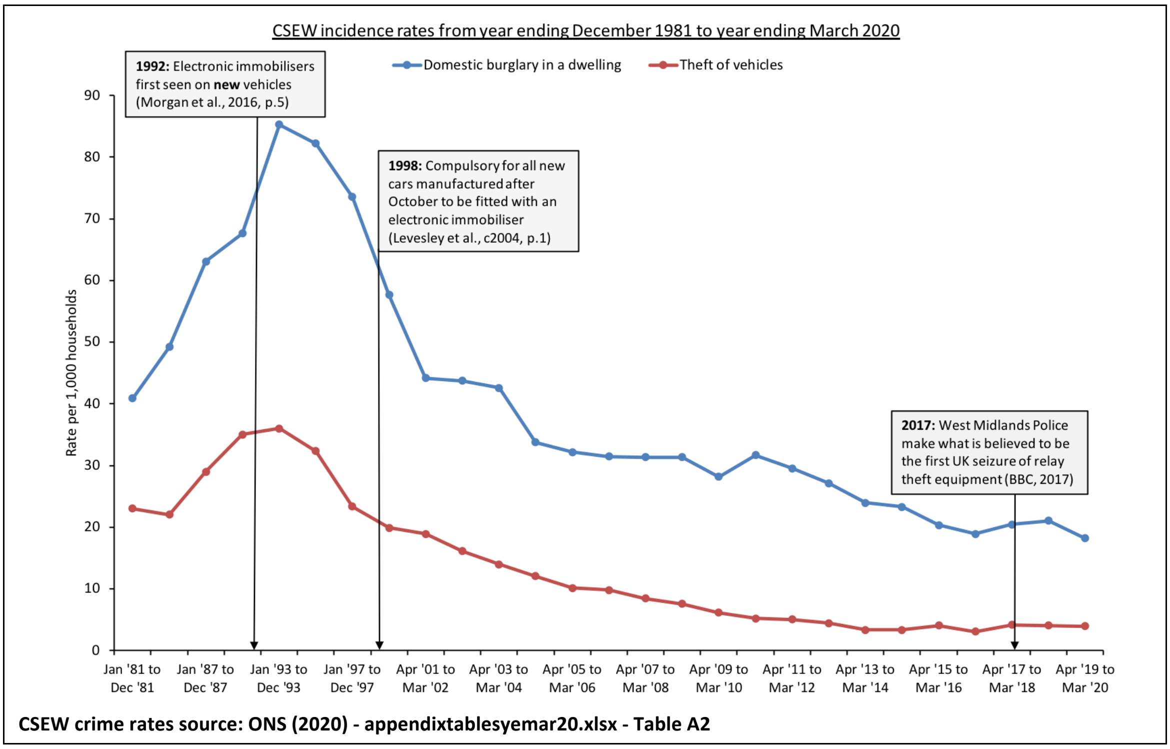
2.4 Car Key Burglary

Since there is no official Home Office classification for a ‘Car Key’ burglary, there is no universally accepted definition for the crime type, however, it is generally taken to be when an offender breaks into a dwelling with the specific aim of obtaining the keys for a vehicle, which they then steal (e.g. see: North Yorkshire Police, 2019; West Mercia Constabulary in Shaw et al., 2010, p.452). There is very little academic literature on the subject, which is probably a reflection of the fact that it is not currently recognised as a standalone offence by the HO; this is despite research identifying notable differences between the characteristics of Burglary Dwelling offences where motor vehicles are stolen and those of conventional, or 'Regular' burglaries. Conversely, there is an extensive grey literature on the subject, including newspaper articles, social media posts, and information on police force websites, the latter indicating that practitioners also view the offence as being distinct from other forms of burglary.

2.4.1 The Growing Importance of Car Key Crime

The EU passed legislation in 1995 that made it compulsory for all new cars manufactured after October 1998 to be fitted with an electronic immobiliser, the purpose of which was to prevent vehicles from being started without keys, thereby reducing thefts (Levesley et al., c2004, p.1). This change in the law has been cited as one of the main reasons for reductions in conventional 'theft of motor vehicle' offences in subsequent years (Shaw et al., 2010, p.451). Figure 2.5 below shows Crime Survey for England and Wales (CSEW) rates per 1,000 households for 'domestic burglary in a dwelling' and 'theft of vehicles' (vehicle-owning households only) from year ending December 1982 to year ending March 2020. The graphs indicate that rates for the two crime types have decreased overall since the early/ mid-1990s. Given that electronic immobilisers were first seen on **new** vehicles in 1992 (Morgan et al., 2016, p.5), the ensuing declines in 'theft of vehicles' rates are likely to be attributable, at least in part, to improved security measures, as explained by the security hypothesis (e.g. see Farrell at al., 2008). However, and particularly in the context of the current research, it is important to note that, following the mandatory implementation of electronic immobilisers, anecdotal evidence suggested that some offenders were focusing on stealing new vehicles using the keys (Levesley et al., c2004, p.1). In response to this, Levesley et al. (c2004) analysed police-recorded thefts/ attempt thefts of R (1997) registration or later cars with a known MO in the Greater Manchester and Northumbria areas between 1998 and 2001. They found that the most common method of acquiring vehicle keys was via a linked burglary (37% of offences), followed by owners leaving the keys in the ignition/ car (18% of offences) (Levesley et al., c2004, p.3). The authors also observed that the proportion of all thefts/ attempt thefts in the Greater Manchester area using keys stolen in a burglary increased from 19% in the first half of 1998 to ~ 44% in the first half of 2001 (Levesley et al., c2004, pp.2-3).

Looking again at Figure 2.5, it is worth highlighting that the rate of decline in domestic burglary appeared to slow around the time that electronic immobilisers became more prevalent in the vehicle fleet, although the extent to which this could be related to vehicle offenders switching MO, i.e. to domestic burglary, cannot be deduced from the CSEW figures. Although it is unlikely that all vehicle offenders responded to the changing opportunity structure by stealing vehicles by other methods, as per Morgan et al's discussion around security measures and crime displacement (2016, pp.7-8), there is evidence to suggest that some did. For example, in Addis et al's (2019) paper exploring the practices of steal-to-order burglars, 'Participant 14' is quoted as saying: "Only started burglaries for car keys. Used to steal cars. But as got more advanced had to have keys, so burgled to steal keys" (Addis et al., 2019, p.465).



CSEW incidence rates from year ending December 1981 to year ending March 2020

1992: Electronic immobilisers first seen on **new** vehicles (Morgan et al., 2016, p.5)

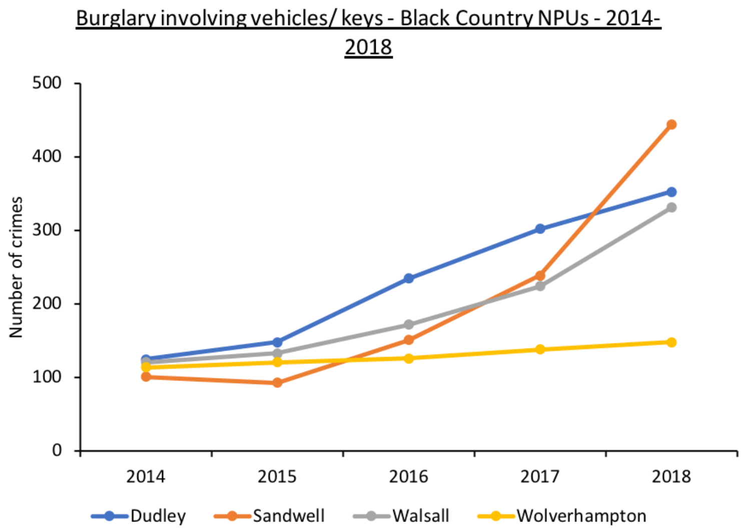
1998: Compulsory for all new cars manufactured after October to be fitted with an electronic immobiliser (Levesley et al., c2004, p.1)

2017: West Midlands Police make what is believed to be the first UK seizure of relay theft equipment (BBC, 2017)

CSEW crime rates source: ONS (2020) - appendixtablesyemar20.xlsx - Table A2

Figure 2.5 - Crime Survey for England and Wales (CSEW) rates per 1,000 households for 'domestic burglary in a dwelling' and 'theft of vehicles' (per 1,000 vehicle-owning households), 1981-2020

As mentioned previously, there is no separate HO classification for a Car Key burglary and, as such, there is a dearth of published statistics on the subject, particularly time series data. One data set, obtained via a Freedom of Information request (not submitted by the author of this work), relates to burglary offences involving vehicles/ keys – returned by a search – in Black Country neighbourhood policing units (NPUs) between 2014 and 2018. The data, which carries all of the usual caveats regarding police-recorded crime data, is graphed in Figure 2.6 below and shows that the four NPUs, and particularly Sandwell, experienced an increase in offences over the five year period. Indeed, Car Key burglary was such a problem for West Midlands Police in 2018 that 'Operation Cantil' was set up to tackle it (West Midlands Police, 2019, p.12), which further justifies the choice of PhD research topic.



Crime data source: West Midlands Police (2019b) - https://foi.west-midlands.police.uk/car-key-burglary-2239-19/

Figure 2.6 - Burglary involving vehicles/ keys - Black Country NPUs, 2014-2018

In addition, Car Key burglary is mentioned in the National Crime Agency's (NCA) 'National Strategic Assessment of Serious and Organised Crime (SOC) 2019' in relation to organised acquisitive crime (OAC): "The potential harm to citizens from OAC has increased in the past year. In some cases, offenders show a growing propensity for violence if confronted. For example, confrontations during car key burglaries are increasing" (NCA, c2019, p.34). A review of the grey literature, including social media sites and online newspapers, also indicates that Car Key burglaries present an ongoing issue – Table 2.1 below presents some recent news headlines to illustrate.

Headline	**Author and publication title**	**Publication date**
Laptop computer and Nissan stolen during car-key burglary in Kenilworth	Elofson, M. **The Leamington Spa Courier**	20/09/2020
Police warning after "car key burglary" in Wrexham	n/a **Wrexham.com**	13/08/2020
Call for CCTV footage after a car key burglary in Amblecote	Holder, B. **Stourbridge News**	09/08/2020
Man arrested after Mansfield car key burglary	Day, L. **Chad**	28/07/2020
Audi and Polo targeted in Hanoi-style burglary in Cross Hills	n/a **Keighley News**	16/07/2020
Police warn Rushcliffe residents to be vigilant after 'hook and cane' car key burglary	n/a **West Bridgford Wire**	28/05/2020

Table 2.1 - Selection of recent news headlines relating to Car Key burglary offences

It is also expedient to note at this point that a previous review of the grey literature identified different terminologies that the police use when referring to Car Key burglaries; these are listed in Table 2.2 below and include 'Hanoi', so called after 'Operation Hanoi', the first police operation to tackle the problem (BBC, 2005), and ‘2-in-1’ (North Yorkshire Police, 2018), i.e. two offences committed in close temporal proximity (Burglary Dwelling and theft of motor vehicle). This potential range of vocabulary is certainly something to be aware of when deriving the Car Key burglary sample for the current research because, in the absence of an associated stolen vehicle record for an offence, it will be necessary to rely on data input fields to establish if an offence is a Car Key burglary.

Terminology and (reason)	Example source
‘Hanoi’ burglary (Operation Hanoi)	West Yorkshire Police (2019)
‘Hook and cane’ burglary (hook or magnet on end of cane)	Greater Manchester Police (2019)
‘Millennium’ burglary (when problem emerged)	Essex Police (ca. 2015)
‘2-in-1’ burglary (effectively two offences at once)	North Yorkshire Police (2018)

Table 2.2 - Alternative police terminology for a Car Key burglary

As long as cars remain a desirable commodity for criminals, Car Key burglaries are likely to continue, especially as newer vehicles, i.e. those with immobilisers, become more prevalent within the UK fleet. However, it is important to recognise that technology is constantly changing and, thus, offenders’ responses to this, a contemporaneous example being the recent emergence of so called ‘relay’ thefts. A ‘relay’ theft/ attack is when two electronic devices (‘relay’ boxes) are used to trick a keyless entry vehicle into thinking that the key is nearby – by extending its signal – which causes the doors to unlock; offenders can then enter the vehicle, push the start button, and drive it away (Keeling, 2018; Harding, 2020). The ‘relay’ theft MO is illustrated in Figure 2.7 below (Harding, 2020).

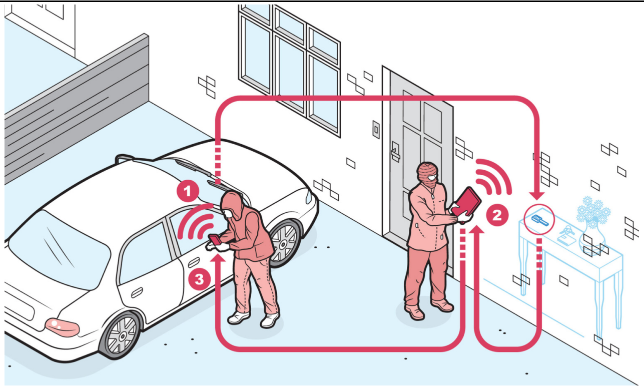
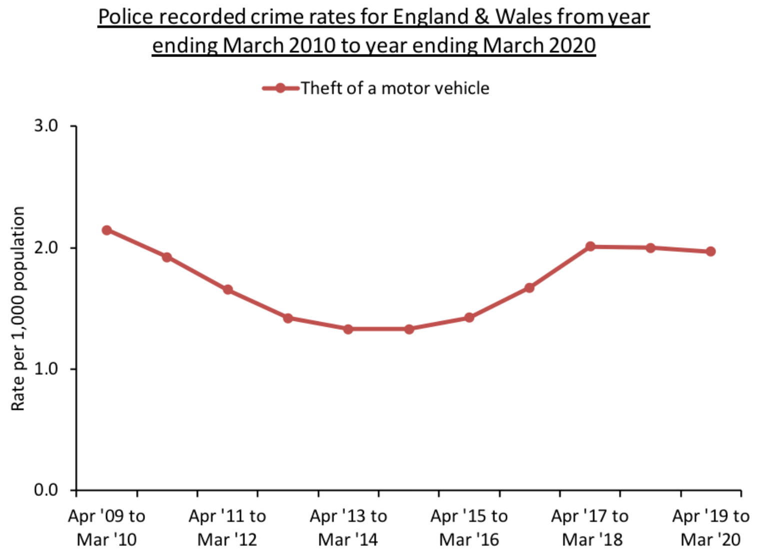


Image source: Harding (2020)

Figure 2.7 - Relay theft MO

As depicted earlier in Figure 2.5, in 2017 West Midlands Police made what is believed to be the first UK seizure of relay theft equipment (BBC, 2017), and in March 2020 a male was arrested for stealing 10 keyless cars in Peterborough between June 2019 and October 2019 (ITV, 2020). Notably, there was an uptick in police-recorded theft of motor vehicle rates starting approximately two years before the West Midlands Police seizure, as can be seen in Figure 2.8 below, although it is not possible to ascertain from the data if this can be explained by the emergence of the 'relay' theft MO.



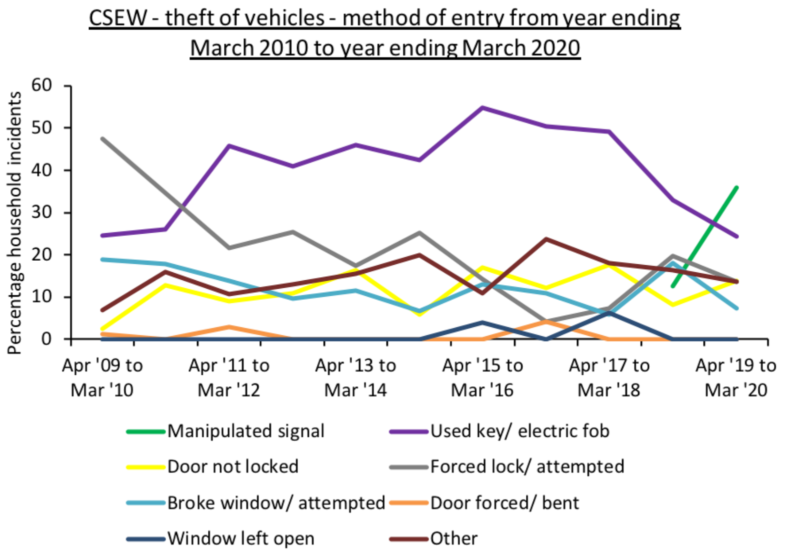
Police recorded crime rates source: ONS (2020) - appendixtablesyemar20.xlsx - Table A7

NB Greater Manchester Police offences are excluded from the last two years as they were unable to supply figures due to the introduction of a new IT system.

Figure 2.8 - England and Wales police-recorded crime rates per 1,000 population for 'theft of a motor vehicle', 2009-2020

Although there was a slight decrease in theft of motor vehicle rates between year ending March 2019 and year ending March 2020, it is worth considering that, if car manufacturers do not implement a fix for the 'relay' attack problem on all affected vehicles – some have on certain models (Harding, 2020) – and keyless entry also becomes more common, the MO may spread, potentially increasing thefts. This could also see car thieves switch from stealing vehicles via linked domestic burglary offences to via theft of motor vehicle offences. Since a vehicle can be stolen in around one minute using the 'relay' method (GMP in Keeling, 2018), this presumably renders it a less risky option for offenders than committing two offences in one, i.e. a Burglary Dwelling offence (to procure the vehicle keys) followed by a theft of motor vehicle offence. Further, the 'relay' offence "can be completed in near silence" (GMP in Keeling, 2018) as compared to a Burglary Dwelling where it might be necessary for offenders to gain entry to a secure, and possibly occupied, property e.g. by snapping a door lock.

So, what are the potential implications of this for the current research? At the time of writing, Car Key burglaries are still being reported in the press, indicating that the PhD subject matter is still relevant. In terms of the 'relay' theft MO, it is pertinent to note that from year ending March 2019 the ONS introduced a 'manipulated signal' category in their CSEW 'theft of vehicles' method of entry statistics, as shown in Figure 2.9 below, which suggests that the MO is becoming more popular amongst thieves. 'Offender manipulated signal from remote locking device' overtook 'offender used a key/electric fob' in year ending March 2020, and the year-on-year change in percentage household incidents for the former was statistically significant at the 5% level, increasing from 13% in year ending March 2019 to 36% in year ending March 2020.



CSEW - theft of vehicles - method of entry source: ONS (2020b) - nocvehiclethefttablescorrection.xlsx - Table_3b

NB There can be more than one MO recorded per incident.

Figure 2.9 - CSEW - theft of vehicles - method of entry, 2010-2020

However, even if offenders switch from stealing newer cars via burglaries to via 'relay' attacks, the latter are also likely to be predominantly residentially-based, that is, cars will probably still be targeted whilst parked in the vicinity of domestic properties. This is because relay thieves (i) need a vehicle to be located within close proximity of its keys and (ii) preferably in the absence of a capable guardian. For example, if offenders try to relay a signal from vehicle keys stored on someone's person, say in a car park, they will be at a very high risk of detection due to the minimum distance required to do so. It is also useful to note here that the CSEW found that 72% of 'theft of vehicles' household incidents were committed at home (private/ semi-private, e.g. unconnected garage/ street) in year ending March 2020 (ONS, 2020b). Thus, the takeaway point from this section is that the routine activities of potential victims determine the spatio-temporal distribution of car keys, and that one of the lowest risk opportunities for 'relay' thieves to target these is when vehicles are parked close to dwellings and the occupants are rendered incapable guardians, e.g. through sleep, as described in Figure 2.10 below. Based on this, the likelihood is that the spatio-temporal patterning of offences will be relatively similar whether a car is stolen via a linked Burglary Dwelling offence or via a 'relay' MO theft of motor vehicle, and that the PhD research rationale and any subsequent findings should therefore hold true in either scenario. Despite the aforementioned, it is also important to acknowledge that changing opportunity structures (e.g. see Copes and Cherbonneau, 2006) could, at some point, render the current research findings outdated, for example, if technological developments mean that newer vehicles can be stolen from non-domestic locations with little risk to, and effort on the part of, offenders.

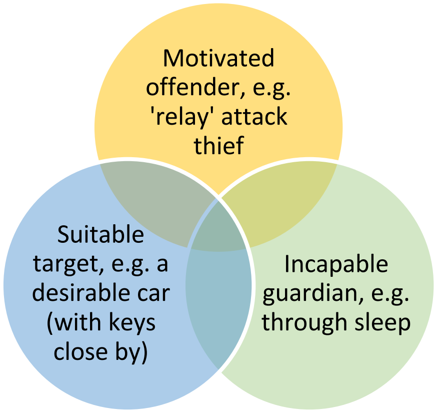

Motivated offender, e.g. 'relay' attack thief
Suitable target, e.g. a desirable car (with keys close by)
Incapable guardian, e.g. through sleep

Figure 2.10 - Application of Routine Activity Theory (Cohen and Felson, 1979) to 'relay' attack theft of motor vehicle offences

2.4.2 Offence Characteristics

Shaw et al. (2010) analysed 514 car key burglaries and 514 conventional burglaries reported to Northamptonshire Police between May 2007 and March 2009 with a view to identifying if there were any notable differences between the characteristics of these. The authors performed logistic regression analysis on a number of predictor variables, including 'time of day' and 'IMD' (Index of Multiple Deprivation), and found three of these to be statistically significant in terms of their capacity to differentiate between car key burglaries and conventional burglaries:

Variable 1: Time of Day

The 'time of day' variable showed that car key burglaries are more likely to be committed overnight than conventional burglaries, which is as might theoretically be expected given that vehicles tend to be parked in the vicinity of dwellings during occupied periods. Further, the analysis also identified that burglaries committed between the hours of 20:00 and 07:59 are approximately five times more likely to be car key offences.

Variable 2: IMD

The 'IMD' (Index of Multiple Deprivation) variable showed that car key burglaries are more likely to occur in less deprived neighbourhoods than conventional burglaries.

Variable 3: Search Type

The 'search' type variable showed that car key burglars are more likely to commit a tidy search of the premises than conventional burglars. This finding might be attributable to the fact that properties are likely to be occupied if there is a vehicle parked outside so offenders need to keep noise to a minimum to reduce the risk of apprehension (Shaw et al., 2010, p.455).

In a similar vein, Carden (2012) conducted an exploratory analysis of car key burglaries in Merseyside using police recorded crime and offender data from 2010. As per Smith et al. (2010), the author found "distinct differences" between car key offences and conventional burglaries (p.76), including that the former are more likely to be committed overnight between 22:00-07:59 (p.60) and offenders are more likely to conduct a tidy search of the premises (p.69). It should be noted, however, that the research findings did not support the idea that car key burglaries are more likely to be committed in affluent areas, with Carden instead citing proximity to fast transport links as a possible determining factor (p.79). Further to the findings of Smith et al. (2010), the author identified the following points of difference between car key offences and conventional burglaries, including variations in offender behaviour for both crime types:

Modus Operandi

Car keys burglars are more likely to use the 'hook and cane' or 'remove/ cut door/ window' MO than conventional burglars but they are far less likely to enter a property by 'force/ smash door/ window', or by means of deception (pp.76-77). As discussed earlier, this is probably to reduce the risk of apprehension associated with occupied properties. Also related to this point, Carden (2012) identified that car key burglars are less likely to be seen during an offence even though properties are more likely to be occupied (p.77).

Entry Point

Car key offenders are more likely to enter a property at the front, as opposed to conventional burglars who typically enter dwellings via the rear (p.77). This might be due to the fact that car keys are often located just inside the front door of a property e.g. 'hook and cane' MO. Further, car key burglars typically operate overnight so they will be afforded a level of cover during hours of darkness.

Target Property Type

Car key burglars are less likely to break into flats than conventional burglars, but they are more likely to target semi-detached properties (p.77). This is probably because flats tend to have communal parking areas whereas semi-detached houses usually have driveways, the latter enabling offenders to match cars to properties more easily (Carden, 2012, pp.80-81).

Distance Travelled

Car key burglars are willing to travel further from home base to crime location than conventional burglars (p.78), possibly because offenders tend to reside in more deprived areas, whereas high-value vehicles are typically located in affluent neighbourhoods. The median distance travelled by car key offenders in the Carden (2012) study was 4.9 km (maximum 23.41 km) *versus* 2.09 km (maximum 14.67 km) for the conventional burglars (n = 140) (p.74).

2.4.3 Offender Characteristics

Shaw et al. (2010), Allcock et al. (2011) and Chapman et al. (2012) all assert that car key burglaries should be considered separately to conventional burglary dwelling offences. Adding further weight to this argument, Chapman et al. (2012) examined the offending history of 110 car key burglars and 110 conventional burglars who were detected to a residential burglary offence in Northamptonshire between January 2008 and January 2009. The authors performed logistic regression analysis on a number of predictor variables, including 'previous for theft of a motor vehicle' and 'co-offending group', and found three of these to be statistically significant in terms of their capacity to differentiate between car key burglaries and conventional burglaries:

Variable 1: Previous for theft of a motor vehicle

Car key burglars are more likely than conventional burglars to have a previous conviction for theft of motor vehicle (p.939). This finding indicates that, following the introduction of immobilisers, car thieves adapted their offending behaviour in order to circumvent vehicle security.

Variable 2: Detection method – on information

Car key burglars are more likely than conventional burglars to be detected on information provided to police (p.939). The authors suggest that this finding reflects the complexity of car key burglary offences, specifically that they are likely to involve a chain of individuals, thus making them more susceptible to detection (p.944).

Variable 3: Previous shoplifting offence

Conventional burglars are more likely than car key burglars to have a previous conviction for shoplifting (p.939). This finding again indicates that car key burglaries and conventional burglaries are not directly comparable, thus challenging the assumptions that are inherent to current crime recording methods i.e. that car key burglaries are committed by conventional burglars.

Although the analysis showed there to be slightly more co-offending groups involved in the car key burglary offences than the conventional burglaries, the difference was not found to be statistically significant (p.944), which is surprising. The authors do, however, suggest that the lack of statistical significance for this variable might be explained by detection bias within the data (p.944). Two possible reasons are given for this, the first being that those individuals who are involved in the disposal of stolen vehicles might not commit the original thefts, so they are less likely to be detected. The second is that the police might not identify every offender who was present at the scene of a crime so these individuals will not appear in the data.

2.4.4 Summary

There is clearly a need for more empirical research in relation to car key burglaries, particularly since some of the existing findings are contradictory (e.g. see Carden, 2012; Shaw et al., 2010). Similarly, Chapman et al. (2012, p.945) highlight the importance of validating current findings through replication, since a fuller understanding of the subject area would be of benefit to organisations that are concerned with the prevention and detection of crime. Further, Bernasco (2006 **in** Bernasco et al., 2015, p.127), speaking in the context of repeat burglary victimisation, suggests that more research is needed into the spatio-temporal decision-making of co-offending groups, a point which is extremely relevant given the current research topic.

It is also worth noting at this point that the current Home Office recording method for car key burglaries receives much criticism within the related literature, both from an analytical perspective and also because it is leading to a flawed interpretation of vehicle crime statistics, i.e. that theft of motor vehicle offences have apparently declined over recent years (Chapman et al., 2012, p.945).

2.5 Existing Models

2.5.1 ProMap and PROVE – Boost

The ‘boost’ element for the ProMap combined risk surfaces, referred to here as the ‘basic ProMap algorithm’, is derived by overlaying a grid of equally sized cells onto a study area and drawing a circle of radius 400 m around each cell centroid. Recent crimes that fall within 400 m of a particular grid cell are then weighted according to distance from the cell and number of days from the date of the prospective map for which risk intensity values are being generated. For each crime within a 400 m radius, the product of the inverse of the distance and time weightings is calculated, and these are then summed to give a risk intensity value for the cell (Bowers et al., 2004). This is an advancement of the moving window KDE approach outlined in Gatrell and Bailey (1995) because temporal decay is represented in the resulting risk surface. Although the basic ProMap algorithm outperformed all of the retrospective crime mapping techniques that were included in Johnson et al.’s 2009 study of general residential burglary in Merseyside (p.188), the method is not expected to be appropriate for estimating future Car Key burglary locations because the associated spatio-temporal decay function is designed to reflect conventional foraging behaviours, i.e. the risk of revictimisation increases the closer a potential target is to a recent crime location.

Ratcliffe et al.’s PROVE software also incorporates spatio-temporal decay in its predictions, but, rather than drawing a buffer around recent offences, near-repeat odds ratios are instead allocated to grid cells. For example, taking an identified contagion period of up to 28 days, the last 28 days’ offences are plotted onto a grid and relevant odds ratios are then assigned to cells based on the spatial and temporal proximity[2] of these to the crime points. So, for a distance-time band of 251-500 m and 2 days having an associated odds ratio of 1.17, any grid cells situated within this range of a recent offence would be stamped ‘1.17’. A key point to note here is that the PROVE software only produces risk maps for a minimum duration of seven days (Ratcliffe et al., 2016b, pp.12-13), which means that any offences committed during a live prediction period will not be captured in the associated risk surface. A potential implication of this is that emerging hot spots might not be addressed until the next prediction period. This was also a limitation of the Trafford approach in that patrol maps were only issued on a weekly basis

[2] It is not clear from the associated documentation if PROVE assigns odds ratios based on a single summary contagion period, e.g. 28 days, and several distance bands, or if it disaggregates the distance bands into shorter time periods, e.g. 0-100 m and 0-7 days, 0-100 m and 8-14 days..., etc. (the latter is assumed here).

(Fielding and Jones, 2012, p.32) and so offences committed in the six days post-dissemination did not contribute to current risk levels.

2.5.2 ProMap and PROVE – Flag

It is useful here to consider how risk heterogeneity was incorporated in each of the models. Johnson et al. (2009) created three static risk variables at the grid cell level, namely roads, buildings, and social barriers, multiplying the respective values for each of these by the basic ProMap algorithm risk values, and in various combinations. Looking at the median results that were achieved when the cells were ordered by risk intensity value, and the proportion of future offences captured recorded, the combined risk surfaces outperformed the 'boost' surface (basic ProMap algorithm) in some instances, for example, the basic ProMap algorithm captured 50 per cent of future offences in 14.3 % of grid cells, whereas the basic ProMap algorithm * Houses * Roads captured 50 per cent of future offences in 12.5 % of grid cells (Johnson et al., 2009, p.188).

In PROVE, the risk heterogeneity surface ('long-term indicator') is generated using negative binomial GLM regression and three independent variables – socio-economic status (SES), total population, and one year's worth of prior offences – to produce predicted crime counts at the census block level. There is also the option to include a race variable in the regression analysis. The resulting crime counts are disaggregated spatially and temporally to align with the grid cell level 'boost' output ('short-term indicator'), and the model is then calibrated to identify relevant weightings for the long-term (flag) and short-term (boost) indicators. Information from the model calibration stage is subsequently used to generate risk predictions for a user-specified time period, with the software outputting a map of expected crime counts for either 7, 14, 21, or 28 days (Ratcliffe et al. 2016; Ratcliffe et al., 2016b).

2.6 Application of the Literature to the PhD Research

Shaw et al. (2010) found that car key burglaries are more likely to occur in less deprived/ more affluent neighbourhoods, Ratcliffe and McCullagh (1999) found that properties in deprived areas are more likely to suffer repeat victimisation than properties in affluent areas, and Johnson and Bowers (2004) assert that more affluent areas experience more near-repeats. Taking all of this, together with the target property type of desirable vehicles, the research on journey-to-crime (JTC), and offenders' likely responses to macro-level community controls, such as collective efficacy, it is inferred that Car Key burglaries will typically present as either lone offences or temporally clustered crime sprees.

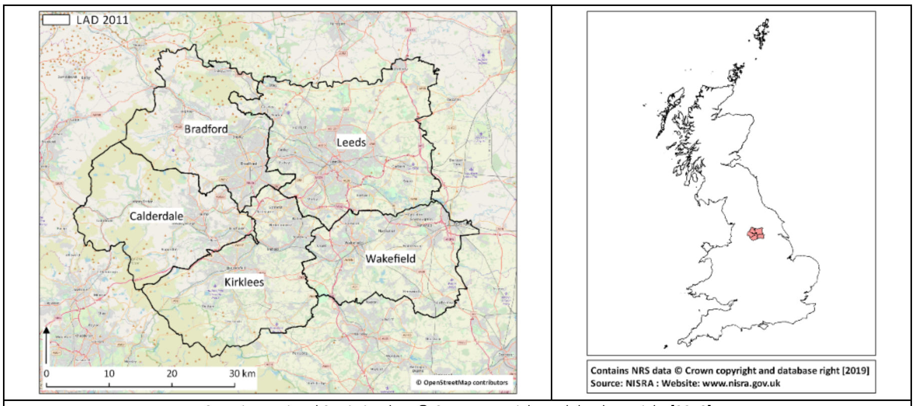
Chapter 3 Data Selection and Exploratory Analysis

3.1 Introduction

This chapter will first provide an overview of the study area, followed by a brief description of the core spatial data that are used in the thesis. Three sources of crime data for England will then be reviewed, after which the crime samples selection method that was employed in the current work will be explained. There is currently no official Home Office classification for a Car Key burglary and so it was not simply a case of separating the two burglary types by name, i.e. Car Key/ Regular. Further, and to the best of the author's knowledge, there are no official/ universally accepted guidelines for the pre-processing of crime data for analysis or, indeed, the input fields that should be considered. For example, selecting crime records for analysis based on the reported date, as opposed to the committed date(s), runs the risk of including late reports in a sample, and thus the potential for data alignment issues. The findings of two normality tests that were performed on the burglary rates are next presented, together with a discussion of how these impacted on choice of statistical tests, i.e. parametric/ non-parametric. Section 3.5 introduces the independent variables that were subsequently analysed with the crime data, these being chiefly socio-demographic in nature, as well as how they were operationalised. Some potential issues for consideration when working with different types of data are set out in Section 3.6, including the Modifiable Areal Unit Problem (MAUP). Aside from the crime samples selection method, perhaps the most important part of this chapter is the exploratory spatial analysis included in Section 3.7 since this gives an initial empirical indication as to whether, or not, the research rationale is likely to be valid, i.e., do the characteristics of high Car Key burglary rate areas differ from those of high Regular burglary rate areas? If shown to be true, then there is a good chance that the spatio-temporal signatures of the two burglary types will also differ, with each generating distinct repeat/ near-repeat victimisation patterns, and thus potentially impacting the efficiency of prevailing burglary modelling techniques, such as buffering of recent crime events. Section 3.8 presents the results of 'aoristic' analysis that was undertaken to identify hot time periods for the two burglary types, the idea being to incorporate these into the current research models. Key observations and opportunities for further work are summarised in Section 3.9.

3.2 Study Area

The study area for the current research, West Yorkshire Metropolitan County, was determined by the spatial extent of the available crime data. The County is centrally located within Great Britain and is sub-divided into five Metropolitan Boroughs, or LADs (local authority districts): 'Bradford', 'Calderdale', 'Kirklees', 'Leeds', and 'Wakefield' (see Figure 3.1 below). Each Borough has its own council responsible for administering local government services, such as education and planning, however, some functions, including transportation, are delivered in partnership with the West Yorkshire Combined Authority. Key administrative centres are the cities of Bradford, Leeds, and Wakefield, and the towns of Halifax (Calderdale) and Huddersfield (Kirklees). The LAD boundaries also demarcate the geographical extents of five 'routine' policing districts within the West Yorkshire Police force area, each of which is subdivided into two or more Neighbourhood Policing Team neighbourhoods. There are also specialist police departments that operate at the cross-district level, e.g. Roads Policing Unit (RPU). The county covers an area of 2,029 km^2 and is surrounded by a mix of rural and urban areas, the latter including Greater Manchester to the south-west and Sheffield to the south-east, which could render it particularly vulnerable to cross-border mobile offenders.


LAD 2011
Bradford
Leeds
Calderdale
Wakefield
Kirklees
0
10
20
30 km
© OpenStreetMap contributors
Contains NRS data © Crown copyright and database right [2019]
Source: NISRA : Website: www.nisra.gov.uk

Contains National Statistics data © Crown copyright and database right [2019]
Contains OS data © Crown copyright [and database right] (2019)
Contains public sector information licensed under the Open Government Licence v3.0
[http://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/]
This work is based on data provided with the support of the ESRC and JISC and uses boundary material which is copyright of the Crown, the Post Office and the ED-LINE consortium.

Figure 3.1 - The West Yorkshire study area and its location within Great Britain

At the 2011 Census UK, the usual resident population of West Yorkshire was 2,226,058 and the number of households was 922,452 (per cent in each LAD shown in Table 3.1 below). Leeds LAD contained the highest proportion of the total resident population with 33.8%, whereas Calderdale was the least populated with 9.2%. The district rank orders were the same for both variables, i.e. Leeds (1), Bradford (2), Kirklees (3), Wakefield (4), and Calderdale (5).

LAD	Res. population	Per cent of total	Households	Per cent of total
Leeds	751,485	*33.8*	320,596	*34.8*
Bradford	522,452	*23.5*	199,296	*21.6*
Kirklees	422,458	*19.0*	173,525	*18.8*
Wakefield	325,837	*14.6*	140,414	*15.2*
Calderdale	203,826	*9.2*	88,621	*9.6*
Total	**2,226,058**	***100.0***	**922,452**	***100.0***

Table 3.1 - Distribution of total usual resident population and households across West Yorkshire LADs as at the 2011 Census UK (source: 2011 Census, ONS)

3.3 Spatial Data

Since a number of census-derived variables were planned to be analysed in the current work, it seemed sensible to align the crime rates data with the official statistics (ONS) geographies for these variables, specifically Lower Layer Super Output Areas (LSOAs) and Output Areas (OAs). Output Areas (OAs) are the smallest geography for which census estimates are released, having been introduced in the UK for the 2001 Census (ONS, ca. 2016). They were designed to have similar population sizes, avoid urban/ rural mixes, and to be as socially uniform as possible with regards to household tenure and dwelling type (ONS, ca. 2016). This means that they are likely to be conducive to unmasking ‘local’ drivers of crime in the current research. LSOAs are derived from aggregations of OAs and it is also possible to nest both of these geographies within Local Authority District (LAD) boundaries. The minimum population count for a West Yorkshire OA, 2011 was 101, with an average of 312, and for a LSOA it was 1,011, with an average of 1,604 (Table KS101EW). The equivalent household counts were OA minimum 39, with an average of 129, and LSOA minimum 399, with an average of 665 (Table KS105EW). Minimum count thresholds are set by ONS to prevent disclosure (ONS, ca. 2016). Given that the LADs also mirror the five West Yorkshire Police districts, these boundaries will be overlaid onto maps to provide police relevant context.

3.4 Crime Data

3.4.1 Sources Of

The main sources of crime data for England are:

(i) The Crime Survey for England and Wales;

(ii) police-recorded (open source); and

(iii) police-recorded (secure source).

Although the current research will utilise secure source police-recorded crime data, it is useful here to briefly outline the positive and negative aspects inherent to each of these three sources.

The Crime Survey for England and Wales

The Crime Survey for England and Wales (CSEW), formerly the British Crime Survey (BCS), is a face-to-face, computer assisted victimisation survey that is conducted on behalf of the Office for National Statistics (ONS) (Tseloni and Tilley, 2016). The Survey asks approximately 35,000 people annually, aged 10 and over (from 2009, aged 16 and over before), in private households in England and Wales, about their experiences of crime in the last year (Tseloni and Tilley, 2016), thus potentially capturing offences not reported to the police (Ratcliffe, 2016). Given that the first CSEW was conducted in 1982, it is a useful source of longitudinal data, however, from the point of view of the current research, the results are not released at a spatially disaggregate enough level to be of any practical use.

Police-Recorded (Open Source)

Street-level crime data is available under Open Government licence v3.0 from the Home Office's 'data.police.uk' website (Home Office, ca. 2018). Due to the open nature of the site, crimes are aggregated temporally by year and month, and anonymised spatially through the assignment of actual crime coordinates to anonymised map points, including street centres and public spaces. Crime types are also aggregated, for example, there is no means of differentiating between a residential burglary and a commercial burglary. The website has a 'custom download' interface where users can select crime data for individual police forces and for a specific time period. Archived data is also available on the site. All requested data is supplied in .csv format and includes, amongst other information: 'crime type', 'latitude', 'longitude', 'location', and 'LSOA code', the latter facilitating analysis with census-derived data. A key benefit of this source is the relative ease by which large volumes of crime data can be acquired, however, this is accompanied by loss of spatio-temporal detail, which is fundamental to much criminological research, including repeat victimisation analysis.

Police-Recorded (Secure Source)

Access to spatially and temporally disaggregated crime data is imperative to much of the analysis in this thesis, especially the identification of repeat/ near-repeat victimisation patterns. However, it is important to note that this is not a panacea for all, with associated issues including missing data (Davies et al., 2012, p.282, p.292; Shaw et al., 2010, p.453), variable free text entry (Adderley and Musgrove, 2003, p.270), local grammar (Rogerson, 2016, p.77), and potentially inaccurate crime locations, for example, the Home Office (ca. 2018) states that "estimates of geocoding accuracy in different forces range from 60% to 97%". More generally, it should be noted that police-recorded crime data is subject to future change, for example if a crime is reclassified, meaning that it should be viewed as a snapshot in time.

According to Dodd et al. (2004, p.34), the England and Wales reporting rate for burglary with loss, based on 2003/2004 British Crime Survey interviews, was 78%. This indicates that the PhD burglary samples should be representative enough from which to draw fairly reliable conclusions, although it is important to acknowledge the possibility of systematic bias within these, i.e. the types of people who might not report crime are likely to live in particular areas.

3.4.2 Crime Samples Selection Method

One of the most time consuming aspects of the current research was the derivation of relevant data sets for analysis, particularly the burglary samples. This section will describe how individual crime records were classified as being either a Car Key burglary or a Regular burglary, and also how the two resulting crime samples were subjected to further manipulation depending on the type of analysis that was to be undertaken. Key outputs from the data preparation stage are listed below, all of which relate to offences with a mid-point committed date between 01/01/2010 and 31/12/2014:

1. Spreadsheet of individual burglary records classified as either Car Key/ Regular burglary
2. Point shapefiles of the Car Key burglary and Regular burglary offences (from s/sheet)
3. Polygon shapefiles of LSOA and OA Car Key burglary rates per 1,000 households
4. Polygon shapefiles of LSOA and OA Regular burglary rates per 1,000 households

Initial Raw Crime Data Set

Data for all Burglary Dwelling offences recorded by West Yorkshire Police between 1st January 2004 and 30th June 2015 (11.5 years) was provided to the author in Excel spreadsheet format; this was split across various workbooks, each of which contained information on either crime (individual offence details, including earliest date committed and MO) or property (items stolen from individual offences). Unique reference numbers were also supplied with the data to facilitate cross-referencing, i.e. so that stolen property could be linked to associated crime records. Table 3.2 below summarises the crime and property workbook structures, including column header descriptions. The author subsequently combined the individual crime workbooks, which resulted in an initial raw crime data set comprising 197,538 records (rows). A quick inspection of the crime data identified some potential issues, including blank input fields, which will be discussed in more detail later.

Workbook type	**Column headers**	**Description**
Crime	URN date_first_crimed date_earliest date_latest ho_class ho_offence easting northing crime_location crime_notes mo system_code	Unique Reference Number Date 1st Crimed Earliest Date Committed Latest Date Committed HO class HO Offence Easting Northing Occurrence Address Crime Notes MO Source System
Property	URN property_type property_sub_type	Unique Reference Number Property Type Property Sub Type

Table 3.2 - Crime and property workbook structures (information provided by West Yorkshire Police)

NB Earliest Date Committed and Latest Date Committed include time in hh:mm:ss format

Although crime data was available for 11.5 years, it was decided to only use those records with a mid-point committed date between 2010 and 2014, i.e. the most recent full five years. Reasons for this were: (i) to more closely align the data with the most recent Census UK year and IMD 2015, (ii) to account for changing terminology, for example, vehicles were recorded as Property Sub Type “MOTORVEHICLE” in the property workbooks up until late 2007, as opposed to “CAR”, “VAN”, etc. for the more recent offences, and (iii) to reduce the amount of manual processing required at the data preparation stage. The mid-point year for each offence was calculated by adding the ‘earliest committed date and time’ to the ‘latest committed date and time' and then dividing the result by two (the aoristic nature of residential burglary is covered in detail in Section 3.8 of this chapter). Disregarding those offences with a mid-point committed date outside the five year study period meant that 76,554 of the original 197,538 records (38.8%) were carried forward to the next processing stage (referred to henceforth as the 'current research Burglary Dwelling data set').

Fundamental to the analysis that follows was the ability to differentiate between 'Car Key' burglaries and 'Regular' burglaries within the current research Burglary Dwelling data set. Recalling that the Home Office does not provide an official classification for Car Key burglaries, it was necessary to employ a manual selection method to derive the two burglary samples, and one that was also repeatable. As mentioned in Chapter 2, West Yorkshire Police sometimes use 'Hanoi' to refer to a Car Key burglary (after an operation to tackle the problem) and so a search was conducted, as outlined in Figure 3.2 below, to select any records that contained this word in either the *crime_notes* or *mo* column. This search identified 2,919 unique records – these were subsequently recorded as 'Car Key burglary' and also used to identify relevant stolen vehicle types.

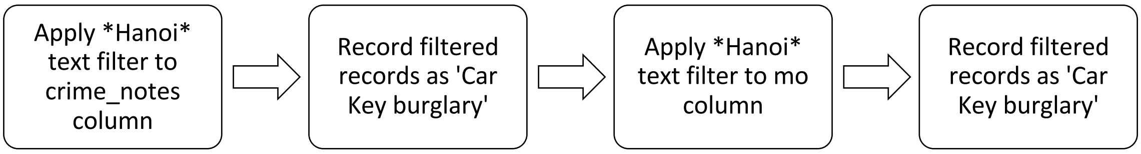

Apply *Hanoi* text filter to crime_notes column
Record filtered records as 'Car Key burglary'
Apply *Hanoi* text filter to mo column
Record filtered records as 'Car Key burglary'

Figure 3.2 – Hanoi keyword search criteria for selecting Car Key burglary offences within the current research Burglary Dwelling data set

Different vehicle types were included in the property workbooks (*property_type* column) and so the VLOOKUP function was used in Excel to identify any that were linked to the 2,919 'Hanoi' records, i.e. to understand which of these types the police considered to be a 'Hanoi' offence. Of the 3,087 vehicles that were linked to the 2,919 'Hanoi' records (more than one vehicle stolen in some offences), 92.4% were recorded as *property_type* = "CAR" and the remainder were either "GOODS VEHICLE", "LIGHT 4 X 4 UTILITY", "MOTORCYCLE", or "VAN". These vehicle types were subsequently used to identify any Car Key burglary offences within the current research Burglary Dwelling data set that had not been picked up by the 'Hanoi' keyword search. An additional 4,073 burglaries were selected via this method, with a total of 3,607 linked cars (88.6%) – this figure is similar to the 'Hanoi' "CAR" proportion. The final Car Key burglary classification criteria was therefore *hanoi* in *crime_notes* and/ or *mo* field and/ or ≥ 1 stolen vehicle; all other records were labelled as Regular burglary (69,562). Final record counts are shown in Table 3.3 below.

Search criteria	Record count
hanoi keyword in crime_notes and/ or mo field	2,919
≥ 1 stolen vehicle linked to offence	6,831
***hanoi* keyword and/ or stolen vehicle (i.e. <u>unique</u> records)**	**6,992**

*Table 3.3 - Record counts for *Hanoi* keyword and stolen vehicle linked to offence searches*

Unfortunately, the only other data that was available to the author in relation to the vehicles in the property workbooks was the "Property Sub Type", for example, "HATCHBACK" or "SALOON". The absence of any information regards vehicles' makes, models, and ages is a key limitation of the current research because some of the offences that were subsequently classified as being Car Key burglaries might not actually have been committed with the sole intention of stealing a vehicle to sell on. For example, some vehicles could have been stolen to carry stolen goods away from the scene of a burglary, to joy ride, to commit another offence, or simply as a means of transport. This could have a bearing on subsequent analysis, for example, different makes/ models/ ages of vehicle might be targeted in different areas and for different reasons, but this cannot be tested with the available data.

A second limitation of the crime samples selection method is that some attempt Car Key burglaries might be present in the Regular burglary sample, i.e. a vehicle was not stolen and 'Hanoi' was not included in the crime record but there was evidence of intent to steal a vehicle. Although manual keyword/ phrase searches, e.g. for "car keys", "vehicle on driveway", etc., might at first appear to be a suitable option by which to identify possible attempts, after having read a number of the crime MOs it was decided that there was a high risk of misclassification. For example, a daytime burglary offence with the hypothetical MO "handbag containing purse and car keys stolen from kitchen – suspect made off on foot" is, on the balance of probability, more likely to be a Regular burglary than a Car Key burglary but a search for "car keys" would result in this being classified as Car Key burglary. In light of this potential ambiguity, a more sophisticated approach, supervised text classification, is proposed in Chapter 7 of this thesis.

The final steps of the crime samples selection method included deleting any records with missing/ incorrect geocoding, invalid committed between times, and all conspiracy offences[3], which equated to 1.0% (1 DP) of the current research Burglary Dwelling data set. Given the size

[3] Conspiring to commit an offence – disregarded from the current research because intended target location not necessarily known.

of the data set, this was deemed to be an acceptable proportion, and also because deletion of incomplete crime records appears to be a conventional approach in academic crime studies, for example, Adderley and Musgrove (2003, p.271) omitted crimes with missing location information from their spatial analysis and Shaw et al. (2010, p.453) disregarded offences with insufficient data from their final analysis. Since the LSOA and OA crime rates were to be presented in terms of number of offences per 1,000 households, burglaries committed in communal establishments, e.g. nursing homes and halls of residence, were removed from the data set because these were not considered to be households in the 2011 Census UK (ONS, 2014), thus better aligning the numerator (police-recorded) and denominator (Census-recorded). It should be noted that this process was not an exact science, being entirely reliant on the quality of "DWELLING" information recorded in the *mo* column.

In a further attempt to more closely align the rates calculation data, duplicate crimes, identified using the criteria: same *date_earliest*, same *date_latest*, same *easting*, and same *northing*, but excluding bedsits and flats, were also deleted. For example, several rooms might be let out on an individual basis in a student house, which would incur one offence per victim if more than one of these was burgled over the same time period (Home Office, 2015), but the property would be viewed as a single household in the Census (ONS, 2014), thereby artificially inflating any crime rates derived using the associated figures. As before, this process relied on "DWELLING" information recorded in the *mo* column, and it should also be noted that the definition of a bedsit is somewhat vague/ contradictory depending on the source consulted so these were left in the data set.

As can be seen from Table 3.4 below, the final current research Burglary Dwelling data set was comprised of 75,088 records, with just over 9 per cent of these being classified as a Car Key burglary. Despite the data limitations already mentioned, the author was relatively confident in the success of the samples selection method because the proportion of Car Key burglaries was similar to that in the following statement from West Yorkshire Police: "Car Key burglaries account for **just under 10%** of burglaries in West Yorkshire" (West Yorkshire Police, 2019, emphasis added). The two crime samples were subsequently used to create the two point shapefiles (outputs 2 & 3), which are analysed later in this chapter, and also to derive the LSOA and OA household rates.

Crime type	**Count**	**Percent of total**
Car Key burglary	6,911	9.2
Regular burglary	68,177	90.8
Total	**75,088**	**100.0**

Table 3.4 - Final burglary sample sizes (Excel spreadsheet and point shapefiles)

Rates Sample Selection Methodology

It is usual when conducting crime analysis to standardise offence counts by some appropriate denominator, thus rendering them directly comparable, as well as helping to mitigate the issues inherent to the Modifiable Areal Unit Problem (MAUP), which is discussed in more detail later. Burglary rates per 1,000 households were therefore calculated for the 1,388 LSOAs and 7,131 OAs in the current work, as is conventional in official statistics reporting, e.g. ONS, and also heeding the advice of Boggs (1965, p.900) that "A valid crime rate (...) should be based on the risk or target group appropriate for each specific crime category". Burglary counts were generated for the LSOA and OA geographies by applying the "Select By Location" tool and count rule "intersect the source layer feature" to the crime point shapefiles in ArcMap 10.4.1. Resulting counts were then divided by the relevant number of households, obtained from the Census 2011 UK, and multiplied by 1,000 to convert them to rates. It should be noted that the intersect method results in double counting where a point is located on a polygon boundary, however, observed differences between the point and intersect counts were deemed negligible.

Repeat and Near-repeat Victimisation Samples Selection Methodology

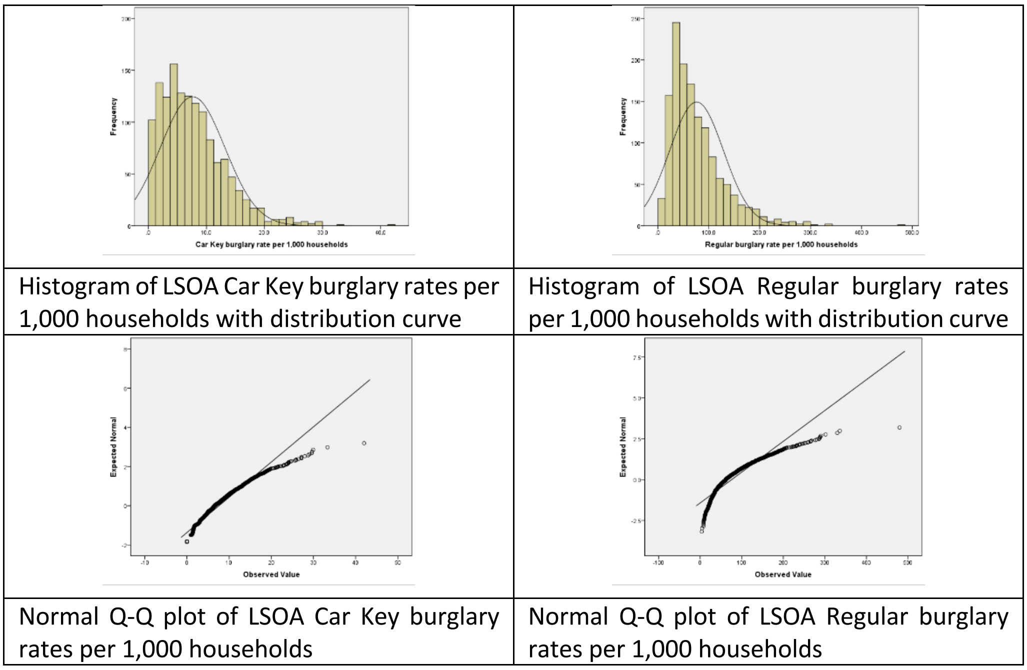
Finally, the police-recorded crime data was subjected to further manipulation for the repeat victimisation/ near-repeat victimisation (RV/ N-RV) analysis that is covered in Chapter 5. Since a key aim of the analysis was to look for overrepresentation of genuine repeats, that is, where offenders return to the scene of a previous crime, any offences that appeared to be part of a temporally clustered crime spree at the same address, such as an apartment block or shared student house, were deleted from the associated samples. The author started with the same 76,554 offences as before (mid-point committed year between 2010 and 2014), deleted any records with missing/ incorrect geocoding, invalid committed between times, and all conspiracy offences (but left communal establishments in this time), sorted the offences by order ascending 'easting', 'northing', 'earliest date & time', 'latest date & time', and then deleted any duplicates, i.e. any offences that matched another offence on all four criteria. Hopefully this approach will have reduced the weighting afforded to non-genuine 'repeats' in the analysis. The final sample sizes for the R V/ N-R V analysis were 6,917 x Car Key burglaries and 68,460 x Regular burglaries.

3.4.3 Impact on Choice of Statistical Tests

Prior to undertaking any analysis, it was necessary to determine whether the crime rates samples followed a normal distribution as this would impact on choice of statistical tests, namely parametric (normally distributed data) or non-parametric (non-normally distributed data) tests. For example, Pearson's Correlation Coefficient is a parametric test, whereas Spearman's Rank Correlation Coefficient is a non-parametric test. 'Kolmogorov-Smirnov' and 'Shapiro-Wilk' normality tests were therefore performed on the crime rates data in IBM SPSS Statistics 22, and all of the distributions were found to be significantly different from normal (see Table 3.5 below – all p values < 0.05 so rejected the null hypothesis of normality). The burglary rates were also analysed in conjunction with histograms and Q-Q plots (see Figure 3.3 below), on the recommendation of Samuels and Marshall (no date), who assert that the two normality tests are overly conservative, i.e. for samples larger than 100 there is a risk that normality might be rejected unnecessarily (Samuels and Marshall, no date, pp.2-3). As can be seen from Figure 3.3, the histograms have long right tales, being positively skewed towards lower values, and the crime rates also deviate from the expected lines of normality on the Q-Q plots.

		LSOA		**OA**	
		Statistic	**Sig.**	**Statistic**	**Sig.**
Car Key burglary	Kolmogorov-Smirnov	.086	.000	.245	.000
	Shapiro-Wilk	.928	.000	-	-
Regular burglary	Kolmogorov-Smirnov	.119	.000	.128	.000
	Shapiro-Wilk	.864	.000	-	-

Table 3.5 - Tests of normality for LSOA and OA Car Key and Regular burglary rates per 1,000 households



Histogram of LSOA Car Key burglary rates per 1,000 households with distribution curve

Histogram of LSOA Regular burglary rates per 1,000 households with distribution curve

Normal Q-Q plot of LSOA Car Key burglary rates per 1,000 households

Normal Q-Q plot of LSOA Regular burglary rates per 1,000 households

Figure 3.3 - Histograms and Q-Q plots of LSOA and OA Car Key burglary and Regular burglary rates

These findings of non-normality are as might reasonably be expected, for example, Ratcliffe (2002, p.32), states that “spatial and temporal crime distributions are rarely, if ever, normally distributed due to the non-random nature of aggregate criminal behaviour”. To explain, assuming that there are only a limited number of locations where a crime can be committed, then this is likely to result in a higher incidence of smaller associated values, than of higher. Rather than transforming the crime rates data, it was decided to use non-parametric tests in the current work, as described in Table 3.6 below. NB All outliers were retained in the crime rates samples and thus included in subsequent analysis.

What testing for	**Dependent data type**	**Independent data type(s)**	**Non-parametric test**
Strength and direction of relationships between crime rates and potential risk factors	Continuous	Continuous/ categorical ordinal	Spearman’s Rank Correlation Coefficient
Significant differences in crime rate distributions across area type groups	Continuous	Categorical nominal	Kruskal-Wallis

Table 3.6 - Non-parametric tests to be performed on crime rates and independent variables

Strength of Association and Significance Levels

It is useful here to state some rules for the interpretation of Spearman’s correlation coefficients and significance levels. Although the directional aspect of coefficients is easy enough to interpret, i.e. > 0 = positive, < 0 = negative, the strength of association is somewhat subjective, for example, should we view a correlation of .50 as being moderate or strong? With this in mind, and in order to standardise the comparison of results throughout the thesis, Cohen’s (1988, pp.79-80) conventions will be used – these are set out in Table 3.7 below.

Correlation coefficient value (+/-)	**Strength of association**
.10 to .29	Small/ weak
.30 to .49	Medium/ moderate
.50 to 1.0	Large/ strong

Table 3.7 - Conventions to be used for correlation coefficient values (Source: Cohen, 1988, pp.79-80)

A significance level is the probability of rejecting the null hypothesis (H_0) when it is actually true. The minimum threshold typically adopted in the social sciences is 0.05 (Glasner et al., 2018, p.4), which equates to a 5% chance of incorrectly rejecting the H_0 (.01 = 1% chance of incorrectly rejecting the H_0 and .001 = 0.1%). The current work will thus view p values ≤ 0.05 as significant.

3.5 Socio-Demographic Data

Having created the two burglary samples, the next step was to identify suitable data sets for testing the hypothesis that Car Key burglaries occur in different area types to Regular burglaries. The availability of socio-demographic data is essential to the identification of residential burglary risk heterogeneity across a study area since population characteristics can determine spatio-temporal convergences of would-be offenders and potential victims.

3.5.1 Census Data

A key source of such data in the UK is the census, a population survey conducted every ten years with the aim of providing a snapshot in time of the nation, including number of people and households. The most recent Census of England & Wales took place on 27th March 2011 and included questions on various topics, including employment, ethnicity, health, marital status, and religion. As well as providing a general description of the population, census-derived information is also used to inform the allocation of public funding and to plan service provision, for example, number of school places. One thing to be mindful of regards census count data, especially for small areas, is that some records are swapped to maintain personal confidentiality.

3.5.2 Geodemographic Classification – 2011 OAC

Although it would have been possible to include every census variable in the PhD analysis, it was expedient to utilise an existing, rigorously derived geodemographic classification, namely the 2011 Output Area Classification (2011 OAC). Since the 2011 OAC was created using K-means clustering (Gale et al., 2016), the final 60 input variables on which it is based can be assumed to parsimoniously differentiate between area types. Further, the classification only uses open source input data, that is, from the Census UK 2011, meaning that the author could easily recreate the variables at LSOA level[4], i.e. to align these with the LSOA burglary rates samples. As per the OAC approach, the LSOA variables were calculated as rates, thus standardising associated counts, and using an appropriate denominator, e.g. % Households with full-time students (denominator households). Spearman's Rank Correlation Coefficient was subsequently used to test the strength of relationships between the LSOA and OA burglary rates and the LSOA and OA 2011 OAC variables respectively, the results of which are presented in Chapter 5.

[4] Except for the Standardised Illness Ratio (SIR) due to unknown method.

Figure 3.4 below illustrates the spatial distributions of a selection of 2011 OAC input variables that were expected to be positively correlated with Car Key burglary rates, e.g. % Households with two or more cars or vans. The general pattern is that the highest rates for each of these variables are situated beyond the main urban centres, e.g. City of Bradford, these being more centrally located within the study area. We might, therefore, expect to find the highest Car Key burglary rates closer to, or in, the more peripheral areas of West Yorkshire.

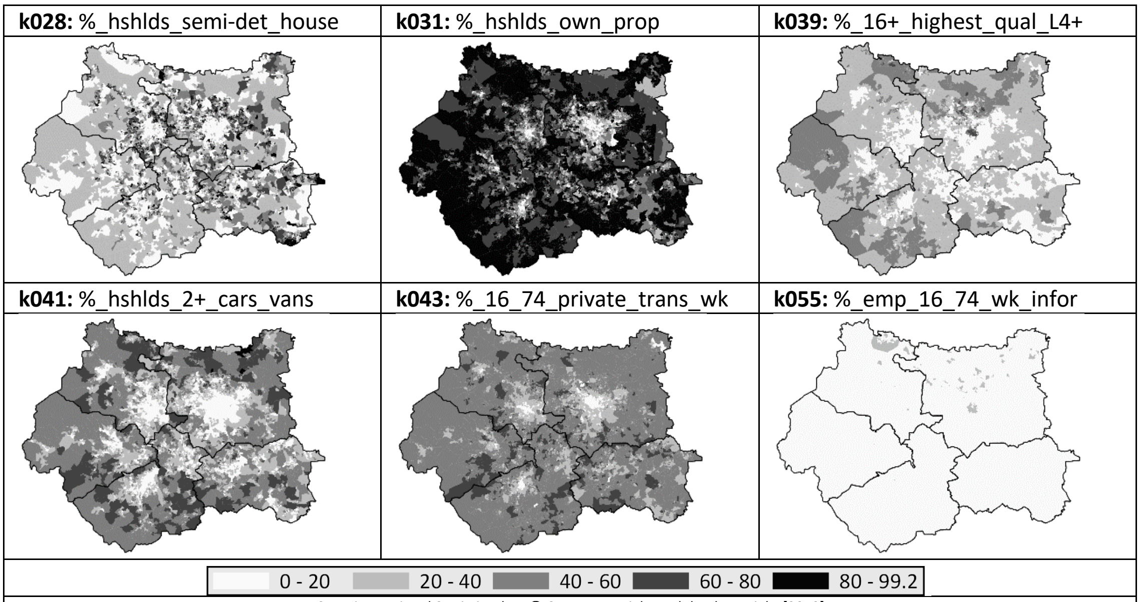
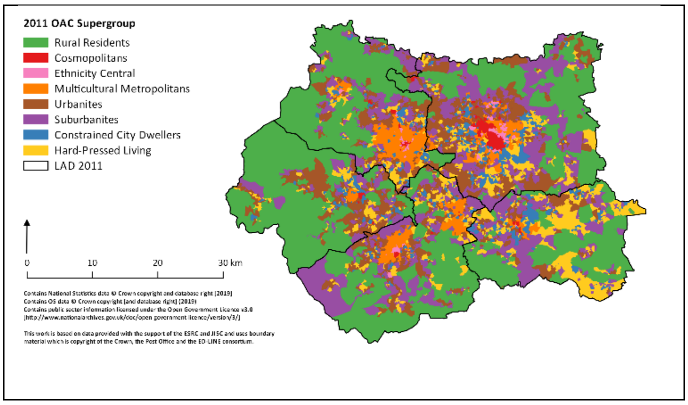

k028: %_hshlds_semi-det_house
k031: %_hshlds_own_prop
k039: %_16+_highest_qual_L4+
k041: %_hshlds_2+_cars_vans
k043: %_16_74_private_trans_wk
k055: %_emp_16_74_wk_infor
0 - 20
20 - 40
40 - 60
60 - 80
80 - 99.2

Figure 3.4 - Spatial distribution within West Yorkshire of select census-derived OAC 2011 input variables

The 2011 OAC is a hierarchical classification, comprised of three tiers, with Supergroups being the most aggregate (mapped in Figure 3.5 below), followed by Groups and then Subgroups. Although the current work will focus primarily on the 2011 OAC Supergroups, information relating to the other two tiers, e.g. Subgroup descriptions, can be found at: http://geogale.github.io/2011OAC/. Table 3.8 below shows the top five variables for each 2011 OAC Supergroup that have a higher value than the standardised UK mean for the variable – this is intended to be viewed in conjunction with Figure 3.5, i.e. to provide an initial indication as to the likely spatial distribution of highest Car Key burglary rates. The information in the table was derived from the 2011 OAC Radial Plots (ONS, 2015). Using the 'two or more cars or vans' variable as an example, and holding all else equal, we might expect to find the highest Car Key burglary rates in the brown, green, and purple areas of the map because the variable has a higher value than the standardised UK mean for the associated Supergroup clusters (Urbanites, Rural Residents, and Suburbanites).



Contains National Statistics data © Crown copyright and database right [2019]
Contains OS data © Crown copyright [and database right] (2019)
Contains public sector information licensed under the Open Government Licence v3.0
[http://www.nationalarchives.gov.uk/doc/open government licence/version/3/]

This work is based on data provided with the support of the ESRC and JISC and uses boundary material which is copyright of the Crown, the Post Office and the ED-LINE consortium.

Figure 3.5 - 2011 Output Area Classification Supergroups

🟩		**Rural Residents**	🟥		**Cosmopolitans**
	1	Work in agriculture, forestry and fishing		**1**	Flats
	2	Detached house/bungalow		**2**	Full-time students
	3	2+ cars/vans		**3**	Private renting
	4	Age 65-89		**4**	Born in old EU
	5	No children		**5**	Overcrowding
▨		**Ethnicity Central**	🟧		**Multicultural Metropolitans**
	1	Black/African/Caribbean/Black British		**1**	Black/African/Caribbean/Black British
	2	Flats		**2**	Pakistani
	3	Mixed ethnic group		**3**	Indian
	4	Overcrowding		**4**	English or Welsh not main language
	5	Social renting		**5**	Chinese and other
🟫		**Urbanites**	🟪		**Suburbanites**
	1	Flats		**1**	Detached house/bungalow
	2	Terrace or end-terrace		**2**	2+ cars/vans
	3	Private renting		**3**	Owned or shared ownership
	4	Work in information and communication		**4**	Semi-detached house/bungalow
	5	2+ cars/vans		**5**	Married or civil partnership
🟦		**Constrained City Dwellers**	🟨		**Hard-Pressed Living**
	1	Flats		**1**	Social renting
	2	Social renting		**2**	Terrace or end-terrace
	3	Overcrowding		**3**	Semi-detached house/bungalow
	4	Unemployed		**4**	Unemployed
	5	Divorced or separated		**5**	Non-dependent children

Table 3.8 - Top five variables for each 2011 OAC Supergroup that have a higher value than the standardised UK mean for the variable (based on ONS, 2015)

In addition to the variables already discussed, the author felt that some Regular burglary risk factors, including those inherent to the concepts of social disorganisation and collective efficacy, as discussed in Chapter 2, were not sufficiently represented within the 60 x 2011 OAC variables. A further twelve high-level variables were therefore created for analysis with the burglary rates, as listed in Table 3.9 below. It was also deemed that, due to the inherently mobile nature of Car Key burglary, area accessibility should also be incorporated into the analysis. For example, some parts of West Yorkshire might be more vulnerable to Car Key burglary due to their relative 'closeness' to all other locations within the study area (measured using street node closeness centrality), or because of their proximity to assumed high offender rates areas – Chapter 4 is wholly dedicated to these two variables so no more will be said about them here.

What measuring	Variable	Data source
Household vulnerability	% Households lone parent with dependent children (also represents family disruption re social disorganisation)	Census UK 2011
	% Household reference persons aged 24 and under	
Residential mobility	% Persons lived at different address one year ago *Derived from lived at same address one year ago (usual residents) – Table UKMIG008*	
Ethnic heterogeneity	Index of Diversity *(created by author)*	
Socio-economic classification (proxy affluence)	% Persons aged between 16 and 74 NS-SeC Higher	NS-SeC *(author employed three classes approach)*
	% Persons aged between 16 and 74 NS-SeC Intermediate	
	% Persons aged between 16 and 74 NS-SeC Routine & Manual	
Relative disadvantage	Index of Multiple Deprivation 2015 Decile (categorical ordinal data)	IMD 2015 *(author assigned LSOA level data to nested OAs)*
	IMD 2015 Employment Deprivation Domain score	
	IMD 2015 Income Deprivation Domain score	
Area accessibility	Area type juxtapositions (7 x sub-variables)	Census UK 2011
	Average closeness centrality (4 x sub-variables)	OpenStreetMap

Table 3.9 - Additional twelve high-level variables for analysis with burglary rates

Of the variables listed in Table 3.9 above, those that probably warrant further explanation are Index of Diversity, NS-SEC, and IMD 2015.

3.5.3 Index of Diversity

Census data (Table KS201EW) was used to create the Index of Diversity following the method outlined in Norman (no date). An Index of Diversity describes the level of ethnic variation within an area. The Index is relatively simple to calculate; the within-area (LSOA/ OA) percentage of each ethnic group is divided by 100, squared, and the resulting values summed and then subtracted from 1. So, taking the five ethnic groups used in the current work, 'White', 'Mixed/multiple ethnic groups', 'Asian/Asian British', 'Black/African/Caribbean/Black British', and 'Other ethnic group', assuming an equal split across these groups, the resulting Index of Diversity score would be 0.80. Low values indicate ethnic homogeneity and the lowest possible value is zero, i.e. no variation (Norman, no date).

3.5.4 NS-SeC

In the absence of any census data relating to income, the occupation-based 'National Statistics Socio-economic Classification (NS-SEC)' was instead used a proxy for affluence. Associated data, which was downloaded from https://www.nomisweb.co.uk/census/2011/ks611ew, relates to the usual resident population aged 16 to 74 as at census day, 2011, by current/ last main job. The NS-SeC can be operationalised as either eight, five, or three analytic classes (ONS, ca.

2016b), with the latter version used here for simplicity, namely: (i) Higher, (ii) Intermediate, and (iii) Routine & Manual. The eight to three classes aggregation process was informed by ONS (ca. 2016b).

3.5.5 Indices of Deprivation 2015

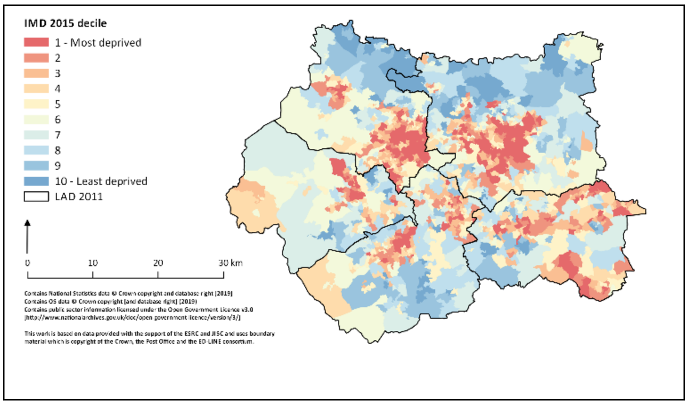
The Index of Multiple Deprivation 2015 (IMD 2015) is one of a number of relative deprivation measures that make up the English Indices of Deprivation 2015 for small areas (LSOAs). The IMD 2015 provides a summary measure of seven different deprivation domains, including Income and Employment, which are listed in Table 3.10 below. Each domain is weighted in the overall Index, with Income and Employment being afforded the greatest weights at 22.5% each. Various indicators are referenced within each domain, for example, the Income Deprivation Domain includes 'adults and children in Income Support families' and the Employment Deprivation Domain includes 'Claimants of Jobseeker's Allowance' (Smith et al., 2015, p.19). The majority of the data on which the IMD is based is from the 2012/ 2013 tax year (Smith et al., 2015, p.13).

Domain	**Domain weight (%)**
Income Deprivation Domain	22.5
Employment Deprivation Domain	22.5
Health Deprivation & Disability Domain	13.5
Education, Skills & Training Deprivation Domain	13.5
Crime Domain	9.3
Barriers to Housing & Services Domain	9.3
Living Environment Deprivation Domain	9.3

Table 3.10 - Weights for individual domains in the IMD 2015 (Source: Smith et al., 2015, p.18)

The IMD 2015 is available in rank, decile, and score format, however, from an analytical perspective, it should be noted that the scores are less meaningful than the first two measures. This is because the ranks and deciles represent relative deprivation, for example, a LSOA ranked at 50 in the Index can be viewed as being more deprived than a LSOA ranked at 500, with 1 being the most deprived and 32,844 being the least deprived, but we cannot say by how much. Decile 1 relates to the most deprived 10% of LSOAs in the country, and Decile 10 to the least. The Employment and Income Deprivation Domains, but not the other five Domains, are rates. This means that they represent the proportion of the associated population that is experiencing that type of deprivation (Smith et al., 2015, p.21), for example, a score of 0.35 on the Income Deprivation Domain equates to 35% being income deprived. Because the IMD 2015 scores cannot be interpreted simply as the proportion of the population that is deprived, Smith et al. (2015, p.21) state that "It is recommended that ranks and deciles, but not scores, are used in the case of the Index of Multiple Deprivation". For these reasons, the current work will utilise the IMD 2015 Deciles and the Income and Employment Deprivation Domain scores. Relevant data was downloaded from https://www.gov.uk/government/statistics/english-indices-of-deprivation-2015, and the Deciles were subsequently mapped, as shown in Figure 3.6 below.

Given the target property type, it was anticipated that as the Income and Employment Deprivation Domain scores (proportions) increased, Car Key burglary rates would decrease because desirable vehicles were expected to be less prevalent in the associated areas. Conversely, as the IMD 2015 Deciles increased, Car Key burglary rates were expected to increase because the target property type was expected to be more prevalent in the associated areas (recall that lower Deciles = more deprived so the higher deciles might, therefore, better reflect the distribution of the target property type). With this in mind, and holding all else equal, the light yellow and blue LSOAs, including to the north of Leeds and Bradford, might be most vulnerable to Car Key burglary offences.



Contains National Statistics data © Crown copyright and database right [2019]
Contains OS data © Crown copyright [and database right] (2019)
Contains public sector information licensed under the Open Government Licence v3.0
(http://www.nationalarchives.gov.uk/doc/open government licence/version/3/)

This work is based on data provided with the support of the ESRC and JISC and uses boundary material which is copyright of the Crown, the Post Office and the ED-LINE consortium.

Figure 3.6 - 2015 Index of Multiple Deprivation Deciles (LSOA level)

Final Variables Summary

Since numerous variables have been covered in this section, a high-level summary of these is provided in Table 3.11 below. Further to the variables already discussed, the eight 2011 OAC Supergroup classifications will be analysed in conjunction with the OA burglary rates to ascertain if there are any statistically significant variations between area types and their associated rates. Kruskal-Wallis will be used to this end because the data is categorical nominal and thus not suitable for analysis with Spearman's Correlation.

Final LSOA level variables	**Final OA level variables**
Car Key and Regular burglary rates	Car Key and Regular burglary rates
59 x 2011 OAC variables (not SIR)	60 x 2011 OAC variables
All variables in Table 3.9	All variables in Table 3.9
n/a	2011 OAC Supergroup classification

Table 3.11 - Final variables to be analysed with the burglary rates at LSOA/ OA levels

3.6 Potential Issues for Consideration

3.6.1 Modifiable Areal Unit Problem (MAUP)

The Modifiable Areal Unit Problem (MAUP) relates to the risk of false inference that can arise from the aggregating of different phenomena, such as crime counts and population attributes, to socially-constructed areal units, e.g. Output Areas (OAs) and Local Authority Districts (LADs). Fotheringham and Rogerson identify the problem as being a key impediment to spatial analysis (1993, p.4) and Flowerdew (2011) notes different correlation coefficients between pairs of census variables depending on the type of areal delineator used. Recognising potential impacts of MAUP on the current research, including loss of detail regards the underlying population, correlation analysis will be performed at both the OA and LSOA levels, and the results examined in the context of the literature, i.e. do the observed correlation coefficients reflect the theory? It should also be noted that boundary locations can affect hot spot identification, for example, a police force boundary might intersect a crime hot spot, but those offences committed in a different police force area might not be identified as belonging to this. Given that crime data is only available for West Yorkshire in the current work, this will be an issue, e.g. for KDE mapping.

3.6.2 Ecological Fallacy

The ecological fallacy again relates to the risk of false inference that can arise when individuals within a population are aggregated to abstract areal units, and specifically the assumption that the prevailing characteristics of a particular area can be applied to all persons within that area (Chainey and Ratcliffe, 2005, pp.152-153). Since publicly available socio-demographic data sets, including UK census data, are typically aggregated to areal units to maintain the privacy of the individuals to which they relate, the implication of this is that associated area level classifications/ indices can mask within-area heterogeneity. For example, the IMD 2015 might identify a LSOA as being more deprived than other LSOAs within a study area, but it does not necessarily follow that all individuals within that area experience the same level of deprivation. MAUP and the ecological fallacy both result from the requirement to balance spatio-temporal disaggregation, and the potential insights that this can offer, with an individual's right to privacy, as well as the need to present information in a way that is relevant to the intended audience, for example, the police might align crime statistics to operational geographies, such as districts.

3.6.3 Edge Effects

Edge effects can be spatial or temporal. A spatial edge effect was described in the MAUP section, namely that areal boundaries, such as police force areas, can intersect crime hot spots. Another example, provided by Bailey and Gatrell (1995, p.90), is that nearest neighbour distance will be biased – typically greater – for points near the boundary of a study area because neighbours situated beyond the boundary are not considered. Although Bailey and Gatrell (1995, pp.76-77; p.90) advise leaving a “guard area” to mitigate spatial edge effects, that is, analyse a sub-area within a study area but incorporate the guard area in calculations, this was not deemed suitable here because a smaller spatial extent was unlikely to fully capture mobile offending patterns. This is supported by Youstin et al. (2011, p.1043), who, referencing Wiles and Costello (2002), suggest that vehicle theft near repeat patterns might be more “widespread spatially” due to specific cars being targeted under more organised/ professional operations, e.g. ‘chop shops’. For this reason, the thesis uses data relating to the whole of the West Yorkshire study, but not for any neighbouring areas because crime data was not available for these.

Temporal edge effects are conceptually similar to spatial edge effects in that choice of boundary can lead to the exclusion of pertinent information, for example, a crime hot spot might span two time periods, but relevant offences would be missed if only one was analysed. In terms of repeat victimisation, addresses that are targeted both before and during a study period will not be identified as exact repeats unless data for a longer time period than is being analysed is used to identify initial events, as per the approach of Bernasco et al., 2015 (pp.123-124). Hopefully, the decision to analyse five years’ worth of crime data in the current research should go some way to mitigating such issues.

3.6.4 Privacy

Privacy of the individuals to which a research data set relates should be of optimum concern to a researcher and, for that reason, crime rates data will only be mapped at the LSOA level here. Individual crime points will only be mapped in KDE format, which is deemed by the author to be safe because actual density values are not supplied with the maps and the crime point values have been ‘spread’ by the kernel function. The crime data to which this research relates has been stored on a secure drive and relevant protocol has been followed regards its handling and the dissemination of relevant findings.

3.7 Exploratory Spatial Analysis

This section presents the findings of exploratory spatial analysis that was undertaken to obtain an initial empirical indication as to whether or not the research rationale was likely to be valid. Acknowledging that crime is not randomly distributed, and also considering the characteristics of typical target area types for Regular burglary, if the spatial distributions for Car Key burglary and Regular burglary differ, then it is reasonable to infer that the associated spatio-temporal patterns (RV/ N-RV) for each of these burglary types will also differ. A potential impact of this, and one that is particularly relevant here, is that prevailing burglary modelling techniques, such as heuristic buffering of recent offences, might not be as effective at predicting future Car Key burglary locations as at predicting future Regular burglary locations.

Prior to performing the analysis, it was anticipated that Regular burglaries would exhibit more spatial clustering than Car Key burglary offences, this being the result of locally-anchored offenders being well-placed to return to the locations/ general vicinity of their past offences. Recall that Boggs (1965, p.904) identified a strong, positive correlation (.762) between residential day burglary rates and burglary offender rates for 128 census tracts in St Louis, US. Car Key burglaries were expected to be more spatially dispersed than Regular burglaries, with moderating factors, such as collective efficacy levels within previously victimised areas, impacting offenders' longitudinal mobility patterns, i.e. causing them to seek out new target locations.

Two relatively simple measures, standard deviational ellipses and average nearest neighbour, will first be used to look for evidence of spatial clustering in the two burglary point data sets. Standard Deviational Ellipses show the directional dispersal of an XY coordinate data set about the mean (geographical) centre of the data set (Chainey and Ratcliffe, 2005, pp.120-126) – for a spatial normal distribution, i.e. where there are more features at the centre than at the periphery, 1 standard deviation ellipse polygon will cover approximately 68% of the points (ESRI, c2019). Average Nearest Neighbour is fairly self-explanatory, being simply the mean of the distances between each point's centroid and its nearest neighbour's centroid in an XY data set. Global Moran's I will then be used to look for evidence of spatial-autocorrelation in the burglary rates data sets. This statistic is inferential in nature so it will not provide any information regards local clustering patterns, simply whether burglary rates are clustered within the study area. Supplementary to the statistical measures, LSOA rates for the two burglary types will be scaled between 0-1 and mapped as a choropleths, thus facilitating visual comparisons between these, and a further three map pairs (referred to hereafter as 'shine through' maps) will be created to show underlying distributions of the IMD 2015 Deciles, 2011 OAC Supergroups, and 2011 Urban Rural Classification (OA) for those LSOAs with a scaled crime rate in one of the top three classes (the 0.00-0.25 class will be masked white). These maps were influenced by the 'DataShine' mapping work of O'Brien and Cheshire (2016).

Ranked LSOA burglary rates per 1,000 households by IMD 2015 decile will next be presented in table format, employing the same colour-scheme used as in the other IMD maps, and a further two tables will show ranked OA burglary rates per 1,000 households by the 2011 OAC Supergroups and 2011 Urban Rural Classification (OA) classes. The associated rates were fairly straightforward to derive, being simply the total number of burglary offences in each area in each class, divided by the total number of households in each area in each class, and then multiplied by 1,000. Although the information in the tables is useful from a summary perspective, it should be noted that the approach is vulnerable to the Ecological Fallacy.

After a brief review of the Kruskal-Wallis test results – performed with the OA burglary rates and the 2011 OAC Supergroups in IBM SPSS Statistics 22 – the last part of this section will be used to discuss the findings of two **hot spot identification** techniques, namely Kernel Density Estimation (KDE) (applied to point data) and Getis-Ord Gi* statistic (applied to area data). KDE was explained in Chapter 2. The maps presented here were produced using the QGIS Desktop 2.14.0 Heatmap plugin – note that the Heatmap kernel function visits individual crimes, as opposed to cell centroids (QGIS, no date). The Gi* statistic identifies statistically significant hot spots and cold spots within a study area by comparing the sum of values for a given feature ***i*** and its neighbours ***j***, up to a user-specified distance, to all values in the study area (Chainey and Ratcliffe, 2005, p.165; Chainey, 2014, p.62). Gi* differs from the Gi statistic only on the basis that ***i*** is included in the calculations (Chainey and Ratcliffe, 2005, p.166). Resulting Gi* z-score significance levels for LSOAs will be presented in choropleth map format, and shine through maps will again be used to show underlying distributions of the 2011 OAC Supergroups, IMD 2015 Deciles, and 2011 Urban Rural Classification (OA), but this time for those LSOAs that are significant hot spots (cold spots/ not significant areas will be masked white).

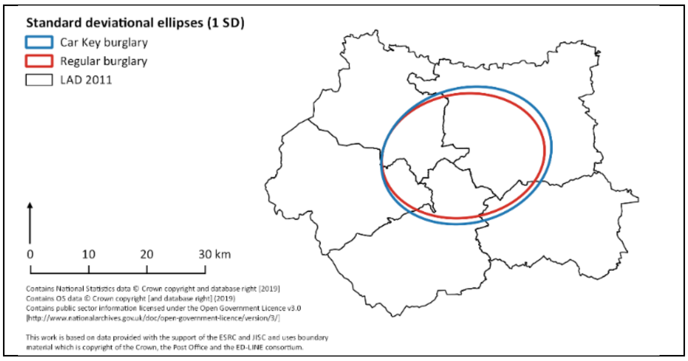
The Standard Deviational Ellipses, Average Nearest Neighbour, Global Moran's I, and Getis-Ord Gi* statistic analysis was performed in ArcMap 10.4.1, and spatial weights matrices were created to define approximate neighbourhoods for the Global Moran's I and Getis-Ord Gi* tools. On the advice of ESRI (no date), the threshold distance for each matrix was set to the average distance to eight nearest neighbours per feature, namely 1612.1 metres for LSOAs and 604.3 m for OAs (see Table 3.12 below). These distances generated an average of ~ 11 neighbours per feature, at least one, and always less than 50, which was thought acceptable (the threshold distance was automatically extended where necessary to ensure at least one neighbour per feature).

Distance (m)	**LSOAs – 1NN**	**LSOAs – 8NN**	**OAs – 1NN**	**OAs – 8NN**
Min	64.9	562.2	23.8	186.1
Ave	688.0	**1612.1**	255.9	**604.3**
Max	4415.5	6986.2	3983.3	5215.4

Table 3.12 - Nearest neighbour distances for the spatial weights matrices

3.7.1 Clustering of Crime Points

Figure 3.7 below shows the standard deviational ellipses (1 SD) that were generated for the Car Key burglary (blue) and Regular burglary points (red). As expected, the Car Key burglary points are more dispersed about the mean centre than the Regular burglary points, particularly around north/ north-east/ east Leeds, which could signal offenders targeting more suburban locations. The approximately central position of the two ellipses within the study area might be an artefact of the underlying resident population distribution, i.e. more people = more properties to burgle.



Contains National Statistics data © Crown copyright and database right [2019]
Contains OS data © Crown copyright [and database right] (2019)
Contains public sector information licensed under the Open Government Licence v3.0
[http://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/]

This work is based on data provided with the support of the ESRC and JISC and uses boundary material which is copyright of the Crown, the Post Office and the ED-LINE consortium.

Figure 3.7 - Standard deviational ellipses (1 SD) for Car Key burglary and Regular burglary

Table 3.13 below contains the average nearest neighbour (NN) results. As expected, the mean nearest neighbour distance for the Car Key burglary points (133.5 metres) is greater than for the Regular burglary points (32.1 metres). This suggests that Car Key burglaries are more spatially dispersed than Regular burglaries, although the possible influence of edge effects should be borne in mind. The NN Ratio figures were calculated by dividing the observed mean by the expected mean (derived from a random distribution of the same number of points) – a result less than 1 represents clustering (ESRI, c2018). The NN Ratios and associated p-values for each data set indicate that both of these exhibit clustering, however, the Car Key burglary points would appear to be less clustered than the Regular burglary points because the associated NN Ratio figure is closer to 1.

	Observed mean (m)	**Expected mean (m)**	**NN Ratio**	**z-score**	**p-value**
Car Key burglary	133.509470	312.195392	0.427647	-91.025941	0.000000
Regular	32.053393	101.214188	0.316689	-341.325243	0.000000

Table 3.13 - Average nearest neighbour results for Car Key burglary and Regular burglary

3.7.2 Clustering of Crime Rates

The Global Moran's I results shown in Table 3.14 and Table 3.15 below indicate that the rates for both burglary types exhibit more spatial clustering of high values and/or low values than would be expected on the basis of random chance (p = 0.00) (ESRI, c2018b). However, because the Moran's Index ranges from -1 to +1, it would again appear that the clustering is less pronounced within the Car Key burglary sample (0.3 vs 0.7 for Regular burglary).

	Moran's Index	**z-score**	**p-value**
Car Key burglary	0.324754	21.100452	0.000000
Regular	0.662368	43.026154	0.000000

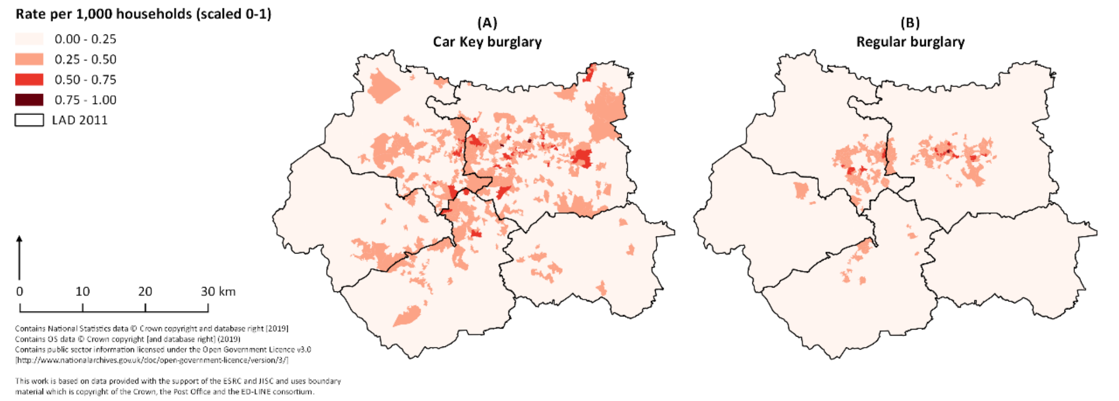
Table 3.14 - Global Moran's I LSOA results for Car Key burglary and Regular burglary

	Moran's Index	**z-score**	**p-value**
Car Key burglary	0.162384	24.974819	0.000000
Regular	0.546632	84.077050	0.000000

Table 3.15 - Global Moran's I OA results for Car Key burglary and Regular burglary

Due to some of the Regular burglary rates being much higher than for Car key burglary, it was decided to min-max scale all of the rates between 0 and 1 (newx = (x – xmin)/(xmax – xmin)), thus rendering the associated spatial distributions directly comparable on maps. As can be seen from Figure 3.8 below, there are notable differences between the LSOA distributions of highest/ lowest Car Key burglary rates (map A) and Regular burglary rates (map B). The distribution of LSOAs in classes 2, 3, and 4 (scaled rates 0.25-1.00) is far more clustered for Regular burglary than for Car Key burglary, with the highest rates in areas around Leeds and Bradford city centres. The LSOA with the highest Regular burglary rate falls within the Hyde Park and Woodhouse ward, Leeds, which is in line with the findings of similar research conducted by Malleson (2010, pp.60-61), and gives additional confidence in the data samples selection method. It should be noted that, for the LSOA in question, the per cent of households with full-time students as at the 2011 Census was 30.8%, versus a study area average of 0.7%, which is not surprising given that the variable was identified as a household burglary risk factor by Budd (2001, p.2). Interestingly, the highest Car Key burglary rate LSOA is also located in north Leeds, albeit farther from the city centre than the highest Regular burglary rate LSOA. As at 2011, the per cent of households in the LSOA with full-time students was 0.2%, it had a population density greater than the study area average, a below average per cent of households who were social renting, and an above average per cent of households who owned or had shared ownership of property.

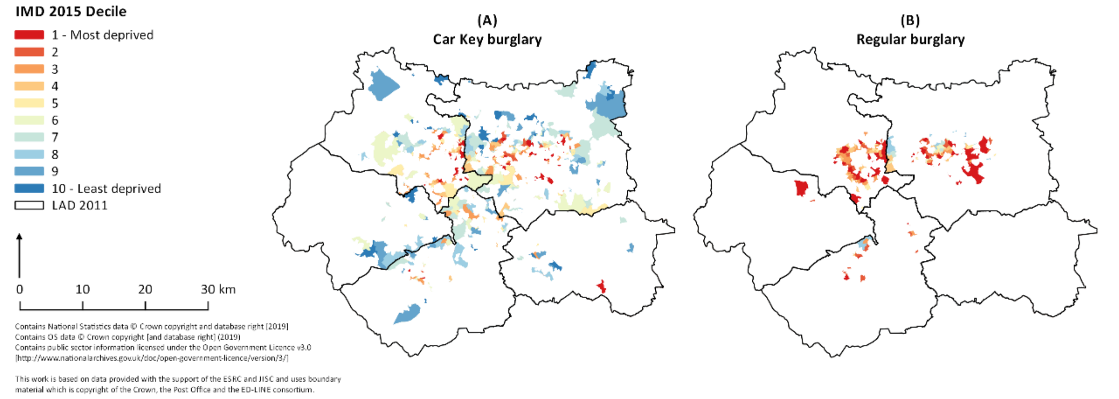
Looking next at the shine through maps for the LSOA scaled rates classes 2, 3, and 4 (0.25-1.00), there are also clear differences between the two IMD 2015 shine through maps (Figure 3.9), with less deprived LSOAs appearing to dominate for Car Key burglary, but more deprived for Regular burglary. However, it should be noted that there are some anomalies to this pattern, particularly for Car Key burglary, with some of the more deprived LSOAs making an appearance. This might be a result of the crime samples selection method, i.e. ‘non-professional’ Car Key burglaries ending up in the associated sample, for example, if a vehicle was stolen for the purpose of joy riding, or to transport goods away from the scene of an offence (the absence of any data re vehicles’ ages, makes, and models was discussed earlier). A possible impact of this is that the correlation coefficients between crime rates and independent variables might not be as strong/ weak as expected, but this cannot be avoided with the data that is currently available. Given that the 2011 OAC and Rural Urban maps depict output area classifications, they are included to provide background context only. Taking the two 2011 OAC maps first (Figure 3.10), the Supergroups that seem to stand out the most for Car Key burglary are “Rural Residents” (green), “Suburbanites” (purple), and “Urbanites” (brown). For Regular burglary, “Constrained City Dwellers” (blue), “Cosmopolitans” (red), “Hard-Pressed Living” (yellow), and “Multicultural Metropolitans” (orange) appear more dominant. It is important to note, however, that the observed patterns could be an artefact of the choropleth mapping technique. The Rural Urban maps (Figure 3.11) indicate that the highest Regular burglary rates are likely to be found in urban major conurbation areas, whereas Car Key burglars appear to target both rural and urban areas.



Contains National Statistics data © Crown copyright and database right [2019]
Contains OS data © Crown copyright [and database right] (2019)
Contains public sector information licensed under the Open Government Licence v3.0
[http://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/]

This work is based on data provided with the support of the ESRC and JISC and uses boundary material which is copyright of the Crown, the Post Office and the ED-LINE consortium.

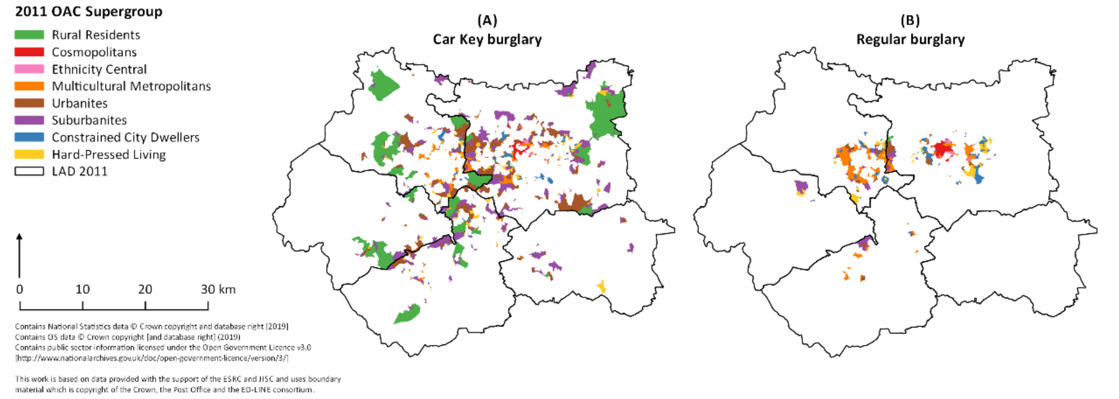
Figure 3.8 - LSOA (A) Car Key and (B) Regular burglary rates per 1,000 households scaled 0-1



Contains National Statistics data © Crown copyright and database right [2019]
Contains OS data © Crown copyright [and database right] (2019)
Contains public sector information licensed under the Open Government Licence v3.0
[http://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/]

This work is based on data provided with the support of the ESRC and JISC and uses boundary material which is copyright of the Crown, the Post Office and the ED-LINE consortium.

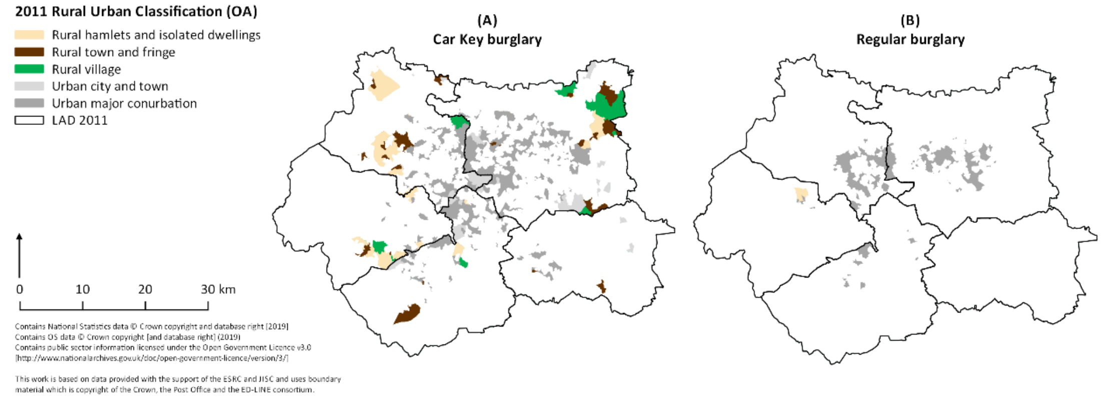
Figure 3.9 - IMD 2015 Decile shine through for scaled classes 0.25-1.00 (0.00-0.25 masked white)



Contains National Statistics data © Crown copyright and database right [2019]
Contains OS data © Crown copyright [and database right] (2019)
Contains public sector information licensed under the Open Government Licence v3.0
[http://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/]

This work is based on data provided with the support of the ESRC and JISC and uses boundary material which is copyright of the Crown, the Post Office and the ED-LINE consortium.

Figure 3.10 - 2011 OAC Supergroup shine through for scaled classes 0.25-1.00 (0.00-0.25 masked white)



Contains National Statistics data © Crown copyright and database right [2019]
Contains OS data © Crown copyright [and database right] (2019)
Contains public sector information licensed under the Open Government Licence v3.0
[http://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/]

This work is based on data provided with the support of the ESRC and JISC and uses boundary material which is copyright of the Crown, the Post Office and the ED-LINE consortium.

Figure 3.11 - 2011 Rural Urban Classification (OA) shine through for scaled classes 0.25-1.00 (0.00-0.25 masked white)

Since the interpretation of shine through maps is subjective, supplementary figures are provided in Table 3.16, Table 3.17, and Table 3.18 below. Table 3.16 confirms the general IMD 2015 patterns identified previously, with LSOAs in the least deprived deciles experiencing the highest overall Car Key burglary rates and those in the most deprived deciles the lowest, whereas the reverse is true for Regular burglary. Table 3.17 and Table 3.18 contain Output Area burglary rates per 1,000 households and are therefore a better representation of reality than the respective shine through maps, which were based on LSOA burglary rates.

IMD 2015 decile	**Car Key burglary rate per 1,000 households (1 DP)**	**Rank**		**Regular burglary rate per 1,000 households (1 DP)**	**Rank**
1 – most deprived	5.4	10		108.3	**1**
2	5.7	9		84.2	**2**
3	7.1	8		81.0	**3**
4	7.5	7		76.6	4
5	8.4	5		65.0	5
6	8.3	6		59.8	6
7	9.1	4		56.4	7
8	9.2	**2**		55.2	8
9	9.2	**2**		46.2	9
10 – least deprived	9.4	**1**		41.9	10

Table 3.16 - Ranked LSOA Car Key burglary and Regular burglary rates per 1,000 households by IMD 2015 decile

As can be seen from Table 3.17 below, OAs in the 2011 OAC Supergroup “Suburbanites” (1) experienced the highest overall Car Key burglary rate, followed by “Urbanites” (2), whereas for Regular burglary the order was “Multicultural Metropolitans” (1) and then “Cosmopolitans” (2). Recalling from Table 3.8 that the “Suburbanites” and “Urbanites” Supergroups both had a higher value than the standardised UK mean for the ‘2+ cars/vans’ variable, these two classes having the highest overall Car Key burglary rates makes sense. Another potential insight here is that the only other Supergroup to have a higher value than the standardised UK mean for the ‘2+ cars/vans’ variable was “Rural residents” – interestingly, this class ranked 6th in terms of its overall Car Key burglary rate at OA level, which suggests that area accessibility and density of potential victims might be influencing offender mobility patterns. Many of the “Rural Residents” LSOAs are located towards the edge of the study area and, thus, some distance from the urban centres where we might expect to find the highest offender rates. Also considering the other four variables with a higher value than the standardised UK mean for this Supergroup (work in agriculture, detached house, aged 65-89, no children), it might be that desirable vehicles are not prevalent enough to warrant the associated journey to crime, for example, pickup trucks might be more common in agricultural areas than performance vehicles.

Looking at the Pen Portrait (summary description) for “Multicultural Metropolitans”, these OAs are likely to be “concentrated in larger urban conurbations in the transitional areas between urban centres and suburbia” (ONS, 2015b, p.10). Thus, this Supergroup ranking 1st for Regular burglary is perhaps unsurprising when considered in the context of the Chicago School’s findings regards delinquency and the ‘zone in transition’ (e.g. see Burgess, 1925 in Chainey and Ratcliffe, 2005, p.83; Shaw and McKay, 1942). Given the earlier discussion about student households, the fact that “Cosmopolitans” also ranked highly for Regular burglary is again as expected, with the associated 2011 OAC Pen Portrait stating that “There are also higher proportions of full-time students” (ONS, 2015b, p.6). Leeds, Bradford, and Huddersfield all have universities, and students typically possess ‘CRAVED’ items (Clarke, 1999), such as laptops and mobile phones, which are easy to conceal, remove, and dispose of. Further, students might leave their residences without a capable guardian for long periods of time, for example, when attending lectures or socialising, and the urban settings of the study area universities are also likely to render them more accessible to offenders.

2011 OAC Supergroup	**Car Key burglary rate per 1,000 households (1 DP)**	**Rank**		**Regular burglary rate per 1,000 households (1 DP)**	**Rank**
Rural Residents	5.9	6		34.8	8
Cosmopolitans	7.3	**3**		117.5	**2**
Ethnicity Central	2.8	8		95.5	**3**
Multicultural Metropolitans	6.9	4		120.1	**1**
Urbanites	8.8	**2**		60.3	6
Suburbanites	10.2	**1**		54.5	7
Constrained City Dwellers	4.2	7		73.7	4
Hard-Pressed Living	6.3	5		62.4	5

Table 3.17 - Ranked OA Car Key burglary and Regular burglary rates per 1,000 households by 2011 OAC Supergroup

As might reasonably be expected, the highest overall Regular burglary rate per 1,000 households was for OAs in urban major conurbations, as shown in Table 3.18 below. Interestingly, this was also the case for Car Key burglary, which again indicates that area accessibility, both generally, and relative to offenders’ home locations, might exert a strong influence on crime distributions.

2011 Rural Urban Classification	**Car Key burglary rate per 1,000 households (1 DP)**	**Rank**		**Regular burglary rate per 1,000 households (1 DP)**	**Rank**
A1 – Urban major conurbation	8.1	**1**		86.3	**1**
C1 – Urban city and town	5.4	**3**		40.4	**2**
D1 – Rural town and fringe	6.3	**2**		30.0	5
E1 – Rural village	5.0	4		35.5	**3**
F1 – Rural hamlets and isolated dwellings	4.9	5		34.2	4

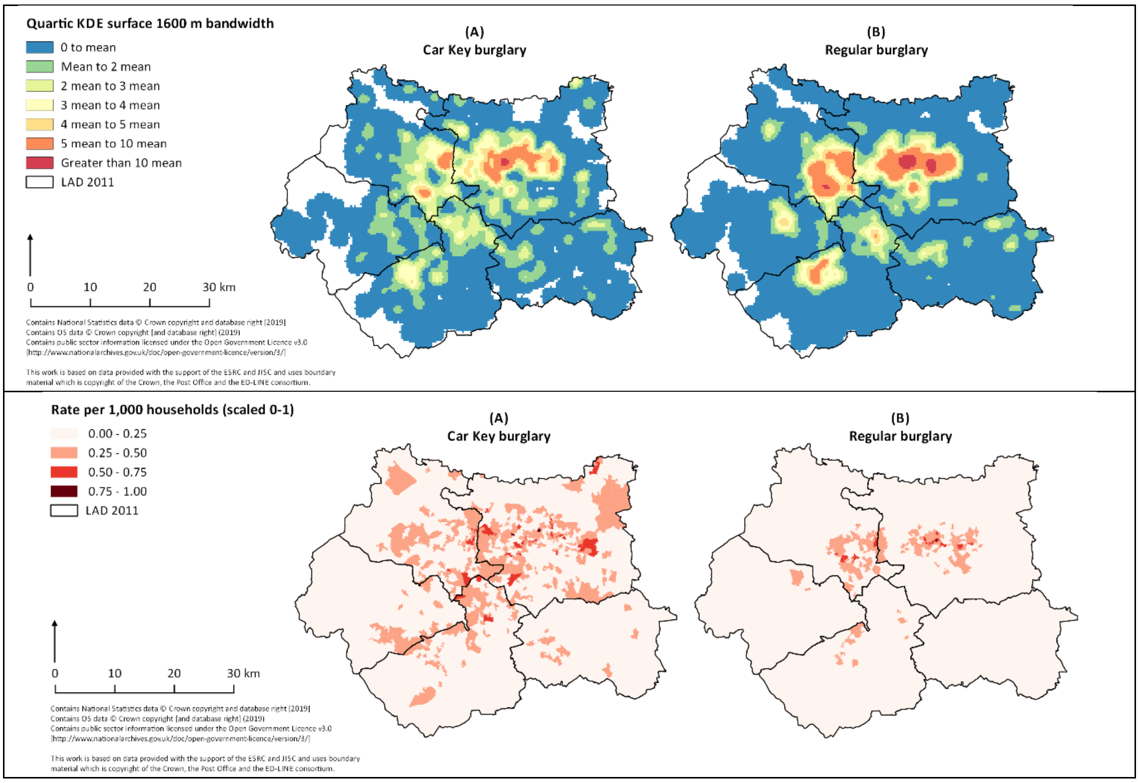
Table 3.18 - Ranked OA Car Key burglary and Regular burglary rates per 1,000 households by 2011 Rural Urban Classification

The Kruskal-Wallis test identified that the distribution of OA Car Key burglary rates was not the same across each of the eight 2011 OAC Supergroup classes, with a p-value of .000 allowing the null hypothesis of no significant difference between the mean ranks to be rejected (McDonald, 2014, pp.157-164; Samuels and Marshall, no dateb). This was also the case for the OA Regular burglary rates. Focusing only on the Car Key burglary rates and the "Suburbanites" and "Urbanites" classes, Dunn-Bonferroni post-hoc testing found significant differences between these two classes and all other classes, including each other (Samuels and Marshall, no dateb). It is also useful to note that the median Car Key burglary rate varied across the classes, with a median of 8.0 for "Suburbanites", 7.4 for "Urbanites", 4.9 for "Multicultural Metropolitans", and 0.0 for all others.

The 2011 OAC Pen Portrait for "Suburbanites" states that "the population of this supergroup is most likely to be located on the outskirts of urban areas" (ONS, 2015b, p.14). It also points to proxy indicators of affluence, and thus the likely presence of desirable vehicles, including home ownership (residents more likely to own their own home), property type (semi-detached/ detached), qualification level (higher), unemployment levels (below national average), job sector (information and communication, financial, public administration, and education), and transport to work (private) (ONS, 2015b, p.14). The "Urbanites" Supergroup is similar to the "Suburbanites" Supergroup in terms of job sector and unemployment level, although private renting seems to prevail (ONS, 2015b, p.12). Further, the Pen Portrait for the "Urbanites" Supergroup also states that "the population of this group are most likely to be located (...) in large urban areas" (ONS, 2015b, p.12). Taking all of this together, and recalling the characteristics of high offender rate/ high offender count areas (e.g. see Boggs, 1965, p.904; Bottoms and Wiles, 1986, p.105; Malleson, 2010, p.79), it is assumed that those areas that are similar in makeup to the "Suburbanites" Supergroup and that are also within relatively easy reach of large urban areas, such as Leeds and Bradford, will be most vulnerable to “professional” Car Key burglary offences.

3.7.3 Hot Spot Identification

Kernel Density Estimation (KDE) and Getis-Ord Gi* statistic were used to compare hot spot locations for the two burglary types. Looking first at the KDE maps in Figure 3.12 below, the KDE classification scheme was informed by Chainey et al. (2002 in Eck et al., 2005, p.29) and Eck et al. (2005, p.29), that is, the mean cell value, μ, was calculated using only those grid cells with a value, $x > 0$ and located within the study area, and all cells were then classified relative to this, e.g. $0 < x \leq \mu$, $\mu < x \leq 2\mu$, ..., and so on. Employing a standardised classification scheme such as this helps to address the subjective nature of KDE hot spot identification, as well as making it easier to compare maps for different crime types. Cell size was determined by dividing the shorter side of the study area’s minimum bounding rectangle by 150 (Chainey and Ratcliffe, 2005, p.159), and bandwidth by dividing the shorter side of the study area’s minimum bounding rectangle by 150 and then multiplying the result by 5 (Chainey, 2011 in Chainey, 2013, p.10). The respective KDE patterns are not too dissimilar to those for the LSOA scaled rates (maps repeated below), with the Car Key burglary hot spots being generally more dispersed than the Regular burglary hot spots. Given that there are more areas with a density value $x \geq 5\mu$ for Regular burglary, particularly around the urban centres of Leeds, Bradford, and Huddersfield, it might follow that these suffer high levels of repeat/ near-repeat victimisation. The main Car Key burglary hot spots to the north of Leeds and north-east/ south-west of Bradford again suggests that area accessibility and area type juxtapositions might influence crime distributions, both of which will be examined in detail in the next chapter.



Contains National Statistics data © Crown copyright and database right [2019]
Contains OS data © Crown copyright [and database right] (2019)
Contains public sector information licensed under the Open Government Licence v3.0
[http://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/]

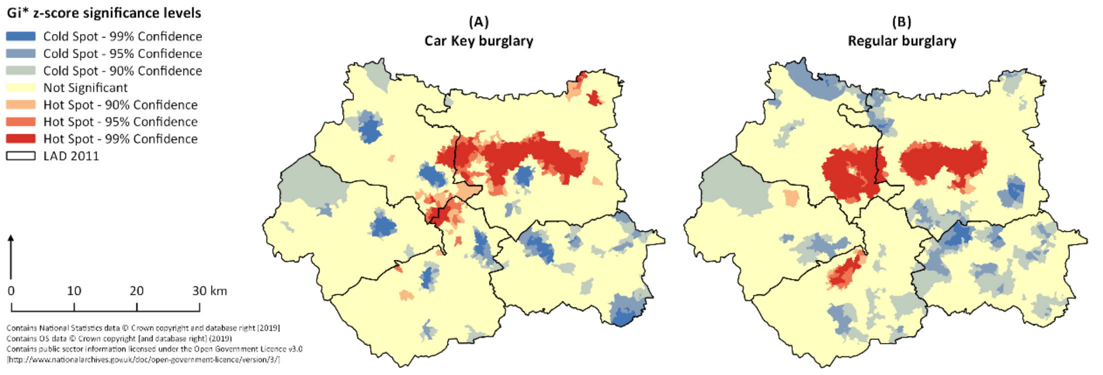
This work is based on data provided with the support of the ESRC and JISC and uses boundary material which is copyright of the Crown, the Post Office and the ED-LINE consortium.

Figure 3.12 - KDE maps for (A) Car Key and (B) Regular burglary crime points (LSOA scaled rates maps repeated here for comparison purposes)

Figure 3.13 below shows the mapped Gi* z-score significance levels for LSOA burglary rates per 1,000 households, these being broadly similar to the KDE maps. As a general rule, the town and city centres within the study area are significant cold spots for Car Key burglary, which is as expected, i.e. the overnight distribution of desirable vehicles is unlikely to coincide with CBDs. Significant hot spots are mainly clustered towards the peripheries of the Leeds and Bradford urban areas, although there are also 99% confidence hot spots in north Kirklees, Wetherby and Boston Spa. Conversely, for Regular burglary, numerous LSOAs from Bradford centre outwards show up as significant hot spots, and the same is also true for Huddersfield, albeit across a smaller spatial extent. The majority of LSOAs to the immediate north of Leeds City centre, and also to the north-east and north-west, are classified as 99% confidence hot spots, i.e. high crime rates surrounded by other high crime rates (relative to the global average).

Shine through maps are again provided to show underlying distributions of the IMD 2015 Deciles (Figure 3.14), 2011 OAC Supergroups (Figure 3.15), and 2011 Urban Rural Classification (OA) (Figure 3.16), but this time for those LSOAs that are significant hot spots (cold spots/ not significant areas are masked white). In a similar vein to previous discussions, summary observations for the Car Key burglary maps are that, for the 2015 IMD deciles, some of the significant hot spots are in areas that make theoretical sense, i.e. less deprived LSOAs within easy access of urban centres (“urban major conurbation”/ “urban city and town” in Figure 3.16). However, there are also some significant hotspots in more deprived LSOAs, which could be a result of the crime samples selection method, MAUP, the Ecological Fallacy, or all three, for example, there might still be some affluent households in LSOAs in the IMD 2015 lowest deciles (recall the findings of Trickett et al., 1995). A similar pattern is evident for the 2011 OAC Supergroups, i.e. some significant hot spots over “Rural Residents” (recall the 2+ cars/vans variable), “Suburbanites” and “Urbanites” OAs within easy access of urban centres, but also over unexpected classes, including Hard-Pressed Living. Just to reiterate, the OA level classifications are provided for background context only. The Regular burglary significant hot spots appear to be much more in line with the literature on conventional burglary risk factors. For example, the prevalence of more deprived LSOAs in Figure 3.14 map (B), particularly around urban centres, might represent the distribution of highest offender count areas (Malleson, 2010, p.79), these also being spatially concurrent/ juxtaposed with 2011 OAC Supergroup areas that are likely to contain a high proportion of potential victims, such as “Cosmopolitans”.

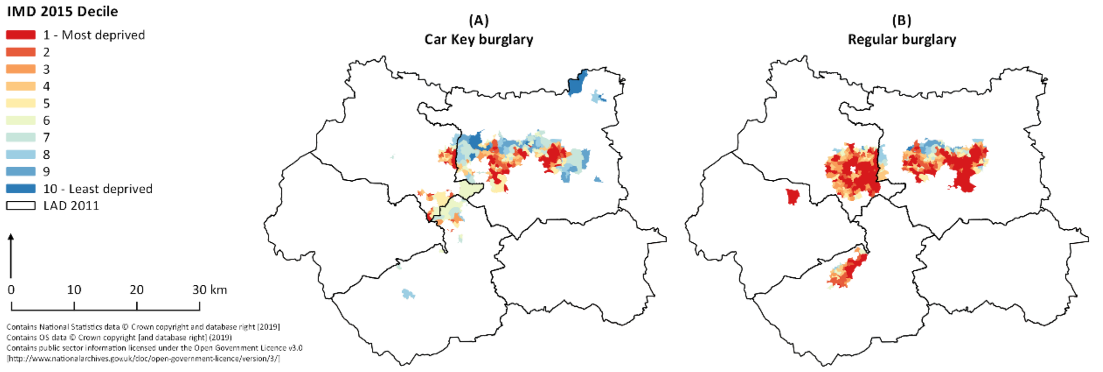
This exploratory analysis has uncovered some differences between the spatial distributions of Car Key burglary and Regular burglary, but also some similarities. Possible explanations for the latter have already been discussed, however, the likely impact of this going forwards is a positive correlation between the two burglary types, together with weaker than expected relationships between the Car Key burglary rates and the study variables.


Gi* z-score significance levels
Cold Spot - 99% Confidence
Cold Spot - 95% Confidence
Cold Spot - 90% Confidence
Not Significant
Hot Spot - 90% Confidence
Hot Spot - 95% Confidence
Hot Spot - 99% Confidence
LAD 2011
0
10
20
30 km
(A)
Car Key burglary
(B)
Regular burglary

Contains National Statistics data © Crown copyright and database right [2019]
Contains OS data © Crown copyright [and database right] (2019)
Contains public sector information licensed under the Open Government Licence v3.0
[http://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/]

This work is based on data provided with the support of the ESRC and JISC and uses boundary material which is copyright of the Crown, the Post Office and the ED-LINE consortium.

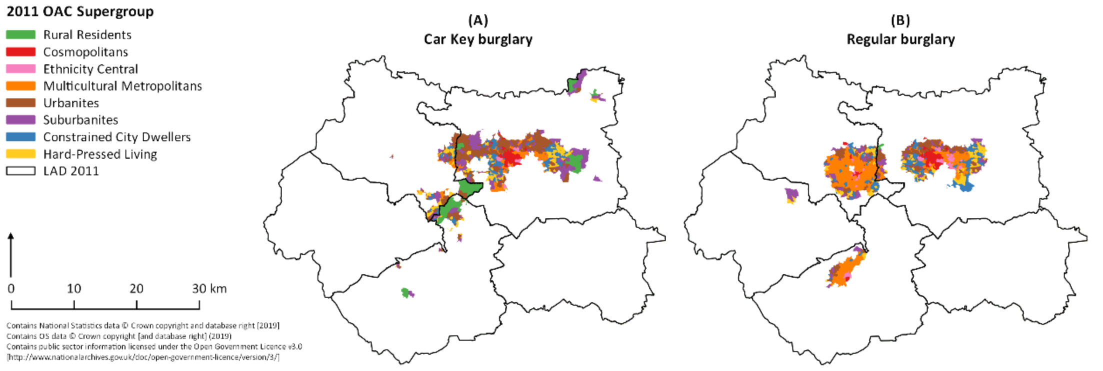
Figure 3.13 - Gi z-score significance levels for LSOA (A) Car Key and (B) Regular burglary rates per 1,000 households*



Contains National Statistics data © Crown copyright and database right [2019]
Contains OS data © Crown copyright [and database right] (2019)
Contains public sector information licensed under the Open Government Licence v3.0
[http://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/]

This work is based on data provided with the support of the ESRC and JISC and uses boundary material which is copyright of the Crown, the Post Office and the ED-LINE consortium.

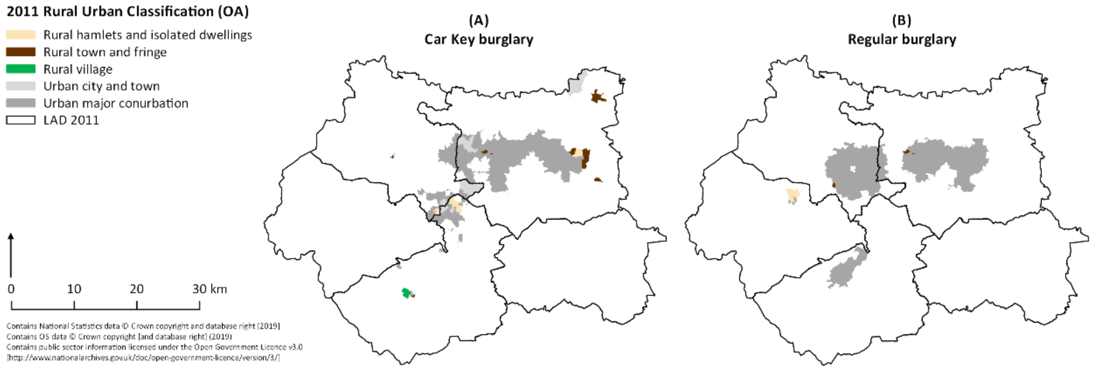
Figure 3.14 - IMD 2015 Decile shine through for significant hot spots (non-significant/ cold spots masked white)



Contains National Statistics data © Crown copyright and database right [2019]
Contains OS data © Crown copyright [and database right] (2019)
Contains public sector information licensed under the Open Government Licence v3.0
[http://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/]

This work is based on data provided with the support of the ESRC and JISC and uses boundary material which is copyright of the Crown, the Post Office and the ED-LINE consortium.

Figure 3.15 - 2011 OAC Supergroup shine through for significant hot spots (non-significant/ cold spots masked white)



Contains National Statistics data © Crown copyright and database right [2019]
Contains OS data © Crown copyright [and database right] (2019)
Contains public sector information licensed under the Open Government Licence v3.0
[http://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/]

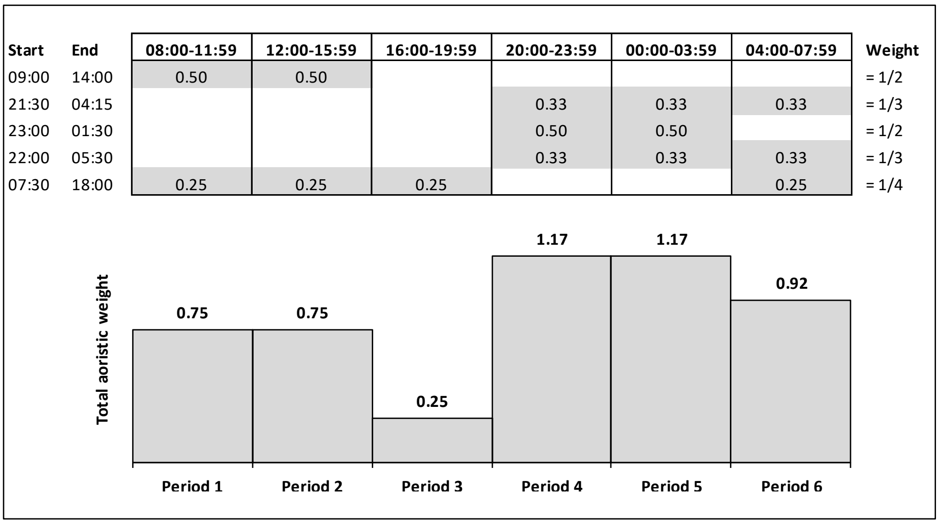
This work is based on data provided with the support of the ESRC and JISC and uses boundary material which is copyright of the Crown, the Post Office and the ED-LINE consortium.

Figure 3.16 - 2011 Rural Urban Classification (OA) shine through for significant hot spots (non-significant/ cold spots masked white)

3.8 Aoristic Analysis

Burglaries are often aoristic in nature, that is, the exact timings of offences are unknown, for example, because occupant(s) were not in residence when an offence occurred, or simply that they did not hear or see the offender(s), as might be the case with an insecure/ sneak-in entry. This typically results in 'earliest' and 'latest' committed date and time ranges, e.g. 01/01/2015, 09:30 until 01/01/2015, 14:15, as opposed to exact offence timings, e.g. 02/01/2015, 03:20. 'Aoristic' is derived from the Greek word 'aorist' meaning an undefined occurrence in time (Ashby and Bowers, 2013, p.2; Ratcliffe and McCullagh, 1998, p.751). It is very important for crime practitioners to understand when offences are most likely to have been committed, particularly if they are making decisions around how best to tackle hot spot areas, i.e. it is no good deploying resources in the right place but at the wrong time. Aoristic analysis can be used in such circumstances to ascertain the likelihood that a crime occurred during a user-specified time period, e.g. within a one hour search block.

Figure 3.17 below depicts the aoristic analysis technique that was employed in the current work, closely following the approach outlined in Ratcliffe (2000; 2002). In this example, the 24 hour time period has been broken down into six, four hour search periods, e.g. 08:00-11:59, 12:00-15:59, etc. Start and end times for five hypothetical offences are shown to the left of the table and associated weights to the right. Because the timings for the first offence (09:00-14:00) span two search periods, the associated weight value was calculated by dividing 1 by 2 to give 0.50 (all rows sum to 1). 0.50 was then assigned to the two periods, i.e. 08:00-11:59 and 12:00-15:59. After repeating the process for every offence, the resulting weights were then summed for each four hour search period to give a total aoristic weight value for that period, e.g. Period 1 = 0.75. In a real-world scenario, we would probably want to focus resources in Periods 4 and 5 as these both achieved the highest total aoristic weight value of 1.17.



Start	End	08:00-11:59	12:00-15:59	16:00-19:59	20:00-23:59	00:00-03:59	04:00-07:59	Weight
09:00	14:00	0.50	0.50					= 1/2
21:30	04:15				0.33	0.33	0.33	= 1/3
23:00	01:30				0.50	0.50		= 1/2
22:00	05:30				0.33	0.33	0.33	= 1/3
07:30	18:00	0.25	0.25	0.25			0.25	= 1/4

Figure 3.17 - Aoristic analysis method; weights rounded to 1DP (adapted from Ratcliffe, 2000, p.671)

One hour search blocks were chosen as the temporal unit in the current work and only those offences that were definitely known to have been committed during a single 24 hour period were included in the analysis; this is because longer time ranges were expected to be limited in terms of the insights that they could offer when determining the most likely timings of offences. The analysis was conducted in Excel, using an aoristic calculator created by the author specifically for the purpose of uncovering any variations in the temporal distribution of Car Key burglary and Regular burglary offences over the 24 hour period. The resulting "aoristic weight histograms" (Ratcliffe, 2002, p.28) for each burglary type are shown in Figure 3.18 below.

As anticipated, given the target property type, events in the Car Key burglary sample occur predominantly in the overnight period with an associated peak time of 02:00-04:00. This finding indicates that Car Key burglars' temporal offending patterns are determined primarily by the routine activities (Cohen and Felson, 1979) of their victims, i.e. they target residential properties when vehicles are most likely to be parked in close proximity to the keys. The peak time of 02:00-04:00 also suggests that Car Key burglars employ a degree of rational choice (Cornish and Clarke, 1987) when carrying out their offences, appearing to minimise the risk of apprehension by stealing vehicles when their owners are most likely to be rendered incapable guardians through sleep. Assuming that cars are generally parked outside houses from circa 18:00 onwards – when occupants return home from work – the lower aoristic weight values for the mid-late evening period compared to the overnight period signal that the availability of "suitable targets" does not generally outweigh the potential presence of "capable guardians" (Cohen and Felson, 1979, p.588) for Car Key burglars. The hot time period identified here, i.e. overnight, is consistent with the findings of similar studies undertaken by Shaw et al. (2010) and Carden (2012), which in turn gives confidence in the crime samples selection method.

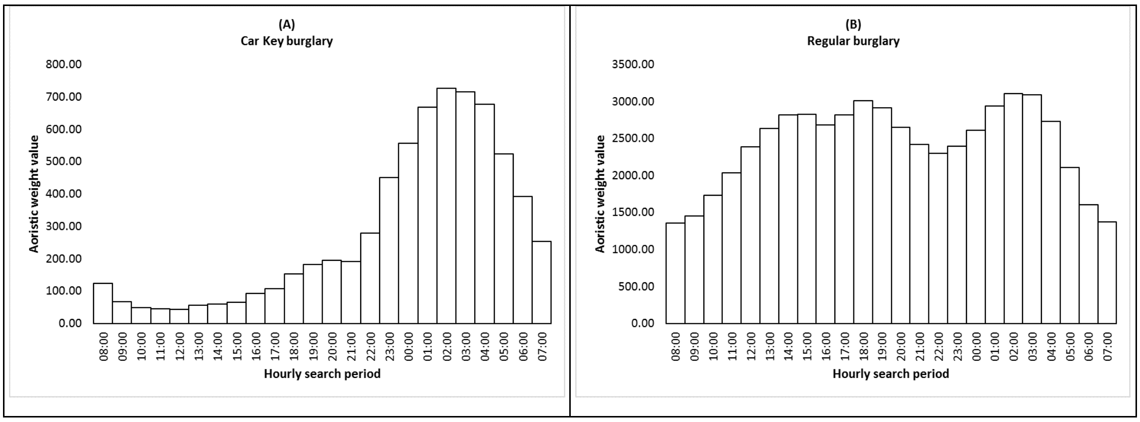


Figure 3.18 - Aoristic weight histograms for (A) Car Key and (B) Regular burglary

The approximate 50/50 split between the daytime and overnight periods for Regular burglary is broadly in line with the related literature, for example, Budd (2001, p.3) observed a ~ 50/50 split for weekdays and a 33.3/66.7 split for weekends, with 'daytime' defined as being from 6am to 6pm and 'night-time' from 6pm to 6am. It is important to note, however, that due to the manual crime samples selection method employed in the current work, some Car Key burglary offences, particularly attempts, might be present in the Regular burglary sample, thus contributing to the overnight hot time period, particularly the 02:00-04:00 peak. Assuming that the identified overnight hot time period is not entirely as a result of misclassified Car Key burglary offences, a possible explanation for this distribution is that some offenders might choose to operate under the cover of darkness as a means of reducing the risk of apprehension. As with Car Key burglary, the 02:00-04:00 peak suggests that Regular burglars also employ rational choice in their temporal decision-making, choosing to target properties at a point during the overnight period when victims are most likely to be asleep.

The relatively high aoristic weight values that are apparent during the afternoon period for Regular burglary indicate that associated offenders also target properties when they are less likely to be occupied, with probable explanations for this provided by Routine Activity Theory (Cohen and Felson, 1979) and Rational Choice Theory (Cornish and Clarke, 1987). Surprisingly, there is also a peak between 18:00 and 20:00 for Regular burglary, a time of the day when homes might reasonably be expected to be occupied, however, it could be that the convergence of certain stealable property types together with a specific set of victim behaviours is responsible for generating this pattern. To use the example of a bag, if the owner takes the bag to work during the day, then it cannot be stolen from their house, but if the owner returns home from work, unlocks the front door, places the bag on the hall table and then moves to a different part of the house, the opportunity presents for a sneak-in entry and subsequent theft of the bag.

In order to better understand the temporal distributions discussed here, future work could include analysis of the two burglary types by the weekday and weekend periods. Since the routine activities of potential victims are likely to be different at weekends than on weekdays, this might produce different spatio-temporal distributions of target property types, particularly for Car Key burglary, for example, cars might be parked outside houses for longer durations. Disaggregate analysis such as this could potentially unmask variations in the general temporal patterns identified thus far. Further, recognising that accurate temporal information is fundamental to the efficient tasking of resources, it would be expedient to incorporate the aoristic analysis findings into the current research dynamic and combined models.

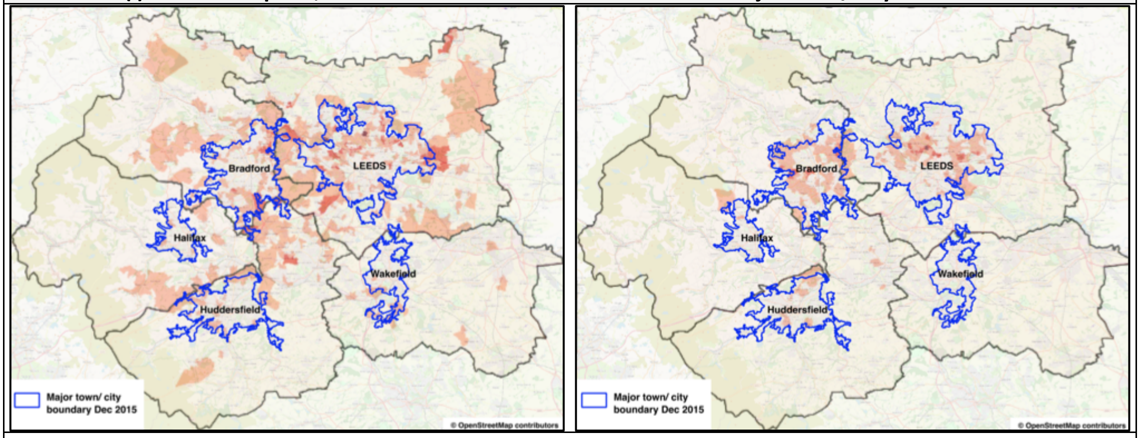
3.9 Summary

Having been through the process of preparing a large, securely held, and police-recorded crime data set for analysis, the absence of any universally accepted guidelines on which to base the associated decision-making, including criteria for deletion, and choice of statistical tests, was somewhat challenging – perhaps this could be considered as a potential area for future work with a view to improving both the comparability and repeatability of different crime studies. Two notable findings from this chapter are: (i) the overnight hot time period for Car Key burglary, and (ii) differences between the characteristics of high Car Key burglary rate areas and high Regular burglary rates areas (supplementary LSOA-level burglary rate maps are provided for the interested reader in Figure 3.19 and Figure 3.20 below, together with a brief commentary). Given these findings, it is anticipated that opposite directional relationships will be observed between the Car Key burglary rates and the current research independent variables and the Regular burglary rates and the independent variables. Car Key burglary is also expected to generate a different repeat/ near-repeat victimisation pattern to Regular burglary.

Further to the discussions in Section 3.7.2 of this chapter on the clustering of burglary rates, Figure 3.19 below shows the LSOA Car Key and Regular burglary rates per 1,000 households scaled 0-1 and overlaid with major town/ city and neighbourhood policing team boundaries. However, it is important to note that, due to social distancing measures, the author was unable to access the original (securely stored) crime rates shapefiles and so the burglary rates layers that are presented in Figure 3.19 and Figure 3.20 have been constructed from the images in Figure 3.8 of Section 3.7.2 – these were georeferenced in a GIS, meaning there could be some inaccuracies. The maps in Figure 3.19 (i) support the observation made in Section 3.7.2 that "the highest Regular burglary rates are likely to be found in urban major conurbation areas, whereas Car Key burglars appear to target both rural and urban areas". All of the highest Regular burglary rates – defined by the author as scaled classes 0.50-0.75 and 0.75-1.00 – are located in the more built-up parts of the study area, namely Leeds and Bradford, whereas the pattern for Car Key burglary is less well defined, with higher rate LSOAs found both within and beyond the major urban areas.

(A) Car Key burglary | **(B) Regular burglary**

(i) LSOA rates per 1,000 households scaled 0-1 overlaid with major town/ city boundaries:

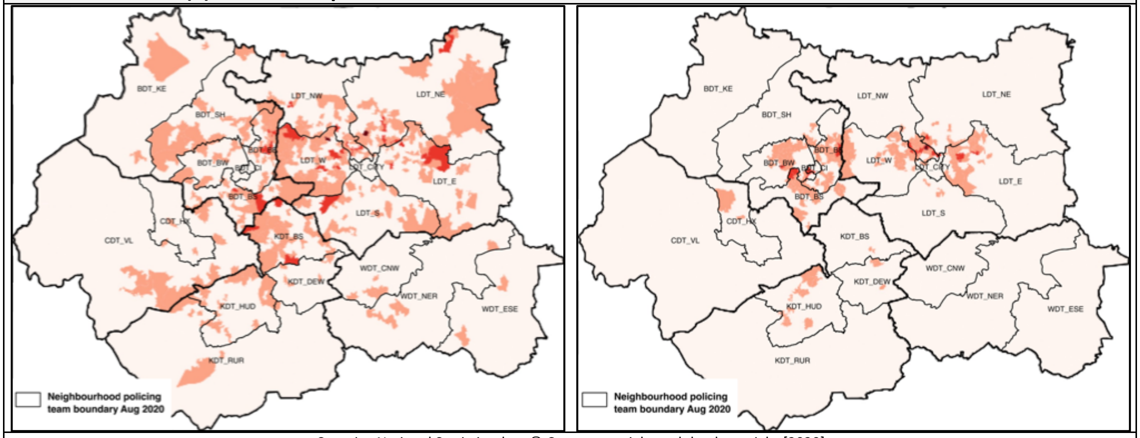

Bradford
LEEDS
Halifax
Wakefield
Huddersfield
Major town/ city boundary Dec 2015
© OpenStreetMap contributors
Bradford
LEEDS
Halifax
Wakefield
Huddersfield
Major town/ city boundary Dec 2015
© OpenStreetMap contributors

NB The crime rates layers are partially transparent so that the more rural areas of the study area 'shine' through.

Rate per 1,000 households (scaled 0-1)

- ▢ 0.00 - 0.25
- 🟧 0.25 - 0.50
- 🟥 0.50 - 0.75
- ■ 0.75 - 1.00
- □ LAD 2011

(ii) LSOA rates per 1,000 households scaled 0-1 overlaid with NPT boundaries:



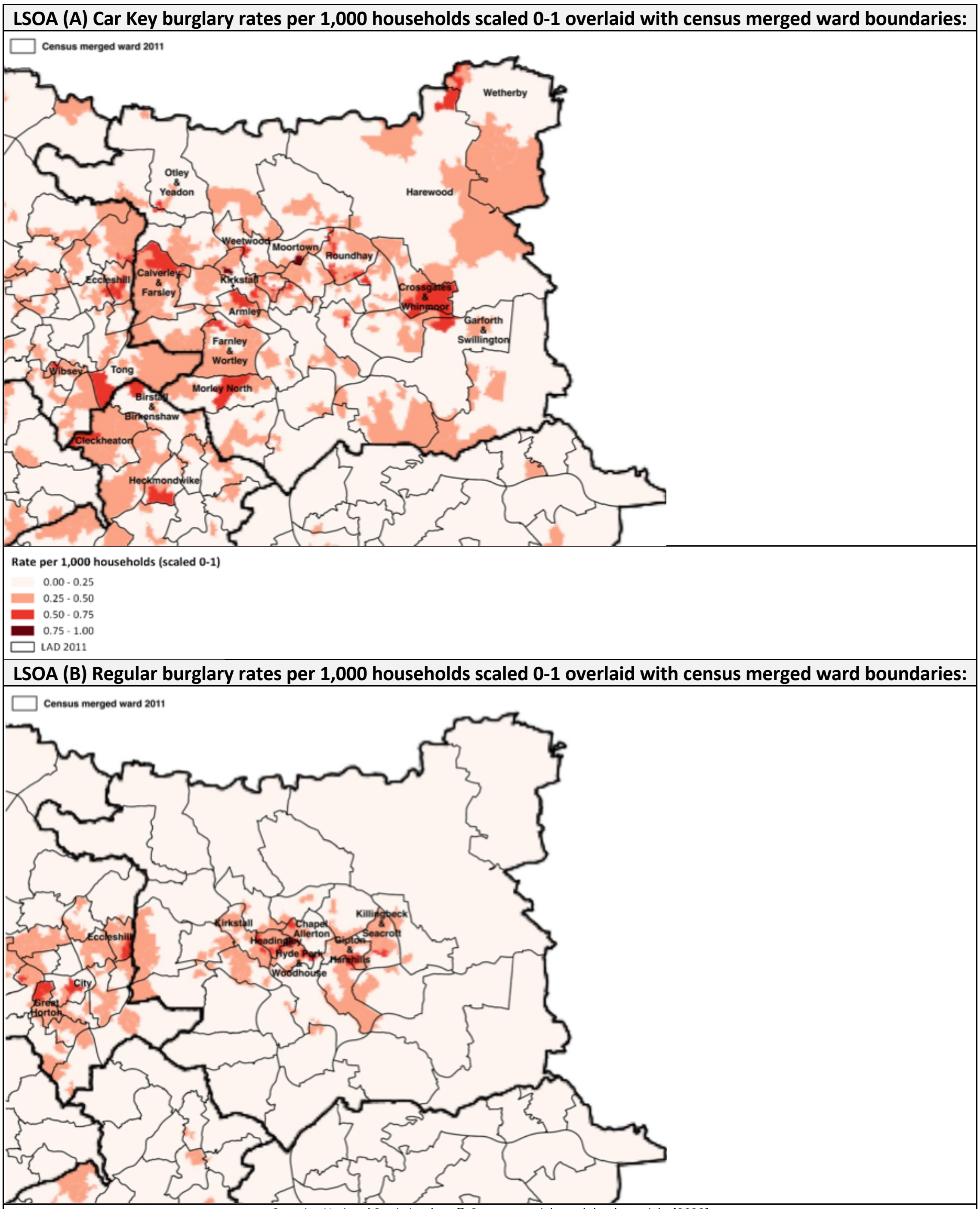
Major town/ city boundary Dec 2015 source: Office for National Statistics licensed under the Open Government Licence v3.0

Figure 3.19 - LSOA (A) Car Key and (B) Regular burglary rates per 1,000 households scaled 0-1 overlaid with: (i) major town/ city boundaries and (ii) neighbourhood policing team (NPT) boundaries

Since some of the higher Car Key burglary rate LSOAs are quite spatially disparate, the maps in Figure 3.19 (ii) will be used to relate these to a larger, more police-relevant geography, specifically neighbourhood policing team (NPT) areas, which are used to direct police resources. In a similar vein, census merged wards are used in Figure 3.20 to provide some 'place' context. The maps in Figure 3.20 contain all LSOAs in the study area that have a burglary rate in one of the top two scaled classes (0.50-0.75 and 0.75-1.00) – the geographical extent is the same for both maps to facilitate comparison, and some of the wards have been assigned name labels for reference purposes.

Looking at the maps in Figure 3.20, it is evident that the ('hot') higher rate LSOAs – top two scaled classes – for Car Key burglary are generally different to those for Regular burglary, for example, there are three 'hot' Car Key burglary LSOAs in the 'Batley and Spen' NPT area ('KDT_BS' in Figure 3.19), but none for Regular burglary. Similarly, there are a couple of 'hot' Car Key burglary LSOAs to the north east of the study area, in Harewood and Wetherby wards (LDT_NE in Figure 3.19), but none for Regular burglary – notably, the two LSOAs in question are in the least deprived decile for the IMD 2015 (decile 10). Wetherby is also well served by the road network, as well as being located approximately central to the Leeds, Harrogate, and York 'golden triangle', thus potentially rendering it vulnerable to both West Yorkshire offenders and those travelling cross-border, e.g. from North Yorkshire. The only wards with an apparent overlap between 'hot' Car Key and 'hot' Regular burglary LSOAs are Headingley and Hyde Park & Woodhouse (LDT_NW in Figure 3.19) – this could be due to both students and professionals residing in these areas, hence the presence of different stealable property types, i.e. 'CRAVED' items (Clarke, 1999) belonging to students and newer vehicles belonging to professionals, and/or that some of the Car Key burglaries relate to students' cars, which might not be 'professional' thefts as such. In terms of general trends, less deprived areas – proxy desirable vehicles – that are relatively accessible by car from more deprived areas, especially around Leeds and Bradford, appear to be especially vulnerable to Car Key burglary offences. In addition to parts of Harewood and Wetherby wards, other 'hot' areas for Car Key burglary include parts of Moortown, Roundhay, Weetwood, Morley North, and Calverley & Farsley wards (Leeds LA) and Birstall & Birkenshaw and Heckmondwike wards (Kirklees LA), with some LSOAs being in the least deprived deciles for the IMD 2015, including Leeds 024D (Roundhay ward) and 095B (Morley North ward) – decile 9. Relative to the West Yorkshire study area, the highest Regular burglary rate LSOAs are far more spatially clustered than for Car Key burglary, as can be seen in Figure 3.20, map (B). In addition to parts of Headingley and Hyde Park & Woodhouse wards, other 'hot' areas for Regular burglary include parts of Kirkstall, City & Hunslet, and Gipton &

Harehills wards (Leeds LA) and Eccleshill, City, and Great Horton wards (Bradford LA), with some LSOAs being in the most deprived decile for the IMD 2015 (decile 1), including Leeds 050D (Killingbeck & Seacroft ward) and 060C (Gipton & Harehills ward). Recalling from Chapter 2 that Malleson (2010, p.79) identified a Pearson's correlation coefficient of 0.56 between IMD Score and LSOA offender count, this might explain some of the observed burglary patterns in the current study. For example, Leeds and Bradford LAs each have a higher proportion of LSOAs in decile 1 of the IMD 2015 than the other three LAs, which could point to larger offender populations, these potentially generating both within-area Regular burglary hot spots and surrounding area (less deprived LSOAs) Car Key burglary hot spots – note the high Car Key burglary rate LSOAs located near to the Bradford/ Leeds/ Kirklees border in Figure 3.20, map (A).


LSOA (A) Car Key burglary rates per 1,000 households scaled 0-1 overlaid with census merged ward boundaries:

LSOA (B) Regular burglary rates per 1,000 households scaled 0-1 overlaid with census merged ward boundaries:

Figure 3.20 - LSOA (A) Car Key and (B) Regular burglary rates per 1,000 households scaled 0-1 overlaid with census merged ward boundaries

Chapter 4 will now describe how the two area accessibility variables were derived, and also present the results of associated analysis. Chapter 5 will then review the results of the correlation and spatio-temporal analysis that was performed on the data from this chapter, as well as describing how the static risk surface was developed for the combined model; recall that this will be used to filter the areal extent of event-based risk.

Chapter 4 Measuring the Impact of Area Accessibility on Burglary Rates

4.1 Introduction – Area Type Juxtapositions

This first part of this chapter will examine the extent to which the characteristics of surrounding areas influence within-area crime rates, a subject that has received only limited attention in the criminological literature thus far. Recalling the theories that were discussed in Chapter 2, including Routine Activity Theory, and journey-to-crime (JTC), it is assumed that attractive areas within offenders' operational spaces will be at a greater risk of victimisation than those located elsewhere. For example, Samson and Wooldredge (1987, p.372) note that: "a major assumption of the opportunity model is that the closer the ecological proximity of potential targets to motivated offenders, the greater the risk of victimization". Despite this, and as noted by Bernasco and Luykx (2003, pp.982-983), researchers seeking to explain burglars' crime location choices have tended to focus on either: (i) the distance between offender residence and offence, or (ii) within-area characteristics (e.g. ethnicity and deprivation). Here, housing tenure will be used as a proxy variable to infer spatial juxtapositions of high offender rate (less affluent) areas and low offender rate (more affluent) areas, and the distances over which these juxtapositions might exert an influence on within-area crime rates will be controlled through inverse weighting. The second half of the chapter will employ a measure of street network centrality, namely 'closeness centrality', to examine the role of area accessibility on within-area residential burglary rates. To aid the flow of the chapter, a separate introduction will be provided for this.

4.2 How Might Area Type Juxtapositions Influence Within-Area Crime Rates?

Notwithstanding the respective contributions of the "offender-based" (distance) and "target-based" (within-area) approaches to crime location analyses (Bernasco and Nieuwbeerta, 2005, p.300), it is important to recognise that the characterisation of area attractiveness without reference to BOTH of these factors runs the risk of misspecifying potential target areas. For example, an area that possesses many attributes of criminal attractiveness might be situated too far from an offender population to make it a viable target, whereas an area that is easily accessible to offenders could present few/ no criminal opportunities. Figure 4.1 below uses two hypothetical scenarios to illustrate how area type juxtapositions might be expected to influence within-area crime rates.

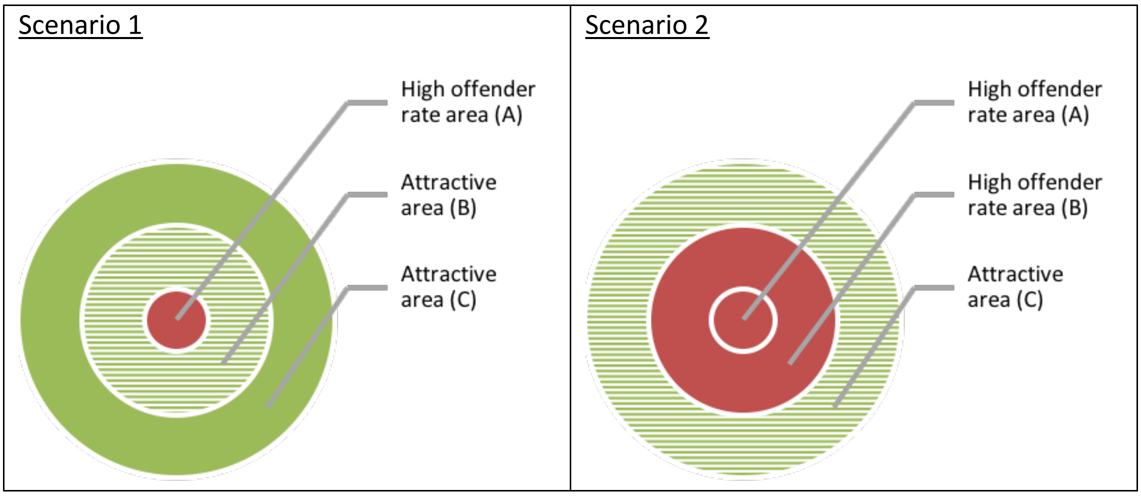

Scenario 1
High offender rate area (A)
Attractive area (B)
Attractive area (C)
Scenario 2
High offender rate area (A)
High offender rate area (B)
Attractive area (C)

Figure 4.1 - Hypothetical scenarios to illustrate how area type juxtapositions might be expected to influence within-area crime rates (based on: Hakim, 1980; Hirschfield et al., 2014)

Informed by the work of Hirschfield et al. (2014), and referencing Hakim (1980, p.265) who suggests that:

> property crimes are 'imported', or attracted, to a community in direct relation to its relative wealth [attractiveness]; and in inverse relation … to its distance from the areas containing the greatest concentration of potential criminals Hakim (1980, p.265),

it would be reasonable to infer that, in Scenario 1, attractive area (B) is more likely than attractive area (C) to be targeted by offenders from high offender rate area (A). In effect, area (B) acts as a protective "buffer zone" (Hirschfield et al., 2014, p.1060) between area (A) and area (C). Conversely, in Scenario 2, attractive area (C) is placed at an increased risk of victimisation through the substitution of area (B) for a high offender rate area. It should be noted, however, that factors such as the "least-effort principle" (Zipf, 1949 in Ackerman, 2015, p.240) will still continue to exert a restrictive influence on offender travel between areas (A) and (C) in Scenario 2. These examples, although much simplified versions of reality, demonstrate how an area can be rendered more/ less vulnerable to victimisation depending on the characteristics of its periphery.

An early example of distance and within-area characteristics being considered concurrently in the context of crime is Smith's (1976) application of gravity modelling to between-area crime trips (Bernasco and Nieuwbeerta, 2005, pp.301-302). The classic form of the gravity model predicts the level of interaction between two locations based on the size of their respective populations and the distance between them (Smith, 1976, pp.802-803). However, assuming that factors other than population size might pull offenders to an area, Smith decided to weight the 'attractiveness' variable by a number of within-area attributes, including 'wealth' and 'percent unemployed', albeit none were found to be successful predictors of crime flows (p.810).

An alternative approach is provided by the discrete spatial choice and conditional logit models. Here an offender chooses a target area from a set of alternatives based on expected benefits, such as proximity to home address, and affluence, with resulting odds ratios denoting the effect of a unit increase in each independent variable on the risk of victimisation (for detailed explanation, see: Bernasco and Nieuwbeerta, 2005, pp.302-304). Although Bernasco and Nieuwbeerta (2005, p.311) remark that the framework could be used to test the impact of area type juxtapositions, it was not deemed appropriate for this work due to an absence of offender data, i.e. no information regards the locations of offenders' residences and their linked offences. This was also the case for gravity modelling, that is, no offender data from which to model flows.

Having discounted the gravity modelling and discrete spatial choice approaches, a further search of the literature was conducted to locate any studies where residential burglary rates/ counts were analysed in relation to the periphery of the spatial unit of study and some socio-demographic variable/s of interest therein. This identified just four unique examples (for details, see: Hirschfield and Bowers, 1997; Ratcliffe and McCullagh, 2001; Bernasco and Luykx, 2003; and Hirschfield et al., 2014), the key points of which are summarised in Table 4.1 below. Given that offenders have the capacity to transcend socially-constructed areal boundaries, the apparent neglect of hinterland effects in spatial crime analyses is somewhat surprising.

Study	Summary
Hirschfield, A. and Bowers, K.J. 1997 The development of a social, demographic and land use profiler for areas of high crime	**Periphery:** 1 km, 2 km, & 3 km of area type cluster perimeter (p.114) Calculate the percentage population living in affluent and disadvantaged enumeration districts (ED) (as identified using the 1991 census-derived Super Profiles classification) around affluent and disadvantaged area type clusters in Merseyside to ascertain the impact of different area type arrangements on residential burglary rates.
Ratcliffe, J.H. and McCullagh, M.J. 2001 Crime, repeat victimisation and GIS	**Periphery**: 200 m, 350 m, & 500 m radii of an address (p.80) Generate an areally-weighted average of a socio-demographic indicator (Index of Local Conditions 1994 [ILC]) for the vicinity of residential burglary locations in Nottinghamshire to better understand the area type distribution of repeat victimisation by mitigating the peninsula effect and inaccurate geocoding of offences (pp.79-80).
Bernasco, W. and Luykx, F. 2003 Effects of attractiveness, opportunity and accessibility to burglars on residential burglary rates of urban neighborhoods	**Periphery:** Study area Derive a “spatially weighed burglar exposition rate” (SWEBER) (p.990) from 1996-2001 police recorded data for neighbourhoods in The Hague, Netherlands and include the resulting variable in a multiple regression model to simultaneously test the effect of within-area attributes and offender proximity on residential burglary rates.
Hirschfield, A. et al. 2014 How places influence crime: The impact of surrounding areas on neighbourhood burglary rates in a British city	**Periphery:** Contiguous areas Develop a set of spatial models, including spatial predictor lag (p.1065), to establish whether different Output Area (OA) type juxtapositions (as identified using the 2001 census-derived Output Area Classification [OAC]), e.g. Core: ‘City living’, Periphery: ‘Prospering suburbs’ (p.1067), in The City of Leeds influence residential burglary rates.

Table 4.1 - Studies that relate residential burglary to characteristics of the periphery

Of the three studies in Table 4.1 that relate to residential burglary rates, all identified relationships between these and characteristics of the periphery, for example:

- **Hirschfield and Bowers (1997)** observed that the Affluent Area Cluster (AAC) with the highest burglary rate also had the highest percentage population living in disadvantaged areas in its periphery. Conversely, the AAC with the lowest burglary rate had no disadvantage in its periphery (pp.114-115).

- **Bernasco and Luykx (2003)** identified a significant positive correlation of .77 between their offender proximity variable (SWEBER) and neighbourhood burglary rates (p.993). As a single independent variable in a multiple regression model, SWEBER accounted for 60% of the variation in rates (pp.992-993).

- **Hirschfield et al. (2014)** found that different OAC super-group collocations influenced OA burglary rates, with all four of their spatial models outperforming a standard ANOVA (pp.1065-1066). It should be noted, however, that even with the preferred model, there was still some unexplained variance (p.1066).

Perhaps the most notable of these findings in terms of identifying areas that might be particularly vulnerable to Car Key burglary is that affluent peripheries can provide "protective 'buffer zones'" for affluent "cores" (Hirschfield and Bowers, 1997; Hirschfield et al., 2014, p.1060, p.1067). A possible explanation for this is that offenders do not, as a general rule, reside in more affluent areas, and so the impeding effects of distance come into play, as illustrated in Figure 4.1. Assuming that desirable vehicles will be most prevalent in more affluent/ less deprived areas, then it is hypothesised that these areas will also experience the highest Car Key burglary rates but ONLY where they are 'bordered' by less affluent/ more deprived areas (proxy: high offender rate areas). Table 4.2 below shows how different area type juxtapositions are expected to influence Car Key burglary rates.

	DOMINANT PERIPHERY TYPE	
WITHIN-AREA TYPE	**Less affluent**	**More affluent**
Less affluent	Low Car Key burglary rate	Low Car Key burglary rate
More affluent	High Car Key burglary rate	Low Car Key burglary rate

Table 4.2 - Anticipated effects of different area type juxtapositions on Car Key burglary rates

4.2.1 Calibrating the Periphery Relative to Offenders' Average Journey-to-Crime Distances

Although it could be argued that any crime-related study of the periphery is inherently linked to the spatial distribution of would-be offenders and potential victims, none of the studies in Table 4.1 explicitly reference offenders' JTC parameters when defining the areal extent of a surrounding area's influence on within-area burglary rates/ counts. Since burglars' JTC distances can vary with offence characteristics, a more empirical approach might be to calibrate the periphery to reflect the average distance travelled by offenders from home base to offence for a specific crime type/ crime type sub-category. Table 4.3 below presents the findings of a review into burglars' average JTC distances, although it should be noted that comparisons are difficult due to different/ unspecified methods used, e.g. Euclidean/ non-Euclidean. Some noteworthy points are that Snook (2004, p.60) identified that median JTC distance increased with target property value, that the median JTC distance for burglars on foot was 1.2 km, and that the median distance for burglars in a vehicle was 5.5 km. In a similar vein, Carden (2012, p.74) observed a median JTC distance for Regular burglars of 2.1 km and a median JTC distance for Car Key burglars of 4.9 km.

--- = no data provided

Source	**Study area**	**Summary**	**Median**	**Mean**	**Min.**	**Max.**
Snook, 2004, p.60	The City of St John's and surrounding areas, Canada	Burglars travelling on foot to offend.	1.2 km (Metric not specified)	1.9 km	0 km	8 km
White, 1932, p.507	The City of Indianapolis, USA	Resident and non-resident burglars offending in study area.	---	2.8 km	---	---
Costello and Wiles, 2001, p.37	The City of Sheffield	Resident domestic burglars offending in study area.	---	3 km	---	---
Carden, 2012, p.74	Merseyside	Resident 'Regular' burglars offending in study area.	2.1 km	---	0 km	14.7 km
Vandeviver et al., 2015, p.406	East and West Flanders, Belgium	Resident domestic burglars offending in study area.	2.6 km	8.2 km	0 km	128 km
Ackerman and Rossmo, 2015, p.249	The City of Dallas, USA	Resident domestic burglars offending in study area.	4 km (Non-Euclidean)	7.6 km	0 km	44.6 km
Carden, 2012, p.74	Merseyside	Resident 'Car Key' burglars offending in study area.	4.9 km	---	0 km	23.4 km
Snook, 2004, p.60	The City of St John's and surrounding areas, Canada	Burglars travelling by vehicle to offend.	5.5 km	6.3 km	0 km	20 km

Table 4.3 - Average, minimum, and maximum burglar journey-to-crime distances

NB All distances rounded to 1 DP, converted to km if originally given in miles, and sorted by Median ascending (where applicable).

Taking all of this together, it is hypothesised that as within-area attractiveness (affluence) increases relative to surrounding area attractiveness, so too will Car Key burglary rates. In order to test this theory, a method will be developed to generate a summary 'attractiveness' measure for the periphery of every areal unit in the study area. Novelty will be introduced by weighting the contribution of every within-area value to every surrounding area value according to burglars' average JTC parameters. Spearman's correlation will then be used to explore the relationship between within-area burglary rates (dependent variable) and the difference between within-area 'attractiveness' and surrounding area 'attractiveness' values (independent variable). Given the median JTC distances for vehicular-based burglars, it is inferred that if a more affluent area is juxtaposed with less affluent areas up to a distance of approximately 5 km, then it will be especially vulnerable to Car Key burglary offences.

Central to the proposed method is the idea that offenders do not typically reside in areas where the target property type for Car Key burglary is expected to be most prevalent. It is important, therefore, to ratify this assumption with a view to ensuring that any resulting differences between within-area attractiveness and surrounding area attractiveness can reasonably be considered to represent spatial juxtapositions of would-be offenders and potential victims (desirable vehicles). Given that no offender data was provided to the author, it was not possible to identify high offender rate areas based on offenders' home addresses and so a review of the associated literature, including that already covered in Chapter 2, was used to determine a suitable proxy indicator, the findings of which are outlined next.

4.3 Inferring High Offender Rate Areas

In common with the assertion of Bowers and Hirschfield that "More research is needed to identify the residential profiles of offenders" (1999, p.164), only a small literature could be located on the subject, however, the following four variables were identified as being indicative of offender distributions at the areal unit level:

1. **Social disorganisation** (Shaw and McKay, 1942; Sanders, 1943; Sampson and Groves, 1989; Andresen, 2014)
2. **Within-area crime rates** (Boggs, 1965)
3. **Deprivation** (Malleson, 2010)
4. **Housing tenure** (Baldwin et al., 1976; Bottoms and Wiles, 1986; Costello and Wiles, 2001)

4.3.1 Social Disorganisation

Social disorganisation was discussed at length in Chapter 2; however, it is useful here to recall Shaw and McKay's finding that delinquency was most pronounced in Zone II of the city (Wortley and Mazerolle, 2008, p.5). This 'zone in transition', characterised by the structural factors inherent to social disorganisation, provided an environment particularly conducive to crime and criminality (Chainey and Ratcliffe, 2005, p.83; Sampson and Groves, 1989, p.774; Wortley and Mazerolle, 2008, p.5). Although a social disorganisation measure could have been created for differentiating area attractiveness/ inferring high offender rate areas in the current research, the choice of variables to include, together with the sourcing of relevant data, was deemed overly complex for the intended purpose (Bursik, 1988 in Andresen, 2014, pp.21-22).

4.3.2 Within-area Crime Rates

Boggs (1965, p.904) identified a strong positive correlation of .762 between burglary offender rates and residential day burglary rates for census tracts in the City of St. Louis, USA. However, whilst the observed relationship between the two variables might well be attributable to the same offender behaviours that are integral to Crime Pattern Theory, the average size of a census tract was not given. If the census tracts were very large, then the associated offenders might not actually have resided in/ close to the areas in which they committed their offences (MAUP).

4.3.3 Deprivation

Malleson (2010, p.79) presented strong empirical evidence that more burglars were likely to live in the more deprived areas of Leeds, West Yorkshire, including a Pearson's correlation coefficient of 0.56 between IMD Score and LSOA offender count.

4.3.4 Housing Tenure

Baldwin et al. (1976) classified ≈ 200 EDs in the City of Sheffield according to dominant tenure type, i.e. > 50% of households held their tenure in a particular way, else recorded as 'mixed' (p.102). Male offender rates were subsequently calculated per 10,000 relevant population (age group, e.g. males aged 10-19) for individual districts, grouped by dominant tenure type, and then averaged (p.105, p.108). As can be seen from Figure 4.2 below, the highest mean offender rates were generated by those areas in which either 'public' (council) or 'private' renting was dominant, however, the authors did note large variation within these sectors (Bottoms and Wiles, 1986, p.105). Although the observed trends are clearly vulnerable to the ecological fallacy, related in-depth analyses at the time did not identify any high offender rate 'owner-occupied' EDs (Bottoms and Wiles, 1986, pp.104-105).

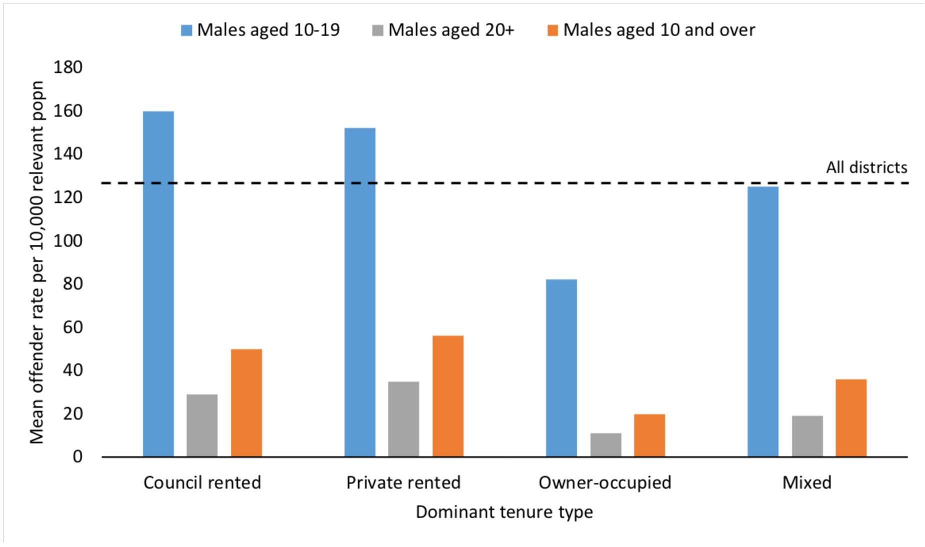


Figure 4.2 - Mean male offender rates by dominant tenure type for EDs in Sheffield, 1966 (adapted from: Baldwin et al., 1976, p.105, p.108)

In a similar study, Costello and Wiles (2001) ascertained that the majority of (residential) neighbourhood offence rates in the City of Sheffield, 1995, were influenced by local offender rates. The authors went on to examine nine of these neighbourhoods in more detail, including prevailing characteristics (derived using the 'GB Profiler' lifestyle analysis – see Table 4.4 and Table 4.5 below), total within-area offender movements, and average JTC distances. Table 4.4 supports Baldwin et al.'s (1976) finding that the highest mean offender rates were generated by those areas in which either 'public' or 'private' renting was dominant. Contrasting the different lifestyle descriptions in Table 4.4 and Table 4.5, it is assumed that 'high offender and high offence rate' neighbourhoods will be less attractive to Car Key burglars than 'low offender and low offence rate' neighbourhoods.

Lifestyle category	**Description**
1	Struggling: Multi-ethnic areas, pensioners and single parents, high unemployment, **LA rented** flats.
2	Struggling: **Council tenants**, blue collar families and single parents, **LA rented** terraces.
4	Struggling: Multi-ethnic areas; less prosperous **private renters**, young blue collar families with children, **privately renting** terraces and bedsits.

Table 4.4 - Lifestyle categories for four high offender and high offence rate neighbourhoods in Sheffield, 1995 (adapted from: Costello and Wiles, 2001, p.41)

Lifestyle category	Description
8	Established: Rural farming communities, mature well-off self-employed couples and pensioners, **owning or privately renting** large detached houses.
9	Prospering: Affluent achievers, mature educated professional families, **owning and buying** large detached houses.
10	Established: Comfortable middle agers, mature white collar couples and families, **owning and buying** semi's.

Table 4.5 - Lifestyle categories for two low offender and low offence rate neighbourhoods in Sheffield, 1995 (adapted from: Costello and Wiles, 2001, p.41)

Using data collected by Costello and Wiles (2001), Table 4.6 gives the total number of *offender movements* – straight-line distance between an offender's home address and offence location (Costello and Wiles, 2001, p.34) – in each of the 'high/high' and 'low/low' offender/offence rate neighbourhoods, together with the proportion of these attributable to either locally-residing or other Sheffield offenders. In line with Boggs' (1965) identification of a positive relationship between within-area crime rates and within-area offender rates, the greatest number of movements occurred in the 'high/high' areas, with at least half being related to local offenders. Unsurprisingly, the longest average JTC distances were generated by the 'low/low' neighbourhoods, perhaps a reflection of the fact that around three quarters of associated movements were connected to other Sheffield offenders (c.f. ≤ 45% for 'high/high' neighbourhoods).

Within-area offender /offence rate	Area	Total offender movements	% within-area offenders	% other Sheffield offenders	Average distance travelled (km)
High/high	F	438	60	36	1.3
	G	375	55	42	1.6
	H	399	50	45	1.8
	I	339	55	40	1.2
Low/low	A	135	24	75	6.4
	B	95	8	79	5.6

Table 4.6 - Offender movements in four 'high/high' and two 'low/low' offender and offence rate neighbourhoods in Sheffield, 1995 (adapted from: Costello and Wiles, 2001, pp.41-42).

NB Distances converted from miles to km and data relating to non-Sheffield offenders omitted from the table with a view to parsimony.

Key findings from this section regards the characteristics of high/ low offender rate residential areas are summarised in Table 4.7 below.

HIGH OFFENDER RATE RESIDENTIAL AREAS	**Example source**	**LOW OFFENDER RATE RESIDENTIAL AREAS**	**Example source**
High crime rate	Boggs, 1965, p.904	**Low crime rate**	Boggs, 1965, p.904
Public and private renting	Baldwin et al., 1976, p.105, p.108	**Owner-occupied and private renting**	Costello and Wiles, 2001, p.41
Less prosperous	Costello and Wiles, 2001, p.41	**Affluent**	Costello and Wiles, 2001, p.41

Table 4.7 - Key findings in relation to the characteristics of high offender and low offender rate residential areas

Notably, Bottoms and Wiles (1986) suggest that access to housing and the housing market is a key determining factor for offence and offender rates in residential areas of Britain. Although this is generally well supported in the literature, as discussed in this section and Chapter 2, it is important to note that some anomalies were identified within the overall trends. Nevertheless, in the absence of any offender data, housing tenure does appear to be an acceptable proxy by which to infer high and low offender rate neighbourhoods in the study region and, by association, area attractiveness in the context of Car Key burglary. Data relating to housing tenure is also readily available at both the LSOA and OA geographies through UK census outputs.

4.4 Current Research Approach

Having identified a proxy indicator for inferring high offender rate areas/ area attractiveness, namely ‘percentage households social renting or private renting’, the next step in the current research was to develop an interpolation method with which to generate a summary value of this variable for the periphery of every LSOA and OA in the study area. Referencing technical elements of the ‘periphery’ studies that were reviewed in Section 4.2, the final method employed here draws primarily on the work of Bernasco and Luykx (2003) and Ratcliffe and McCullagh (1999; 2001), as well as on some aspects of geographic profiling and gravity modelling. The following factors were considered as part of the method development process, all of which will be discussed in more detail below:

- Defining the periphery, e.g. buffer, contiguity

- Choice of distance metric, i.e. Euclidean or Manhattan
- Start and end points for distance calculations
- Replicating offender mobility behaviours
- Computing power (7,131 OAs in West Yorkshire)

Figure 4.3 below is included as a high level aide memoire to some of the steps mentioned earlier in this chapter, and also to introduce some of the steps that will be outlined next.

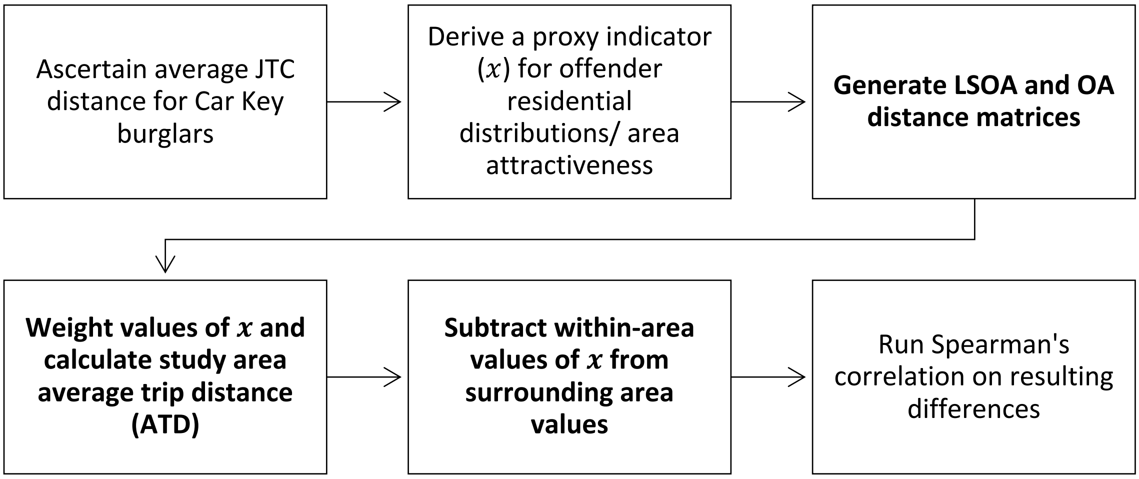

Ascertain average JTC distance for Car Key burglars
Derive a proxy indicator (x) for offender residential distributions/ area attractiveness
Generate LSOA and OA distance matrices
Weight values of x and calculate study area average trip distance (ATD)
Subtract within-area values of x from surrounding area values
Run Spearman's correlation on resulting differences

Figure 4.3 - Process map of current research approach (emphasis indicates process step in R).

4.4.1 Interpolation Methods

Recall from Table 4.1 that Hirschfield and Bowers (1997) used conventional rate calculations in their analysis of the periphery, whereas Ratcliffe and McCullagh (2001) employed an areally-weighted interpolation technique to generate summary values of the Index of Local Conditions (ILC), an early area deprivation measure, for the periphery of individual residential burglary locations in Nottinghamshire. As described in Figure 4.4 and Equation 4.1 below, Ratcliffe and McCullagh (1999, p.39; 2001, pp.78-80) created a buffer of radius (r) around a crime location (i), multiplied the proportional contribution of every intersecting region ($a_j(r)$) by the corresponding ILC figure (x), summed the resulting values and then divided the total by the buffer's overall area to give a "'vicinity' value" (V) (2001, p.80). This is a version of the 'weighted average (weighted mean)', as presented in Equation 4.2.

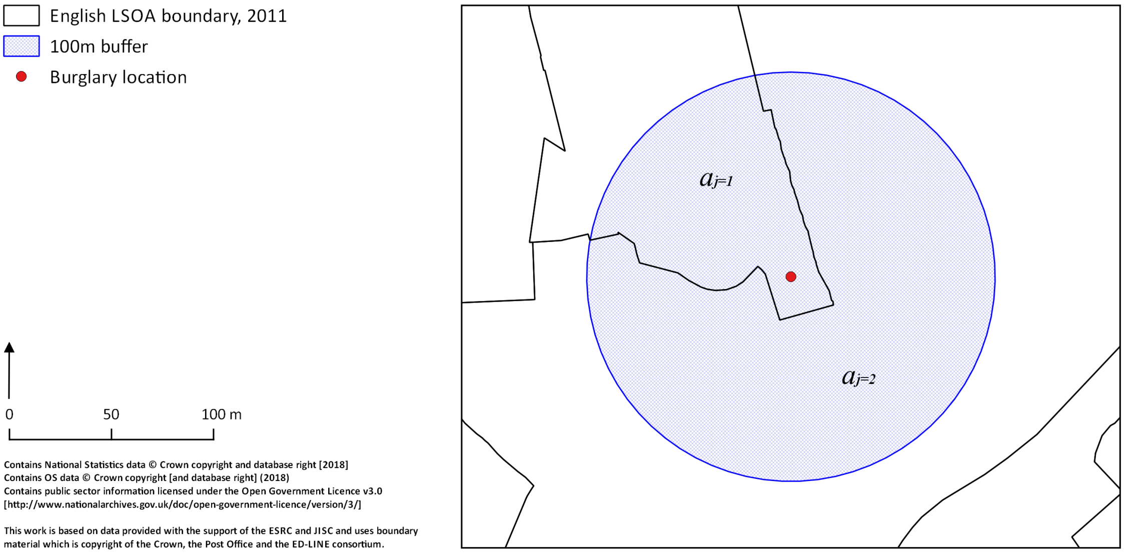


Figure 4.4 - Areally-weighted interpolation method (adapted from: Ratcliffe and McCullagh, 1999, p.39; Ratcliffe and McCullagh, 2001, p.79; anonymised burglary location from `https://data.police.uk`*)*

$$V_i = \frac{\displaystyle\sum_{j=1}^{j=n} x_j a_j(r)}{\displaystyle\sum_{j=1}^{j=n} a_j(r)}$$

Where:

$\boldsymbol{V_i}$ = ‘Vicinity’ value for crime location i

$\boldsymbol{x_j}$ = value of x for region j

$\boldsymbol{a_j(r)}$ = area of j within given radius of i

Equation 4.1 - Vicinity calculation (Source: Ratcliffe and McCullagh, 2001, p.80)

$$\frac{(w_1x_1 + \ w_2x_2 + \cdots + w_nx_n)}{(w_1 + \ w_2 + \cdots + \ w_n)}$$

Where:

$\boldsymbol{w_1}$ = weight for observation n=1

$\boldsymbol{x_1}$ = value of x for observation n=1

Equation 4.2 - Weighted average calculation (Source: Upton and Cook, 2014, p.456)

Although Ratcliffe and McCullagh (1999; 2001) did not explicitly incorporate the effects of distance decay in their interpolation method, Equation 4.1 provided a useful basis for the current research and was adapted as follows:

- Rather than using a fixed radius buffer to define the periphery, the whole of the West Yorkshire study area was instead considered.

- Values of x were weighted according to distance, as opposed to proportional areal contribution (for details, see: Figure 4.5 and Equation 4.3).

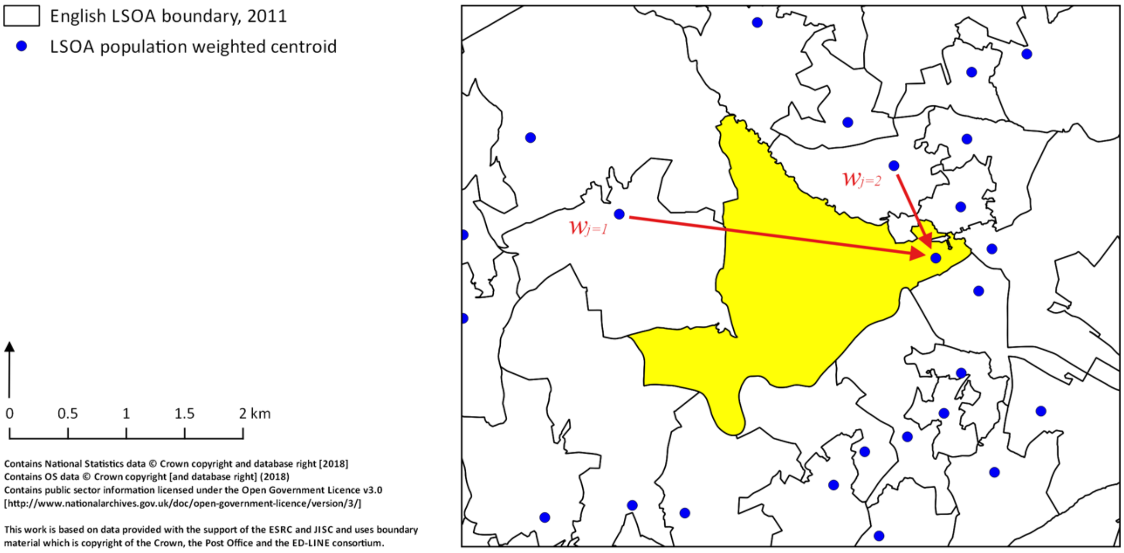
As already mentioned, gravity models were not included in the analysis *per se*, however, a core element of these, namely ‘inverse distance’, was utilised.

4.4.2 Inverse Distance

The choice of IDW interpolation to replicate offender mobility behaviours in the current research is well supported in the related literature, for example, Bernasco and Luykx (2003, p.991) employed an inverse distance function of the form ${D_{ij}}^{-2}$ to derive their SWEBER measure, stating that:

> In order to implement the (limited) mobility of burglars, it is assumed that the threat that a burglar poses to a potential target neighbourhood is some inverse function of the distance between the burglar’s home and the target neighborhood (Bernasco and Luykx, 2003, p.991).

Another example is Kent et al. (2006) who used the negative exponential and truncated negative exponential functions to model homicide offenders’ JTC behaviours, finding both of these to be relatively successful in terms of reproducing associated distance travelled frequency distributions; R^2 ranged from .720 to .851 depending on the type of distance function and distance metric tested, the latter being either ‘direct-path Euclidean’, ‘travel-path’, or ‘temporal path’ (p.194). Distance decay functions are also used in retail sector gravity models to predict flows of people/ demand between residential areas and retail outlets. The function $\exp(-\beta)$ features in the most common of these gravity models, the ‘production-constrained’ spatial interaction model (SIM), with beta calibrated to reflect customers’ known travel behaviours (Birkin et al., 2017, pp.74-75).


English LSOA boundary, 2011
● LSOA population weighted centroid
0
0.5
1
1.5
2 km
Wj=1
Wj=2
i

Contains National Statistics data © Crown copyright and database right [2018]
Contains OS data © Crown copyright [and database right] (2018)
Contains public sector information licensed under the Open Government Licence v3.0
[http://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/]

This work is based on data provided with the support of the ESRC and JISC and uses boundary material which is copyright of the Crown, the Post Office and the ED-LINE consortium.

Figure 4.5 - Current research approach 'inverse distance weighted' (IDW) interpolation method

$$P_i = \frac{\sum_{j=1}^{j=n} x_j w_j}{\sum_{j=1}^{j=n} w_j}$$

Where:

$\boldsymbol{P_i}$ = 'Periphery' value for areal unit i

$\boldsymbol{x_j}$ = value of x for areal unit j

$\boldsymbol{w_j}$ = inverse distance weights matrix of the form $\boldsymbol{exp}(-\boldsymbol{\beta d_{ij}})$

Where:

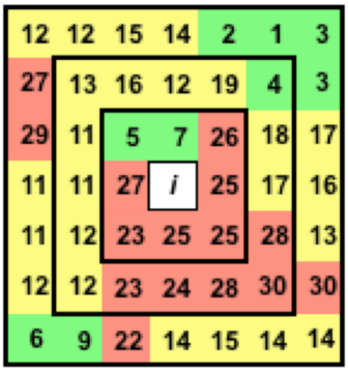
$\boldsymbol{\beta}$ = beta (controls rate of distance decay)

$\boldsymbol{d_{ij}}$ = distance between i and j

Equation 4.3 - Current research approach periphery calculation (adapted from: Birkin et al., 2017, p.74; Ratcliffe and McCullagh, 2001, p.80; Upton and Cook, 2014, p.456)

4.4.3 Calibrating the Model

Figure 4.6 below highlights the importance of conceptualising an area's surroundings in a way that is relevant to the phenomenon being considered, for example, if the average JTC distance for a specific crime type falls within the immediate (inner) periphery of i, then we should attach the most weight in any associated IDW calculations to those cells. Given that Carden identified a median journey-to-crime distance for Car Key burglars of 4.9 km (2012, p.74), it was anticipated that the strongest Spearman's ρ value would be achieved for Car Key burglary rates and the 'attractiveness' (housing tenure) variable when β was calibrated to reflect this.



12	12	15	14	2	1	3
27	13	16	12	19	4	3
29	11	5	7	26	18	17
11	11	27	*i*	25	17	16
11	12	23	25	25	28	13
12	12	23	24	28	30	30
6	9	22	14	15	14	14

Area of interest = i

Unweighted average values:

- Inner periphery (8 cells) = **20.4**
- Inner and middle periphery (24 cells) = **18.4**
- Inner, middle, and outer periphery (48 cells) = **15.9**

Figure 4.6 - Example to show how choice of peripheral extent can impact summary values

Figure 4.7 shows how different values of β can alter the rate of distance decay in a SIM; $exp(-\beta)$ was applied to the integer range 0 to 25 (e.g. distance) for β = 0.2, 0.5, and 1.

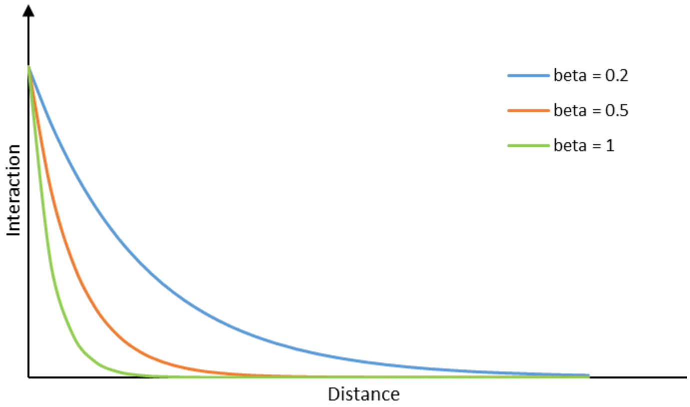


Figure 4.7 - Distance decay curves for the integer range 0 to 25 and different values of beta

Hansen (1959, p.74) lists exponent (β) values for vehicular travel relating to the following activities: 'work', 'social', 'shopping', and 'school' (for details, see Table 4.8). A lower exponent value reflects a greater propensity to travel, for example, people are willing to travel further for work (0.9) than they are for socialising (2.0), or school (2.0+). The figures referred to above do not include terminal time, however, this is unlikely to be a major issue during the hot time period for Car Key burglary i.e. evening/ overnight. We might, therefore, expect the exponent value for mobile offenders to be roughly similar to that for workers, since the purpose of the travel is the same; financial gain. Also, Car Key burglars might reduce the risk of apprehension by undertaking less frequent trips, so they might actually be willing to travel longer distances than legitimate workers.

Trip type	Intraurban travel exponent
Work	0.9
Social	1.1
Shopping	2.0
School	2.0+

Table 4.8 - Exponent values identified by Hansen (1959, p.74)

Birkin et al. (2010, p.440) note that SIMs are traditionally calibrated using average trip distance (ATD), with Batty and Mackie suggesting that this is the most appropriate method for models that incorporate an exponential distance function (1972 in Newing et al., 2014, p.22). Calibration of the current model will therefore be undertaken by altering the value of β in the P_i equation until the difference between the observed ATD of 4.9 km for Car Key burglars is as close as possible to the ATD that is generated by the IDW method (Birkin et al., 2010, p.440; Newing et al., 2014, pp.22-25). Equation 4.4 presents the ATD calculation for individual LSOAs/ OAs, whereas the ATD for the whole of the West Yorkshire study area will be obtained by taking a simple average of the resulting areal unit ATDs.

$$ATD_i = \frac{\sum_{j=1}^{j=n} w_j d_{ij}}{\sum_{j=1}^{j=n} w_j}$$

Where:

$\boldsymbol{ATD_i}$ = Average trip distance for areal unit i

$\boldsymbol{w_j}$ = inverse distance weights matrix

$\boldsymbol{d_{ij}}$ = distance between i and j

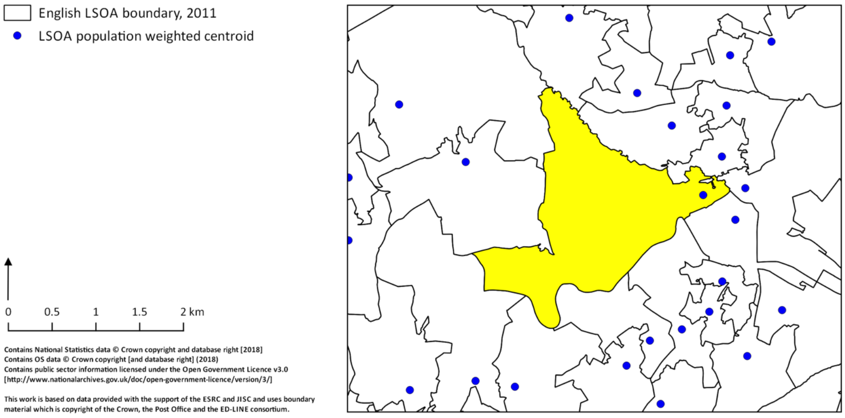
Equation 4.4 - Current research approach ATD calculation (adapted from: Newing et al., 2014, p.23)

4.4.4 Calculating Distance

Population weighted centroids (PWC), as opposed to zone/ polygon centroids, were chosen as the start and end points for distance calculations in the model because these were expected to most accurately reflect the real-world arrangement of would-be offenders and potential victims. PWCs were created for the 2011 OA, LSOA, and MSOA geographies, using the Median Center function in ArcGIS 10.0, to provide a single summary description of the spatial distribution of the population in each areal unit (ONS, 2017, p.1). The Median Center algorithm iterates over potential locations until it identifies the geographical point that minimises the Euclidean distance to all other features in the dataset. It is also less vulnerable to outliers than the Mean

Center (ESRI, 2016). Where a PWC fell outside, or < 2 metres from, an area's boundary, it was moved to the nearest point ≥ 2 metres inside the boundary (ONS, 2017, p.1).

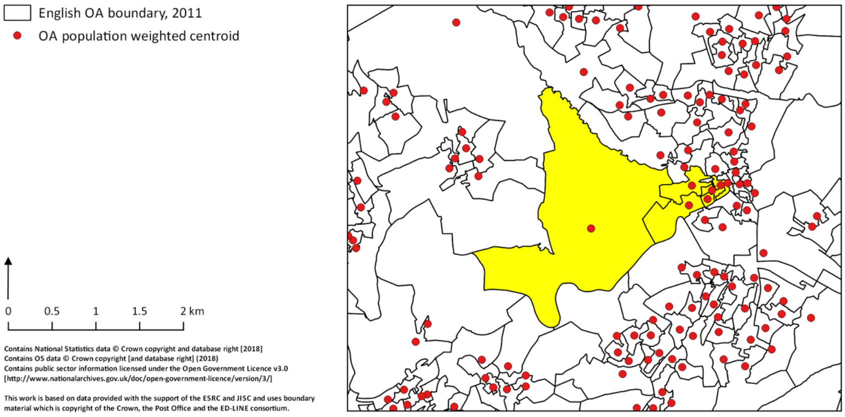
The 2011 output area PWCs were generated using 2011 Census derived household coordinates and populations, a key purpose being to facilitate best-fitting of national statistics to different geographies (ONS, 2017; 2017b; 2017c). 2011 Census outputs and statistical tables for OAs, LSOAs, and MSOAs are exact estimates (ONS, 2017c). To protect the privacy of individuals, OAs are never split or apportioned during the fitting process (ONS, 2017c). Figure 4.8 and Figure 4.9 demonstrate how the distribution of PWCs can vary depending on the level of output area that is considered because this is likely to have an impact on the accuracy of interpolated values in the current research. For example, there is just one PWC in the LSOA highlighted in Figure 4.8, as would be expected, but an additional five when the area is disaggregated into its constituent OAs.


English LSOA boundary, 2011
LSOA population weighted centroid
0
0.5
1
1.5
2 km

Contains National Statistics data © Crown copyright and database right [2018]
Contains OS data © Crown copyright [and database right] (2018)
Contains public sector information licensed under the Open Government Licence v3.0
[http://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/]

This work is based on data provided with the support of the ESRC and JISC and uses boundary material which is copyright of the Crown, the Post Office and the ED-LINE consortium.

Figure 4.8 - Example distribution of LSOA population weighted centroids



Contains National Statistics data © Crown copyright and database right [2018]
Contains OS data © Crown copyright [and database right] (2018)
Contains public sector information licensed under the Open Government Licence v3.0
[http://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/]

This work is based on data provided with the support of the ESRC and JISC and uses boundary material which is copyright of the Crown, the Post Office and the ED-LINE consortium.

Figure 4.9 - Example distribution of OA population weighted centroids

Although some studies have used Manhattan distance to represent offender mobility, particularly in block-based US cities, it was decided that direct-path Euclidean distance would be more suitable for calculating distances between PWCs in the current study, i.e. in a UK context. Further, Kent et al. (2006, p.197) asserts that offenders conceptualise the route between two points as a straight line, which is deemed especially relevant for mobile Car Key burglary offenders who are hypothesised to commit their criminal activities within spatially extensive, and thus more generalised awareness spaces, i.e. beyond the confines of their residentially-anchored (neighbourhood-centred) routine activity spaces.

4.4.5 Model Inputs and Outputs

Initially, two matrices were created in Microsoft Excel to calculate the straight line distance (km) between every LSOA/ OA PWC and every other in the study area, together with weights. However, the size of the problem, including 50,851,161 calculations to generate the OA weights (7,131 x 7,131 units), was found to be intractable in Excel and so a coding algorithm was instead developed and the statistical package, R, was used to perform the final calculations. The coding implementation included a β variable that could be reset every time the algorithm was run, i.e. to test different values of beta, and the diagonal of the weights matrix was set to zero so that the within-area area values were not included in the average calculations, i.e. the aim was to capture differences between areas. Sample model output is shown in Table 4.9 below.

	LSOA code	**(A)**	**(B)**	**Difference**	**LSOA ATD**	**Study area**
		LSOA	**Periphery**	**(B minus A)**	**(km)**	**ATD**
		% renting	**% renting**			**(km)**
1	E01010568	22.7	25.4	2.6	2.0	1.9
2	E01010569	37.7	30.2	-7.5	1.9	1.9
3	E01010570	2.9	29.9	26.9	2.1	1.9
4	E01010571	35.6	30.5	-5.1	1.9	1.9
5	E01010572	9.5	27.9	18.3	1.9	1.9

Table 4.9 - Sample model output from coding algorithm for LSOAs, beta = 1.0 (figures rounded to 1 DP)

Table 4.10 below shows the three 2011 Census UK inputs that were used to derive the 'percentage households social renting or private renting' variable in the current work. These are the same as were used for the 2011 Output Area Classification variables K032 (% Households who are social renting) and K033 (% Households who are private renting).

Variable	Variable table	EW code	Column header 1	Column header 2
Total households	Household Composition	KS105EW0001	All categories: Household composition	n/a
Households who are social renting	Tenure	KS402EW0005, KS402EW0006	**Social rented:** Rented from council (Local Authority)	**Social rented:** Other
Households who are private renting	Tenure	KS402EW0007, KS402EW0008	**Private rented:** Private landlord or letting agency	**Private rented:** Other

Table 4.10 - 'Percentage households social renting or private renting' variable derivation inputs

Figure 4.10 and Figure 4.11 below show the spatial distributions of the 'percentage households social renting or private renting' variable for LSOAs and OAs in the West Yorkshire study area. As expected, the highest LSOA rates (dark blue) coincide with the main urban centres of Leeds and Bradford, although the distribution is more nuanced at the OA level. The other two maps, Figure 4.12 and Figure 4.13, show the spatial distributions of the percentage point difference values when beta was set to 0.5 (ATD ~ 3.5 km). Recall that the differences were calculated by subtracting within-area values of the 'percentage households renting' variable (x) from IDW values of the variable (x). Therefore, a positive difference represents a higher proportion of households renting in the periphery (on average), whereas a negative difference represents a lower proportion of households renting in the periphery (on average). The areas with the highest positive differences are shown in red and dark orange on the maps, and with the highest negative differences in dark green and bright green. In terms of Car Key burglary, the red and dark orange areas are expected to be particularly vulnerable to the offence type given that they potentially represent spatial juxtapositions of would-be offenders and the target property type.

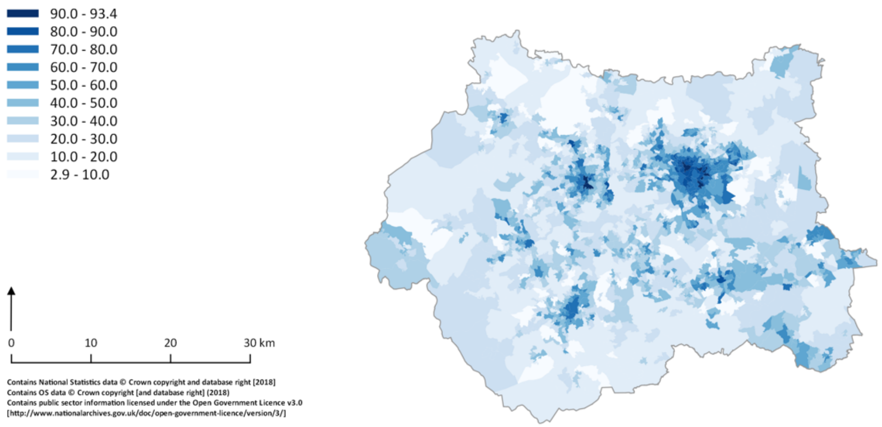


Figure 4.10 - LSOA percentage households social or private renting, 2011

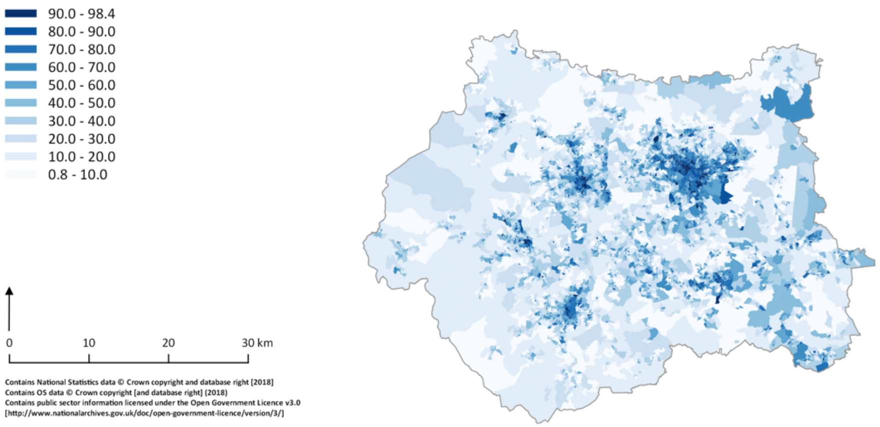


Figure 4.11 - OA percentage households social or private renting, 2011

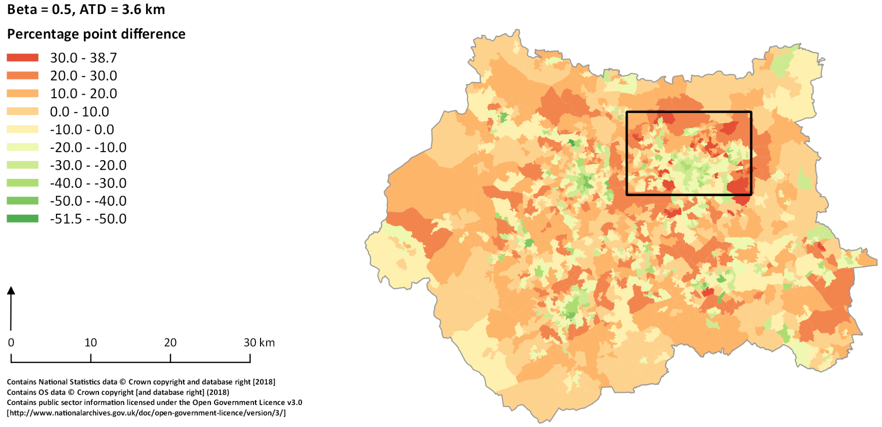


Figure 4.12 - Percentage point difference between LSOA percentage households social or private renting, 2011 and surrounding LSOAs IDW average

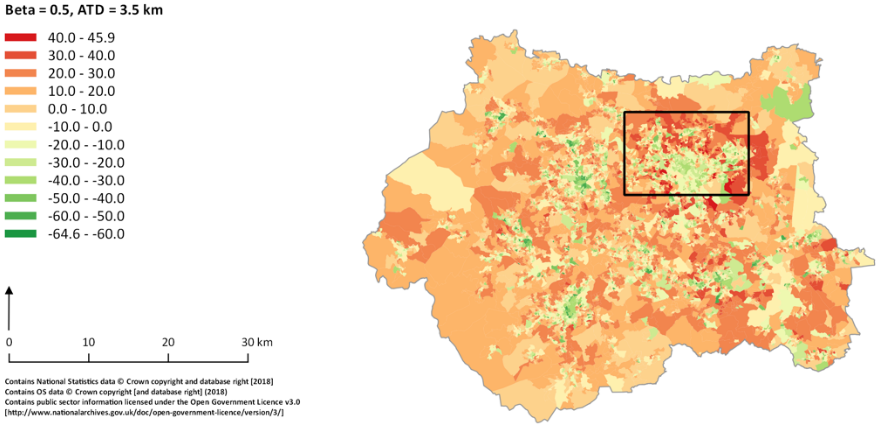


Figure 4.13 - Percentage point difference between OA percentage households social or private renting, 2011 and surrounding OAs IDW average

Figure 4.14 and Figure 4.15 below show detailed views of the LSOAs and OAs located within the black rectangle (Leeds centre and periphery) in Figure 4.12 and Figure 4.13. Interestingly, there are a number of high positive percentage point difference areas on the outskirts of Leeds, particularly at the OA level, which might explain some of the burglary distributions that were discussed in Chapter 3.

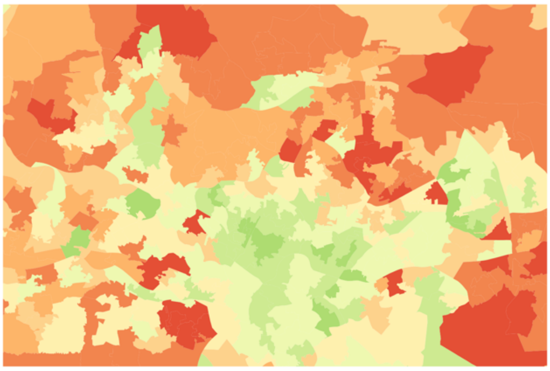


Figure 4.14 - LSOA detailed view (area in rectangle Figure 4.12)

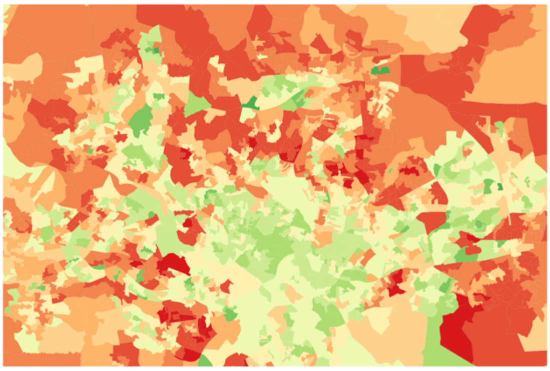


Figure 4.15 - OA detailed view (area in rectangle Figure 4.13)

4.4.6 Results and Discussion

Table 4.11 below presents the LSOA percentage point difference Spearman's Correlation coefficients for each beta value tested – seven in total – and each burglary type, together with the average trip distances (ATDs) generated. From the point of view of interpreting the results, note that a positive correlation coefficient indicates that as the percentage point difference increases, that is, there is a higher proportion of households renting in the periphery (on average), so too do within-area burglary rates. As expected, all of the Car Key burglary coefficients are both positive and significant, which strengthens the idea that area type juxtapositions have some bearing on this crime type, i.e. more attractive areas targeted by offenders from less attractive areas. Also of note is that the strongest Car Key burglary correlations are observed for the 3.6 km and 5.0 km ATDs, which supports the hypothesis about calibrating the extent of the periphery in line with offender mobility behaviours. Conversely, all of the LSOA Regular burglary coefficients are negative, which indicates that, as within-area attractiveness increases relative to surrounding area attractiveness, within-area Regular burglary rates decrease. This relationship might seem surprising given Hirschfield and Bower's (1997) finding that crime rates in Affluent Area Clusters (AACs) in Merseyside increased with disadvantage in the periphery, and that this was especially pronounced for burglary. However, the authors did note that only a few area clusters were examined in their study and, thus, a more systematic approach would be needed to validate the results. Referring back to Figure 3.9

and the IMD 2015 shine through map for Regular burglary, a possible explanation for the negative correlations between more attractive/ less attractive area type juxtapositions and Regular burglary in the current study is that some of the highest Regular burglary rates in West Yorkshire are in the more deprived LSOAs. The related literature also indicates that more deprived areas are likely to suffer higher Regular burglary rates; recall that Malleson (2010, p.79) uncovered positive correlations between LSOA deprivation scores and offender counts, and Boggs (1965, p.904) between US census tract residential day burglary rates and burglary offender rates. Therefore, it would appear that, in the West Yorkshire study area, the fact that an LSOA is surrounded by less attractive LSOAs has little impact on Regular burglary rates within that LSOA because the highest rates would not typically be found in such areas anyway. In this respect, the hinterland measure is acting as a proxy for within-area deprivation, i.e. a positive percentage point difference for an LSOA indicates that it is less deprived. A second point to note here is that the method by which the area differences were operationalised might have influenced the results (to be discussed in more detail later). The Spearman's ρ values and associated ATDs from Table 4.11 are also presented in graph format in Figure 4.16 below, and the OA results are shown in Table 4.12 and Figure 4.17 below – the OA trends are not that dissimilar to the LSOA trends, although the correlations are weaker.

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

Beta value	**3.0**	**2.0**	**1.0**	**0.5**	**0.35**	**0.2**	**0.1**
ATD (km)	0.8	1.1	1.9	3.6	5.0	8.0	11.9
Car Key burglary ρ	.287**	.318**	.360**	.382**	.381**	.362**	.330**
Regular burglary ρ	-.043	-.054*	-.090**	-.156**	-.200**	-.265**	-.317**

Table 4.11 - ρ values for percentage point difference between LSOA percentage households social or private renting, 2011 and surrounding LSOAs IDW average, and burglary rates

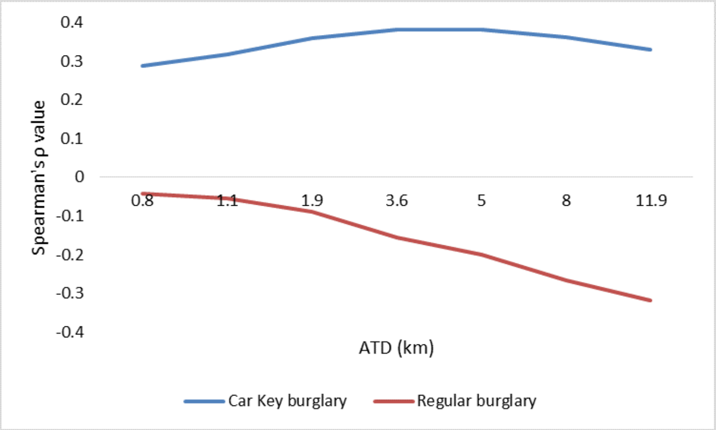


Figure 4.16 - Spearman's ρ values as presented in Table 4.11 and associated ATDs (km)

Beta value	3.0	2.0	1.0	0.5	0.35	0.2	0.1
ATD (km)	0.6	0.9	1.7	3.5	4.9	7.9	11.9
Car Key burglary ρ	.231**	.247**	.270**	.282**	.281**	.272**	.254**
Regular burglary ρ	.010	.004	-.018	-.064**	-.095**	-.143**	-.183**

Table 4.12 - ρ values for percentage point difference between OA percentage households social or private renting, 2011 and surrounding OAs IDW average, and burglary rates

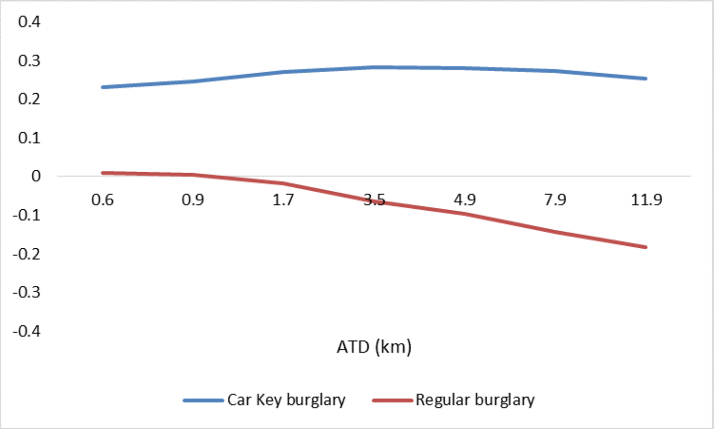


Figure 4.17 - Spearman's ρ values as presented in Table 4.12 and associated ATDs (km)

Some limitations of the method employed here include that: (i) only those areas within West Yorkshire were considered, i.e. there was no allowance for possible cross-border effects, (ii) the housing tenure variable is susceptible to both the ecological fallacy and MAUP – perhaps a more robust approach would have been to estimate within-area offender counts based on the information contained in Figure 4.2 (Baldwin et al., 1976) and using housing tenure ratios, (iii) averaging the housing tenure variable might have diluted some local effects, e.g. directionality, and (iv) the percentage point difference values due not capture relative area differences *per se*. To explain, as can be seen from Table 4.13 below, areas A and B have different within-area and surrounding area values of x, but each generates a percentage point difference value of 40.

Area	Attractiveness values		Difference (surrounding area renting minus within-area renting)
	Within-area renting	**Surrounding area renting**	
A	35	75	40
B	10	50	40

Table 4.13 - Potential issue with percentage point difference variable

The Car Key burglary results presented in this chapter do support the hypothesis that more attractive areas surrounded by less attractive areas are especially vulnerable to the crime type, particularly when area type juxtapositions occur within the spatial extent of offenders' average journey-to-crime distances. There is, however, clear opportunity for future development of the approach, including how best to delineate area attractiveness in the model. Notably, Bernasco and Luykx (2003, p.997) state that "the use of spatially weighted measures of accessibility to offenders offers a new and promising approach to research on intra-urban spatial dynamics of crime."

4.5 Introduction – Calculating Average Area Accessibility (Closeness Centrality)

To first define closeness centrality (CC), assuming that we have a graph comprised of nodes (points) and edges (connecting lines), as per Figure 4.18 below, CC can be used to ascertain for each node how well connected it is to every other node in the graph – the node that is most 'centrally' located (closest to all other nodes, on average, in the graph) will receive the highest CC score (Bavelas, 1950; Freeman, 1979; Cohen et al. 2014; Boeing, 2017). In the current research, CC scores will be calculated for the driveable West Yorkshire street network, with junctions/ terminating points (e.g. ends of cul-de-sacs) acting as nodes, and streets forming the edges between these. The resulting scores will then be averaged at the LSOA and OA levels to provide area measures for analysis with the crime rates. The basic CC algorithm calculates the total distance from the 'source' node, e.g. n=1 in Figure 4.18 below, to every other connected 'target' node in the network, and then divides the number of connected nodes considered (minus n=1, i.e. source node) by the total distance. This process is repeated until every node has been assigned a score, that is, has acted as a source node. However, the current work will employ a slightly more sophisticated version of the algorithm, namely 'scaled' CC, which penalises the resulting scores relative to the number of nodes in the entire graph (to be explained in detail later). Additional novelty will also be introduced through the use of distance thresholds to examine the effects of closeness centrality at different spatial scales on within-area Car Key burglary and Regular burglary rates. For example, an area might appear to be fairly well connected when only the 'local' street network is considered but it could actually be relatively disconnected from more distant areas in the study area.

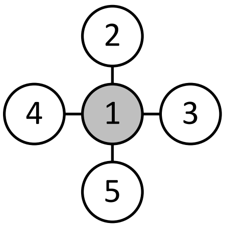


Figure 4.18 - Basic graph comprised of nodes and edges

4.6 Background

It is hypothesised that closeness centrality will exhibit a stronger positive relationship with Regular burglary than with Car Key burglary, primarily because the literature indicates that delinquency clusters close to urban centres (e.g. recall Shaw and McKay, 1942), which should, by definition, be relatively well connected to other locations in the West Yorkshire study area. However, a second consideration is whether any observed relationships between average CC and crime rates reflect offenders actively choosing to target some locations over others due to the permeability of the associated street network, that is, the ease by which it can be navigated (Johnson and Bowers, 2010, p.90). For example, an area containing lots of highly connected roads should be easier to traverse than one comprised of dead-ends and/ or cul-de-sacs, as well as offering more potential escape routes. Well-connected areas are also likely to experience more non-local traffic than poorly connected areas, meaning that offenders will be less likely to be challenged (Newman, 1972 in Johnson and Bowers, 2010, p.91), although some authors do argue that permeability actually reduces the risk of victimisation (Jacobs, 1961 in Johnson and Bowers, 2010, p.91). Generally speaking, street network permeability has been shown to increase the risk of burglary/ crime risk, with Johnson and Bowers (2010, pp.92-93) citing the following studies: Bevis and Nutter (1977) – census tract level, White (1990) – neighbourhood level, Beavon et al. (1994) – street segment level, and Armitage (2007) – household level. Further, Johnson and Bowers (2010) observed in their own analysis that burglary risk was higher on major roads and adjacent street segments, but lower on cul-de-sacs.

Although a number of studies already exist on the subject of area accessibility and residential burglary (e.g. see: Bernasco and Luykx, 2003; Johnson and Bowers, 2010; Frith et al., 2017; Rosser et al., 2017), the current research is the first to consider Car Key burglary in this context. Note that due to the absence of any offender data in the current research, it is not possible to control for this a regression model, other than by inferring the characteristics of high offender rate areas, and so the approach will not attempt to quantify the unique contribution of accessibility, but rather calculate the different impacts that it has on the two types of burglary.

4.7 Graph Theory and Centrality Measures

Graph theory is a field of mathematics in which networks are represented in abstract form, that is as graphs comprised of ‘nodes’ (points/ vertices) and ‘edges’ (lines/ links), the main purpose being to facilitate a better understanding of the connectivity between elements (Freeman, 1979; Boeing, 2017; Rodrigue and Ducruet, ca. 2017). The theory can be traced to the early 18th century when mathematician Leonhard Euler employed a series of nodes and edges to prove that the ‘Seven Bridges of Konigsberg’ problem could not be resolved; it was impossible to walk a continuous path across the city crossing each bridge once only (Boeing, 2017; Rodrigue and Ducruet, ca. 2017). Graph theory has since been applied in various research areas, including analysis of communication patterns in groups (e.g. see Bavelas, 1950; Freeman, 1979), and brain network analysis (e.g. see Fornito et al., 2016). Commonly used graph terminology is provided in Table 4.14 below for reference purposes.

Graph terminology	**Definition**
G	A graph (can be aspatial or spatial)
Node	Subject of interest e.g. person, location, telephone number
Edge	Link between two subjects
N	Number of nodes in a graph
Adjacent nodes	Pair of nodes sharing a single edge
Path	Successive edges connecting a pair of nodes
Distance	Number of edges on a path (one edge = distance 1)
Weighted distance	Sum of edge attributes along a path e.g. total length in km
Geodesic	Shortest-path between a pair of nodes
Connected graph	Every node can reach every other
Disconnected graph	Not all nodes can reach every other
Undirected graph	All edges are bi-directional (two-way travel)
Directed graph	Edges can be bi-directional or uni-directional (one-way travel)
Multidigraph	Directed graph that allows multiple edges between a pair of nodes

Table 4.14 - Commonly used graph terminology (Sources: Freeman, 1979, p.218; Fornito et al., 2016; Boeing, 2017, p.134)

Bavelas (1948) is credited with being the first to consider centrality in the context of human communication (e.g. see Freeman, 1979, p.215; Cohen et al., 2014), having theorised that influence in group processes might be related to structural position (Freeman, 1979, p.215). The notion of centrality in terms of a node's relative 'closeness' to all other nodes in a network was also introduced by Bavelas (1950), the thinking being that more central locations are better-placed than more peripheral locations to both send and receive information, thus enabling them to exert a level of control over associated flows (Fornito et al., 2016, pp.147-148). Beauchamp (1965 in Freeman, 1979, p.226; Fornito et al., 2016, p.148) formally defined the 'closeness centrality' of a node as the inverse of the average shortest-path distance from the node to all other nodes.

Referencing prior research on node centrality, Freeman (1979, pp.218-219) uses a star-shaped graph to identify the following structural attributes that render the central point in a network unique: (i) maximum 'degree centrality', (ii) maximum 'betweenness centrality', and (iii) maximum 'closeness centrality'. The 'degree centrality' of a node is the number of nodes to which it is adjacent (Freeman, 1979, p.219), and the 'betweenness centrality' of a node is the fraction of all shortest-paths that pass through it (Boeing, 2017, p.134). As can be seen in Figure 4.19, node u_1 is adjacent to all other nodes (v1, v2, v3, v4), it lies on the greatest number of shortest-paths (v1-v2, v2-v3, v3-v4, v4-v1, v1-v3, v2-v4), and it is closest to all other nodes ([5-1]/4 = 1). The CC value of a node will always be unity (1) if it is maximally close to all other nodes in the aspatial sense i.e. unweighted edges with individual distance values equal to one (Freeman, 1979, p.226; Fornito et al., 2016, p.138).

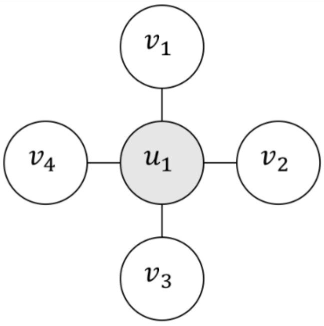


Figure 4.19 - Simple undirected 'star' graph comprised of five nodes and four edges (adapted from: Freeman, 1979, p.219; Fornito et al., 2016, p.138)

As mentioned above, edges in an aspatial graph each have a distance attribute value of one, for example, connections between individuals in a criminal organisation, but there are some circumstances where a spatial graph is more appropriate, including in transport network analysis. Here, edges can be weighted according to some measure of impedance (resistance) between two adjacent nodes, e.g. distance, travel time, cost. Since the current research is concerned with the proximity, or 'closeness', of areas to other areas, street network edges are weighted by their length in metres. The use of real-world distances is particularly relevant in the context of offender mobility because these are likely to influence journey-to-crime behaviours and associated crime patterns.

4.7.1 Closeness Centrality

The basic formula for calculating the closeness centrality score of an individual node u is shown in Equation 4.5; this is commonly referred to as the 'normalised' version. To summarise, the number of target nodes (v) to which a source node (u) is connected is divided by the sum of the shortest-path distances from u to each of these. The order of v, u is irrelevant in an undirected graph, but in the case of a directed graph it might not necessarily follow that the distance between a pair of connected nodes is the same in both directions (Wasserman and Faust, 1994, p.200). It is important, therefore, to state the direction of travel for the type of shortest-path algorithm that has been employed in a CC calculation, i.e. outbound from source node u to target node v, or inbound from target node v to source node u.

$C(u) = \dfrac{n-1}{\sum_{v=1}^{n-1} d(v,u)},$	Where: $\boldsymbol{C(u)}$ = normalised node closeness centrality score $\boldsymbol{n}$ = number of nodes v to which node u is connected, plus u $\boldsymbol{d(v,u)}$ = shortest-path distance between v and u

Equation 4.5 - Normalised node closeness centrality formula (Source: Hagberg et al., c2015)

4.7.2 Node CC Normalisation

Given that the sum of shortest-path distances in a closeness centrality calculation depends on the number of nodes being considered (Hagberg et al., c2015), this presents an issue for anyone seeking to compare CC scores derived from different sized graphs (Freeman, 1979, p.226). As subsequently noted by Freeman (1979, p.226), however, "Beauchamp (1965) has already solved this problem" by proposing that the number of nodes v to which a source node u is connected should be divided by the sum of associated shortest-path distances, thus generating an average measure of proximity. Another aspect of Beauchamp's formula to mention here is the inverse element, the effect of this being that larger scores indicate greater closeness centrality.

4.7.3 Node CC Scaling

One of the problems with only using the normalised version of the closeness centrality calculation is that, in disconnected graphs, smaller components might appear to be better connected than they actually are; this is because no account is taken of the size of the graph from which they are drawn. Wassermann and Faust's (1994, pp.200-201) solution to this is that CC scores should be scaled according to the number of nodes in G; "One can see that this index is a ratio of the fraction of the actors in the group who are reachable (…), to the average distance that these actors are from the actor (…)'' (Wassermann and Faust, 1994, p.201). A version of the Wassermann and Faust formula is given in Equation 4.6.

$C_{WF}(u) = \frac{n-1}{G-1} \frac{n-1}{\sum_{v=1}^{n-1} d(v,u)},$	Where: $\boldsymbol{C_{WF}(u)}$ = scaled node closeness centrality score $\boldsymbol{n}$ = number of nodes v to which node u is connected, plus u $\boldsymbol{d(v,u)}$ = shortest-path distance between v and u $\boldsymbol{G}$ = number of nodes in graph

Equation 4.6 - Scaled node closeness centrality formula (adapted from: Hagberg et al., c2004-2018)

Figure 4.20 below is included to illustrate the effects of node closeness centrality normalisation and scaling. Three different sized networks are presented, namely (A), (B), and (C), each with a source node (u), and 1, 3, and 5 target nodes (v) respectively. An assumption is made that u_1 is located exactly 50 m from every instance of v.

EXAMPLE 1: Basic/ normalised method – (A1), (B1), and (C1)

This example shows normalised closeness centrality scores calculated for each node u_1. As expected, all three networks produce the same CC score (0.02) despite them being different sizes. Although node CC normalisation does mitigate the 'sum of shortest-path distances problem', if each of the networks was a component part of a larger directed graph, i.e. one in which node u_1 cannot reach every other, as is likely to be the case for street networks, we would probably want to consider scaling the CC scores as well.

EXAMPLE 2: Improved/ scaled method – (A2), (B2), and (C2)

In this example, the scores are scaled relative to the size of G (assuming 1500 nodes). NB If a source node can reach every other node in a graph, then the normalised and scaled CC scores will be the same.

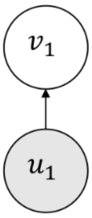
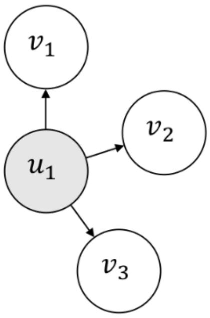
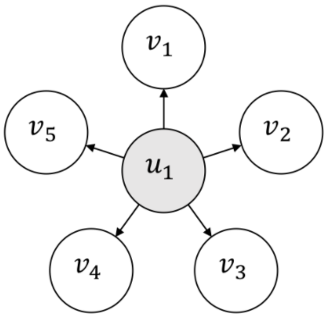
<table>
<tr><td><b>Network (A)</b>
 </td><td><b>Network (B)</b>
 </td><td><b>Network (C)</b>
 </td></tr>
<tr><td colspan="3"><b><u>EXAMPLE 1:</u> Basic/ normalised method</b> (assuming 50 m from node u_1 to every node v)</td></tr>
<tr><td><b>(A1)</b>
n minus 1 = 1
Total path length = 50 m
= 1/50
<b>CC normalised = 0.02</b></td><td><b>(B1)</b>
n minus 1 = 3
Total path length = 150 m
= 3/150
<b>CC normalised = 0.02</b></td><td><b>(C1)</b>
n minus 1 = 5
Total path length = 250 m
= 5/250
<b>CC normalised = 0.02</b></td></tr>
<tr><td colspan="3"><b><u>EXAMPLE 2:</u> Improved/ scaled method</b> (assuming G = 1500 nodes)</td></tr>
<tr><td><b>(A2)</b>
G minus 1 = 1499
= (1/1499)*(1/50)
<b>CC scaled = 0.00001</b></td><td><b>(B2)</b>
G minus 1 = 1499
= (3/1499)*(3/150)
<b>CC scaled = 0.00004</b></td><td><b>(C2)</b>
G minus 1 = 1499
= (5/1499)*(5/250)
<b>CC scaled = 0.00007</b></td></tr>
</table>

Figure 4.20 - Effects of node CC normalisation and scaling on different sizes of network (my diagram – based on: Freeman, 1979, p.219; Wassermann and Faust, 1994, pp.200-201; Fornito et al., 2016, p.138; Hagberg et al., c2004-2018; Hagberg et al., c2015)

4.7.4 Normalisation and Scaling with a Distance Threshold

Figure 4.21 below shows closeness centrality scores for a small street network (400 nodes) in Elland, West Yorkshire. A 500 m distance threshold was used to select target nodes for inclusion in the CC calculations, an approach that will be outlined in more detail in section 4.8.1 below. Map (A) shows the CC scores following normalisation and map (B) shows the CC scores following scaling. The results for both maps are classified into five equal intervals, the warmer colours indicating higher closeness centrality and the cooler colours indicating lower closeness centrality. In map (A), the **two red nodes** (north west of the network) are identified as being the most central, however, the use of a 500 m threshold has meant that only one target node has been taken into consideration in each of the associated calculations for these. In reality, we would probably consider the two nodes in question to be poorly connected relative to others in the graph, which is why the scaled method is more appropriate here. Map (B) appears to be a far more realistic representation of the network's structure.

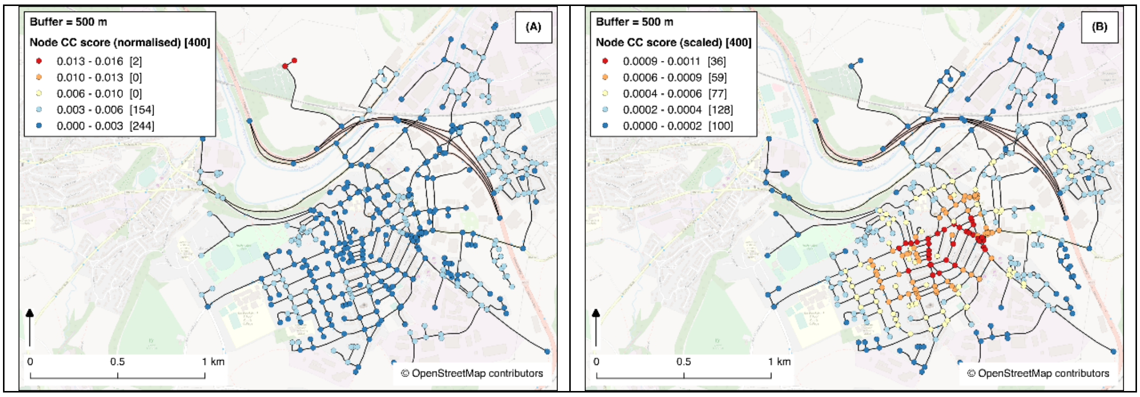


Figure 4.21 - Node CC score normalisation methods on a small network: (A) basic/normalised, (B) improved/scaled

4.8 Current Research Approach

OSMnxOSMnx (Boeing, 2017) is a Python tool that can be used to download and analyse OpenStreetMap (OSMnx) street networks. Streets can be obtained based on a 'place', 'address plus distance', 'lat-long point plus distance', 'polygon', or 'bounding box' (Boeing, 2017, p.131), and available network measures include 'betweenness centrality' and 'closeness centrality' (Boeing, 2017, p.134). The type of street network can also be specified e.g. 'drive', 'walk', 'bike', 'all' (Boeing, 2017, p.131). In the current research, West Yorkshire was passed to the 'place' variable and 'drive' was chosen as the network type (*drivable public streets but not service roads*). OSMnx then created a polygon from OSM's boundary geometry, buffered this by a distance of 500 m, simplified the street network (removed none-true network nodes), truncated the network to the bounds of the original polygon, and returned the West Yorkshire drivable street network as a multidigraph (G) (Boeing, 2014, pp.131-132). The effect of graph simplification is shown in Figure 4.22 below.

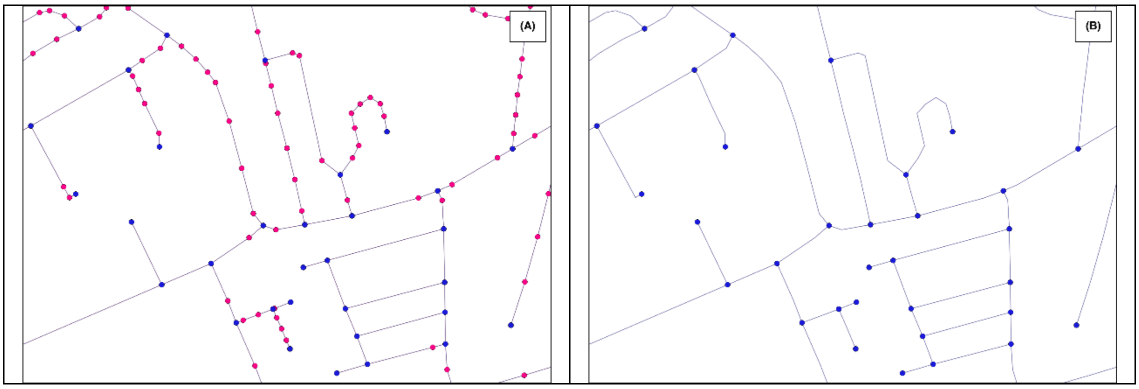


Figure 4.22 - Graph simplification method: (A) not simplified, (B) simplified (pink nodes removed) (based on: Boeing, 2017, p.132)

4.8.1 Adapted CC Algorithm

The OSMnx code (and associated NetworkX code – Hagberg et al., 2008) was adapted so that only nodes (v) within a user-specified distance threshold of a source node (u) were considered in the associated closeness centrality calculation. OSMnx uses Dijkstra's shortest-path algorithm (Hagberg et al., c2015) to calculate the distance between every source node u and every reachable target node v in a street network graph (outbound travel). Given the size of the West Yorkshire study area (approx. 2000 km^2), the algorithm outlined in Table 4.15 below was deemed to be more appropriate than the original in terms of being able to identify local variations in accessibility.

Other studies to have employed a cut-off in centrality calculations include Mahfoud et al. (2018) and Porta et al. (2009). The former used travel time thresholds of between one and five minutes to generate closeness centrality scores for the street network in Amsterdam; these were then averaged at the four-digit postal code level and included as predictors in a residential burglary model. Porta et al. (2009) used a distance threshold of 800 m to calculate closeness centrality scores for the street network in Bologna, Italy; these were then transformed into KDE surfaces, using different bandwidths, and correlations run between these and retail and service activity KDEs.

Step	Description
1	Download 'driveable' OSM street network for West Yorkshire
2	For node u_1, u_2 … calculate shortest-path to every reachable target node (v)
3	For node u_1, u_2 … disregard any target nodes located >= distance threshold
4	For node u_1, u_2 … calculate scaled CC score using selected target nodes

Table 4.15 - Adapted CC algorithm (based on: Hagberg et al., 2008; Boeing, 2017)

The adapted Python code outputs a .csv file where every row represents an individual node in the West Yorkshire street network, each having an associated closeness centrality score. Since the use of a distance threshold artificially creates lots of disconnected component parts, and also because the West Yorkshire street network contains many one-way edges, the improved/ scaled CC method was used. The OSMnx multidigraph of West Yorkshire was saved as two shapefiles (nodes and edges), as shown in Figure 4.23.

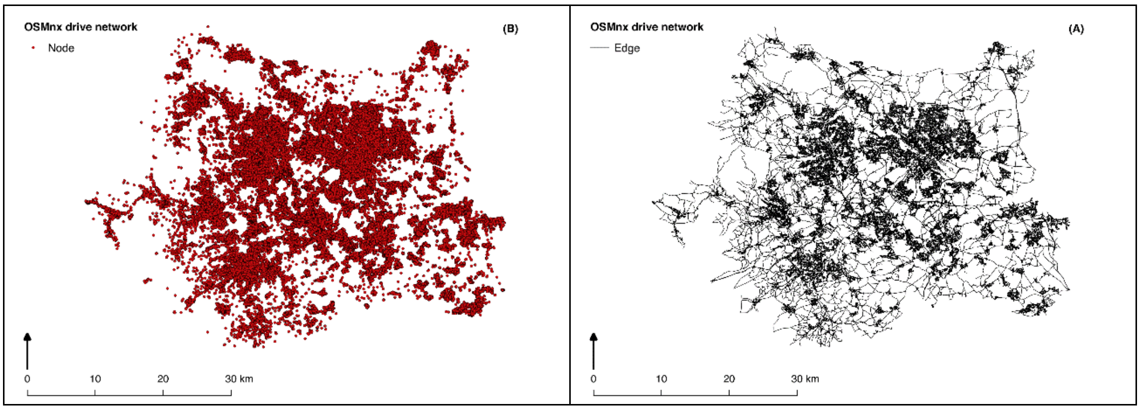


Figure 4.23 - OSMnx 'drive' network for West Yorkshire: (A) nodes, (B) edges

Map (A) in Figure 4.24 below shows a small section of the West Yorkshire street network (the Burley area of Leeds) as it appears in OpenStreetMap. Map (B) shows the same street network represented in graph format – the nodes are classified according to CC scores; these having been derived using a 500 m distance threshold. Maps (C) and (D) show the scores aggregated to LSOAs and OAs respectively; a GIS was used to perform a spatial join between the nodes and the areal units based on match option 'intersect' and merge rule 'median', the latter because the scores were positively skewed. Comparing maps B, C, and D, it is clear that the LSOA averaging has diluted the highest scores, whereas they are more accurately represented in (D).

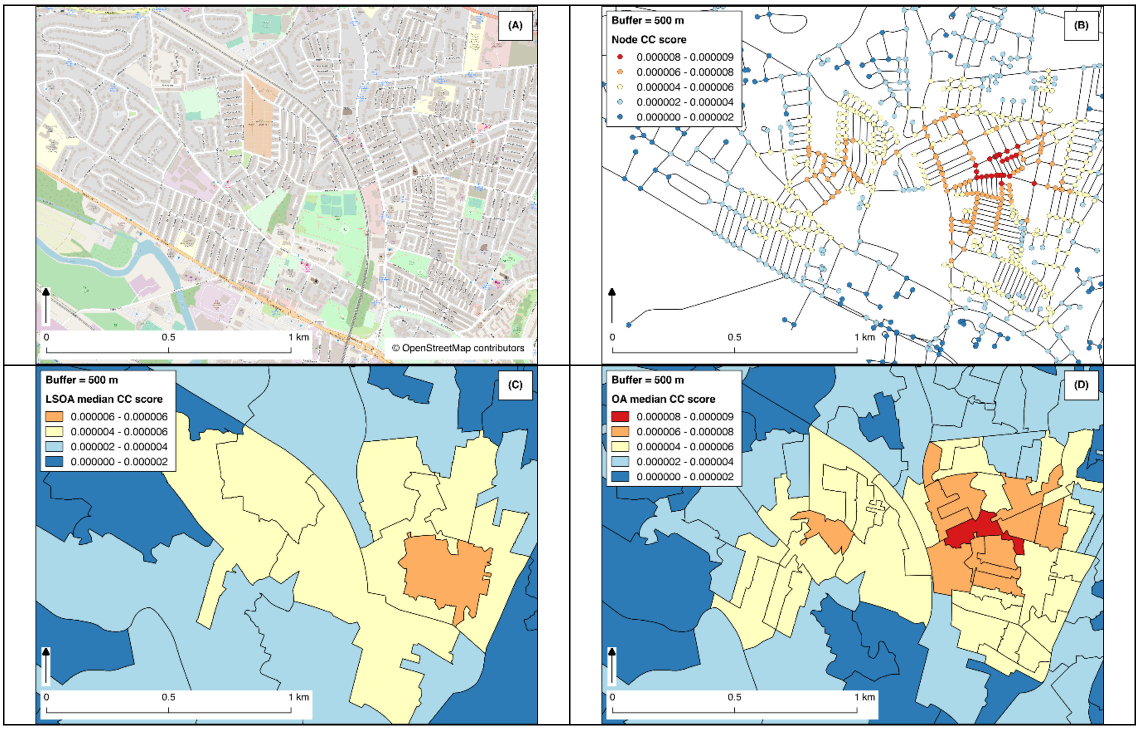


Figure 4.24 - (A) OpenStreetMap (Burley area, Leeds), (B) Node CC scores, (C) LSOA median CC scores, (D) OA median CC scores

Maps (A)-(D) in Figure 4.25 below show LSOA median closeness centrality scores for the whole of the West Yorkshire study area. The main difference between the maps is the distance threshold that was used to select target nodes for inclusion in the closeness centrality calculations, for example, in map (A), the CC scores were generated for individual nodes (e.g. u_1, u_2 …) by only selecting those reachable target nodes (v) located < 500 m along the drivable street network. Each median CC data set was subsequently classified into five equal intervals because the resulting scores varied depending on the distance threshold used. The warmer areas in each of the maps indicate that nodes in those LSOAs are, on average, 'closer' (scaled) to all reachable nodes within the specified distance threshold than are the nodes in other areas.

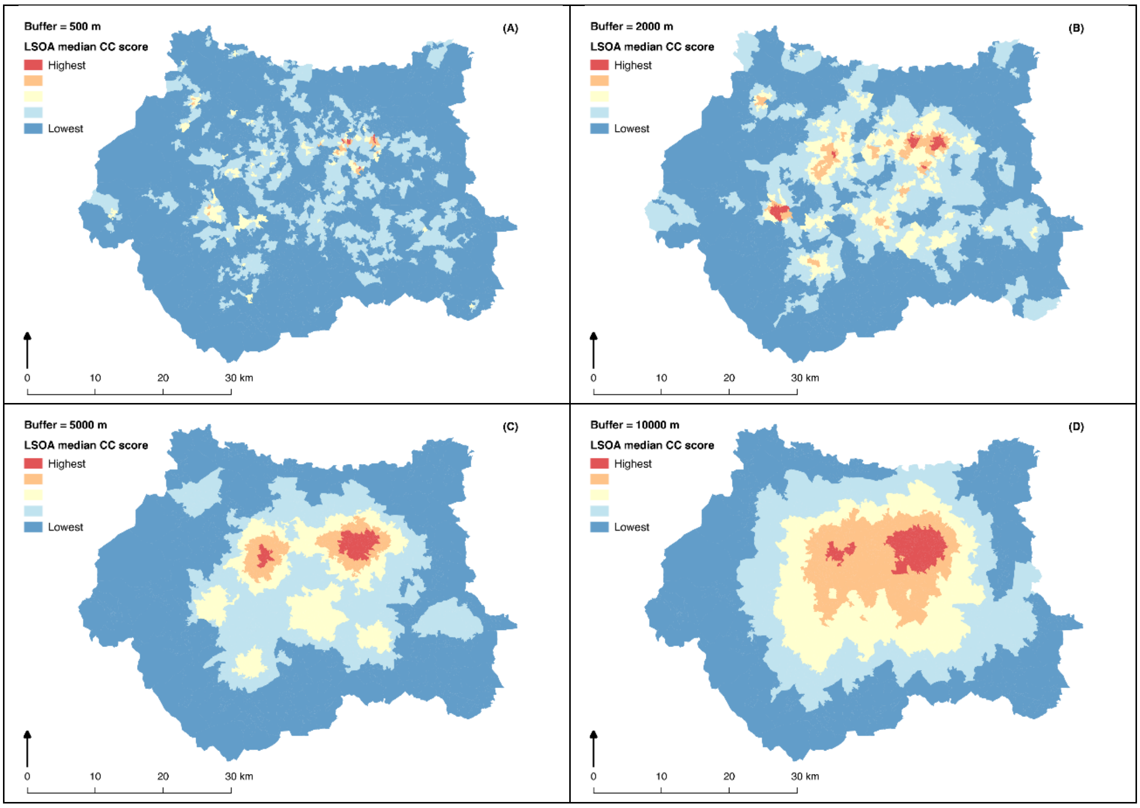


Figure 4.25 - West Yorkshire LSOA median CC scores for: (A) 500 m, (B) 2000 m, (C) 5000 m, and (D) 10000 m distance thresholds

Maps (A)-(D) in Figure 4.26 below show a zoomed in section of the West Yorkshire street network (Leeds, including city centre and North West areas). These maps illustrate in more detail how choice of distance threshold influences median closeness centrality scores at the LSOA level (relative to the whole of West Yorkshire). The same classification scheme has been used as in Figure 4.25, with the exception that the lowest category has been changed to transparent so that the underlying street network can be seen.

- **Map (A) – 500 m threshold:** the least connected areas contain, amongst other things, cul-de-sacs, rural spaces, and physical barriers, e.g. river, rail.
- **Map (B) – 2000 m threshold:** areas appear better connected but one LSOA did drop down to the lowest class (New Farnley – bordered by fields).
- **Map (C) – 5000 m threshold:** no areas in the lowest class.
- **Map (D) – 10000 m threshold:** all areas now fall into one of the top two classes.

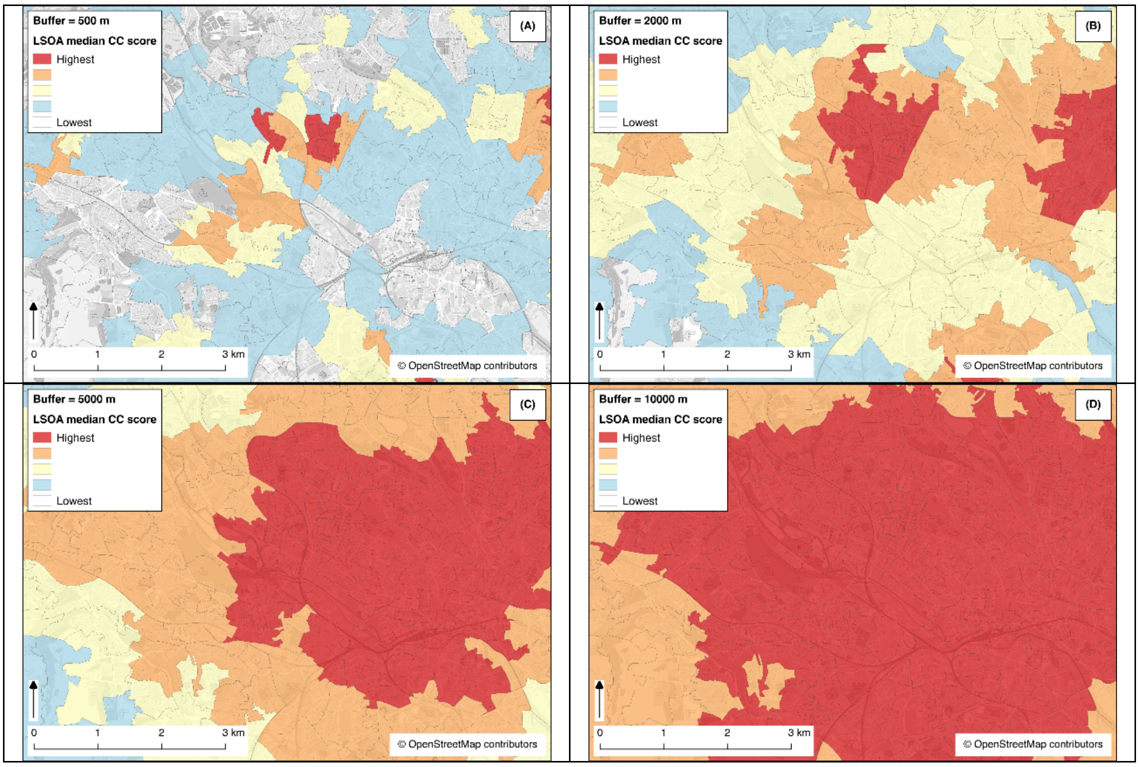


Figure 4.26 - Leeds, incl. city centre and North West areas, LSOA median CC scores for: (A) 500 m, (B) 2000 m, (C) 5000 m, and (D) 10000 m distance thresholds

4.9 Results and Discussion

Table 4.16 below contains the Spearman's Correlation coefficients for the Car Key burglary rates and LSOA and OA median closeness centrality scores for the four buffer distances tested, and Table 4.17 below contains the coefficients for the Regular burglary rates. As expected, the Regular burglary rates are more positively correlated with the median area closeness centrality scores than are the Car Key burglary rates. The strongest correlation (.690) was between Regular burglary and the 5000 m distance threshold scores, which might reflect both the spatial characteristics of high offender rate areas (recall Shaw and McKay, 1942), and that higher risk populations, e.g. students, typically reside in more central areas. Further, because more centrally located areas are likely to attract greater volumes of pedestrian and vehicular traffic, they are more likely to fall within offenders' awareness spaces. Given that the weakest correlations for Regular burglary were generated by the 500 m distance threshold, another factor to consider here is the scale at which offenders conceptualise street network permeability. To use the example of New Farnley (residential streets surrounded by fields), which was mentioned in relation to Figure 4.26, when a longer distance threshold was used (2000 m), the median closeness centrality for the associated LSOA was poorer relative to others parts of the study area than when a shorter distance threshold was used (500 m). Perhaps, therefore, Regular burglars first select potential target areas based on their assumed accessibility at larger spatial scales, considering such factors as ease of access, and how readily they might escape, and then seek out the most permeable streets within these, i.e. a hierarchical process as per Bernasco and Luykx, 2003.

Looking at Table 4.16 below, median area closeness centrality appears to have far less of a bearing on Car Key burglary rates than on Regular burglary rates, a possible explanation for this being that the type of properties that desirable vehicles are likely to be parked outside overnight are perhaps less prevalent in more central areas. However, that is not to say that Car Key burglars do not consider accessibility at all, for example, the strongest correlation for the crime type (.230) was generated by the 10000 m distance threshold, i.e. this might reflect Car Key burglars conceptualising area accessibility in relation to their (assumed) far more extensive criminal activity spaces. Also, even though Car Key burglars' area-level target selection choices are likely to be determined primarily by the supposed prevalence of desirable vehicles, they might still filter potential target locations within these based on perceived ease of access/ egress (street network permeability), however, it is not possible to test this hypothesis at the area level.

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

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

	Distance threshold			
Unit	**500 m**	**2000 m**	**5000 m**	**10000 m**
LSOA	-.061*	-0.005	.112**	.230**
OA	-0.003	0.011	.075**	.136**

Table 4.16 - Spearman's rho for Car Key burglary rates and median closeness centrality scores at LSOA and OA level

	Distance threshold			
Unit	**500 m**	**2000 m**	**5000 m**	**10000 m**
LSOA	.235**	.508**	.690**	.646**
OA	.162**	.407**	.559**	.525**

Table 4.17 - Spearman's rho for Regular burglary rates and median closeness centrality scores at LSOA and OA level

It is worth noting here some limitations of the approach employed, including that: the closeness centrality variable might be strongly correlated with other explanatory variables, e.g. within-area offender rates; only the driveable street network was included in the analysis, meaning that walking routes were not considered; only the West Yorkshire street network was analysed, which could have resulted in edge effects – Frith et al. (2017) created a buffer around their study area to address this issue; only outbound closeness centrality was considered; some of the OpenStreetMap edge attributes were incorrectly coded, e.g. bi-directional roundabouts, which will have introduced error into the distance calculations and also highlights potential issues with using crowdsourced data; the adapted algorithm took a long time to generate results (15 hours and 20 mins for the 10000 m threshold), hence, only the West Yorkshire street network was analysed; and the analysis is based on street network distances, whereas Kent et al. (2006, p.197) suggest that offenders think in terms of straight line distances.

Chapter 5 Developing the Static and Dynamic Risk Surfaces

5.1 Introduction

The main purpose of this chapter is to determine suitable spatio-temporal parameters for the current research dynamic model, and to create the static risk surface for the combined model. Figure 5.1 below shows the key steps by which this will be achieved; the first part of the chapter will describe the process of deriving the static risk surface, and the second the dynamic surfaces. Spearman's Rank Correlation analysis will also be performed between each of the LSOA and OA Car Key burglary and Regular burglary rates samples and independent variables from Chapters 3 and 4 to understand if relationships vary by crime type. The best performing independent variables for Car Key burglary will be then used as inputs in two types of regression model, namely: OLS and *k*-NN, to ascertain which approach is able to best explain the variation in Car Key burglary rates at the LSOA level. It is anticipated that the non-normal distribution of the crime rates will invalidate some of the assumptions of the OLS model, hence the inclusion the non-parametric *k*-NN model. Once an appropriate static risk surface has been created, the second part of the chapter will present the findings of RV/N-RV analysis undertaken using Ratcliffe's Near Repeat Calculator. The RV/N-RV results will be used to determine the size of the buffers in the dynamic model and the frequency by which they should be updated. To summarise, the two key outputs from this chapter will be (i) a static (flag account) risk surface, and (ii) spatio-temporal parameters for the dynamic (boost account) risk surfaces.

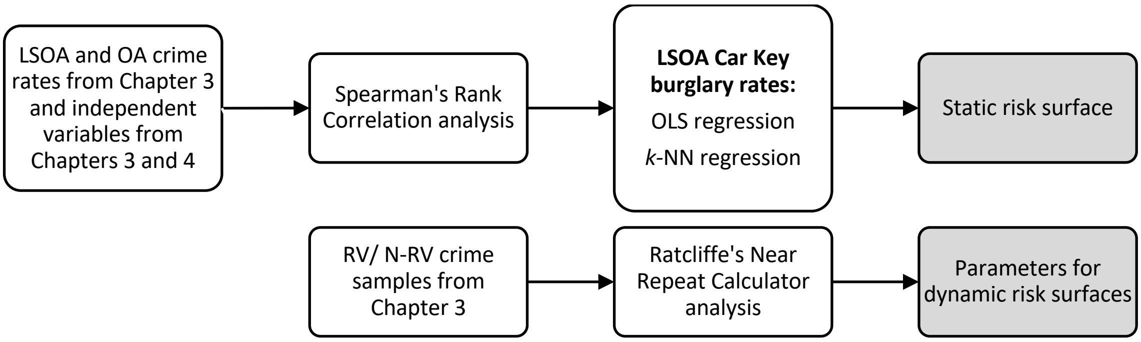

LSOA and OA crime rates from Chapter 3 and independent variables from Chapters 3 and 4
Spearman's Rank Correlation analysis
LSOA Car Key burglary rates:
OLS regression
k-NN regression
Static risk surface
RV/ N-RV crime samples from Chapter 3
Ratcliffe's Near Repeat Calculator analysis
Parameters for dynamic risk surfaces

Figure 5.1 - How the risk surface outputs will be derived for the current research dynamic model and combined model

5.2 Correlation Analysis

This section will present the results of the Spearman's Rank Correlation analysis, and then relate the main findings to the existing literature. However, because the Car Key burglary correlations that were calculated at the Lower Super Output Area (LSOA) geography, which contains larger areas than output areas (OAs), were not overly strong, it was decided to create the static risk layer at the LSOA level, as this geography was expected to generate a better fitting model. Hence, the results discussed in this chapter all relate to LSOAs. With a view to parsimony, only the top twenty variables are included for each burglary type and each directional relationship (positive/ negative). Further, to facilitate additional comparison between the two crime types, for each variable in a top 20, the equivalent correlation is provided between the variable and the other burglary type rates. For example, in Table 5.1 below, which shows the top 20 positively correlated Car Key LSOA variables, the Regular burglary equivalent correlations for each of these are shown in the adjacent column. So, using the '% Households 2+ cars/vans' variable as an example, this falls within the top 20 most positively correlated variables for the LSOA Car Key burglary rates at 0.272, whereas the Regular burglary equivalent for this variable is -0.450. The coefficients have also been shaded according to the strength and direction of the relationships, with dark green representing the most positively correlated variables and dark red the most negatively correlated. The shading is based on the max and min coefficients in the results, i.e. dark green does not represent a perfect correlation of + 1.00, rather the strongest identified positive correlation of 0.690 (Table 5.2 – Regular burglary - CC_5000_m). Recall that this 'average area accessibility' variable (closeness centrality based on street network nodes) was created by the author specifically for the current work and discussed at length in Chapter 4.

Reviewing the general trends first, it is apparent that a large proportion of the significant correlations run in opposite directions for the two crime types. For example, looking at the 20 most negatively correlated variables for Car Key burglary (Table 5.3), 90% of the Regular burglary equivalents are positive, whereas for the 20 most negatively correlated variables for Regular burglary (Table 5.4), 95% of the Car Key burglary equivalents are positive (although one is not significant). It is also pertinent to mention the moderate positive correlation of 0.337 between Car Key burglary rates and Regular burglary rates (Table 5.1), although this is not wholly unexpected given the findings of the exploratory spatial analysis. For example, if a LSOA with a high Car Key burglary rate is situated near to a high offender rate/ high offender count area type, then it is also likely to be attractive to Regular burglars. Further, the crime samples selection method might have caused some attempt Car Key burglary offences to be misclassified as

Regular burglaries, thus causing a correlation between the two. The overall patterns observed here, plus the notably stronger maximum correlations for the Regular burglary rates, were all hypothesised in Chapter 3 of the thesis.

Other than the crime samples selection method, another factor that might have influenced the generally weaker correlations for Car Key burglary is the inability of some census variables to adequately differentiate between relevant area characteristics for the crime type in question. For example, it could be that desirable vehicles are frequently stolen from semi-detached households in more affluent areas, however, much of the UK's social housing stock is also semi-detached, which is perhaps less likely to be targeted by offenders seeking desirable vehicles. Similarly, and as noted by Lansley (2016, p.264), due to car ownership becoming more widespread over recent years, and also a wide range of vehicle values, car ownership might now be less relevant as a proxy indicator of affluence.

TOP 20 POSITIVE CAR KEY VARIABLES:		**CKB**	**Sig.**	**REG**	**Sig.**
1	IDW_0.5	0.382	0.01	-0.156	0.01
2	IDW_0.35	0.381	0.01	-0.200	0.01
3	IDW_0.2	0.362	0.01	-0.265	0.01
4	IDW_1.0	0.360	0.01	-0.090	0.01
5	**Regular_burg_rate**	**0.337**	**0.01**	**1.000**	**N/A**
6	IDW_0.1	0.330	0.01	-0.317	0.01
7	IDW_2.0	0.318	0.01	-0.054	0.05
8	%_16_74_NS-SeC_Intermed	0.313	0.01	-0.352	0.01
9	%_hshlds_own_prop	0.298	0.01	-0.349	0.01
10	IMD_2015_Decile	0.287	0.01	-0.453	0.01
11	IDW_3.0	0.287	0.01	-0.043	**NS**
12	%_emp_16_74_wk_infor	0.285	0.01	-0.191	0.01
13	%_16_74_private_trans_wk	0.278	0.01	-0.445	0.01
14	%_hshlds_2+_cars_vans	0.272	0.01	-0.450	0.01
15	%_16_74_NS-SeC_Higher	0.264	0.01	-0.411	0.01
16	%_emp_16_74_wk_finan	0.259	0.01	-0.119	0.01
17	%_hshlds_semi-det_house	0.252	0.01	0.036	**NS**
18	%_emp_16_74_wk_educa	0.249	0.01	-0.045	**NS**
19	%_16+_highest_qual_L4+	0.244	0.01	-0.263	0.01
20	%_16+_highest_qual_L3	0.240	0.01	-0.177	0.01

Table 5.1 - Top 20 positive Car Key burglary LSOA variables with Regular burglary equivalents

TOP 20 POSITIVE REGULAR VARIABLES:		**REG**	**Sig.**	**CKB**	**Sig.**
1	CC_5000_m	0.690	0.01	0.112	0.01
2	Index_of_Diversity	0.675	0.01	0.025	**NS**
3	%_mixed	0.648	0.01	0.090	0.01
4	CC_10000_m	0.646	0.01	0.230	0.01
5	%_Black	0.613	0.01	0.042	**NS**
6	%_not_speak_Eng	0.576	0.01	-0.126	0.01
7	%_Pakistani	0.542	0.01	-0.019	**NS**
8	%_Chinese	0.542	0.01	-0.000	**NS**
9	%_Indian	0.516	0.01	0.140	0.01
10	CC_2000_m	0.508	0.01	-0.005	**NS**
11	%_16+_school_FT_students	0.506	0.01	-0.019	**NS**
12	%_Arab	0.501	0.01	0.006	**NS**
13	%_born_new_EU	0.472	0.01	-0.181	0.01
14	Popn_density	0.467	0.01	0.003	**NS**
15	%_hshlds_less_rooms	0.460	0.01	-0.234	0.01
16	%_16+_single	0.454	0.01	-0.101	0.01
17	%_Bangladeshi	0.451	0.01	-0.055	0.05
18	%_16_74_unemployed	0.448	0.01	-0.259	0.01
19	%_16_74_public_trans_wk	0.445	0.01	0.072	0.01
20	%_HRPs_aged_24_or_less	0.427	0.01	-0.199	0.01

Table 5.2 - Top 20 positive Regular burglary LSOA variables with Car Key burglary equivalents

TOP 20 NEGATIVE CAR KEY VARIABLES:		CKB	Sig.	REG	Sig.
1	IMD_Employ_Dep_Score	-0.304	0.01	0.380	0.01
2	IMD_Income_Dep_Score	-0.282	0.01	0.424	0.01
3	%_hshlds_social_rent	-0.274	0.01	0.206	0.01
4	%_16_74_unemployed	-0.259	0.01	0.448	0.01
5	%_emp_16_74_wk_trans	-0.243	0.01	0.288	0.01
6	%_16_74_NS-SeC_Routine	-0.238	0.01	0.108	0.01
7	%_hshlds_less_rooms	-0.234	0.01	0.460	0.01
8	%_emp_16_74_wk_whole	-0.230	0.01	0.124	0.01
9	%_16+_divorced	-0.229	0.01	-0.011	**NS**
10	%_HRPs_aged_24_or_less	-0.199	0.01	0.427	0.01
11	%_emp_16_74_wk_accom	-0.195	0.01	0.361	0.01
12	%_emp_16_74_wk_manuf	-0.192	0.01	-0.070	0.01
13	%_16_74_walk_cycle_wk	-0.188	0.01	0.271	0.01
14	%_born_new_EU	-0.181	0.01	0.472	0.01
15	%_hshlds_flat	-0.170	0.01	0.233	0.01
16	%_hshlds_terraced_house	-0.158	0.01	0.130	0.01
17	%_hshlds_lone_par_dep_childn	-0.153	0.01	0.300	0.01
18	%_emp_16_74_wk_PT	-0.145	0.01	0.266	0.01
19	%_diff_address_one_year_ago	-0.145	0.01	0.323	0.01
20	%_aged_0_to_4	-0.132	0.01	0.391	0.01

Table 5.3 - Top 20 negative Car Key burglary LSOA variables with Regular burglary equivalents

TOP 20 NEGATIVE REGULAR VARIABLES:		REG	Sig.	CKB	Sig.
1	%_white	-0.654	0.01	0.004	**NS**
2	%_born_UK_Ireland	-0.599	0.01	0.076	0.01
3	%_hshlds_no_childn	-0.528	0.01	0.183	0.01
4	%_aged_45_to_64	-0.506	0.01	0.144	0.01
5	IMD_2015_Decile	-0.453	0.01	0.287	0.01
6	%_hshlds_2+_cars_vans	-0.450	0.01	0.272	0.01
7	%_16_74_private_trans_wk	-0.445	0.01	0.278	0.01
8	%_hshlds_detached_house	-0.441	0.01	0.152	0.01
9	%_16_74_NS-SeC_Higher	-0.411	0.01	0.264	0.01
10	%_aged_65_to_89	-0.403	0.01	0.106	0.01
11	%_emp_16_74_wk_minin	-0.397	0.01	0.062	0.05
12	%_prov_unpaid_care	-0.376	0.01	0.113	0.01
13	%_16_74_NS-SeC_Intermed	-0.352	0.01	0.313	0.01
14	%_hshlds_own_prop	-0.349	0.01	0.298	0.01
15	%_16+_married	-0.343	0.01	0.151	0.01
16	%_emp_16_74_wk_agric	-0.330	0.01	-0.065	0.05
17	IDW_0.1	-0.317	0.01	0.330	0.01
18	%_emp_16_74_wk_publi	-0.312	0.01	0.198	0.01
19	%_emp_16_74_wk_FT	-0.266	0.01	0.145	0.01
20	IDW_0.2	-0.265	0.01	0.362	0.01

Table 5.4 - Top 20 negative Regular burglary LSOA variables with Car Key burglary equivalents

Now to look in more detail at some of the relationships between the two burglary rates and individual variables. For the Car Key burglary rates, the strongest positively correlated variable is 'IDW_05' (moderate, 0.382). Recall from Chapter 4 that this variable represents the percentage point difference between the percentage households social or private renting, 2011 within a LSOA and the inverse distance weighted average of the same for surrounding LSOAs (beta = 0.5; average trip distance = 3.6 km). Note that the coefficient for the IDW_0.35 variable (ATD = 5.0 km) is very similar at 0.381. To summarise earlier discussions regards the IDW variables, these are hypothesised to reflect spatial juxtapositions of the target property type and possible high offender rate/ high offender count areas, which, as mentioned above, might also go some way to explaining the moderate correlation observed between the two burglary rates. Looking at the top 20 positively correlated variables for Car Key burglary that are not IDW-derived, these would all appear to be proxy indicators of the likely presence of desirable vehicles within a LSOA. For example, as the IMD 2015 Decile increases for a LSOA, we might also expect the number of desirable vehicles to increase (recall that for the IMD 2015 Deciles, 1 = most deprived 10% of LSOAs in the country and 10 = least deprived 10%).

Given the observed moderate positive correlation of 0.313 between Car Key burglary rates and the '% Persons aged between 16 and 74 **NS-SeC Intermediate**' variable, it is useful to again reference Lansley (2016) who segmented cars registered in England and Wales in 2011, based chiefly on physical structure, and then analysed LSOA proportions of these by 8 NS-SeC groups. The ten car segments in Lansley's study, together with example makes and models, are listed in Table 5.5 below.

Car segment	**Example make and model**
City	Ford Ka
Superminis	Ford Fiesta
Small family	Ford Focus
Large family	Ford Mondeo
MPVs	Vauxhall Zafira
Compact executive	BMW 3 Series
Executive cars	Mercedes-Benz E Class
Sports	Audi TT
SUVs	Land Rover Discovery
Luxury	Mercedes-Benz S Class

Table 5.5 - Car segments and example makes and models in Lansley's study (Source: Lansley, 2016, p.269)

Although Lansley (2016) used the 8 class version of the NS-SeC, the classification is nested and so the NS-SeC Intermediate variable in the current work equates to classes 3 and 4 in Lansley's analysis, namely *'intermediate occupations'* and '*small employers and own account workers'*. Looking at the car segments that are positively correlated with these two NS-SeC groups might help to explain the positive correlation between the NS-SeC Intermediate variable and Car Key burglary rates in the current work, i.e. do the correlated segments represent desirable vehicles? Interestingly, the positively correlated car segments for class 3 were 'City', 'Supermini', and 'Sports', and for class 4 'City', 'Compact executive', 'Executive', 'Sports', 'SUV', and 'Luxury' (Lansley, 2016, p.280). Lansley (2016) also analysed three vehicle age groups, namely 0-3, 4-10, and 11+ years, and ascertained that neither of the Intermediate NS-SeC classes were positively correlated with the 11+ group (p.280).

Although the majority of the aforementioned correlations make theoretical sense in relation to the positive correlation between Car Key burglary rates and the NS-SeC Intermediate variable in the current work, i.e. desirable cars and relatively new cars, it is surprising that the 'Small family' segment (e.g. Ford Focus) was not positively correlated with either of the Intermediate NS-SeC classes in the Lansley (2016) study. For example, West Yorkshire Police (2020) state in reference to Car Key burglaries that: “Some of the makes and models stolen include **VW Golf, Vauxhall Astra, Ford Focus,** Ford Fiesta, Audi A3 and Vauxhall Corsa” (my emphasis). Given that the ‘Small family’ segment was most positively correlated with the ‘semi-routine’ (0.395) and ‘routine occupations’ (0.391) NS-SeC classes in Lansley’s analysis, and that these classes were also positively correlated with the 11+ age group, but not with the other two age groups, it is possible that some area type trends might have been masked due to the data disaggregation method. To explain, if the ‘Small family’ cars segment had been further sub-divided, such as into older/ newer models, or average performance/ high-performance models, e.g. Golf GTI and Focus RS, then different NS-SeC group correlations might have emerged. It is also worth mentioning that Lansley’s equivalent classes for the **NS-SeC Higher** variable used in the current work – 0.264 correlation with the Car Key burglary rates – were both positively correlated with the ‘City’, ‘Compact executive’, ‘Executive’, ‘Sports’, ‘SUV’, and ‘Luxury’ car segments, as well as the two younger age groups (Lansley, 2016, p.280). Considering all of this in conjunction with the fact that the ‘Executive’, ‘Sports’, ‘SUV’, and ‘Luxury’ car segments were all negatively correlated with the bottom four NS-SeC classes in the Lansley (2016) study, it is likely that the positive correlations observed between the LSOA Car Key burglary rates and the NS-SeC Intermediate and Higher groups in the current work are, at least in part, attributable to the presence of certain desirable vehicle types.

Turning now to the top 20 most negatively correlated variables for the Car Key burglary rates (Table 5.3), the strongest of these, at -0.304, is the IMD 2015 Employment Deprivation Domain Score, followed by the Income Deprivation Domain Score at -0.282. These correlations are as might theoretically be expected, i.e. as the proportion of a LSOA population experiencing employment/ income deprivation increases, thefts of desirable vehicles decrease. It appears that the majority of the Car Key burglary variables in Table 5.3 represent the prevailing characteristics of those area types where desirable vehicles might not typically be prevalent, assumed here to be mainly for reasons of affordability, but also perhaps because motor vehicles are not the preferred mode of transport in some LSOAs. For example, professionals living in city centre flats might find cycling to work to be more convenient/ cost-effective than driving (%_16_74_walk_cycle_wk = -0.188). It is also interesting to note that, for the 20 most negatively correlated Car Key burglary variables, 50 per cent of the equivalent Regular burglary variables are moderately positively correlated, and some of these are indicative of social disorganisation. Examples include ‘per cent usual residents lived at a different address one year ago’ (residential mobility) and ‘per cent households lone parent with dependent children’ (family disruption) (e.g. see Tseloni, 2002, p.110). Although Regular burglary is not the focus here, it is notable that the related correlations are generally in line with the existing literature (e.g. see Budd, 2001, pp.2-3), and also that two of the variables created by the author are the most strongly correlated with the LSOA Regular burglary rates, namely ‘CC_5000_m’ (0.690) and ‘Index_of_Diversity’ (0.675).

To summarise, the results presented in this section have highlighted some notable differences between the two burglary types and their relationship with a selection of independent variables, and the findings therefore support the research rationale regards heterogeneity of offence characteristics within the Home Office’s ‘Burglary Dwelling’ (residential burglary) classification.

5.3 Creating the Static Risk Surface

This section will now describe how variables from Chapters 3 and 4 were used to create the static risk surface for the current research combined model. Section 5.3.1 will report the findings of the OLS regression models that were initially tested, followed by a detailed explanation of the *k*-NN regression approach that was ultimately chosen.

5.3.1 OLS Regression

Given the highly skewed nature of the crime rates data, it was expected that some of the assumptions of OLS, including normally distributed residuals (Charlton and Fotheringham, 2009, p.1), would be violated. This proved to be the case, as will be demonstrated below. Nevertheless, the OLS regression method and results will be discussed here, both to justify the subsequent choice of a non-parametric modelling approach and also to establish a benchmark for the *k*-NN regression results that are discussed in the next section. Because the top 20 positive and top 20 negative Spearman’s Rank Correlation coefficients for the study variables were not overly strong for LSOAs – moderate/ weak based on Cohen (1988) – and even less so for OAs, it was decided to create the static risk layer at the LSOA level only. Due to the large number of potential predictor variables in the research, only the thirty most strongly, and significantly, correlated (in either direction) variables with Car Key burglary rates were considered for inclusion in the multivariate regression analysis. To limit multicollinearity, the selected variables were then ordered from high to low in terms of the strength of their relationship with the dependent variable and any that had a correlation of 0.750 or greater with another variable that was more strongly correlated with the Car Key rates variable were deleted.

This gave 14 variables, although one (%_16+_highest_qual_L3) was subsequently deselected because its relationship with the dependent variable did not appear to be linear when graphed. The thirteen remaining variables, shown in Table 5.6 below, were added sequentially by strength of correlation to an OLS regression in IBM SPSS Statistics 22, and any that became non-significant when added, or where there was evidence of multicollinearity (Condition Index value > 15) (School of Geography, UOL, no date), were disregarded. Note that a stepwise regression, following the method outlined in School of Geography, UOL (no date), was performed in SPSS prior to employing the manual variable selection method but the final model did not make much theoretical sense to the author.

Variable short name	**Variable description**	**Denominator**	**Coefficient**
IDW_0.5	Percentage point difference between LSOA % Households social/ private renting and surrounding LSOAs IDW average (beta = 0.5, ATD = 3.6 km)	n/a	0.382
Regular_burg_rate	Regular burglary rate per 1,000 households, mid-point year 2010-2014	All households	0.337
%_16_74_NS-SeC_Intermed	% Persons aged between 16 and 74 NS-SeC Intermediate (current/ last main job)	All usual residents aged 16 to 74	0.313
IMD_Employ_Dep_Score	IMD 2015 Employment Deprivation Domain Score	n/a	-0.304
%_emp_16_74_wk_infor	% Employed persons aged between 16 and 74 who work in the information and communication or professional, scientific and technical activities industries	All usual residents aged 16 to 74 in employment the week before the census	0.285
%_emp_16_74_wk_finan	% Employed persons aged between 16 and 74 who work in the financial, insurance or real estate industries	All usual residents aged 16 to 74 in employment the week before the census	0.259
%_hshlds_semi-det_house	% Household spaces semi-detached house or bungalow	All household spaces	0.252
%_emp_16_74_wk_educa	% Employed persons aged between 16 and 74 who work in the education sector	All usual residents aged 16 to 74 in employment the week before the census	0.249
%_emp_16_74_wk_trans	% Employed persons aged between 16 and 74 who work in the transport or storage industries	All usual residents aged 16 to 74 in employment the week before the census	-0.243
%_hshlds_less_rooms	% Households who have one fewer or less rooms than required	All households	-0.234
CC_10000_m	Average area closeness centrality based on street network nodes within 10000 m buffer	n/a	0.230
%_emp_16_74_wk_whole	% Employed persons aged between 16 and 74 who work in the wholesale and retail trade; repair of motor vehicles and motor cycles industries	All usual residents aged 16 to 74 in employment the week before the census	-0.230
%_16+_divorced	% Persons aged over 16 who are divorced or separated	All usual residents aged 16 and over	-0.229

Table 5.6 - Thirteen LSOA variables selected for multivariate OLS regression testing - All sig. at the 0.01 level (2-tailed)

The five independent variables in the initial best fitting OLS model, with no obvious issues regards multicollinearity and/ or non-significance, are shown in Table 5.7 below. However, because the Spearman's Rank coefficients were very similar for the IDW_05 and IDW_0.35 variables, and also recalling Carden's finding that Car Key burglars travel a median distance of 4.9 km from home base to offence (2012, p.74), a second model was tested, replacing the IDW_0.5 (ATD = 3.6 km) variable with the IDW_0.35 (ATD = 5.0 km) variable, but leaving the other four the same. This increased the adjusted R square very slightly to 0.394, and so the variables in Table 5.8 below were assumed to be the best at explaining LSOA level variations in Car Key burglary rates. Adjusted R square is preferable for measuring performance because it penalises non-parsimony (Fogarty, 2019, p.198).

Given earlier discussions around the limitations of the crime data and crime samples selection method, a model that is able to explain just under 40% of the variation in the dependent variable is encouraging. It is also interesting to note that, after the Regular burglary rates, the next most important variable, based on standardised coefficients for the model, was created by the author specifically for the purposes of the current work, namely IDW_0.35. Looking at the five variables in the best performing OLS model (Table 5.8), it can be inferred that, together, these pinpoint more affluent LSOAs that are also accessible to offenders.

Variable short name	**Variable description**	**R square**	**Adjusted R square**
IDW_0.5	Percentage point difference between LSOA % Households social/ private renting and surrounding LSOAs IDW average (beta = 0.5, ATD = 3.6 km)	**0.393**	**0.391**
Regular_burg_rate	Regular burglary rate per 1,000 households, mid-point year 2010-2014		
IMD_Employ_Dep_Score	IMD 2015 Employment Deprivation Domain Score		
%_emp_16_74_wk_infor	% Employed persons aged between 16 and 74 who work in the information and communication or professional, scientific and technical activities industries		
%_hshlds_semi-det_house	% Household spaces semi-detached house or bungalow		

Table 5.7 - Initial OLS model, with associated R square and adjusted R square values

Variable short name	Variable description	R square	Adjusted R square
IDW_0.35	Percentage point difference between LSOA % Households social/ private renting and surrounding LSOAs IDW average (beta = 0.35, ATD = 5.0 km)	**0.396**	**0.394**
Regular_burg_rate	Regular burglary rate per 1,000 households, mid-point year 2010-2014		
IMD_Employ_Dep_Score	IMD 2015 Employment Deprivation Domain Score		
%_emp_16_74_wk_infor	% Employed persons aged between 16 and 74 who work in the information and communication or professional, scientific and technical activities industries		
%_hshlds_semi-det_house	% Household spaces semi-detached house or bungalow		

Table 5.8 - Final OLS model, with associated R square and adjusted R square values

Unfortunately, the final model violated some of the assumptions of OLS regression, for example, when the regression standardised residuals and regression standardised predicted values were plotted together, heteroscedasticity was evident in the form of a cone/ funnel-shaped pattern, i.e. the residuals (error terms) were not equally distributed (Frost, 2017). Further, when Kolmogorov-Smirnov and Shapiro-Wilk tests were performed on the unstandardised residuals, the H0 of normality was rejected. Recognising that spatial data can also violate another assumption of OLS, namely independence of the residuals (Charlton and Fotheringham, 2009, p.1), it was decided to trial an approach that has no prior assumptions, namely *k*-NN regression, which will now be discussed in Section 5.3.2.

5.3.2 *k*-NN Regression

Overview

Recall from the preceding section that some of the assumptions of OLS regression, including homoscedasticity of the residuals (Frost, 2017), were violated when the method was used to predict LSOA Car Key burglary rates. As a non-parametric approach, *k*-NN has no prior assumptions (Raschka, 2018, p.2) and was therefore deemed suitable for generating the static risk surface for the current research combined model.

When we think of nearest neighbours in a geographical context, it is usually in terms of Euclidean distance on the ground, for example, the point within a (global) data set that is most spatially proximate to a given target point, such as a LSOA centroid, or a crime location. The *k*-NN algorithm in machine learning considers nearness in terms of 'feature space'. To explain using Figure 5.2 below, and informed by Raschka (2018) and Singh (2018), assume that we have two classes; 'Class 1 = LSOA < median crime rate' and 'Class 2 = LSOA ≥ median crime rate', and we want to predict the class for a LSOA with an unknown crime rate (red shape on plot), then we can look at how similar ('near') the variable (feature) values are for this LSOA to those for the LSOAs with a known class. In this hypothetical example we will assume fifteen LSOAs and two independent variables of interest, e.g. '% Households 2+ cars/vans' and '% Persons white'. Looking at the graph in Figure 5.2, points representing the fifteen hypothetical LSOAs, including the one with an unknown class, have been plotted at the intersections of the variable values for each of these, for example, the X and Y values for LSOA 'C' are 5 & 9 and for LSOA 'D' 88 & 74. Based on Euclidean distance, in terms of independent variable values, and two nearest neighbours, we would thus classify the LSOA with an unknown crime rate as belonging to 'Class 2 = LSOA ≥ median crime rate', i.e. the same class as its two nearest neighbours, LSOA 'A' and LSOA 'B'. Note that when just one nearest neighbour is considered, the approach is simply termed the 'Nearest Neighbor Algorithm (NN)' (Raschka, 2018, p.1).

As per the OLS modelling approach presented in the previous section, the crime rates are already known for all LSOAs in the current study, and so the intention here is simply to identify the best-fitting model, as opposed to estimating rates for LSOAs with unknown crime rates. However, if the underlying independent variables data for the static risk surface was to change/ be updated in the future, e.g. new census data was released, then the existing *k*-NN regression model could be used to produce contemporaneous crime rate estimations.

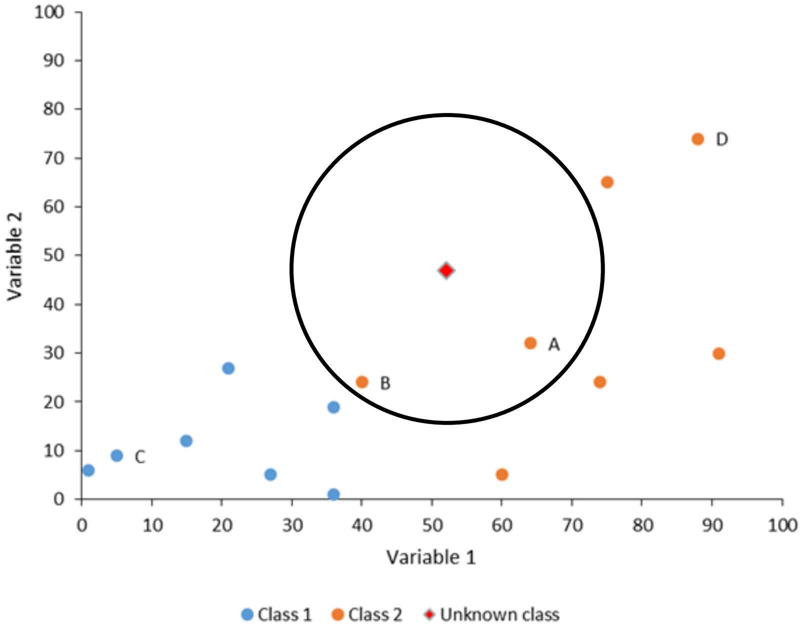


Figure 5.2 - Hypothetical example to illustrate how k-Nearest Neighbor (k-NN) classification works (adapted from: Raschka, 2018, p.2; Singh, 2018)

The main difference between *k*-Nearest Neighbour classification and k-Nearest Neighbour regression is that, for regression, the average of a continuous dependent variable for each *k* nearest neighbour is used as the predicted value for the target point with an unknown value (Raschka, 2018; Singh, 2018). So, to use the example in Figure 5.2 above again, assuming a crime rate of 70 offences per 1,000 households for LSOA 'A' and 54 for LSOA 'B', then a crime rate of 62 offences per 1,000 households ((70+54)/2 = 62) would be assigned to the target LSOA (red shape on plot).

This leads nicely onto three key considerations for *k*-NN classification and regression, namely:

1. How to prevent different variable scales from weighting the *k*-NN calculations (Raschka, 2018, p.15)
2. How to reduce/ limit the effects of the 'curse of dimensionality' (Raschka, 2018, pp.7-8)
3. How to determine the most appropriate *k* order (number of neighbours) (Singh, 2018)

Each of these considerations will now be discussed below.

Standardising Variable Scales

If the independent variables to be used in *k*-NN classification/ regression are on different scales, then those with smaller scales will be afforded more weight in the associated k-NN calculations (Raschka and Mirjalili, 2017, p.121). For example, a value of 5 on a scale of 0 to 5 is actually farther from 0, relatively speaking, than a value of 400 on a scale of 0 to 500, but the feature space distance between 0 and 400 is more. One solution to this problem is to min-max scale the independent variable values between 0 and 1 (Raschka and Mirjalili, 2017, p.121), thus ensuring that distance calculations capture any relative differences between these. Min-max scaling was explained in Chapter 3, but a reminder is provided in Figure 5.3 below. Although it is not feasible to show the 'not scaled' graph (A) at its actual feature space size, the min-max scaled graph (B) shows how scaling has affected each of the respective variable values. Note that min-max scaling of the independent variable values was chosen here for simplicity, however, depending on the distributions of these, i.e. presence/ absence of outliers, Z-scores might have been more appropriate (Raschka, 2014; Raschka and Mirjalili, 2017, p.122).

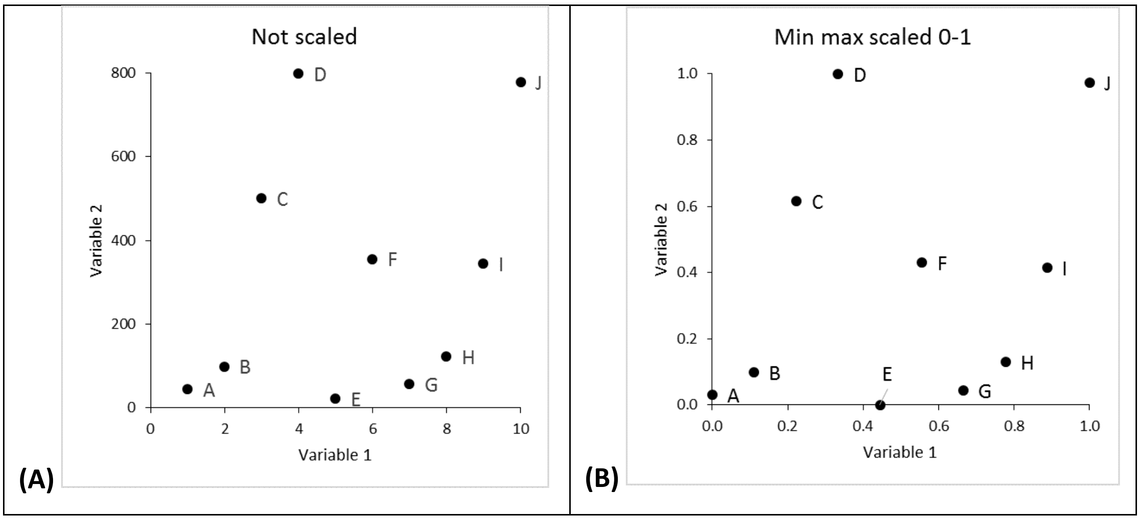


Figure 5.3 - Min-max scaling variable values between 0 and 1

Curse of Dimensionality

Having a large number of independent variables in a *k*-NN classification/ regression can render the process vulnerable to the so called 'curse of dimensionality'. As the number of independent variables, and thus dimensions, increases, so too does the volume of the feature space, i.e. the area that needs to be searched in order to locate the *k*-nearest neighbours for a target point. This means that as points become more distant from one another, they also become less similar to their nearest neighbours, which contradicts the underlying assumption of *k*-NN classification (Raschka, 2018, pp.7-8). Further, as the volume of the feature space increases, there is an increased risk of over-fitting, which can make it difficult to generalise a model to unseen data (Raschka and Mirjalili, 2017, p.123). With this in mind, only the five variables from the final OLS regression model will be included in the *k*-NN regression.

Choosing an Appropriate Value of k

The value of *k* is perhaps the most important decision to be made in *k*-NN classification/ regression given that it ultimately determines the model's performance. It is usual to identify the most appropriate *k* by performing the *k*-NN algorithm many times on a training data set, as was the case in the current work – see Figure 5.4 below. Key references that were used to inform the development of the R coding algorithm in the current work are: Saxena (2017), Kassambara (2018), Singh (2018), and Kuhn (2019). An 80% subset of the Car Key burglary rates data set was selected and repeated cross-fold validation (to be explained shortly) was then performed on this to ascertain which value of *k* generated the lowest Root Mean Square Error (RMSE) (Singh, 2018). The best performing model was then used to make predictions on a test data set, i.e. the 20% of the rates data set that was not included at the training stage of the process. Because there is no more recent census data available than was used to train the *k*-NN regression model, the same input data had to be used to generate the LSOA predictions, which was also the case with the OLS – hopefully the cross-fold validation approach will have gone some way to mitigating any potential issues inherent to this.

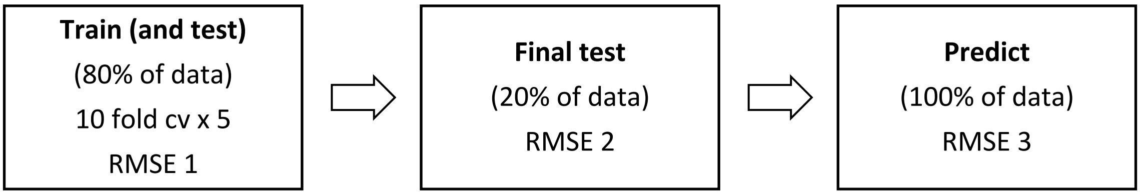


Figure 5.4 - k-NN training and testing current research approach

The following explains in more detail the method that was employed in the current work. *k*-NN is a lazy learner in that it simply memorises all of the points in a training data set and then identifies which of these are nearest to a given target point for which a value is not known (Raschka, 2018, pp.1-2), with *k* having been ascertained at the model training stage. Recall that in *k*-NN regression, the predicted value for a target point is the average of the dependent variable values for the *k* neighbouring points. Nearest neighbour algorithms are often trained by splitting a data set into two according to some ratio, e.g. 65:35, and then using the larger of the data splits as training data and the smaller as testing data. However, in the absence of any additional validation, there is a risk that the data that is held out for testing will not be representative of general trends within the entire data set (Scikit-learn, c2007-2019), which is where repeated cross-fold validation comes in.

With cross-fold validation, a dataset is split into a user-specified number of approximately equal parts, known as folds, e.g. 10 fold = 10 subsets, and each time a training algorithm is run, one of these folds is held out for testing (Raschka and Mirjalili, 2017, p.192; Scikit-learn. c2007-2019). The difference between cross-fold validation and repeated cross-fold validation is that, for the latter, once all folds have been used as a test set, the entire process is repeated again *n* times (Kuhn, 2019). Figure 5.5 below is included to visualise the process. Assuming 5 folds (1 to 5) and 3 repeats, a training algorithm would be run five times (A-E), with a different fold acting as the test set. So, for run 'A', the first fold would be used as the test set and folds 2-5 would be used as the train set, whereas for run 'B', the second fold would be used as the test set and folds 1 & 3-5 as the train set. Once the initial five runs had been completed, the process would then be run its entirety for another two times (5 folds x 3). The 'Caret' package in R, which stands for

"**C**lassification **A**nd **RE**gression **T**raining" (Kuhn, 2019), was used to perform the repeat cross-validation in the current research, with 10 folds and 5 repeats for every value of *k* (1:50) tested.

A	B	C	D	E
1				
2	2			
3		3		
4			4	
5				5

Figure 5.5 - Cross-fold validation example (adapted from: Raschka and Mirjalili, 2017, p.192; Scikit-learn, c2007-2019)

Input Variables

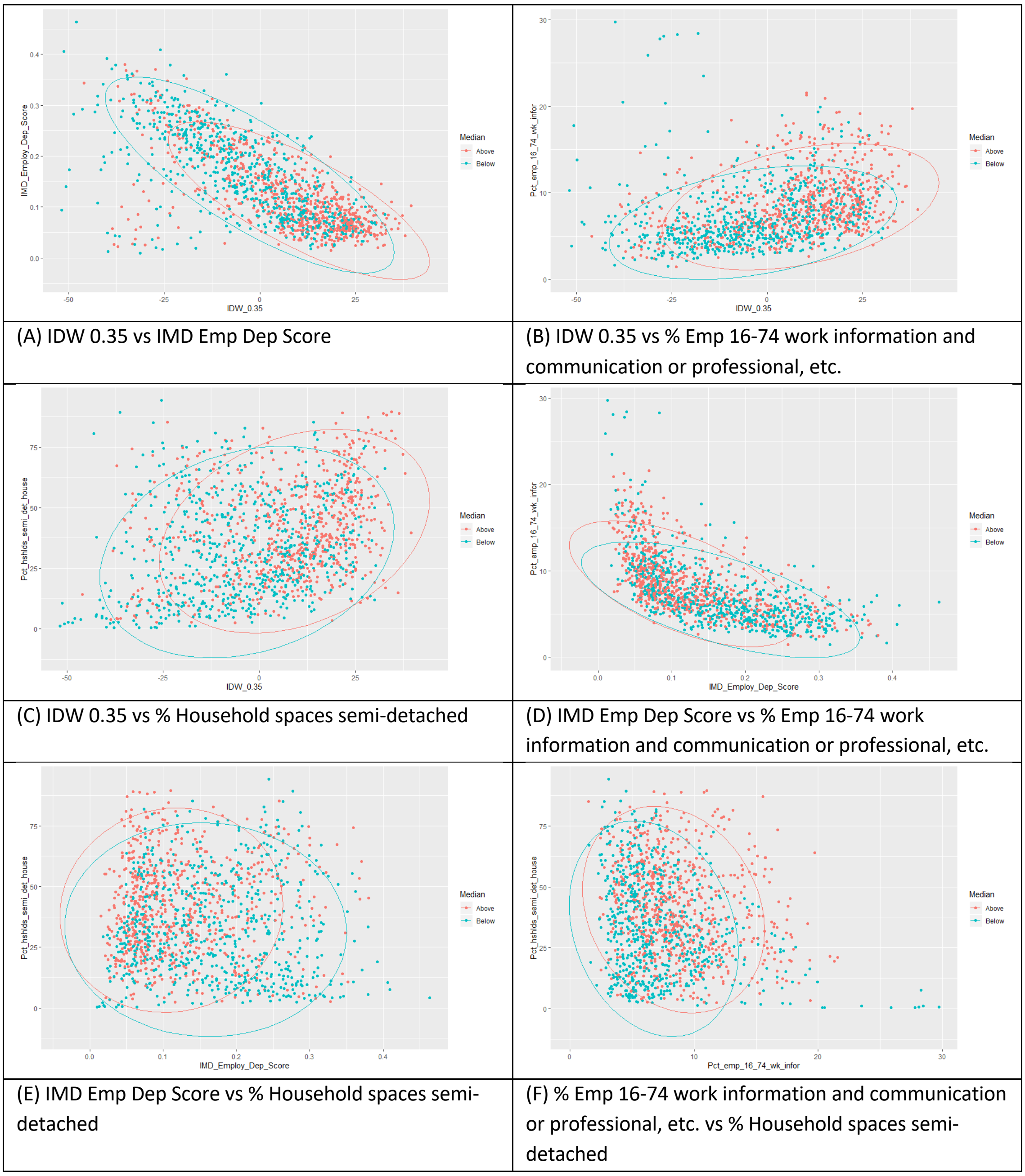
As previously mentioned, it was decided to use the same five input variables for the *k*-NN regression model as were used in the final OLS model, namely:

1. Percentage point difference between LSOA % Households social/ private renting and surrounding LSOAs IDW average (beta = 0.35, ATD = 5.0 km)
2. Regular burglary rate per 1,000 households, mid-point year 2010-2014
3. IMD 2015 Employment Deprivation Domain Score
4. % Employed persons aged between 16 and 74 who work in the information and communication or professional, scientific and technical activities industries
5. % Households spaces semi-detached house or bungalow

A selection of these input variables have been plotted, together with 95% confidence ellipses, in Figure 5.6 below, the associated points having been classified according to a LSOA Car Key burglary rate per 1,000 households either '**above/ equal to'** or '**below'** the study area median (mid-point year 2010-2014). The plots are intended to provide a visual indication as to the range of feature values that might typically present with different Car Key burglary rates, and also which of these is likely to be most effective at differentiating between classes/ predicting rates. To highlight a couple of examples in Figure 5.6, looking first at plot (A), there is evidence of some dissimilarity between the two feature variables and the associated Car Key burglary rate classes. For the 'Above' median class (red points), there is notable clustering towards the bottom right of the graph, which represents a higher proportion of households renting in surrounding LSOAs BUT low within-area employment deprivation. Attempting to situate this observation in theory, the IDW % households renting variable potentially points to a proximate offender supply (recall Bottoms and Wiles, 1986, p.105), whereas the low employment deprivation score suggests that the target property type (desirable vehicles) might be more prevalent in the related LSOAs.

Therefore, the identified clustering might reflect spatial juxtapositions of offenders and victims. It seems that similar factors are at play in plot (B), with apparent clustering of 'Above' median Car Key burglary rate points (red) in the section of the graph that represents LSOAs with a higher proportion of employed persons aged between 16 and 74 who work in the information and communication or professional, scientific and technical activities industries AND a higher proportion of households renting in surrounding LSOAs.

Note that the input data in the plots has not been scaled – this was subsequently undertaken during the modelling process (min-max, 0-1). All input data were also rounded to three decimal places for the *k*-NN modelling.



(A) IDW 0.35 vs IMD Emp Dep Score

(B) IDW 0.35 vs % Emp 16-74 work information and communication or professional, etc.

(C) IDW 0.35 vs % Household spaces semi-detached

(D) IMD Emp Dep Score vs % Emp 16-74 work information and communication or professional, etc.

(E) IMD Emp Dep Score vs % Household spaces semi-detached

(F) % Emp 16-74 work information and communication or professional, etc. vs % Household spaces semi-detached

Figure 5.6 - Plots of select input variables (features) classified according to below (blue) or above/ equal to (red) median LSOA Car Key burglary rate (coding based on by LearnByExample.org, c2019)

Generating the Predictions

Just to reiterate, the model training was performed using repeated cross-fold validation with 10 folds and 5 repeats for *k* = 1:50. Although the Caret package in R automatically outputs the optimum value of *k* at the end of the training algorithm, it is also useful to look at the associated elbow plot of RMSE values for all *k* iterations (Singh, 2018). Caret identified the optimum value of *k* as being 14, however, this information could also have been ascertained by reading off the corresponding number of neighbours for the point on Figure 5.7 below with the lowest RMSE value. Looking at the graph in more detail, the model performed poorly when only one nearest neighbour was considered, probably due to over-fitting, but improved markedly as *k* increased, up to 14, after which the performance gradually worsened again, although not to the same level as for *k* = 1. Note that each point in the study area had a potential maximum of 1,387 *k* nearest neighbours (1,388 LSOAs minus 1), so testing up to 50 of these seemed reasonable.

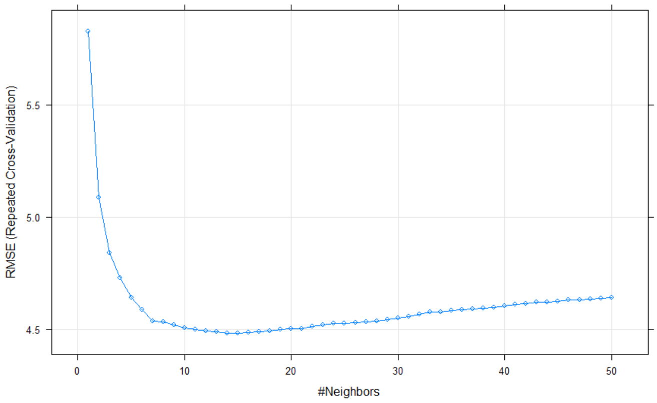


Figure 5.7 - Elbow graph for k-NN training model

Figure 5.8 below shows the variable importance plot for the training model (Brownlee, 2014), again produced using Caret. Since the '% Household spaces semi-detached' variable was not important, if the model was to be refined in the future, this could be removed to reduce the feature space extent.

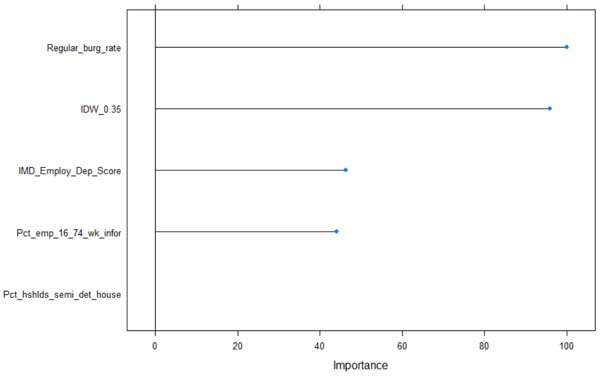


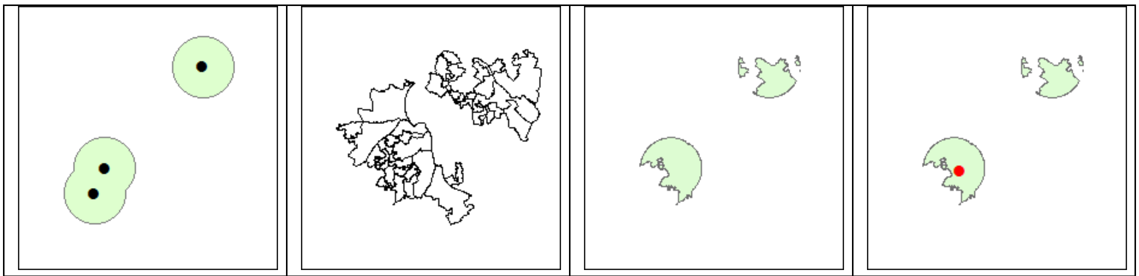
Figure 5.8 - Variable importance plot for training model

Table 5.9 below contains the RMSE results for *k = 14* for the train & test data, the final test data, and the predicted Car Key burglary rates. The RMSE for the best performing training model was 4.48 (2 DP), decreasing to 4.28 for the final test data (unseen data set), and 4.21 for the predicted crime rates. It is not ideal that the rate predictions were generated from the same input data set that was used to train the model but, as mentioned earlier, this was unavoidable. However, the fact that the training model actually performed slightly better on the unseen 20% of the data is reassuring. The R square value for the *k*-NN training model was slightly lower than for the OLS regression model (*c.f.* 0.396), explaining circa 35 per cent of the variation in Car Key burglary rates but, importantly, with no violated assumptions.

RMSE for train & test data (80% of data set)	**RMSE for final test data (20% of data set)**	**RMSE for predictions (100% of data set)**
Model error RMSE: K = 14 RMSE = 4.482266 **R^2** = 0.3517307 **0.352 (3 DP)**	Prediction error RMSE: RMSE = 4.275746	Prediction error RMSE: RMSE = 4.207409

Table 5.9 - k-NN regression RMSE results for optimum value of k (14)

Once the LSOA Car Key burglary rate predictions had been generated from the *k*-NN regression, the median of these was calculated. Any LSOAs with a Car Key burglary rate above/ equal to the median were included in the static risk surface for the current research combined model and all others were disregarded. The static risk surface will be used to clip the dynamic risk surfaces that are generated by the dynamic model, thus reducing the spatial extent of these, as illustrated in Figure 5.9 below – this will be covered in more detail in Chapter 6.



(A) Buffer past offences to create dynamic risk surface **(BOOST ACCOUNT)**

(B) Static risk surface (created from *k*-NN regression results)

(C) Clip dynamic risk surface using static risk surface **(FLAG ACCOUNT)**

(D) Plot ‘future’ offences over combined risk surface

Figure 5.9 - How the static risk surface will be used to clip the dynamic risk surfaces (hypothetical crime points)

5.4 Spatio-temporal Clustering

This section will now report on the spatio-temporal clustering analysis using Ratcliffe's Near Repeat Calculator Version 1.3 (May 2008?) to identify parameters for the dynamic risk surfaces. The Near Repeat Calculator (NRC) implements a version of the Knox Test, which was developed by Knox (1964) for the purpose of analysing space-time clusters of childhood Leukaemia cases. The Test works by pairing individual cases of a subject, e.g. crime events, within a contingency table based on spatio-temporal proximity of these (Johnson and Bowers, 2004) – Table 5.10 below is a hypothetical example of a Knox contingency table. The actual number of cases within each spatio-temporal bandwidth is then compared to a computed random distribution of the total cases to understand if there are more in a given cell than would be expected on the basis of chance, i.e. significant clustering (Ratcliffe, ca. 2009; Youstin et al., 2011).

	Weeks between events			
Metres between events	*1*	*2*	*3*	*4*
0-100	n=5	n=3	n=4	n=6
100-200	n=2	n=1	n=3	n=1
200-300	n=4	n=2	n=0	n=0

Table 5.10 - Hypothetical example of a Knox contingency table (adapted from Johnson and Bowers, 2004, p.247)

As explained in Ratcliffe (ca. 2009), the NRC program takes a .csv file of X, Y coordinates and date stamps for individual offences, e.g. "`427727, 434870, 29/06/2017`", with each row representing an individual crime record. Users can then specify spatial and temporal bandwidths for the analysis, that is, the distance and time increments at which offences will be considered paired; these parameters are typically informed by the related theory (Ratcliffe, ca. 2009, p.6), e.g. the expected areal extent of offenders' criminal activities, and also with consideration to the intended practical application of any findings, e.g. the length of time that police can realistically concentrate resources in one area (Youstin et al., 2011, pp.1051-1052). Referring again to Ratcliffe (ca. 2009), the user's choice of significance level determines how many Monte Carlo iterations are performed by the NRC, that is, how many times dates are reassigned to X, Y locations to generate expected patterns, as well as the best p value than can be achieved during the analysis. A rigorous significance level of p = 0.001 (999 iterations) was chosen for the current work because the RV/ N-RV findings are expected to impact on the performance of the two research models. The distance setting was left as 'Manhattan' because this is thought to better represent distances for unknown routes in urban environments (Ratcliffe, ca. 2009, p.9). Once the NRC has performed the required number of iterations, it outputs an 'observed over mean expected frequencies ratio' for each spatio-temporal band tested. The ratios are calculated by dividing the observed number of offence pairs for a given spatio-temporal band, e.g. '0 to 1 days' and '1 to 250 metres', by the average expected number of offence pairs for the same (Ratcliffe, ca. 2009, p.10). Any bands with a significance value ≤ 0.05 and an odds ratio ≥ 1.20 are considered to experience over-representation of crime events (Ratcliffe, ca. 2009, pp.9-10). A ratio of 1.20 for a cell indicates that the associated over-representation is 20% greater than would be expected on the basis of chance (Ratcliffe, ca. 2009, p.10). Table 5.11 below is a simplified version of the NRC 'observed over mean expected frequencies table' output – ratios ≥ 1.20 are shown in bold. To give an example, the odds ratio of 1.89 for the 501 to 750 metres distance band and 3-3 days temporal band (yellow shading) indicates that, given one burglary at a location, the chance of another burglary occurring within 501 to 750 m and 3-3 days of this is 89% greater than if offenders' behaviour was random (Ratcliffe, ca. 2009, p.10).

	0 to 1 days	*2 to 2 days*	*3 to 3 days*	*4 to 4 days*	*5 to 5 days*
Same location	**7.89**	**4.52**	**2.04**	**1.66**	0.38
1 to 250 metres	**5.51**	**1.81**	**1.52**	1.17	**1.41**
251 to 500 metres	**2.78**	**1.51**	**1.51**	**1.58**	**1.39**
501 to 750 metres	**1.73**	**1.23**	**1.89**	**1.35**	**1.40**
751 to 1000 metres	1.19	**1.34**	**1.32**	**1.56**	1.17
1001 to 1250 metres	**1.57**	1.18	1.06	1.07	**1.40**

Table 5.11 - Simplified (hypothetical) NRC output - 'observed over mean expected frequencies table' (all bold taken to be significant at ≤ 0.05, adapted from: Ratcliffe, ca. 2009, p.10)

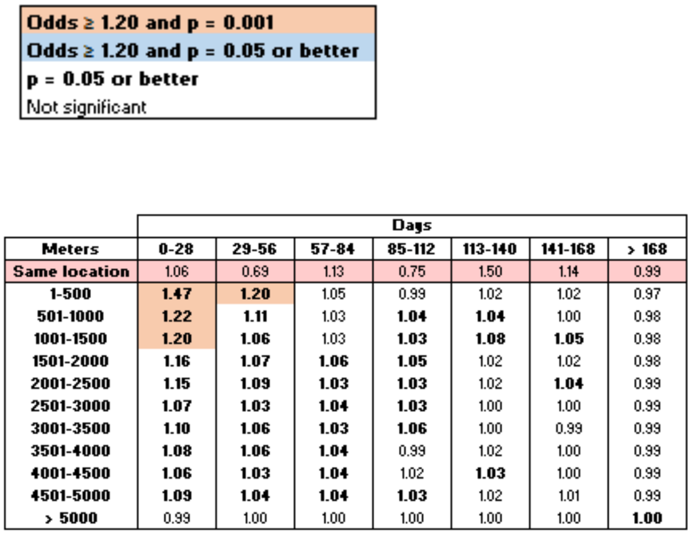
Recall from Chapter 3 that the initial crime sample sizes for the RV/ N-RV analysis were 6,917 x Car Key burglaries and 68,460 x Regular burglaries. The first part of the analysis that follows was conducted using the entire Car Key burglary sample, that is, offences with a mid-point committed year between 2010 and 2014 (five years in total). Both crime types were then analysed by individual mid-point committed year to ascertain if there were any notable differences between the associated spatio-temporal patterns for each of these, and also whether any observed trends were stable over time. However, because there is a limit to the volume of input data that the NRC can process during a single run, Regular burglary, being far more abundant than Car Key burglary, was only analysed for three of the five mid-point years, namely; 2012, 2013, and, 2014.

It is worth noting here that the NRC output, i.e. observed overrepresentation, seems to be extremely sensitive to the user's choice of (arbitrary) spatio-temporal bandwidths, and also that the '0 to 1 days' output actually appears to refer to offences committed on the same date. The latter could potentially lead to the misinterpretation of results because '1 day' implies return visits, i.e. genuine repeats/ near-repeats, as opposed to a temporally clustered crime spree. Having run a dummy data set through the NRC using the same X, Y coordinates but various dates, including some duplicates, this identified that the '2 to 2 days' output refers to crimes committed on adjacent dates, the '3 to 3 days' output to offences committed one day apart, and so on. Although offences in the '2 to 2 days' time band might be a more reliable indicator of genuine repeat/ near-repeat activity than those in the '0 to 1 days' band, there is still a risk that, for those crime types that typically occur overnight, linked offences might be recorded under adjacent dates, i.e. if committed either side of midnight. In light of this, perhaps significant repeat/ near-repeat over-representation at the '3 to 3 days' band, and beyond, is a more reliable indicator of optimal foraging behaviour than over-representation at shorter time bands. Adding further weight to this argument, Youstin et al. (2011, p.1043) suggest that if offences are not disaggregated into suitably short time periods, i.e. they are only analysed by week/ month, then any identified over-representation at longer time bands might actually be being driven by over-representation at much shorter time bands. For example, an offender might commit a number of burglary offences at a student house on the same date, or a group of Car Key burglars might steal a number of cars over a single 24 hour period, both of which could potentially contribute to over-representation at longer time bands.

5.4.1 Aggregate Car Key Burglary Results

Numerous NRC experiments were run using the full Car Key burglary data set, that is, 6,917 offences with a mid-point committed year between 2010 and 2014. As mentioned previously, different patterns emerged depending on how the spatio-temporal bands were conceptualised. Results from a selection of the NRC experiments are shown in Table 5.12 to Table 5.17 below. The pink shaded row in each table represents the same location, i.e. genuine repeats (same X,Y coordinates), and any cells with a significance level ≤ 0.05 are shown in bold text. Blue shaded cells represent an odds ratio ≥ 1.20 and a p value ≤ 0.05 and orange shaded cells represent an odds ratio ≥ 1.20 and a p value of 0.001. Due to the many ways in which the input data could have been split, both spatially and temporally, it was decided to begin the analysis with broad time and distance ranges in order to identify high-level patterns. Being mindful of the '0 to 1 days' issue discussed earlier, any significant RV/ N-RV over-representation was then disaggregated further, including by individual days. So, in Table 5.12, offences were paired if they occurred within 5 km and six months of each other, split into 28 day and 500 metre intervals. Six months was chosen as the maximum temporal parameter for the analysis in case mobile offenders wait a length of time before returning to previously victimised locations. Possible reasons for this include the time taken to replace the stolen property type (desirable vehicles) and also the risk of apprehension that is inherent to return visits to unfamiliar areas. For example, affluent communities might be better equipped than those in more deprived areas to act quickly following an initial event, such as by increasing neighbourhood-level surveillance (Sampson et al., 1997), and installing building security measures (Johnson et al., 1997, p.238). 5,000 metres was chosen as the maximum distance parameter with a view to ensuring that the majority of linked offences were captured, i.e. Car Key burglars are assumed to operate across much larger areas than Regular burglars and so their offences might be situated farther apart – recall that Carden (2012, p.74) identified a median JTC distance of 4.9 km for Car Key burglars. It is also worth mentioning that a maximum distance parameter of 5,000 m is longer than has been examined in similar studies of residential burglary RV/ N-RV, e.g. Adepeju (2017) and Johnson et al. (2007) reported results up to 2,000 m, and Youstin et al. (2011) up to ~ 1,500 m.

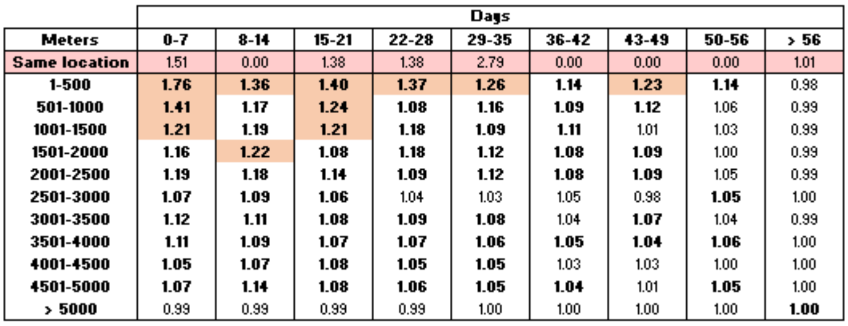
The initial NRC analysis, presented in Table 5.12 below, identified significant spatio-temporal clustering of Car Key burglaries up to 56 days and 1,500 metres for odds ≥ 1.20 and p = 0.001. It was decided, therefore, to disaggregate the two associated 4 week periods for the 56 days maximum into individual one week periods, but to leave the distance bands the same at 500 m increments. As can be seen from the results in Table 5.13 below, the odds ratios for the 36-42 and 50-56 days temporal bands are less than 1.20, which is not evident from Table 5.12.



Odds ≥ 1.20 and p = 0.001
Odds ≥ 1.20 and p = 0.05 or better
p = 0.05 or better
Not significant

	Days						
Meters	**0-28**	**29-56**	**57-84**	**85-112**	**113-140**	**141-168**	**> 168**
Same location	1.06	0.69	1.13	0.75	1.50	1.14	0.99
1-500	**1.47**	**1.20**	1.05	0.99	1.02	1.02	0.97
501-1000	**1.22**	**1.11**	1.03	**1.04**	**1.04**	1.00	0.98
1001-1500	**1.20**	**1.06**	1.03	**1.03**	**1.08**	**1.05**	0.98
1501-2000	**1.16**	**1.07**	**1.06**	**1.05**	1.02	1.02	0.98
2001-2500	**1.15**	**1.09**	**1.03**	**1.03**	1.02	**1.04**	0.99
2501-3000	**1.07**	**1.03**	**1.04**	**1.03**	1.00	1.00	0.99
3001-3500	**1.10**	**1.06**	**1.03**	**1.06**	1.00	0.99	0.99
3501-4000	**1.08**	**1.06**	**1.04**	0.99	1.02	1.00	0.99
4001-4500	**1.06**	**1.03**	**1.04**	1.02	**1.03**	1.00	0.99
4501-5000	**1.09**	**1.04**	**1.04**	**1.03**	1.02	1.01	0.99
> 5000	0.99	1.00	1.00	1.00	1.00	1.00	**1.00**

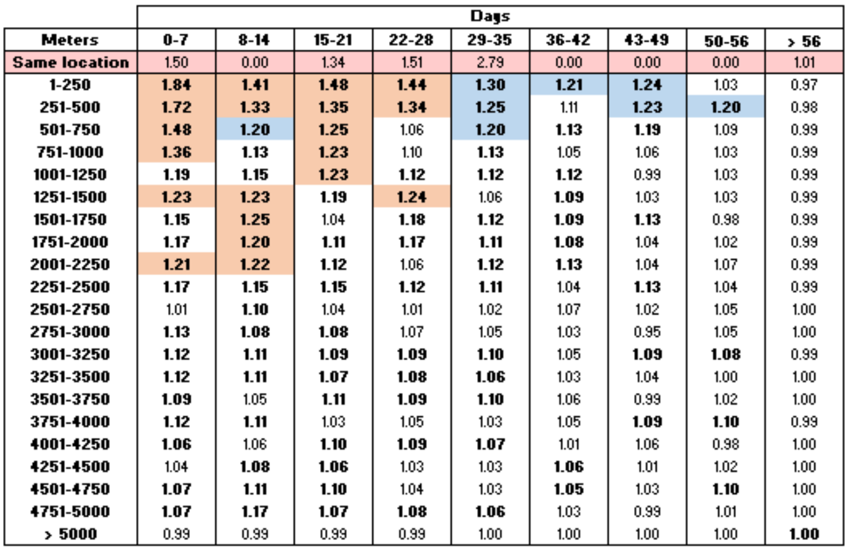
Table 5.12 - 500 m x 10 and 28 days x 6 (999 iterations) = 5 km and ~ 6 months



	Days								
Meters	**0-7**	**8-14**	**15-21**	**22-28**	**29-35**	**36-42**	**43-49**	**50-56**	**> 56**
Same location	1.51	0.00	1.38	1.38	2.79	0.00	0.00	0.00	1.01
1-500	**1.76**	**1.36**	**1.40**	**1.37**	**1.26**	**1.14**	**1.23**	**1.14**	0.98
501-1000	**1.41**	**1.17**	**1.24**	**1.08**	**1.16**	**1.09**	**1.12**	1.06	0.99
1001-1500	**1.21**	**1.19**	**1.21**	**1.18**	**1.09**	**1.11**	1.01	1.03	0.99
1501-2000	**1.16**	**1.22**	**1.08**	**1.18**	**1.12**	**1.08**	**1.09**	1.00	0.99
2001-2500	**1.19**	**1.18**	**1.14**	**1.09**	**1.12**	**1.08**	**1.09**	1.05	0.99
2501-3000	**1.07**	**1.09**	**1.06**	1.04	1.03	1.05	0.98	**1.05**	1.00
3001-3500	**1.12**	**1.11**	**1.08**	**1.09**	**1.08**	1.04	**1.07**	1.04	0.99
3501-4000	**1.11**	**1.09**	**1.07**	**1.07**	**1.06**	**1.05**	**1.04**	**1.06**	1.00
4001-4500	**1.05**	**1.07**	**1.08**	**1.05**	**1.05**	1.03	1.03	1.00	1.00
4501-5000	**1.07**	**1.14**	**1.08**	**1.06**	**1.05**	**1.04**	1.01	**1.05**	1.00
> 5000	0.99	0.99	0.99	0.99	1.00	1.00	1.00	1.00	**1.00**

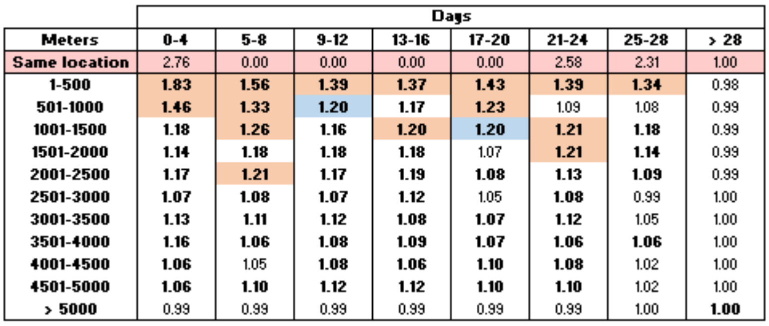
Table 5.13 - 500 m x 10 and 7 days x 8 (999 iterations) = 5 km and 8 weeks

The next table (Table 5.14 below) is based on the same time bands as Table 5.13 above, i.e. 7 days x 8, but the distance bands have been disaggregated into 250 m intervals. An interesting observation in relation to this table is that, for all except one of the cells with an odds ratio ≥ 1.20 in the 0-28 days columns, the associated p value is 0.001, whereas for the cells with an odds ratio ≥ 1.20 in the other four columns it is 0.05. In light of this, and also noting that 28 days is probably more practicable than 56 days in terms of maintaining police/ partner agency presence in specific areas, the data is split into 4 day periods x 7 (total of 28 days) in Table 5.15 below, and then disaggregated by individual days (x 28) in Table 5.16 below, the distance bandwidth having been increased to 500 m for both. It is useful to note here that some of the days that are aggregated within the odds ≥ 1.20 and p = 0.001 cells in Table 5.15 are less significant/ not significant in Table 5.16, thus providing empirical support for Youstin et al's (2011) argument regards the importance of disaggregating crime data in RV/ N-RV analysis.



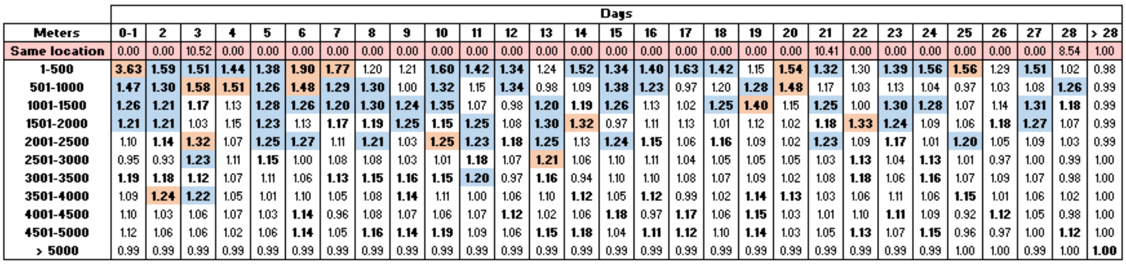
	Days								
Meters	**0-7**	**8-14**	**15-21**	**22-28**	**29-35**	**36-42**	**43-49**	**50-56**	**> 56**
Same location	1.50	0.00	1.34	1.51	2.79	0.00	0.00	0.00	1.01
1-250	**1.84**	**1.41**	**1.48**	**1.44**	**1.30**	**1.21**	**1.24**	1.03	0.97
251-500	**1.72**	**1.33**	**1.35**	**1.34**	**1.25**	1.11	**1.23**	**1.20**	0.98
501-750	**1.48**	**1.20**	**1.25**	1.06	**1.20**	**1.13**	**1.19**	1.09	0.99
751-1000	**1.36**	**1.13**	**1.23**	1.10	**1.13**	1.05	1.06	1.03	0.99
1001-1250	**1.19**	**1.15**	**1.23**	**1.12**	**1.12**	**1.12**	0.99	1.03	0.99
1251-1500	**1.23**	**1.23**	**1.19**	**1.24**	1.06	**1.09**	1.03	1.03	0.99
1501-1750	**1.15**	**1.25**	1.04	**1.18**	**1.12**	**1.09**	**1.13**	0.98	0.99
1751-2000	**1.17**	**1.20**	**1.11**	**1.17**	**1.11**	**1.08**	1.04	1.02	0.99
2001-2250	**1.21**	**1.22**	**1.12**	1.06	**1.12**	**1.13**	1.04	1.07	0.99
2251-2500	**1.17**	**1.15**	**1.15**	**1.12**	**1.11**	1.04	**1.13**	1.04	0.99
2501-2750	1.01	**1.10**	1.04	1.01	1.02	1.07	1.02	1.05	1.00
2751-3000	**1.13**	**1.08**	**1.08**	1.07	1.05	1.03	0.95	1.05	1.00
3001-3250	**1.12**	**1.11**	**1.09**	**1.09**	**1.10**	1.05	**1.09**	**1.08**	0.99
3251-3500	**1.12**	**1.11**	**1.07**	**1.08**	**1.06**	1.03	1.04	1.00	1.00
3501-3750	**1.09**	1.05	**1.11**	**1.09**	**1.10**	1.06	0.99	1.02	1.00
3751-4000	**1.12**	**1.11**	1.03	1.05	1.03	1.05	**1.09**	**1.10**	0.99
4001-4250	**1.06**	1.06	**1.10**	**1.09**	**1.07**	1.01	1.06	0.98	1.00
4251-4500	1.04	**1.08**	**1.06**	1.03	1.03	**1.06**	1.01	1.02	1.00
4501-4750	**1.07**	**1.11**	**1.10**	1.04	1.03	**1.05**	1.03	**1.10**	1.00
4751-5000	**1.07**	**1.17**	**1.07**	**1.08**	**1.06**	1.03	0.99	1.01	1.00
> 5000	0.99	0.99	0.99	0.99	1.00	1.00	1.00	1.00	**1.00**

Table 5.14 - 250 m x 20 and 7 days x 8 (999 iterations) = 5 km and 8 weeks



	Days							
Meters	**0-4**	**5-8**	**9-12**	**13-16**	**17-20**	**21-24**	**25-28**	**> 28**
Same location	2.76	0.00	0.00	0.00	0.00	2.58	2.31	1.00
1-500	**1.83**	**1.56**	**1.39**	**1.37**	**1.43**	**1.39**	**1.34**	0.98
501-1000	**1.46**	**1.33**	**1.20**	**1.17**	**1.23**	1.09	1.08	0.99
1001-1500	**1.18**	**1.26**	**1.16**	**1.20**	**1.20**	**1.21**	**1.18**	0.99
1501-2000	**1.14**	**1.18**	**1.18**	**1.18**	1.07	**1.21**	**1.14**	0.99
2001-2500	**1.17**	**1.21**	**1.17**	**1.19**	**1.08**	**1.13**	**1.09**	0.99
2501-3000	**1.07**	**1.08**	**1.07**	**1.12**	1.05	**1.08**	0.99	1.00
3001-3500	**1.13**	**1.11**	**1.12**	**1.08**	**1.07**	**1.12**	1.05	1.00
3501-4000	**1.16**	**1.06**	**1.08**	**1.09**	**1.07**	**1.06**	**1.06**	1.00
4001-4500	**1.06**	1.05	**1.08**	**1.06**	**1.10**	**1.08**	1.02	1.00
4501-5000	**1.06**	**1.10**	**1.12**	**1.12**	**1.10**	**1.10**	1.02	1.00
> 5000	0.99	0.99	0.99	0.99	0.99	0.99	1.00	**1.00**

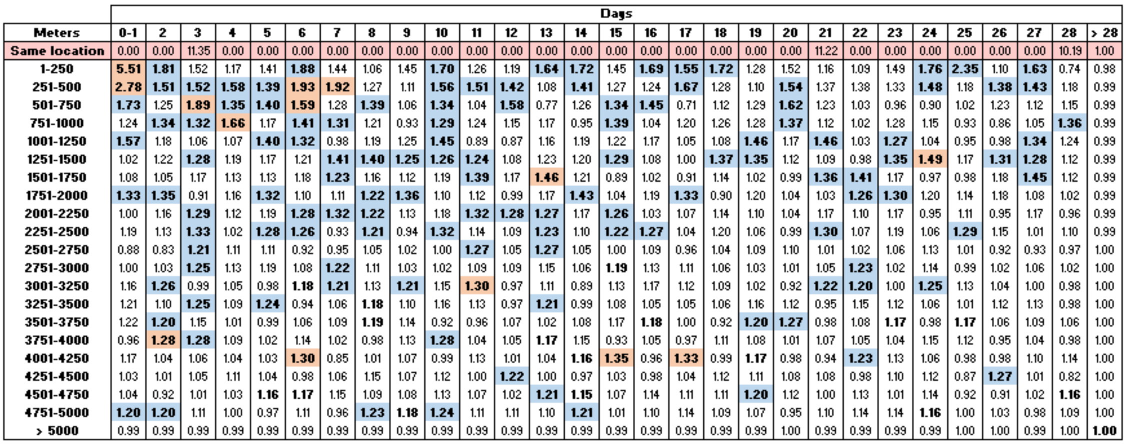
Table 5.15 - 500 m x 10 and 4 days x 7 (999 iterations) = 5 km and 28 days



	Days																												
Meters	**0-1**	**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**	**> 28**
Same location	0.00	0.00	10.52	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	10.41	0.00	0.00	0.00	0.00	0.00	0.00	8.54	1.00
1-500	**3.63**	**1.59**	**1.51**	**1.44**	**1.38**	**1.90**	**1.77**	1.20	1.21	**1.60**	**1.42**	**1.34**	1.24	**1.52**	**1.34**	**1.40**	**1.63**	**1.42**	1.15	**1.54**	**1.32**	1.30	**1.39**	**1.56**	**1.56**	1.29	**1.51**	1.02	0.98
501-1000	**1.47**	**1.30**	**1.58**	**1.51**	**1.26**	**1.48**	**1.29**	**1.30**	1.00	**1.32**	1.15	**1.34**	0.98	1.09	**1.38**	**1.23**	0.97	1.20	**1.28**	**1.48**	1.17	1.03	1.13	1.04	0.97	1.03	1.08	**1.26**	0.99
1001-1500	**1.26**	**1.21**	**1.17**	1.13	**1.28**	**1.26**	**1.20**	**1.30**	**1.24**	**1.35**	1.07	0.98	**1.20**	**1.19**	**1.26**	1.13	1.02	**1.25**	**1.40**	1.15	**1.25**	1.00	**1.30**	**1.28**	1.07	1.14	**1.31**	**1.18**	0.99
1501-2000	**1.21**	**1.21**	1.03	1.15	**1.23**	1.13	**1.17**	**1.19**	**1.25**	**1.15**	**1.25**	1.08	**1.30**	**1.32**	0.97	1.11	1.13	1.01	1.12	1.02	**1.18**	**1.33**	**1.24**	1.09	1.06	**1.18**	**1.27**	1.07	0.99
2001-2500	1.10	**1.14**	**1.32**	1.07	**1.25**	**1.27**	1.11	**1.21**	1.03	**1.25**	**1.23**	**1.18**	**1.25**	1.13	**1.24**	**1.15**	1.06	**1.16**	1.09	1.02	**1.23**	1.09	**1.17**	1.01	**1.20**	1.05	1.09	1.03	0.99
2501-3000	0.95	0.93	**1.23**	1.11	**1.15**	1.00	1.08	1.08	1.03	1.01	**1.18**	1.07	**1.21**	1.06	1.10	1.11	1.04	1.05	1.05	1.05	1.03	**1.13**	1.04	**1.13**	1.01	0.97	1.00	0.99	1.00
3001-3500	**1.19**	**1.18**	**1.12**	1.07	1.11	1.06	**1.13**	**1.15**	**1.16**	**1.15**	**1.20**	0.97	**1.16**	0.94	1.10	1.10	1.08	1.07	1.09	1.02	1.08	**1.18**	1.06	**1.16**	1.07	1.09	1.07	0.98	1.00
3501-4000	1.09	**1.24**	**1.22**	1.05	1.01	1.10	1.05	1.08	**1.14**	1.11	1.00	1.06	1.10	**1.12**	1.05	**1.12**	0.99	1.02	**1.14**	**1.13**	1.03	1.06	1.11	1.06	**1.15**	1.01	1.06	1.02	1.00
4001-4500	1.10	1.03	1.06	1.07	1.03	**1.14**	0.96	1.08	1.07	1.06	1.07	**1.12**	1.02	1.06	**1.18**	0.97	**1.17**	1.06	**1.15**	1.03	1.01	1.10	**1.11**	1.09	0.92	**1.12**	1.05	0.98	1.00
4501-5000	1.12	1.06	1.06	1.02	1.06	**1.14**	1.05	**1.16**	**1.14**	**1.19**	1.09	1.06	**1.15**	**1.18**	1.04	**1.11**	**1.12**	1.10	**1.14**	1.03	1.05	**1.13**	1.07	**1.15**	0.96	0.97	1.00	**1.12**	1.00
> 5000	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	1.00	1.00	0.99	1.00	**1.00**

Table 5.16 - 500 m x 10 and 1 days x 28 (999 iterations) = 5 km and 28 days

Looking finally at Table 5.17 below, the time bands are the same as for Table 5.16 above, but the distance bandwidth has been decreased to 250 m. Notably, the ‘0 to 1 days’ time band is significant at p = 0.001 up to 500 m, even with the shorter bandwidth. Assuming that the author is correct regards the ‘0 to 1 days’ issue, this finding indicates that Car Key burglars are likely to commit temporally clustered offences within relatively close proximity of each other (up to 500 metres). Focusing now on the other cells with an odds ratio ≥ 1.20 and a p value of 0.001 in Table 5.17, these are perhaps more dispersed than we would expect to see if offenders were displaying conventional foraging behaviours, i.e. such as those that are typically seen with Regular burglary. Further, and in a similar vein, no statistically significant over-representation of exact repeat victimisation (same location) was identified in any of the aforementioned RV/ N-RV experiments.



	Days																												
Meters	**0-1**	**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**	**> 28**
Same location	0.00	0.00	11.35	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	11.22	0.00	0.00	0.00	0.00	0.00	0.00	10.19	1.00
1-250	**5.51**	**1.81**	1.52	1.17	1.41	**1.88**	1.44	1.06	1.45	**1.70**	1.26	1.19	**1.64**	**1.72**	1.45	**1.69**	**1.55**	**1.72**	1.28	1.52	1.16	1.09	1.49	**1.76**	**2.35**	1.10	**1.63**	0.74	0.98
251-500	**2.78**	**1.51**	**1.52**	**1.58**	**1.39**	**1.93**	**1.92**	1.27	1.11	**1.56**	**1.51**	**1.42**	1.08	**1.41**	1.27	1.24	**1.67**	1.28	1.10	**1.54**	1.37	1.38	1.33	**1.48**	1.18	**1.38**	**1.43**	1.18	0.99
501-750	**1.73**	1.25	**1.89**	**1.35**	**1.40**	**1.59**	1.28	**1.39**	1.06	**1.34**	1.04	**1.58**	0.77	1.26	**1.34**	**1.45**	0.71	1.12	1.29	**1.62**	1.23	1.03	0.96	0.90	1.02	1.23	1.12	1.15	0.99
751-1000	1.24	**1.34**	**1.32**	**1.66**	1.17	**1.41**	**1.31**	1.21	0.93	**1.29**	1.24	1.15	1.17	0.95	**1.39**	1.04	1.20	1.26	1.28	**1.37**	1.12	1.02	1.28	1.15	0.93	0.86	1.05	**1.36**	0.99
1001-1250	**1.57**	1.18	1.06	1.07	**1.40**	**1.32**	0.98	1.19	1.25	**1.45**	0.89	0.87	1.16	1.19	1.22	1.17	1.05	1.08	**1.46**	1.17	**1.46**	1.03	**1.27**	1.04	0.95	0.98	**1.34**	1.24	0.99
1251-1500	1.02	1.22	**1.28**	1.19	1.17	1.21	**1.41**	**1.40**	**1.25**	**1.26**	**1.24**	1.08	1.23	1.20	**1.29**	1.08	1.00	**1.37**	**1.35**	1.12	1.09	0.98	**1.35**	**1.49**	1.17	**1.31**	**1.28**	1.12	0.99
1501-1750	1.08	1.05	1.17	1.13	1.13	1.18	**1.23**	1.16	1.12	1.19	**1.39**	1.17	**1.46**	1.21	0.89	1.02	0.91	1.14	1.02	0.99	**1.36**	**1.41**	1.17	0.97	0.98	1.18	**1.45**	1.12	0.99
1751-2000	**1.33**	**1.35**	0.91	1.16	**1.32**	1.10	1.11	**1.22**	**1.36**	1.10	1.12	0.99	1.17	**1.43**	1.04	1.19	**1.33**	0.90	1.20	1.04	1.03	**1.26**	**1.30**	1.20	1.14	1.18	1.08	1.02	0.99
2001-2250	1.00	1.16	**1.29**	1.12	1.19	**1.28**	**1.32**	**1.22**	1.13	1.18	**1.32**	**1.28**	**1.27**	1.17	**1.26**	1.03	1.07	1.14	1.10	1.04	1.17	1.10	1.17	0.95	1.11	0.95	1.17	0.96	0.99
2251-2500	1.19	1.13	**1.33**	1.02	**1.28**	**1.26**	0.93	**1.21**	0.94	**1.32**	1.14	1.09	**1.23**	1.10	**1.22**	**1.27**	1.04	1.20	1.06	0.99	**1.30**	1.07	1.19	1.06	**1.29**	1.15	1.01	1.10	0.99
2501-2750	0.88	0.83	**1.21**	1.11	1.11	0.92	0.95	1.05	1.02	1.00	**1.27**	1.05	**1.27**	1.05	1.00	1.09	0.96	1.04	1.09	1.10	1.01	1.02	1.06	1.13	1.01	0.92	0.93	0.97	1.00
2751-3000	1.00	1.03	**1.25**	1.13	1.19	1.08	**1.22**	1.11	1.03	1.02	1.09	1.09	1.15	1.06	**1.19**	1.13	1.11	1.06	1.03	1.01	1.05	**1.23**	1.02	1.14	0.99	1.02	1.06	1.02	1.00
3001-3250	1.16	**1.26**	0.99	1.05	0.98	**1.18**	**1.21**	1.13	**1.21**	1.15	**1.30**	0.97	1.11	0.89	1.13	1.17	1.12	1.09	1.02	0.92	**1.22**	**1.20**	1.00	**1.25**	1.13	1.04	1.00	0.98	1.00
3251-3500	1.21	1.10	**1.25**	1.09	**1.24**	0.94	1.06	**1.18**	1.10	1.16	1.13	0.97	**1.21**	0.99	1.08	1.05	1.05	1.06	1.16	1.12	0.95	1.15	1.12	1.06	1.01	1.12	1.13	0.98	1.00
3501-3750	1.22	**1.20**	1.15	1.01	0.99	1.06	1.09	**1.19**	1.14	0.92	0.96	1.07	1.02	1.08	1.17	**1.18**	1.00	0.92	**1.20**	**1.27**	0.98	1.08	**1.17**	0.98	**1.17**	1.06	1.09	1.06	1.00
3751-4000	0.96	**1.28**	**1.28**	1.09	1.02	1.14	1.02	0.98	1.13	**1.28**	1.04	1.05	**1.17**	1.15	0.93	1.05	0.97	1.11	1.08	1.01	1.07	1.05	1.04	1.15	1.12	0.95	1.04	0.98	1.00
4001-4250	1.17	1.04	1.06	1.04	1.03	**1.30**	0.85	1.01	1.07	0.99	1.13	1.01	1.04	**1.16**	**1.35**	0.96	**1.33**	0.99	**1.17**	0.98	0.94	**1.23**	1.13	1.06	0.98	0.98	1.10	1.14	1.00
4251-4500	1.03	1.01	1.05	1.11	1.04	0.98	1.06	1.15	1.07	1.12	1.00	**1.22**	1.00	0.97	1.03	0.98	1.04	1.12	1.11	1.08	1.08	0.98	1.10	1.12	0.87	**1.27**	1.01	0.82	1.00
4501-4750	1.04	0.92	1.01	1.03	**1.16**	**1.17**	1.15	1.09	1.08	1.13	1.07	1.02	**1.21**	**1.15**	1.07	1.14	1.11	1.11	**1.20**	1.12	1.00	1.13	1.01	1.14	0.92	0.91	1.02	**1.16**	1.00
4751-5000	**1.20**	**1.20**	1.11	1.00	0.97	1.11	0.96	**1.23**	**1.18**	**1.24**	1.11	1.11	1.10	**1.21**	1.01	1.10	1.14	1.09	1.07	0.95	1.10	1.14	1.14	**1.16**	1.00	1.03	0.98	1.09	1.00
> 5000	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	1.00	0.99	0.99	0.99	0.99	1.00	1.00	0.99	1.00	**1.00**

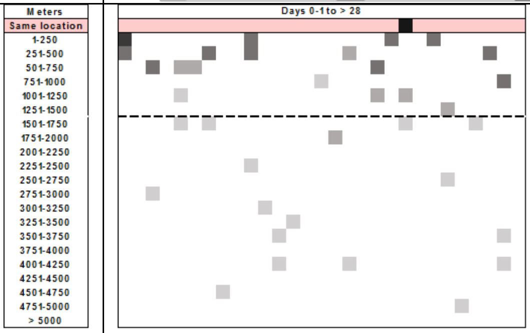
Table 5.17 - 250 m x 20 and 1 days x 28 (999 iterations) = 5 km and 28 days

5.4.2 Disaggregated Car Key Burglary and Regular Burglary Results

Now to compare mid-point year RV/ N-RV patterns for the two burglary types using a spatial bandwidth of 250 m (x 20) and a temporal bandwidth of 1 day (x 28), i.e. the most disaggregated spatio-temporal parameters that were analysed in the preceding section. The results of this analysis are presented in Figure 5.10 below, drawing on a visualisation technique employed by Johnson et al. (2007, p.211), and including only those cells with an odds ratio ≥ 1.20 and p ≤ 0.05. The diagrams in the left hand column relate to Car Key burglary and the right to Regular burglary. Each cell has been shaded according to the size of the associated odds ratio, with black representing the highest values, i.e. an odds ratio of 10.00 and over. As before, the pink rows represent the same location (exact repeats). A dashed horizontal line has been added at 1,500 m to facilitate comparison between the two burglary types. The diagrams are intended to provide a quick means of identifying and comparing spatio-temporal patterns within the burglary data samples, rather than having to look at the detailed output tables generated by the NRC software.

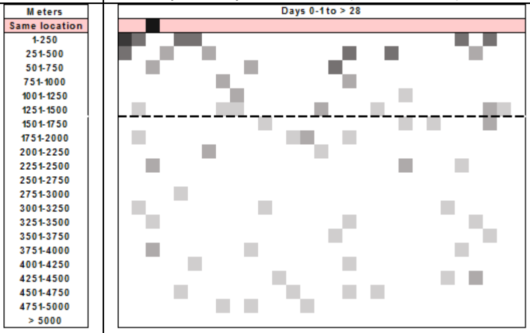
Looking at the diagrams for each crime type, there appears to be a much clearer, and far more stable, spatio-temporal decay pattern for Regular burglary than for Car Key burglary, including for exact repeats. Although there is some significant same location over-representation evident in three of the Car Key burglary diagrams, the reliability of this is questionable given that two of the cells relate to just one repeat offence each and the third is not a genuine repeat (identified by checking the linked addresses and geocoding in the original police-recorded crime data set).

	Car Key burglary	Regular burglary
Odds:	░ 1.20 - < 1.50 ▒ 1.50 - < 1.80 ▓ 1.80 - < 3.00 ■ 3.00 - < 10.00 █ 10.00 and over	



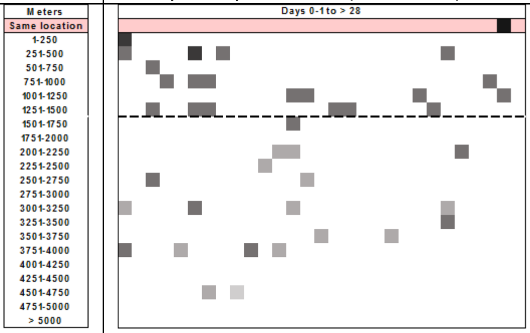
Mid-point year 2010 (n = 1,885)

NRC unable to process due to large sample size (n = 18,204)

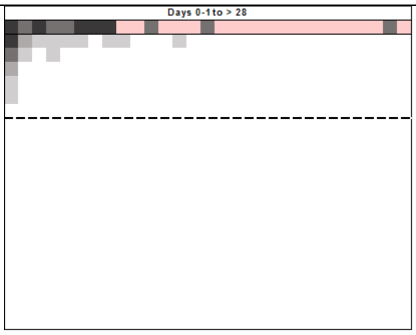


Mid-point year 2011 (n = 1,859)

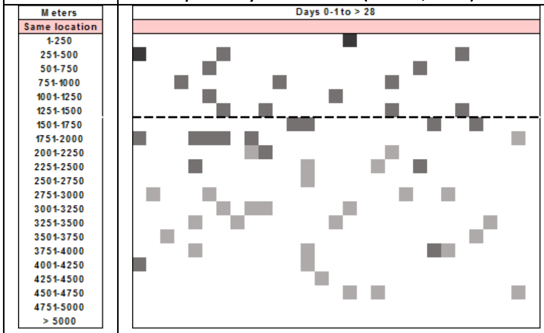
NRC unable to process due to large sample size (n = 17,186)



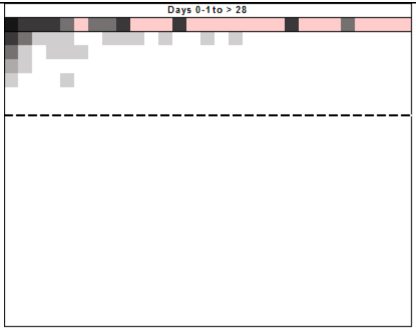
Mid-point year 2012 (n = 1,164)



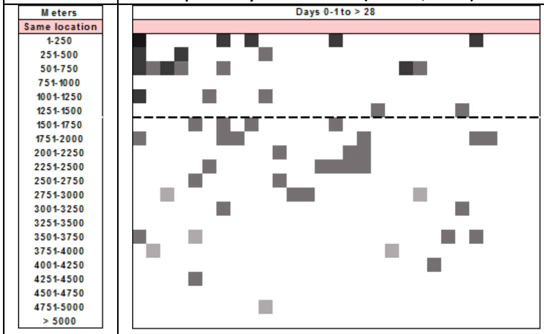
Mid-point year 2012 (n = 12,526)



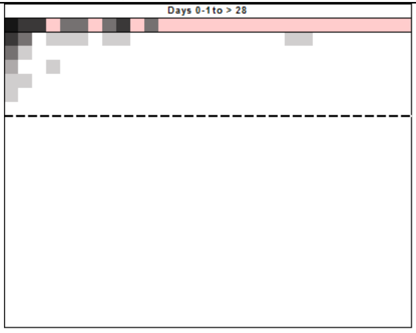
Mid-point year 2013 (n = 1,145)



Mid-point year 2013 (n = 11,146)



Mid-point year 2014 (n = 864)



Mid-point year 2014 (n = 9,398)

Figure 5.10 - Car Key burglary and Regular burglary NRC analysis by mid-point year - 5 km and 28 days

The two diagrams in Figure 5.11 below were created to assess the stability of the observed RV/ N-RV patterns for each burglary type over time. Because only three years' worth of Regular burglary data were analysed with the Near Repeat Calculator, the diagrams in Figure 5.11 were created using information from only the last three mid-point years in Figure 5.10 above (recall that the associated diagrams only relate to those cells with an odds ratio ≥ 1.20 and p ≤ 0.05). Each of the cells that experienced over-representation in the relevant Figure 5.10 diagrams were assigned a new value of 1, stacked for each of the respective burglary types, and then summed. Any cells that achieved a value ≥ 2 were retained, i.e. they experienced over-representation for at least 2 out of the 3 years considered, and all others were disregarded. Those cells that were assigned the maximum value of 3 are highlighted in red in Figure 5.11 (value of 2 shown in grey).

Looking first at the Regular burglary diagram in Figure 5.11, it is interesting to note that all of the enduring over-representation is above the 1,500 metre reference line, up to a maximum of 1,250 m. Even when the '0 to 1 days' column is discounted, there is still evidence of a relatively stable, and conventional RV/ N-RV pattern, the latter extending up to a maximum of 500 m. Assuming that the '0 to 1 days' column relates to offences committed on the same date, stable over-representation of exact repeats presents for a period of 10 days following an initial event. These findings are broadly in line with the related literature, for example, Johnson and Bowers noted communicability of residential burglary risk in Merseyside up to a distance of 300-400 m and for a period of 1-2 months (2004, p.237), and Johnson et al. identified distances of 300 m and 500 m respectively in Bournemouth and Wirral, both for a period of 2 weeks (2007, p.214). There is, to the best of the author's knowledge, no published research on Car Key burglary RV/ N-RV patterns, however, it would appear from Figure 5.11 that offences are typically highly temporally clustered following an initial event, and also that N-RV over-representation occurs at much longer spatio-temporal bands than for Regular burglary. Further, there is no evidence of stable over-representation of exact repeats during the three mid-point years considered here.

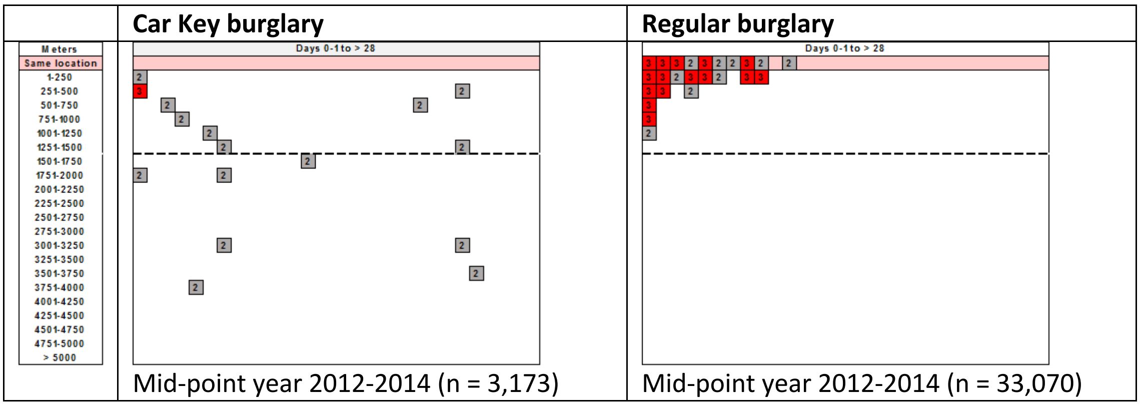


Figure 5.11 - Statistically significant RV/ N-RV over-representation stable for at least 2 out of 3 years - Car Key burglary and Regular burglary - 5 km and 28 days

Since it was possible to analyse five years' worth of Car Key burglary data with the Near Repeat Calculator, the summing process was repeated again using all associated diagrams in Figure 5.10, the only difference being that the maximum cell value was now 5. Again, only those cells that experienced significant over-representation for at least 2 of the years considered were retained. As can be seen from Figure 5.12 below, the only spatio-temporal band that consistently experienced significant N-RV over-representation was 251-500 m and 0-1 days (5 out of 5 years). Given the NRC '0 to 1 days' issue, this pattern can be taken to represent spree-like offending, rather than conventional near-repeat behaviour where offenders return to the locality of their previous targets but on a different date. The most stable over-representation of what might be considered genuine 'near' repeat activity for Car Key burglary was at the 501-750 m and 3 days spatio-temporal band (4 out of 5 years). This suggests that, although Car Key burglars might return to the general vicinity of their most recent offences in the short-term, they will typically leave a safe buffer distance between an originating event and a near-repeat event. Possible reasons for this include: (i) high levels of collective efficacy within target areas (recall the discussion in Section 5.4.1 re post-crime activity in affluent communities), (ii) homogeneity of population characteristics within target areas, that is, offenders run the risk of being identified as 'others' if they return to previously victimised locations in the short-term, and (iii) effective micro-level guardianship within target areas, such as home security (Johnson et al., 1997, p.238). Although a contagion distance of 501-750 m is longer than identified in other general residential burglary N-RV studies in the UK, it does still appear to be a form of near-repeat activity when viewed in the context of the large areal extents that mobile offenders can potentially cover. Therefore, it is inferred that, during a single 24 hour period, Car Key burglars typically target properties within fairly close proximity of each other (≤ 500 m), but do not return to the immediate surroundings of their recent offences in a consistent manner, unlike Regular burglars.

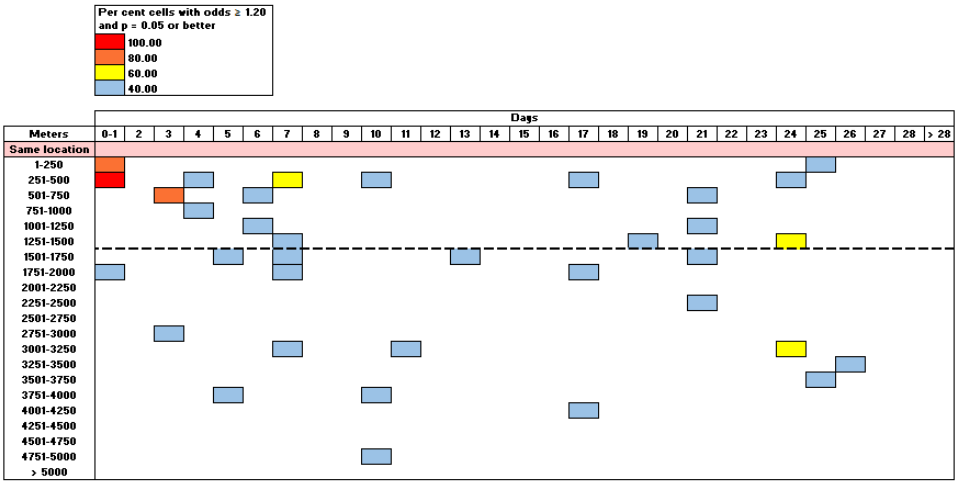


Figure 5.12 - Statistically significant RV/ N-RV over-representation stable for at least 2 out of 5 years - Car Key burglary - 5 km and 28 days

5.5 Summary

The majority of the findings presented in this chapter are as previously hypothesised, with some notable differences identified between Car Key burglary and Regular burglary, including level of spatio-temporal clustering. However, the Spearman's coefficients between the Car Key burglary rates and the independent variables analysed were not as strong as had been previously hoped. As already discussed, a possible explanation for this is that some of the data was not sufficiently detailed to allow for the unmasking of related trends. For example, only basic information was supplied regards the characteristics of stolen vehicles and so it was not possible to ascertain if different makes/ models/ ages are targeted in different areas and for different reasons. Unfortunately, this appears to have impacted on the predictive capacity of the static risk surface, with a best *k*-NN R square value of only 0.352 (3 DP). It is unlikely, therefore, that any related model will be able to predict with 100 % accuracy the locations of Car Key burglary offences. Despite this, the use of *k*-NN regression to estimate crime rates is novel. Although there appears to be quite a lot of information (grey literature) with regards to the technique in general, a search on Google Scholar for "knn regression" + "crime rate*" only returned eight results (as at 17/03/20, 14:20); 3 conference papers, 3 theses, 1 citation, and 1 book chapter – this was not a detailed search, but it is clear that the method is relatively under-explored in crime science. Recalling that the *k*-NN regression approach employed here is non-spatial, the incorporation of an area type juxtaposition variable (IDW_0.35) appears to have been quite successful, and something that is perhaps not routinely considered in crime analysis. This presents an important opportunity for future work, i.e. through the combining of feature space and geographical space. A final consideration in relation to the *k*-NN regression approach employed here is that the dependent variable values could instead be distance-weighted according to their proximity to a target point (Raschka, 2018, pp.15-16). This would ensure that target points are assigned average values that are most representative of their nearest neighbours.

In terms of identifying spatio-temporal parameters for the dynamic risk surfaces, the Near Repeat Calculator results indicate that a much larger proportion of the West Yorkshire study area would need to be searched in order to obtain the same hit rate for Car Key burglary as for Regular burglary in the days immediately following an initial event. This in turn justifies the use of a static risk surface (flag account) to filter dynamic risk (boost account) within more spatially extensive offender activity spaces. Given the significant, albeit relatively unstable, near-repeat over-representation that was evident at longer spatio-temporal bands for Car Key burglary, a variety of buffer distances, up to a maximum of 5,000 metres, will be tested in the next chapter

(Developing and Evaluating a Combined Risk Model) – longer distances than this are expected to be operationally impracticable, especially since the static risk surface can only explain circa 35 per cent of the variation in LSOA Car Key burglary rates.

Chapter 6 Developing and Evaluating a Combined Risk Model

6.1 Introduction

Recalling the core research rationale that all Burglary Dwelling offences are not created equal and thus exhibit different spatio-temporal signatures, evidence of which is provided in earlier chapters of this thesis, the overall aim of this chapter is to assess whether boost account theory, in the form of prospective buffers, is an effective predictor of future Car Key burglary offences. As shown in Figure 6.1 below, there are five main modelling steps to be covered in this chapter. Given the findings of the RV/ N-RV analysis reported in Chapter 5, including that Car Key burglary offences appear to be highly temporally clustered following an initial event, a dynamic risk model will first be developed that is able to make daily predictions over an extended time period. The resulting model will be run initially with Car Key burglary data, and three performance measures will be recorded at each iteration, including the hit rate and prediction accuracy index. Once the dynamic model has been run in completeness for Car Key burglary, that is, for the duration of the test period, the process will then be repeated with Regular burglary data. Resulting daily performance measures for each crime type will subsequently be compared with a view to providing additional empirical evidence in support of the research rationale. The second part of the chapter will first outline the development of the combined risk model, and this will then be tested with the Car Key burglary data only. The purpose of the combined model is to understand if risk heterogeneity can accurately reduce the spatial extent of transient risk whilst still maintaining previously observed hit rates, i.e. those generated by the dynamic model. This approach is expected to be particularly relevant for the policing of crime types where related offences tend to be spatially dispersed over time and which would therefore require extensive geographical buffering in order to capture a substantial proportion of future incidents.

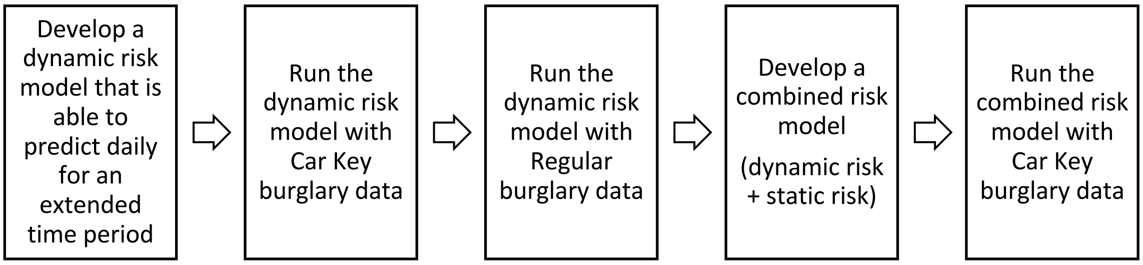

Develop a dynamic risk model that is able to predict daily for an extended time period
Run the dynamic risk model with Car Key burglary data
Run the dynamic risk model with Regular burglary data
Develop a combined risk model (dynamic risk + static risk)
Run the combined risk model with Car Key burglary data

Figure 6.1 - Main modelling steps in this chapter

6.1.1 Novelty of the Current Research Models

It is pertinent to mention here that the present work is, to the best of the author's knowledge, the first to assess the performance of a combined risk model over 24 hour periods. This level of analysis has been made possible due to the availability of a large volume of spatially and temporally disaggregated residential burglary data which, when used in conjunction with open source coding software (R), will allow numerous 'daily' predictions (30,618 days) to be generated and then 'retrospectively' tested, i.e. as though the process is being conducted in real-time. The rationale behind making daily predictions is that, for those offence types, such as Car Key burglary, where the associated hot spots are very transient in nature, crime practitioners should be regularly reassessing the spatio-temporal distribution of risk relative to any identified RV/ N-RV patterns to ensure that resources are being deployed in the right place and at the right time.

Further, at the time of writing, the author is only aware of two other predictive crime models that have combined short-term risk ('boost account') and long-term risk ('flag account'), these being Johnson et al.'s 'ProMap' (2009) and Azavea's commercially available 'HunchLab 2.0' (now ShotSpotter® Missions^{TM}), together with their free to download, and seemingly closely related, 'PROVE' program (this is not currently accessible via the ShotSpotter® website [18/12/2019]). Methodological aspects of ProMap and PROVE were discussed in detail in Chapter 2, however, it was not possible to review HunchLab 2.0 due to a lack of independent, peer-reviewed material about the methods and associated software. Key differences between ProMap and PROVE and the current research models are that Johnson et al. (2009) only tested the accuracy of their combined risk surfaces, including 'ProMap * Houses * Roads', over nine one week periods between September and November 1997 (p.185), and Ratcliffe et al.'s PROVE software (2016b) only outputs aggregate predictions for either 7, 14, 21, or 28 days (p.12), and based on US census data (p.5). In addition, there are no examples of a combined risk model having been applied to a sub-category of an official HO crime classification, or tested over such a long time period, in this case, circa four years.

6.2 Developing the Current Research Models

Given the results of the current research RV/ N-RV analysis, it is hypothesised that a dynamic risk surface constructed from prospective buffers will be far less effective at predicting future Car Key burglary locations than at predicting future Regular burglary locations. This is because statistically significant over-representation of Car Key burglary N-RV was found to be highly temporally clustered following an initial event (spree-like offending), but also that additional, albeit generally unstable, N-RV over-representation was observed at much longer spatio-temporal bands than for Regular burglary. Further, there was little evidence of statistically significant over-representation of exact repeat victimisation for Car key burglary. Therefore, it was inferred that, during a single 24 hour period, Car Key burglars typically target properties within fairly close proximity of each other (≤ 500 m), but do not return to the immediate surroundings of their recent offences in a consistent manner, unlike Regular burglars.

Figure 6.2 below is included to remind the reader of some of the enduring RV/ N-RV patterns that were identified in the previous chapter, and indicates that a much larger proportion of the West Yorkshire study area will need to be searched to obtain the same hit rate for Car Key burglary as for Regular burglary. If true, i.e. the dynamic model performs less well for Car Key burglary, then this will justify the use of a static risk surface to filter transient risk within more spatially extensive offender activity spaces. The overall success of the combined model will lie in its ability to reduce the areal extent of the dynamic surfaces whilst maintaining the previously observed hit rates for these. To place this in an operational context, the delineation of smaller risk (patrol) areas is conducive to the efficient use of police resources, as well as contributing to the prevention and detection of crime, e.g. by helping to focus intelligence gathering activity.

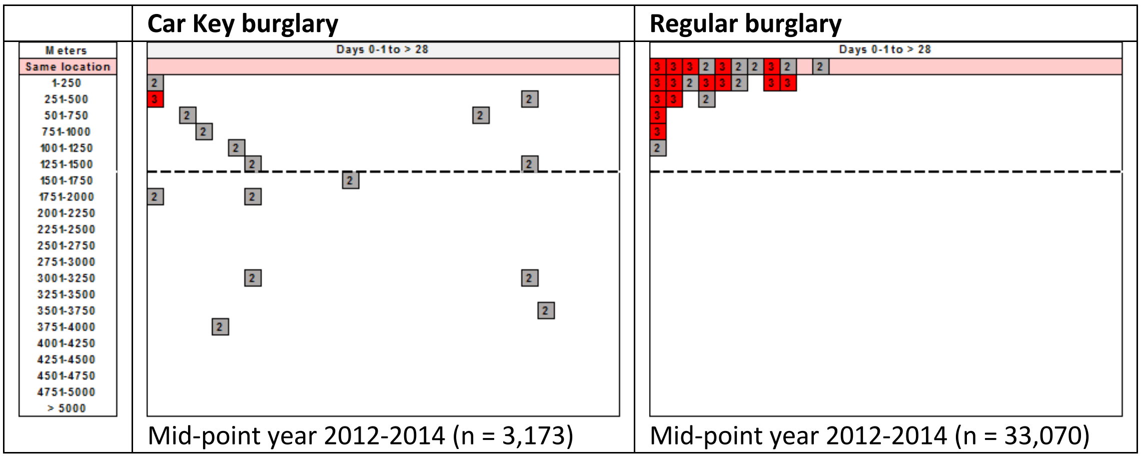


Figure 6.2 - Statistically significant RV/ N-RV over-representation stable for at least 2 out of 3 years - Car Key burglary and Regular burglary - 5 km and 28 days

Importantly, the RV/ N-RV patterns that are typically associated with Regular burglary, as identified in this work, as well as other studies on the subject, appear to be far more conducive to prospective crime modelling because there is a much clearer, and operationally practicable, spatio-temporal decay pattern. As such, there is unquestionable value in drawing a circle around recent Regular burglary offences with a view to estimating where offenders might strike next. Given that Car Key burglaries appear to occur in sprees, as indicated by the enduring N-RV over-representation observed at 251-500 m and 0-1 days, and also because the current research did not uncover any evidence of stable exact repeat over-representation, the spatio-temporal decay function that is associated with this crime type is more likely to take the form of a bell-shaped kernel with a dip in the middle, reminiscent of the buffer zone around an offender's home (Brantingham and Brantingham, 2010, p.236), as illustrated in Figure 6.3 below. Therefore, the centre of a prospective buffer is unlikely to capture many, if any, future car key burglary offences in the days immediately following an initial event, however, it is anticipated that more peripheral areas might capture 'near-repeat' activity occurring at longer distances, e.g. 501-750 m, as offenders move between different target locations. Also, considering Carden's finding that Car Key burglars travel a median distance of 4.9 km from home base to offence (2012, p.74), then there is likely to be additional merit in the use of prospective buffers because this parameter indicates that offenders' mobility patterns are still subject to a level of spatial constraint, albeit less so than Regular burglars. To explain, a single Car Key burglary offence indicates that an offender, or offender group, is active in an area, and the location of this might, therefore, provide a best guess as to the likely maximum spatial extent of their criminal activity.

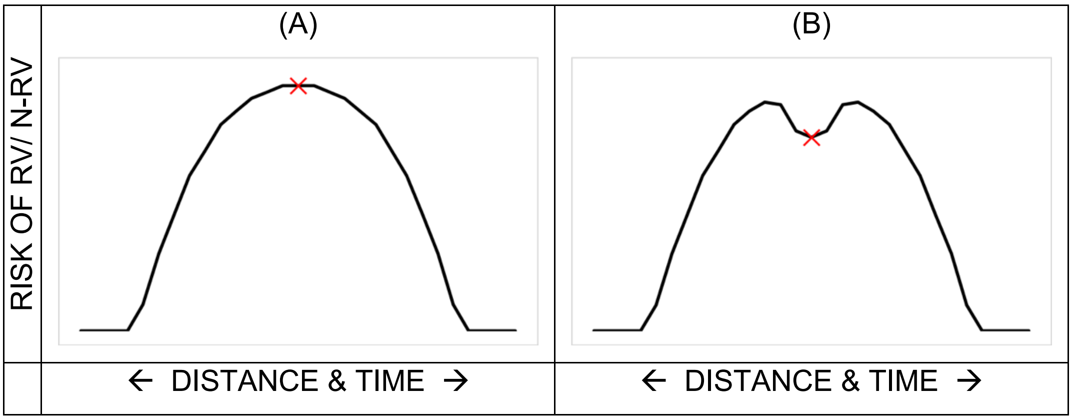


Figure 6.3 - (A) Bell shaped kernel, (B) dipped kernel ('safe' buffer), x = initial event (based on: Bowers et al., 2004; Johnson et al., 2007b, p.19; Brantingham and Brantingham, 2010, p.236)

In light of the above, the more complex prospective modelling techniques that are inherent to ProMap and PROVE are not deemed as relevant for Car Key burglary. Therefore, a simple heuristic buffering method, similar to that employed in Trafford (Fielding and Jones, 2012), will be used to generate the daily risk predictions, and a number of different buffer distances will be tested to assess the impact of these on associated performance measures.

The flag element of the current research combined model will differ from both ProMap and PROVE in terms of the input data sets used, the method by which the risk heterogeneity surface was derived, and the conceptualisation of risk, with areas simply defined as being either at risk, or not at risk, i.e. an on-off/ 0-1 approach. Rather than presenting crime counts or risk intensity values at the grid cell level, the current research will take the buffers from the dynamic model and clip these using the static risk surface. This means that only those parts of a buffer that are predicted to have an above, or equal to, median Car Key burglary rate at the LSOA level will be considered at risk of future victimisation. Recall that *k*-NN regression, i.e. a non-parametric approach, was used to derive the static risk surface with a view to mitigating the issues that are inherent to the use of parametric statistical techniques with non-normally distributed data, and, in the case of the current research, highly positively skewed crime rates. As was discussed in the previous chapter, the *k*-NN model only explained circa 35 per cent (R^2 value) of the variation in Car Key burglary rates at the LSOA level, which means that the combined model might not perform as well as was previously hoped, i.e. had a more representative risk heterogeneity surface been identified. Although such an R^2 value is not unusual in social sciences research (e.g. see: Bartholomew et al., 2008, p.157; Cahill and Mulligan, 2007, p.179), it does present an opportunity for future work, including how the crime data samples on which the regression was performed might be improved, e.g. by obtaining information on vehicle makes/ models, and also by identifying any important omitted predictor variables.

6.2.1 Measuring the Models' Performance

One of the main considerations when developing a predictive crime model is how best to measure its performance. In an operational context, this might equate to something along the lines of: 'were fewer offences committed in high risk areas during a period of intervention?' However, because it is not possible to test the current research models in real-time, performance metrics will instead be used to give an indication of how many offences the models might have captured (prevented) had they been used to inform operational activity, such as disruption patrols and target hardening. A key consideration in the current work will be how the areal extents of any identified high risk areas compare to the size of the study/ policing area (West Yorkshire), i.e. they should be operationally practicable. In light of this, three key indicators will be used to assess the models' performance in the current research, these being:

1. Hit Rate/ Capture Rate
2. Prediction Accuracy Index (PAI)
3. Search Efficiency Rate (SER)

Hit Rate/ Capture Rate

Perhaps the most basic and commonly employed method for assessing the performance of a crime model is the hit rate. This metric is relatively straightforward both to calculate and interpret, being simply the total number of offences within an identified high risk area/ hot spot divided by the total number of offences within the study area (or policing area). For example, if 25 out of 200 offences occurred in a high risk area, then the associated hit rate would be 12.5%. The main issue to be aware of here, however, is that no account is taken of the size of the risk area relative to that of the wider area in which it is situated. For example, if an entire study area is identified as being at risk of future victimisation, then the hit rate would be 100%, which is unlikely to be of any operational use (Bowers et al., 2004, p.645; Chainey et al., 2008, p.12).

Search Efficiency Rate (SER)

This performance measure, proposed by Bowers et al. (2004), provides the hit rate per km^2 (or other area unit) in areas that have been defined as being at risk of victimisation (p.645). So, if 50 offences were captured in an area of 100 km^2, the associated Search Efficiency Rate (SER) would be 0.5. Although this method might be useful, say, from the point of view of assessing the performance of different crime modelling techniques within the same study area, a potential limitation of the approach is that it does not reference the size of the associated study area (Chainey at el., 2008, p.12; Chainey, 2014, pp.98-99). For example, the SER for 20 crimes captured in 2 km^2 of a 50 km^2 study area would be 10 and the SER for 20 crimes captured in 2 km^2 of a 500 km^2 study area would be 10. In the case of the 500 km^2 study area, the risk area is much smaller relative to the overall size of the study area, which should be more informative from a resource deployment perspective (Chainey at el., 2008, p12; Chainey, 2014, p.99).

Prediction Accuracy Index (PAI)

Recognising the potential limitations of these two approaches, Chainey et al. (2008) developed a method known as the Prediction Accuracy Index (PAI), which incorporates the size of the study area in the associated calculation. The formula for calculating the Prediction Accuracy Index (PAI) is shown in Equation 6.1 below; it is the hit rate (number of offences in a risk area / number of offences in a study area) divided by the area percentage (size of a risk area / size of a study area) (Chainey et al., 2008, p.12; p.14). The process of dividing the two percentage figures means that the result is a ratio of the hit rate to the area percentage, which gives a better indication of likely operational practicability than taking the hit rate in isolation. So, to illustrate, a hit rate of 100 per cent in 100 per cent of a study area would generate a PAI value of 1 (Chainey et al., 2008, p.14), as would a hit rate of 25 % in 25 % of a study area, i.e. finding one quarter of future offences in one quarter of a study area is no better than finding 100 % in 100 % of a study area. Ideally, therefore, we should be looking for a PAI value greater than 1. For example, if one method captured 10 % of offences in 25 % of a study area (PAI = 0.40) and another captured 37 % of offences in 25 % of a study area (PAI = 1.48), we should probably be more interested in developing the latter.

$$\frac{\left(\frac{n}{N}\right) * 100}{\left(\frac{a}{A}\right) * 100} = \frac{Hit\ Rate}{Area\ Percentage} = Prediction\ Accuracy\ Index$$

Equation 6.1 - Prediction Accuracy Index (PAI) formula (Source: Chainey et al., 2008, p.14)

Expected Results

The dynamic model is expected to generate better hit rates for Regular burglary than for Car Key burglary relative to the size of the associated risk areas, which should be reflected in larger PAI values for the former. Search Efficiency Rates will also be provided for information but they are not directly comparable with Car Key burglary due to the underlying rate of offences. The PAI will be particularly useful for comparing the performance of the dynamic and combined models for Car Key burglary, that is, the values should be larger for the combined model if the dynamic risk areas have been accurately filtered (reduced in size) using the static risk layer. Results will be presented primarily as median values, which is a typical in the literature, for example, Johnson et al. presented median performance accuracies for ProMap (2009, p.188) and Chainey et al. averaged many individual PAI calculations for different hot spot mapping techniques (2008, pp.14-15).

6.2.2 Crime Input Data Sets

Both of the current research models were tested using Car Key burglary data, and four prediction periods were chosen to capture seasonal variations, etc., these being; census year (six months either side of 27th March 2011), 2012, 2013, and 2014. Only the dynamic model was tested with Regular burglary data, i.e. to ascertain if the approach was more effective for this crime type than for Car Key burglary. For each year and crime type tested, a .csv file of all offences for that year, plus 28 days prior, was read into the coding software. The author started with the same 'clean' data set that was used to undertake the RV/ N-RV analysis (75,377 burglary offences) and then selected sub-sets of this for the two burglary types and four prediction periods, for example, a selection criteria of earliest committed date >= 04/12/2011 and <= 31/12/2012 was used to identify offences for the 2012 Car Key burglary predictions 'all_crimes.csv' (Figure 6.7, step 4). This ensured that all relevant previous offences were available to generate the first risk surface. Table 6.1 below shows the four prediction periods, the number of days predicted for, and the number of model iterations with no 'future' offences ('0 offence days'). So, taking the 2012 data set, the first daily prediction was from 01/01/2012 12:00 until 02/01/2012 11:59 and the last was from 30/12/2012 12:00 until 31/12/2012 11:59. A total of 14,580 daily predictions each were made for the dynamic and combined models (number of days x 10 buffer distances). Once a model had been run for each of the annual prediction periods, the results files were merged by buffer distance to facilitate analysis using the entire four years' worth of data. It is useful here to suggest some possible reasons for the zero offence days, including that they could be an artefact of the strict 24 hour search constraint that was used to select 'future' burglary offences (explained in next section), or that previous/ ongoing operational activity could have prevented offences on certain days, for example, if offenders were detained as a result of this.

Data Set	**From**	**Until**	**Days**	**0 offence days**
Census	26/09/2010, 12:00	26/09/2011, 11:59	365	2 (0.5%)
2012	01/01/2012, 12:00	31/12/2012, 11:59	365	17 (4.7%)
2013	01/01/2013, 12:00	31/12/2013, 11:59	364	29 (7.9%)
2014	01/01/2014, 12:00	31/12/2014, 11:59	364	42 (11.5%)

Table 6.1 - Model prediction periods

6.2.3 Modelling Algorithms

The date and time search parameters for the dynamic and combined models were determined by the findings of the preceding chapter. When five years' worth of Car Key burglary offences (single combined data set) were analysed using Ratcliffe's Near Repeat Calculator (10 x 500 m distance bands and 8 x 7 day time bands), there was statistically significant over-representation of near-repeat victimisation at p = 0.001 for a period of four weeks. 28 days was therefore chosen as the past offences 'dates' parameter, i.e. for every model iteration the dynamic risk surface was constructed from crimes committed <= 28 days prior. The same 'dates' parameter was used for the Regular burglary dynamic risk layer so that direct comparisons could be made between the two offence types at the results stage.

Because the models are geared towards an operational setting, that is, daily predictions with a view to informing tactical resource deployment, the next decision was around how best to define the 24 hour prediction period. Recognising that Car Key burglaries are most likely to occur during the early hours – peak time period of 01:00-05:00 (based on the aoristic analysis findings in Section 3.8) – this was defined as being from 12:00 noon ‘today’ until 11:59 am ‘tomorrow’, thus comfortably encapsulating the overnight hot time period. The end time of 11:59 was used because the committed times on the crime records were in hours and minutes, i.e. 12:00-11:59 = 24 hours. Notably, the choice of 24 hour prediction period meant that prior offences were selected up to 11:59 on a prediction date, thus ensuring that the previous night’s offences were informing the distribution of risk for the next hot time period. Operationally, this would equate to last night’s offences (late evening yesterday/ early hours ‘today’) informing tonight’s activity. Since the ‘future’ offences data was available to the author, it was possible to run each model as though predicting in real-time.

Figure 6.4 and Figure 6.5 below are included to illustrate the intricacies of the date and time search process. For example, assuming a 24 hour prediction period starting 05/02/2015, 12:00 (‘today’), all Car Key/ Regular burglary offences (depending on the crime type being modelled) with a committed date and time range between 08/01/2015 12:00 and 05/02/2015 11:59 would be used to create the prospective dynamic buffers surface. To test how well the model performed during the 24 hour prediction period, all offences with a committed date and time range between 05/02/2015 12:00 and 06/02/2015 11:59 would then be plotted onto the dynamic risk surface and the number of intersecting points recorded. After this, the model would iterate to the next prediction period, i.e. a new ‘today’ (previous ‘today’ date + 1), and the previous 28 days’ offences search parameters would also increment by + 1, thus incorporating yesterday’s ‘future’ offences and disregarding any offences committed between 08/01/2015 12:00 and 09/01/2015 11:59.

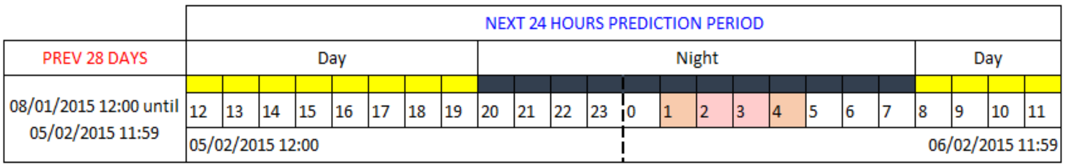


Figure 6.4 - Hypothetical model iteration 1 - previous and future offences search parameters

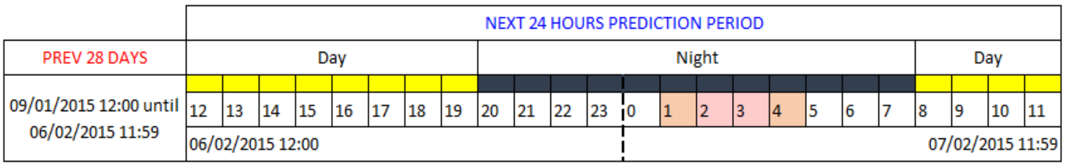


Figure 6.5 - Hypothetical model iteration 2 - previous and future offences search parameters

Only those offences that were definitely known to have occurred during a given prediction period (according to the crime record) were included in the related analysis. For example, a future offence committed between 05/02/2015 12:00 and 08/02/2015 16:55 would be excluded from the Figure 6.4 iteration because it could potentially have occurred after 06/02/2015 11:59, which would not constitute a fair test of the method. Similarly, offences spanning the lunchtime period (morning into afternoon) were disregarded, which might have affected Regular burglaries more than Car Key burglaries. Figure 6.6 below shows how the coding algorithm date and time variables would be initialised for the Figure 6.4 iteration, and then used to subset the crime data.

```
today <- 05/02/2015 12:00
tomorrow <- 06/02/2015 12:00
prev28days <- 08/01/2015 12:00

prev28days_crimes <- subset(all_crimes, EarliestDT >= prev28days & LatestDT < today)
next24hrs_crimes <- subset(all_crimes, EarliestDT >= today & LatestDT < tomorrow)
```

Figure 6.6 - Example variables and R code used to subset relevant offences

Given that the spatial decay was very unstable for Car Key burglary, a number of buffer radiuses were tested (10 x 500 m incrementing distance bands), however, it was the size of the resulting risk areas relative to the associated hit rates that were of most interest, as opposed to the buffer distances *per se*. Since there were far more Regular burglary offences than Car Key burglary offences, 10 x 100 m incrementing distance bands were used for the former to generate more realistically sized risk areas (500 m + buffers would likely have covered most of the study area). The algorithm for the dynamic model is outlined in Figure 6.7 below – all of the models were run in RStudio 3.6.0.

Dynamic model (prospective buffers) coding algorithm

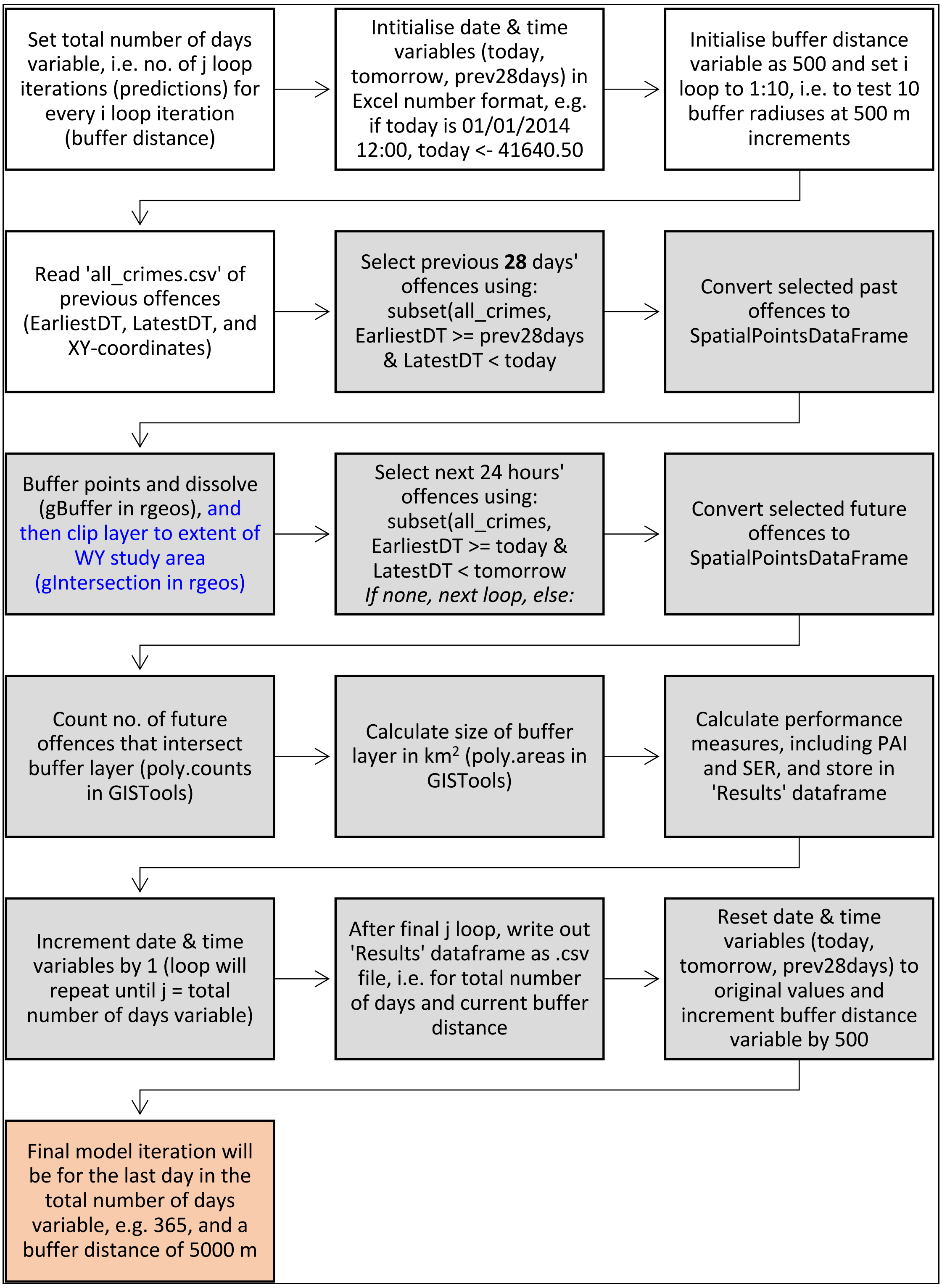

Set total number of days variable, i.e. no. of j loop iterations (predictions) for every i loop iteration (buffer distance)
Intitialise date & time variables (today, tomorrow, prev28days) in Excel number format, e.g. if today is 01/01/2014 12:00, today <- 41640.50
Initialise buffer distance variable as 500 and set i loop to 1:10, i.e. to test 10 buffer radiuses at 500 m increments
Read 'all_crimes.csv' of previous offences (EarliestDT, LatestDT, and XY-coordinates)
Select previous **28** days' offences using: subset(all_crimes, EarliestDT >= prev28days & LatestDT < today
Convert selected past offences to SpatialPointsDataFrame
Buffer points and dissolve (gBuffer in rgeos), and then clip layer to extent of WY study area (gIntersection in rgeos)
Select next 24 hours' offences using: subset(all_crimes, EarliestDT >= today & LatestDT < tomorrow *If none, next loop, else:*
Convert selected future offences to SpatialPointsDataFrame
Count no. of future offences that intersect buffer layer (poly.counts in GISTools)
Calculate size of buffer layer in km^2 (poly.areas in GISTools)
Calculate performance measures, including PAI and SER, and store in 'Results' dataframe
Increment date & time variables by 1 (loop will repeat until j = total number of days variable)
After final j loop, write out 'Results' dataframe as .csv file, i.e. for total number of days and current buffer distance
Reset date & time variables (today, tomorrow, prev28days) to original values and increment buffer distance variable by 500
Final model iteration will be for the last day in the total number of days variable, e.g. 365, and a buffer distance of 5000 m

Figure 6.7 - Dynamic model nested loop algorithm (j loop process steps are shown in grey)

The algorithm for the combined model was the same as for the dynamic model except that, rather than clipping the dynamic risk layer using the West Yorkshire study area extent (Figure 6.7, step 7, blue text), the static risk layer, having already been clipped to the study area extent, was instead used to perform the clip. This meant that only those parts of the dynamic risk layer that were predicted to have an above, or equal to, median Car Key burglary rate at the LSOA level were included in the future offences intersect operation.

6.3 Dynamic Model Results[5]

Although the performance measures for the two crime types were derived using different sized buffers (recollect that this was due to there being far more Regular burglary offences), useful insights can still be gained by comparing these, as is evident from Table 6.2 below. The left hand column of Table 6.2 contains the median daily performance measures for Car Key burglary, and the right hand column the median daily performance measures for Regular burglary. As can be deduced from the figures, the dynamic model produced higher median daily hit rates for Regular burglary than for Car Key burglary relative to the associated median daily risk area sizes (km^2), which is reflected in the higher PAI values. For example, creating 100 m buffers around the past 28 days' Regular burglary offences produced a median daily hit rate of **15 per cent** and a median daily risk area of **25 km^2** (based on 1,458 model iterations). Creating 500 m buffers around the past 28 days' Car Key burglary offences produced a median daily hit rate of **0 per cent** and a median daily risk area of **63 km^2**. Dividing the median daily risk area figures by the size of the West Yorkshire study (2029.3 km^2) area makes it easier to compare these results. For example, the median hit rate of **15 %** for Regular burglary was accompanied by a median risk area equivalent to **1 %** of the West Yorkshire study area, whereas for Car Key burglary, the **0 %** median hit rate was accompanied by a median risk area equivalent to **3 %** of the study area – the former is clearly far more conducive to the efficient deployment of police and partner agency resources.

Dynamic Model – Car Key Burglary All 4 Years (Census, 2012, 2013, 2014)						**Dynamic Model – Regular Burglary** All 4 Years (Census, 2012, 2013, 2014)					
Buffer Distance	**Hit Rate**	**Risk Area (km^2)**	**% WY**	**PAI (1DP)**	**SER (1DP)**	**Buffer Distance**	**Hit Rate**	**Risk Area (km^2)**	**% WY**	**PAI (1DP)**	**SER (1DP)**
500 m	0	63	3	0.0	0.0	**100 m**	15	25	1	11.8	14.0
1000 m	43	218	11	3.4	0.5	**200 m**	36	87	4	7.9	8.9
1500 m	60	417	21	2.7	0.4	**300 m**	52	168	8	6.0	6.8
2000 m	**75**	616	**30**	2.3	0.3	**400 m**	65	259	13	4.9	5.6
2500 m	89	810	40	2.0	0.3	**500 m**	**74**	351	**17**	4.2	4.8
3000 m	100	983	48	1.8	0.3	**600 m**	81	440	22	3.6	4.2
3500 m	100	1137	56	1.6	0.2	**700 m**	86	524	26	3.2	3.7
4000 m	100	1270	63	1.5	0.2	**800 m**	88	605	30	2.9	3.3
4500 m	100	1378	68	1.4	0.2	**900 m**	91	681	34	2.7	3.0
5000 m	100	1475	73	1.3	0.2	**1000 m**	92	753	37	2.5	2.8

*Table 6.2 - Median daily values for **Dynamic Model – Car Key Burglary** and **Dynamic Model – Regular Burglary** (grey column)*

[5] Days with no burglary offences were excluded from the results in this chapter; these could either be genuine 'no offence' days, or an artefact of the exact 24 hour selection criteria.

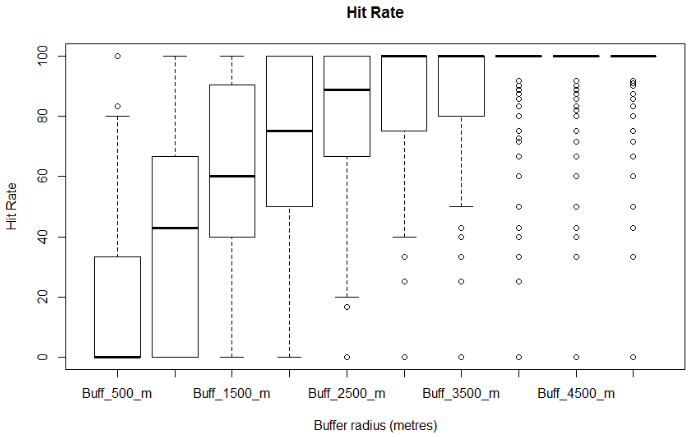
Note that the % WY figures in Table 6.2 above were derived by dividing the median Risk Area (e.g. 63 km^2 for the 500 m buffer) by the area of West Yorkshire, i.e. one calculation per buffer size.

Also of note in relation to Table 6.2 above is that the PAI and SER values for Regular burglary decreased as the size of the buffers increased, which is indicative of conventional distance-decay patterns. Conversely, the model generated a disappointing median daily hit rate and PAI for Car Key burglary at the smallest buffer distance, but these measures improved when 1000 m buffers were used, which supports the idea that Car Key burglars might typically leave a 'safe' distance around recent offences. After 1000 m, the PAI and SER values for Car Key burglary gradually decreased, with the high hit rates being penalised relative to the size of the associated risk areas.

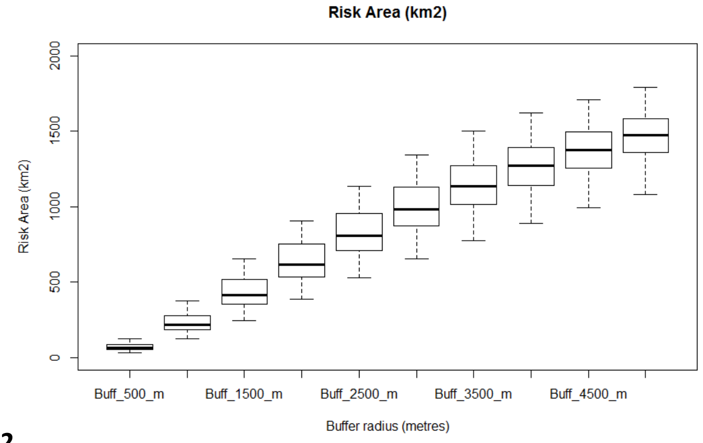
Figure 6.8 below presents multiple box plots of the three performance measures for the dynamic model, namely: Hit Rate, PAI, and SER. The first column relates to Car Key burglary (diagrams A01-A04), and the second to Regular burglary (diagrams B01-B04; grey background). Individual box plots are provided for each of the buffer radius iterations, for example, in diagram A01, the first plot – labelled 'Buff_500_m' – shows the distribution of the hit rates for the 1,458 daily iterations of the model when a 500 metre buffer was created around the past 28 days' offences. As can be seen, the 500 m buffers for Car Key burglary were not a complete failure, with 25 % of the model iterations generating a hit rate > 33.3 %. However, the key message from this section is that prospective buffering appears to be a far more efficient means of capturing future Regular burglary offences than of capturing future Car Key burglary offences.

Dynamic Model – Car Key Burglary

All 4 Years (Census, 2012, 2013, 2014) and incrementing **500 m** buffers



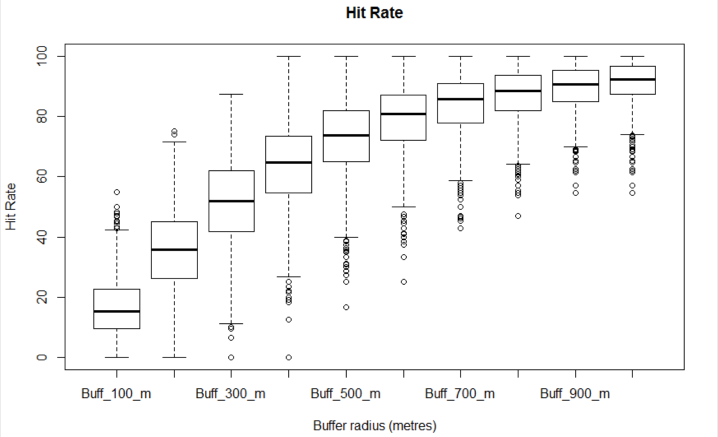
A01



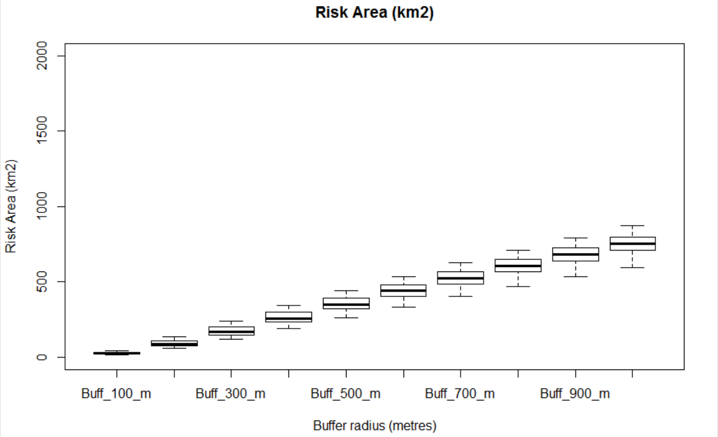
A02

Dynamic Model – Regular Burglary

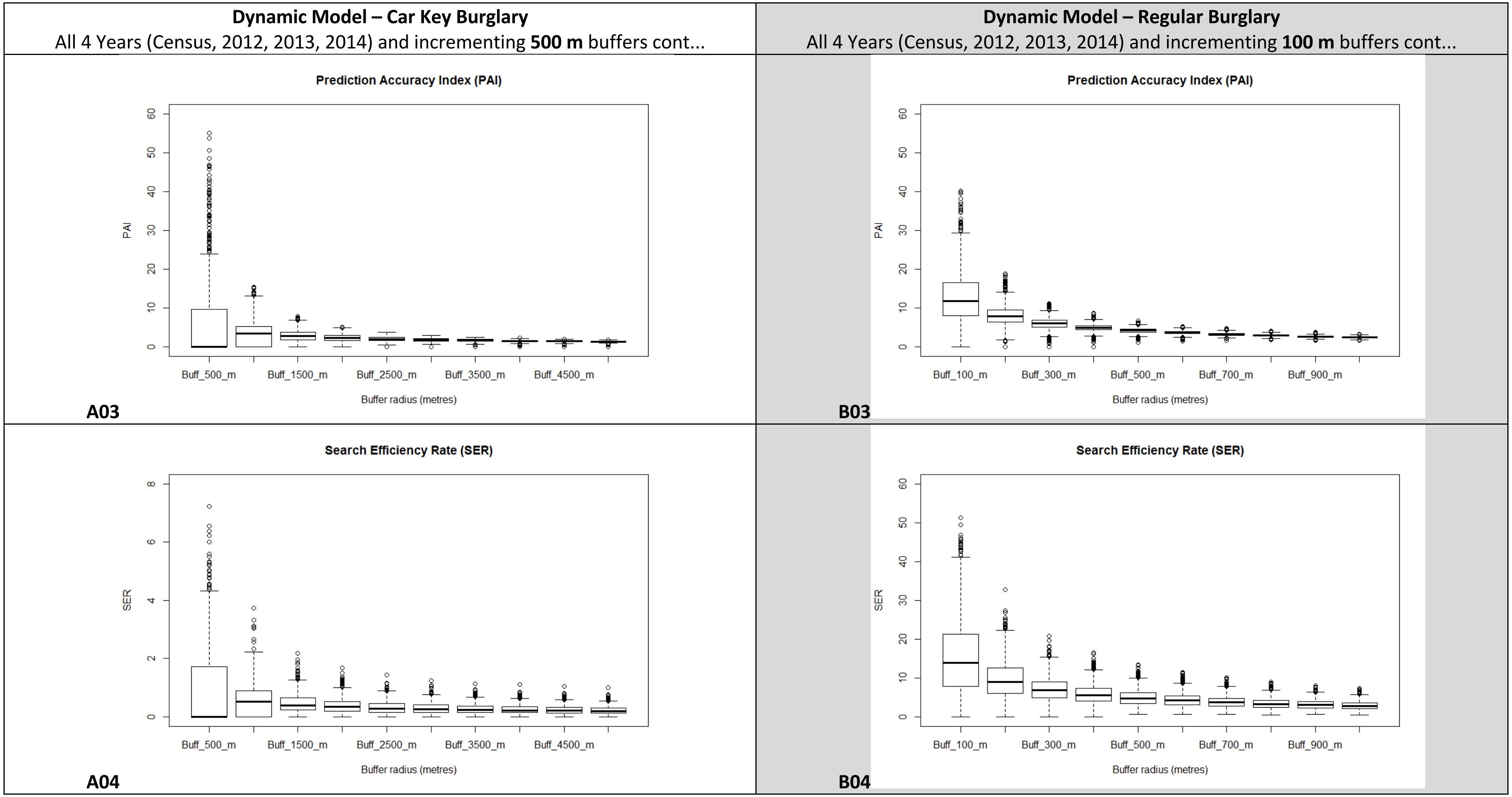
All 4 Years (Census, 2012, 2013, 2014) and incrementing **100 m** buffers



B01



B02



Dynamic Model – Car Key Burglary
All 4 Years (Census, 2012, 2013, 2014) and incrementing **500 m** buffers cont…

Dynamic Model – Regular Burglary
All 4 Years (Census, 2012, 2013, 2014) and incrementing **100 m** buffers cont…

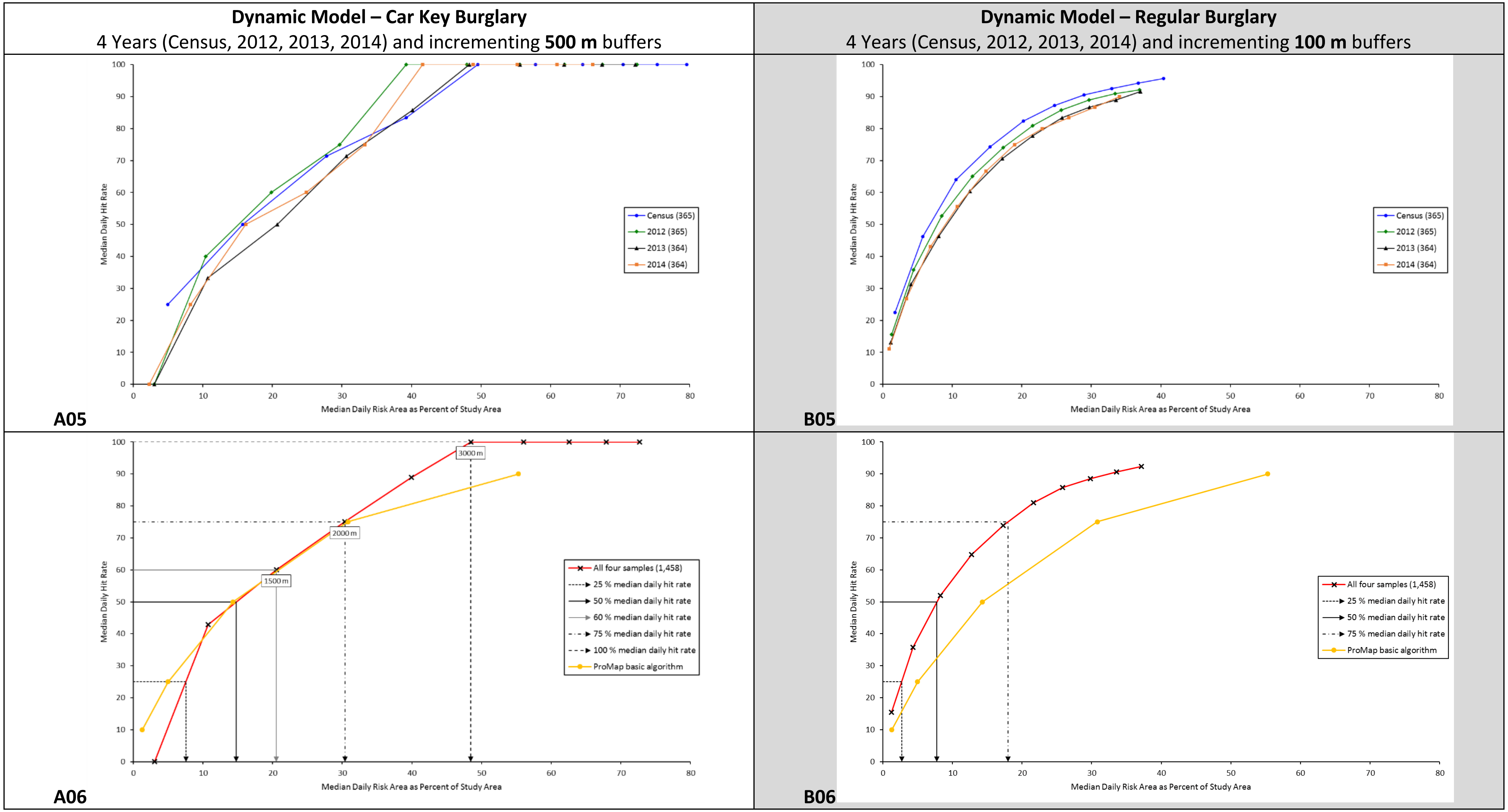
*Figure 6.8 - Box Plots of Hit Rate, Risk Area (km²), PAI, and SER for **Dynamic Model – Car Key Burglary** and **Dynamic Model – Regular Burglary** (grey column)*

The graphs in Figure 6.9 below (A05-A06 and B05-B06) provide a visual representation of the information presented in Table 6.2 (discussed at the beginning of Section 6.3), with each line marker representing a buffer distance iteration (e.g. 500 m, 1000 m, 1500 m..., etc.), each X-value being the associated % WY for that buffer iteration (median daily risk area as per cent of study area), and each Y-value being the median daily hit rate for that buffer iteration. As before, the first column relates to Car Key burglary and the second to Regular burglary (grey background). The first two graphs (A05 and B05) show the dynamic model results broken down by year, with the number of individual predictions shown in brackets, whereas the bottom two diagrams are based on the aggregated results data set. For example, in A06, the median daily hit rate for the 500 m buffer distance is based on 1,458 model iterations (Census year at 500 m, 2012 at 500 m, 2013 at 500 m, and 2014 at 500 m).

Looking at graphs A05 and B05 first, the trends appear to be relatively stable year on year, with similar % WY values generating similar median daily hit rates, particularly for Regular burglary, however, the Census year seemed to perform better than other years when the associated % WY values were comparatively low (note that this year had the highest average number of future offences per day for both crime types based on the model selection criteria; 5.1 for Car Key burglary and 34.8 for Regular burglary). Graphs A06 and B06 are slightly easier to interpret in that all four years are summarised using a single (red) line – drop lines have also been added at certain points to facilitate comparison between the two crime types (if a drop line has a buffer distance label attached then this represents an actual, as opposed to interpolated, value). Note that the 25 %, 50 %, and 75 % median daily hit rate values (indicated on the graphs) are associated with lower median daily risk areas for Regular burglary than for Car Key burglary. As can be seen by referring back to Table 6.2, the 75 % hit rate for Car Key burglary has an associated % WY value of 30 and the 74 % hit rate for Regular burglary has an associated % WY value of 17 (closest comparative, yellow shading). Taking all of this together, it can be inferred that the dynamic model performed much better for Regular Burglary than for Car Key burglary, i.e. higher hit rates observed for smaller risk areas.

It will have been noted that a yellow line is also included on graphs A06 and B06; this shows the median hit rates that were achieved when the basic ProMap algorithm was applied to residential burglaries in Merseyside over 9 separate 1 week test periods (Johnson et al., 2009, p.185, p.188). Note that X-axis values for the ProMap line should be interpreted as ‘per cent of risk ordered grid cells searched’, as opposed to % WY (this is the same principle but a different study area).

Although not directly comparable, the performance of the ProMap algorithm is reassuringly similar to that of the dynamic model developed here for Regular burglary, especially at lower % WY values. Notably, when the ProMap line is overlaid on graph A06, the method appears to have captured more residential burglary offences (proxy Regular burglary offences), on average, relative to the per cent of the study area considered than did the PhD dynamic model for Car Key burglary. Therefore, a study that was conducted completely independently of the current work adds further support to the argument that event-based prospective modelling does not perform as well for Car Key burglary as for Regular burglary.



Dynamic Model – Car Key Burglary
4 Years (Census, 2012, 2013, 2014) and incrementing **500 m** buffers

Dynamic Model – Regular Burglary
4 Years (Census, 2012, 2013, 2014) and incrementing **100 m** buffers

Figure 6.9 - Median Daily Hit Rate vs. Median Daily Risk Area as % of Study Area for ***Dynamic Model – Car Key Burglary*** *and* ***Dynamic Model – Regular Burglary***

6.3.1 Zero Hit Rate Days

Before summarising the dynamic model results section, it is pertinent to mention here that there were 1,578 24 hour prediction periods when the model failed to capture any future Car Key burglary offences (note that these are not 'no offence' days as mentioned earlier). Given that each prediction period spanned two weekdays, e.g. Monday into Tuesday, it was decided, for simplicity's sake, rather than trying to estimate the timing of every non-captured event, to instead bin the data according to the second half of the 24 hour period (00:00-11:59). This meant that all non-captured offences were assumed to have occurred during the hot time period for Car Key burglary, which could have introduced some error into the related analysis. The zero hit rate prediction periods were subsequently graphed according to 'percent of month', 'percent of day of week', and 'percent of month and year' but there were no obvious seasonal or temporal patterns (see Figure 6.10 below). Possible reasons for the zero hit rate days are that: (i) non-captured offences were the first in a new crime series and so there are no previous offences off which to make future predictions; this type of scenario could occur when a prolific Car Key burglary offender is released from prison and they have only just started to re-offend, or (ii) offenders were operating across a wider target area than the prospective buffer distance being considered; the smaller buffer sizes did experience the highest frequency of zero hit rate periods (500 m to 1500 m = 78.1 % of all zero hit rate periods).

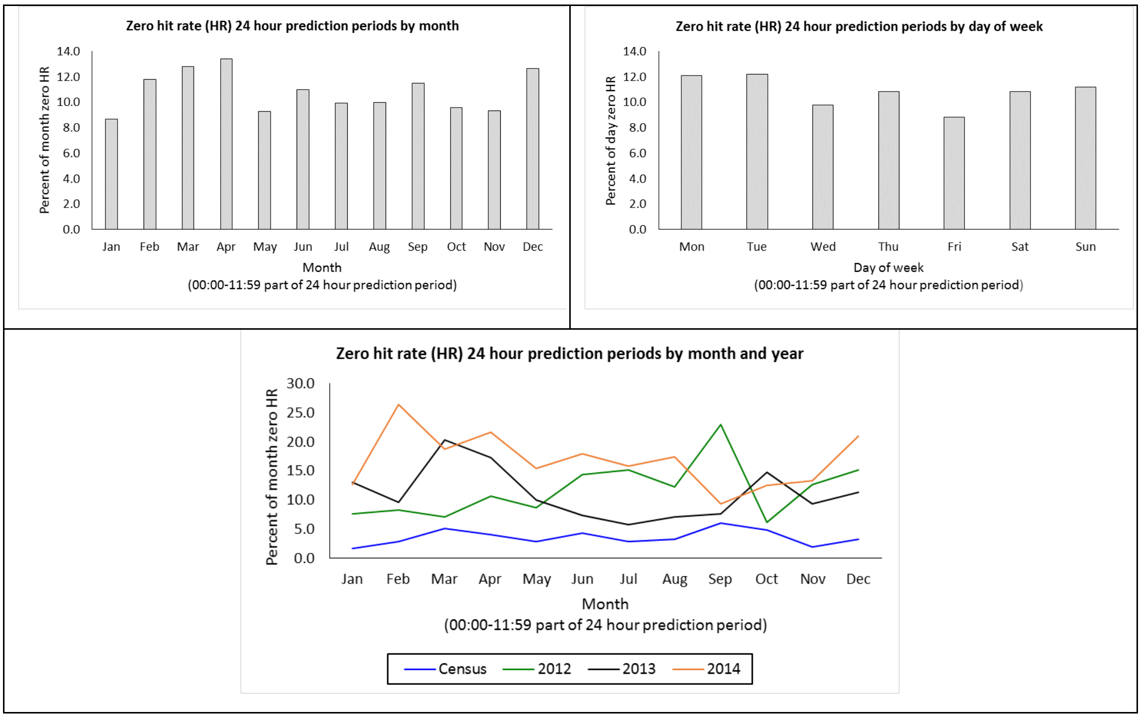


Figure 6.10 - Zero hit rate prediction periods by percent of month, percent of day of week, and percent of month and year

To briefly summarise this Section, the dynamic model did not perform as well for Car Key burglary as for Regular burglary. This result is entirely as expected when viewed in the context of environmental criminological theories, the literature on repeat/ near-repeat victimisation, and the findings of the RV/N-RV that was undertaken in the previous chapter. Although the overall aim of the thesis is to develop a combined risk model for the prediction of temporally clustered offences, this finding is quite exciting from a standalone perspective in that it challenges the prevailing thinking on residential burglars' spatio-temporal offending patterns. Despite Car Key burglars' behaviours being subject to the same moderating factors as Regular burglars, including rational choice (Cornish and Clarke, 1987), it is inferred here that the spatial distribution of the target property type for Car Key burglary relative to offenders' home addresses modifies these to the point where they generate different spatio-temporal crime patterns to Regular burglary. For example, Car Key burglars do not appear to forage in one target area in the same way that Regular burglars do, and they do not generally appear to return to recently targeted properties in the short-term (exact repeats). Taken together, these findings add further support to the argument that all residential burglaries are not created equal, the implication of this being that the spatio-temporal parameters for prospective crime models need to be calibrated relative to the specific crime type, or sub-category of a crime type, that they are predicting for, as opposed to employing a 'one model fits all' approach. The results for the combined model will now be discussed.

6.4 Combined Model Results

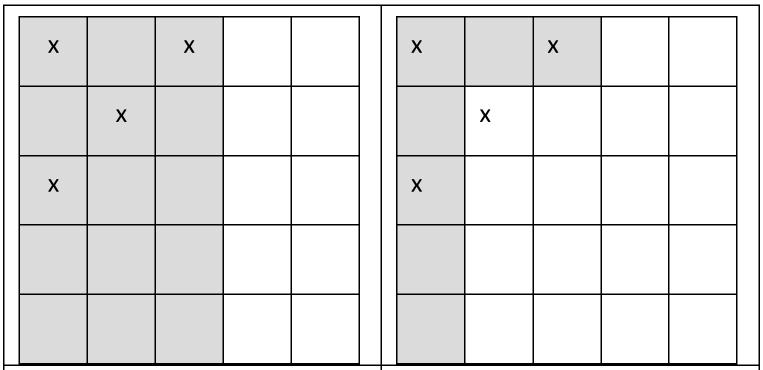
The figures and tables set out below follow the same format as those presented in the dynamic model results section, however, the right hand columns now relate to the combined model results for Car Key burglary (grey background). The PAI and SER scales have been updated on the Car Key burglary dynamic model box plots in Figure 6.12 (discussed later in this Section) to match those for the combined model. Note that the results are directly comparable because the same data sets were used for each model. Recall that the overall aim of the combined model was to reduce the spatial extent of previously identified short-term risk areas whilst maintaining the observed hit rates for each of these. As can be seen from Table 6.3 below, although the median daily hit rates were lower[6] for every buffer distance iteration of the combined model compared to the dynamic model, which initially seems disappointing, the median daily PAI and SER values were higher (also confirmed at 2 DP). Taking the 1500 m buffer distance as an example, although the associated median daily hit rate dropped by 10 percentage points for the combined model (60 % to 50 %), the median daily risk area (km^2) also decreased markedly, resulting in a median daily PAI of 3.2 compared to 2.7 previously (see green shaded row). Similarly, the median daily SER increased from 0.4 to 0.5, i.e. more offences were captured, on average, by the combined model for every km^2 of risk area.

Dynamic Model – Car Key Burglary *(repeated here for comparison)* All 4 Years (Census, 2012, 2013, 2014)						**COMBINED Model – Car Key Burglary** All 4 Years (Census, 2012, 2013, 2014)					
Buffer Distance	**Hit Rate**	**Risk Area (km^2)**	**% WY**	**PAI (1DP)**	**SER (1DP)**	**Buffer Distance**	**Hit Rate**	**Risk Area (km^2)**	**% WY**	**PAI (1DP)**	**SER (1DP)**
500 m	0	63	3	0.0	0.0	**500 m**	0	40	2	0.0	0.0
1000 m	43	218	11	3.4	0.5	**1000 m**	25	135	7	3.7	0.7
1500 m	60	417	21	2.7	0.4	**1500 m**	50	252	12	3.2	0.5
2000 m	75	616	30	2.3	0.3	**2000 m**	50	370	18	2.6	0.4
2500 m	89	810	40	2.0	0.3	**2500 m**	50	475	23	2.3	0.3
3000 m	100	983	48	1.8	0.3	**3000 m**	60	569	28	2.1	0.3
3500 m	100	1137	56	1.6	0.2	**3500 m**	67	649	32	1.9	0.3
4000 m	100	1270	63	1.5	0.2	**4000 m**	67	717	35	1.8	0.3
4500 m	100	1378	68	1.4	0.2	**4500 m**	67	776	38	1.7	0.3
5000 m	100	1475	73	1.3	0.2	**5000 m**	67	829	41	1.6	0.2

*Table 6.3 - Median daily values for **Dynamic Model – Car Key Burglary** and **COMBINED Model – Car Key Burglary** (grey column)*

[6] It was not possible for a hit rate to increase because the same buffers were used for the combined model as for the dynamic model. Also, the median hit rate for the 500 m buffer could not decrease from zero.

Figure 6.11 below is intended to show how a lower hit rate might not necessarily be a bad thing from an operational perspective. For example, viewing the left hand grid first, an area of 15 km^2 would need to be searched to capture 4 offences (100 % hit rate), whereas in the second grid, an area of 7 km^2 would need to be searched to capture 3 offences (75 % hit rate). Although the hit rate is lower for the second grid, this option would probably be preferable when considered relative to the size of the search area, as reflected in the respective PAI values of 1.67 and 2.68. This suggests, that, when making operational decisions based on a prospective crime model, there might need to be some trade-off between likely hit rate and size of risk area.



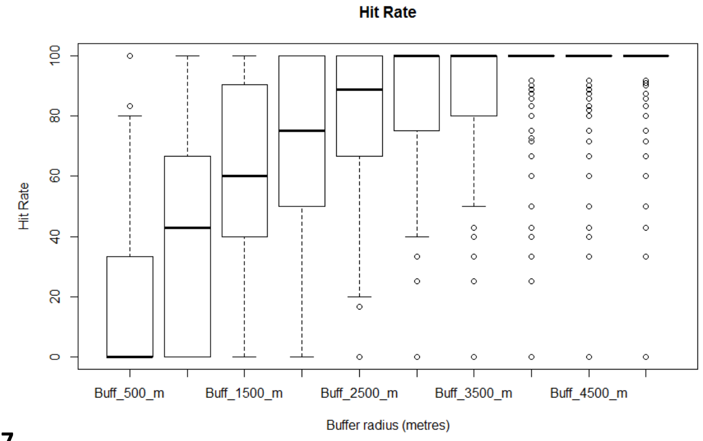
Each cell = 1 km^2	Each cell = 1 km^2
Grey cell = risk area (15 km^2)	Grey cell = risk area (7 km^2)
x = crime	x = crime
Hit rate = 4/4 = 100 %	Hit rate = 3/4 = 75 %
Area percentage = 15/25 = 60 %	Area percentage = 7/25 = 28 %
PAI = 100/60 = 1.67 (2 DP)	**PAI = 75/28 = 2.68 (2 DP)**

Figure 6.11 - Hit Rate vs. PAI illustration

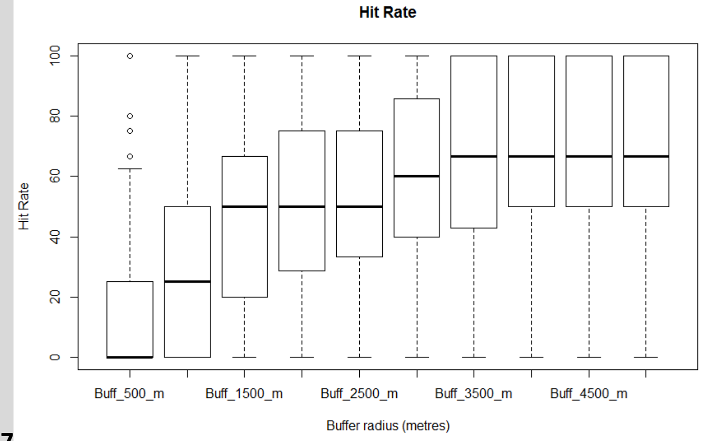
Looking at the individual box plots in diagrams A07, A08, B07, and B08 in Figure 6.12 below, the size of the risk areas generated by the combined model were, on average, smaller than for the dynamic model, but the median daily hit rates for the associated buffer distances were lower. However, when the performance of the combined model is judged solely on changes to the Prediction Accuracy Index (PAI) and Search Efficiency Rate (SER) values (diagrams A09, A10, B09, and B10 in Figure 6.12), the findings are more encouraging. For example, comparing diagrams A09 and B09, and focusing on the 500 m, 1000 m, and 1500 m buffer distances, the whiskers on the combined model plots extend further up the PAI scale than on the dynamic model plots, and there are also more outliers at higher values. This means, that, for some of the combined model iterations, there was an improvement in the hit rate *relative* to the size of the associated risk area, which is a positive in the context of efficient resource deployment. Delving into this a bit deeper, and again taking the 1500 m buffer distance as an example, of the 1191 prediction periods that experienced future offences AND where the dynamic model hit rate was not zero, the combined model maintained the previous hit rate AND improved the associated PAI and SER values in 52 % of these (618 out of 1191), which indicates that the static risk layer successfully disregarded some non-relevant areas in the dynamic risk layers. Considering that the static risk layer was constructed from LSOA level data, and that the associated R^2 value was not overly strong, these results are quite promising in terms of the potential to improve the model's performance in the future, e.g. by utilising predictor variables at more disaggregate geographies.

Dynamic Model – Car Key Burglary ***(repeated here for comparison)***
All 4 Years (Census, 2012, 2013, 2014) and incrementing **500 m** buffers

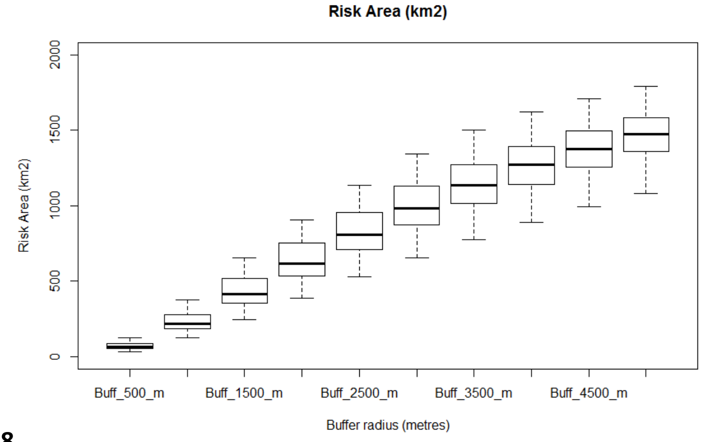
COMBINED Model – Car Key Burglary
All 4 Years (Census, 2012, 2013, 2014) and incrementing **500 m** buffers



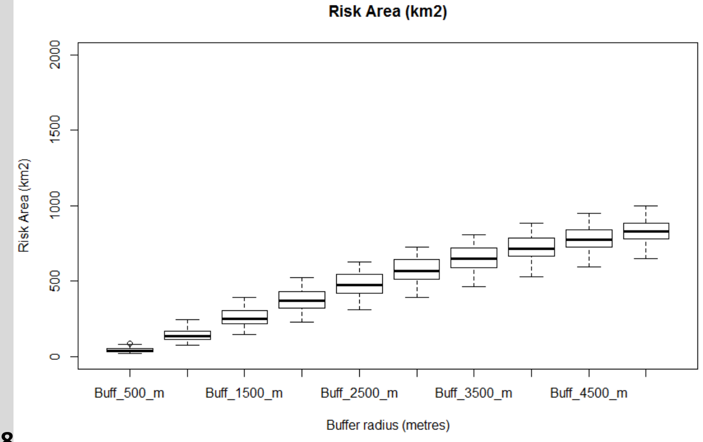
A07



B07



A08



B08

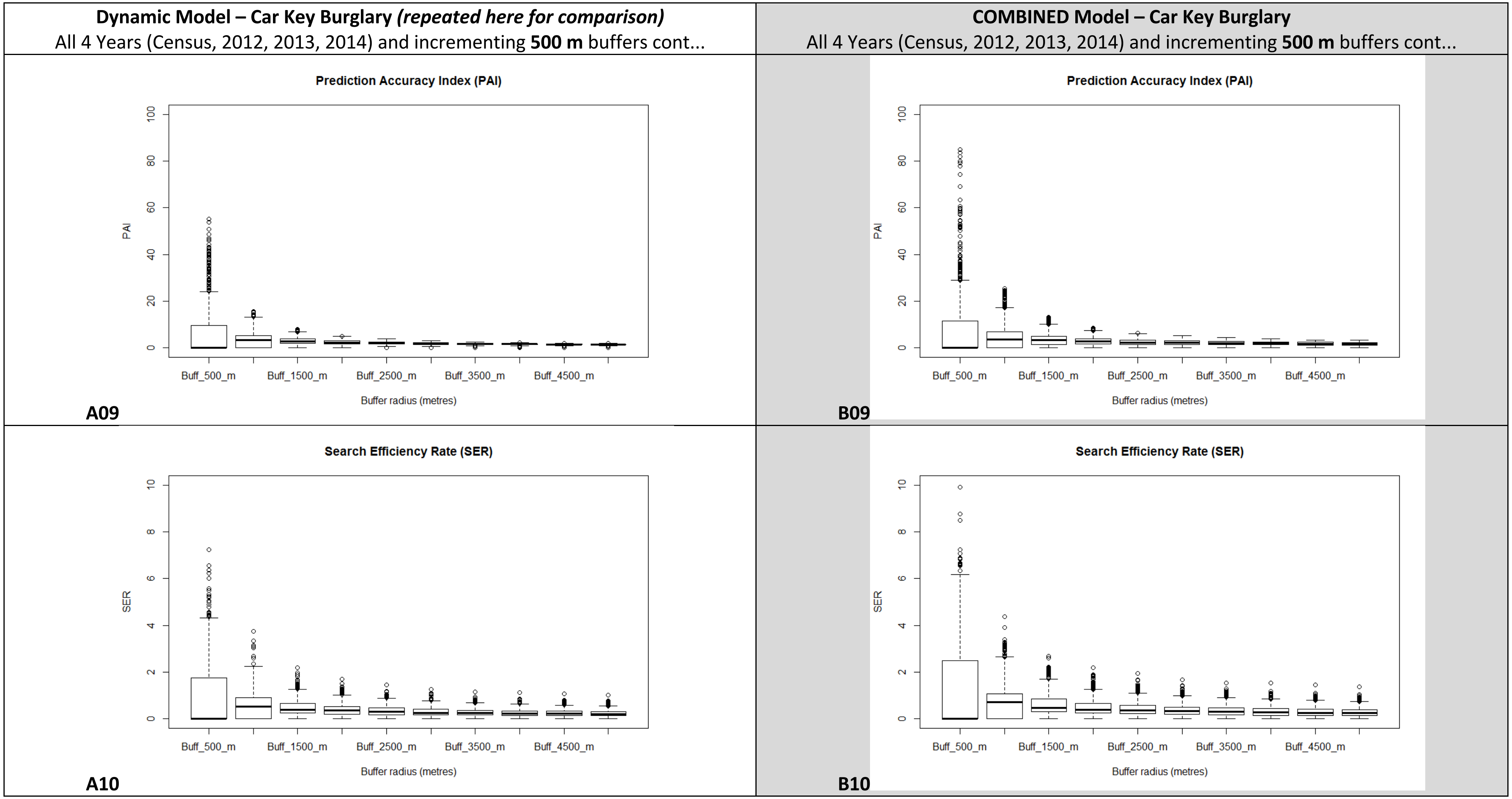
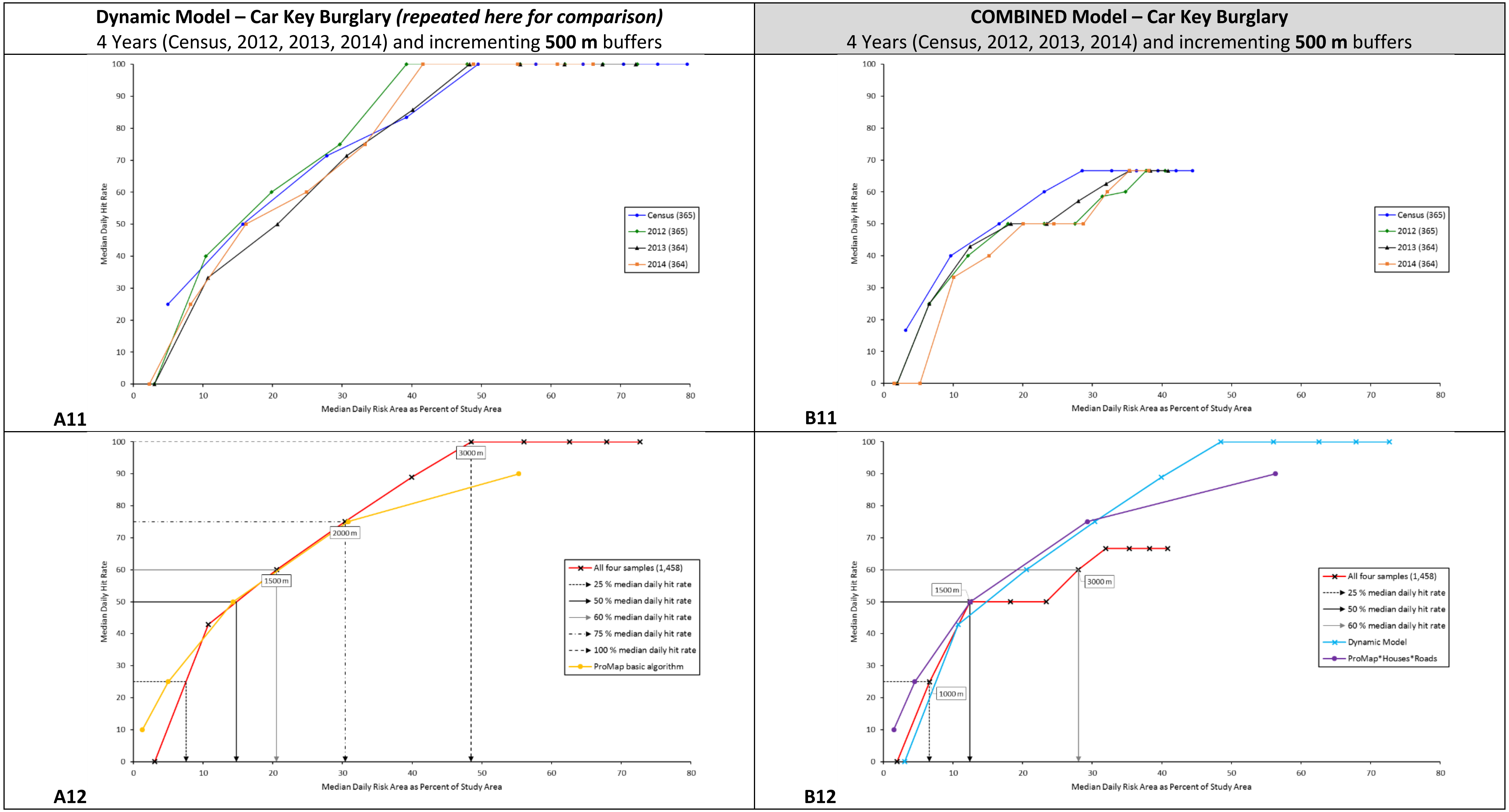


Figure 6.12 - Box Plots of Hit Rate, Risk Area (km²), PAI, and SER for ***Dynamic Model – Car Key Burglary*** *and* ***COMBINED Model – Car Key Burglary*** *(grey column)*

The graphs in Figure 6.13 below (A11-A12 and B11-B12) provide a visual representation of the information in Table 6.3 (discussed at the beginning of Section 6.4), with each line marker representing a buffer distance iteration (e.g. 500 m, 1000 m, 1500 m..., etc.), each X-value being the associated % WY for that buffer iteration (median daily risk area as per cent of study area), and each Y-value being the median daily hit rate for that buffer iteration. On this occasion, the first column relates to the dynamic model results for Car Key burglary and the second to the combined model results for Car Key burglary (grey background). Graphs A11 and A12 (dynamic model) are included again here for comparison purposes. The top right graph (B11) shows the combined model results broken down by year, with the number of individual predictions shown in brackets, whereas the bottom right graph is based on the aggregated results data set. As can be ascertained from graph B11, the trends for the combined model appear to be relatively stable year on year, with similar % WY values generating similar median daily hit rates, however, the Census year performed better than the other three years, possibly because this year was more closely aligned with the Census-derived data in the static risk surface. Note that the combined model failed to generate a median daily hit rate of 100 per cent, unlike the dynamic model, which suggests that the risk heterogeneity surface caused some relevant risk areas to be disregarded, although this is not wholly unexpected given the associated R^2 value.

Looking now at graph B12 (combined model aggregated results), a blue line has been added to remind the reader of the dynamic model results and a purple line has been added to represent the median hit rates that were achieved when the basic ProMap algorithm * Houses layer * Roads layer (combination of 'boost account' and 'flag account') was applied to residential burglaries in Merseyside over 9 separate 1 week test periods (Johnson et al., 2009, p.185, p.188). Recollect that X-axis values for the ProMap line should be interpreted as 'per cent of risk ordered grid cells searched', as opposed to % WY (this is the same principle but a different study area). Although not directly comparable, the performance of the ProMap 'combined' algorithm for residential burglary is not too dissimilar at lower % WY values to that of the current research combined model for Car Key burglary. To illustrate, a 50 per cent median weekly hit rate was achieved by searching 12.5 % of the highest risk ProMap grid cells (Johnson et al., 2009, p.188) and a 50 per cent median daily hit rate was achieved by searching 12 % of the West Yorkshire study area, although it should be noted that the latter figure was derived simply by dividing the median daily risk area (km^2) by the extent of West Yorkshire, i.e. there might be some daily variations on this.

Comparing the lines for the dynamic model (blue) and combined model (red) on graph B12, the combined model does appear to have captured a marginally higher proportion of future Car Key burglary offences at lower % WY values, but not at higher % WY values (based on interpolated values only, e.g. no actual data points at the 25 % and 50 % hit rates for the dynamic model). This might at first appear confusing when considered in relation to the improved PAIs and SERs for each of the individual buffer distance iterations, however, it should be noted that the '% of study area' graph enables size of risk area to be viewed independently of the buffer distance from which it was generated. So, to summarise, although the combined model generated higher median daily PAI and SER values for each of the individual buffer distance iterations – the same buffers were used in both models thus rendering them directly comparable – its performance relative to that of the dynamic model was less cut and dry when size of risk area was considered independently of buffer size.



Dynamic Model – Car Key Burglary *(repeated here for comparison)*
4 Years (Census, 2012, 2013, 2014) and incrementing **500 m** buffers

COMBINED Model – Car Key Burglary
4 Years (Census, 2012, 2013, 2014) and incrementing **500 m** buffers

*Figure 6.13 - Median Daily Hit Rate vs. Median Daily Risk Area as % of Study Area for **Dynamic Model – Car Key Burglary** and **COMBINED Model – Car Key Burglary***

6.5 Conclusions and Further Work

Based on the results presented here, it would seem reasonable to conclude that prospective buffering is not as effective at capturing future Car Key burglary locations as it is at capturing future Regular burglary locations. This is evident from the median hit rates that were observed for Car Key burglary relative to the size of the associated risk areas, versus those for Regular burglary. The use of a risk heterogeneity surface to filter the areal extent of offenders' presumed criminal activity spaces was therefore justified. It is also worth noting that the dynamic risk surfaces generally performed more consistently for Regular burglary than for Car Key burglary, with shorter interquartile ranges observed at the 100 m to 700 m buffer distances (Figure 6.8, diagrams A01 and B01 – Hit Rate). Taken together, this indicates that the parameters of prevailing crime modelling techniques should be calibrated relative to the characteristics of the crime type that they are seeking to predict, and that consideration should also be given to the disaggregation of official HO crime classifications.

Although the combined model did not perform as well as previously hoped, it is likely that a better specified risk heterogeneity surface would go some way to addressing this, which in turn presents an opportunity for future work, including undertaking analysis with more spatially disaggregate predictor variables. Despite the limited performance of the combined model, the median PAI values for all but the 500 m buffer distance indicate that, on average, the model still outperformed a risk area equivalent to 100 % of the study area, especially when 1000 m and 1500 m buffers were used (see Table 6.3).

It is also important to mention here some potential limitations of the methods that were employed in this chapter, including that chance expectation was not factored into the analysis, namely, did any of the models perform better than would be expected on the basis of chance? However, considering the sheer number of model iterations that were run, and different time periods examined, this might have gone some way to mitigating the issue. For example, Chainey (2014, p.100) states that researchers have illustrated *"how chance expectation can be minimised by using the mean PAI results from a large number of experiments across different temporal input and temporal output data periods, and by observing the variation in the standard deviation of the PAI generated from these many experiments."* Note that median PAI values and box plots were used in place of mean and standard deviation in the current research. Another limitation of the analysis is that no consideration was given to how the individual risk areas were distributed across the study area, which could have implications from an operational resourcing perspective, i.e. if the overall risk area (km^2) per model iteration was comprised of many disparate risk areas. Further, the prospective buffering approach that was employed was very simplistic and, as such, could potentially be refined through the recognition of overlapping risk areas (recall that individual buffers were dissolved to a single dynamic risk layer), and it might also be useful to consider using doughnut shaped buffers as a means of representing the assumed 'safe' area around recent Car Key burglary offences. Finally, and at a more general level, the crime data on which the analysis was based might be improved, for example, through the identification of likely attempt Car Key burglary offences; this particular issue will be discussed at length in the next chapter (using Natural Language Processing (NLP), specifically text classification, to label offences automatically).

Generally speaking, the results warrant further investigation, including for other crime types, such as business robbery and distraction burglary. That is, do the associated RV/ N-RV patterns (if shown to exist) reflect locally anchored offending, and can these be policed using conventional crime modelling techniques? Some potential areas for future development have already been suggested, however, an additional one to consider is the identification of risk at much smaller spatial scales, possibly even at the street network level, which would build on the work of Rosser et al. (2017), as well as complementing the inherently mobile nature of Car Key burglary offending.

Chapter 7 Using Machine Learning in an End-to-End Burglary Model

7.1 Introduction

Given the difficulties inherent to manually deriving the two Burglary Dwelling data sets (Car Key burglary and Regular burglary), as discussed in Chapter 3, this chapter is intended to show how the process might be improved by using Natural Language Processing (NLP), specifically text classification, to label offences automatically. However, due to the text analysis methods proposed here still being under development, they were **not** ultimately used to create the burglary data sets for the current research but, nevertheless, they provide a valuable indication of important future work as a means of more accurately classifying some crime types. Therefore, what follows is **largely theoretical**, although its application beyond the thesis could benefit future analysis, with NLP being particularly conducive to research replicability, as well as potentially reducing the need for skilled resources, such as police analysts, to undertake manual crime classification work in order to identify relevant data for analysis.

The chapter begins with an overview of Natural Language Processing and text classification, after which Multivariate Bernoulli Naïve Bayes classification is used, together with a dummy training data set, to classify a hypothetical Burglary Dwelling MO as being either a Car Key burglary or a Regular burglary. A proposal is then outlined for a case linkage model that incorporates text classification, the purpose being to identify unique crime series within a pre-labelled Car Key burglary data set. It is anticipated that outputs from the model would be useful for both the 'TIC' process (when an offender agrees to have additional offences 'taken into consideration') and the estimation of future crime risk. The chapter ends with a suggestion as to how an alternative classification method, Latent Dirichlet Allocation (LDA), could be used in conjunction with police incident logs (calls for service) to inform strategic planning and resource deployment.

7.2 Incorrect Crime Classification

A possible future development of the manual sample selection method that was employed in the current research would be to explore the capacity of a machine learning algorithm, such as Naïve Bayes, to automatically differentiate between Car Key burglary and Regular burglary offences within a general Burglary Dwelling data set (recall that there is currently no separate Home Office classification for a Car Key burglary). This approach is expected to be particularly useful in terms of identifying 'attempt' Car Key burglary offences, i.e. those in which a vehicle was the most likely target property type, as indicated by the MO, but where the offence was not completed (vehicle not stolen), for example, because the offenders were disturbed, or they could not locate the vehicle's keys. Since the current research sample selection method relied on the 'Hanoi' keyword being present in at least one of two searched data input fields per crime occurrence, namely 'MO' and 'Crime Notes', and/ or an offence having as associated stolen vehicle record, it is unlikely that all, if any, attempt Car Key burglaries will have been correctly recorded as such. The risk of misclassifying an attempt Car Key burglary as a Regular burglary is a key limitation of the present work and the majority of this chapter will therefore be used to illustrate how text classification might go some way to addressing this issue.

Having the means to automatically classify offences also has potential real-world application beyond the confines of the current research. It is imperative from a policing perspective that crimes are correctly classified. For example, the omission of an attempt Car Key burglary offence from an investigative data set could mean that vital evidence is overlooked for a related crime series, as well as perhaps masking the full extent of a crime problem, e.g. hot spot size. A non-manual crime classification technique is also predicted to be helpful in practical policing situations where large volumes of historical crime data need to be interrogated, such as when an offender agrees to have additional offences 'taken into consideration' (TICs), or an analyst has to report on longer-term crime trends. Rather than having to manually assess hundreds, or even thousands, of crime records, text classification could be used instead to extract relevant offences. Although not a panacea for all, bearing in mind that computerised crime MO fields are typically free text entry, thus rendering them vulnerable to spelling mistakes and the like, text classification, both as an analytical support tool and time-saving device, undoubtedly warrants further exploration.

7.2.1 Natural Language Processing

Natural Language Processing (NLP) is a sub-field of artificial intelligence (AI) in which computers are used to manipulate and/ or extract meaning from 'natural' languages, that is, the unstructured languages by which humans communicate on a day-to-day basis as opposed to 'artificial' languages, such as mathematical notation and those used in software programming (Perlman, 1984; Bird et al., 2009; Rogerson, 2016; Geitgey, 2018). There are numerous examples of NLP applied in everyday settings, including, but not limited to: predictive typing, internet searches, text translation, sentiment analysis (Bird et al., 2009, ix), and classifying documents by topic type (Bird et al., 2009, pp.221-222). However, it is the latter of these that is most relevant in the current research context because the purpose of a crime classification is to assign a summary label to a particular set of offence characteristics, thus facilitating greater control over, and understanding of, associated information.

Police-recorded MOs can be thought of as relatively short passages of natural language, albeit a localised subset, in that they contain unstructured text and, as noted by Rogerson (2016, p.128), have an average length of 40 to 150 words. As is often the case within organisations, the police tend to communicate using a very context-specific terminology, or 'local grammar', for example, the word 'Hanoi' in a West Yorkshire Police Burglary Dwelling MO is, on the balance of probability, much more likely to refer to a Car Key burglary offence than to the capital city of Vietnam. This could actually be beneficial from the point of view of delineating crime types, particularly if some local grammar is specific only to a single crime classification, or subclass of.

7.2.2 Text Classification

Text classification is the process of assigning a class label to a given input, for example, classifying a movie review as being positive or negative, identifying the main theme of a news report, or deciding if an email is spam or non-spam (Bird et al., 2009, pp.221-222). The approach can be either 'supervised' or 'unsupervised'. The former requires a pre-labelled training corpora[7] to be passed to a machine learning algorithm, e.g. a set of MO text files and associated crime categories, whereas the latter does not involve any pre-labelling, instead searching for clusters in the data (for detailed explanations see: Manning et al., 2009). Given that crime categories are already known for the current research, i.e. 'Car Key' burglary and 'Regular' burglary, a

[7] In this chapter, 'corpus' will be used to describe an individual crime record, and 'corpora' a set of records.

supervised classification method, namely 'Multivariate Bernoulli Naïve Bayes', will be used to allocate the following hypothetical Burglary Dwelling MO to one of these classes:

> Makeshift device used to hook car keys through letterbox. Suspects believed disturbed as new Mercedes still parked on driveway and nothing else stolen.

It is worth noting here that had the above MO, which appears to describe an attempt Car Key burglary, featured in the PhD Burglary Dwelling data set then it would have been classified as a 'Regular' burglary because a vehicle was not stolen and the 'Hanoi' keyword is not present (assumption made for the purposes of illustration that the 'Hanoi' keyword is also absent from the Crime Notes section of the crime record). Potential errors such as this could negatively impact the interpretation of associated results and should thus be afforded due consideration. Given the relatively large size of the current research burglary data samples, it is likely that any prevailing trends will have been captured, even with a level of misclassification, however, the issue is still something to be aware of, and to seek to mitigate wherever possible, hence this thesis chapter.

7.3 Multivariate Bernoulli Naïve Bayes Classification

Multivariate Bernoulli Naïve Bayes was chosen over Multinomial classification because, recalling the average length of police-recorded MOs is approximately 40 to 150 words, the former is considered to perform well for short documents (Manning et al., 2009, p.268). Both approaches represent documents as a 'bag of words', commonly referred to as the bag-of-words model. A bag-of-words vocabulary is constructed by identifying unique words within a set of training documents, these usually having first undergone some form of pre-processing to normalise the text and remove uninformative content, such as punctuation marks and stop words, e.g. '?', ':', 'and', 'or', 'the' (discussed in more detail later). The key difference between a Multinomial and a Bernoulli classification is that Multinomial classification counts the number of times that individual words within a pre-defined vocabulary (|V|) appear in a classification document, whereas the Bernoulli document model only considers the presence or absence of each word (Shimodaira, 2015).

Adding further support to bag-of-words document representation, as a method for identifying possible attempt Car Key burglaries, is the following extract from a piece of 'Burglary Residential' analysis that was undertaken for the West Yorkshire PCC Community Outcomes Meeting April 2018: "**Euro Profile** breaches are still prevalent and often associated with **2 in 1 burglaries** or **likely attempts**" (Abbott-Smith, 2018, p.5, my emphasis). This indicates that anyone seeking to identify attempt Car Key burglaries, in the absence of the 'Hanoi' keyword and a stolen vehicle, should look at burglary offences where a Euro Profile lock[8] was targeted. The inherent difficulty here, however, is by what means to differentiate attempt Car Key burglaries where a Euro Profile lock was targeted from Regular burglaries where a Euro Profile lock was targeted, thus the bag-of-words model is anticipated to be particularly relevant because it captures those words that occur frequently in documents of the same class. For example, in an attempt Car Key burglary MO, we might expect to find the words 'Euro' and 'Profile' presenting concurrently with words that are also prevalent in known Car Key burglary MOs, but not in known Regular burglary MOs. Although this idea would need to be tested empirically, it does offer a potential avenue for further research, including the development of a 'bag-of-bigrams'/ 'trigrams' model, i.e. word collocations, such as 'car keys', 'complainants asleep', and 'tidy search' (recall Shaw et al., 2010).

7.3.1 Overview of Naïve Bayes Classification

Multivariate Bernoulli Naïve Bayes classification uses Bayes' Theorem, as described in Equation 7.1 below, to determine the most appropriate class label for a document (Shimodaira, 2015). Formulated by mathematician Thomas Bayes in the 18th century, the theorem calculates the posterior probability that an event will occur based on the knowledge that a second event has already occurred (Cochrane, 2016). The 'naïve' element reflects the fact that input features, e.g. words, and binary representations of words, are considered independently of one another, i.e. naïve Bayes' assumption (Krishnaveni and Sudha, 2017, p.290; Manning et al., 2009; Shimodaira, 2015).

$$\overbrace{P(A|B)}^{\substack{posterior\\ probability}} = \frac{\overbrace{P(B|A)}^{\substack{conditional\\ probability}} \quad \overbrace{P(A)}^{\substack{prior\\ probability}}}{\underbrace{P(B)}_{evidence}} \tag{7.1}$$

To use an example from Cochrane (2016, pp.169-170), if an unweighted die is rolled once and we are told just that the outcome is an even number, then we can employ Bayes' Theorem to

[8] Used on many UPVC doors and vulnerable to 'snapping', i.e. cylinder compromised to gain entry.

calculate the probability that the rolled number is a ‘6’. Crucially, without any prior information about the outcome, we can only say that there is a one in six chance that the number is a ‘6’, however, because we know that the rolled number is even, we can establish a probability of 1/3, as shown in Equation 7.2 below. Relating this example to crime type classification, we can think of an MO as an outcome event from which the most appropriate class label can be determined.

$$P(six|even) = \frac{P(even|six)P(six)}{P(even)} = \frac{1 * \frac{1}{6}}{\frac{1}{2}} = \frac{1}{3} \qquad (7.2)$$

Adapted from Cochrane, 2016, pp.169-170.

As per Equation 7.3 below, the ‘A’ and ‘B’ terms in the classic Naïve Bayes’ formula are often replaced with ‘C’ and ‘D’ in text classification, representing ‘class’ and ‘document’ respectively (e.g. see: Manning et al., 2009, p.265; Shimodaira, 2015, p.1); this notation will be used in subsequent examples. Note that the denominator, or ‘evidence’ term, is dropped in Equation 7.3 because it is a constant across all classes (Manning et al., 2009, p.265; Pan et al., 2004, p.82).

$$P(C|D) = \frac{P(D|C)P(C)}{P(D)} \propto P(D|C)P(C) \qquad (7.3)$$

The method by which the hypothetical Burglary Dwelling MO will be classified is expressed as per the Naïve Bayes’ formula, (minus $P(D)$), in Equations 7.4 and 7.5 below, i.e. the probability that the MO relates to: (i) a Car Key burglary, and (ii) a Regular burglary. Whichever of these two equations produces the highest posterior probability value will determine the MO classification.

$$P(Car\ Key\ burglary|MO) = P(MO|Car\ Key\ burglary)P(Car\ Key\ burglary) \qquad (7.4)$$

$$P(Regular\ burglary|MO) = P(MO|Regular\ burglary)P(Regular\ burglary) \qquad (7.5)$$

7.3.2 Supervised Classification Process

Figure 7.1 below shows the main steps in the first stage of the supervised classification process, namely training a classifier with pre-labelled corpora. An example of pre-labelled corpora is police-recorded MOs grouped by known crime type, e.g. ‘Regular’ burglary MOs. Figure 7.2

below depicts the second stage of the process, that is, classifying unlabelled corpora with a classifier model. Grey shading denotes that a step is the same at both stages of the process, whereas red indicates that it is different. Each stage will be outlined in detail below. These process maps will be used to classify the hypothetical Car Key burglary MO.

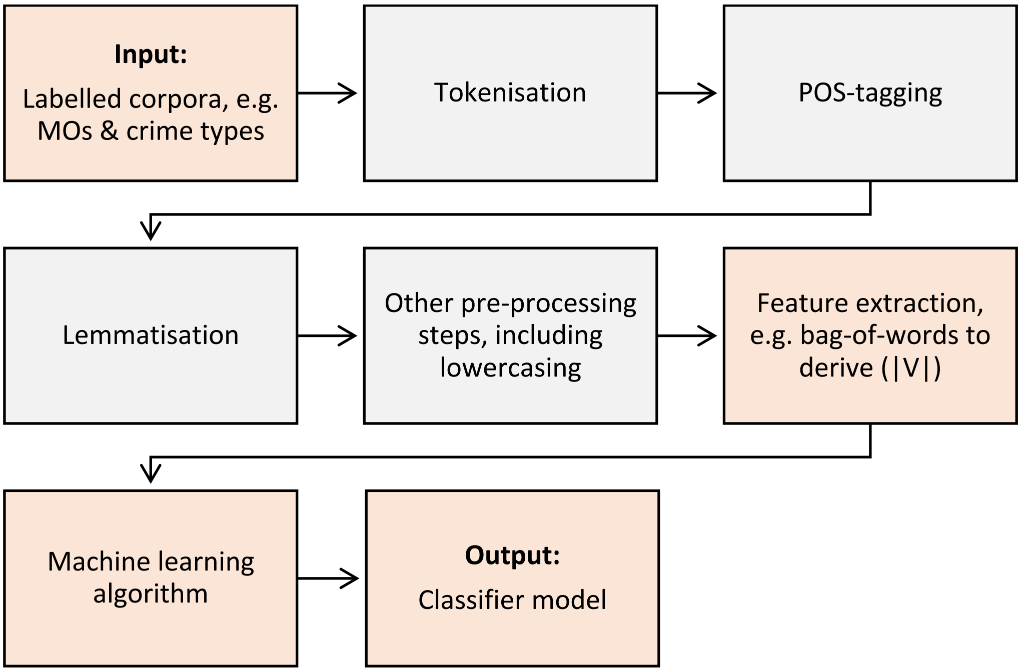

Input:
Labelled corpora, e.g. MOs & crime types
Tokenisation
POS-tagging
Lemmatisation
Other pre-processing steps, including lowercasing
Feature extraction, e.g. bag-of-words to derive (|V|)
Machine learning algorithm
Output:
Classifier model

Figure 7.1 - Process map for training a classifier with pre-labelled corpora

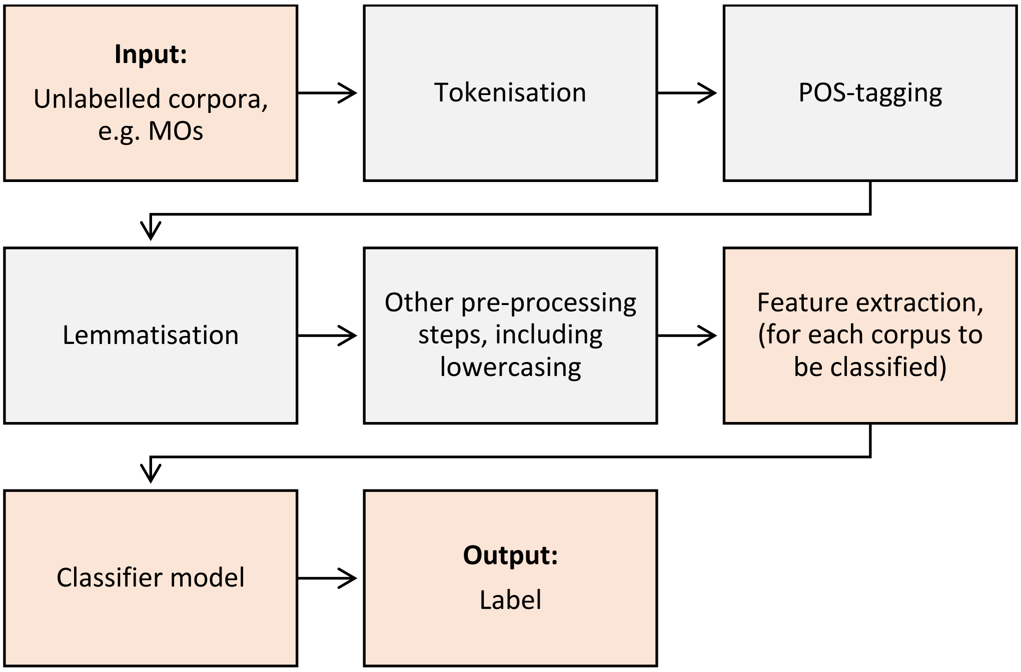

Input:
Unlabelled corpora, e.g. MOs
Tokenisation
POS-tagging
Lemmatisation
Other pre-processing steps, including lowercasing
Feature extraction, (for each corpus to be classified)
Classifier model
Output:
Label

Figure 7.2 - Process map for classifying unlabelled corpora with a classifier model

7.3.3 Pre-Processing Techniques

As mentioned previously, it is usual for documents to undergo some form of pre-processing prior to classification. The most fundamental pre-processing operation is tokenisation, whereby paragraphs and sentences are broken down into their constituent parts to create machine readable 'tokens'. E.g. with word tokenisation, the phrase '*The cat sat on the mat.*' becomes '[The], [cat], [sat], [on], [the], [mat], [.]'. Tokens can be used as an input into a Natural Language Processing algorithm in their own right, and also subjected to other pre-processing methods, including lemmatisation and POS-tagging. With the exception of tokenisation, pre-processing is usually carried out to improve the accuracy of classification results. Six commonly used pre-processing techniques are described in Table 7.1 below.

Pre-processing step	**Description**
Tokenisation (tokenization)	Breaking down a body of text into its constituent parts, e.g. paragraphs to sentences, or sentences to words. The resulting elements, known as 'tokens', can then be used as inputs in other NLP tasks.
POS-tagging	Assigning an appropriate part-of-speech tag, e.g. 'adjective', 'noun', 'verb', to individual tokens in a sentence based on the context in which they appear. This is an important pre-classification step in terms of being able to accurately infer the meaning of a document (see also lemmatisation).
Stemming	Reducing a word to its base form (stem), usually by removing the suffix. This is a fairly crude approach, e.g. 'caring' becomes 'car' with stemming, as opposed to 'care' with lemmatisation. Further, the method does not always produce a recognisable word.
Lemmatisation (lemmatization)	Identifying the base form (lemma) of a word, for example, 'stealing', 'stole', and 'stolen' all have the word 'steal' as their root. Due to the existence of homographs (words that are spelled the same but have different meanings), it is advisable to undertake POS-tagging prior to lemmatisation. Taking the word 'saw', for example, this has at least two lemmas, one relating to the act of seeing (POS = 'verb', lemma = 'see'), and another describing a tool for cutting wood (POS = 'noun',

	lemma = ‘saw’). The main benefit of lemmatisation over stemming is that the original meaning of a word is retained. Two worked examples of POS-tagging and lemmatisation are provided in Table 5.2 below.
Lowercasing	Making all text lowercase, e.g. words at the beginning of sentences. This removes any distinction between upper and lower case versions of the same text, thus preventing them from being analysed separately.
Deletion	Removing punctuation and stop words, e.g. ‘?’, ‘.’, ‘:’, ‘!’, ‘a’, ‘and’, ‘is’, ‘the’, ‘to’, ‘with’. Disregarding tokens that add little/ no value to the interpretation of text reduces noise and computer processing time.

Table 7.1 - Commonly used pre-processing techniques (e.g. see Bird et al., 2009; Manning et al., 2009; Rogerson, 2016)

Undertaking POS-Tagging Prior to Lemmatisation

POS-tagging and lemmatisation are used in two different contexts in Table 7.2 below to highlight the importance of undertaking ‘part of speech’ analysis prior to lemmatisation. In the first text, the word ‘jogging’ is a noun, but in the second it is a verb. Both texts were tagged using NLTK’s standard tagger for English, which in turn uses the ‘Penn Treebank’ tagset (Bird et al., 2009) (see Table 7.3 below for sample extract), and then lemmatised using NLTK’s ‘WordNetLemmatizer’. The pertinent point to take away from this example is that, had the second text been passed to the lemmatiser without first being tagged, then the word ‘jogging’ would have been lemmatised as ‘jogging’, instead of the correct base form, ‘jog’; this is because the default POS for WordNetLemmatizer is ‘noun’. Thus, to ensure that the original context and meaning of words is preserved, it is advisable to undertake POS-tagging prior to lemmatisation.

Text	**POS-tagged**	**Lemmatised**
Experts think that jogging is a good form of exercise	[('Experts', 'NNS'), ('think', 'VBP'), ('that', 'IN'), **('jogging', 'NN')**, ('is', 'VBZ'), ('a', 'DT'), ('good', 'JJ'), ('form', 'NN'), ('of', 'IN'), ('exercise', 'NN')]	Experts think that **jogging** be a good form of exercise

The witness was jogging around the park with a friend	[('The', 'DT'), ('witness', 'NN'), ('was', 'VBD'), **('jogging', 'VBG')**, ('around', 'IN'), ('the', 'DT'), ('park', 'NN'), ('with', 'IN'), ('a', 'DT'), ('friend', 'NN')]	The witness be **jog** around the park with a friend

Table 7.2 - Examples of POS-tagging and lemmatisation

Tag	**Description**
CC	Coordinating conjunction
IN	Preposition or subordinating conjunction
JJ	Adjective
NN	Noun, singular or mass
NNP	Proper noun, singular
NNS	Noun, plural
RB	Adverb
TO	*to*
VB	Verb, base form
VBD	Verb, past tense
VBG	Verb, gerund or present participle
VBZ	Verb, 3rd person singular present

Table 7.3 - Sample extract of POS-tags from the Penn Treebank tagset

(Source: https://www.ling.upenn.edu/courses/Fall_2003/ling001/penn_treebank_pos.html)

7.4 Classifying the Hypothetical Burglary Dwelling MO

Multivariate Bernoulli Naïve Bayes will now be used to classify the hypothetical Burglary Dwelling MO. Please note that, with the exception of the pre-processing steps in Table 7.4 below, the approach outlined in this section closely follows that presented in Shimodaira, 2015.

7.4.1 Pre-Processing

Table 7.4 below shows the output of the pre-processing techniques that were performed on the hypothetical Burglary Dwelling MO; these were also used on the dummy training data MOs, as per Table 7.5 and Table 7.6 below.

<table>
<tr><th>Pre-processing step</th><th colspan="2">NLP output</th></tr>
<tr><td>Tokenisation (word)</td><td colspan="2">['Makeshift', 'device', 'used', 'to', 'hook', 'car', 'keys', 'through', 'letterbox', '.', 'Suspects', 'believed', 'disturbed', 'as', 'new', 'Mercedes', 'still', 'parked', 'on', 'driveway', 'and', 'nothing', 'else', 'stolen', '.']</td></tr>
<tr><td>POS-tagging</td><td colspan="2">[('Makeshift', 'NNP'), ('device', 'NN'), ('used', 'VBN'), ('to', 'TO'), ('hook', 'VB'), ('car', 'NN'), ('keys', 'NNS'), ('through', 'IN'), ('letterbox', 'NN'), ('.', '.'), ('Suspects', 'VBZ'), ('believed', 'VBN'), ('disturbed', 'NNS'), ('as', 'IN'), ('new', 'JJ'), ('Mercedes', 'NNP'), ('still', 'RB'), ('parked', 'VBD'), ('on', 'IN'), ('driveway', 'NN'), ('and', 'CC'), ('nothing', 'NN'), ('else', 'RB'), ('stolen', 'VBN'), ('.', '.')]</td></tr>
<tr><td rowspan="20">Lemmatisation</td><td><b>Token*</b></td><td><b>Lemma*</b></td></tr>
<tr><td>Makeshift</td><td>Makeshift</td></tr>
<tr><td>device</td><td>device</td></tr>
<tr><td>used</td><td>use</td></tr>
<tr><td>hook</td><td>hook</td></tr>
<tr><td>car</td><td>car</td></tr>
<tr><td>keys</td><td>key</td></tr>
<tr><td>letterbox</td><td>letterbox</td></tr>
<tr><td>Suspects</td><td>Suspects</td></tr>
<tr><td>believed</td><td>believe</td></tr>
<tr><td>disturbed</td><td>disturbed</td></tr>
<tr><td>new</td><td>new</td></tr>
<tr><td>Mercedes</td><td>Mercedes</td></tr>
<tr><td>still</td><td>still</td></tr>
<tr><td>parked</td><td>park</td></tr>
<tr><td>driveway</td><td>driveway</td></tr>
<tr><td>nothing</td><td>nothing</td></tr>
<tr><td>else</td><td>else</td></tr>
<tr><td>stolen</td><td>steal</td></tr>
<tr><td colspan="2">*Punctuation and stop words removed for illustration purposes. Red text indicates a different output.</td></tr>
<tr><td>Lowercasing and deletion (punctuation & stop words)</td><td colspan="2">['makeshift', 'device', 'use', 'hook', 'car', 'key', 'letterbox', 'suspects', 'believe', 'disturbed', 'new', 'mercedes', 'still', 'park', 'driveway', 'nothing', 'else', 'steal']</td></tr>
</table>

Table 7.4 - Output of pre-processing performed on hypothetical Burglary Dwelling MO

It is important to note at this point that POS-tagging is not always accurate, for example, the word ‘Suspects’ in the hypothetical MO has been tagged as ‘VBZ’ (Verb, 3rd person singular present), as opposed to a plural noun, i.e. describing persons suspected of committing an offence. Therefore, in situations where numerous crime MOs need to be classified, it might be expedient to develop a police-specific tagger.

Table 7.5 below contains eight dummy MOs that were constructed by the author for use in this case study example; half are intended to reflect typical Car Key burglary MOs, and half Regular burglary MOs. Table 7.6 below shows the MOs after pre-processing.

URN	Content	Class
MO1	Hook and cane implement used to steal car key	Car Key burglary
MO2	Vehicle removed from driveway by two suspects	Car Key burglary
MO3	Keys hooked through letterbox and BMW stolen	Car Key burglary
MO4	VW Golf stolen with keys obtained via burglary	Car Key burglary
MO5	Male entered insecure front door and removed TV	Regular burglary
MO6	Window smashed and laptop stolen from study	Regular burglary
MO7	Front door jemmied and untidy search of lounge	Regular burglary
MO8	Handbag containing car keys taken from hall table	Regular burglary

Table 7.5 - Dummy training data set of unprocessed and labelled MOs

URN	Content	Class
MO1	['hook', 'cane', 'implement', 'use', 'steal', 'car', 'key']	Car Key burglary
MO2	['vehicle', 'remove', 'driveway', 'two', 'suspect']	Car Key burglary
MO3	['keys', 'hook', 'letterbox', 'bmw', 'steal']	Car Key burglary
MO4	['vw', 'golf', 'steal', 'key', 'obtain', 'via', 'burglary']	Car Key burglary
MO5	['male', 'enter', 'insecure', 'front', 'door', 'remove', 'tv']	Regular burglary
MO6	['window', 'smash', 'laptop', 'stolen', 'study']	Regular burglary
MO7	['front', 'door', 'jemmied', 'untidy', 'search', 'lounge']	Regular burglary
MO8	['handbag', 'contain', 'car', 'key', 'take', 'hall', 'table']	Regular burglary

Table 7.6 - Dummy training data set of processed and labelled MOs

7.4.2 Feature Extraction

Table 5.7 below lists all of the unique words (tokens) that were extracted from Table 7.6 above.

Bag-of-words vocabulary (|V|)			
1. bmw	**11.** hall	**21.** lounge	**31.** table
2. burglary	**12.** handbag	**22.** male	**32.** take
3. cane	**13.** hook	**23.** obtain	**33.** tv
4. car	**14.** implement	**24.** remove	**34.** two
5. contain	**15.** insecure	**25.** search	**35.** untidy
6. door	**16.** jemmied	**26.** smash	**36.** use
7. driveway	**17.** key	**27.** steal	**37.** vehicle
8. enter	**18.** keys	**28.** stolen	**38.** via
9. front	**19.** laptop	**29.** study	**39.** vw
10. golf	**20.** letterbox	**30.** suspect	**40.** window

Table 7.7 - Bag-of-words vocabulary derived from dummy training data set MOs (blue text indicates that a word is present in the hypothetical Burglary Dwelling MO)

7.4.3 Calculating Probabilities

Recall that there are three terms in the Multivariate Bernoulli Naïve Bayes document classification formula, a reminder of which is provided in Equation 7.6 below.

$$posterior\ probability = conditional\ probability * prior\ probability \tag{7.6}$$

- **Posterior probability**: this describes the probability that a classification document, e.g. an MO, belongs to a particular class given its component features (words).
- **Conditional probability**: this describes the probability that a classification document belongs to a particular class based on associated word likelihoods for |V|.
- **Prior probability**: this describes the proportion of documents in a set of training documents that belong to a particular training class, e.g. Car Key burglary.

Prior probability is the most straightforward of these terms to calculate; it is simply the number of documents in a particular training class divided by the total number of documents in the associated training set. Both classes in the case study example have a prior probability of 0.5 because there are four MOs in each class, and eight MOs in total (4 divided by 8).

Conditional probability is rather more complex to determine than prior probability. First, individual word likelihoods must be calculated, i.e., the probability that a token, (t), in the vocabulary |V| will occur in a given class, e.g. 'driveway' in a Car Key burglary MO. To this end, an empty feature vector (matrix) is constructed for each training class with number of rows equal to the number of documents in the class, and number of columns equal to the length of

|V|. A binary representation is then used to label each row according to the presence (1)/ absence (0) of a word in |V| in the corresponding document, as shown in Table 7.8 below. E.g. if **|V|** = {quick, brown, fox, jump, over, lazy, dog} and **document** = "dog was too lazy to jump", then the **binary vector** = ([0] (quick), [0] (brown), [0] (fox), [1] (jump), [0] (over), [1] (lazy), [1] (dog)). It is not feasible to show results for all forty tokens here so only those for the first nine and t=40 are included in the examples that follow, however, these are derived from the complete vocabulary.

		t=1	**t=2**	**t=3**	**t=4**	**t=5**	**t=6**	**t=7**	**t=8**	**t=9**	**......**	**t=40**
		bmw	burglary	cane	car	contain	door	driveway	enter	front	......	window
Car Key burglary	**MO 1**	0	0	1	1	0	0	0	0	0	......	0
	MO 2	0	0	0	0	0	0	1	0	0	......	0
	MO 3	1	0	0	0	0	0	0	0	0	......	0
	MO 4	0	1	0	0	0	0	0	0	0	......	0
	Total	**1**	**1**	**1**	**1**	**0**	**0**	**1**	**0**	**0**	**.....**	**0**
Regular burglary	**MO 5**	0	0	0	0	0	1	0	1	1	......	0
	MO 6	0	0	0	0	0	0	0	0	0	......	1
	MO 7	0	0	0	0	0	1	0	0	1	......	0
	MO 8	0	0	0	1	1	0	0	0	0	......	0
	Total	**0**	**0**	**0**	**1**	**1**	**2**	**0**	**1**	**2**	**.....**	**1**

Table 7.8 - Extract of binary vector for dummy training data set of processed and labelled MOs

As illustrated in Table 7.8 above, the binary counts for each word in |V| are summed for each training class and the results divided by the number of documents in the class to produce individual word likelihoods. For example, the word 'bmw' occurs once in the Car Key burglary training documents so the likelihood for this token is 0.25 (1 divided by four). A notable issue with this method, however, is that if a particular word in |V| does not appear in any documents in a training class, then the associated binary count is 0, as per Table 7.9 below, which can affect the subsequent posterior probability calculation because zeros cannot be "conditioned away" (Manning et al., 2009, p.260). This problem is more likely to arise with sparse feature vectors, i.e. those with a large associated vocabulary and some words that only occur in a small number of training documents. A possible solution to this, and one that was used in the current example, is Laplace smoothing. Here, +1 is added to the numerator in the word likelihood calculation and

+2 is added to the denominator (Krishnaveni and Sudha, 2017, p.292; Manning et al., 2009, p.263), thus eliminating zeroes, as shown in Table 7.10 below.

	t=1	**t=2**	**t=3**	**t=4**	**t=5**	**t=6**	**t=7**	**t=8**	**t=9**	**......**	**t=40**
	bmw	burglary	cane	car	contain	door	driveway	enter	front	......	window
Car Key burglary	0.25	0.25	0.25	0.25	0.00	0.00	0.25	0.00	0.00	......	0.00
Regular burglary	0.00	0.00	0.00	0.25	0.25	0.50	0.00	0.25	0.50	......	0.25

Table 7.9 - Word likelihoods for dummy training data set without Laplace smoothing

	t=1	**t=2**	**t=3**	**t=4**	**t=5**	**t=6**	**t=7**	**t=8**	**t=9**	**......**	**t=40**
	bmw	burglary	cane	car	contain	door	driveway	enter	front	......	window
Car Key burglary	0.33	0.33	0.33	0.33	0.17	0.17	0.33	0.17	0.17	......	0.17
Regular burglary	0.17	0.17	0.17	0.33	0.33	0.50	0.17	0.33	0.50	......	0.33

Table 7.10 - Word likelihoods for dummy training data set with Laplace smoothing

The processed hypothetical Burglary Dwelling MO, comprising 18 tokens, will now be referred to as b_1, or the 'classification MO'. b_1 contains just seven words from the bag-of-words vocabulary |V|, these being: 'car', 'driveway', 'hook', 'key', 'letterbox', 'steal', and 'use' (recall highlighting in Table 7.7). Having converted the classification MO to a binary feature vector, as per Table 7.11 below, individual word likelihoods will now be used to calculate conditional probability values for the two burglary MO types, namely Car Key burglary and Regular burglary.

	t=1	**t=2**	**t=3**	**t=4**	**t=5**	**t=6**	**t=7**	**t=8**	**t=9**	**......**	**t=40**
	bmw	burglary	cane	car	contain	door	driveway	enter	front	......	window
b_1	0	0	0	1	0	0	1	0	0	**......**	0

Table 7.11 - Extract of feature vector for hypothetical Burglary Dwelling MO

Equations 7.7 and 7.8 below describe how conditional probability values can be derived for each class respectively in the current example. Since |V| contains forty tokens, the 'classifier' iterates over the b₁ binary feature vector forty times per class, checking at every iteration if the corresponding token in |V| is present in the MO (recall that presence = '1' and absence = '0)'. Where a token is present, the associated word likelihood for the class is allocated to the conditional probability formula, however, where a token is not present, the assigned value is '1 minus the word likelihood', or, to put it another way, the chance of t = n not occurring in a document of the class. In this respect, the Bernoulli Naïve Bayes method is unique because it explicitly penalises the absence of an expected word in a classification document (Hu and Tripathi, 2017, p.79). For example, if a word has a 0.85 likelihood of occurring in a Car Key burglary MO, but it does not appear in the classification MO, then the value that is assigned to the conditional probability formula is 0.15, or [(0) + ((1 - 0)*(1 - 0.85))]; this has the effect of lowering the overall conditional probability score. Conversely, if a word has a high likelihood of NOT occurring in a Car Key burglary MO, and it is NOT present in the classification document, then the conditional probability formula is 'rewarded' with a higher input value.

$$\prod_{t=1}^{40}[b_{1t}\mathrm{P}(w_t|Car) + (1 - b_{1t})\big(1 - P(w_t|Car)\big)] \quad (7.7)$$

$$\prod_{t=1}^{40}[b_{1t}\mathrm{P}(w_t|Reg) + (1 - b_{1t})\big(1 - P(w_t|Reg)\big)] \quad (7.8)$$

Table 7.12 below shows a sample of conditional probability inputs for the current example (derived from Laplace smoothed word likelihoods), and Equations 7.9 and 7.10 below show where these probabilities sit in each of the classification formulas. Finally, posterior probability scores can be calculated for each class by multiplying the prior probability by the product of the conditional probability input values.

	t=1	**t=2**	**t=3**	**t=4**	**t=5**	**t=6**	**t=7**	**t=8**	**t=9**	**……**	**t=40**
	bmw	burglary	cane	car	contain	door	driveway	enter	front	……	window
Car Key burglary	0.67	0.67	0.67	0.33	0.17	0.83	0.33	0.83	0.83	……	0.83
Regular burglary	0.83	0.83	0.83	0.33	0.33	0.50	0.17	0.67	0.50	……	0.67

Table 7.12 - Conditional probability inputs (derived from Laplace smoothed word likelihoods)

$$P(Car|\mathrm{b}_1) = 0.5(.67 \times .67 \times .67 \times .33 \times .17 \times .83 \times .33 \times .83 \times .83 \times \ldots\ldots \times .83) \tag{7.9}$$

$$P(Reg|\mathrm{b}_1) = 0.5(.83 \times .83 \times .83 \times .33 \times .33 \times .50 \times .17 \times .67 \times .50 \times \ldots\ldots \times .67) \tag{7.10}$$

Comparing Equations 7.11 and 7.12 below, the Car Key burglary class has generated the highest P value so we can now classify document b_1 as being a Car Key burglary.

$$P(Car\ Key\ burglary|\mathrm{b}_1) = 5.5 \times 10^{-9} \tag{7.11}$$

$$P(Regular\ burglary|\mathrm{b}_1) = 2.3 \times 10^{-11} \tag{7.12}$$

It is beyond the scope of this thesis to review further document classification techniques, but this case study example has hopefully demonstrated that a relatively simple approach, namely Multivariate Bernoulli Naïve Bayes, could potentially improve the results of the current research manual crime sample selection method, particularly in terms of identifying attempt Car Key burglaries. However, the accuracy of any classifier model would need to be evaluated using real crime data. The next section of this chapter will build on the Bernoulli classification example, setting out a proposal for an end-to-end model in which Burglary Dwelling offences are classified, analysed using case linkage, and then used to inform a dynamic crime estimation model, etc.

7.5 Developing an End-to-End Model

It would be useful in the context of the current research, and also from an operational policing perspective, to have some means of automatically identifying unique crime series[9] within a post-classification Car Key burglary data set. The 'Case linkage' approach is appropriate here because it involves connecting undetected offences to a single offender, or group of offenders, based on the principles of 'behavioural consistency' and 'behavioural distinctiveness', the basic premise being that an individual will respond to similar circumstances in a relatively constant manner, but that these responses will vary from one individual to another (Bernasco et al. 2015, p.123; Davies et al., 2012, pp.274-275; Tonkin et al., 2008, pp.59-60). Three potential real-world applications of a case linkage model are shown in Figure 7.3 below, this depicting a high-level process map of the proposed end-to-end model, and also explained in detail in Table 7.13 below.

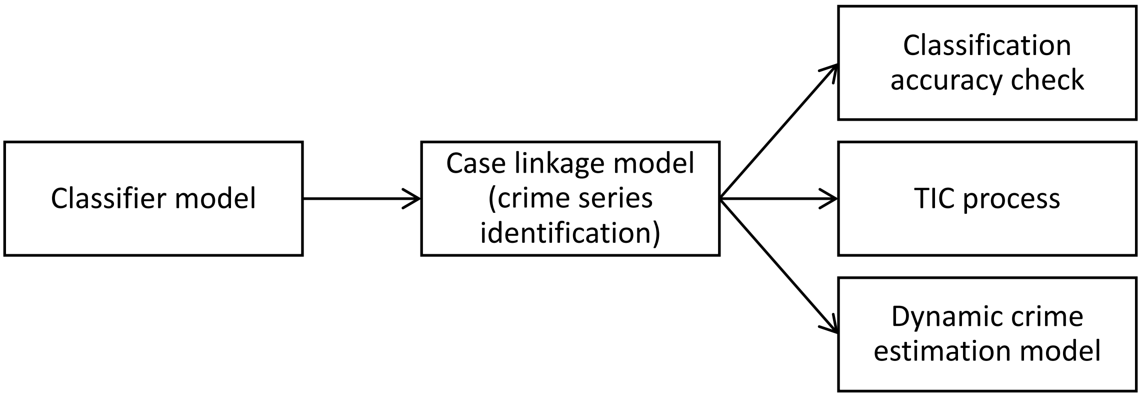

Classifier model
Case linkage model (crime series identification)
Classification accuracy check
TIC process
Dynamic crime estimation model

Figure 7.3 - High-level process map of proposed end-to-end model

[9] When used in the singular sense, this term refers to a collection of offences that bear the hallmarks of having been committed by the same offender, or group of offenders, e.g. an OCG (Organised Crime Group). Somewhat confusingly, the plural form of the term, i.e. more than one crime series, is the same.

Potential application	**Explanation**
Classification accuracy check	Case linkage presents a potential method for checking the reliability of NLP classification results, for instance, when a Burglary Dwelling offence is assigned the label of ‘Car Key burglary’ but a vehicle was not stolen, and the ‘Hanoi’ keyword is not present. Variables that have been shown to accurately discriminate between offenders, including spatio-temporal proximity of offences (Davies et al., 2012, pp.289-290), could thus be used to assess the likelihood of such an offence being an ‘attempt’ Car Key burglary *versus* a misclassified Regular burglary, e.g. is the offence located ‘close’ in time and space to confirmed Car Key burglaries?
TIC process	The CPS states that: “In order to maximise opportunities for suspects to admit TICs, police officers should: … **check all available information sources to establish what other offences the suspect may have committed**” (CPS, 2007, p.9, my emphasis). Since the identification of possible linked offences within a crime data set is fundamental to the TIC process, and also noting that some offenders will have been active over an extended time period, thus requiring many crime records to be searched, an automated case linkage approach is likely to be beneficial in this setting.
Dynamic crime estimation model	Because more than one Burglary Dwelling offender is likely to be operating within a police force area at any given time, it would be difficult to predict when and where each might strike next without having some means of delineating their previous offences, i.e. unique crime series. As before, case linkage could be used to this end, and any results analysed in relation to identified spatio-temporal parameters for the crime type. In other words, given an individual crime series with a start date and start time, how long might we expect the series to continue for, and over what spatial extent?

Table 7.13 - Three potential real-world applications of a case linkage model

7.5.1 Case Linkage of Serial Vehicle Thefts

Only two examples of case linkage having been performed on ‘serial vehicle theft’ offences could be located in the literature, these being Tonkin et al. (2008) and Davies et al. (2012). Both studies used detected car thefts, including thefts of vans, recorded by Northamptonshire police force, East Midlands, UK, and committed between 2004 and 2007, and 2007 and 2011, respectively, to establish which, if any, of a number of different ‘behavioural’, ‘spatial’, and ‘temporal’ variables could accurately discriminate between linked/ unlinked offences. A notable distinction between the two studies is that Tonkin et al. just considered conventional car thefts, whereas Davies et al., having noted the limitations of the former (p.276), also included Burglary Dwelling offences where a car was stolen. Since the current research is primarily focused on Car Key burglary offences, the results of Davies et al. (2012) are likely to be most relevant in terms of their potential application. At the time of writing, there are no examples of case linkage having been performed exclusively on Car Key burglary offences, which suggests that the proposed model would be a novel application of the subject matter.

An initial data sample was created for each of the case linkage studies by first selecting solved car thefts for the associated time period and then isolating individual crime series in the results; both studies defined a series as being two or more offences committed by the same offender. Next, two subset samples were generated, one by pairing an offender’s two most recent offences (‘linked pairs’ subset), and another, comparative sample, by randomly selecting offences detected to different offenders (‘unlinked pairs’ subset). For example, Tonkin et al. (2008) identified 193 serial car thieves in their initial data sample, and thus produced one subset containing 193 linked pairs of offences and another containing 193 unlinked pairs of offences. A similar research method was implemented in both studies and several analytical techniques were employed, including Jaccard’s coefficient, which is a similarity measure, binary logistic regression, and Receiver Operating Characteristic (ROC) curves.

Of the numerous variables that were examined in both studies, two spatial indicators, namely ‘intercrime distance’ (distance between offences) and ‘interdump distance’ (distance between recovered vehicles) consistently performed well in terms of their ability to differentiate between linked/ unlinked offences (e.g. see: Davies et al., 2012, p.282; Tonkin et al., 2008, p.72). Associated negative logit coefficients for these two features suggest that intercrime and interdump distances are shorter for linked car thefts than for unlinked car thefts (Davies et al., 2012, p.282; Tonkin et al., 2008, p.66). Further, Davies et al. (2012) constructed a continuous temporal proximity variable, that is, the number of days between paired offences, and also found this to be a successful case linkage indicator, with an associated negative logit coefficient signifying that linked car thefts are committed closer in time than unlinked car thefts (Davies et al., 2012, p.282). Taken together, these findings are in line with the (closely related) literature on repeat/ near-repeat victimisation and offender mobility patterns, as discussed elsewhere in this thesis, however, it is important to recognise that interdump distance might not be that relevant as a Car Key burglary case linkage indicator because this type of offence is typically committed for material gain and so the intent is to permanently deprive the victim(s) of their property, as opposed to temporary thefts, such as joyriding, where vehicles are usually dumped. Adding further weight to the author's proposal for a standalone Car Key burglary case linkage model is that Tonkin et al. identified < 4.44 km as being the maximum distance at which two car thefts should potentially be considered linked (2008, p.71), whereas Davies et al. observed a maximum distance of 10.55 km (2012, p.289). One possible explanation for the wide variation in these figures is that Davies et al. included Burglary Dwelling offences in their data sample so the presence of more mobile/ professional offenders could have increased the linkage distance. Davies et al. also identified a maximum temporal distance of 227 days for linked cases (p.289). Unfortunately, because no offender data was supplied for the current research, it is not possible to determine Car Key burglary-specific case linkage parameters from detected offences, however, a potential workaround would be to use the Knox test, in conjunction with Monte Carlo simulation, to ascertain over what time and distance repeats/ near-repeats typically occur. For example, assuming that (hypothetically), following an initial Car Key burglary offence, there is over-representation of near-repeats for 14 days and 5 km, this information could be used to conditionally link other crimes committed within the same spatio-temporal parameters relative to an initial event, although a couple of points to consider are: (i) how to identify initial events, and (ii) how to delineate overlapping crime series, i.e. if more than one offender/ offender group is active in an area. It might be useful, therefore, to first classify each case by modus operandi, and then to identify initial events using some sort of spatio-temporal 'search window' algorithm, e.g. given an offence with an earliest committed date and geocoded location, are there any other offences within 14 days and 5 km of this? If not, then the offence could be an initial event, if yes, then the offence might be a linked crime in a series. Further, if two offences are spatio-temporally 'close', and they share the same MO, then these should probably receive higher conditional case linkage scores than two spatio-temporally 'close' offences with different MOs. A weighted scoring method would thus be preferable to a simple 'linked'/ 'unlinked' approach, i.e. we would not wish to entirely disassociate two spatio-temporally 'close' offences on the basis of them having different MOs, but we would somehow need to reflect the fact that they could have been committed by different offenders. The same reasoning applies if we consider that an offender might vary their MO depending on situational characteristics, and also that pertinent details, such as ‘point of entry’, might not be present in/ easily elicited from some MO fields. The suggested algorithm and Figure 7.4 below illustrate how the aforementioned process might work.

Suggested algorithm (based on hypothetical near-repeat parameters):

1. Use Multivariate Bernoulli Naïve Bayes to classify each offence in a Burglary Dwelling data set as being either a ‘Car Key’ burglary or a ‘Regular’ burglary.
2. Create a ‘Car Key’ burglary subset from the results and then classify each offence according to MO characteristics, e.g. ‘Euro Profile’, ‘hook and cane’, ‘insecure’, etc.
3. Geocode the ‘Car Key’ burglary subset, assign URNs, and then load into a GIS.
4. Use a spatio-temporal search window to identify any offences committed within 14 days and 5 km of target offence URN1; if none found, label URN1 as a possible ‘initial event’.
5. Visit URN2 and check stipulated case linkage conditions relative to target offence URN1, e.g. temporally ‘close’? spatially ‘close’? MO the same? Calculate the conditional case linkage score for URN1 and URN2 and then repeat for any remaining URN1 offence pairs.
6. Repeat steps 4 and 5, choosing a new target offence at every model iteration, until there are no offence pairs left to be scored.
7. Output a list of possible ‘initial events’ and a list of URNs with conditional linkage scores; these can then be used as inputs in other processes, such as those shown in Table 5.13.

A possible method for assigning conditional case linkage scores to pairs of offences could be to add (1 / number of linkage conditions stipulated) for every condition met, e.g. ‘temporally close’ = + 33.3 (1 DP), ‘spatially close’ = + 33.3, and ‘same MO’ = + 33.3, meaning that any paired cases where all conditions are met would receive a score of 100.0. To explain using Figure 7.4, assuming that all offences within the 5 km buffer occurred within 14 days of target offence URN1 (yellow X), then a case linkage score of 100.0 would be assigned to the pair ‘URN1 and hook and cane’ because all conditions are met, i.e. spatially ‘close’, temporally ‘close’, and same MO, whereas a case linkage score of 66.6 would be assigned to the pair ‘URN1 and Euro Profile’ because two out of a possible three conditions are met. Only those offences within the 5 km buffer could ever receive the maximum score of 100.0 when paired with URN1 because of the spatial proximity constraint. The scoring method could potentially be developed further by incorporating additional offence elements, such as make & model of vehicle stolen, and suspect description(s), and it might also be worth employing a nominal labelling scheme, e.g. ‘high’, ‘medium’, ‘low’, to avoid 100.0 from being interpreted literally, i.e. still only conditionally linked.

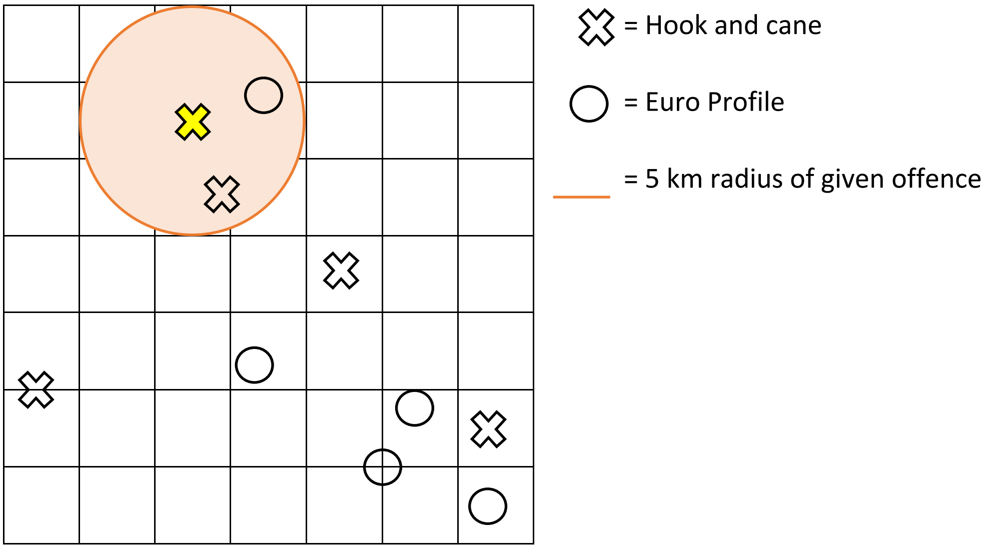


Figure 7.4 - Hypothetical Car Key burglary offences by MO type (URN1 highlighted in yellow)

Both Tonkin et al. (2008) and Davies et al. (2012) analysed a number of MO behaviours (some different), these having first been grouped under relevant ‘domain’ headings, for example, **domain = ‘target selection’**, variables = ‘age of car’, ‘time of day’, etc., **domain = ‘disposal behaviour’**, variables = ‘light damage’, ‘burnt out’, etc.. Tonkin et al. concluded that these domains “performed relatively poorly” as case linkage indicators (2008, p.72), some possible reasons for this being a lack of behavioural distinctiveness between car theft MOs, i.e. there are only so many ways to steal a car, and that individual variables were not analysed separately, potentially masking relationships (p.75). Davies et al.’s (2012) findings, however, were less clear-cut (p.290), which again justifies the need for a separate Car Key burglary study, especially since associated MOs might be far more nuanced given that a dwelling is targeted as well as a vehicle.

7.5.2 Potential Benefits of Automation

The availability of an automated method for interrogating crime data, particularly MO fields, is expected to be most advantageous in those situations where the process needs to be completed quickly, for example, when a potential suspect is in custody and the identification of potentially linked offences could uncover further lines of enquiry. Adding further support to the author’s suggestion for an automated case linkage model is the work of Adderley and Musgrove (2003), who established that their MO data-mining system, ‘Clementine’, was able to produce a list of crime records potentially linked to a specific offender network, and based on 3 months’ of crime data, in just five minutes, vs. between 1 and 2 hours for a manual search (p.275). The provisional accuracy rate of the results was 85 per cent, as opposed to 5 to 10 per cent for a manually produced list (p.275). Taken together, these findings indicate that an automated case linkage model for Car Key burglary would have notable resourcing benefits, as well as providing useful insights for the investigative process. Such a model is also expected to integrate effectively with NLP document classification.

7.6 Topic Modelling of Police-Recorded Incident Logs

This chapter has primarily been concerned with the supervised classification of recorded crime MOs, however, another possible application of natural language processing (NLP) is the use of unsupervised classification to identify emerging issues within police-recorded incident logs. Given that West Yorkshire Police’s Customer Contact Centre “handles an average **2,800** 101 calls per day” (WYP, 2019b, my emphasis and underlining), it would be impractical for an employee to read through every one of these in order to identify emerging issues, and a supervised classification method would not be appropriate either because classes cannot be defined for a subject that is not yet known to exist. Thus, in situations where relevant information is potentially ‘concealed’ within a mass of unstructured textual data, e.g. incident logs, a probabilistic topic model, such as Latent Dirichlet Allocation (LDA) (Blei, 2012), is likely to be of immense value. Without going into too much technical detail, the general idea behind unsupervised classification is that the ‘topics’ are determined by the data, for example, LDA identifies clusters of words within a document set that can then be labelled (manually) according to their prevailing theme (Blei, 2012), e.g. ‘anti-social behaviour’ (ASB), ‘missing person’, ‘suspicious vehicle’. LDA generates a fixed number of topics per document set (Blei, 2012, p.79) and classifications are then assigned depending on how these topics are distributed within individual documents (Kelechava, 2019). It should be noted that LDA has recently been applied to recorded crime MOs (for details see Coleman et al., 2019).

Other potential applications of LDA in the context of incident logs:

To uncover seasonal fluctuations in call for service types, e.g. firework-related ASB.

To capture non-crime issues that still warrant police/ partner agency attention.

To provide evidence of distinct topics in support of new crime classifications.

Table 7.14 below is included to give an indication as to the type of topics that might emerge if LDA is applied to police-recorded incident logs. Three news articles were chosen on different crime and ASB subjects and a random sample of ten words was then extracted from each.

Topic	**News article and source**	**Randomly sampled words**
Thefts of catalytic converters from Honda vehicles	https://www.examinerlive.co.uk/news/west-yorkshire-news/driver-finds-car-jacked-up-16384877 (Robinson, 2019, ExaminerLive).	Car, catalytic, converter, Honda, jacked, Jazz, model, propped, steal, up.
Moped gang robberies	https://www.telegraph.co.uk/news/2019/05/21/moped-robbers-tackled-heroic-shoppers-armed-traffic-cone/ (Lowe, 2019, The Telegraph).	Chase, gang, getaway, helmets, jewellers, moped, mopeds, motorcycle, robbers, stole.
'Spice' drug use	https://www.independent.co.uk/news/uk/home-news/outcry-drug-abuse-photo-spice-zombies-slumped-bench-bridgend-a8402266.html (Staff Reporter, 2018, Independent).	Bench, centre, drug, drugs, slumped, spice, sprawled, town, unconscious, zombies.

Table 7.14 - Example crime topics, associated news articles, and randomly sampled words

To use the example of thefts of catalytic converters from Honda cars, this type of offence is likely to be highly correlated with global metal prices meaning that the problem will probably emerge at specific points in time, potentially generating just a few offences per policing district initially, i.e. until the trend takes off/ there are sufficient crimes to attract attention. In such situations, LDA might be able to flag the problem earlier than conventional crime pattern analysis, for example, if a classifier was implemented at force-level as per Figure 7.5 below, then there could be enough related incident logs, including reports of suspicious activity, etc., to generate a distinct topic. A force-level classifier would thus provide a single source method for capturing the sum of the whole and could be run daily to check for emerging/ ongoing topics within logs.

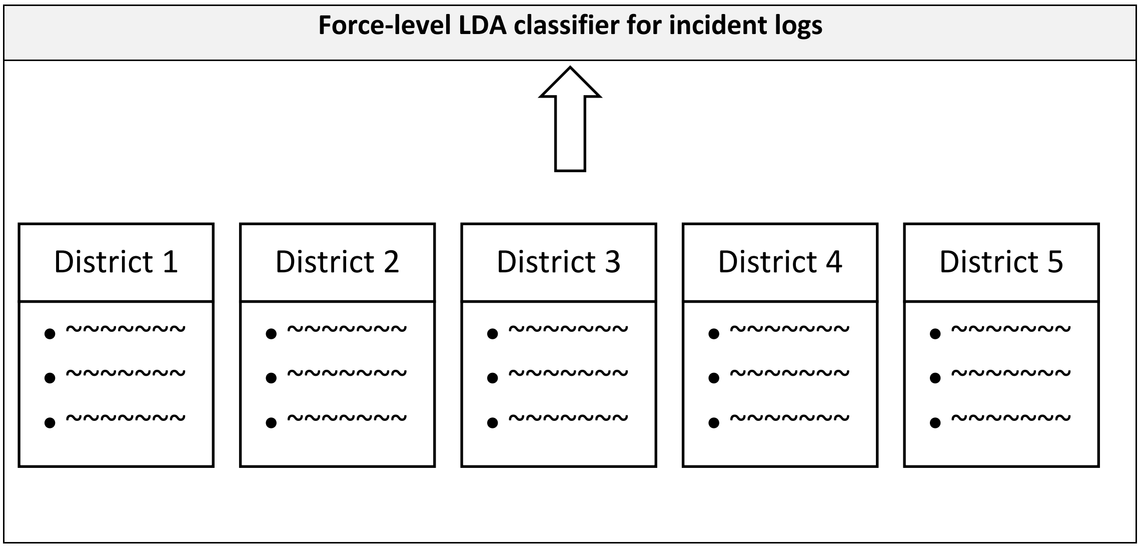
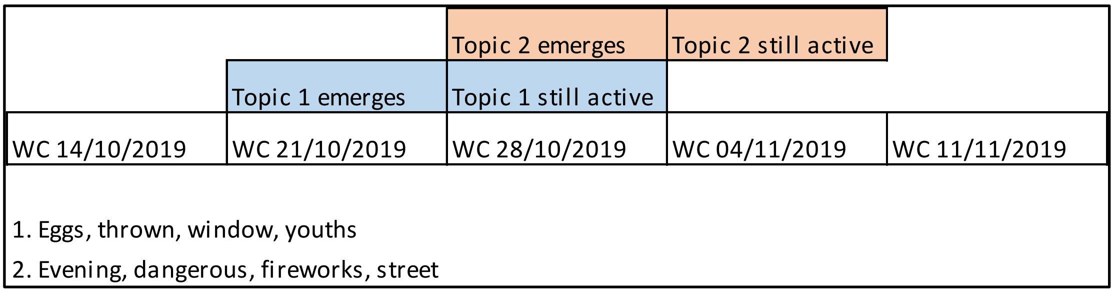


Figure 7.5 - Proposed force-level LDA classifier for incident logs

Figure 7.6 below shows how topic modelling might be used to inform strategic planning and resource deployment, for example, if distinct themes emerge at certain times of the year, such as Halloween (31st October) and Bonfire Night (5th November).



		Topic 2 emerges	Topic 2 still active	
	Topic 1 emerges	Topic 1 still active		
WC 14/10/2019	WC 21/10/2019	WC 28/10/2019	WC 04/11/2019	WC 11/11/2019

1. Eggs, thrown, window, youths
2. Evening, dangerous, fireworks, street

Figure 7.6 - Using topic modelling to inform strategic planning and resource deployment

7.7 Concluding Thoughts

Supervised classification and automatic case linkage appear to be most useful for offences that sit independently under an 'official' crime category, whereas unsupervised classification seems more relevant from a strategic planning perspective, i.e. what is likely to be an issue for policing in the future? Although it has not been possible within the time frame of the PhD to test the ideas discussed in this chapter, there is scope for future novel contribution.

Chapter 8 Discussion and Conclusions

8.1 Introduction

The overall aim of this research was to develop a combined risk model for the prediction of temporally clustered offences, the core rationale being that, for Car Key burglary offences, conventional repeat/ near-repeat victimisation (RV/ N-RV) patterns would not apply, and, thus, some predictive crime modelling methods would be less effective at capturing future offences. Because it was not feasible within the research timescale to evaluate the performance of all existing crime models for Car Key burglary, it was decided to focus on an approach that is inherently linked to spatio-temporal parameters, and that has also been utilised operationally, namely prospective buffering of recent offences. The majority of the results presented in this thesis do support the core research rationale, which is exciting from both a theoretical and a practical perspective, effectively challenging the prevailing narrative within the literature that 'all residential burglary events are created equal'. Perhaps the most valuable learning, however, is that any dominant spatio-temporal parameters that are apparent when offences are analysed at the aggregate level of official Home Office crime recording classes, e.g. Burglary Dwelling, might not apply to all offences that are recorded under these 'summary' categories. Recalling that spatio-temporal specificity is key to the success of both problem-oriented and intelligence-led policing, it is suggested that official crime categories are disaggregated as far as is reasonably possible when undertaking related analysis, including by such factors as property type and MO. The remainder of this chapter will be used to review the key research findings in the context of the core research rationale and the overall aim and objectives that were set out in Chapter 1, to discuss some of the limitations of the research, including those inherent to the data sets used, and to outline some potential opportunities for future development. The chapter will conclude with closing thoughts.

8.2 Key Findings

8.2.1 Conventional RV/ N-RV Patterns Do Not Apply to Car Key Burglary

The repeat/ near-repeat victimisation results presented in Chapter 5 suggest that Car Key burglars do not behave in the same way as Regular burglars, i.e. they do not generally return to the immediate vicinity of their most recent offences in the short-term. It was hypothesised in Chapter 6 that Car Key burglars leave a 'safe' buffer zone around their most recent offences, similar to the buffer zone that is postulated to exist around offenders' home addresses (Brantingham and Brantingham, 2010, p.236). Possible explanations for this are provided by the analysis in Chapters 3, 4, and 5, including that Car Key burglars risk being identified as 'other' when offending in area types that exhibit different characteristics to their own neighbourhoods. It is possible that some Car Key burglary offences will present as lone offences, however, the Near Repeat Calculator results presented in Chapter 5 indicate that when near-repeats do occur, they are highly temporally clustered, i.e. committed on the same mid-point date (crime sprees). The practical implication of this is that there is less opportunity for intervention activity on the part of police and partner agencies. Further, recalling that Johnson and Bowers noted communicability of residential burglary risk in Merseyside up to a distance of 300-400 m and for a period of 1-2 months (2004, p.237), and Johnson et al. identified distances of 300 m and 500 m respectively in Bournemouth and Wirral, both for a period of 2 weeks (2007, p.214), the application of these parameters to recent Car Key burglary offences would probably prove unfruitful, i.e. because offenders are likely to target more distant locations on future crime trips.

8.2.2 Prospective Buffers Are Less Effective for Car Key Burglary Than for Regular Burglary

As expected, the prospective buffers method did not perform as well for Car Key burglary as for Regular burglary, with much higher median daily PAI values achieved for Regular burglary (recall that PAI is the hit rate divided by the area percentage [size of risk area/size of study area*100]). This again supports the research rationale and is ultimately a consequence of the finding that conventional RV/ N-RV patterns do not appear to apply to Car Key burglary offences in the same way that they do Regular burglary. Therefore, the use of a static risk surface to filter areally extensive buffers was justified, albeit the results were slightly disappointing, most likely due to the predictor variables only explaining circa one third of the variation in Car Key burglary rates. However, it is important to note that, although the previously observed hit rates for Car Key burglary decreased when the static risk surface was used (combined risk model), the median daily PAI and SER (hit rate per km^2) values increased, indicating that the hit rates were actually better, on average, relative to the sizes of the identified risk areas. Using risk heterogeneity to filter offenders' general criminal activity spaces clearly warrants further investigation, although a better fitting regression model would need to be identified. If so, it would then be worth developing a static risk surface with more spatially disaggregate input data, e.g. at the Output Area or postcode unit level, with a view to delineating more operationally actionable patrol areas that are able to capture/ prevent a high proportion of future offences.

8.3 Limitations of the Research

8.3.1 Implications of Data Choices

The crime samples selection method that was employed in the current research meant that any residential burglary offences that were not classified as a Car Key burglary were subsequently assigned to the Regular burglary data set. Consequently, there is likely to be some variation of offence characteristics within the Regular burglary data set, for example, if distraction burglaries and/or 'attempt' Car Key burglaries are present, as illustrated in Figure 8.1 below.

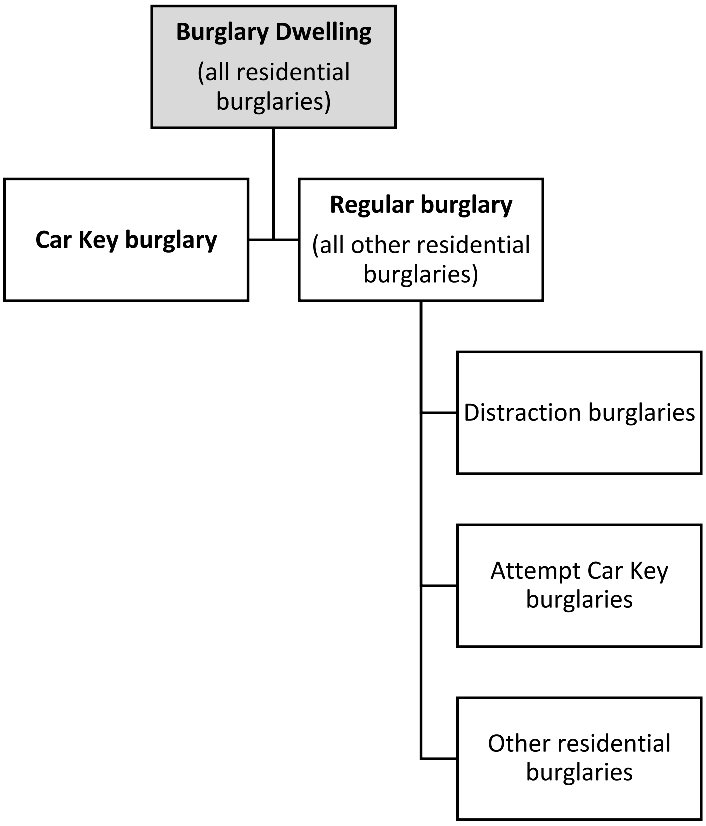

Burglary Dwelling
(all residential burglaries)
Car Key burglary
Regular burglary
(all other residential burglaries)
Distraction burglaries
Attempt Car Key burglaries
Other residential burglaries

Figure 8.1 - Possible crime type (characteristics) heterogeneity within the Regular burglary data set

Possible implications of this for the current research are that different spatio-temporal signatures could have masked/ weakened prevailing trends within the Regular burglary data set, and that any findings might not be applicable to all offences within this. Thus, in hindsight, it would have been expedient to have removed any burglaries from the Regular burglary data set where the associated spatio-temporal signatures were not expected to reflect 'conventional' burglars' offending patterns, i.e. locally-anchored suspects generally operating within the confines of their routine activity spaces. For example, a study by Steele et al. (2001) observed that distraction burglars typically travel long distances from their home base to commit offences, with one participant (bogus offender) stating: "Good bogeymen travel long distances to do jobs" (Steele et al., 2001, p.58). Distraction burglaries and Car Key burglaries are the two sub-categories of residential burglary that spring to mind in the context of mobile offending and, thus, potential crime sprees, although there could be other crime types within the Regular burglary data set that the author is unaware of that also generate sprees. However, it is pertinent to mention here that ~ 2% of police-recorded domestic burglaries in England and Wales from year ending March 2010 to year ending March 2015 – approximate study period for the thesis – were classified as distraction/ attempt distraction MO (ONS, 2020), and also to recall from Chapter 3 that West Yorkshire Police (2019) estimate that < 10% of burglaries are Car Key burglaries. Therefore, the relatively low prevalence of distraction burglary and Car Key burglary offences, together with the findings of the Literature Review and the spatial analysis presented in Chapter 3, indicate that 'conventional' residential burglary is likely to be well-represented within the Regular burglary data set. It is also worth mentioning here the possibility that some Car Key burglaries might have been misclassified as vehicle crime offences, i.e. theft of motor vehicles, although this is not expected to be a common occurrence given that victims and/or the police are likely to know if car keys were stolen during a linked Burglary Dwelling offence.

Had offender data been made available to the author, this could have been used to determine the average journey-to-crime distance for Car Key burglars in West Yorkshire, i.e. the distance measure that was used to calibrate the extent of the periphery for the surrounding areas work in Chapter 4. This would have been preferable to relying wholly on Carden's (2012, p.74) finding – for offences in Merseyside – that Car Key burglars travel a median distance of 4.9 km from home to crime location. Further, a distance measure derived from the PhD study area might have been more appropriate given that the geography of West Yorkshire differs to that of Merseyside, which could affect travel patterns. A second average journey-to-crime distance would also have facilitated a better understanding of Car Key burglars' journey-to-crime behaviours and how these may generate particular crime patterns. However, and perhaps more crucially, the offender data could have been used to validate the findings of the repeat victimisation/ near-repeat victimisation analysis that was undertaken using Ratcliffe's Near Repeat Calculator and discussed in Chapter 5 of the thesis. The author subsequently inferred, based on the results of this analysis, that: "during a single 24 hour period, Car Key burglars typically target properties within fairly close proximity of each other (≤ 500 m), but do not return to the immediate surroundings of their recent offences in a consistent manner, unlike Regular burglars". Given that eastings and northings were supplied for the offence locations in the current research, these could have been used in conjunction with offender data, i.e. for detected offences, to ascertain if the aforementioned statement holds true for the majority of Car Key burglars. That is, how often does an offender/ offender group return to the exact location/ immediate vicinity of a previous Car Key burglary offence, and over what time period? This analysis could have been performed in either a GIS, or using coding software, such as R. Another data set that was not available to the author at the time of writing, but which could potentially have offered some useful insights, is make and model of vehicles owned at LSOA level. However, and as mentioned in Chapter 5, some of the independent variables that were included in the PhD analysis, such as NS-SeC, might have successfully acted as a proxy for this. Nevertheless, assuming that data on vehicles at LSOA level had been available, this could have been categorised by the author if necessary, for example, using Lansley's (2016) approach, and the resulting groupings then included as variables in the Spearman's Rank Correlation analysis. This would have identified any significant relationships between Car Key burglary rates and Regular burglary rates and LSOA prevalence of different vehicle makes and models. In addition to make and model, other vehicle-related data that might have proved useful includes registration year and value.

In terms of data for the dependent and independent variables in the study, the ideal for the dependent data would have been a pre-classified Burglary Dwelling data set, including attempt Car Key burglaries, or, failing this, the availability of automated crime/ burglary classification software, hence Chapter 7. It would also have been useful if offender data been provided for any detected offences so that this could have been used to identify dominant spatio-temporal behaviours for different offender types, e.g. Car Key burglars vs. 'conventional' Regular burglars vs. distraction burglars. As already mentioned elsewhere, it would have been helpful if more information had been provided on the characteristics of vehicles stolen from Car Key burglaries because that which was included was extremely limited. This information could then have been used to further disaggregate the Car Key burglary data set, for example, to distinguish more organised/ professional offences from those that appeared to be less so. The ideal for the independent variables would have been more data on vehicle ownership at LSOA level, as discussed in the previous paragraph, together with more spatially disaggregate, e.g. postcode unit level, versions of this and the independent variables that were used in the Spearman's Correlation analysis. The inclusion of more spatially disaggregate data in the analysis might have helped to mitigate any issues inherent to the ecological fallacy and Modifiable Areal Unit Problem (MAUP), as well as potentially reducing the areal extent of any identified 'patrol' areas in the combined risk model.

8.3.2 Lack of Information Regards Stolen Vehicles

A key limitation of the current research is that detailed information regarding the characteristics of stolen vehicles was not available to the author – this meant that it was not possible to distinguish more 'professional' offences, i.e. those where high-value vehicles were stolen, from those where vehicles were stolen for the sole purpose of providing transport, etc. Perhaps if 'professional' offences had been analysed separately from all other Car Key burglary offences, there might have been a much stronger relationship between the associated crime rates and the prevailing socio-demographic characteristics of target areas. Recognising that the research was conducted in the absence of any information on the makes, models, ages, and values of the stolen vehicles, the moderate Spearman's coefficients that were observed between the Car Key burglary rates and some of the independent variables are quite encouraging, i.e. a better fitting *k*-NN regression model could probably be developed if more detailed vehicle information was available.

8.3.3 What If the Technology Moves On?

A question that was frequently put to the author in relation to the choice of crime type for the research, namely Car Key burglary, was how relevant the findings would be if technological advances rendered the associated MO redundant. My usual response to this was that the research is not about Car Key burglary *per se*, rather it is about evidencing how the spatio-temporal signatures of different crime types can vary depending on associated juxtapositions of would-be offenders and a target property type. The PhD hypotheses, and research findings, are also viewed by the author to be transferrable to other crime types, for example, bank robberies. To explain, professional bank robbers are unlikely to target the same branch more than once within a short time period due to the inherent risks, however, an initial robbery might flag the risk of future events at other locations within offenders' criminal activity spaces. As per the approach employed in this study, the average JTC distance for professional robbers could be used to infer the spatial extent of a general criminal activity space, based on the location of a recent bank robbery, and the resulting buffer then filtered to identify future high risk locations.

8.4 Opportunities for Future Work

Aside from the potential opportunities for future work that have already been discussed, other areas to consider include developing an end-to-end predictive crime model, refining the area types juxtapositions variable, further investigating the suitability of *k*-NN regression as a means of estimating crime rates, and testing the dynamic and combined models on unseen crime data, i.e. in an operational setting.

Chapter 7 outlined a proposal for how machine learning could be used to automate the crime prediction process, i.e. from initial crime recording through to risk area identification. Although an end-to-end (single push button) model would perhaps be the panacea, the text classification element of the proposed approach presents an opportunity for a quick win in terms of the efficient use of public resources. Noting the difficulties that were inherent to manually deriving the crime samples in the current research, if this is replicated in the real-world, i.e. crime practitioners also have to manually differentiate between offence types within crime data sets, then there is unquestionable value in pursuing the supervised text classification approach, although this would need to be tested empirically using real crime data.

Perhaps one of the most exciting findings of the current research is that the area type juxtaposition variable – 'IDW_05' – exhibited the strongest correlation with the Car Key burglary rates. This indicates that there would be real value in undertaking further analysis on the influence of surrounding area characteristics on Car Key burglary rates, albeit whilst employing a more robust method of defining relative area attractiveness. Given that 'percentage households social renting or private renting' was used to infer the location of area attractiveness/ high offender rate areas, it might have been preferable, in the absence of actual offender data, to estimate West Yorkshire LSOA/ OA offender counts using the offender rate and housing tenure data shown in Figure 4.2; these could then have been used as the input data in the inverse distance weighting. Another variable that could have been examined to this end is within-area deprivation. The PhD approach might also have been improved by using road network distances between areas, as opposed to Euclidean distances, to weight the attractiveness variable. It is also worth considering that Car Key burglars may well avoid certain categories of road due to the risk of detection – this could be investigated further by speaking to practitioners with a view to more accurately representing offenders' mobility patterns. A final refinement to mention here is that areas beyond the West Yorkshire boundary should be included in future analysis to capture the potential for cross-border offending.

The apparent dearth of literature regards *k*-NN regression having been applied to the estimation of crime rates is somewhat surprising – it is the author's view that this method warrants further investigation, particularly since the k-NN model performed almost as well as the final OLS model in the current research. In terms of other models/ methods that could have been employed in the current research, Andresen's (2016) 'area-based nonparametric spatial point pattern test' might have proved useful in the Chapter 3 analysis. This method tests if the proportional distributions of two point data sets vary globally (*S*-Index) and locally (indicator of spatial similarity) across the same areal units (Andresen, 2016), and could thus have been used to compare the Car Key burglary and Regular burglary points with a view to boosting the research rationale. Unfortunately, the author was not aware of the test until recently, however, it would probably have been quite easy to perform due to there being a graphical user interface. A different method that could potentially have been used to predict the locations of future burglary offences in the research is self-exciting point process (SEPP) modelling. SEPP is based on the idea that a background rate of events generates initial events which then trigger future events (e.g. see Mohler et al., 2011), and this has been applied to spatio-temporal clustering in epidemiology and seismology, the latter to predict aftershocks (Reinhart, 2018). Unsurprisingly, given that it replicates repeat/ near-repeat victimisation behaviours, SEPP has also been employed in the field of crime prediction, and apparently underpins *PredPol,* a commercial crime risk prediction software, that is used in some policing departments in the USA and UK (Rosser et al., 2017, p.574). Further, the findings of the Near Repeat Calculator analysis that were used to inform the dynamic risk model inputs could have been tested/ validated using a spatio-temporal clustering algorithm, such as 'Spatial-Temporal Density Based Spatial Clustering of Applications with Noise' (ST-DBSCAN), to ascertain which space-time parameters displayed the greatest similarity, say among modus operandi or offender characteristics (e.g. see Chen et al., 2020), in identified hot spots for each burglary type. This approach might, for example, have uncovered homogeneity of Car Key offence/ Car Key offender characteristics at much shorter distances than for Regular burglary due to the assumed spree-like offending behaviours of Car Key burglars.

Since the 'boost' – dynamic buffers – approach performed less well for Car Key burglary than for Regular burglary, probably because Car Key burglars appear to commit temporally clustered crime sprees, then the potential application of the dynamic risk model for future crime pattern analysis is limited. However, if offender data was to be made available to the author for any detected Car Key burglaries in the crime data set, then it would then be possible to ascertain if the overrepresentation of near-repeat victimisation that was apparent at longer spatio-temporal bands for Car Key burglary was in fact linked to the same offenders/ offender groups as previous sprees, as opposed to simply reflecting risk heterogeneity across the study area. If so, then this would justify the use of a static risk surface to filter dynamic risk in the model and would also give greater confidence in the usefulness of the combined model for law enforcement purposes. For example, identified risk areas could be used to inform police activity, including intelligence gathering, crime prevention work, and disruption patrols. However, the two models that were developed in the current research would need to be tested on unseen crime data, and ideally in an operational setting, to understand if the observed results are a true representation of reality because desk-based performance is likely to be 'ideal-world'. For example, if identified risk areas are not both contiguous and small in extent, then they are likely to be impractical for the police to resource. Since the performance of the combined model for Car Key burglary was not that dissimilar to the ProMap 'combined' algorithm for residential burglary, and also assuming that the current research near-repeat victimisation findings for Car Key burglary hold true, then the author sees potential value in developing the combined risk model with a view to it being used as a tool in the prevention and detection of future Car Key burglary offences, as well as any other crime types displaying similar spatio-temporal signatures.

8.5 Closing Thoughts

One of the most novel aspects of this PhD is that it generated daily crime predictions over an extended time period and for an unofficial crime category – to the best of the author's knowledge there are no other examples of the HO Burglary Dwelling classification having been disaggregated and then analysed in relation to repeat/ near-repeat victimisation patterns. Perhaps the closest example is Adepeju (2017), who examined RV/ N-RV patterns for three official burglary categories (burglary-in-residence, burglary-of-hotels, and burglary-in-stores) in

San Francisco. The current work not only contributes to, but supports, the existing (limited) literature on Car Key burglary, as well as adding a new dimension to that on repeat/ near-repeat victimisation and boost account theory.

List of References

Abbott-Smith, C. 2018. *Community outcomes meeting April 2018: Report on serious acquisitive crime*. [Online]. West Yorkshire Police. [Accessed 26 April 2019]. Available from: http://www.westyorkshire-pcc.gov.uk/sites/default/files/2019-11/item_7-_sac_cover_report.pdf

Ackerman, J. M. and Rossmo, D. K. 2015. How far to travel? A multilevel analysis of the residence-to-crime distance. *Journal of Quantitative Criminology.* **31**(2), pp.237-262.

Adderley, R. and Musgrove, P. 2003. Modus operandi modelling of group offending: A data-mining case study. *International Journal of Police Science & Management.* **5**(4), pp.265-276.

Addis, N. 2012. *Exploring the impact and effectiveness of the 'Project Optimal' burglary reduction initiative in Leeds: A spatio-temporal approach*. [Online]. [Accessed 12 February 2016]. Available from: http://www.geos.ed.ac.uk/

Addis, N., Evans, A. and Malleson, N. 2019. Exploring the practices of steal-to-order burglars: A different brand of offender? *Security Journal*. **32**(4), pp.457-475.

Adepeju, M. 2017. Investigating the repeat and near-repeat patterns in sub-categories of burglary crime. In: *Proceedings of the International Conference on GeoComputation, 04-07 Sep 2017, Leeds, UK*. Leeds: Centre for Computational Geography, University of Leeds.

Allcock, E., Bond, J. W. and Smith, L. L. 2011. An investigation into the crime scene characteristics that differentiate a car key burglary from a regular domestic burglary. *International Journal of Police Science & Management.* **13**(4), pp.275-285.

Andresen, M. A. 2014. *Environmental criminology: Evolution, theory, and practice.* Abingdon: Routledge.

Andresen, M. A. 2016. An area-based nonparametric spatial point pattern test: The test, its applications, and the future. *Methodological Innovations*. **9**, pp.1–11.

Ashby, M. P. J. and Bowers, K. J. 2013. A comparison of methods for temporal analysis of aoristic crime. *Crime Science.* **2**(1), pp.1-16.

Bailey, T. C. and Gatrell, A. C. 1995. *Interactive spatial data analysis*. Harlow: Longman Scientific & Technical.

Baldwin, J., Bottoms, A. E. and Walker, M. A. 1976. *The urban criminal: A study in Sheffield.* London: Tavistock.

Bartholomew, D. J., Steele, F., Moustaki, I. and Galbraith, J. I. 2008. *Analysis of multivariate social science data.* Second ed. Boca Raton, FL: CRC Press

Bavelas, A. 1950. Communication patterns in task-oriented groups. *The Journal of the Acoustical Society of America.* **22**(6), pp.725-730.

BBC. 2005. *Hanoi-style car theft gang jailed*. [Online]. 30 June 2005. [Accessed 18 February 2016]. Available from: http://news.bbc.co.uk/1/hi/england/west_yorkshire/4639421.stm

BBC. 2017. *Car theft 'relay' devices seized in Birmingham*. [Online]. 15 December 2017. [Accessed 14 September 2020]. Available from: https://www.bbc.co.uk/news/uk-england-birmingham-42370086

Bedfordshire Police. 2019. *What is burglary*. [Online]. [Accessed 26 December 2019]. Available from: https://www.bedfordshire.police.uk/information-and-services/Crime/Burglary/What-is-burglary

Bernasco, W. 2006. Co-offending and the choice of target areas in burglary. *Journal of Investigative Psychology and Offender Profiling.* **3**(3), pp.139-155.

Bernasco, W. and Luykx, F. 2003. Effects of attractiveness, opportunity and accessibility to burglars on residential burglary rates of urban neighborhoods. *Criminology.* **41**(3), pp.981-1001.

Bernasco, W. and Nieuwbeerta, P. 2005. How do residential burglars select target areas? A new approach to the analysis of criminal location choice. *British Journal of Criminology.* **45**(3), pp.296-315.

Bernasco, W., Johnson, S. D. and Ruiter, S. 2015. Learning where to offend: Effects of past on future burglary locations. *Applied Geography.* **60**(June), pp.120-129.

Bird, S., Klein, E. and Loper, E. 2009. *Natural language processing with Python: Analyzing text with the Natural Language Toolkit.* Sebastopol, CA: O'Reilly.

Birkin, M., Clarke, G. and Clarke, M. 2010. Refining and operationalizing entropy-maximizing models for business applications. *Geographical Analysis.* **42**(4), pp.422-445.

Birkin, M., Clarke, G. and Clarke, M. 2017. *Retail location planning in an era of multi-channel growth.* Abingdon: Routledge.

Blei, D. M. 2012. Probabilistic topic models: surveying a suite of algorithms that offer a solution to managing large document archives. *Communications of the ACM.* **55**(4), pp.77-84.

Boeing, G. 2017. OSMnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks. *Computers, Environment and Urban Systems.* **65**(September), pp.126-139.

Boggs, S. L. 1965. Urban crime patterns. *American Sociological Review.* **30**(6), pp.899-908.

Bottoms, A. E. and Wiles, P. 1986. Housing tenure and residential community crime careers in Britain. *Crime and Justice.* **8**, pp.101-162.

Bowers, K. and Hirschfield, A. 1999. Exploring links between crime and disadvantage in north-west England: An analysis using geographical information systems. *International Journal of Geographical Information Science.* **13**(2), pp.159-184.

Bowers, K. J., Johnson, S. D. and Pease, K. 2004. Prospective hot-spotting: The future of crime mapping? *British Journal of Criminology.* **44**(5), pp.641-658.

Bowers, K. J. and Johnson, S. D. 2005. Domestic burglary repeats and space-time clusters: The dimensions of risk. *European Journal of Criminology.* **2**(1), pp.67-92.

Brantingham, P. and Brantingham, P. 1995. Criminality of place: Crime generators and crime attractors. *European Journal on Criminal Policy and Research.* **3**(3), pp.5-26.

Brantingham, P. L. and Brantingham, P. J. 1993. Nodes, paths and edges: Considerations on the complexity of crime and the physical environment. *Journal of Environmental Psychology.* **13**, pp.3-28.

Brantingham, P. L. and Brantingham, P. J. 2010. Notes on the geometry of crime (1981). In Andresen, M. A., Brantingham, P. J. and Kinney, J. B. eds. *Classics in environmental criminology.* Boca Raton, FL: CRC Press, pp.231-256.

Brantingham. P. J. and Brantingham. P. L. 2013. The theory of target search. In: Cullen, F. T. and Wilcox, P. eds. *The Oxford Handbook of Criminological Theory.* New York: Oxford University Press, pp.535-553.

Brown, N. 2002. *Robert Park and Ernest Burgess: Urban ecology studies, 1925*. CSISS Classics. [Online]. UC Santa Barbara: Center for Spatially Integrated Social Science. [Accessed 10 August 2015]. Available from: https://escholarship.org/uc/item/6f39q98d

Brownlee, J. 2014. *Feature selection with the caret r package*. [Online]. [Accessed 18 February 2020]. Available from: https://machinelearningmastery.com/feature-selection-with-the-caret-r-package/

Budd, T. 2001. *Burglary: Practice messages from the British Crime Survey.* (Briefing Note 5/01). London: Research, Development and Statistics Directorate, Home Office.

Cahill, M. and Mulligan, G. 2007. Using geographically weighted regression to explore local crime patterns. *Social Science Computer Review.* **25**(2), pp.174-193.

Carden, R. J. 2012. *Car key burglaries: An exploratory analysis*. MSt thesis, University of Cambridge.

Chainey, S. and Ratcliffe, J. 2005. *GIS and crime mapping.* Chichester: John Wiley & Sons Ltd.

Chainey, S., Tompson, L. and Uhlig, S. 2008. The utility of hotspot mapping for predicting spatial patterns of crime. *Security Journal.* **21**(1-2), pp.4-28.

Chainey, S. 2013. *Examining the influence of cell size and bandwidth size on kernel density estimation crime hotspot maps for predicting spatial patterns of crime.* Bulletin of the Geographical Society of Liege 60:7-19.

Chainey, S. 2014. *Examining the extent to which hotspot analysis can support spatial predictions of crime*. Ph.D. thesis, University College London (UCL).

Chapman, R., Smith, L. L. and Bond, J. W. 2012. An investigation into the differentiating characteristics between car key burglars and regular burglars. *Journal of Forensic Sciences.* **57**(4), pp.939-945.

Charlton, M. and Fotheringham, A. S. 2009. *Geographically weighted regression: A tutorial on using GWR in ArcGIS 9.3.* [Online]. Maynooth, Ireland: National Centre for Geocomputation. [Accessed 24 October 2017]. Available from: https://www.geos.ed.ac.uk/~gisteac/fcl/gwr/gwr_arcgis/GWR_Tutorial.pdf

Chen, T., Bowers, K., Cheng, T., Zhang, Y. and Chen, P. Exploring the homogeneity of theft offenders in spatio-temporal crime hotspots. *Crime Science*. **9**(1), pp.1-13.

Clarke, R. V. 1999. *Hot products: Understanding, anticipating and reducing demand for stolen goods*. Police Research Series Paper 112. London: Policing and Reducing Crime Unit, Research, Development and Statistics Directorate, Home Office.

Cochrane, R. 2016. *The secret life of equations: The 50 greatest equations and how they work.* London: Cassell.

Cohen, L. E. and Felson, M. 1979. Social change and crime rate trends: A routine activity approach. *American Sociological Review.* **44**(4), pp.588-608.

Cohen, J. 1988. *Statistical power analysis for the behavioral sciences.* 2nd ed. Hillsdale, NJ: L. Erlbaum Associates.

Cohen, E., Delling, D., Pajor, T. and Werneck, R. F. 2014. Computing classic closeness centrality, at scale. In: *Proceedings of the second ACM Conference on Online Social Networks, 01-02 Oct 2014, Dublin, Ireland*. New York, NY: ACM, pp.37-50. [Accessed 26 November 2018]. Available from: https://dl.acm.org/

Coleman, A., Birks, D., Malleson, N. and Farrell, G. 2019. *Extracting actionable insights from free text police data*. [Online]. [Accessed 02 July 2019]. Available from: https://lida.leeds.ac.uk/research-projects/extracting-actionable-insights-from-free-text-police-data/

Copes, H. and Cherbonneau, M. 2006. The key to auto theft: Emerging methods of auto theft from the offenders' perspective. *British Journal of Criminology.* **46**(5), pp.917-934.

Cornish, D. B. and Clarke, R. V. 1987. Understanding crime displacement: An application of rational choice theory. *Criminology.* **25**(4), pp.933-948.

Cornish, D. B. and Clarke, R. V. 2008. The rational choice perspective. In: Wortley, R. and Mazerolle, L. eds. *Environmental criminology and crime analysis.* Cullompton: Willan Publishing, pp.21-47.

Costello, A. and Wiles, P. 2001. GIS and the journey to crime: An analysis of patterns in South Yorkshire. In: Hirschfield, A. and Bowers, K. eds. *Mapping and analysing crime data: Lessons from research and practice.* London: Taylor & Francis, pp.27-60.

CPS. 2007. *Prosecution Team Guidance: Offences to be taken into consideration.* [Online]. London: Crown Prosecution Service. [Accessed 23 August 2019]. Available from: https://www.cps.gov.uk/sites/default/files/documents/legal_guidance/chapter_o_ptg.pdf

Davies, K., Tonkin, M., Bull, R. and Bond, J. W. 2012. The course of case linkage never did run smooth: A new investigation to tackle the behavioural changes in serial car theft. *Journal of Investigative Psychology and Offender Profiling.* **9**(3), pp.274-295.

Day, L. 2020. Man arrested after Mansfield car key burglary. *Chad*. [Online]. 28 July 2020. [Accessed 24 September 2020]. Available from: https://www.chad.co.uk/news/crime/man-arrested-after-mansfield-car-key-burglary-2926573

Dodd, T., Nicholas, S., Povey, D. and Walker, A. 2004. *Crime in England and Wales 2003/2004.* (Home Office Statistical Bulletin). London: Research, Development and Statistics Directorate, Home Office.

Eck, J. E. and Weisburd, D. 1995. Crime places in crime theory. In: Eck, J. E. and Weisburd, D. eds. *Crime and Place: Crime Prevention Studies, Volume 4*. Monsey, NY: Criminal Justice Press, pp.1-33.

Eck, J. E., Chainey, S., Cameron, J. G., Leitner, M. and Wilson, R. E. 2005. *Mapping crime: Understanding hot spots.* [Online]. Washington, DC: National Institute of Justice. [Accessed 18 November 2015]. Available from: https://www.researchgate.net/publication/32894301_Mapping_Crime_Understanding_Hot_Spots

Elofson, M. 2020. Laptop computer and Nissan stolen during car-key burglary in Kenilworth. *The Leamington Spa Courier*. [Online]. 20 September 2020. [Accessed 24 September 2020]. Available from: https://www.leamingtoncourier.co.uk/news/crime/laptop-computer-and-nissan-stolen-during -car-key-burglary-kenilworth-2977573

ESRI. [no date]. *Hot spot analysis (Getis-Ord Gi*) (spatial statistics).* [Online]. [Accessed 14 December 2019]. Available from: http://resources.esri.com/help/9.3/arcgisengine/java/gp_toolref/spatial_statistics_tools/hot_spot_analysis_getis_ord_gi_star_spatial_statistics_.htm

ESRI. 2016. *How median center works.* [Online]. [Accessed 04 September 2018]. Available from: http://desktop.arcgis.com/en/arcmap/10.3/tools/spatial-statistics-toolbox/h-how-median-center-spatial-statistics-works.htm

ESRI. c2018. *How average nearest neighbor works.* [Online]. [Accessed 04 February 2020]. Available from: https://desktop.arcgis.com/en/arcmap/10.3/tools/spatial-statistics-toolbox/h-how-average-nearest-neighbor-distance-spatial-st.htm

ESRI. c2018b. *How spatial autocorrelation (Global Moran's I) works*. [Online]. [Accessed 04 February 2020]. Available from: https://desktop.arcgis.com/en/arcmap/10.3/tools/spatial-statistics-toolbox/h-how-spatial-autocorrelation-moran-s-i-spatial-st.htm

ESRI. c2019. *Directional distribution (standard deviational ellipse).* [Online]. [Accessed 04 February 2020]. Available from: https://desktop.arcgis.com/en/arcmap/latest/tools/spatial-statistics-toolbox/directional-distribution.htm

Essex Police. ca. 2015. *Essex Police performance update December 2015*. [Online]. Essex: Corporate Services, Essex Police. [Accessed 18 April 2019]. Available from: http://www.essex.pfcc.police.uk/wp-content/uploads/2015/08/Item-4a-Essex-Police-Performance-Update-December-2015-V3.pdf

Farrell, G., Tilley, N., Andromachi, T. and Mailley, J. 2008. The crime drop and the security hypothesis. *British Society of Criminology Newsletter.* **62** (Winter 2008), pp.17-21.

Fielding, M. and Jones, V. 2012. 'Disrupting the optimal forager': Predictive risk mapping and domestic burglary reduction in Trafford, Greater Manchester. *International Journal of Police Science & Management.* **14**(1), pp.30-41.

Fogarty, B. J. 2019. *Quantitative social science data with R: An introduction*. [Online]. London: Sage. [Accessed 04 February 2020]. Available from: https://books.google.co.uk/books?id=YUNeDwAAQBAJ

Fornito, A., Zalesky, A. and Bullmore, E. T. 2016. *Fundamentals of brain network analysis.* Amsterdam: Academic Press.

Flowerdew, R. 2011. How serious is the modifiable areal unit problem for analysis of English census data? *Population Trends.* (145), pp.102-114.

Fotheringham, A. S. and Rogerson, P. A. 1993. GIS and spatial analytical problems. *International Journal of Geographical Information Science.* **7**(1), pp.3-19

Freeman, L. C. 1979. Centrality in social networks: Conceptual clarification. *Social Networks.* **1**(3), pp.215-239.

Friendly, M. 2007. A.-M. Guerry's moral statistics of France: Challenges for multivariable spatial analysis. *Statistical Science.* **22**(3), pp.368-399.

Frith, M. J., Johnson, S. D. and Fry, H. M. 2017. Role of the street network in burglars' spatial decision-making: Offender spatial decision-making. *Criminology.* **55**(2), pp.344-376.

Frost, J. 2017. *Heteroscedasticity in regression analysis*. [Online]. [Accessed 09 February 2020]. Available from: https://statisticsbyjim.com/regression/heteroscedasticity-regression/

Gale, C. G., Singleton, A. D., Bates, A. G. and Longley, P. A. 2016. Creating the 2011 area classification for output areas (2011 OAC). *Journal of Spatial Information Science.* **2016**(12), pp.1-27.

Geitgey, A. 2018. *Natural language processing is fun! How computers understand human language*. [Online]. [Accessed 23 March 2020]. Available from: https://medium.com/@ageitgey/natural-language-processing-is-fun-9a0bff37854e

Glasner, P., Johnson, S. D. and Leitner, M. 2018. A comparative analysis to forecast apartment burglaries in Vienna, Austria, based on repeat and near repeat victimization. *Crime Science.* **7**(1), pp.1-13.

Greater Manchester Police. 2019. *Crime prevention advice (D182aT): Any vehicle can be a target*. [Leaflet]. [Online]. Manchester: Greater Manchester Police. [Accessed 29 December 2019]. Available from: https://www.gmp.police.uk/SysSiteAssets/media/images/greater-manchester/campaigns/2019/local-advice/stockport/anyvehiclecanbeatarget.pdf

Hagberg, A., Schult, D. and Swart, Pieter. c2004-2018. networkx.algorithms.centrality.closeness_centrality. [Online]. [Accessed 15 October 2018]. Available from: https://networkx.github.io/

Hagberg, A., Schult, D. and Swart, Pieter. c2015. closeness_centrality. [Online]. [Accessed 15 October 2018]. Available from: https://networkx.github.io/

Hagberg, A., Schult, D. and Swart, Pieter. 2008. Exploring network structure, dynamics, and function using NetworkX. In: Varoquaux, G., Vaught, T. and Millman, J. eds. *Proceedings of the 7th Python in Science Conference (SciPy 2008), 19-24 August 2008, Pasadena, CA.* [Online]. SciPy, pp.11-16. [Accessed 17 December 2018]. Available from: https://conference.scipy.org/

Hakim, S. 1980. The attraction of property crimes to suburban localities: A revised economic model. *Urban Studies.* **17**(3), pp.265-276.

Hansen, W. G. 1959. How accessibility shapes land use. *Journal of the American Institute of Planners.* **25**(2), pp.73-76.

Harding, J. 2020. Keyless car theft – why aren't car manufacturers doing more? *Which?* [Online]. 03 May 2020. [Accessed 02 September 2020]. Available from: https://www.which.co.uk/news/2020/05/keyless-car-theft-why-arent-car-manufacturers-doing-more/

Hirschfield, A., Birkin, M., Brunsdon, C., Malleson, N. and Newton, A. 2014. How places influence crime: The impact of surrounding areas on neighbourhood burglary rates in a British city. *Urban Studies.* **51**(5), pp.1057-1072.

Hirschfield, A. and Bowers, K. J. 1997. The development of a social, demographic and land use profiler for areas of high crime. *British Journal of Criminology.* **37**(1), pp.103-120.

Holder, B. 2020. Call for CCTV footage after a car key burglary in Amblecote. *Stourbridge News*. [Online]. 09 August 2020. [Accessed 24 September 2020]. Available from: https://www.stourbridgenews.co.uk/news/18639122.call-cctv-footage-car-key-burglary-amblecote/

Home Office. 2015. *Home Office counting rules for recorded crime – burglary (with effect from April 2015)*. [Online]. London: Home Office. [Accessed 01 February 2016].

Home Office. ca. 2018. About data.police.uk. [Online]. [Accessed 07 January 2020]. Available from: https://data.police.uk/about/

Hu, T. and Tripathi, A. 2017. The performance evaluation of machine learning classifiers. In: Fan, M., Heikkilä, J., Li, H., Shaw, M. J. and Zhang, H. eds. *Internetworked World: 15th Workshop on e-Business, WeB 2016, Dublin, Ireland, December 10, 2016, Revised Selected Papers.* Basel: Springer International Publishing, pp.74-83.

ITV. 2020. *Man stole 10 keyless cars in Peterborough crime spree*. [Online]. 06 April 2020. [Accessed 16 September 2020]. Available from: https://www.itv.com/news/anglia/2020-04-06/man-stole-10-keyless-cars-in-peterborough-crime-spree

Johnson, S. D., Bowers, K. and Hirschfield, A. 1997. New insights into the spatial and temporal distribution of repeat victimization. *British Journal of Criminology.* **37**(2), pp.224-241.

Johnson, S. D. and Bowers, K. J. 2004. The burglary as clue to the future: The beginnings of prospective hot-spotting. *European Journal of Criminology.* **1**(2), pp.237-255.

Johnson, S. D., Bernasco, W., Bowers, K. J., Elffers, H., Ratcliffe, J., Rengert, G. and Townsley, M. 2007. Space–time patterns of risk: A cross national assessment of residential burglary victimization. *Journal of Quantitative Criminology.* **23**(3), pp.201-219.

Johnson, S. D., Birks, D. J., McLaughlin, L., Bowers, K. J. and Pease, K. 2007b. *Prospective crime mapping in operational context: Final report.* [Online]. London: Home Office. [Accessed 17 December 2019]. Available from: http://library.college.police.uk/docs/hordsolr/ rdsolr1907.pdf

Johnson, S. D., Bowers, K. J., Birks, D. J. and Pease, K. 2009. Predictive mapping of crime by ProMap: Accuracy, units of analysis, and the environmental backcloth. In: Weisburd, D., Bernasco, W. and Bruinsma, G. J. N. eds. *Putting crime in its place.* New York: Springer, pp.171-198.

Johnson, S. D. and Bowers, K. J. 2010. Permeability and burglary risk: Are cul-de-sacs safer? *Journal of Quantitative Criminology.* **26**(1), pp.89-111.

Kassambara, A. 2018. *Knn: K-nearest neighbors essentials*. [Online]. [Accessed 17 February 2020]. Available from: http://www.sthda.com/english/articles/35-statistical-machine-learning-essentials/142-knn-k-nearest-neighbors-essentials/

Keeling, N. 2018. Thieves are 'hacking' keyless cars. These are the models at risk - and the simple thing you can do to protect yourself. *Manchester Evening News*. [Online]. 08 November 2018. [Accessed 02 September 2020]. Available from: https://www.manchestereveningnews.co.uk/news/greater-manchester-news/keyless-car-relay-theft-advice-14496158

Keighley News. 2020. Audi and Polo targeted in Hanoi-style burglary in Cross Hills. *Keighley News*. [Online]. 16 July 2020. [Accessed 24 September 2020]. Available from: https://www.keighleynews.co.uk/news/18586366.audi-polo-targeted-hanoi-style-burglary-cross-hills/

Kelechava, M. 2019. *Using LDA topic models as a classification model input*. [Online]. [Accessed 02 July 2019]. Available from: https://towardsdatascience.com/unsupervised-nlp-topic-models-as-a-supervised-learning-input-cf8ee9e5cf28

Kent, J., Leitner, M. and Curtis, A. 2006. Evaluating the usefulness of functional distance measures when calibrating journey-to-crime distance decay functions. *Computers, Environment and Urban Systems.* **30**(2), pp.181-200.

Knox, G. 1964. Epidemiology of childhood Leukaemia in Northumberland and Durham. *British Journal of Preventive and Social Medicine.* **18**(1), pp.17-24.

Krishnaveni, G. and Sudha, T. 2017. Naïve Bayes text classification: A comparison of event models. *Imperial Journal of Interdisciplinary Research (IJIR).* **3**(1), pp.290-294.

Kuhn, M. 2019. *The caret package*. [Online]. [Accessed 16 February 2020]. Available from: https://topepo.github.io/caret/

Lansley, G. 2016. Cars and socio-economics: understanding neighbourhood variations in car characteristics from administrative data. *Regional Studies, Regional Science.* **3**(1), pp.264-285.

LearnByExample.org. c2019. *R scatter plot – ggplot2*. [Online]. [Accessed 14 February 2020]. Available from: https://www.learnbyexample.org/r-scatter-plot-ggplot2/

Levesley, T., Braun, G., Wilkinson, M. and Powell, C. c2004. *Emerging methods of car theft – theft of keys.* (Findings 239). London: Research, Development and Statistics Directorate, Home Office.

Mahfoud, M., Bhulai, S., van der Mei, R., Erkin, D. and Dugundji, E. 2018. Forecasting burglary risk in small areas via network analysis of city streets. In: Bhulai, S., Kardaras, D. and Semanjski, I. eds. *DATA ANALYTICS 2018: The Seventh International Conference on Data Analytics, 18-22 Nov 2018*, *Athens, Greece*. Wilmington, DE: IARIA, pp.109-114.

Malleson, N. 2010. *Agent-based modelling of burglary*. Ph.D. thesis, University of Leeds.

Manning, C. D., Raghavan, P. and Schütze, H. 2009. *Introduction to information retrieval.* [Online]. Cambridge: Cambridge University Press. [Accessed 02 July 2019]. Available from: https://nlp.stanford.edu/IR-book/pdf/irbookonlinereading.pdf

McDonald, J. H. 2014. *Handbook of biological statistics*. 3rd ed. Baltimore, MD: Sparky House Publishing.

Mohler, G. O., Short, M. B., Brantingham, P. J., Schoenberg, F. P. and Tita, G. E. 2011. Self-exciting point process modeling of crime. *Journal of the American Statistical Association*. **106**(493), pp.100-108.

Morgan, N., Shaw, O., Feist, A. and Byron, C. 2016. *Reducing criminal opportunity: Vehicle security and vehicle crime*. (Research Report 87). [Online]. London: Home Office. [Accessed 04 September 2020]. Available from:

https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/489097/horr87.pdf

NCA. c2019. *National strategic assessment of serious and organised crime (SOC) 2019*. [Online]. NCA. [Accessed 14 September 2020]. Available from: https://www.nationalcrimeagency.gov.uk/who-we-are/publications/296-national-strategic-assessment-of-serious-organised-crime-2019/file

Newing, A., Clarke, G. and Clarke, M. 2014. Developing and applying a disaggregated retail location model with extended retail demand estimations. *Geographical Analysis.* **47**(3), pp.219-239.

Norman, P. [no date]. *Unit 3 practical: Accessing and using census data.* [Handout for lecture practical accessed through Minerva]. GEOG5240 Applied Population & Demographic Analysis. University of Leeds.

North Yorkshire Police. 2018. *Local policing operation put in gear after York '2-in-1' burglaries*. [Online]. [Accessed 18 April 2019]. Available from: https://northyorkshire.police.uk/news/local-policing-operation-put-in-gear-after-york-2-in-1-burglaries/

North Yorkshire Police. 2019. *A car key burglary is where a criminal breaks into your home with the specific aim of removing your car* […]. [Facebook]. 12 March. [Accessed 18 May 2019]. Available from: https://www.facebook.com/NorthYorkshirePolice/photos/a.280174491069/10156844362426070/?type=3

O'Brien, O. and Cheshire, J. 2016. Interactive mapping for large, open demographic data sets using familiar geographical features. *Journal of Maps.* **12**(4), pp.676-683.

ONS. 2014. *Households and household composition in England and Wales: 2001-11.* [Online]. [Accessed 03 March 2016]. Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/families/articles/householdsandhouseholdcompositioninenglandandwales/2014-05-29

ONS. 2015. *Radial plots for the 2011 Area Classification for Super Output Areas*. [Online]. [Accessed 27 October 2016]. Available from: https://www.ons.gov.uk/methodology/geography/geographicalproducts/areaclassifications/2011areaclassifications/penportraitsandradialplots

ONS. 2015b. *Pen portraits for the 2011 Area Classification for Super Output Areas*. [Online]. [Accessed 27 October 2016]. Available from: https://www.ons.gov.uk/methodology/geography/geographicalproducts/areaclassifications/2011areaclassifications/penportraitsandradialplots

ONS. ca. 2016. *Census geography: An overview of the various geographies used in the production of statistics collected via the UK census.* [Online]. [Accessed 15 January 2020]. Available from: https://www.ons.gov.uk/methodology/geography/ukgeographies/censusgeography

ONS. ca. 2016b. *The national statistics socio-economic classification (NS-SEC).* [Online]. [Accessed 13 January 2017]. Available from: https://www.ons.gov.uk/methodology/classificationsandstandards/otherclassifications/thenationalstatisticssocioeconomicclassificationnssecrebasedonsoc2010

ONS. 2017. *Information note: Population weighted centroids guidance.* [Online]. [Accessed 04 September 2018]. Available from: https://data.gov.uk/dataset/2c5695f2-39d0-457f-a03c-1f4d3617bb48/population-weighted-centroids-guidance

ONS. 2017b. *Methodology note on production of small area population estimates.* [Online]. [Accessed 04 September 2018]. Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/methodologies/methodologynoteonproductionofsmallareapopulationestimatesoctober2017

ONS. 2017c. *Best-fit policy.* [Online]. [Accessed 04 September 2018]. Available from: http://webarchive.nationalarchives.gov.uk/

ONS. 2017d. *Overview of burglary and other household theft: England and Wales (LATEST RELEASE AS AT 24/04/19).* [Online]. [Accessed 06 April 2018]. Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/crimeandjustice/articles/overviewofburglaryandotherhouseholdtheft/englandandwales

ONS. 2020. Crime in England and Wales, year ending March 2020: Appendix tables. *ONS*. [Online]. [Accessed 22 September 2020]. Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/crimeandjustice/datasets/crimeinenglandandwalesappendixtables

ONS. 2020b. Nature of crime: Vehicle-related theft. *ONS*. [Online]. [Accessed 25 September 2020]. Available from:

https://www.ons.gov.uk/peoplepopulationandcommunity/crimeandjustice/datasets/natureofcrimevehiclerelatedtheft

Pan, L., Yu, S. and Ma, F. 2004. SIMON: A multi-strategy classification approach resolving ontology heterogeneity on the semantic web. In: Yu, J. X., Lin, X., Lu, H. and Zhang, Y. eds. *Advanced Web Technologies and Applications. APWeb 2004. Lecture Notes in Computer Science, vol 3007.* Berlin: Springer-Verlag, pp.79-88.

Perlman, G. 1984. Natural artificial languages: Low level processes. *International Journal of Man-Machine Studies.* **20**(4), pp.373-419.

Polvi, N., Looman, T., Humphries, C. and Pease, K. 1991. The time course of repeat burglary victimization. *British Journal of Criminology.* **31**(4), pp.411-414.

Porta, S., Strano, E., Iacoviello, V., Messora, R., Latora, V., Cardillo, A., Wang, F. and Scellato, S. 2009. Street centrality and densities of retail and services in Bologna, Italy. *Environment and Planning B: Urban Analytics and City Science.* **36**(3), pp.450-465.

QGIS. [no date]. *Heatmap plugin*. [Online]. [Accessed 24 September 2019]. Available from: https://docs.qgis.org/2.18/en/docs/user_manual/plugins/plugins_heatmap.html

Quetelet, M. A. 1842. *A treatise on man and the development of his faculties*. [Online]. Edinburgh: W. and R. Chambers. [Accessed 18 March 2020]. Available from: https://books.google.co.uk/books?id=X95DAQAAMAAJ

Raschka, S. 2014. *About feature scaling and normalization – and the effect of standardization for machine learning algorithms*. [Online]. [Accessed 15 February 2020]. Available from: https://sebastianraschka.com/Articles/2014_about_feature_scaling.html

Raschka, S. and Mirjalili, V. 2017. *Python machine learning*. [Online]. 2nd ed. Birmingham: Packt Publishing. [Accessed 15 February 2020]. Available from: https://books.google.co.uk/books?id=_plGDwAAQBAJ

Raschka, S. 2018. *STAT 479: Machine learning lecture notes*. [Online]. Madison, WI: Department of Statistics, University of Wisconsin-Madison. [Accessed 14 February 2020]. Available from: https://sebastianraschka.com/pdf/lecture-notes/stat479fs18/02_knn_notes.pdf

Ratcliffe, J. H. and McCullagh, M. J. 1998. Aoristic crime analysis. *International Journal of Geographical Information Science.* **12**(7), pp.751-764.

Ratcliffe, J. and McCullagh, M. 1999. Burglary, victimisation, and social deprivation. *Crime Prevention and Community Safety.* **1**(2), pp.37-46.

Ratcliffe, J. H. and McCullagh, M. J. 2001. Crime, repeat victimisation and GIS. In: Hirschfield, A. and Bowers, K. eds. *Mapping and analysing crime data: Lessons from research and practice.* London: Taylor & Francis, pp.61-92.

Ratcliffe, J. H. 2000. Aoristic analysis: The spatial interpretation of unspecific temporal events. *International Journal of Geographical Information Science.* **14**(7), pp.669-679.

Ratcliffe, J. H. 2002. Aoristic signatures and the spatio-temporal analysis of high volume crime patterns. *Journal of Quantitative Criminology.* **18**(1), pp.23-43.

Ratcliffe, J. H. 2016. *Intelligence-led policing*. 2nd ed. Abingdon: Routledge.

Ratcliffe, J. H., Taylor, R. B. and Perenzin, A. 2016. *Predictive modeling combining short and long-term crime risk potential: Final report.* [Online]. Philadelphia, PA: Temple University, Center for Security and Crime Science. [Accessed 28 March 2018]. Available from: https://www.ncjrs.gov/pdffiles1/nij/grants/249934.pdf

Ratcliffe, J. H., Taylor, R. B. and Perenzin, A. 2016b. *PROVE manual.* [Online]. Philadelphia, PA: Temple University, Center for Security and Crime Science. [Accessed 28 March 2018]. Available from: https://www.ncjrs.gov/pdffiles1/nij/grants/249934.pdf

Ratcliffe, J. H. ca. 2009. *Near repeat calculator: Program manual for Version 1.3.* [Online]. Philadelphia, PA: Temple University. [Accessed 30 January 2017]. Available from: https://www.temple.edu/cj/misc/nr/ (NB Page not found as at 06/04/20)

Ratcliffe, JH, Near Repeat Calculator (version 1.3). Temple University, Philadelphia, PA and the National Institute of Justice, Washington, DC. May 2008?

Reinhart. A. 2018. A review of self-exciting spatio-temporal point processes and their applications. *Statistical Science*. **33**(3), pp. 299-318.

Rodrigue, J.-P. and Ducruet, C. ca. 2017. *Graph theory: Definition and properties.* [Online]. [Accessed 18 January 2019]. Available from: https://transportgeography.org/

Rogerson, M. 2016. *The utility of applying textual analysis to descriptions of offender modus operandi for the prevention of high volume crime*. Ph.D. thesis, University of Huddersfield.

Rosser, G., Davies, T., Bowers, K. J., Johnson, S. D. and Cheng, T. 2017. Predictive crime mapping: Arbitrary grids or street networks? *Journal of Quantitative Criminology.* **33**(3), pp.569-594.

Sampson, R. J. and Groves, W. B. 1989. Community structure and crime: Testing social-disorganization theory. *American Journal of Sociology.* **94**(4), pp.774-802.

Sampson, R. J. and Wooldredge, J. D. 1987. Linking the micro- and macro-level dimensions of lifestyle-routine activity and opportunity models of predatory victimization. *Journal of Quantitative Criminology.* **3**(4), pp.371–393.

Sampson, R. J., Raudenbush, S. W. and Earls, F. 1997. Neighbourhoods and violent crime: A multilevel study of collective efficacy. *Science*. **277**(5328), pp.918-924.

Samuels, P. and Marshall, E. [no date]. *Checking normality in SPSS*. [Online]. [Accessed 19 October 2017]. Available from: https://www.sheffield.ac.uk/polopoly_fs/1.579181!/file/stcp-marshallsamuels-NormalityS.pdf

Samuels, P. and Marshall, E. [no dateb]. *Kruskal-Wallis in SPSS*. [Online]. [Accessed 19 October 2017]. Available from: https://www.sheffield.ac.uk/polopoly_fs/1.714567!/file/stcp-marshall-KruskalSPSS.pdf

Sanders, W. B. 1943. Juvenile delinquency and urban areas. *Social Forces*. **21**(4), pp.487-488.

Saxena, R. 2017. *Knn r, k-nearest neighbor implementation in r using caret package*. [Online]. [Accessed 17 February 2020]. Available from: https://dataaspirant.com/2017/01/09/knn-implementation-r-using-caret-package/

School of Geography, UOL. [no date]. *Stepwise linear regression*. [Online]. [Accessed 03 February 2020]. Available from: http://www.geog.leeds.ac.uk/courses/other/statistics/spss/stepwise/

Scikit-learn. c2007-2019. *3.1. Cross-validation: Evaluating estimator performance*. [Online]. [Accessed 17 February 2020]. Available from: https://scikit-learn.org/stable/modules/cross_validation.html

Shaw, C. R. and McKay, H. D. 1942. *Juvenile delinquency and urban areas*. Chicago: University of Chicago Press.

Shaw, S. E., Smith, L. L. and Bond, J. W. 2010. Examining the factors that differentiate a car key burglary from a regular domestic burglary. *International Journal of Police Science & Management.* **12**(3), pp.450-459.

Shimodaira, H. 2015. *Text classification using Naive Bayes. Learning and data note 7: Informatics 2B.* [Online]. University of Edinburgh. [Accessed 20 June 2019). Available from: https://www.inf.ed.ac.uk/teaching/courses/inf2b/learnnotes/inf2b-learn07-notes-nup.pdf

Singh, A. 2018. *A practical introduction to k-nearest neighbors algorithm for regression (with Python code)*. [Online]. [Accessed 13 February 2020]. Available from: https://www.analyticsvidhya.com/blog/2018/08/k-nearest-neighbor-introduction-regression-python/

Smith, T. S. 1976. Inverse distance variations for the flow of crime in urban areas. *Social Forces.* **54**(4), pp.802-815.

Smith, T., Noble, M., Noble, S., Wright, G., McLennan, D. and Plunkett, E. 2015. *The English indices of deprivation 2015: Research report.* London: Department for Communities and Local Government.

Snook, B. 2004. Individual differences in distance travelled by serial burglars. *Journal of Investigative Psychology and Offender Profiling.* **1**(1), pp.53-66.

Steele B, Thornton A, McKillop C. and Dover, H. 2001. *The formulation of a strategy to prevent and detect distraction burglary offences against older people*. [Online]. London: Home Office. [Accessed 01 October 2020]. Available from: https://www.tradingstandards.uk/media/documents/news--policy/research/strategy_burglary.pdf

Tonkin, M., Grant, T. and Bond, J. W. 2008. To link or not to link: A test of the case linkage principles using serial car theft data. *Journal of Investigative Psychology and Offender Profiling.* **5**(1-2), pp.59-77.

Tseloni, A. and Pease, K. 2003. Repeat personal victimization. *British Journal of Criminology.* **43**(1), pp.196-212.

Tseloni, A. and Tilley, N. 2016. Choosing and using statistical sources in criminology: What can the Crime Survey for England and Wales tell us? *Legal Information Management.* **16**(2), pp.78-90.

Trickett, A., Osborn, D. R. and Ellingworth, D. 1995. Property crime victimisation: The roles of individual and area influences. *International Review of Victimology.* **3**(4), pp.273-295.

Tseloni, A., Osborn, D. R., Trickett, A. and Pease, K. 2002. Modelling property crime using the British Crime Survey: What have we learnt? *British Journal of Criminology.* **42**(1), pp.109-128.

Upton, G. and Cook, I. 2014. A dictionary of statistics. 3rd ed. Oxford: Oxford University Press.

Vandeviver, C., Van Daele, S. and Vander Beken, T. 2015. What makes long crime trips worth undertaking? Balancing costs and benefits in burglars' journey to crime. *British Journal of Criminology.* **55**(2), pp.399-420.

Wasserman, S. and Faust, K. 1994. *Social network analysis: Methods and applications.* Cambridge, UK: Cambridge University Press.

West Bridgford Wire. 2020. Police warn Rushcliffe residents to be vigilant after 'hook and cane' car key burglary. *West Bridgford Wire*. [Online]. 28 May 2020. [Accessed 24 September 2020]. Available from: https://westbridgfordwire.com/police-warn-rushcliffe-residents-to-be-vigilant-after-hook-and-cane-car-key-burglary

West Midlands Police. 2019. *Tackling serious acquisitive crime*. [Online]. West Midlands Police and Crime Commissioner. [Accessed 04 September 2020]. Available from: https://cl-assets.public-i.tv/westmidspcc/document/SPCB_190219___Item_8c___Tackling_Serious_Aquisitive_Crime.pdf

West Midlands Police. 2019b. Car key burglary (2239/19): Crimes broken down by year force and Black Country NPU for burglary where the search has returned vehicles / keys for the years 2014 - 2018. *West Midlands Police*. [Online]. [Accessed 03 September 2020]. Available from: https://foi.west-midlands.police.uk/car-key-burglary-2239-19/

West Yorkshire Police. 2019. *Hanoi burglaries - crime prevention advice*. [Online]. [Accessed 15 February 2020]. Available from: https://www.wypcommunityalert.co.uk/da/286286/Hanoi_Burglaries_-_Crime_Prevention_Advice.html

West Yorkshire Police. 2019b. *news/ appeals – West Yorkshire Police*. [Online]. [Accessed 02 July 2019]. Available from: https://www.westyorkshire.police.uk/news-appeals/police-seek-publics-help-reducing-demand-999-and-101-services

West Yorkshire Police. 2020. *Car key burglaries*. [Online]. [Accessed 25 February 2020]. Available from: https://www.westyorkshire.police.uk/advice/home-security/home-security/car-key-burglaries

White, R. C. 1932. The relation of felonies to environmental factors in Indianapolis. *Social Forces.* **10**(4), pp.498-509.

Wortley, R. and Mazerolle, L. 2008. Environmental criminology and crime analysis: Situating the theory, analytic approach and application. In: Wortley, R. and Mazerolle, L. eds. *Environmental criminology and crime analysis.* Cullompton: Willan Publishing, pp.1-18.

Wrexham.com. 2020. Police warning after “car key burglary” in Wrexham. *Wrexham.com*. [Online]. 13 August 2020. [Accessed 24 September 2020]. Available from: http://www.wrexham.com/news/p olice-warning-after-car-key-burglary-in-wrexham-191279.html

Youstin, T. J., Nobles, M. R., Ward, J. T. and Cook, C. L. 2011. Assessing the generalizability of the near repeat phenomenon. *Criminal Justice and Behavior.* **38**(10), pp.1042-1063.