Criminal Activity Spaces and Crime Linkage Analysis:

Towards a Suspect Prioritization System

Benjamin Louis Goldy

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

The University of Huddersfield

October 2018

Copyright Statement

Abstract

Offender spatial behaviour has been utilized extensively by practitioners and researchers alike as a vector to study criminal behaviour. Psychologists and criminologists alike have been utilizing spatial behaviour as a means to describe crime patterns, which has resulted in several areas of specific research. Several prominent theories of criminal behaviour resulted from this work including routine activity theory and crime pattern theory. This thesis builds upon past research by reintroducing activity space as a means to study criminal spatial movement and the progression of criminality in general

First, this thesis addresses the lack of specific methodologies for a standard estimation process for criminal activity spaces. There is no consensus among the existent activity space literature as to how to calculate such spaces, which is further compounded by the unique challenges associated with working with crime data. As such, the current thesis introduces methods for estimating activity spaces as a single or composite geometric surface given the entirety of a given offenders criminal history. From this standardized activity space construct, general properties of spatial involvement (size, dispersion) were calculated as well as the proportion of offenders for whom the activity space for a given number of crimes included the next temporally sequential crime in their series.

Furthermore, this thesis demonstrates the utility of an activity space based framework in studying criminal spatial behaviour through direct example via crime linkage analysis. Activity spaces for all serial offenders (n = 997) within the available sample were calculated and used within a crime linkage exercise. Results indicate that such methods can be greatly improved by incorporating specific activity space based measures via two distinct studies: the first utilizing

logistic regression models in accordance with past research; the second employing a more sophisticated machine learning technique of random forest modelling.

Finally, this thesis demonstrates how activity spaces can be used to extend crime linkage methods away from the traditional crime-focused approach to an offender-focused approach. Past studies of crime linkage analysis have used spatial proximity to assess strict crime-to-crime relationships, however results from a suspect prioritization simulation indicated that an activity space-based one-to-many comparison resulted in significantly improved suspect prioritization performance: traditional models correctly prioritized the correct offender into the top five suspects 24% of the time compared to 32% of the time for activity space-based approaches. This body of work serves to illustrate not just the suitability of the proposed method of modelling criminal activity spaces, but also the theoretical implications contained therein.

Table of Contents

Total Word Count: 58,715

List of Figures

List of Tables

Chapter 1

Spatial Cognition and Criminal Activity Space

1.1 Introduction

In 1855 John Snow published *On the Mode of Communication of Cholera* wherein he described how observed cases of cholera were traced back to a single contaminated well pump in Broad Street, London. In his words: "…it will be observed that the deaths either very much diminished, or ceased altogether, at every point where it becomes decidedly nearer to send to another pump than to the one in Broad Street" (Snow, 1855, p. 47). Thus cholera related deaths were observed to cluster around a particular pump, and as one moved away from this pump two things happened. First, the number of deaths decreased, and second the likelihood an individual used a different, nearer, pump increased. Snow's work highlights a fundamental spatial relationship which has been observed in a number of behavioural domains including criminal activity.

Brantingham and Brantingham's criminality of place (1981) first popularized the notion of a 'criminal activity space' whereby criminal activity could be conceptualized as a by-product of an individual's daily routine (Cohen & Felson, 1979). Much like how individuals would travel to the nearest pump for water in Snow's study, offenders were thought to commit crime in and around their day-to-day activity centres. Conceptually, both of these phenomena can be related through Zipf's principle of least effort (Zipf, 1949), which suggests that individuals will choose the most efficient path when completing a given task. In Snow's case, individuals exhibited least effort by travelling to the nearest well. Criminals, on the other hand, exhibit least effort by offending in and around their domicile or places of leisure rather than traveling far afield.

Sometimes referred to as propinquity, (Canter & Youngs, 2009), this observation has been noted in the journey to crime literature under the broad term of ‘distance decay’. Distance decay describes the negative relationship between the likelihood of an offender engaging in criminal activity and travel distance required to commit said crime (O’Leary, 2011). Recently, notions of distance decay have been found to be highly indicative of linked crime series within the crime linkage literature (Slater, Woodhams, & Hamilton-Giachritsis, 2015; M. Tonkin, Woodhams, Bull, Bond, & Palmer, 2011; Woodhams & Bennell, 2015), which is an area of research focused on developing methods for identifying sets of crimes which share a common perpetrator (Woodhams & Bennell, 2015).

While both crime pattern theory and crime linkage analysis each describe criminal spatial patterns, they have not as of yet been explored in tandem. Activity spaces from crime pattern theory have been used conceptually to describe how offenders may encounter their victims, which in turn give rise to the various observed patterns of crime *in aggregate*. There has not, however, been any formal empirical work on individual criminal activity spaces and how they may develop or relate to future crime. Similarly, crime linkage analysis has focused on identifying behavioural domains that can be used to differentiate linked crime from unlinked crime *in aggregate*. Identifying methods for linking crimes to specific individuals has thus far been under-researched.

This thesis seeks to address these shortcomings by addressing two distinct areas of focus. The first is exploring activity spaces and how they may be conceptualized and subsequently empirically measured or otherwise approximated. The second is utilizing activity spaces as a means for exploring individual variation within a crime linkage framework. The predictive models typically employed in crime linkage analysis are derived from aggregate trends with little to no acknowledgement of the individuals who drive those variations. Furthermore, there is a common practice within the crime linkage literature to employ data sets that are limited to either a single crime type or a very small subset of crime types (Tonkin & Woodhams, 2015; Woodhams & Bennell, 2015). Chapter 1, in part, challenges this practice by introducing the major theories surrounding criminal activity spaces specifically and general spatial cognition in general. This will provide the necessary framework from which to begin a discussion surrounding how to construct formal measures for activity spaces and beyond.

1.2 Spatial Morphology: An Activity Space Perspective

The concept of an “activity space” has been of interest to researchers and practitioners from as early as the 1950s (Lewin, 1951; Ren, 2016), and can be thought of as a means to describe how and where an individual interacts with their environment. In that time, there has been little comprehensive research done in the specific application of activity spaces to the study of crime. The significance of criminal spatial patterns have long been understood and documented in a number of contexts including: sexual assault, rape, burglary, general property crime, and serial murder to name a few (Bennell, Bloomfield, Snook, Taylor, & Barnes, 2010; D. Canter & Larkin, 1993; Canter, 1995; Canter & Hammond, 2006; Hazelwood & Warren, 2003; Lundrigan, Czarnomski, & Wilson, 2010; Slater et al., 2015; Snook, Cullen, Mokros, & Harbort, 2005; Van Daele & Vander Beken, 2011a). Despite these consistent findings of the importance of location of criminal activity, there have been few studies that attempt to quantify criminal behaviour formally in terms of a “criminal activity space”.

The activity space framework can have its roots traced back to Lewin’s field theory (1951), which describes behaviour in terms of a “life space”. This life space, Lewin argued, is contingent on the totality of personal and environmental factors from which it is built, and summarises the possible extent of behaviours available. More modern interpretations of activity space are more conservative in their formulations, but maintain the importance of the person-to-environment relationship. Horton and Reynolds, for example, defined activity space as the "subset of all urban locations with which the individual has direct contact as the result of day-to-day activities" (Horton & Reynolds, 1971, p. 37). Over the course of the last 60 years, researchers have proposed a number of specific definitions for what an activity space is, and in general these working definitions all fall within a similar vein (Horton & Reynolds, 1971; Johnston, 1972; Lee, Davis, Yoon, & Goulias, 2016; Perchoux, Chaix, Cummins, & Kestens, 2013). Namely, activity space is related to, and defined by, a geographical representation of an individual's spatial behaviour over time.

The activity space construct has traditionally been described in terms of routine activities. Perchoux et al. outline: "The activity space, in reflecting daily mobility, is an individual measure of spatial behaviour" (Perchoux et al., 2013, p. 88). Past studies of activity space have utilised travel diaries and geospatial data in an effort to estimate where and how often individuals travel to any given area (some examples: Dijst, 1999; Lee et al., 2016; Sherman, Spencer, Preisser, Gesler, & Arcury, 2005; Wong & Shaw, 2011). In this way, activity spaces serve as a means to quantify an individual's interaction with their environment. Activity spaces have been used to explore such questions as inter-urban movement and familiarity (Brown & Moore, 1970; Horton & Reynolds, 1971), residential preference (Johnston, 1972), daily activity (Lee et al., 2016), to evaluate segregation (Wong & Shaw, 2011) and as the foundation for a measure of exposure to foodscapes (Kestens, Lebel, Daniel, Thériault, & Pampalon, 2010) to list a few examples.

Given the discussion so far, activity spaces can be described as a means quantifying individual morphology. Studies that employ activity space measures do so by comparing differences between individual activity spaces – either in terms of size, exposure or some other quantifiable metric. However the literature is not clear on how such activity spaces can be or should be measured. Furthermore the literature is undecided as to how activity spaces can be formally applied in criminal contexts. From here it is useful to introduce the specific views on criminal spatial movement and how they may relate to the activity space construct.

1.3 Offender Spatial Movement

One of the most consistent and often perplexing findings within the criminal geography literature is what Canter and Youngs (2009) describe as the "locatedness" of crime. That is that for whatever reason, crimes seem to occur in a relatively predictable manner in and around offenders' homes. Such a consistent finding is often portrayed as counter-intuitive: as Canter and Youngs point out: "at the very least, many would assume that criminals would take care to ensure that the place they choose to offend would have no obvious association with other aspects of their lives, especially with such crucial information as where they live." Yet, despite such conventional wisdom, evidence has mounted in a number of areas that indicates the exact opposite: the locations where offenders choose to commit their crimes have some relationship to their non-criminal lives; including where they live.

This observation has not gone unnoticed, however, and serves as the foundation upon which many areas of criminal behaviour research are focused. These include: journey-to-crime (Bernasco, 2010; Van Daele & Vander Beken, 2011b), crime linkage (Bennell & Canter, 2002; Slater et al., 2015; M. Tonkin et al., 2011), and geographic offender profiling (D. Canter, Hammond, Youngs, & Juszczak, 2013; Canter & Youngs, 2009; Rossmo, 2000). While many of these areas leverage the observation of "locatedness", few go into detail regarding why such patterns are observed. This is the cause without the effect and is most evident in areas of applied research like crime linkage analysis. In crime linkage analysis, the proximity of crimes is often reported as the most stable predictor; yet there is rarely ever any investigation as to why such patterns occur or under what conditions such relationships would not apply (if any).

At its core, the discussion surrounding the study of offender spatial behaviour is one that revolves around two areas of focus: aggregate - or nomothetic - approaches versus individual - or ideographic - approaches. One of the most widely reported findings within offender geography is the distance decay phenomena: as distance from the home increases the likelihood of an offender committing a crime decreases. These studies often report similar distance-decay functions which serve to describe behaviour at the aggregate level. Such studies often take the nomothetic or aggregate perspective. However, it is well understood that individual variation can be “hidden” within such aggregate trends, and there is some evidence to suggest that similar masking is occurring within offender spatial research (Townsley & Sidebottom, 2010). Similarly, Canter et al. (2013) presented work that evaluated the efficacy of ideographic models built from an individual’s crime history only. They found that the ideographic approach was comparable in terms of precision while also providing smaller search areas on average. The implication being that different offenders utilize their surrounding environment differently, and that tapping into these differences my produce more representative models of specific offender spatial behaviour (D. Canter et al., 2013). The question then becomes how best to model such individual differences and preferences in a practical and meaningful way.

Canter and Youngs (2009) present two key concepts for describing spatial movement that may be of use in describing spatial tendencies: (1) propinquity and (2) morphology. Propinquity corresponds to the “locatedness” observation mentioned previously and literally means proximity. The propinquity argument encapsulates the idea that offenders travel relatively short distances to commit crimes. The propinquity observation has culminated in widespread support for the “distance-decay” observation of criminal range (Emeno & Bennell, 2013; Kent, Leitner, & Curtis, 2006; O’Leary, 2011). The distance decay function represents the likelihood of committing a crime as a function of distance. That is, as distance travelled (from home for example) increases, the likelihood of offending decreases. As previously mentioned, the propinquity property of crime has been exploited in a number of areas of research including: geographic offender profiling where it serves as the basis for constructing probability surfaces for offender home locations (Canter & Hammond, 2006; Canter et al., 2013; Rossmo, 2000); as well as crime-linkage analysis where inter-crime distance has been routinely demonstrated to be among the most consistent predictors of same-offender crime (Bennell & Canter, 2002; M. Tonkin et al., 2011). Thus propinquity can be understood to represent the range any given offender is willing or able to travel in order to commit their crimes.

If propinquity can be thought of as the average tendency of criminal range across a population – as again it is usually an aggregate measure, then morphology corresponds to the individual variations of target selection. In other words, morphology can be thought to describe the shape or pattern in space a specific offender’s crime series takes. Put another way, if propinquity describes the magnitude of criminal range (i.e. travel distance), then morphology describes the direction.

Morphology as a parameter of spatial movement occupies an ambiguous space within criminal geography research. Many studies that look at spatial relationships, such as crime linkage analysis, often ignore specific offender morphologies and rely solely on the aggregate propinquity relationships. Research into criminal and non-criminal individual environmental use alike highlights the difficulty in capturing individual morphologies in a systematic and measurable way (Chaix et al., 2012). Geographic profiling work, for example, highlights the challenges in differentiating between so-called “commuters” and “marauders”.

First proposed by Canter and Larkin (1993), and later expanded upon by a number of researchers (Paulsen, 2007; Rossmo, 2000), the commuter/marauder typology highlights two distinct spatial movement patterns of offenders. The two proposed types correspond with explicit differences in morphology: marauders offend outward from a base in all directions, whereas commuters travel to a specific area. This typology is of particular importance as it highlights the difficulties associated with generalizing spatial movement of individuals: the propinquity observation tends to describe marauders much better than it does commuters (Canter & Larkin, 1993; Rossmo, 2000; Paulsen, 2007). This is because the relationship between the home, or “base”, does not appear to be the same for both types, which would indicate a different underlying decision process at play.

Furthermore, several other morphology centric observations have been made with regard to offender spatial behaviour that are related to this typology: for example, buffer zones have been suggested as a means to explain instances whereby offenders travel outward from a base to offend, while simultaneously avoiding offending within a certain area of said base (Rossmo, 2000).This observation is somewhat at odds with the pure propinquity observation which would suggest that an offender would be most likely to offend in and around their base. Criminological theory (covered in section 1.4) argues that offender morphology is necessarily tied to the underlying environment (or “backcloth” – see 1.4.3). For example, commuters necessarily have to show some directional bias – otherwise they would be marauders. While such directional biases have been postulated to exist for some time (Brantingham et al., 1981), empirical works demonstrating the effect have only recently been undertaken (Frank, Andresen, & Brantingham, 2012, 2013). Their work investigated criminal trips undertaken by burglars (Frank et al., 2012) and general property crime offenders (Frank et al., 2013). Their work demonstrates that, at least in urban environments, property crime oriented offenders exhibit some amount of directional bias. This bias is heavily dependent on the underlying urban structure; in particular the distribution of residential and major shopping districts. They note that, at least in the case for general property crime, there was not always a strong global (municipality-wide) directional bias. In some circumstances, offenders would exhibit a more localized directional convergence (Frank et al., 2013).

This discussion serves to illustrate the importance of both propinquity as well as morphology. Frank et al. succinctly conclude that '…it should be clear that direction is an important aspect when considering the spatial dimension of criminal activity" (Frank et al., 2012, p. 42). Together propinquity and morphology provide useful vectors of measurement and assessment whose real worth can most likely be appreciated when considered in tandem. Propinquity and morphology remain, however, only descriptive attributes of criminal behaviour, and a number of theories have been proposed that seek to explain the observed spatial patterns of criminals thus far discussed. These include: routine activity theory, rational choice theory and crime pattern theory / the geography of crime.

1.4 Criminological Theories of Criminal Spatial Behaviour

This section presents three commonly cited criminology theories of offender spatial behaviour: routine activity theory, rational choice theory and crime pattern theory. By identifying why some of the crime patterns occur in the way that they do, then more informed applications of individual mobility can be developed. Furthermore, work in these areas can help to understand factors that may contribute to observing the spatial patterns presented by offenders, which in turn allows for more accurate modelling of an empirical activity space.

1.4.1 Routine Activity Theory

Routine activity theory was first presented by Cohen and Felson (1979). It states that criminal activity is the result of the convergence of motivated individuals (likely offenders), suitable targets and a lack of capable guardians (Clarke & Felson, 1993). This convergence occurs when the routine activities – that is “recurrent and prevalent activities which provide for basic population and individual needs” (Cohen & Felson, 1979, p. 593) – of offenders and victims coincide at a time and place where capable guardians are absent. In other words, a bad guy has to encounter a victim in an area with no good guys. If any of these three conditions are not met, Cohen and Felson argue, a crime will not occur. Furthermore, they specify that these “recurrent” activities can occur in one of three places: “(1) at home, (2) in jobs away from the home, and (3) in other activities away from home” (Clarke & Felson, 1993, p. 593). Routine activity theory frames crime in mundane terms: the typical offender goes about their daily business in much the same way as non-offenders and only commits crime when the opportunity and impulse arise.

Routine activity theory can also be used as a means to describe criminal mobility (Van Daele & Vander Beken, 2011b). The convergence of offender and victim in space and time is contingent on how far either party is willing to travel at a given time. For example, assume that offender B is willing to travel twice the distance as offender A; then obviously they cover more raw area and, assuming equal distribution of suitable victims, are far more likely to encounter said victims. Conversely, if a would-be victim travels greater distances on average in their daily activities, then it would be reasonable to expect that they would encounter more potential offenders on average and thus have a higher likelihood of being targeted. Importantly, it should be noted that the opportunity to fulfil one's needs are not universally distributed in space; if one needs to go shopping for food, for example, shopping centres are not typically evenly distributed. Thus an individual's travels and subsequent offending pattern depend, at least in part, on the offending opportunities afforded by their surrounding environment. Routine activity theory posits that the observable distribution of crime is reflective of the environment in which it occurs as well as the specific needs and biases of the individual in question.

Fundamentally, routine activity theory assumes motivated offenders. That is, routine activities differentiates between individuals who have the capacity to commit crime – given an appropriate opportunity presents itself – and everyone else. As pointed out by Daele and Beken (2011b), however, it is not always necessary to understand such motivations; rather the study of the location of the crime itself is sufficiently revealing. According to routine activities, the crime location itself is worthy of study as it represent locations where motivated offenders and suitable victims converge in time and space. In other words, while we may not always be able to know what a particular offender's crime motivations are, by studying their criminal history patterns we may be able to identify known areas of activity unique to their own routine activities. Such a 'signature area' may be uniquely indicative of said offender.

1.4.2 Rational Choice Theory

Rational choice theory has roots in economic thinking and unlike routine activity theory is primarily focused on capturing the specific motivations of offenders (Clarke & Cornish, 1985; Clarke & Felson, 1993). The early iterations of rational choice theory treated criminality as an occupation alternative and heavily emphasized the materialistic gains associated with only specific types of crimes. For example, some rational choice interpretations would say an individual would turn to burglary when the profits of doing so outweighed the risks of capture and the benefits of alternatives (such as traditional employment).

The rational choice theory interpretation put forth by Clarke and Felson (1993) proposes that "crime is purposive behaviour designed to meet the offender's common place needs for such things as money, status, sex and excitement, and that meeting these needs involves the making of (sometimes quite rudimentary) decisions and choices, constrained as these are by limits of time and ability and the availability of relevant information" (Clarke & Felson, 1993, p. 6). They also outline a distinction between being involved in a crime – that is choosing to commit any given type of crime over another – and the specific manner in which any given crime is committed.

The central argument is that the choice of being criminally involved occurs over a longer timescale and is more likely to be a result of a considered position when compared to the situationally volatile nature of specific crime actions. Thus rational choice posits that what crimes offenders choose to engage in reflect the needs they seek to fulfil and is the result of some calculus surrounding their knowledge and ability to fulfil that need. Unfortunately, a key assumption is that the individual has sufficient knowledge and ability to perform such a cost/benefits analysis and are aware of competing alternatives. Similar assumptions are often made in economic theory, but as was the case there, criminals rarely behave as rationally as the models propose.

Despite this, however, rational choice theory is important to consider as it provides agency to the individual and highlights the distinction between committing a crime at all and how the crime is committed. This is important as the decision to commit certain types of crimes over time indicates some consistent decision making process, whereas the specific actions during any given offense may be the result of circumstance and may be exceptionally volatile. Thus when attempting to quantify spatial behaviour over time, rational choice theory implies that consistency is perhaps more important to focus than any single event.

1.4.3 Crime Pattern Theory and the Geometry of Crime

As the name suggests, crime pattern theory is primarily interested in providing explanations for why crime occurs where it does; for example why crimes cluster in certain areas of a city over others and was first proposed by Brantingham and Brantingham (1993). Crime pattern theory represents a ‘meta-theory’ bringing together concepts introduced by routine activities, rational choice, and the Brantingham’s own geography of crime. Brantingham and Brantingham’s geography of crime drew specific attention to what they called the “environmental backcloth” (Brantingham & Brantingham, 1995). Specifically, that for any given area there exists some intrinsic structure; roadways, bus routes, store locations, physical barriers etc. that would influence and limit how people travel. This “backcloth” includes crime generators and crime attractors: areas that either attract large numbers of people some of whom will be criminals versus areas whose reputation for criminal opportunity attracts offenders directly.

This backcloth, Brantingham and Brantingham (1993), argue, necessarily influences how offenders go about committing their crimes. They describe offender mobility via nodes, paths and edges. Nodes represent discrete areas of activity: the home or places of work and/or leisure. Conversely, paths represent the specific routes that an individual travels between nodes. Nodes and paths are semantically similar to those areas described in routine activity theory as those specific locations that people frequently interact with. Edges are described as “places where there is enough distinctiveness from one part to another that the change is noticeable” (Brantingham & Brantingham, 1993, p. 17). Barriers, like rivers or major motorways, are some examples of edges. The significance of edges is argued due to the apparently higher risk of victimization that occurs along such edges and has been termed the “edge effect” (Brantingham & Brantingham, 1993, p. 18).

Brantingham and Brantingham’s (1995) portrayal of crime pattern theory is one in which crime occurs when the “activity space” of a would be offender intersects with the “activity space” of a would be victim. The activity space described here is the summation of an individual’s specific nodes and paths. In this context a node is a specific location, such as the home, work or a recreation centre, where routine activity occurs. Paths are the specific routes individuals take to reach nodes. Individuals would have specific nodes they frequented and specific paths that connect those nodes, and this relationship forms the basis of their activity space. This formulation of an activity space is a reinterpretation of routine activity theory: the nodes make up the specific locations that individuals frequently travel (for whatever reason). Crime pattern theory differs from routine activity theory by introducing the notion of more purposive crime by way of an “awareness space”. Such an awareness space, it is argued, contains an individual’s understanding of their environment and its apparent criminal opportunity. Such an awareness space is built up over time and can be thought to be generated from the routine activity space. That is during the course of their routine activities, individuals build understanding of some areas over others and this understanding represents their awareness space (Rossmo, 2000).

Crime pattern theory describes crime as occurring when the activity spaces of motivated offenders and suitable victims intersect. These intersections are described as crime ‘attractors’ or crime ‘generators’ which are conceptually linked to the underlying usage of the space. For example a busy bus terminal or subway station could be considered a crime generator as the primary use of the space involves a large flux of individuals. Brantingham and Brantingham describe crime as being generally opportunistic, and as such the active search for victims in relatively unknown areas is unlikely (Brantingham & Brantingham, 1993, p. 10). Brantingham and Brantingham provide the following summary of the most succinct postulates of crime pattern theory:

(1) Individuals build cognitive maps; knowledge of spatial relations influence crime location.
(2) The cognitive representations reflect high activity nodes and the paths between them and through those representations shape the location of crime.
(3) The type of crimes are varied, but some are highly opportunistic and highly dependent on daily activities and the physical availability of suitable targets and suitable crime situations, frequently including lack of surveillance or a feeling of anonymity.
(4) Those who commit crimes have their own behaviour settings, influenced by social surrounds but also by simpler characteristics and as safe access.
(5) Within a major urban area the city planners and other decision makers use zoning, transportation planning and site review to shape travel paths, store locations, school locations, parks and special activity centres and consequently create crime generator locations and produce some of the actual pattern of crime.

Crime pattern theory is severely limited, however in that it is typically descriptive in nature. The original publications are exceedingly light in any substantial empirical support for the claims summarised above. Furthermore, like routine activity theory, crime pattern theory is primarily a macro level model of criminality. The patterns described typically only hold when information is heavily aggregated, and as such provides little explanatory power when applied to any given individual offender. For example, the notion of crime generators is one that only makes sense in the context of general trends. Being an offender does not necessarily imply that all or even any of one's crimes would occur at a crime generator location; however such an offender may very easily commit multiple crimes at the same non-crime generator location. Furthermore, while crime pattern theory uses the concept of an awareness space to explain individual level variation, it provides no guidance into how or why awareness spaces between individuals can be the same or different. The concept of an activity space, however, has been pursued within the literature in non-criminal contexts as a means to study individual movement within an environment (Johnston, 1972; Kestens et al., 2010; Lee et al., 2016; Perchoux et al., 2013; Wong & Shaw, 2011).

The criminological theories explored here provide a framework from which to begin to understand how criminal activity may be shaped according to the unique interplay between a given offender and the environment in which they operate. Specifically, the unique combination of individual tendencies (rational choice) combine with the specific environmental exposures resulting from one's routine activities within their environmental backcloth to create a unique individual-to-environment dynamic which can, in theory, be conceptualized as an activity space. What follows is a discussion surrounding how and why one may expect such activity spaces to vary over time from individual to individual as well as cognitive processes which may be driving such variations.

1.5 Spatial Cognition and the Development of an Awareness Space Model

Activity spaces are generally described as a geographic extent or surface that encapsulates an individual's interactions with specific locations. However, such a working definition fails to provide context for the cognitive processes at play which determine an individual's understanding of their environment and subsequently what they can or cannot do in said environment. Heft, for example, argues that most psychological research in spatial cognition falls into one of three perspectives: "1) an information-processing approach; (2) a tradition stemming from Piaget and Inhelder's work on children's development of spatial cognition; and (3) a nativist approach to spatial cognition with its roots in Cartesian and Kantian thought" (Heft, 2013, p. 18). These three perspectives, Heft argues, summarise the different ways in which researchers have attempted to model how humans and animals develop a mental schema of their environment, of which activity spaces are thought to be a subset. Generally speaking the first two perspectives argue that spatial cognition is developed iteratively through experience and observation which in turn generates a mental construct that represents the surrounding environment, while the third posits some form of inherent ability. Where the Tolman and Piagetian perspectives differ is how those constructs are organized: the Tolman viewpoint being one of a set of possible actions and their alternatives relative to a place (Tolman, 1948), and the Piagetian perspective being one of cultivating a logical mental construct of the environmental structure in strict relational terms (Piaget, Inhelder, & Piaget, 1997). The nativist perspective, finally, proposes that spatial cognition occurs *a priori*; it is not a process that is learned but is instead inherent. However, spatial cognition research very rarely adheres to any one perspective strictly, instead typically being a combination of all three (Heft, 2013).

These perspectives all differ to various degrees regarding how spatial information is processed and subsequently stored; the behaviourist perspective describes it as a set of actions and their alternatives whereas the Piagetian perspective emphasises the malleable nature of spatial cognition and its development; and finally the nativist perspective simply offers that somehow spatial cognition is an inherent quality. Where these three perspectives seem to agree is that the highest form of spatial cognition lies in the configurational understanding of the environment – i.e. the individual understands how components of the environment are related. This configurational understanding is typically described as “survey knowledge” or a cognitive map (Heft, 2013). Thus these perspectives disagree on the exact processes involved, but agree that the “end product” of spatial cognition is some mental construct that stores the individual’s understanding of their environment and its relationships in memory for later recall.

The exact process by which the environment gets “parsed” or understood by an individual and subsequently stored is not yet agreed upon. Empirical research from animal studies as well as child orientation tasks point towards competing processes. For example, in child orientation tasks it has been found that geometric cues were more influential than “non-geometric” or “local” cues, such as coloured walls, for child participants (Hermer & Spelke, 1996). Children were placed into an enclosure and shown a toy, which was then hidden behind a sheet. The participants were then gently spun around to disorientate them, and their search behaviour for the toy was observed. Typically children either searched the correct corner of the enclosure, or the geometric similar opposite corner - even when local cues were present. This implies that the children were utilizing geometric cues – corner locations – to orientate themselves and not local cues such as coloured walls.

Subsequent studies have illustrated that the size of the enclosure, among other things, influences whether or not subjects utilize geometric or local cues to orientate themselves. Learmonth, Nadel and Newcombe (2002) illustrated that once the experimental enclosure reaches a certain size, children began to use local cues and would no longer confuse diagonally adjacent corners. Thus orientation is not an either or scenario with regard to the importance of geometric cues vs local landmarks, but rather that circumstance dictates which orientation method is used. Finally, Heft (2013) draws attention to one other important consideration in the development and manifestation of spatial cognition: personal history.

A number of animal studies investigated whether or not the rearing conditions of the animal influenced whether geometric or local cues were more important in orientation. The empirical results of which imply that the method of orientation used was in fact contingent on their rearing conditions (such as circular vs rectangular enclosures (Twyman, Newcombe, &

Gould, 2009). This body of work serves to illustrate the plasticity associated with spatial cognition. The relative importance of geometric cues versus local cues appear to vary according to circumstance and an individual's past experience.

Recent work within neurophysiology describes specific mechanistic theories for spatial cognition and provides some context for interpreting the adaptive nature of spatial cognition and its method of storage outlined above. Hartley and Burgess (2002) describe a cognitive model based around neurobiology. They argue that different components of the brain handle different spatial representations based on their unique demands. For example, they describe how: "Hippocampal processes are concerned with large distances and long timescales, whereas parietal processes are concerned with short timescales and the space immediately surrounding the body" (Hartley & Burgess, 2002, p. 2).

These observations lead to the development of a cognitive model based on frames of reference. An egocentric frame of reference is centred on the self; whereas an allocentric frame of reference is centred on the "world". These two reference frames correspond to two different mental representations of the environment corresponding to short term and long term respectively. The reason for this division is best illustrated by example: an egocentric representation of the environment would be one that held the self as central to all relationships; as the individual moves through space such representations would require constant updating and obviously would be a very "expensive" means of representing the world. The allocentric representation sidesteps this issue by providing a means to store long-term spatial relationships between entities irrespective of their location to the self. If we return to the orientation task example, whether or not an individual employs local cues vs geometric cues may be thought of as invoking either an egocentric frame of reference or an allocentric frame of reference depending on the contextual demands of the task. This short – long term mechanist model of spatial cognition shares many similarities to the modal model of memory of Atkinson and Shiffrin (Atkinson & Shiffrin, 1968).

Thus far, however, the examples and discussion have only dealt with spatial knowledge resulting from direct experiences and interactions. Participants in the examples provided have only needed to process one source of data: their immediate surroundings and by mediation of their past experiences, utilize that information to complete a simple task. Obviously in more complex behaviour evaluating more data streams becomes a necessity and the resulting conceptualization becomes more complex.

This complexity was explored by Brown and Moore (1970), who were interested in studying inter-urban migration. They identified three domains of information important in any search activity: 1) the total information available to the searcher; (2) the initial information possessed initially by the searcher; and (3) the way in which the searcher combines these two sources of information. Thus for a given house-seeker that individual's ability to find a new house is restricted by (1) the available information regarding new houses, because they cannot consider a house they do not know about (i.e. they are not aware of it as a possibility); (2) knowledge about houses they have generated through experience (i.e. their specific taste in house or perceived needs); and (3) the strategy they put into action given the information they have aggregated together; as not all individuals search for houses in the same way. Brown and Moore describe the syntheses of these three domains as an *awareness space*.

An *awareness space* is "the locations within the total urban space about which the intended migrant household has knowledge (or knowledge above some threshold level) before search begins" (Brown & Moore, 1970, p. 8). From this a more generic formulation of an *awareness* *space* can be thought of as the collection of knowledge an individual holds about their environment and its apparent opportunities. The concept of a mental representation of the environment is not a unique one within the literature. Dijst (1999) describes this construct as the "perceptual action space"; Horton and Reynolds named it the "action space" (1971), and Lynch (1984) called it a "mental map". For consistency's sake, "Awareness space" will hereafter be used as the term to describe the construct that holds an individual's understanding of their environment.

This formulation is consistent with Dijst who said: "Theoretically, the actual action space is situated within the potential action space. The perceptual action space covers the actual action space. The potential action space can be covered entirely by the perceptual action space (Dijst, 1999, p. 196)". While Dijist's terminology can be confusing, it succinctly relates three important concepts together which are consistently mentioned within the literature: (1) that there exists some construct that encapsulates the full understanding an individual holds about their environment [awareness space], (2) that there exists some subset of locations from said construct where the individual *could be* at any given time [potential activity space], and (3) those locations where the individual *actually is or has been* [activity space]. The potential activity space described by Dijst relates the spatial and temporal limitations (see: 1.6 Time-Space Geography) of an individual together. Thus the awareness space can be thought to be synonymous to the allocentric representation of the environment from the mechanistic theory of spatial cognition, and also can be related to the survey knowledge described by Heft.

According to Brown and Moore this survey knowledge is derived from two sources: day-to-day activities which yield information through direct contact and information from secondary sources through indirect contact. That is, that such survey knowledge is created from the repeated exposure to the environment via routine activities. Thus, an activity space is the observable subset of locations of an individual's awareness space. Figure 1 outlines how the unique combination of individual traits and environmental traits are combined to form an individual's awareness space construct and the observable activity space subset.

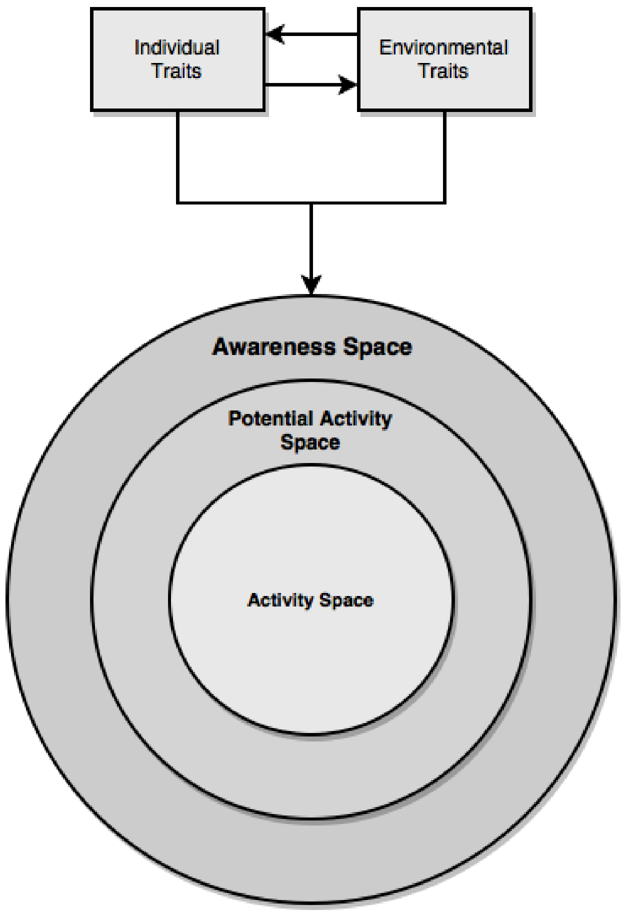


Figure 1: Awareness Space // Activity Space Spatial Cognition Development Model

Horton and Reynolds described an individual's understanding of their environment as a dynamic process. Specifically, they outline a multi-stage process of environmental learning that revolves around a continuum of location familiarity. Initially, individuals new to an environment have a small awareness space that is centred on the home and place of work. The individuals awareness and activity spaces rapidly expand as new opportunities are discovered until they reach a stage of “equilibrium” at which point development rapidly slows (Horton & Reynolds, 1971). This process was also documented, at least informally, by Canter (1977) whereby a colleague new to the London area was tasked with drawing their mental representation of their new environment; the drawings started off rather crude as the relationships between places was new but grew in sophistication and precision over time.

A similar learning continuum could be applicable to crime and criminal opportunity. An offender with a rudimentary understanding of criminal opportunities could “build up” understanding in much the same way as someone who recently moved to a new location. Such an interpretation would have much in common with routine activity and crime pattern theories. As an offender’s understanding of criminal opportunities matures, they are more able to make informed decisions about where and how to offend. There is some empirical evidence to support this hypothesis: Bernasco (2010) describes a tendency for offenders who have recently moved, to target areas around their current and previous home locations, as opposed to similar locations they have never visited. The culmination of these observations suggests that a rudimentary understanding of offender spatial behaviour can be achieved by framing spatial activity in terms of: (1) a limited physical activity space mediated by (2) a cognitive awareness space comprised of the entirety of knowledge an individual has generated concerning their local environment, which (3) itself is developed iteratively over time and experience. Figure 2 provides a thematic outline of how the criminal awareness space may be influenced by the development of an offender’s understanding of their environment. While Figure 2 is presented as a dichotomy, it is more accurate to think of the process in terms of a continuum between low and high environmental familiarity. Such a formulation provides ample avenues to explore specific empirical inquiries into the spatial tendencies of offenders.

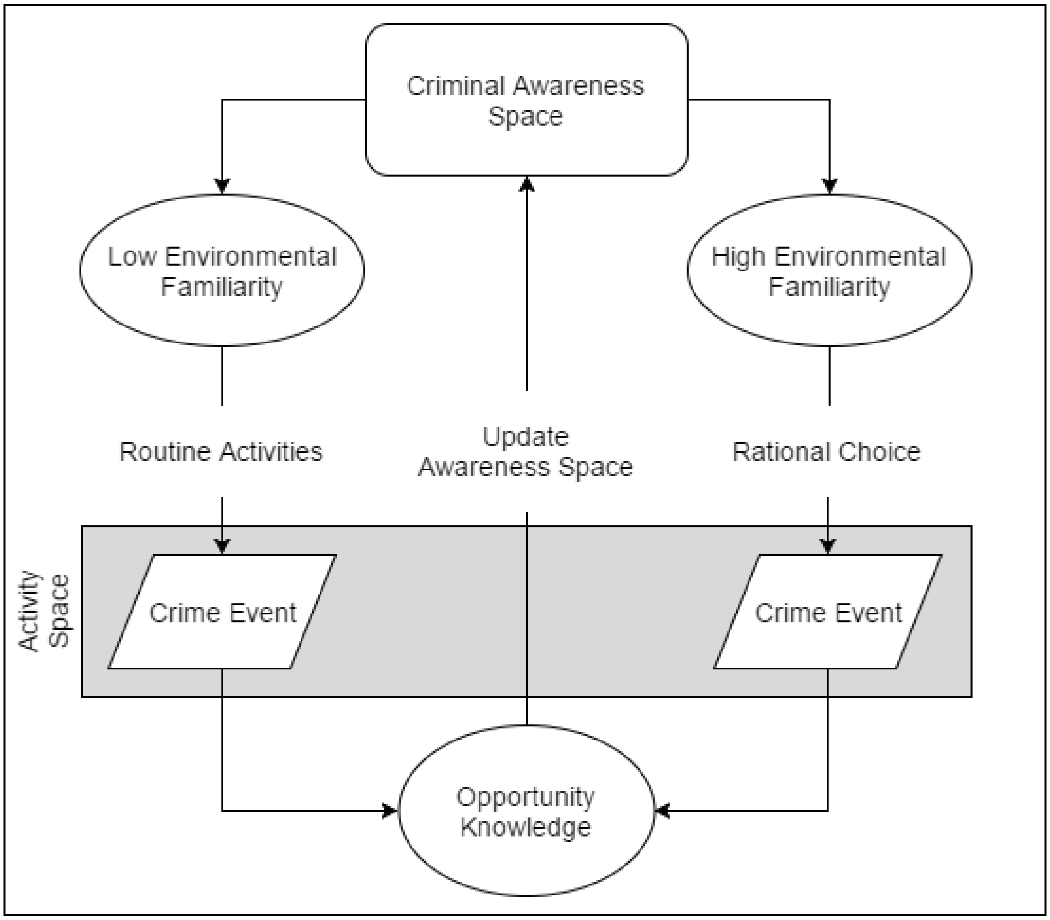


Figure 2: Criminal Awareness Space Development

1.6 Time-Space Geography

All activity space measurements, regardless of their specific formulation, are predicated on the limiting natures of time and space. Hägerstraand (1970) proposed the initial time-space geography framework, which was further developed by Miller (2005) into a more formal measurement based proposition. The conceptual framework presented by Hägerstraand details how the life-path – how an individual has chosen to move through time and space – can be described as a time-space prism, which is a construct that describes all possible destinations an individual can travel at a given point in time. The framework is based on the observation that an individual's ability to travel is limited by various resources, the most important of which being time. For example, the amount of time an individual can "spend" to engage in a given activity is to some extent contingent on the amount of time required to "spend" to get to the location of said activity. Thus where activities are, and how long they are engaged with, are of particular importance.

Miller, on the other hand, took the initial informal work of Hägerstraand and attempted to transform it into a formal measurement theory. Miller's formal definition of the life-path relates two key entities: (1) control points and (2) path segments. Miller describes a control point as being one of three possible types of locations: "(i) known activity locations; (ii) locations where the path experiences a change in direction (e.g., a turn at a street intersection) or velocity; and (iii) arbitrary locations where a spatial reference and time stamp are recorded (e.g., from a GPS receiver)" (Miller, 2005, p. 26). Importantly, a control point is not necessarily the location of an activity, as can be seen from Lee et al (2016) where many of the geo-recordings occurred en-route to the destination.

Path segments represent specific routes taken between temporally adjacent control points. Conceptually, these represent the travel network, be it road or otherwise, between control points. Path segments are by Miller's definition *unobserved*, as otherwise they would be control points. This uncertainty is the foundation of the time-space prism, which encapsulates the range of possible destinations or routes surrounding a path segment. An example more clearly illustrates the concept: consider you know where an individual is, at two distant points in time; the time-space prism is said to encapsulate all possible routes available to that individual that could deliver them from the first location to the second within the time-limit. This range of possible routes can be represented as a 2-D surface; most typically some form of ellipse (Miller, 2005).

The general proposals of time-space geography can be summarised as relating the relationship between observed events, or control points, to one another for a given individual while accounting for the uncertainty in how those points are explicitly connected. Furthermore, even though control points occur in three-dimensional space (Euclidean X and Y and the temporal Z axis), a 2-D surface can be sufficient to represent the observed temporal-spatial behaviour. Finally, the life-path, and by definition the time-space prism, are idiographic in nature; a given life-path is only descriptive of the individual who created it.

The time-space geography framework thus far presented has operated from a general application perspective due to Miller's emphasis on a focus on measurement. Insofar as crime is concerned, time-space geography is applicable as criminal events occur amidst the temporal demands of daily life. This is not a new observation, as similar propositions have been put forth in the already described criminology theories, as well as by investigative psychology, and even by crime analyst practitioners (Brantingham & Brantingham, 1995; Canter & Youngs, 2009; Cohen & Felson, 1979; *Exploring Crime Analysis*, 2004), however few studies incorporate a holistic view of any given offender's offending history counter to what this view might suggest.

The problem with life-paths or time-space prisms when applied to criminality, however, is one of resolution. Time-space prisms as presented by Miller are generic in their formulation – meaning they are free to be applied at any temporal resolution – however, they quickly lose descriptive power as the temporal span of the unknown path between crime events grows. For example, while it is reasonable to estimate the area an offender is located between two crimes that are a number of hours apart, it becomes nearly meaningless to do so when the two control points are weeks or months apart. Furthermore, the mode of travel has a direct impact on the upper limit of any given offender's range; as does any number of extenuating circumstances that impact travel such as traffic or public transportation service interruptions. Placing an emphasis on establishing the potential extent an offender can travel or be located for a given time window is thus problematic.

It is not the specific whereabouts of a given individual between crimes that is of importance so much as it is the fact that that individual chose to offend in the same area or not. Activity spaces may be used to assess offender spatial consistency, which has been proposed to be one of the foundational assumptions in areas of applied criminal behaviour research including crime linkage analysis (Tonkin & Woodhams, 2015).

1.7 Thesis Goals

Where past studies of criminal spatial movement have focused on exclusively aggregate measures of offender mobility the current thesis seeks to expand offender spatial modelling by establishing the 'criminal activity space' as a reasonable measure of specific individual mobility. Such a measure is argued here to be an important descriptive feature of offending behaviour that has thus far remained under-explored.

Activity spaces have already been demonstrated to be a useful tool for studying individual engagement with their environments in non-criminal contexts, and their existence within criminal contexts has been suggested by a number of theories including the often cited criminology theory of crime patterns. However, to-date there has been little empirical work performed that evaluates criminal activity spaces in a comprehensive manner. This thesis seeks to address such shortcomings within the literature by presenting a comprehensive review of activity space estimation methods and how such methods can be utilized to study offender spatial behaviour. To that end, the following research aims and the specific research questions therein will be explored.

1.7.1 Research Aims & Questions

The following research aims and their targeted research questions represent the identified gaps in the existent criminal spatial-behavioural literature; drawing heavily from theories of opportunity from criminological theory as well as applied cognitive theory from both psychology generally as well as from the crime-linkage literature specifically.

Aim 1: To explore and develop an empirical activity space measure for offender criminal activity.

Research Question Set 1:

1.1. Can diverse individual crime patterns be described via a generalized 'activity space'?

1.2. Are such activity spaces predictive of future criminal spatial activity?

Question set 1 is primarily focused on establishing the groundwork from which all subsequent analysis will focus. Question 1.1 focuses on the absolute base-line understanding of what it means to model offender spatial activity in an empirical way by establishing a fixed definition for what constitutes an offender activity space and how such a space can be calculated from crime location data. Question 1.2 builds upon this definition by exploring the most simplistic theoretical relationship of whether a space is in some way predictive of future activity.

Aim 2: Examine the validity of the activity space construct by exploring the predictive gains afforded by the inclusion of activity space based measures into crime linkage analysis tasks and models.

Research Question Set 2:

2.1. Can the inclusion of activity space measures improve established models of crime linkage analysis?

2.2. Can the findings from (1) be replicated using alternative modelling methodologies?

Research question set 2 is focused on bridging the gap from the theoretical understanding of offender spatial behaviour - represented by the established activity space construct from question set 1 – to an applied context represented here by crime linkage analysis. In this way, crime linkage analysis acts as a suitable and established testing bed from which the predictive power of the outlined activity space construct can be evaluated. Question 2.1 targets this specifically by investigating whether activity space based measures meaningfully improve model predictive ability by replicating and comparing to past crime-linkage study methodologies. Finally Question 2.2 begins to address both validity and reproducibility by replacing the established linking model from the literature (logistic regression) with a generalized machine-learning algorithm (random forests). It is assumed here that if the theoretical relationships are stable then both methods should yield similar results.

Aim 3: Assess the stability of the activity space based approach of crime linkage by exploring the possibility of a 'suspect prioritization' system which surfaces likely individuals linked to a crime event rather than simply other individual events.

Research Question Set 3:

3.1. Can offenders be meaningfully prioritized – i.e. ranked – for a given candidate set such that the responsible offender appears near the top of an automatically generated list?

3.2. Does the inclusion of activity space measures improve rank performance?

3.3. Does the aggregation method significantly impact rank performance?

3.4. Do other systematic factors, such as the underlying geography, disproportionately impact ranking results?

Research question set 3 introduces an approach to crime linkage which shifts the focus of the linkage task away from classification (linked vs unlinked) and towards a ranked outcome. As the activity space construct is an ideographic measure in nature, Question 3.1 investigates whether a more individual focused testing methodology – as opposed to event focused – could address several limitations outlined in response to Questions 2.1 and 2.2; such as the large false positive rate for classifying linked crime pairs (see: Table 22, p. 265). Questions 3.2 through 3.4 represent further validity testing given a new ranking based methodology. Specifically they focus on assessing whether several readily identifiable free-parameters influence the final ranking as much or more than the inclusion of the activity space measure itself.

Chapter 2

Data and Design

2.1 Introduction

The current research is centred on exploring several facets related to criminal activity spaces. While specifics surrounding said activity spaces will be introduced in subsequent chapters, it is sufficient for now to state that the manner in which such spaces can be reasonably described are dependent on available data. As will be shown, a fundamental assumption made as part of the current research is that individual differences in behaviour and victim targeting will be manifested as differences in criminal spatial activity. In the current research, ‘criminal spatial activity’ is taken to be specific crime locations. This places several specific demands on the data: specifically crimes need to be able to be attributable to specific individuals and furthermore the crimes need to have suitable geographic data attached. While the data made available for use in the current research can satisfy these two requirements, there were a number of issues that warrant specific consideration. The specific issues encountered as part of the current research as well as a discussion regarding working with real-world crime data in general will follow.

2.2 Design Philosophy

As indicated in Chapter 1, the literature is sparse when it comes to applied theory of criminal spatial activity. In order to address this limitation, the research seeks to accomplish two overarching goals. The first is to establish an empirical measure that can represent the activity space construct. The second is to test the measure’s validity. The empirical chapters each focus on one aspect of these two goals. Chapter 3 provides an in-depth discussion surrounding activity space estimation methods and ultimately provides what is assumed to be a reasonable starting point for what such a measure may look like. Recall that the activity space measure – whatever it ends up being – would be representative of the underlying awareness space and decision making process of an individual in question. Thus the empirical activity space measures put forth by Chapter 3 are assumed to be representative measures of the awareness space construct. Chapter 4 and 5 explore the construct validity via the ‘nomological network’ (Cronbach & Meehl, 1955, p. 290). A ‘nomological network’ is the “interlocking system of laws which constitute a theory” (Cronbach & Meehl, 1955, p. 290). In other words, the activity space measure is used to test the awareness space construct by identifying related but distinctly testable subspaces.

Chapter 4 introduces the activity space measures into a crime linkage framework. Crime linkage analysis makes two fundamental assumptions concerning offender behaviour: 1) offenders are consistent in their behaviour and 2) these behaviours are distinctive from others. Crime linkage analysis itself is a classification task; a classifier is trained to identify pairs of crimes as likely to be committed by a common offender or not. Assuming that the awareness space construct is valid, then construct validity is tested by looking at the predictive validity of such a classifiers containing the activity space measure vs those without. Improved performance would be assumed to indicate that the activity space measure is correctly capturing relevant information about where individuals commit crimes vs their peers.

Chapter 5, on the other hand, approaches validation from a different perspective. Where crime linkage analysis focuses on crime-pair classification, Chapter 5 introduces a ranking approach that evaluates the consistency and distinctiveness of individual offenders to crimes. Again, models are evaluated both with activity space measures and without and again it is generally hypothesized that models with activity space measures will perform ‘better’. While it is unlikely that any one of the validation tests proposed can be considered sufficient to establish validity, when taken together the sum of the evidence may indicate a level of confidence in the measure. Obviously, however, the outcome of such tests relies heavily on the formulation of the measure in the first place.

There are numerous ways in which activity spaces have been described within the literature (See: Sherman et al., 2005 for some such examples). However, there has not been a generally accepted ‘best practice’ when it comes to specific measures. In actual fact, the methods employed for deriving such measures are generally driven by the nature of the data available to any given study. Such data sources range from self-reported travel diaries (Kwan, 1998), to time-stamped geo data from social media (Lee et al., 2016). From these two examples alone it is readily obvious that the precision and inferences afforded can vary drastically. In the case of Kwan, for example, the actual destination data will be prone to reporting error, whereas mode of transportation is most likely very reliable. This dynamic is almost certainly flipped in the case of Lee who has relatively accurate geo-data but very inaccurate or altogether missing data concerning mode of transportation. Issues such as these are further exacerbated when one considers crime and crime data. Subsequently, many of the assumptions made throughout the current research can be directly attributable to the underlying data.

2.3 Data

The current research is centred on exploring several facets related to criminal activity spaces. While specifics surrounding said activity spaces will be introduced in subsequent chapters, it is sufficient for now to state that the manner in which such spaces can be reasonably described are dependent on available data. As will be shown, a fundamental assumption made as part of the current research is that individual differences in behaviour and victim targeting will be manifested as differences in criminal spatial activity. In the current research, ‘criminal spatial activity’ is taken to be specific crime locations. This places several specific demands on the data: specifically crimes need to be able to be attributable to specific individuals and furthermore the crimes need to have suitable geographic data attached. While the data made available for use in the current research can satisfy these two requirements, there were a number of issues that warrant specific consideration.

2.3.1 Data Collection

The current research has been made possible by a data sharing agreement between the International Research Centre for Investigative Psychology (IRCIP) and the Greater Manchester Police (GMP). As part of this agreement, law enforcement data pertaining to crime incidents was extracted and shared for research purposes. The data itself represents a static cross-section view of the crime condition from a sub-section of the Greater Manchester area. As far as can be determined after-the-fact, data manipulations on behalf of GMP were minimal: sensitive data such as specific individual’s names were removed but very little else was done prior to data being handed off. Finally it should be noted that data was accessed in accordance with IRCIP and university guidelines and that the research itself was approved by the ethics panel (SREP).

The data itself covered a wide range of criminal activity for the period: it was not limited to only solved crimes for example; nor was it limited to only specific crime sub-types. This both opened many interesting lines of inquiry while also introducing several important challenges that needed to be addressed. First, as the design philosophy has a distinct focus on predictive outcomes, then it was paramount to identify the subset of records for which the responsible party was correctly identified. This was done by referring to a case closing field which had a number of structured outcomes including: charged, undetected, identified by non-charged among others. For the current research, only charged crimes were considered as these were thought to hold the highest burden of proof for attribution within the available data set. It is very important to note, however, that a police charge does not itself constitute a guilty verdict. From the data that was made available it was not always possible to determine with certainty that a given crime event could be attributed to the individual indicated within the data. This represents a fundamental limitation within the data.

Limitations such as these are common when working with second-hand data sources as is the case here. Contrary to data from classical laboratory experiments, 'real world' data are often saddled with a number of issues. First it must be understood that often secondary data are being used to investigate questions it was not initially collected for which in turn can cause a host of problems. Law enforcement records, for example, are often simply a log of events: what happened, where it happened, what did witnesses say, etc. The information that is collected as part of on-going casework is inherently biased towards achieving a given agency's goals which may or may not align with subsequent research questions down the road.

At a more practical level, law-enforcement officers typically do not sign-up because of their love of forms and paper work. Record keeping is often a secondary concern behind resolving active incidents and ensuring public safety in general. Thus records may be recorded well after-the-fact and data contained therein may not always be as specific or accurate as possible. These challenges are not themselves new or unique to the current research presented here, nearly all researchers working with secondary data – and crime data specifically – acknowledge the various limitations that arise from working with such sources (see: Canter &

Youngs, 2009 for a general discussion and Woodhams & Bennell, 2015 for a more targeted discussion within crime linkage analysis).

Fundamentally, all of these concerns stem from the fact that researchers consuming data second-hand were not present for data collection and may not be aware of the various ‘quirks’ – that is business rules – that have impacted how said data was collected in the first place. A primary example of this is how crime events are recorded which often results in duplicated or inflated incidence numbers. For example, consider a burglar breaks into student housing, and in so doing gains access to a number of unlocked rooms. This one criminal event could lead to many individual crime records. The business rules for collecting such complaint data may dictate one record be generated for every victim. From the law-enforcement perspective this makes sense: one offender committed a string of acts that impacted multiple people. However, from the researcher’s perspective, the fact that there were multiple victims may be the result of circumstance and that that one evening’s set of events should not be more heavily weighted than another where the same perp only victimized one individual.

The key question for the current research and researchers in general is how to rectify law-enforcement’s representation of criminal events and what they may or may not represent for their perpetrator. For this particular example, such incidents were generally aggregated into a singular event in the current research. The rational for doing so is that the number of rooms that were actually unlocked is a matter of circumstance and was not believed to be the result of any specific action taken on behalf of the offender. Further methodological reasons for dropping such ‘duplicate’ events are covered in more detail in later chapters. A general discussion of the data made available to the current research follows which is intended to provide context for all subsequent chapters and analysis.

2.3.2 *Data Descriptive Statistics*

This section provides a general summary of the crime data sample made available to the current research. In total, the sample contained information pertaining to 97,878 total crime incidents from August 2011 through August 2015, committed by 22,370 unique offenders. It should be noted that this total sample was a composite of several, smaller data dumps during that period and it is not clear what data may have been omitted as a result (see: Figure 7). Due to the sensitive nature of the data in question, special procedures were necessary to follow in order to acquire said data. Most importantly, the data could not be transmitted electronically. This necessitated the hand-off of physical disc across a number of occasions that then needed to be merged back into a master data set. As the main project involving this data was on-going for several years, several of these hand-offs took place. This dataset included the following information:

Variable	**Description**
Crime Type	Example: Shoplifting, Burglary (Dwelling / Commercial)
Offender Details	Age, Sex, Ethnicity
Offender Home Location	Recorded in Easting // Northing (EPSG 27700, OGSB 1936 Projection)
Crime Location	Recorded in Easting // Northing (EPSG 27700, OGSB 1936 Projection), Premise Type
Date / Time of Crime Event	Standard Date-time stamp
Entry & Exit	Structured field, for example: 'rear', 'front', 'window', etc.
Victim Details	Age, Sex, Ethnicity
MO' Crime Actions	Recorded as dichotomous variables, for example: 'MO Knife' - yes or no

2.3.2.1 Coordinate Data

Missing data was a constant and consistent problem. 65% (n = 63,802) of records had no coordinate information for offender home locations. Furthermore, 28% (n = 27,302) of records had no coordinate information for crime events. Figure 3 provides a scatterplot view of the crime locations data. Importantly, the data provided is not a representative sample of the entirety of the subject city. Instead, only a small 'slice' of the city is represented. Note that even within the sample of records for which there was coordinate information provided, approximately 34% had to be further dropped as they contained nonsensical values; for example, placing crime locations in the middle of the ocean.

Table 1 shows the distribution of the crimes within the sample across the various police-designated sub-divisions. While these sub-divisions can be a convenient way to group the geographic data, it is unclear what characteristics separate one zone from another – if any. This complicates any analysis that seeks to identify differences in outcomes based on these zones. For example, social disorder theory postulates that an individual's likelihood to engage in street-level criminal activity is dependent on where they live and the social cohesion observed within their home neighbourhood (Sampson & Raudenbush, 1999). Without relying on external sources, the enforcement sub-divisions would allow for at least a rudimentary exploration of what effect the surrounding environment may have on individual offending patterns.

Figure 4 provides a coloured view of the crime coordinate data according to zone. It can be clearly seen from this plot that the data contains a mix of some kind of old zone system – encompassing zones B1 through C4 – and the currently used system (E1 through E4). However it does appear that there is some subset of data belonging to zone B2 that is not otherwise covered by any of the E zones. Finally, Figure 5 provides a map view of a subset of the crime data, coloured by crime sub-division. It can be clearly seen that E1 acts as a buffer around the city centre from zones E2 and E3. Zone E4 is considerably removed the other three; separated by a major motorway. If there are localized differences in activity spaces, then it is most likely that zone E4 is the outlier. Comparisons between the different zones in terms of their offender's activity spaces is covered in more depth in Chapter 3.

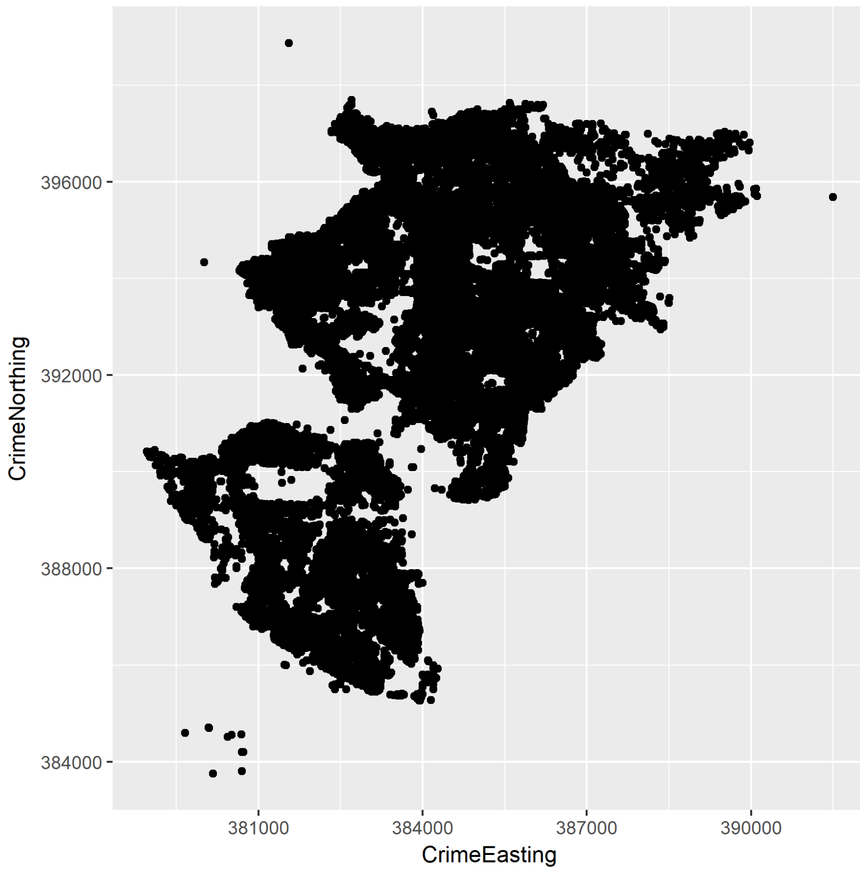


Figure 3: Scatter Plot for Crime Easting / Northing Positions

Table 1: Police Designated Geographic Sub-Division Sample Distribution

Sub-Division	Percent of Sample
E1	19.0%
E2	12.3%
E3	11.0%
E4	13.1%
Other	16.0%
Missing	28.5%

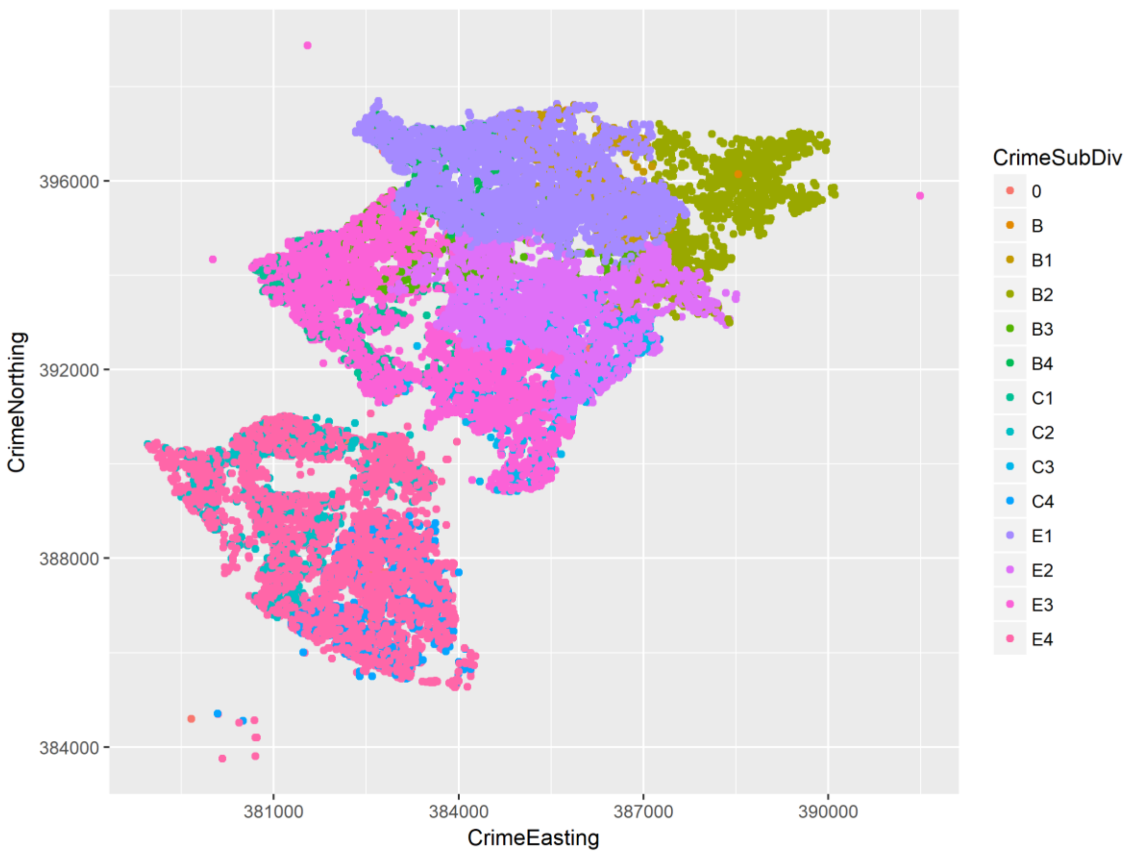


Figure 4: Crime Location Scatter Plot - Coloured by Police Designated Sub-Division

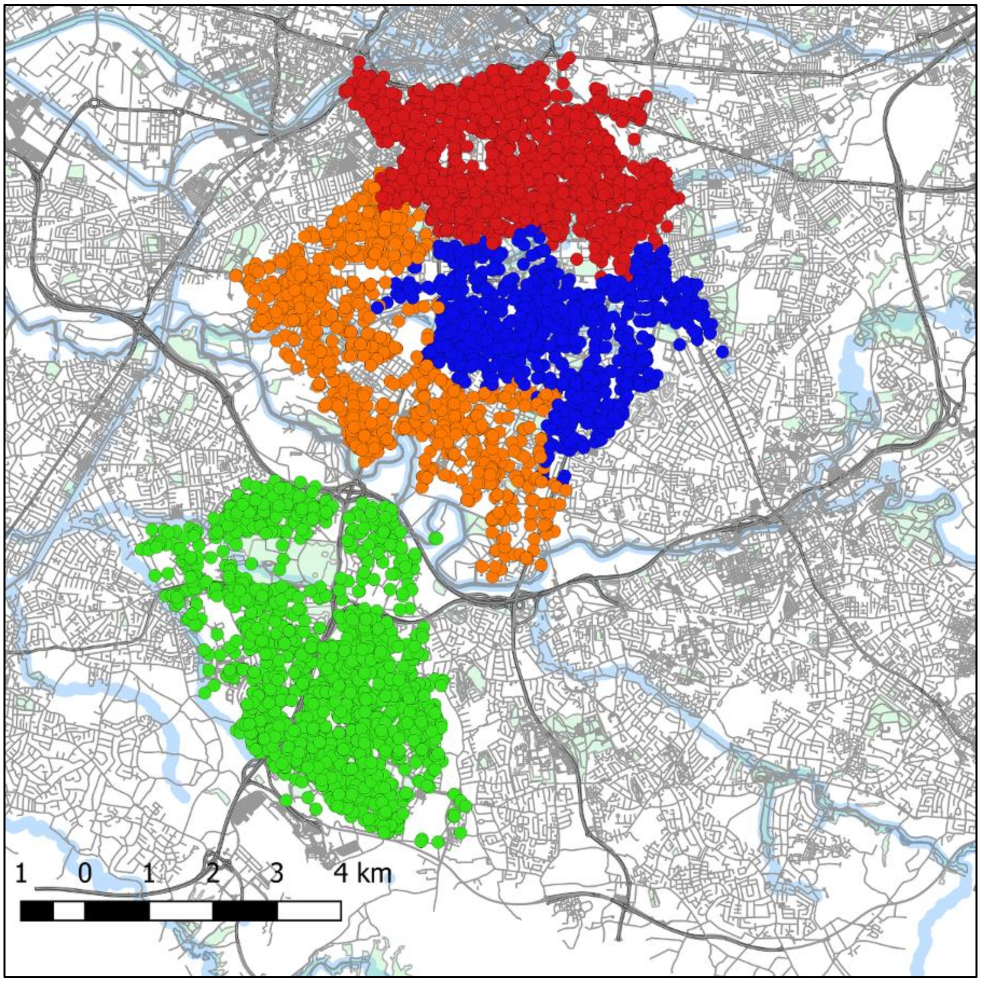


Figure 5: Crime Zone Sub-Divisions Map View

Note: Red denotes E1, Blue E2, Orange E3, and Green E4.

2.3.2.2 Crime Solved Codes

Figure 6 shows the distribution of crimes according to their clearance code. The largest single group is 'Undetected' which can have obvious implications for any subsequent analysis that focuses on cleared crimes only. It cannot be said for certain, for example, that there is not some inherent bias within the crime data here whereby the solved crime has some property not present in the unsolved crime that accounts for why it was able to be solved in the first place. As for what constitutes a 'solved' crime, this is unfortunately a subjective decision. For the current research, a 'solved' crime was taken to be any crime that was ultimately charged or otherwise received some punitive measure – such as restorative justice – or an official warning / caution. Cases where the suspect is only named by the witness, or otherwise had some evidential difficulty, attribution was deemed to be too ambiguous.

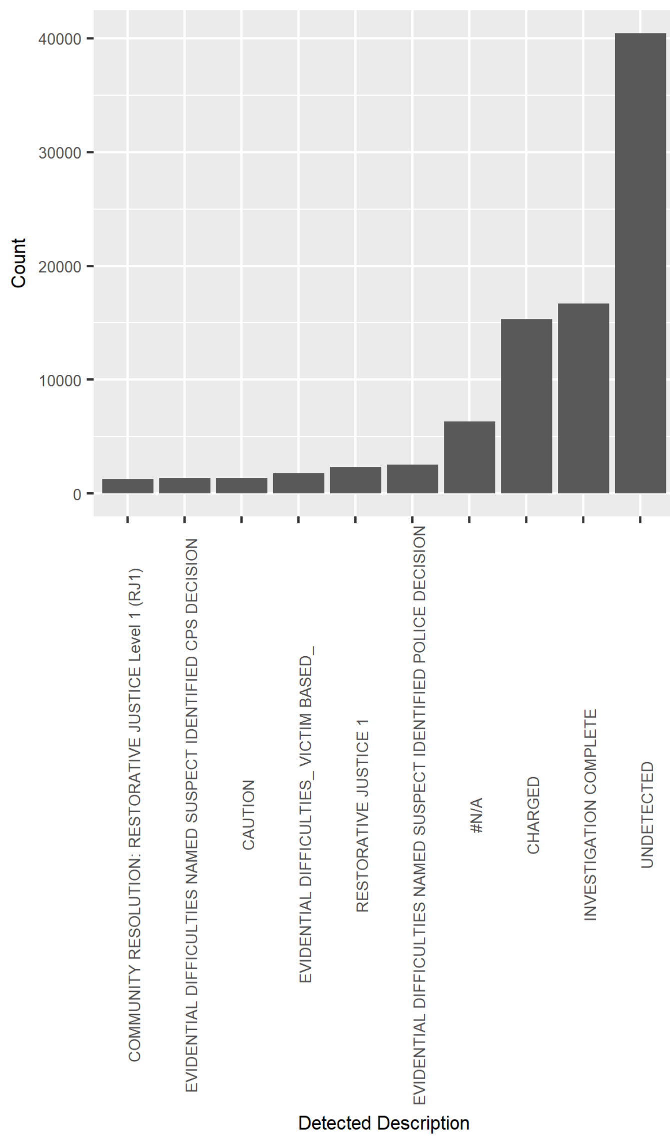


Figure 6: Solved vs Unsolved Code Frequency

2.3.2.3 Sample Time Period – Solved vs Unsolved

Figure 7 and Figure 8 show year over year incidence totals. Figure 7 provides these totals for the entirety of the sample whereas Figure 8 shows those totals for solved (i.e. charged) crimes only. Here it becomes obvious that some data quality issues were introduced over the course of acquiring the data. Recall that the data was provided through a series of physical hand-offs; it is likely that over the course of these hand-offs the data was not always queried in such a way to facilitate a ‘seamless’ transition. Given this data quality issue, it is difficult to discern any meaningful patterns from these data; there does appear to be a general downward trend in the total number of incidents but it is difficult to comment on definitively. Finally, Figure 8 shows a tremendous gap in detected crime between 2013 and 2014. It is unlikely that this is reflective of the true rates of crime for this time and is most likely some form of systematic error in how the data was collected from the master database. It is, however, important to be aware of as it may impact the reliability of measures dependant on date of occurrence. Because subsequent analysis is based on solved crimes only, the data gap identified here is problematic as it is indicative of some form of systematic error within the data. For example, offenders who have committed crimes on either side of the gap would appear the same as those offenders who regularly committed crimes throughout the sampled time period. As will be discussed in later chapters, the frequency with which offenders commit crimes – specifically the temporal delay from one crime to the next – is a common predictor used in linkage analysis, and this censorship of the data will impact such measures.

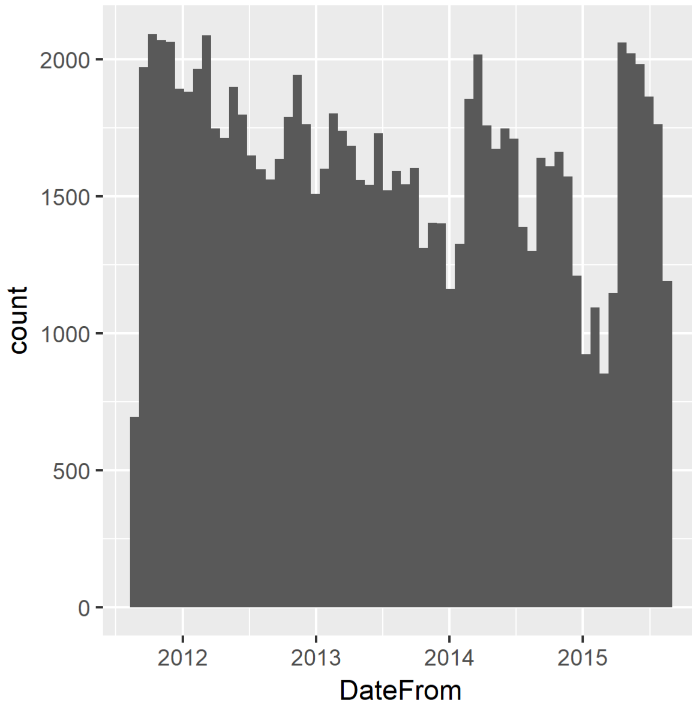


Figure 7: Sample Occurrence Date for both Solved and Unsolved

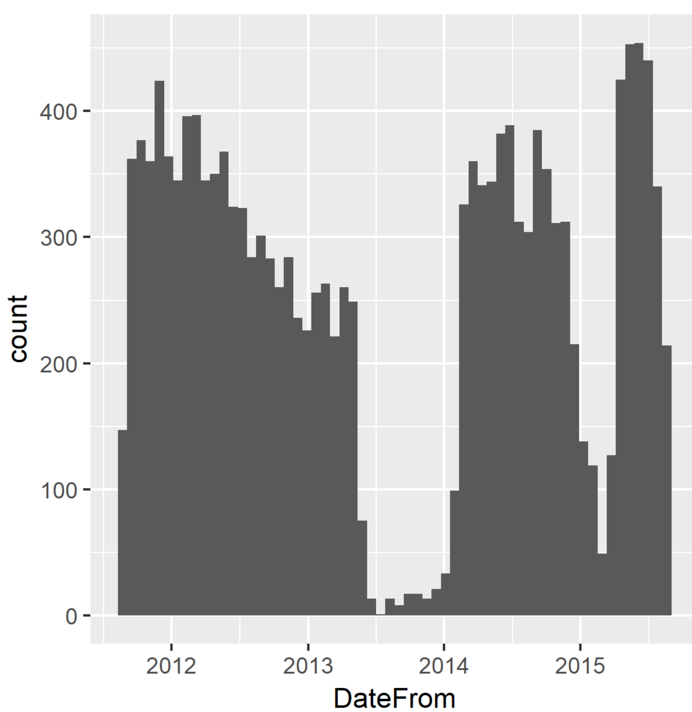


Figure 8: Date of Occurrence for Solved Crimes Only

2.3.2.4 Crime Type Distribution

Figure 9 and Figure 10 show frequency counts by crime group from the sample. Appendix E - Crime Group to Subgroup Dictionary provides a lookup table for the various crime-sub type classifications and the generalized crime groupings used here. Figure 9 shows these counts across the entire sample whereas Figure 10 only shows the counts for solved (detected) crimes. Overall, the distribution of crime types is consistent for both solved and unsolved crime. This suggests that limiting analysis to only solved crimes is reasonable. However, it is still uncertain as to whether there is yet an underlying bias within these crime groups as to certain sub-types of crime being more readily ‘solvable’ than others. As far as the current study is concerned, however, this was not deemed to be a significant issue as there is likely to be a greater degree of variability in how any given crime is classified into a specific sub-type as opposed to belonging into a certain crime grouping. As can be seen from appendix, the sub-type crime descriptions are themselves subject to a number of data quality errors such as spelling mistakes and shorthand that could make rectifying records difficult.

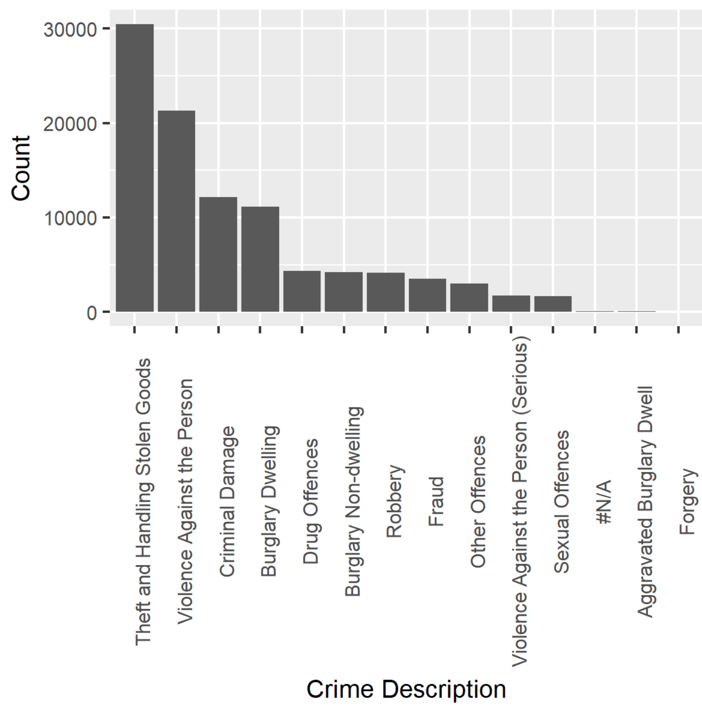


Figure 9: Number of Offenses by Crime Type for both Solved and Unsolved Crimes (Top 20)

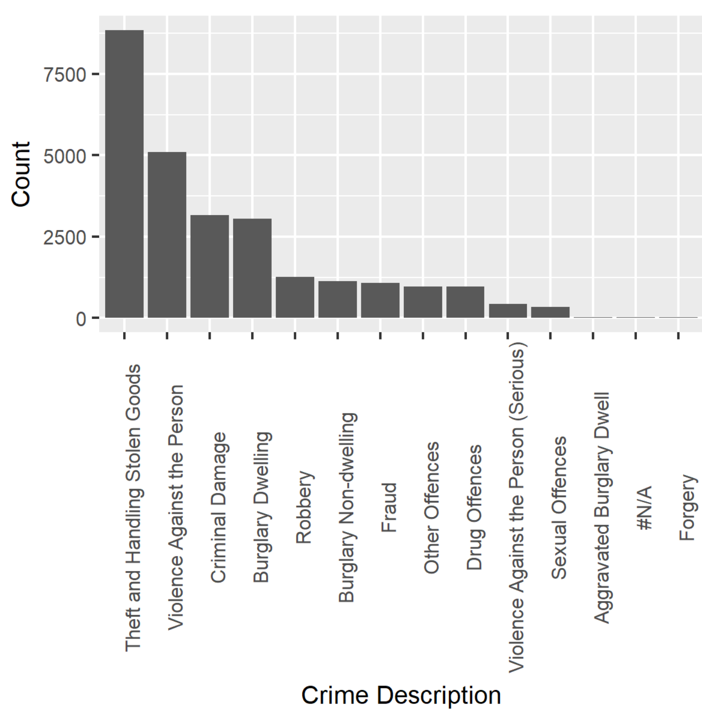


Figure 10: Number of Offenses by Crime Type for Solved Crimes Only (Top 20)

2.3.2.5 Offender Characteristics

To reiterate, the sample consisted of 97,878 unique offences. Of these, 34.7% (n=33,957) were committed by males, 7.6% (n=7,432) by females, and 57.7% (n=56,489) had missing data. Table 2 shows the age distribution of crimes by offender age group within the sample. Unsurprisingly, age data are unknown for the majority of cases which is consistent with the large number of undetected cases present within the overall sample. Of the crimes with a reported offender age, 63% of offences fell within either the Young Adult category (n=11,380; 26.3%) or the Adult category (n=16,138; 37.3%).

Table 2: Offender Age Distribution

Age Group	Total Crimes	Percent of Sample
Youth (10-17)	5,721	5.8%
Young Adult (18-24)	11,380	11.6%
Adult (25-39)	16,138	16.5%
Older Adult (40+)	10,035	10.3%
Missing	54,604	55.8%

2.3.2.6 Criminal Histories

Figure 11 and Figure 12 provide an overview of the offending histories observed within the sample. Figure 11 shows the total number of offenders whom have been observed to have committed a given number of unique offenses. Note these statistics are reported after offenses with null or invalid coordinate data as well as co-charges have been removed. As can be seen, the vast majority of offenders observed within the sample only have one crime event attributed to them. Given that the current sample only contains data from a three-year time span, however, it cannot be known for certain how many of these single-offense offenders have in actual fact committed crimes outside of the current sample. This represents a significant upfront challenge and limitation to activity spaces in general. It is, after all, exceedingly difficult (i.e. impossible) to quantify a 'space' from a single data point. While this is a potentially significant issue (depending on intended use-cases for activity spaces) the current research proceeds under the assumption that if activity spaces are representative of individual activity, then this fact should become evident from the subsequent analysis. If, however, activity spaces are *not* representative in the ways theorised here, then the 'single offense offender' problem is irrelevant. In other words, the current research is concerned with demonstrating the concept first and foremost; leaving further challenges to future research.

Finally, Figure 12 shows the distribution of offenders according to the number of unique crime groups (see: Figure 10) they were observed to have committed within the sample. Again, if one were to focus on a single crime type or crime domain only, then nearly half of all repeat offenders represented here would be missing some subset of attributable crimes, thereby biasing any offender level measures as the researcher would then be working with non-representative sample. In the interest of constructing the most representative total spaces as possible, as well attempting to include as many unique offenders as possible, all subsequent activity space analysis includes all crimes attributable to offenders with no sub-sampling based on crime sub-type or crime group.

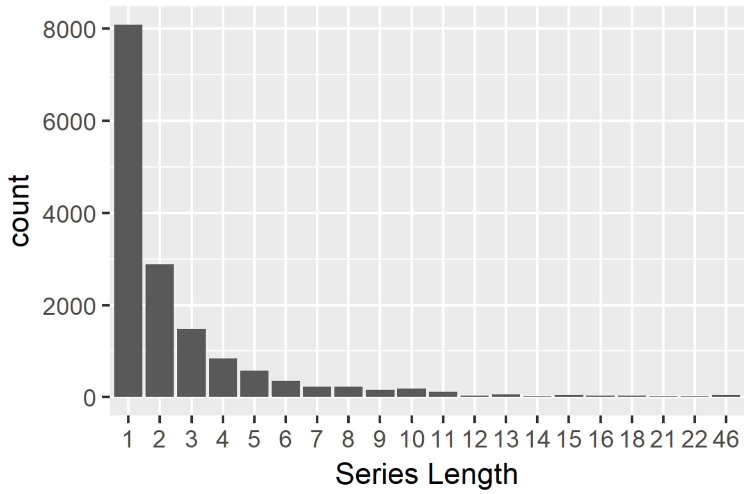


Figure 11: Number of Offenders per Observed Series Size

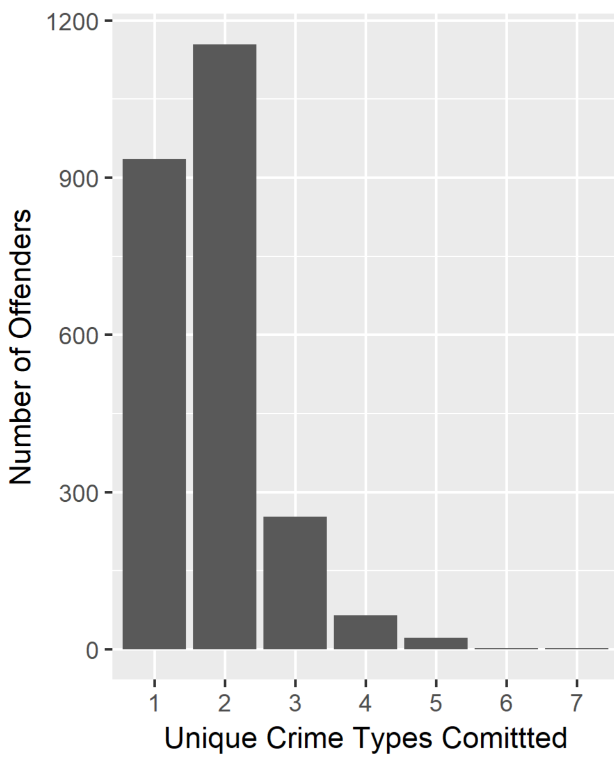


Figure 12: Offender Distribution over Total Number of Unique Crime Types Committed for 2,436 Unique Offenders with 2 or more Crimes on Record

2.3.3 *Modelled Data Descriptive Statistics*

The previous section provided an overall summary view of the data available to the present study. This section, however, provides similar studies for the subset of data which is focused upon for the remainder of the thesis. As will be discussed in subsequent chapters, only a relatively small fraction of data was ultimately used for analysis. Specifically only data that matched the following criterion were included: (1) the crimes had to be solved so that the offender was known, (2) offenders were excluded if they only had two or fewer solved crimes on record, (3) crime location data had to be present, and (4) the crime was a solo (no co-offenders) offense.

Criterion 2 requires some up-front justification: offenders were only included if they had three or more crimes in the data set due to the specific train / test data splitting methodologies adopted in later studies of this thesis. To briefly summarize: for each offender within the dataset, their last recorded crime is set aside to act as a ‘validation’ record, and all other known crimes were included in the model training set. One of the fundamental assumptions being made as part of this thesis is that previous crimes will be indicative in some way of future crimes, and that the activity space will be a useful way to describe this relationship. However, in order to estimate an activity space, the methods outlined in Chapter 3 require at a minimum of two data points. Thus each offender needs at least two data points to estimate an activity space from then then a final data point to act as a validation record; hence the three observed crimes per offender requirement. Overall, these criterion result in 3,449 total crimes for 997 unique offenders.

Generally speaking, this (much) smaller subset of data did not appreciably differ from larger data set; one might expect, for example, that the types of crimes committed by these more prolific offenders may systematically differ from the larger set of crimes which did not appear to be the case. Table 3 and Table 4 show the recorded crime-group distribution for the complete data set as well as the modelled data subset only. Again, differences are small as both datasets were dominated largely by theft and stolen goods offences (31% for all data, 26% for modelled data) and violence offences (21% for all data, 39% for modelled data).

Figure 14 shows that offenders with multiple offences in the data set offended at rates in-line with the overall dataset, and again ranges from August 2011 through August 2015. Furthermore, the 2013-2014 gap is still present. For some subset of offenders this gap will be significant: they will fall out of the inclusion criterion due to the data set having omitted some of their crimes, and for others it may introduce a slightly bias to longer crime-to-crime delay intervals due to the gap masking more regular offending patterns. Finally, Figure 15 shows that offenders that offenders within the modelled data engaged in similar levels of speciality regarding their offences with the majority of offenders having committed two or more distinct types of crime over the observed time frame.

Table 3: Complete Data Set - Crime Group Distribution

Crime Group	Total	Count	%
Theft and Handling Stolen Goods	97878	30432	31.09
Violence Against the Person	97878	21276	21.74
Criminal Damage	97878	12142	12.41
Burglary Dwelling	97878	11151	11.39
Drug Offences	97878	4352	4.45
Burglary Non-dwelling	97878	4236	4.33
Robbery	97878	4123	4.21
Fraud	97878	3534	3.61
Other Offences	97878	2988	3.05
Violence Against the Person (Serious)	97878	1763	1.8
Sexual Offences	97878	1661	1.7
Missing	97878	87	0.09
Aggravated Burglary Dwell	97878	81	0.08
Forgery	97878	52	0.05

Table 4: Modelled Data Only - Crime Group Distribution

Crime Group	Total	Count	%
Violence Against the Person	3449	1334	38.68
Theft and Handling Stolen Goods	3449	894	25.92
Criminal Damage	3449	316	9.16
Drug Offences	3449	275	7.97
Burglary Dwelling	3449	167	4.84
Other Offences	3449	108	3.13
Fraud	3449	101	2.93
Robbery	3449	93	2.7
Violence Against the Person (Serious)	3449	63	1.83
Sexual Offences	3449	55	1.59
Burglary Non-dwelling	3449	36	1.04
#N/A	3449	3	0.09
Aggravated Burglary Dwell	3449	3	0.09
Forgery	3449	1	0.03

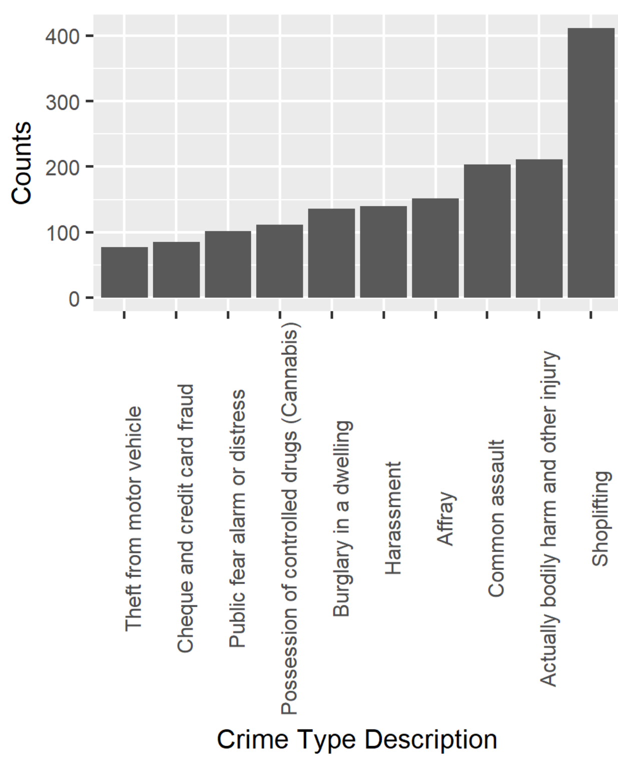


Figure 13: Modelled Data - Crime Types Distribution

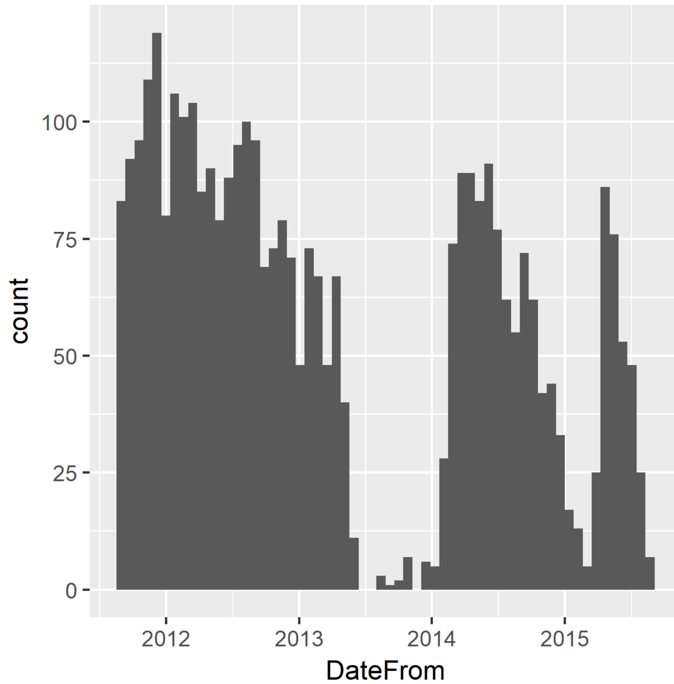


Figure 14: Modelled Data - Crime Occurrence Date Distribution

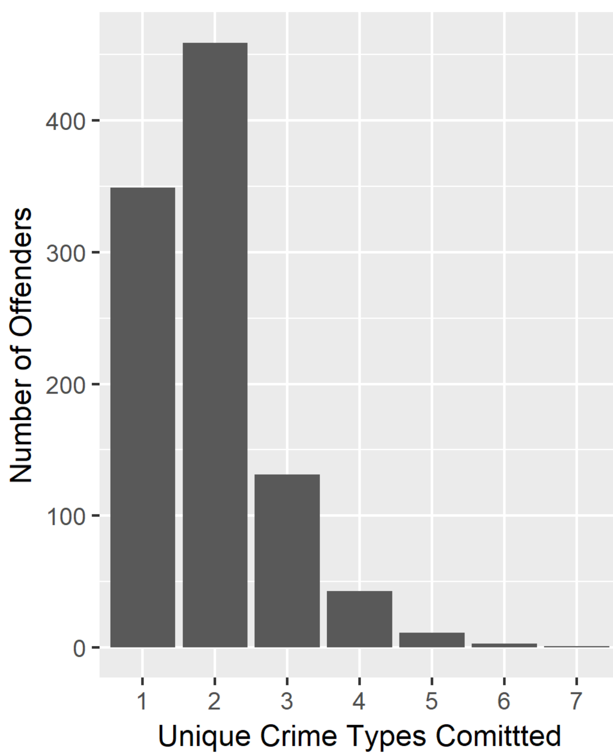


Figure 15: Modelled Data - Distribution of the Number of Unique Crime Types per Offender

Chapter 3

Criminal Activity Space

3.1 Introduction

This chapter focuses on addressing research question set 1 by introducing activity spaces formally. A brief introduction into the current activity space literature is provided which is then followed by an extensive discussion of several possible methods for calculating activity spaces. This chapter concludes with a brief study investigating the predictive power of a chosen activity space estimation method which is used throughout all subsequent empirical chapters.

Activity space based studies have begun to see a resurgence in recent years in a number of fields as a means to more meaningfully evaluate individual exposure and engagement with their environments (Greenberg Raanan & Shoval, 2014; Lee et al., 2016; Parthasarathi, Hochmair, & Levinson, 2015). This increase in interest can be attributed to the increase in availability of accurate GPS data, as used by Lee at al. (2016), as well as the increased availability of GIS (geographic information systems) software (Parthasarathi et al., 2015). The widespread use of GIS within policing has seen a similar uptick in popularity in recent years (Fitterer, Nelson, & Nathoo, 2015; Ratcliffe, 2004; Wang, 2012). GIS and crime mapping for crime, however, has remained a largely administrative exercise focused on the identification of aggregated trends of crime for given *places*. As has been argued in other fields such as health and epidemiology, an increased focus on *individual*-based measures, such as activity space, could provide more representative measures of criminal spatial behaviour.

Activity space based research has typically been conducted in one of two ways: (1) by the self-reporting of trips via a travel diary or equivalent, and (2) via the direct acquisition and analysis of GPS or other Location-based service (LBS) generated data. Which data collection method to use depends heavily on the specific goals of the research project, but generally small-scale qualitative studies are better served by the former and large-scale quantitative studies by the latter. Data used for the current research was provided by a U.K. metropolitan police force and due to its size and scope it would have been impractical to conduct individual interviews. This creates several specific challenges not necessarily faced elsewhere in the literature concerning activity space calculation and estimation. These challenges include: inaccurate or incomplete data as it relates to the timing or duration of the criminal offence, or even inaccuracies in the location of the event itself. Furthermore, unlike in similar non-criminal oriented studies, the specific goal – *why the criminal event took place* – is often unknown.

Uncertainties such as those above, while not unusual within the literature, have direct impact on how and why a given estimation method for activity space is used, and what the resulting geometry of that space implies about an offender’s behaviour. The literature is exceptionally sparse when it comes to the direct measurement or formulation of activity space as it relates to offenders and criminality. While it is generally accepted that crime can often be understood as occurring as a result of the rhythms of otherwise non-criminal day-to-day living, there are no specifics as to how to capture and represent such patterns. For example, Brantingham and Brantingham’s (1995) crime pattern theory, while stipulating that individual activity spaces play a role in creating the conditions for a criminal event, offer no empirical methods for calculating or evaluating such a “criminal activity space”. Addressing this shortcoming is the primary goal of this chapter.

3.2 Activity Space Estimation

The propositional definition of an activity space has changed little since the early presented works of Horton and Reynolds (1971), and Johnston (1972). The methodologies employed to study activity spaces, however, have been wide and varied (Chaix et al., 2012; Dijst, 1999; Sherman et al., 2005). These range from the conceptual formulations that relate places to each other as presented by Perchoux (2013), to the more rigorous mathematical formulations of Sherman (2005) and Chaix (2012). In both cases, it is largely accepted that there exists an important relationship between conceptual “anchor points” – typically the home or workplace – and other areas of daily activity. Furthermore, it is largely accepted that destinations individuals indicate they prefer tend to follow a rule of distance decay: namely as distance increases, the likelihood of engaging with a particular location diminishes (Golledge & Stimson, 1997). Criminal events tend to follow similar patterns of distance decay (D. Canter et al., 2013; Emeno & Bennell, 2013; O’Leary, 2011; Rossmo, 2000), and the assumed relationship between criminal events and day-to-day activities serves as the foundation upon which both routine activity and crime pattern theories reside (Brantingham & Brantingham, 1995; Cohen & Felson, 1979).

These observations – distance decay and routine activities – underpin the various operationalized formulations of activity space. Broadly speaking, these operational approaches fall into one of several categories: (1) elliptical methods with the standard deviational ellipse (SDE) being the standard, (2) network based measures such as the road network buffer (RNB) from Sherman (2005), and finally (3) other polygonal definitions such as the minimum convex hull. These categories can be thought to vary as a function of their complexity and the number of assumptions made in their formulation. These three approaches relate to the foundational theories from Chapter 1 in specific ways: the elliptical and polygonal surfaces describe a generalized area which relates directly to the awareness space construct; the network based measures have the most in common with crime pattern theory which emphasises crime in relation to the common routes travelled by offenders. It is useful to inspect visual examples of activity space surfaces, and Figure 16 provides location data for an example crime series to serve as a reference; this crime series will be used in a number of examples throughout this section.

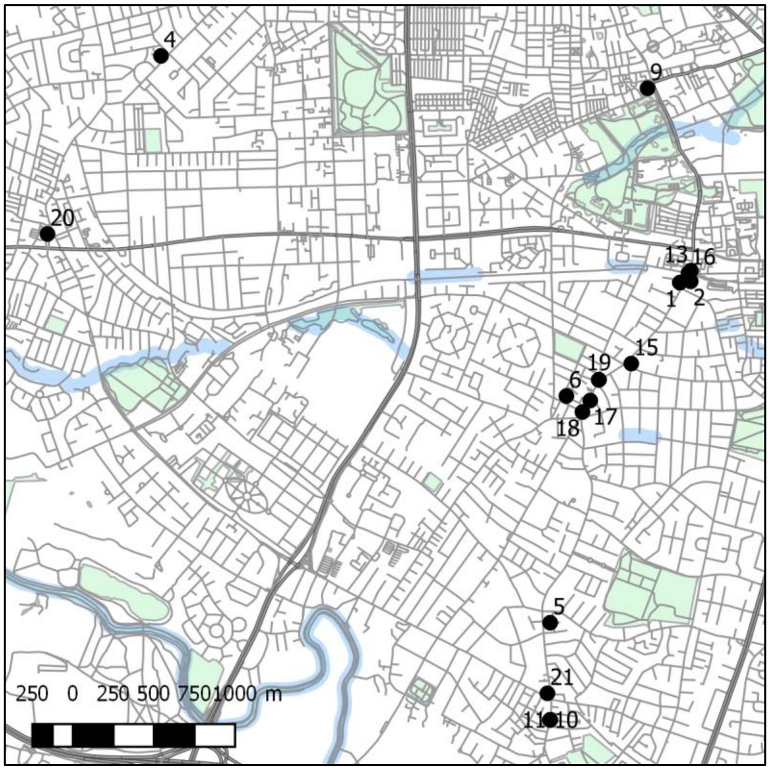


Figure 16: Crime Series Example

3.2.1 Convex Hull

The least complex of the three is the convex hull which is simply the smallest encompassing polygon for a set of locations. This measure is described as “simple” because it makes no implicit assumptions about the relationship between the home and activity locations, nor does it take into account any travel networks. Furthermore, by definition when ignoring the possibility of spatial outliers, the convex hull will always include all observed activity locations. The disadvantage of such an approach, however, is that it can imply regions of familiarity for individuals where no such familiarity exists (Chaix et al., 2012, p. 448). The convex hull for the reference data are provided in Figure 17.

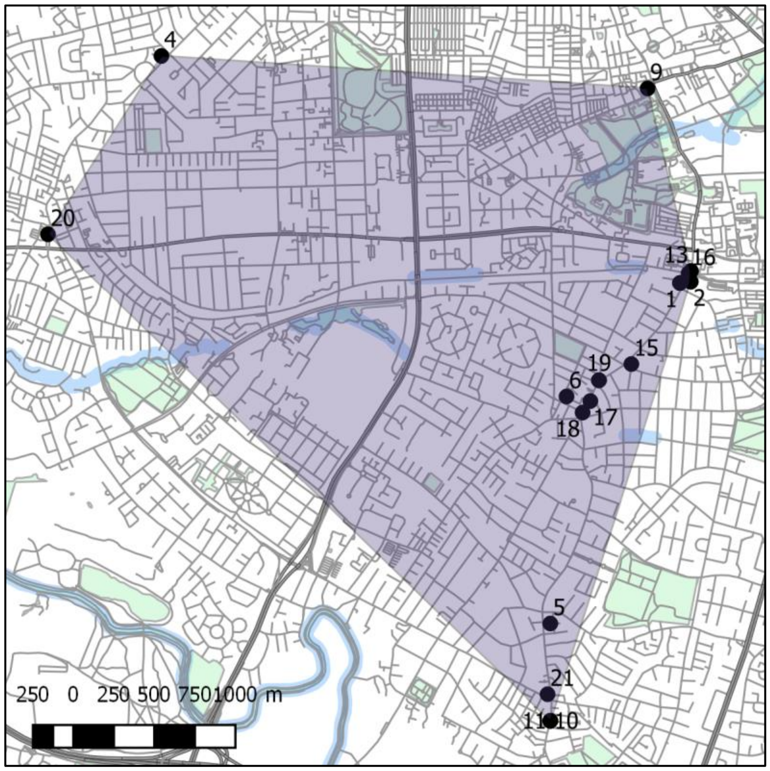


Figure 17: Convex Hull Example

3.2.2 Elliptical Methods

Elliptical methods are another approach to generating a simple geometric shape to describe an activity space. There have been a number of methods explored within the literature, including circular methods (D. Canter & Larkin, 1993), and standard deviational ellipses (Buliung & Kanaroglou, 2006; Lee et al., 2016). Due to the mathematical requirements of defining an ellipse or circle, elliptical methods make a number of implicit assumptions that convex hulls do not.

Specifically, they require some central point around which to draw the circle or ellipse as well as some indication of range which represents either the radius of a circle, or the major and minor axes of an ellipse. Typically, the centre of the ellipse is calculated as the geometric mean of all observed activity locations (including the home). In the case of the SDE, the major and minor axes are calculated as the standard deviation distance in the X and Y coordinate-direction for all points from the mean centre. As Sherman et al. (2005) observe, the SDE does not necessarily encompass all observed activity locations, and like the convex hull it does not make any implicit assumptions regarding travel and how the observed activity locations may be connected.

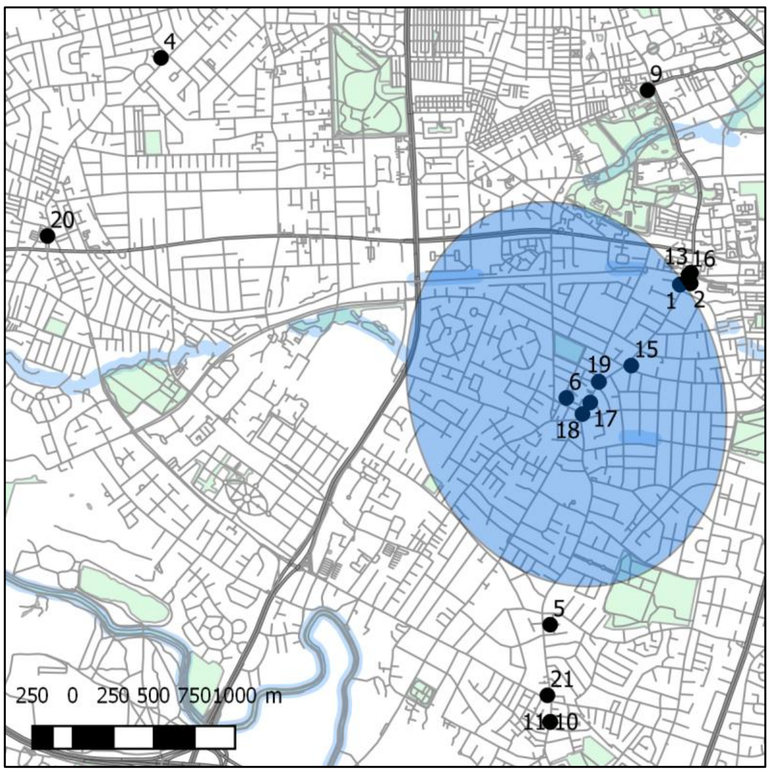


Figure 18: Standard Deviation Ellipse Example

3.2.3 *Network Based Measures*

Network based measures, unlike the previously mentioned measures of convex hulls and SDEs, are an attempt to account for travel routes between observed activities. Such methods typically attempt to estimate the shortest travel path between temporally adjacent activities and then apply some buffer distance around this route. Like the convex hull, road network buffers (RNBs) typically cover all observed activity locations. This is because the RNB space is defined by three parts: origins, destinations and travel paths. The RNB has the most in common with time-space geography principles of the three general approaches discussed thus far, but it comes at cost. In instances where perfect travel information is not known, as is typically the case with crime data, there is no guarantee that the route chosen reflects actual travel patterns. From a practical perspective, RNB like calculations require extensive data on underlying travel networks such as roads, busses, and trains that may not always be available, accurate or complete.

Network based measures like the road network buffer are all a subset of analysing means of travel and pathways. This form of analysis is particularly difficult with crime data as it can often be difficult or impossible to not only identify a specific time of occurrence but also identify the means and route of travel. The RNB, for instance, assumes that individuals follow roads to reach their destinations. The appropriateness of this assumption will obviously vary from data set to data set as the underlying geography changes. To further complicate matters, any pathway level analysis necessitates that a pathway be made up of a start and end point. Geographic offender profiling (GOP) literature argues that such 'anchor' points are often an offender's home or equivalent (D. Canter et al., 2013; Canter & Hammond, 2006; Rossmo, 2000).

The relationship between the home and crime locations is not always clear even within the GOP literature - see: Canter and Larkin (1993) for an early example of the 'commuter /

marauder' typology. And so if the home location is unreliably inaccurate or otherwise unrelatable to a set of crimes, then what is left for establishing a pathway to study? The only data points that remain would be the crime locations themselves, but it seems highly unlikely – except in instances of extremely rapid 'spree' attacks – for any given crime location to be directly related to one another by direct travel.

3.3 Activity Space Uncertainty

There is no consensus within the literature as to which of these approaches is "best" (Buliung & Kanaroglou, 2006; Dijst, 1999). This is because, as Buliung and Kanaroglou point out, "the behavioural insights drawn from activity/travel measures are not necessarily consistently observed across different measures." (Buliung & Kanaroglou, 2006, p. 44). This is easily observed from the two provided surface examples: given the difference in absolute size alone it is obvious that convex hulls and SDEs would give different results of environmental exposure.

Studies of crime and criminality can be conceptualized into two primary areas of focus: (1) the criminality of place which is primarily interested in understanding why crimes occur in the areas in which they do for a given population, and (2) studies of specific criminal motivation which is primarily interested in answering why do offenders commit the crimes they do in the ways in which they do. The study of criminal activity spaces occupies a grey zone that lies in the gap between these two areas of focus; as attempting to study criminal motivation in relation to target choice is inextricably linked to the environment in which they operate.

Despite this complication, however, target location choice is nevertheless important. Recall the earlier provided example concerning the predictive impact of target choice given crime type: shoplifting offences are often thought to provide less information regarding target selection than burglary due to the fixed distribution of shopping centres. However, unless there is only one shop to victimize, for example, then the choice of a specific shop *is* significant. This line of reasoning can also be applied to the underlying environment. The natural conclusion is that the crime location is itself of paramount importance.

What about those crime instances which arise seemingly by chance: a predisposed offender encounters an unlocked door to an apartment for example? The implicit difference between opportunistic crime and premeditated crime and how those two crimes may or may not be spatially manifest is of particular importance when attempting to construct a valid measure of criminal activity space. This observation suggests that crime events, even for the same offender, may come about as the result of multiple different motivations. In such cases, it is not obvious that a single geometric surface is the “best” choice in representing an activity space.

The discrepancies highlighted here are primarily attributable to different research focuses; as they are by and large tailored to answer specific research questions which are not always universal. For example, RNB measures may be employed when attempting to study travel patterns and traffic or congestion, whereas the SDE may be used to study travel dispersion. With regard to crime, the usability of the outlined approaches is dependent largely on the specific research question as well as data availability. Specific route information, and / or accurate home locations of offenders are not always readily available which would render RNB and other travel network based measures very difficult to reliably and meaningfully implement.

At their core, the various activity space based measures are estimation methods of individual spatial movement. As such, they are often inaccurate or incomplete snapshots of individuals in space and time. Thus there is a certain degree of uncertainty in any geometry used to describe mobility. The various buffering techniques used, such as in RNBs and SDEs reflect this uncertainty. By borrowing on concepts from time-space geography, these uncertainties can be viewed as features which describe the hypothesized potential access an individual holds to a given area.

The buffering techniques outlined above can be said to describe areas for which individuals possess sufficient knowledge to engage with the environment, but as of yet have not been observed to do so. As previously mentioned, this is not a straight forward process as there are many confounding variables which can unfairly inflate an activity space such that it is not representative of an individual's true awareness space (Buliung & Kanaroglou, 2006; Chaix et al., 2012). Chaix et al. (2012) provide the example of high speed trains connecting two distant geographic areas together and the resulting activity space geometry covering a large area the individual has no appreciable knowledge of. The use of a buffer distance, however, is necessary in some capacity as without one activity spaces could only describe specific locations where individuals have been observed in the past and thus have little future predictability.

This discussion leads to the conclusion that the size of a given activity space reflects two properties: first is that it is reflective of some observed behaviour – typically travel range, and second it reflects some degree of uncertainty in the awareness space as specified by the researcher. Put another way, smaller spaces could be indicative of either smaller actual travel ranges or less approximation on behalf of the researcher or measure. This is not to say, however, that activity space size is indicative in any way of precision or utility of a given measure. Rather, the estimation method used will impact the relative size of an activity space and this in turn will impact what behavioural inferences can be generated. In a study of accessibility, for example, using a measure that generates large activity spaces (larger uncertainty) necessarily results in larger potential exposures, but there is no guarantee that such a measure would be in any way accurate. This implicit trade-off between certainty and uncertainty given size requires careful up-front specification.

3.4 Data Clustering and Activity Space

A fundamental problem associated with the various measures of activity space discussed previously is the loss of information. In attempting to describe an activity space in terms of a single shape, important local structural entities can be lost. This was exemplified by Sherman et al. (2005) who used a variety of activity measures to study accessibility to healthcare centres. The authors found that which activity space measure used resulted in sharp differences in both the total area of the measure, as well as differences in the total numbers of accessible locations captured by the measure. All of the measures provided by Sherman et al. were non-composite single geometry surfaces – that is that they consisted of a single geometric shape with a single centre point. In the context of an accessibility study this approach makes sense; data collection is controlled and respondents can be interviewed to ascertain what their motivations were when they travelled to a given location; non-routine activities or atypical locations are not likely to be included in such a study. In the case of crime data, however, it is not always known if an offender travelled to a given location with the express intent to commit an offence, or if they were there for some other reason – routine or otherwise. Clustering crime locations can be a beneficial way to capture different local spatial structures if there are in fact different focal areas at play.

Clustering locations is not a common practice within the activity space literature; most researchers prefer to describe an activity space with a single surface (Kestens et al., 2010; Lee et al., 2016; Sherman et al., 2005). However, clustering has been proposed at least conceptually numerous times: Brantingham and Brantingham (1993) presented the idea of a “node” which represents a notional area to which an individual may travel in order to fulfil a specific activity, such as entertainment. The schematic relationship between nodes, paths and anchors as described by Brantingham and Brantingham is also mentioned by Perchoux et al. (2013) in non-criminal contexts. Furthermore, while Dijst (1999) used clustering to categorize different activity spaces observed over time into notional categories of shape, his use of clustering differs in intent from what is implied by crime pattern or routine activity theory. Routine activity theory, like crime pattern theory, suggests crimes, like other activities, may themselves "cluster" around routine activity destinations.

The goal of any clustering algorithm is to identify groups, or clusters, for which observations are most "similar". Notionally, this similarity is analogous to "closeness" in space – in other words points within a cluster have shorter distances between themselves than with points outside of a cluster. This results in two key concepts for understanding the efficacy of a given clustering approach: (1) compactness which describes how dispersed *points* within a cluster are, and (2) separation which describes how dispersed *clusters* within the data are (Kovács, Legány, & Babos, 2005). Conveniently, compactness and closeness correspond to the concepts of propinquity and morphology respectively from Chapter 1. Plainly speaking a "good" clustering is achieved when each cluster is highly compacted and each of those clusters is well separated from each other.

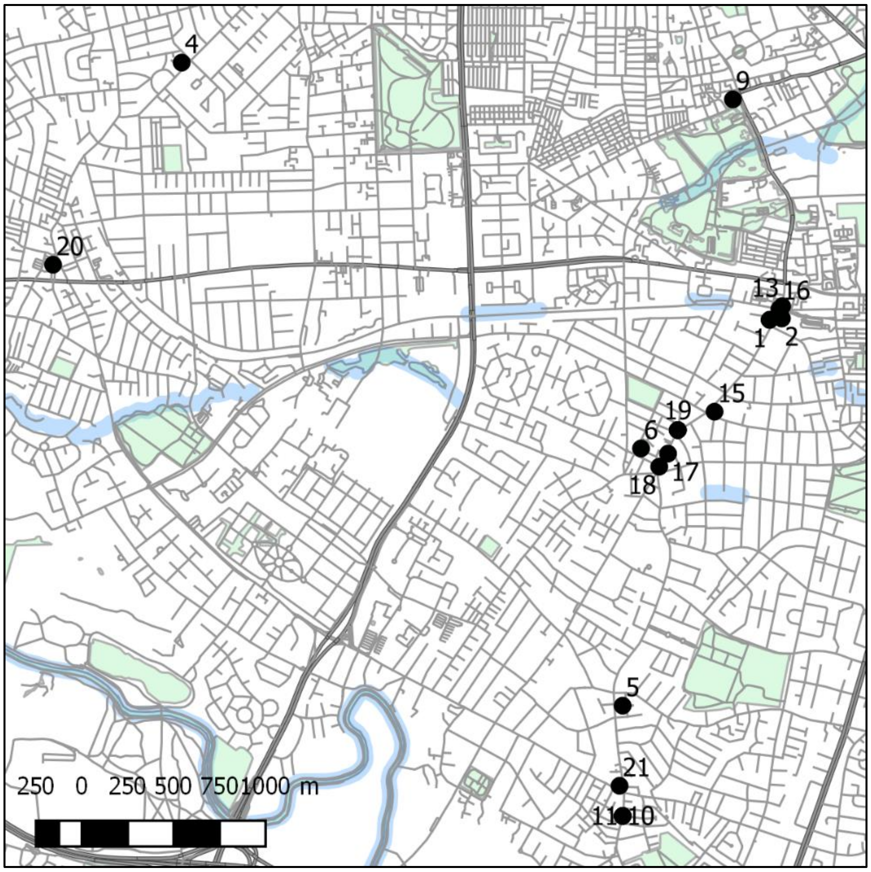


Figure 19: Example Crime Series - How Best to Cluster?

Figure 19 provides a visual example of a crime series where clustering may be warranted. Intuitively, the crime series above may be divided into three or four possible clusters corresponding to three or four possibly distinct activity nodes. Conversely, if no clustering were performed, and a single geometric surface was used to describe this crime series, then the resulting surface would be very large. A single geometric surface implies that all points belong to the same cluster, however this is not very satisfying as it would encompass a very large area for which there is no direct evidence to indicate the offender engages in activities there (see: Figure 17, p. 74 for an example). Clustering crime locations can be used as a means to reduce uncertainty in the final geometry used to describe an activity space while ensuring that observed locations are fully represented. Furthermore, clustering crime locations can help to identify focal areas for a given offender.

There are a number of methods available to perform cluster analysis. These include: hierarchical, k-means, density models and distribution models. Much like the choice of which geometry to represent an activity space, there is no definitive "best" clustering approach to take. Again like the choice of which geometry to use, the choice of which clustering approach to use will often be guided by application. Of the four listed approaches, hierarchical and k-means clustering are the "classical" clustering approaches, with the two former approaches being more modern algorithmic optimizations. However, as hierarchical and k-means clustering methods are distance based measures as opposed to density or distribution based models, they are the most intuitive of the methods to understand in the context of clustering locations data.

3.4.1 Hierarchical Clustering

Hierarchical clustering is a clustering process by which a dendrogram, a type of tree diagram, is created from a set of points which relates the spatial distance between observations. The name "hierarchical clustering" comes from the fact that the dendrogram will merge groups of closely spaced observations together as one moves up the tree; or split groups into distinct clusters as one moves down the tree. There are two ways to perform hierarchical clustering: (1) divisive clustering and (2) agglomerative clustering. These correspond to a "top-down" approach (divisive) versus a "bottom-up" approach (agglomerative). Agglomerative clustering is performed by first treating each observation as its own unique cluster at the base of a tree diagram. As one moves up the tree, each cluster is merged with its nearest neighbour until all points are placed into the same cluster. The resulting tree diagram is referred to as a dendrogram. This process of starting from each individual observation and subsequently merging nearest points is where the “bottom-up” description and hierarchical structure originates. The advantage of this approach is that for a given data set there is one dendrogram, i.e. the method is deterministic (Chidananda Gowda & Krishna, 1978). An example of agglomerative clustering using data from Figure 19 is provided in Figure 20.

Divisive clustering, as the name suggests, involves dividing observations into subsequently smaller groups. Unlike in agglomerative clustering, divisive clustering first starts with all observations as one cluster. From there, clusters are separated into smaller groups until each observation is in its own unique cluster. The dividing criterion used in divisive clustering is not as straight forward as the merging criterion used in agglomerative clustering and can change depending on context or use (Dhillon, Mallela, & Kumar, 2003).

The primary disadvantage of both of these methods is that they do not implicitly indicate at which point within the tree to “stop” the clustering – all possible clustering formulations are presented. The decision of how many clusters to use to describe the data are left to the researcher. Furthermore, hierarchical clustering, especially agglomerative clustering, is sensitive to spatial outliers. Handling outliers in the context of data clustering will be revisited later.

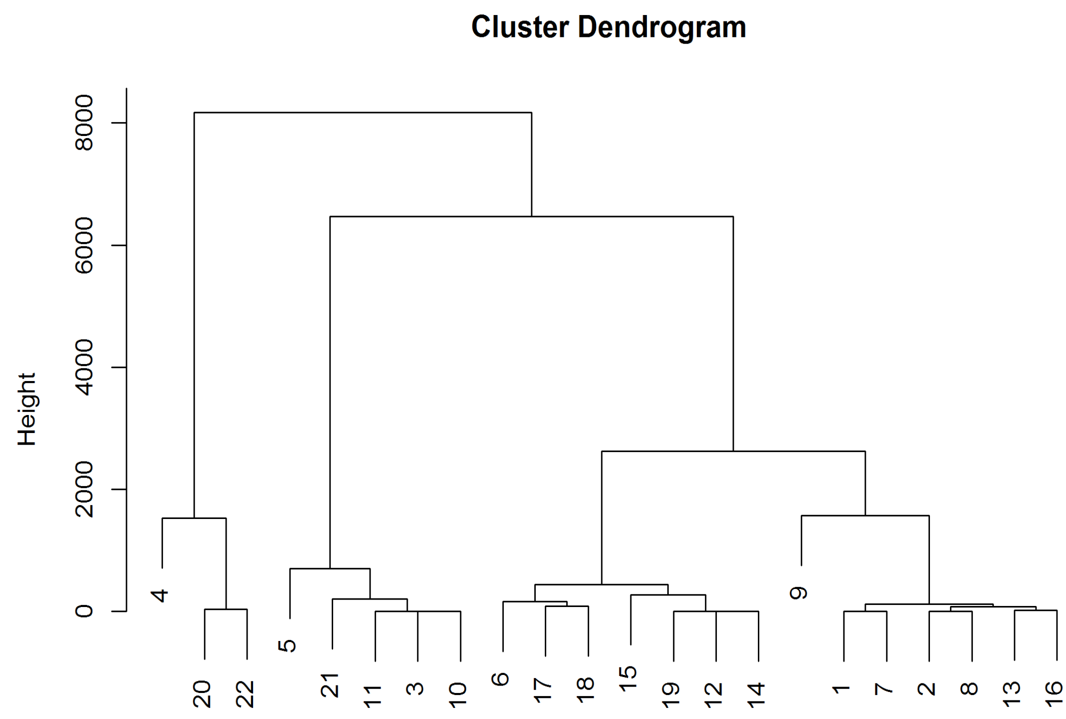


Figure 20: Hierarchical Clustering Dendrogram Example

Note: Numbers here denote node names; in this example node 20 and 22 are closest to each other. Different 'heights' denote different tolerances for what constitutes a 'cluster'.

3.4.2 K-means Clustering

K-means clustering is a variance-based form of clustering (Kanungo et al., 2002). Specifically, k-means clustering involves finding *k* centres from which the mean squared distance from the centre to each data point within the cluster is minimized. In other words, k-means is a method by which the observed space is divided into *k* regions where each region represents a cluster. K-means clustering is related to nearest neighbour analyses and is a common clustering approach (Kanungo et al., 2002). A fundamental problem with k-means, however, is that the number of clusters, *k*, must be specified up-front (though there do exist techniques and measures to help in choosing reasonable values of *k*). This is because unlike in hierarchical clustering, k-means algorithms only identify the number of clusters for which the researcher specifies. Furthermore, identifying where the centres should be placed is not deterministic, and thus implementation of k-means requires multiple iterations to ensure that centres are reasonably placed. Figure 21 provides an example for the data used in Figure 19 clustered by k-means when k = 3 and k = 4 respectively.

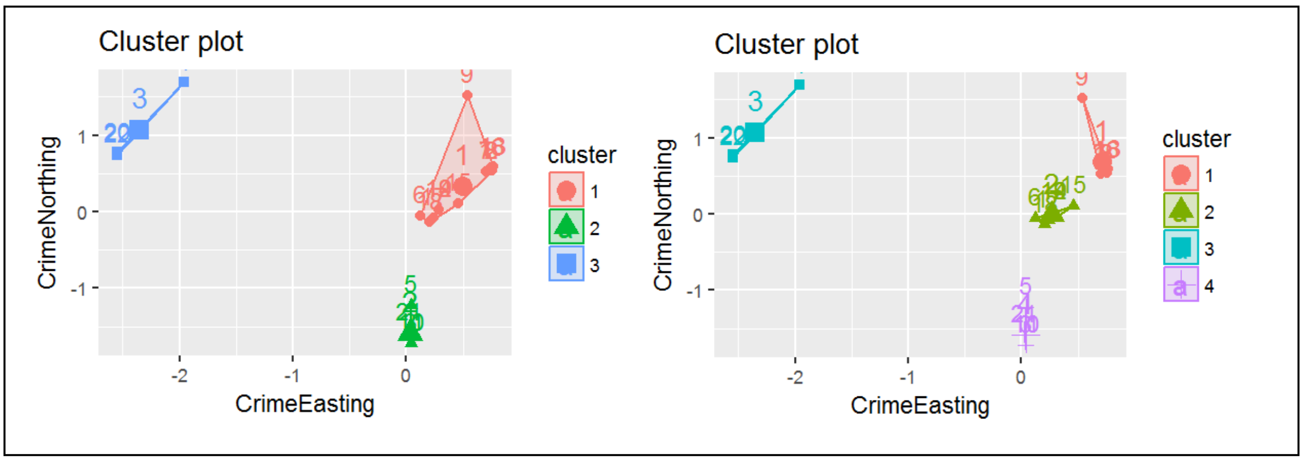


Figure 21: K-means Clustering Example; k = 3 and 4 respectively

3.4.3 Cluster Validation

There are two primary concerns with any clustering technique: the first is the specification of the number of clusters. This value is either explicitly described by the researcher “up-front”, as in k-means, or it is specified implicitly by the optimization function used in the algorithm which determines clusters. The second is assessing how “successful” the clustering was in identifying meaningful clusters. Recall that clusters can be assessed by their “compactness” and their separation; that is how tightly observations within a cluster are grouped together and how distant the different clusters are from one another.

Rousseeuw (1987) presents a method by which the quality of an identified cluster structure can be assessed in terms of their compactness and separation. He presents this in the form of the silhouette plot, which offers a visualization, as well as an empirical metric, of how well observations are clustered. The silhouette plot is constructed using a dissimilarity measure calculated for reach point in a cluster. The dissimilarity score quantifies how “well” the points are clustered, and ranges from -1, the observation belongs in a different cluster, to 1, perfectly clustered. Observations with a dissimilarity score close to 0 are located in between two clusters and it is not obvious to which cluster they belong. Because the dissimilarity score is based on average distances between observations, it offers a very direct and intuitive way to assess the effectiveness of clustering on the offense location data available (Rousseeuw, 1987).

The argument presented here is that generally there is less structural information loss associated with many smaller clusters for a criminal activity space than a single larger cluster. This is hypothesized to be due to the specific nodal attractors associated with a given cluster differ from the attractors present in other clusters. These differences are thought to be either environmental differences in target dispersion, or differences in the routine non-criminal goals the offender associates with a particular node.

3.5 Spatial Outliers

Much like the buffer zone and clusters described previously, outliers have a behavioural analogue. In the case of buffer zones, their use was justified through time-space geography as capturing a likely area for which that individual may interact with in the future – i.e. the potential activity space (see: Figure 19). Outliers, on the other hand, are obviously different in that they are locations at which the individual has already been observed to engage in activities. The justification for their potential removal stems from the idea that not all activities are indicative of future behaviour, which coincides with the traditional definition of an outlier. Behaviourally speaking, an outlier can be understood as arising from one of three possible scenarios: (1) it is a one-off event, (2) it is one of many (to come) events, or (3) it is an artefact of incomplete data.

The concept of an outlier in statistical terms is well documented (Leys, Ley, Klein, Bernard, & Licata, 2013), and the *Encyclopaedia of Mathematics* defines an outlier as "any observation in a set of data that is inconsistent with the remainder of the observations in that data set." (Hazewinkel, 1988). Hawkins (1980) defines an outlier as: "an observation that deviates so much from other observations as to arouse suspicion that it was generated by a different mechanism." An inference can only be considered valid once one removes possible outliers which may be skewing results. Most are familiar with the influence outliers have on non-robust measures such as the mean, and in the case of the clustering activity space framework being presented here their effect is no less noticeable. The process by which one goes about detecting outliers, however, is not straight forward.

Buliung and Kanaroglou describe several problems concerning "infrequently occurring distant activities." While not dubbing such observations as "outliers" per se, the authors acknowledge such locations have a pronounced impact on measures such as the convex hull which often results in overly large (highly uncertain) spaces. Methods for identifying outliers vary considerably, but most activity space estimation methods acknowledge the need to minimize their impact in some way; either by choice of measure or by removing problematic observations.

Outlier detection methods can be loosely categorized into categories: (1) distribution-based, and (2) depth-based. Distribution-based methods are the most familiar and widely used within the social sciences (Leys et al., 2013). These approaches assume some distribution, often normal, and outliers are identified based on the probability distribution (Breunig, Kriegel, Ng, & Sander, 2000). Leys et al. (2013) found that a majority of researchers identified outliers using two or three standard deviations around the mean. This is an example of a distribution-based approach to outlier detection, specifically a normal distribution of the data are assumed and values greater than two or three standard deviations are unlikely to be caused by the same mechanism (to use Hawkin's terminology).

Depth-based approaches are used as a means to attempt to identify outliers in high dimensional space, the number of dimensions corresponding to the number of features (i.e. variables). The problem, however, is the "curse of dimensionality" which loosely describes the impracticality – that is high computational cost – associated with doing high dimensional calculations. Furthermore, depth-based methods are not intuitive and difficult to conceptualize.

Breunig et al. (2000) provides a compelling argument for the use of a method termed the Local Outlier Factor (LOF). This is a distance-based measure that assess the degree to which a point is an outlier based on a "local density". Rather than assigning a binary yes/no decision to whether a point is an outlier, the LOF instead provides a score indicative of its relative isolation from other points in its neighbourhood. This "local" approach to identifying possible outliers differs from the many traditional "global" outlier approaches that consider the dataset in its entirety.

The choice of this particular outlier detection method is two-fold: first by providing a relative measure of isolation the LOF allows for specific tuning of how strict outlier detection is, and second a local test of isolation makes intuitive sense given the underlying assumption that crimes cluster around given areas according to routine activities or some other preference of the offender. In such cases, it would not make sense to categorize a point in one cluster as an outlier due to dispersions observed in other clusters when the driving forces forming those two clusters differ.

It is important to note that the order of these two steps is not arbitrary. Consider the case when they are reversed: outliers are probed first and clusters are identified after the fact. The number of local neighbourhoods would not be known and subsequently remote observations that may be a part of a distant and small cluster could be inadvertently identified as outliers. With these points removed, the subsequent cluster analysis would identify fewer clusters and the resulting composite activity space would be the more inaccurate for it.

3.6 Method - Criminal Activity Space Estimation

The developed approach for calculating a criminal activity space employed here and throughout the thesis is summarised in Figure 22.

The specific goals of this approach are:

1. Identify any local clusters present in a given offender's crime series
2. Within each local cluster, identify any crime locations that are sufficiently distant to be considered an outlier
3. Affix a geometric surface to each observed cluster that reasonably describes the spatial extent an offender has committed crimes within that area

This method results in a composite activity space consisting of one or many clusters. The first step is the unsupervised clustering of crime locations into distinct areas of focus for each offender. The second step of the overall activity space process involves outlier detection per cluster per offender. Outliers in this context are defined as locations where offenders have been observed to commit a crime, but are not likely to be representative of future targeting behaviour due to the location's spatial isolation. Finally, a geometric surface is calculated for the resulting clusters; the combination of one or many surfaces symbolizes the criminal activity space.

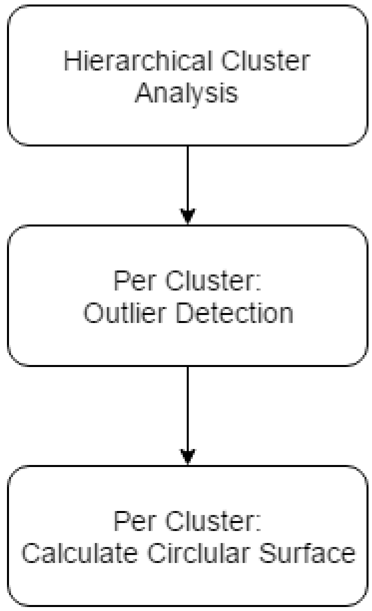


Figure 22: General Activity Space Estimation Method

3.6.1 Data

The data used to construct the criminal activity space construct consists of 3,449 crime locations committed by 997 unique offenders in a major U.K. metropolitan area. These crimes occurred from August 2011 to August 2015. These data were selected for analysis based on several criterion: (1) the crimes had to be solved so that the offender was known, (2) offenders were excluded if they only had two or fewer solved crimes on record, (3) crime location data had to be present, and (4) the crime was a solo (no co-offenders) offense.

These criterion ensured that only offenders who had committed multiple crimes of their own volition would be included in analysis, and there would be a reasonable amount of certainty that the crimes were legitimately committed by the offenders to whom they were attributed.

Importantly, crimes were not chosen based on a specific crime type contrary to common practice within the literature. This departure is consistent with several of the discussed crime occurrence theories, such as routine activity; as well as being necessary to ensure sufficient data were available for analysis.

The specific data of interest for the construction of the criminal activity space construct consists of the easting and northing geo-codes collected at the time of the offense. These geo-locations are based on the EPSG 27700 / OSGB 1936 British National Grid projection, and is the standard reported format within the U.K. for spatial data.

3.6.2 Methodology

3.6.2.1 Clustering

Clustering was performed using agglomerative hierarchical clustering. Specifically, the package factoextra (Kassambara & Mundt, 2016) on the R platform was used. The specific code used is provided in Appendix A – Code Snippets. The clustering process was unsupervised due to the number of unique offenders. Furthermore, because the number of optimal clusters, *k*, cannot be known in advance an iterative approach with a variable initialization value for *k* was used. Even though hierarchical clustering evaluates all possible cluster assignments, evaluating a set number of *k* clusters was still necessary in order to represent an offender's activity space as a singular geometry.

For each offender, let $X = \{x_1, x_2, \ldots, x_n\}$ denote the spatial set of their crime locations, where x_n denotes a single crime location. Let *k* denote the number of clusters tested for, where $k = \sqrt{n\ crimes}$ rounded to the nearest whole number for a given offender. For example, if the offender had committed 9 crimes, then $k = 3$. This formulation is consistent with similar initialization parameters within the machine learning literature (see: http://docs.h2o.ai/h2o/latest-stable/h2o-docs/data-science/drf.html, "mtries") and was chosen to strike a balance between the minimum number of points required to classify a group as a "cluster" and the maximum number of clusters that could be identified for a set of *n* crimes (Deming & Morgan, 1993). The hierarchical clustering procedure was then performed such that each x_n location begins in its own cluster (see: Figure 20, p.85 for an example). Observations are then merged together based on spatial proximity until *k* clusters is reached. By default the "hclust" clustering method within R implements Ward's clustering criterion (Ward, 1963).

For each cluster, each observation was assigned a silhouette width via the "fviz_silhouette" function (Kassambara & Mundt, 2016). This assigns each location observation a silhouette score (Rousseeuw, 1987). This score ranges from -1, indicating the observation is classified under the wrong cluster, and 1 which indicates the observation is well clustered. By averaging the silhouette width across all observations within each identified cluster, the average silhouette width can be calculated. Clusters with higher average silhouette widths are well clustered (Rousseeuw, 1987). Average silhouette widths of -1 to 0 indicate poor or incorrect cluster assignment. Thus, if the average silhouette width of any cluster is less than or equal to zero, *k* is decreased by one until either: (1) the average silhouette width for all clusters is greater than zero, or (2) *k* equals one which indicates only a single cluster. These represent the stopping criterion used in the clustering procedure. Next, each cluster for each offender was probed for outliers using the LOF.

3.6.2.2 Outliers

Outlier detection was performed by using the Rlof package which provides implementation of the LOF algorithm (Breunig et al., 2000). The LOF algorithm computes the local outlier factor as "the average of the ratio of the local reachability density of p and those of p's nearest neighbours" (Breunig et al., 2000, p. 96). The reachability distance is calculated as the maximum radius required to construct a circle which covers *k* points from a centre point *p*. So if $k = 4$, then the reachability distance of *p* would be the radius required such that a circle centred on *p* would encompass four points. Informally, as the reachability distance of *p* increases, and the reachability distance of its *n* nearest neighbours decreases, the LOF increases. This corresponds to the point *p* residing in less dense space than that of the closest points to *p* within the spatial set. The value *k* was implemented similarly as in the clustering step, namely $k = \sqrt{n\ crimes}$ within a cluster rounded to the nearest whole number.

Due to the way in which the LOF is calculated, multiple data points with the same location are problematic. Thus for each cluster for each offender, each unique crime location was randomly sampled once resulting in a single data point for a given spatial location. The LOF was then calculated for each x_n crime location for each cluster of each offender. Finally, if the ratio of unique crime locations present within a cluster divided by the total number of crimes was less than or equal 0.5, or if *k* equals one, then all locations within that cluster are assigned an LOF score of one. This corresponds to not having enough unique information to meaningfully identify points as outliers and so all points are retained.

3.6.2.3 Buffers

Once all crime locations for a given offender were assigned to a cluster and checked for outlier status, the final process of calculating the activity space surface was performed. Locations with an LOF score of greater than 1.5 were excluded from this phase of analysis. Because both the hierarchical clustering and LOF methods are best at identifying circular spatial patterns, the activity space surface was calculated as a circle. Furthermore, because the alternative methods convex hull and SDE require more than two points to estimate, the use of either would have resulted in significant loss of data in subsequent sections. The circular estimation method was done in two steps and is similar to calculating the reachability distance in LOF.

First, the geometric centre, or centroid, was calculated for each cluster for each offender. Let C represent the cluster centroid, then $C = \frac{x_1+x_2+\cdots+x_n}{n}$, where n is the number of crimes within the given cluster and x_n is a unique crime's location. Next, the distance from the centroid to each non-outlying point within each cluster was calculated. The radius of the circle was taken to be the maximum observed centroid to crime distance for a given cluster. An example of a final output is provided in Figure 23.

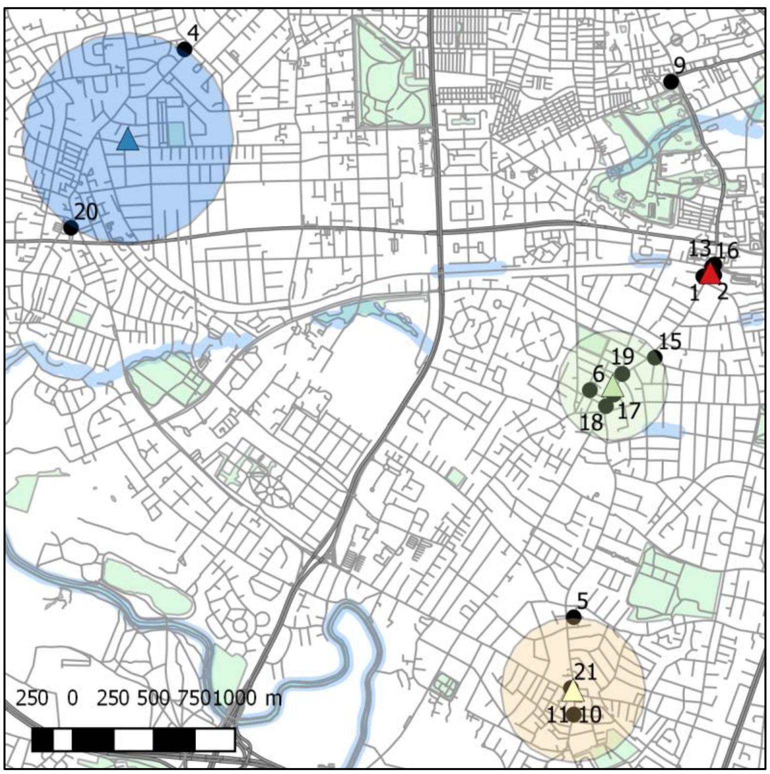


Figure 23: Example Output from Activity Space Calculation Process

Cursory inspection of Figure 23 shows that the process outlined above is behaving in a manner in-line with intuition insofar as the provided example is concerned. The process identified four unique cluster locations of varying sizes along with a single outlier – point nine – for this example.

3.6.2.4 Activity Space Development

The criminal activity space construct as currently described can be used to assess offender spatial activity over time. Specifically, do offenders generally "spread out" over the course of their crime series or remain constant. Given the proposed framework, exploring crime dispersion over time can be done in two ways: (1) by assessing how often the next crime in sequence occurs within an already identified cluster, and (2) assessing how "often" new clusters are identified. Furthermore, the current activity space estimation method makes no explicit assertions regarding an offender's home location. Past research indicates that the home most likely plays an important role in determining where offenders commit crimes as it serves as a fixed hub from which they must travel (Brantingham & Brantingham, 1995; D. Canter et al., 2013; D. Canter & Larkin, 1993; Rossmo, 2000). Geographic offender profiling has proposed a commuter / marauder dichotomy whether offenders travel outwards, "maraud", or travel to specific locations, "commute", to commit their crimes. This leads to three basic questions regarding criminal activity space development and how it relates to the home and future crime:

1. How often does a new crime in a sequence fall within an already identified cluster?
2. How often does the offender's home fall within the identified cluster(s)?
3. How do the number of identified clusters vary as a function of the number of observed crimes?

Kwan (1998) demonstrated the utility in using time-space restraints to model spatial data, and a similar approach is applicable here. While the specific time and date of an offence is not always available, the relative sequence of crimes is. By applying the activity space estimation method thus far outlined to temporally sequential subsets of crimes from crime series, it is possible to get a rudimentary view of how the activity space construct develops over the course of a crime series.

Let $X = \{x_1, x_2, \dots, x_n\}$ be a spatial set of crimes, a crime series, for an offender of n length such that x_3 occurs temporally after x_2 but before x_4. Then for a given crime x_n, that offender's crime series can be subdivide such that $X_n = \{x_1, x_2, \dots, x_{n-1}\}$. For example, the spatial subset for crime four of a series would be the first three crimes. In this way, crimes can be iterated through in sequence for each offender. For each spatial subset, clusters are created and outliers are checked according to the process described in previous sections. A point is located within the activity space if the distance from that point to the activity space centroid is less than or equal to the radius of the circle. Each cluster has a unique circle attributed to it, and so this check is performed for each cluster within the composite criminal activity space. Figure 24 provides an example of a point that resides within one cluster but not within another; this point would be considered "within" the composite activity space as it falls within at least one cluster's circular surface.

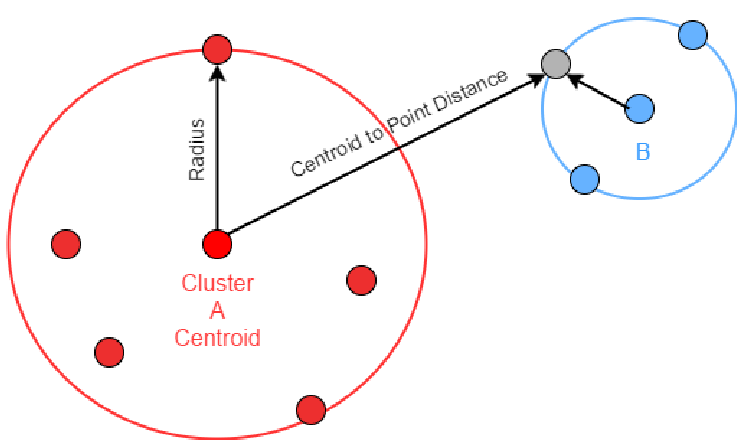


Figure 24: Point-Cluster Inclusion / Exclusion Example

Thus for each offender, a criminal activity space was calculated for each spatial subset. For each of these activity spaces, the area was calculated as well as whether or not the offender's home and or their next crime in sequence fell within the derived criminal activity space. Finally, for each spatial subset, the number of crimes used to estimate the criminal activity space was recorded, as well as the number of resulting clusters.

3.7 Results

3.7.1 Activity Space Descriptives

As previously mentioned, absolute size of an activity space has a direct impact on how that measure can be used. Larger spaces cover more absolute area and thus have more uncertainty. Figure 26 shows the size distribution of criminal activity spaces given the number of crimes present within the data for all 997 offenders. Observations that fall outside of the inter-quartile range are displayed as dots. It can be clearly seen that activity spaces calculated from three points have the widest distribution in sizes. After three crimes, there is significant drop-off in total size that appears to increase until around 8 crimes are used in estimation. Sample sizes at the various crime series lengths is an issue: there are 491 activity spaces calculated from two crimes (49.2%) to 10 activity spaces (~1%) calculated from 10 crimes. Figure 25 shows the distribution of offenders according to the length of their crimes series.

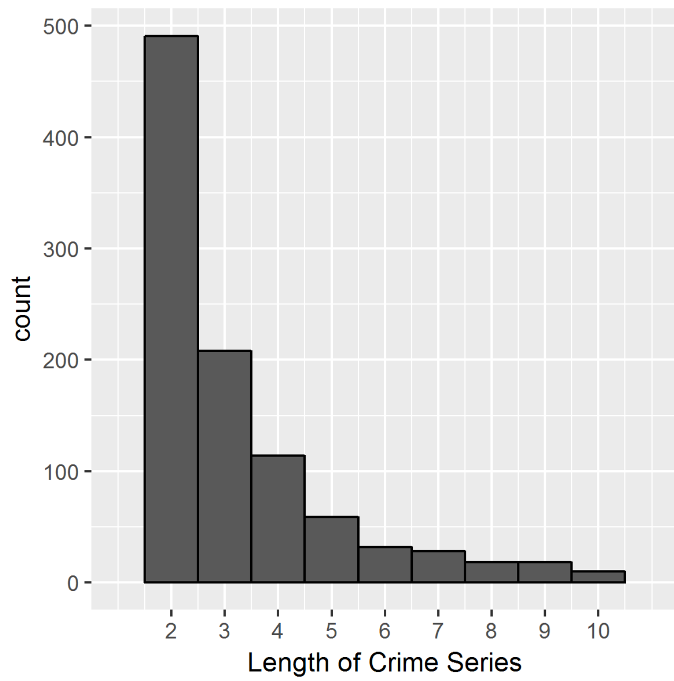


Figure 25: Crime Series Length Distribution

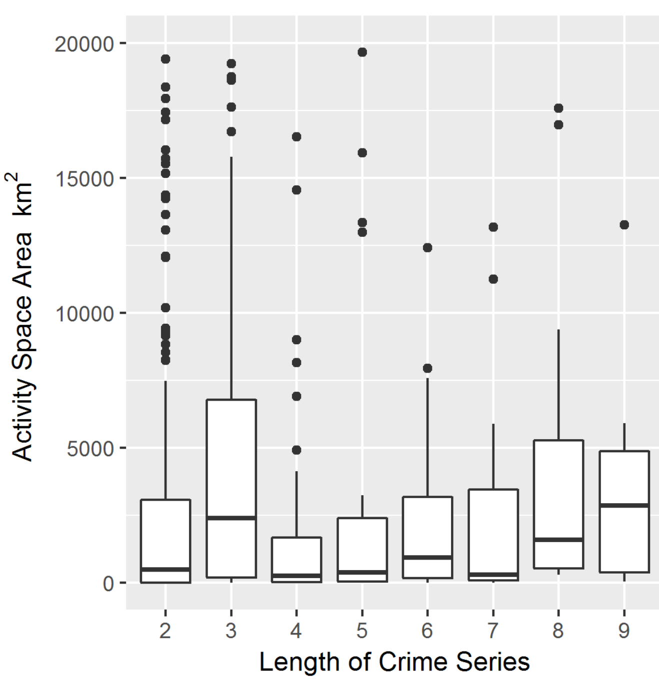


Figure 26: Activity Space Size Distribution by Crime Series Length

Figure 26 shows the distribution of criminal activity spaces for each offender's entire crime series only, and Figure 27 shows the size distribution when one only looks at the number of crimes used to estimate criminal activity space. Using three crimes for estimation still results in the largest spread of sizes, but starting with four crimes there is an apparent growth pattern not present when one only looks at the final space sizes. As in Figure 26, Figure 27 shows a similarly sharp drop-off in criminal activity space size when four crimes are used in estimation. Furthermore, there is a sharp *increase* in sizes associated with criminal activity spaces when calculated from three crimes as opposed to two crimes. When taken together, these two trends imply that the third crime is generally spatially isolated from the first two. Adding the fourth crime allows the algorithm to either create two smaller clusters and reduce the absolute size, or classify one of the four crime locations a spatial outlier which would also create a smaller surface using the three remaining crimes. As more crimes are added, the absolute size of the criminal activity space seems to increase asymptotically to a value that is overall less than the sizes observed when three crimes were used.

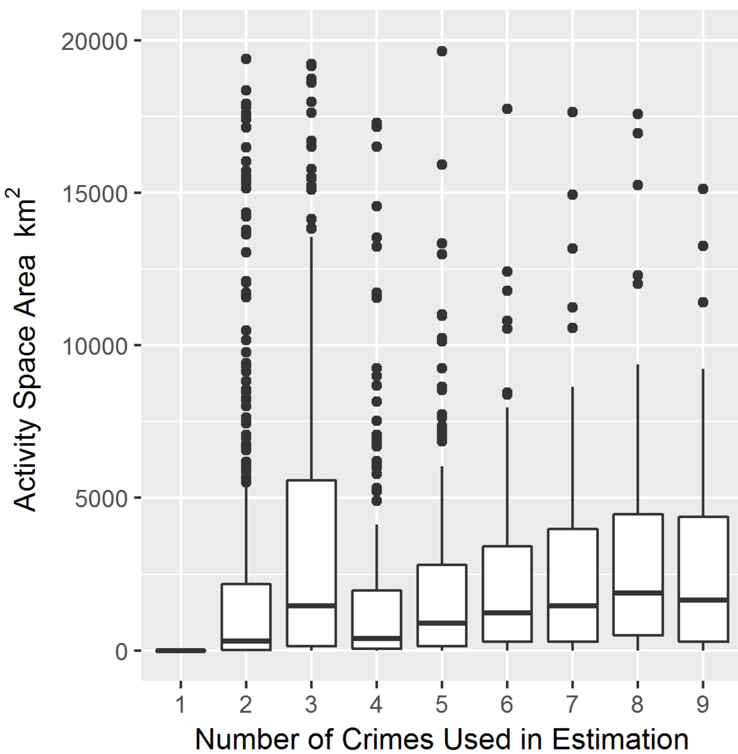


Figure 27: Criminal Activity Space Size Distribution by the Number of Crimes Used in Estimation

Next, Figure 28 shows the distribution of the number of clusters generated from *n* crimes used in estimating the criminal activity space construct. Due to the manner in which *k* is initialized in the cluster procedure, for a given *n* crimes from which to estimate an activity space, there can only be a maximum of $\sqrt{n}$ clusters. Thus for three or less crimes, the algorithm can only designate one cluster; for four to six crimes only a maximum of two clusters can be identified and so on. Results indicate, however, that for seven or fewer crimes it is most likely that only one cluster will be identified. The median number of clusters only increases to two clusters for eight crimes or more.

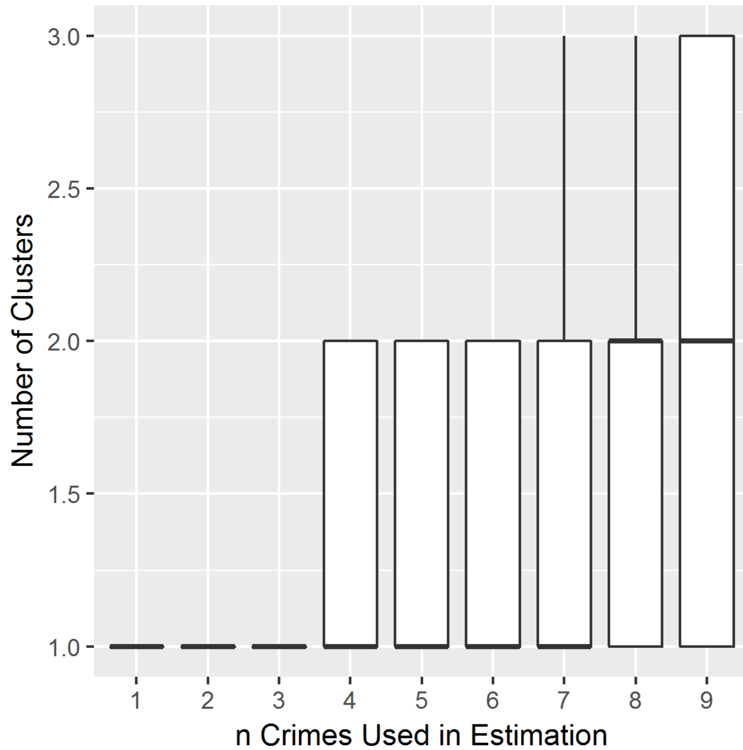


Figure 28: Number of Clusters by the Number of Crimes Used in Estimation

3.7.1.1 Activity Space Size Distributions

The results thus far have explored how activity spaces vary according to changes in the input variables, such as number of crimes used to estimate the space vs the total number of crimes present in an offender's criminal history. This sub-section focuses on exploring whether there are any significant differences in activity space size according to other facets of the data including: offender sex and crime sub-division. In order to do so, however, the underlying data – the activity space size (in km^2) needs to be checked for a normal distribution.

Figure 29 depicts four quantile-quantile plots stratified by several values of λ. This approach is known as power transformations as they all revolve around raising the set of observed data to some exponential power, λ. The approach was popularized by Box and Cox

(Box & Cox, 1964) and the set of standardized power transformations are often referred to as 'Box and Cox Transformations'. The quantile-quantile plots below place the quantiles of a normal distribution on the x-axis, and the corresponding quantiles from the sample on the y-axis. In a normal quantile-quantile plot if both samples come from the same distribution, then the resulting scatter plot should correspond to a y = x normal 45 degree line; this corresponds to the black reference line in the plots below. As can be seen, the un-transformed data are highly non-normal with warping at the tails. Of the three depicted λ transformation values – (0, $\frac{3}{20}$, and ¼), $\frac{3}{20}$ or 0.15 yields the most 'normal' quantile-quantile plot. However there is still slight warping at the tails which indicates that it is unlikely that the underlying distribution can be truly described as normal. However for the sake of comparison, the λ = 0.15 transformation will be used in subsequent charts.

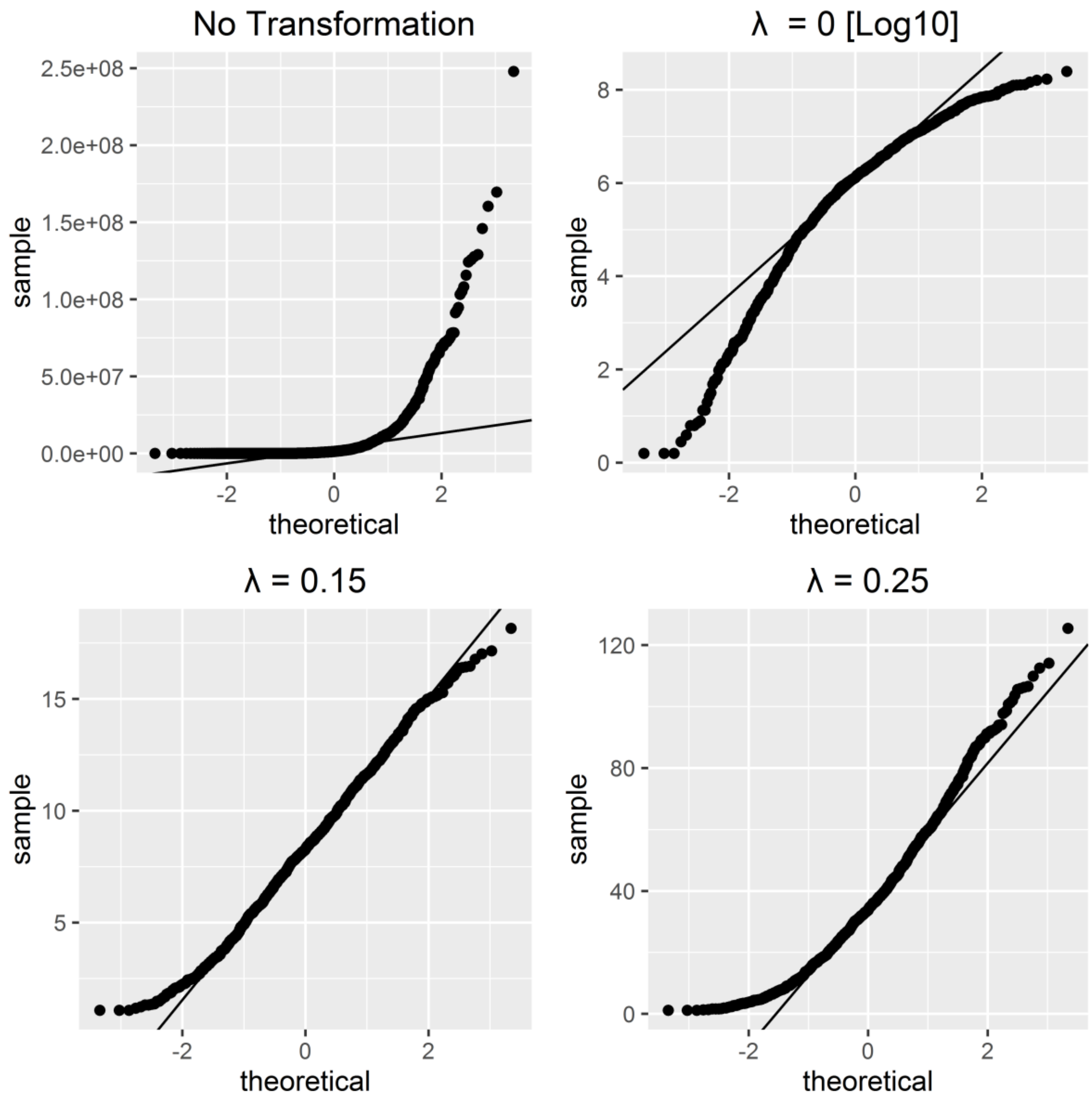


Figure 29: Activity Space Area Quantile-Quantile Plots - Stratified by Lambda Power Transformations

Figure 30 shows the density distribution of the transformed activity space size variable stratified by offender sex. The accompanying histogram provides an indication into the total number of observed cases for any given size bucket. There does not appear to be any significant difference between sexes in terms of their observed activity space size distributions.

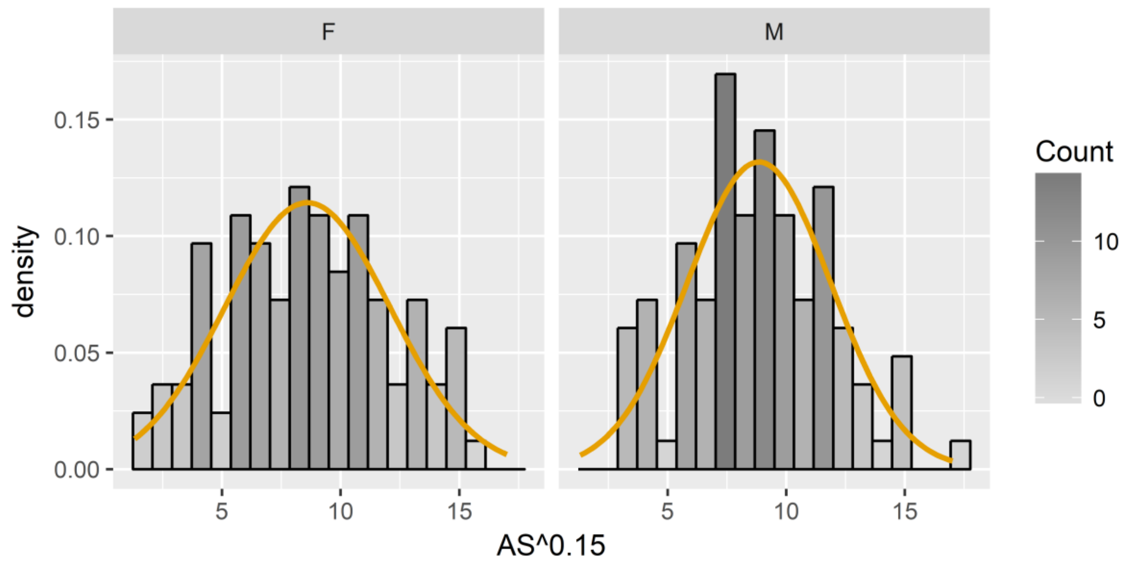


Figure 30: Transformed Activity Space Area Distribution Stratified by offender Sex

Finally, there is the as of yet unresolved issue regarding differences in the underlying geography which could influence how an offender's activity space is shaped. Unfortunately the only convenient variable for grouping the observations spatially within the data are the police derived crime sub-division. As there was no documentation or official explanation for how the borders of these zones were derived the following analysis is severely limited when it comes to discussing cause. However it is still possible to investigate whether there are differences in general.

Figure 31 shows the density distributions for the transformed activity space size variable stratified by crime sub-division. Again, the underlying histogram is also provided for reference. While there does appear to be differences present, it is difficult to quantify what those differences may be. Thus Figure 32 depicts the density distributions for the four crime sub-divisions overlaid onto a single plot. What is immediately obvious is that the distributions to appear to be subtly different. In particular, activity spaces from E4 appear to be somewhat smaller than the other three zones. E3, on the other hand, appears to have the largest overall activity spaces. Finally E1 and E2 are somewhere in-between with very similar distributions.

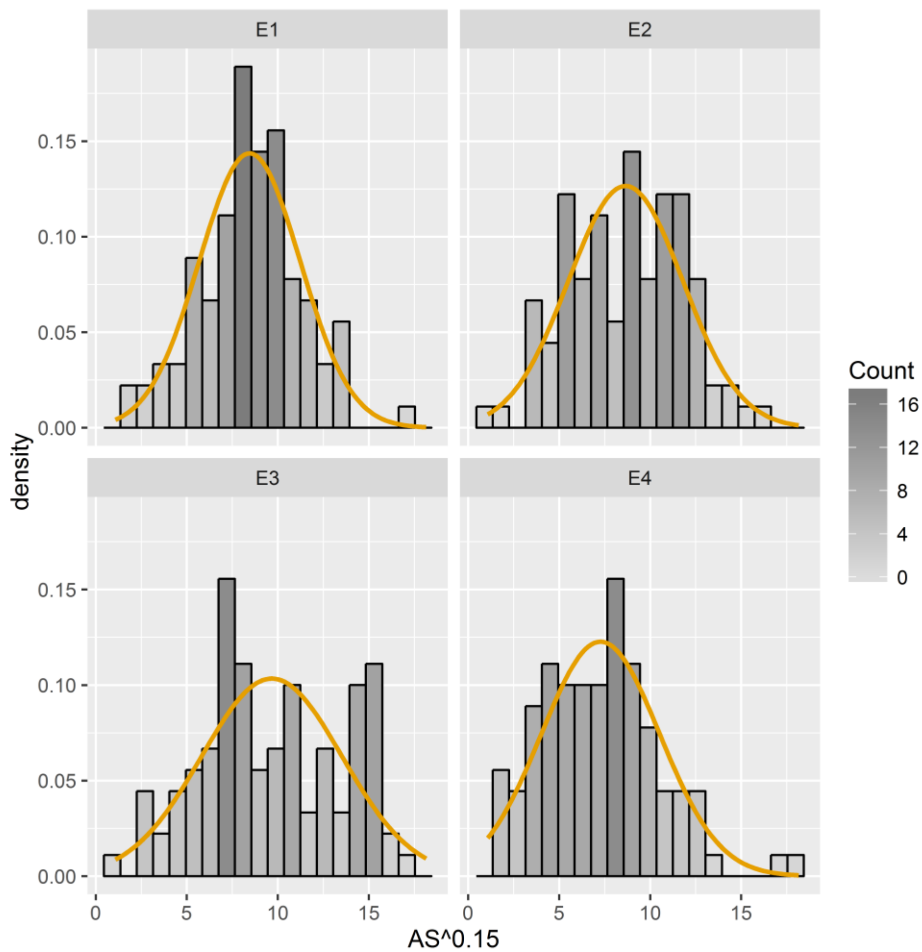


Figure 31: Activity Space Size Distribution Stratified by Crime Sub Division Location

In order to formally test the differences observed in Figure 31, an ANOVA was performed comparing mean values of activity space sizes for the four sub-divisions. The ANOVA indicated there was a statistically significant difference ($F(3) = 11.82$, $p < 0.001$). In order to determine where these differences lie, Tukey's honest significant differences (HSD) post-hoc test was performed. Results are shown graphically in Figure 33 with the group being compared shown on the y-axis and mean difference on the x-axis. As can be seen, the E4 sub-division is significantly different from E3 and E2 and nearly different from E1 as well. In all cases, activity spaces from E4 tend to be smaller than the other three crime sub-divisions.

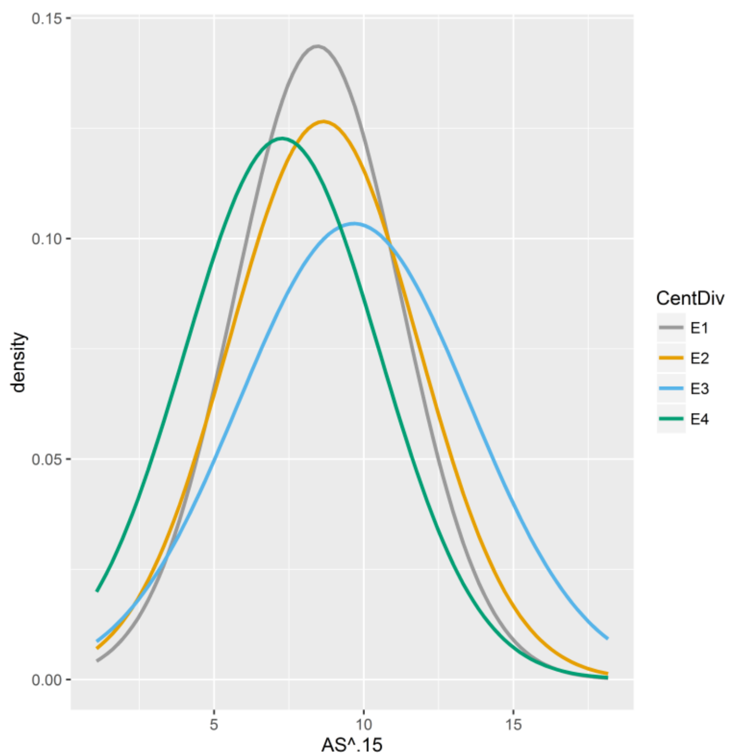


Figure 32: Activity Space Transformed Area Density Plots by Centroid Sub-Division Location

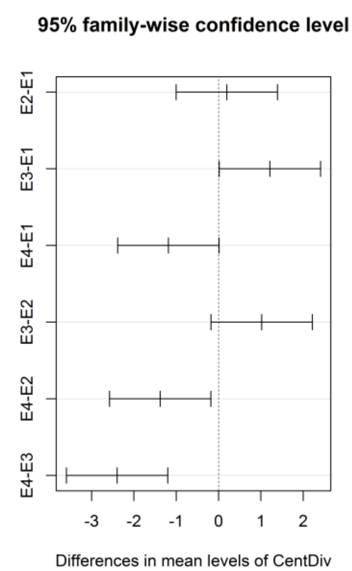


Figure 33:Tukey Honest Significant Difference (HSD) Between Sub-Division Activity Space Size Mean Levels

3.7.2 Activity Space Hit Rate

Table 5 shows the hit percentage as a function of the number of crimes used in calculating a criminal activity space. A “hit” is defined as occurring if the next crime in sequence falls within the criminal activity space derived from the preceding crimes. For example, if three crimes are used, then the hit percentage corresponds to the ratio of the number of offender’s who’s fourth crime falls within the activity space over the total number of offenders with at least four crimes recorded. The average total size, in square meters, is also provided for reference as high hit percentages could be achieved with arbitrarily large activity space sizes. This is exemplified in row three of the table which simultaneously has the highest hit percentage but also the largest average activity space size. Overall, there is some evidence to suggest that past crimes are spatially predictive of future crimes. The hit percentage ranges from 17.9% to 60.7% with an average of 47.5%. The average increases to 51.2% when the hit rate from one crime only is removed.

Table 5: Frequency Activity Space Predicts n Crime in Series from n-1 Previous Crimes

N Crimes in Series	# of Offenders	Hits	Hit %	Average Activity Space Area km^2
2	997	178	17.9	0
3	506	207	40.9	4692.7
4	298	181	60.7	11923.8
5	184	83	45.1	6398.338
6	125	58	46.4	5409.821
7	93	49	52.7	5955.915
8	65	35	53.8	3266.479
9	47	24	51.1	3595.258
10	29	17	58.6	3365.669

Finally, Table 6 shows the home hit percentage as a function of the number of crimes used in calculating the criminal activity space. A "home hit" is defined here as being when the home location associated with some crime is located within the activity space constructed from said crime and all crimes committed before it for a given offender. For example, if the home location associated with the 4th crime in a series is located within the activity space calculated from that crime and the three crimes preceding it, then it is considered a "hit". Similar to the results of predicting areas of future crimes, the most success was seen when three crimes were used. Given the present activity space estimation method, it is rare for an offender's home to be located in an area immediately surrounded by their crime locations. Home hit rates range from 20.0% to 42.6% with an average of 35.4%. The differences in the number of reported offenders between Table 5 and Table 6 are due to missing home location records for a number of offenders.

Table 6: Frequency Activity Space Covers an Offender's Home

# of Crimes Used	# of Offenders	Hits	Hit %	Average Activity Space Area km^2
2	776	13	1.7	0
3	381	79	20.7	5007.483
4	223	95	42.6	9739.327
5	126	42	33.3	5509.85
6	91	31	34.1	4902.253
7	56	22	39.3	4852.467
8	36	15	41.7	2868.912
9	26	10	38.5	3297.529
10	18	6	33.3	3504.587
11	10	2	20.0	2958.658

3.8 Discussion

This chapter set out to answer two fundamental questions presented in research question set 1: can diverse individual crime patterns be described via a generalized ‘activity space’, and 2) are such activity spaces predictive of future criminal spatial activity? To address these questions, a generalised approach to estimating criminal activity space was presented, as well as several established alternatives including convex hulls, elliptical methods and network based measures. Upon initial inspection, the established estimation methods were not satisfactory, and so an approach that combined location clustering, outlier detection, and circular geometric surface fitting – was presented. This method was initially evaluated in terms of the general size distribution of the composite activity space construct that results from it, as well as how predictive such a composite activity space is of future crimes in general within the population.

The observed distribution in activity space sizes highlighted an interesting pattern: the absolute size of the composite activity space distribution was approximately equal for two crimes and four crimes; yet for three crimes the activity space sizes ballooned considerably. Due to the specific implementation of the clustering algorithm, there could only be one cluster with no outliers for three crimes. The large increase in the size distribution implies that the third crime in a series is more spatially distant from the first two crimes than those crimes are from each other. This can be confirmed by the smaller distribution associated with two crimes than with three crimes. Interestingly, the size distribution for four crimes is reduced considerably; and as previously mentioned there exists two distinct possibilities for explaining this phenomena. First, the fourth crime location occurs near, i.e. more compactly, with the first two crimes than the third and subsequently the third crime location is classified as an outlier and removed. Alternatively, the fourth crime can occur in a location that is more spatially compact with the third "distant" location and the algorithm subsequently identifies two distinct regions – i.e. clusters – of activity whose composite area is smaller than the single area attributed to the first three crimes.

Behaviourally, this pattern is interesting because it suggests offenders target or are otherwise active in a relatively confined space, after which they target a relatively new space. This change in morphology – how the crime locations are dispersed - is of particular interest. From a routine activity perspective, changes in morphology should be rare. By definition a change in routine activities should not happen frequently as they would no longer be "routine".

However, the present data implies that crimes occur fairly regularly at locations not spatially consistent with earlier crimes.

There are several possible explanations for this phenomena, but the most likely is that the activity space estimate is overly conservative when calculated from two locations. In other words, the first two crimes committed are not sufficiently representative of an offender's awareness space to provide a reliable estimate. Evidence for this interpretation can be seen in Figure 27 whereby after the size distribution is reduced for four crimes, it slowly creeps back up to larger sizes for eight or more crimes. From a routine activity space perspective, this pattern would imply that not all routine activity locations are represented from the first two crimes' locations which seems intuitively obvious. The same logic applies to the awareness space interpretation: namely that the first two crimes' locations only draw attention to a small subarea of a given offender's awareness space. In both cases, as more crimes are added to the estimation process, one would expect the estimate to become more accurate. This is exemplified by the stabilizing of the total size distribution of the space.

Unfortunately, this interpretation is not fully supported by the observed predictive utility of the derived criminal activity spaces from Table 5. Overall, future crimes were observed to lay within the activity space calculated from the preceding crimes on average ~51% of the time. This percentage does not appear to appreciably change according to the number of crimes used in the estimation process. From a routine activity perspective, this is troubling because it implies that crimes are not occurring simply around routine activity nodes. If they were, one would expect the predictive utility of activity spaces to increase systematically with the number of crimes used in estimation which does not appear to be the case.

Unlike routine activity theory, however, an awareness space interpretation places far more agency on the offender. As described in Chapter 1, routine activity theory predominately paints the offender as reacting to the environment and the opportunities presented. Awareness space interpretations, on the other hand, incorporate the active learning of new opportunities which would then be manifested as novel locations. The predictive utility results reported here may suggest an awareness space interpretation more so than a purely routine activity theory based interpretation due to the consistently novel locations travelled to over the course of the various sampled crime series.

There are a number of limitations that affect the ability to draw any definitive conclusions from the methods and data presented here. First, the results for all offenders over a rather large geographic area have been aggregated together. If one assumes that opportunities are equally distributed in space for all offenders this is not a problem, however such an assumption is obviously unrealistic. Next, activity space development as presented in this chapter has been simplified in that it ignores actual time delays between crimes. For example, a crime series consisting of three crimes occurring over the course of seven days was treated the same as a crime series of three crimes occurring over three months. Furthermore, the types of crimes have been completely ignored. A crime series consisting entirely of burglaries was treated the same as a crime series consisting of many different crime types. It is not clear what impact, if any, these generalizations would have on the observed properties – size and number of clusters – of criminal activity spaces. Issues such as what influence – if any – a criminal's unique offending history has on their observed activity space, as well as what impact the environment itself has on activity spaces is the topic of exploration in the next chapter.

Chapter 4

Activity Space and Crime Linkage Analysis

4.1 Introduction

Thus far this thesis has focused on defining what a criminal activity space construct represents behaviourally, and what the apparent properties of such a construct are for a given sample offending population. However the question remains as to whether activity spaces as presented here have any empirical merit and can be applied in practical contexts.

Criminal activity spaces operate under the assumption that past crime locations are indicative of future crime locations. This relationship has been presented as the result of routine activities and rational choices emerging from an offender's unique understanding of their environment and its criminal opportunities – i.e. awareness space. Given that it is unlikely that any two individuals would build an identical awareness space, a given offender's resulting criminal activity space should be unique. After all, if the activity space construct is in fact descriptive of spatial behaviour, then one would expect it to be predictive of future spatial behaviour; a finding that would have direct practical applications. In order to test such an assertion, we now turn to incorporating activity space based measures into a crime linkage framework outlined in research questions 2.1 and 2.2

4.2 Crime Linkage Analysis and Behavioural Crime Linkage

Crime linking analysis, or behavioural crime linkage (BCL), is an area of research that has grown extensively in the last ten years (Bennell & Canter, 2002; Tonkin, Woodhams, Bull, Bond, & Palmer, 2011; Woodhams & Bennell, 2015, and others), and in its most basic interpretation can be understood as a process (or processes) by which a set of two or more crimes is determined to have been committed by the same (linked) offender or not (un-linked). Such tasks play a key role in many aspects of investigative decision making: linking tasks can help identify emerging crime series or help to provide guidance for potential suspects for already established crime links (such as in instances when DNA evidence is available) (Woodhams & Bennell, 2015).

Crime linkage research provides a rich environment to assess a variety of offender behaviour(s) with the possibility of direct decision support feedback for participating police agencies. For example, one of the more direct arguments for the importance of the linking task comes as a result of the common finding that a disproportionate amount of crime is committed by a small percentage of the overall criminal population (Wolfgang, 1973). Cohort studies within the criminal career literature and elsewhere have consistently outlined how a small proportion of known repeat ("chronic") offenders is responsible for a disproportionately large percentage of recorded crime in the US (Wolfgang, 1973; Wolfgang, Filio, & Tracey, 1983), UK (Kershaw, Nicholas, & Walker, 2008) (Farrington et al., 2006), and Sweden (Falk et al., 2013). These studies reported between 1-8% of the studied offending population were responsible for between 50-63% of recorded crimes within their respective samples.

Such findings raise several important questions about how such chronic offenders differ from their non-chronic counterparts. Are there important developmental factors? Personality factors? Environmental or socio-economic factors? Of particular interest to practitioners, that researchers are now focused, is the question of if chronic offenders differ in *how* they offend from non-chronic offenders and whether such differences can be detected. Answering such a question lies at the heart of crime linkage research.

4.3 Offender Consistency and Distinctiveness

The primary focus of much of the crime linkage literature deals with identifying and assessing what has been called the consistency and distinctiveness hypothesis (Woodhams & Bennell, 2015). This assumption arises from the *a priori* argument that *if* crime linkage is possible, *then* offenders must differ from one another in such a way as to be identifiable (distinctive) and non-random (consistent). In other words, for crime linkage to be possible, offenders must both be consistent and distinct in their actions (Woodhams and Bennell, 2014). Logically, if an offender is inconsistent, then it would be very difficult to identify a pattern, if any, which could be used to describe that offender. Conversely, if an offender is consistent but not distinctive then there would be no way to differentiate that offender from a pool of other similar offenders.

Consider a scenario, for example, where a group of offenders commit a series of burglaries. These hypothetical offenders' M.O. consist of breaking into houses via the front window and stealing laptops. These offenders would be exceptionally *consistent* over time – same type of crime, type of location, method of entry and items stolen – but also wholly indistinct from each other. It should be noted, however, that even in this scenario there are other components that would allow us to discriminate between these offenders: for unless they all target the same house on the same day, then the specific time and area would be revealing. It is identifying and testing these descriptive features that is the primary focus of crime linkage research. Thus assumptions of offender distinctiveness and consistency necessitates a methodology that allows for the comparison of a set of features whose predictive utility is such that future observations of those features can be predicted utilizing past observations, and that such features, or combination of features, is in some way discriminatory.

The issue of consistency is one that has been discussed for some time, not just in crime linking literature, but personality literature as well (Canter, 2000; Mischel & Shoda, 2000). Personality psychology seeks to describe an individual's personality by means of observing and quantifying their behaviour. An example may include observing how often a participant interacts with a confederate while "waiting" for the experiment to start as an indication for extroversion. A person's personality, it is argued, influences how they will behave. The so-called "personality paradox", however, outlines a key challenge to behavioural consistency: given similar or identical circumstances, the same individual may not display similar or identical behaviours.

Finally, with regard to distinctiveness, there is a question of underlying base rates for behaviours within a crime. For any given crime type, one would expect a specific grouping of behaviours to be present; for example: items of value to be stolen during a burglary. Social cognitive theory posits that humans learn behaviours by observing others and not simply by constantly trying and failing by themselves (Bandura, 1986). Many criminology theories of learned behaviour echo similar views and state that a component of criminality is the learned deviant behaviour from peers or external sources (Akers, 1985; Sutherland, Cressey, & Luckenbill, 1992). Behaviourally speaking, this implies that there are two sets of actions observable within a criminal offence: those that are typical of the offence type, and those that are individualistic or otherwise atypical.

While there is an on-going debate within criminology with regards to what comes first: delinquency or delinquent peers; this question is mostly ignored in crime linkage which focuses exclusively on the proposed learned behaviours. From this perspective, a proportion of the observable behaviour engaged by an offender during a crime will be those almost scripted, learned behaviours – behaviours that are effectively crime-type specific and not offender specific. Thus any research into crime linkage should consider very carefully the distinctiveness of behaviours and just how unique they really are or are not.

4.4 Crime Linkage Methodologies

As previously pointed out, much of the literature surrounding crime linkage research has revolved around testing the distinctiveness and consistency of behavioural domains in order to determine the extent to which they can be used to identify linked crimes from unlinked crimes. The most widely adopted method for doing so was first proposed by Bennell and Canter (2002) in which they argued the similarity between crime linkage assessment and other diagnostic tasks from fields such as medical research. Specifically, they employed statistical models and processes to assess the degree to which modus operandi, that is observable behaviour investigators actively track, could be employed to differentiate between linked and unlinked crime. Due to the ubiquity with which Bennell’s approach has been employed since its initial appearance, a detailed review of its steps as well as a summary review of the previous research in which the method was adopted will follow.

4.4.1 Origins

One of the first primary contributions to what could be considered a ‘standard’ approach to crime linkage analysis was put forth by Bennell and Canter (2002). Their approach was simple revolved around three primary phases: 1) Identifying effective linking features, 2) Constructing predictive models from those linking features and 3) Assessing the performance of the derived models (Bennell & Canter, 2002; Woodhams & Bennell, 2015). The specific steps undertaken during each phase varies slightly from study to study as data allows, but conceptually these three phases are consistent within the literature. A discussion of these three broad phases follows.

4.4.1.1 Identifying Effective Linking Features

The first phase of the generalized approach encompasses the multitude of collation and data preparation steps necessary for assessing crime linkage. The goal of this phase is to identify behavioural features and domains or other useful constructs that would show differentiation between serial and non-serial offending. As available data varies greatly not only from study to study, but also police force to police force, an initial assessment of available information is necessary.

Previous studies of crime linkage analysis have employed content analysis to identify salient behaviours for the given sample (Tonkin et al., 2011; Woodhams & Bennell, 2015). However as content analysis is extremely time intensive, this method may not always be feasible. In such cases, researchers would have to rely on previous research and available data fields to guide their inclusion criterion. As of yet, there is no consensus as to what constitutes necessary predictive information surrounding crime linkage.

Studies have, however, outlined a number of domains that appear to routinely provide useful results including: crime location and crime date. Whether these two domains provide useful results due to the consistency with which they are collected by police agencies, or they provide useful insight into offender behaviour remains a contested topic (Woodhams & Bennell, 2015). Typically, these two domains (location and time) are combined with a behavioural domain that assesses similarity in distinct behaviours undertaken by the offender during the offense.

Ultimately, the choice of effective features is guided by underlying theories of offending behaviour and their consistency and distinctiveness. Crime location, for example, is posited to be an effective feature because it describes offender targeting behaviour. Several theories of offender spatial behaviour provide context for why offenders may consistently target a given area over another. Such theories include routine activity theory and offender awareness as discussed in previous chapters.

The entire discussion surrounding salient features for crime linkage becomes exceptionally more complicated in instances where one attempts to assess case linkage across crime types. Recall that there is evidence that a small proportion of offenders have been found to be responsible for a large percentage of crimes committed by their cohort (Wolfgang, 1973); these studies also outlined that such offenders did not necessarily specialize in the crimes they committed. For such offenders, there is an argument to be made for attempting to conduct case linkage between a burglary and an assault, as an example. Doing so, however, is exceptionally difficult from an M.O. perspective. How does one evaluate the similarity in behaviours between two unrelated crime types such as assault and burglary? This is a key area of concern within the literature, with only a few studies having been published that actively attempt to mediate this problem (Tonkin et al., 2011; Tonkin & Woodhams, 2015).

Tonkin et al.'s 2011 study ignored specific behaviours altogether and instead focused only on temporal and spatial similarity. Tonkin and Woodhams 2015 study side-stepped the issue by focusing on the related crime types of burglary and robbery; they were able to devise several behaviour domains that could be applied across crime types. Generally speaking, even in these "extreme" cases, the general premise remains the same: utilise available information to determine factors which could differentiate between linked and unlinked crimes. These features can then be passed to the next phase and utilized in creating predictive models of case linkage.

4.4.1.2 Linking Assessment Models

The earliest presentation of Bennell's work described the linking task in terms of classification (Bennell & Canter, 2002). For a given set of crimes, a corresponding set of crime pairs could be constructed which could then be classified as either linked or unlinked. This binary classification scheme lends itself most readily to traditional logistic regression models, which remains the most widely employed model to date (Bennell, 2005; Bennell & Canter, 2002; M. J. Tonkin, 2012; M. Tonkin et al., 2011; Matthew Tonkin & Woodhams, 2015; Woodhams & Bennell, 2015). However, as reported by Bennell et al. (2014), a handful of studies have employed other model techniques including classification trees (M. J. Tonkin, 2012), discriminant function analysis and naïve Bayesian classifiers (Winter et al., 2013). It should be noted that the sample sizes of various studies, discussed more below, has varied within the literature from the relative small n=86 up to n=386 crimes. Part of logistic regression's extensive use lies in its stability in the face of misuse or overly muddy data (Harrell, 2001). A further advantage to logistic regression is its ability to provide insight into the relative impact of any individual predictive domain. Regardless of the model used, the ultimate goal of crime linkage research is to assess how successfully linked cases can be identified from unlinked cases utilizing available data. Making the final assessment as to whether a given model is "good" or not is a nuanced one and falls under the purview of phase three.

4.4.1.3 Assessing Model Performance

Assessing model performance is a complex undertaking that is highly dependent on the nature of the question a given model is designed to answer. Bennell's early work outlined the linking task as one of strict classification of a binary outcome: whether the crimes are linked or not. A two-step approach was adopted that first assessed how well the model could differentiate between linked and unlinked crimes. The most common method for doing so within the literature has been ROC analysis. Receiver operating characteristic (ROC) analysis can be notionally understood as providing a performance metric in the AUC, or area under the curve, that assesses how discriminate a continuous score is at any given discrete decision threshold. Bennell and others have argued for the use of the AUC as a means to provide a comparable statistic across studies utilizing different samples and different behavioural domains (Bennell & Canter, 2002; Bennell & Jones, 2005).

4.4.2 Past Research

Over the past decade there has been numerous studies conducted that look at the feasibility of behavioural case linkage. To date, there have been several published works that summarize the current levels of predictive performance, one of the earliest of which being done by Woodhams et al. (2007). Further summary work can be found by Bennell et al. (2014) and most recently by Bennell and Woodhams (2015). These collections outline several recurring themes within the literature: 1) Linkage analysis appears to be possible across a range of crime types utilizing a variety of behavioural predictors, 2) Of said predictors (as previously mentioned) crime location and date often appear as the most useful factors, 3) Consensus regarding "best practice" is still in debate. This section will serve as an introduction to a sample of the body work outlined by these summaries with a focus on methods employed and the general findings the studies seem to suggest.

Bennell and Canter's 2002 study was the earliest publication instance of Bennell's "standard" method as outlined in section 4.4.1. This study was conducted utilizing a relatively small sample of 84 domestic burglaries and included inter-crime distance, temporal proximity and a Jaccard behavioural coefficient as model predictors. Logistic regression was employed as the statistical model of choice and found that while all the predictors contributed to model performance, inter-crime distance was the most useful. This was one of the first studies to employ ROC analysis in crime linkage and the AUC was reported. The results from the regression analysis as well as the subsequent ROC analysis suggests that the predictors used held at least some utility in determining the linkage status of crime sets. Given the small sample set, however, further replication was suggested.

Standing in sharp contrast to the very systematic approach of Bennell and Canter (2002), Hazelwood and Warren (2003) adopted a very thematic approach to linking. Their study focused on linkage analysis with regard to sexual offenses. Their approach consisted of five steps: "1) Gathering detailed, varied and multisource documentation; 2) reviewing the documentation and identifying significant features of each crime individually across the series; 3) classifying the significant features of the crime as either MO and/or ritualistic constructs; 4) comparing the combination of MO and ritualistic features across the crimes to determine if a signature exists; 5) compiling a written analysis that details the conclusions derived from the available information." The study provides a case example of the proposed methodology, but no formal statistical analysis were conducted or provided. While it would be difficult to adopt a general approach based upon this study, the similarities between Hazelwood and Warren (2003) and Bennell and Canter's (2002) study in identifying MO and other behavioural similarities between crime sets implies a level of validity to the approach.

Bennell and Jones (2005) in many ways is a replication and extension of Bennell and Canter's (2002) study. Again the study focused on burglary, though the sample was much larger with 634 commercial burglaries and 660 domestic burglaries. The method was much the same; crime-pairs were created from the individual crime sample and various behavioural similarity measures were created including: entry behaviours, target characteristics, items stolen and inter-crime distance. Again each individual behavioural characteristic was found to be significant, but inter-crime distance was the single best performing predictor. Overall linking performance was assessed by ROC analysis and again was found to suggest that offenders appeared to be sufficiently consistent and distinctive to allow for successful linkage assessment.

Departing briefly from strict crime linkage, Snook et al.'s (2006) study took a different approach to offender inference. Snook and colleagues attempted to empirically test the ability of journey-to-crime distance as a means for generating potential at-large suspects for unsolved crimes. Their methodology comes about from the observation that offenders typically choose to commit offenses close to home (D. Canter & Larkin, 1993; Canter & Youngs, 2009; Rossmo, 1999, as some examples). Thus their method was straight forward: 1) generate a list of possible suspects by assembling a pool of possible suspects who had committed the same type of crime in past; 2) Measure crow-flight distance from each possible offender's home to the crime location; and 3) Rank offenders according to those who lived closest to the unsolved crime. Their sample consisted of 36 commercial robberies committed in St Jonn's, Newfoundland between 1995 and 1999. Their results indicated a moderate rate of success with a third of the cases having the actual perpetrator in the top 35 suspects, and two thirds of the cases had the perpetrator in the top 255 suspects. This study highlights the apparent utility in critically evaluating the offense and offender location relationship as well as providing a unique look at what a potential decision support system may be when based upon empirically derived analysis.

Baumgartner, Ferrari and Palermo (2008) approached the linking problem from a different perspective. Rather than assessing the similarity between two or more crimes, the authors instead employed statistical analysis to infer offender characteristics of unsolved crimes utilizing a database of known solved crimes. Specifically, the authors advocated the use of a machine learning technique known as a Bayesian networks to assess the relationship between "hidden" – that is unobservable – offender characteristics and observable offense characteristics. Such "hidden" offender characteristics included employment status, gender and criminal history. The results of their study indicated that such machine learning techniques could be very useful as their Bayesian network was reported to predict the hidden variables accurately between 80 and 95% of the time. These results suggest that offenders can be identified utilizing MO behaviours; albeit in a different manner from the "traditional" crime linkage approach.

Bennell, Jones and Melnyk (2009) continue to publish supporting work in favour of the use of the AUC as the standardized performance metric within crime linkage research. This is in response to the observed disagreement within the literature surrounding how best to assess crime linkage. Bennell et al. argue that previous works fail to consider appropriate use cases for practitioners and further fails to assess what constitutes a sufficient amount of similarity between crimes to justify a linked classification. The authors argue at length the similarity between the linking task and other diagnostic decisions; that is for any given linking task based around an empirically derived score of similarity between crimes then the performance (correct classification rate) will vary depending on where researchers decide to place the decision cut-off threshold along the continuum of similarity scores. The authors argue that the AUC, from ROC analysis, is singularly well positioned to provide an unbiased estimate of overall performance across all such cut-offs. The sample used to illustrate this point consisted of 126 rape offences committed across the UK.

Tonkin et al. (2011) provides one of the first publications extending Bennell's empirical approach to crime linkage to diverse offenders. Prior studies had focused on discriminating behaviours surrounding a single given crime type; of which the entire sample being used was the same (see above for examples). Such crimes previously researched included domestic and commercial burglary, sexual offenses and robbery (not exclusive). Tonkin et al.'s 2011 study sought to assess whether linked crimes could be identified when they did not share the same crime type and/or specific behaviours.

Due to the lack of specific behaviours shared across overly dissimilar crime types, Tonkin et al. assessed how discriminatory temporal and spatial proximity were in identifying linked crimes. Further, they divided their sample of crime pairs into three levels: across crime categories, across crime types and within crime types. By subdividing in this way, they were able to assess whether linkage analysis performed better or worse given the type of links being made. Their results indicated that there was no significant difference between the three levels; the linkage analysis performed roughly as well in all cases. Furthermore, the authors reported similar performance to that which had been observed in previous studies. The authors concluded that that, again, inter-crime distance was a robust predictor of crime linkage and even go so far as to suggest that temporal proximity was not necessary.

Finally, Tonkin and Woodhams (2015) continue to develop cross-crime type linkage with this more focused study. Rather than assessing linkage status across completely dissimilar crime-types as the 2011 study – or perhaps in response to the inability to incorporate more specific behaviours into the analysis – the 2015 study looks at crime linkage across burglary and robbery crime types. These two crime types shared several behavioural domains while also being distinctly different: robbery offenses necessitate direct offender-victim interaction whereas burglary (often) does not. Furthermore, this study employed a much larger data set than is typical of most crime linkage studies - see (Bennell et al., 2014) for a comprehensive review – with a total n of 749 solved commercial burglaries and robberies. The results of this study echo results found in earlier linkage studies adopting Bennell's approach: high AUC levels were achieved and the full model combination of inter-crime distance, temporal proximity and Jaccard behavioural similarity between crimes provided the best performing model. Similar to Tonkin et al.'s 2011 study, Tonkin and Woodhams' 2015 study also subdivided crime pairs into three levels: burglary-burglary pairs, robbery-robbery pairs, and burglary-robbery pairs. They found no significant performance difference between these levels. Finally, the authors again note the relative strength of inter-crime distance as a crime linkage predictor.

Several key ideas can be gleaned from the preceding research summary: first, researchers appear largely confident that crime linkage is possible given observable behavioural information commonly collected during investigations. Second, of the collection of possible behavioural domains to focus on, inter-crime distance most commonly surpassed other predictors in its contribution to linkage predictions. Thirdly, there is a semi-standardized approach in the Bennell method to crime linkage. Fourth, there is emerging research into cross-crime type linking and the challenges associated with such methods. Finally, the AUC is often the most cited performance metric of the linkage task, but there is little mentioned in how such measures can be adopted by practitioners to improve investigative outcomes.

4.4.3 Limitations of Past Crime Linkage Research

Although past crime linkage research has routinely espoused the consistency and distinctiveness of the spatial and temporal relationship within linked crime series, such research has routinely over simplified the propinquity relationship between linked crimes while also simultaneously ignoring offenders' unique crime morphology. Both propinquity and morphology have been proposed as fundamental properties of offender geographic behaviour, and their oversimplification or omission stems from a tradition within the literature to ignore the primary driving component behind linked crime: specific individuals. Due to the manner in which studies within crime-linkage studies typically attempt to demonstrate predictive utility, the emphasis has been placed on the accurate prediction of discrete crime-pairs using aggregate crime trends. The end result has been a literature that consistently reports similar findings but in a way that fails to provide actionable insights on the part of practitioners as well as failing to provide explanations as to why the trends observed offer the utility described by researchers.

Canter and Youngs (2009) described propinquity and morphology as the basic attributes that can be used to describe offender spatial behaviour. Propinquity relates to the idea that the probability of a crime being committed by an offender decreases over distance (D. Canter et al., 2013; Canter & Youngs, 2009). Morphology on the other hand, describes the specific distribution of crimes for a given offender in Euclidean space. In other words, propinquity describes the proximity of crimes, either to a home or to each other, and morphology describes their shape. Crime linkage studies that have included distance measures have all affirmed the propinquity of crime in some fashion, with such measures often outperforming other predictive measures such as temporal proximity or behavioural similarity (M. Tonkin et al., 2011; Matthew Tonkin, Grant, & Bond, 2008). Due to the manner in which crime linkage data are usually generated, the morphology of crimes for a given individual is rarely discussed and is usually outright omitted.

The lack of including morphology results is a fundamental flaw in how crime linkage studies employ distance measures and by extension how they interpret the results of such measures. To understand why, consider the following example: a crime occurs that one would be interested in conducting linkage analysis upon. There are two crimes in the immediate area, each committed by a different offender, which may be linked to our crime of interest. Keeping in mind the propinquity property of linked crime, one would assume that the closer of these two crimes would have the higher likelihood of being committed by the same offender. This is the interpretation that is most often presented within the linkage literature. However, assume in our example both crimes are equidistant. In such a case, the propinquity argument does not provide sufficient information to differentiate between the two candidate crimes. This failure is a consequence of ignoring individual offender's morphology, which itself is a consequence of the typically employed research strategy which places crimes within a vacuum and ignores the individual who committed them. In the provided example, further information about other crimes an offender has committed and their spatial relationship with the crime of interest could help to differentiate between the two candidate offenders.

The most widely used approach within the crime linkage literature is predicated upon the initial works of Bennell and Canter (2002). This initial work, which has since been widely replicated and marginally iterated upon, presented the linkage task as one of binary classification (Bennell & Canter, 2002; Bennell & Jones, 2005; Bennell et al., 2009, 2014; M. J. Tonkin, 2012; Matthew Tonkin et al., 2008). To this end, Bennell and Canter presented a supervised learning approach to investigate the properties of linked crime. This was achieved by constructing individual crime pairs, some of which were committed by the same offender – i.e. positive for the linked case – and a comparison group of unlinked crime pairs – i.e. negative for the linked case. The way in which these crime pairs is constructed is a source of much debate within the literature, as there is no consensus on how these two groups should be properly sampled from a larger dataset of individual crimes (Bennell & Woodhams, 2012; Matthew Tonkin & Woodhams, 2015). The most common method involves randomly selecting two crimes per offender to create the linked case data, and an equal number of crimes committed by different offenders to represent the unlinked case.

This “sub-sampling” approach results in three distinct limitations that have direct consequences on the practical applicability of results. First, by randomly selecting a sub-selection of crimes per serial offender, past studies fail to account for offender learning and targeting preference. The consistency hypothesis suggests that where crimes occur for a given offender should be relatively stable, but there is empirical evidence to suggest that there is significant within series variation (D. Canter et al., 2013). Second, the random sub-selecting method implicitly assumes that all crimes within a series relate to each other in the same way. This assumption does not appear to be reasonable for all types of predictors, the most obvious of which being temporal proximity – a commonly used predictor within the literature. The temporal proximity relationship between crimes will obviously be different between a “spree” offender and a chronic recidivist for example. Practitioners have observed that offending patterns by prolific offenders can often fall into a predictable temporal rhythm, however such observations are only intuitive between temporally sequential crimes (Gwinn, Bruce, Hick, Cooper, & International Association of Crime Analysts, 2008). This relationship has been ignored by past studies as the model validation schemes used are not sensitive to time series information. Finally, in an effort to reduce the influence of overly prolific offenders, the sub-selecting approach results in models that are inherently nomothetic. Specifically, they reduce the problem space from one that asks how offenders differ from each other to one that asks how linked crimes differ from unlinked crimes. This aggregate approach ignores important between-subject variations such as their specific morphologies.

The present study proposes that the limitations outlined above can be at least partially addressed by introducing offender morphology into established linkage methods. Offender morphology has been explored in the context of geographic profiling as an ideographic approach to studying the relationship between crime locations and an offender's home (Canter et al., 2013). A similar approach could be taken within crime linkage analysis by introducing an ideographic measure of offender morphology which would relate crime locations to other crime locations within a series. One such ideographic measure that could be directly estimated from offender spatial activity could be an activity space. Activity spaces have been suggested both directly and indirectly within the criminology literature including classic theories such as crime pattern theory and routine activity theory as a means to describe offender spatial activity and targeting preferences (Brantingham & Brantingham, 1993; Cohen & Felson, 1979).

4.5 Criminal Activity Space and Crime Linkage

The discussion up to this point has focused on presenting the history and "standard practice" of crime linkage research, and it is from here that the criminal activity space construct can begin to be incorporated. As previously mentioned, crime linkage research is predicated on two fundamental assumptions: (1) offenders are generally consistent in their offending behaviour, and (2) that this behaviour is distinctive across individual offenders. Chapter 3 provided some initial evidence for offender spatial consistency: namely that future crimes could be predicted by the locations of past crimes for around 50% of cases.

Activity space based research seeks to quantify spatial activity into a measurable construct which can then be used to evaluate difficult questions regarding an individual's engagement with their environment. The activity space framework can have its roots traced back to Lewin' field theory (1951), which describes behaviour in terms of a "life space". This life space, Lewin argued, is contingent on the totality of personal and environmental factors from which it is built, and summarises the possible extent of behaviours available. More modern interpretations of activity space are more conservative in their formulations, but maintain the importance of the person to environment relationship.

Horton and Reynolds defined activity space as the “subset of all urban locations with which the individual has direct contact as the result of day-to-day activities” (Horton & Reynolds, 1971, p. 37). Over the course of the last 60 years, researchers have proposed a number of specific definitions for what an activity space is, and in general these working definitions all fall within a similar vein (Horton & Reynolds, 1971; Johnston, 1972; Lee et al., 2016; Perchoux et al., 2013). Namely, activity space is related to, and defined by, a geographical representation of an individual’s spatial behaviour which describes the relationship between an individual’s home and their frequently visited areas of activity.

The activity space construct has traditionally been described in terms of routine activities. Perchoux et al. outline: “The activity space, in reflecting daily mobility, is an individual measure of spatial behaviour” (Perchoux et al., 2013, p. 88). Past studies of activity space have done so through the use of travel diaries and geospatial data in an effort to estimate where, and how often, individuals travel to any given area (some examples: Dijst, 1999; Lee et al., 2016; Sherman, Spencer, Preisser, Gesler, & Arcury, 2005; Wong & Shaw, 2011). Activity spaces are typically a means to an end, and have been used to explore such questions as inter-urban movement and familiarity (Brown & Moore, 1970; Horton & Reynolds, 1971), residential preference (Johnston, 1972), daily activity (Lee et al., 2016), to evaluate segregation (Wong & Shaw, 2011) and as the foundation for a measure of exposure to foodscapes (Kestens et al., 2010).

The routine activity theory of crime as well as crime pattern theory both suggest that criminal events stem from the non-criminal day to day activities of offenders. As such, criminal behaviour would be expected to follow similar trends as non-criminal behaviours observed in the previously mentioned studies. Studies such as Canter et al.'s (2013) study would seem to suggest that this is indeed the case. Furthermore, crime pattern theory explicitly names an individual's morphology as an activity space (Brantingham & Brantingham, 1995). If offender spatial activity follows similar morphology and propinquity trends as observed in non-criminal activity space studies, then it would serve as a very powerful ideographic measure of individual offender spatial variation and targeting preference.

4.6 Study - Can Activity Spaces Improve Crime Linkage Analysis? (RQ 2.1)

This study's primary objective is to assess the predictive utility of criminal activity spaces in a crime linkage context using established linkage assessment practices. As the literature review outlined in the previous section has shown, several aspects of the crime event have been demonstrated to provide predictive utility in assessing whether a pair or set of crimes can be attributed to the same offender. These include: the spatial proximity of crimes as well as the temporal proximity of crimes. These two primary crime attributes have been demonstrated to be significant even when applied to crime pairs of dissimilar crime types (Matthew Tonkin & Woodhams, 2015).

The methodology used will follow Tonkin and Woodhams' (2015) study methodology closely as it is one of the few studies within the crime linkage literature that attempts to incorporate diverse crime type offending. This is important for assessing the efficacy, if any, of the criminal activity space construct has in predicting crime linkage status. Recall that the criminal activity space construct is predicated on the assumption that the spatial distribution of crimes for a given offender is reflective of some individual decision making process, and that the resulting spatial pattern of crime for a give individual would be consistent regardless of the specific crimes that comprise it.

Past studies of crime linkage research have proceeded thusly: first salient features of crime are chosen from which to build the linkage models, next the data are split into training and testing data sets, then a linkage model – typically logistic regression but other approaches have been described – is fit to the training data set, finally model performance is assessed – usually via ROC analysis and the area under the curve (AUC) statistic – on the validation data set.

The current study will follow a similar procedure whereby several linkage models will be produced each with slight variations in what predictors are used. The central hypothesis is that: if the criminal activity space construct is representative of offender spatial behaviours, then models that include criminal activity space measures will be more reliable (better apparent performance) in their linkage predictions than models without criminal activity space measures.

4.6.1 Method

4.6.1.1 Data

The data used in the present study consisted of the same sample used in the previous chapter with one exception. Where the previous chapter included all offenders whom had committed at least two crimes the present study had to further restrict this sample to offenders who had committed at least three crimes. This was due to the specific train vs test methodology adopted (explained below). Thus the sample used in the present study consisted of 4,446 crimes committed by 997 offenders. This sample was pulled from a larger collection of solved crimes (n = 15,409). These data consisted of a large variety of crime types committed from August 2011 to

August 2015 committed in a major U.K metropolitan city. The data were provided directly from the police force with individually identifying information removed (i.e. no names etc.).

Due to the manner in which the police force handled concurrent crimes – multiple offence charges – an “offence” as its used here was defined to be a criminal event that occurred at either a unique location, or on a unique day. This reduced bias towards locations for offenders who received many multiple co-charges for a single crime event. For example, if an offender was arrested for a crime on Monday the 23rd and charged with multiple counts of burglary for a number of different flats within the same building, then the police would typically log each of those burglaries as a unique crime event. From the burglar’s perspective, however, they only visited one building of flats. In such instances where crimes for a given offender share the same geo-location and date, one of the crimes would be randomly selected and included in the analysis. Once this was complete, offenders whom only had two or less crimes on record were removed from the data.

Next, to facilitate the validation procedure employed in the current study, the data was split such that the last crime committed by each offender was removed and set aside into a “validation set”. This resulted in 3,449 crimes for model fitting, and 997 crimes for validation. This specific splitting formulation was chosen as it most closely resembles a potential use-case for practitioners and because it provides direct assessment of how predictive a linkage process can be of future crimes given past linked crimes.

Finally, for each crime, a number of salient data fields was selected for inclusion in the linkage analysis. These selected fields were consistent with past studies of crime linkage analysis and included: (1) the offender ID number, (2) the Crime ID number, (3) the easting / northing (geo-codes) of the offense, (4) the date the offence occurred, (5) the type of offence (burglary, robbery, assault etc.), (6) the type of location the offence occurred (street, driveway, shop etc.; this was a provided field by the police). These data fields represent a small selection of all available fields; specific M.O. behaviour fields were not included for several reasons. While there were upwards for 300 unique M.O. data fields provided by the police, for many of the crimes these fields were largely absent and the computational cost of including them relative to the benefits likely to be gained from them was low. Tonkin and Woodhams previously demonstrated that in instances of cross-crime type linkage the M.O. data fields were consistently the weakest predictors, even when the crime types were highly similar in motivation (burglary and robbery). Given the wide-net approach of the criminal activity space construct, it made little logical sense to assess the behavioural similarity of an assault to that of a burglary for example. By limiting the scope of included predictors primarily to spatial and temporal factors, the efficacy of the criminal activity space construct could be more directly observed.

4.6.1.2 Analytical Strategy

The procedure used in this study follows the general approach originally outlined by Bennell and Canter (2002). All data analysis and feature construction was done using the R platform, code can be found in . Briefly, the Bennell method involves creating crime pairs from which to build a predictive model that estimates the linked status for a pair of crimes. These crime pairs represent the unit of analysis and require some initial feature construction.

4.6.1.2.1 Data Set-up and Feature Creation

First, the crime-pair data was constructed from the 3,449 crimes present in the training data set. This corresponds to an *n* choose k problem, where k = 2 and n = 3,449. This results in 5,946,076 total unique crime pairs. For convenience, and for later steps, these crime pairs were constructed in form of (crime1, crime2) where crime2 always occurs *after* crime1. For each of these crime pairs, a dichotomous variable “Linked” was created where if both crime1 and crime2 shared the same offender ID then it took the value 1 and 0 otherwise. From the 5,946,076 crime-pairs, there were 8,021 unique linked (committed by the same offender) crime-pairs and the rest were unlinked (committed by different offenders). Given that the goal is to predict linked crimes and that the total number of crime-pairs was so large, the unlinked crime-pair data was randomly sampled such that the final crime-pair training data set consisted of a 5:1 mix of unlinked crime-pairs to linked crime-pairs. This process resulting in 48,126 unique crime-pairs for model fitting.

The data validation set was constructed in a similar manner. Each of the 997 crimes (the final crime for each offender) was combined as crime2 with each of the 3,449 crimes from the training set. Because the crime-pair itself represents the unit of analysis, and not the individual crimes, this does not violate the assumptions of independence. This results in 3,438,653 unique crime-pairs for validation. In order to assess whether over fitting was occurring, a third data set was created as a sub-set of this validation set. Like the training set, all linked crime pairs in the validation set were selected, as well as 5x this number of unlinked crime pairs. This resulted in a “test set” of 20,694 crime pairs. The smaller sub-sampled “test set” allows for direct comparisons of model performance between training and testing, and the “full” validation set provides an indication of performance when sub-sampling does not occur – as it would in a “real life” scenario.

4.6.1.2.2 Predictor Construction

For all crime-pairs in the training and validation sets, the spatial proximity, temporal proximity, same location type, crime type combination linked base rate, and finally two criminal activity space measures. Spatial proximity was calculated as the Euclidean distance between the two crimes within each crime-pair (measured in meters). Temporal proximity was calculated as the number of days separating crime2 from crime1. The “same location type” variable was a dichotomous variable which equalled 1 if both crimes occurred in the same type of location (i.e. “street”, “alleyway”, “store” etc.) and 0 otherwise.

The crime type combination linked base rate was slightly more involved. For each crime type combination, the probability that that crime type combination would be observed given that the crime-pair was linked was calculated. This was calculated as the ratio of the number of linked crime-pairs with a given crime-type combination over the total number of observed crime-pairs within the training set (n = 48,126). This measure was included as a means to help differentiate crime-pairs with rare crime-type combinations which otherwise may be overly similar spatially or temporally.

Two measures of criminal activity space were calculated. This process required the formal sequence definition of the crimes within the crime-pair. For each crime pair, the criminal activity space for the offender of crime1 within the crime pair was calculated. This process utilized the same procedure outlined in Chapter 3. In order to not expose the predictors to unintended data, the crimes used to estimate the criminal activity space was limited to those crimes that occurred before crime2 within a given crime-pair. In practice this means that if crime1 within the crime-pair was the third crime in sequence for a given offender, then the criminal activity space calculated for that crime-pair was generated from the first three crimes of crime1’s offender’s crime series only. From this criminal activity space, two measures were then calculated.

Because criminal activity spaces can have multiple clusters, the following measures were calculated for each cluster and then aggregated in some fashion. The first measure was the centroid to crime distance. This was calculated as the Euclidean distance from the criminal activity space centroid of each cluster of crime1's offender's space to crime2. From these distance measures, the minimum distance was retained. This corresponds to the centroid-to-crime distance from the closest cluster.

Finally, an "activity space inclusion" variable was calculated. This was performed by comparing the centroid-to-crime distance to the radius of the criminal activity space. If the centroid-to-crime distance was less-than-or-equal-to the criminal activity space's radius, then the crime must be located within the circular space. The final inclusion variable took on a value of 0 if crime2 did not fall within any clusters of crime1's offender's criminal activity space, 1 if it fell within one cluster; 2 for two clusters and so on.

4.6.1.2.3 Linkage Model Construction

Before Logistic regression was used to generate the linkage models, a non-parametric estimation method was used to assess the apparent impact of the identified predictors on a crime-pair's linked status. This method is adopted from the modelling strategies outlined by Harrell (2001). Outputs of this initial analysis can be seen in B. It can be seen from these outputs that the continuous predictors, inter-crime distance, days elapsed, centroid-to-crime distance, do not vary linearly with the dependent variable Linked. This is not surprising given the literature surrounding distance decay functions - all of which describe a similar negative exponential curve (O'Leary, 2011). This means that it is unlikely that a single linear coefficient will sufficiently capture the observed variation within the data (Harrell, 2001). Regression splines can be used to allow for non-linearity in a continuous variable (Harrell, 2001). To handle the observed non-linearity here, a restricted cubic-spline with five knots (corresponding to four unique intervals) was used.

The use of splines is not standard practice within the crime linkage literature and indeed none of the studies presented in this chapter's literature review make use of them at all. Thus in order to assess what impact splining the continuous predictors has, as well as to isolate any potential impact the criminal activity space measures may have, three primary logistic regression models were created.

Model One most closely resembles those presented by Tonkin whereby linkage status was assessed across a diverse set of crimes much like it is here via spatial and temporal measures only (M. J. Tonkin, 2012). This "no-spline, no-AS" (AS corresponds to "Activity Space") represents the "control" model against which gains are evaluated. Thus model one consists of crime-to-crime distance, days elapsed, linked crime-pair probability, and a same location check.

Model Two – "Spline, No AS" – adds cubic splines to the continuous predictors of crime-to-crime distance and days elapsed. Like model one, model two also incorporates linked crime-pair probability and a same location check.

Model Three – "Spline, AS" – adds the two criminal activity space measures, total activity space inclusion and minimum centroid to crime distance – to the predictors outlined in model two. Again the continuous variables are modelled with restricted cubic splines with five knots.

The logistic regression models were generated using the R package "rms" (Harrell, 2001). The modelling strategy progressed in three phases: phase one consisted of sub-dividing the test set into five "folds" for five-fold cross validation. Five-fold cross validation is a process by which the data are randomly divided into five "folds" and the model in question is generated using data from four of the five folds and then tested on the fifth fold. This process is repeated until n models are built, where n is the number of folds created. For each model, the AUC, area under the ROC curve, was recorded. This was done in order to gauge initial model performance and stability, and in particular assess whether the number of identified knots in the spline functions produced a stable model.

Phase two consisted of fitting the specified models using the entirety of training data set described previously. Several performance metrics were captured on the training set including the linked probability distributions as well as the reported AUC values. The models generated here were then applied to the test set in order to assess the stability of the models on data not yet seen. Because the test set was downsized in a similar method as the training set, the predictive performance observed during the initial five-fold validation should be similar to the performance observed on the test set.

Finally, phase three consisted of applying the three models previously described to the entirety of the validation set (n = 3,438,653 crime pairs). This phase was included as it allows direct inspection of how the down-sampling procedure influences the distribution of linked predictions. Because the test set described in the previous paragraph includes all the linked pairs (by construction) and a mix of randomly sampled unlinked pairs, it is expected that apparent performance on the full validation set will be worse.

4.6.1.2.4 Assessing Performance

Model performance was assessed using several measures. First, apparently model performance statistics were collected for each model including: pseudo R^2 and the AUC. The use of the AUC has become standard practice within the crime linkage literature, and notionally it provides a quantitative assessment of how discriminatory a diagnostic test is given a continuous predictor variable (Bennell & Jones, 2005). In this case, the predictive value is the linked probability assigned to a crime pair from the fit logistic regression models. The AUC is a value that ranges from 0 to 1 and comes from the area under the ROC (receiver operating curve) curve; though for practical purposes. Within the crime linkage literature, AUC values have been broadly interpreted following Swets (1988) guidelines: namely values between 0.5 – 0.7 are "low", 0.7-0.8 are "moderate" and 0.9 and above are "high". Thus the AUC provides an indication for not only the raw performance of a diagnostic test, but also a means by which to compare competing tests. Informally, models with higher AUC values are thought to be more discriminatory.

Classification performance was also recorded in terms of the binary classification outcomes. Because logistic regression typically provides a continuous "score", a decision threshold was developed using the Youden's J statistic which attempts to maximize the sensitivity (proportion of correct positive classifications or true positive rate) and specificity (proportion of correct negative classifications or true negative rate) (Youden, 1950). Linked probability scores for a given model above their respective decision thresholds (derived from their respective ROC curves), were classified as "linked" and crime-pairs whose scores fell below said threshold were classified as "unlinked". This allowed for the construction of a truth table which compares the predicted classification rate versus the actual class (linked vs unlinked) within the data.

The binary classifications yield four possible outcomes: true positives (TP), false positives (FP), true negatives (TN) and false negatives (FN). From these four outcomes, five measures of model performance were calculated: sensitivity, specificity, precision, accuracy and Mathew's correlation coefficient. The specific forms for each of these five measures are presented in Table 7. Sensitivity, specificity, precision and accuracy can all take values from 0 to

1, where 1 represents no errors and 0 represents complete error. The MCC can take a value between -1 to 1, where -1 is total disagreement between predicted and actual values, 0 is random prediction and 1 is perfect prediction. Of these five measures, given the extremely unequal group sizes between linked and unlinked crime pairs, the MCC provides the most "balanced" representation of overall classification performance, and accuracy provides the least "balanced" representation. These five measures were calculated for all three linkage models for both the test set and the full validation set.

Table 7: Confusion Matrix Metrics

Sensitivity	$\frac{TP}{TP + FN}$
Specificity	$\frac{TN}{TN + FP}$
Precision	$\frac{TP}{TP + FP}$
Accuracy	$\frac{TP + TN}{TP + TN + FP + FN}$
MCC	$\frac{TP * TN - FP * FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$

4.6.2 Results

Three logistic regression models were fit to the training data. These models corresponded to three different mixes of predictors with each subsequent model growing in complexity. The first model – "No Spline, No AS" – included no activity space based measures and the continuous variables, ICD (inter-crime distance) and temporal proximity, were not splined; i.e. they were assumed to have a linear relationship with the linked likelihood of a given crime pair. This model corresponded to the nearest "gold standard" from past crime linkage works that included cross crime-type associations. The second model – “Spline, No AS” – also included no activity space based measures, but differed in that the continuous variables were splined with cubic splines with five knots. Finally, the third model – “Spline, AS” – included both splining as well as two activity space based measures: minimum centroid to crime distance and activity space inclusion.

The three models described above were fit to the full training set, summary descriptions for the three fit logistic regression models can be seen in Table 8. These represent the “final” models. For each model, the model’s likelihood ratio, R^2 and AUC - with 95% confidence intervals - from ROC analysis are all reported. All three models were significant as evidenced by their likelihood ratios, and all three models achieved moderate AUC scores. From the results of the models alone, the “Spline, AS” model achieved the highest likelihood ratio (LR = 18247.5), R^2 (0.531) and AUC (0.895) values of the three models. The differences in AUC values were statistically significant across the three models as evidenced by the 95% confidence intervals. Again the “Spline, AS” model displayed the strongest performance according to the AUC (95% CI = 0.8912-0.8995) compared to the other two models (M2 95% CI = 0.882-0.8907 and M1 95% CI = 0.8777-0.8865).

Table 8: Logistic Regression Model Summaries

Model	Predictor	Logit (SE)	Wald	Model X^2(df)	R2	AUROC (95% CI)
No Spline, No AS (M1)	Intercept	0.1469 (0.0313)	4.69*	15838.2 (4)	0.472	0.882 (0.8777-0.8865)
	ICD	-0.0006 (0)	-66.44*			
	TP	-0.0014 (0)	-30.83*			
	CT-P	129.0179 (3.4208)	37.72*			

	SLT	0.726 (0.0375)	19.36*			
Spline, No AS (M2)	Intercept	1.7686 (0.0589)	30.01*	17274.24 (10)	0.508	0.886 (0.882-0.8907)
	ICD	-0.0016 (0)	-36.86*			
	ICD’	0.0073 (0.0004)	17.82*			
	ICD’’	-0.0183 (0.0012)	-14.77*			
	ICD’’’	0.0154 (0.0014)	11.29*			
	TP	-0.0107 (0.0005)	-21.55*			
	TP’	0.1307 (0.0088)	14.84*			
	TP’’	-0.2264 (0.0166)	-13.64*			
	TP’’’	0.1067 (0.01)	10.64*			
	CT-P	120.2632 (3.5249)	34.12*			
	SLT	0.651 (0.0398)	16.34*			
Spine, AS (M3)	Intercept	1.7171 (0.0671)	25.6*	18247.5 (15)	0.531	0.895 (0.8912-0.8995)
	ICD	-0.0009 (0.0001)	-13.86*			
	ICD’	0.0058 (0.0006)	9.75*			
	ICD’’	-0.0161 (0.0018)	-8.94*			

	ICD’’’	0.0155 (0.002)	7.72*			
	CtoCD	-0.0008 (0.0001)	-11.82*			
	CtoCD’	0.0018 (0.0007)	2.6			
	CtoCD’’	-0.002 (0.002)	-1			
	CtoCD’’’	-0.0015 (0.0021)	-0.72			
	TP	-0.0101 (0.0005)	-19.98*			
	TP’	0.1246 (0.009)	13.84*			
	TP’’	-0.2163 (0.017)	-12.76*			
	TP’’’	0.1029 (0.0102)	10.05*			
	CT-P	116.6325 (3.6411)	32.03*			
	SLT	0.6055 (0.0409)	14.79*			
	AS	0.5262 (0.045)	11.7*			

Note:
‘ = Piecewise intervals for continuous predictors
ICD = inter-crime distance (in meters)
CtoCD = Centroid to crime distance (in meters)
TP = temporal proximity (Days elapsed between crimes)
CT-P = Linked Crime Type Combination Base Probability
SLT = Same offence location type (Street, house etc.)
AS = Activity Space inclusion (0-4)

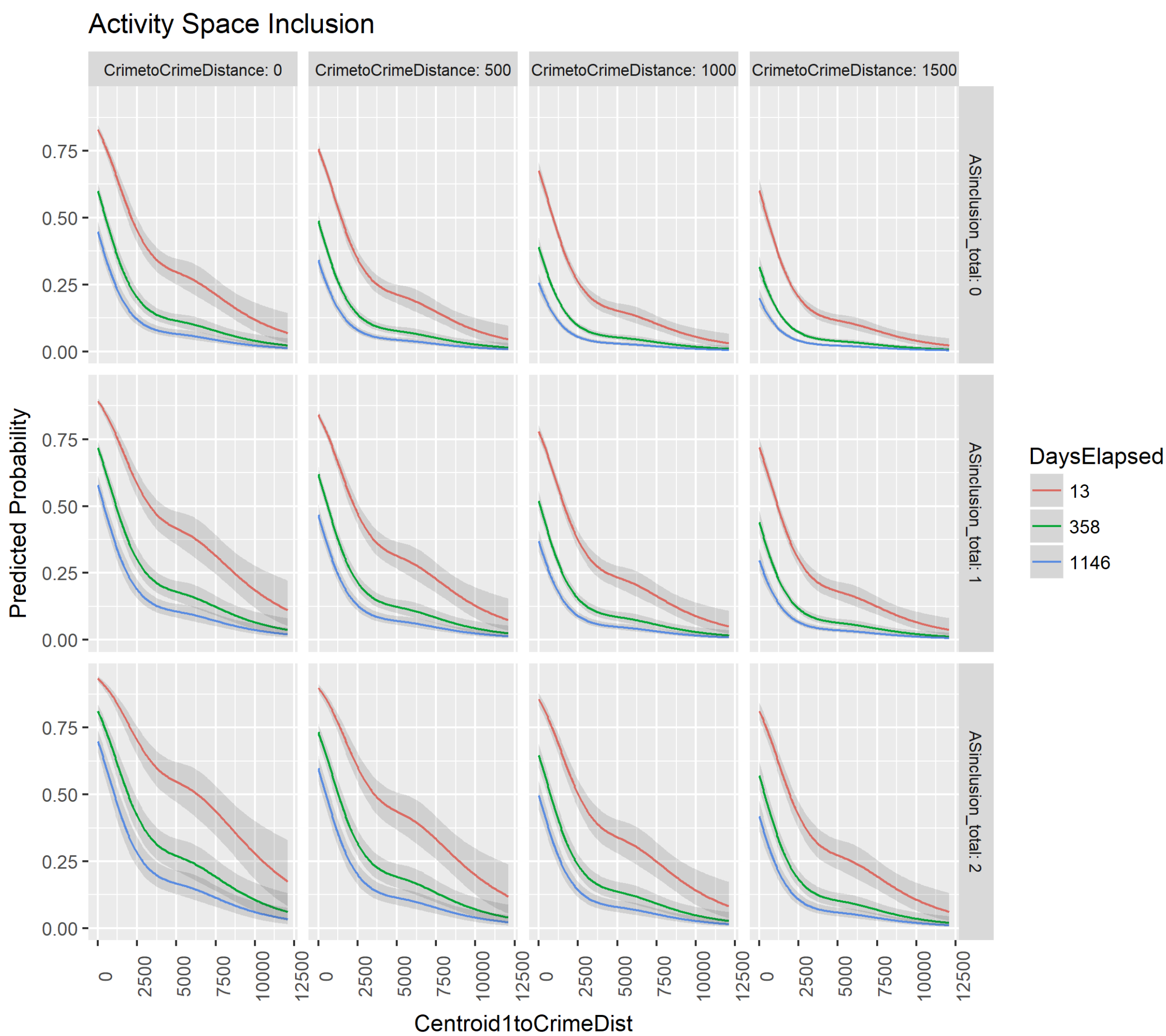


Figure 34: Activity Space, Spline Model Predicted Linked Probability Plot

Figure 34 shows the predicted probability over a range of input variable levels for the strongest model – the 'AS – Spline' model (variables not shown are held constant at median values). This provides a more intuitive graphical display of how the various predictors interact within the final model. Each line corresponds to a fixed number of days elapsed between a candidate crime and the target crime; these correspond to the 25th percentile (13 Days), the 50th percentile (346 Days), and the 75th percentile (1146 Days). As can be seen, the general patterns found in past studies are present. As the number of days elapsed increase, the linked probability decreases, though it appears to decrease slower the farther part two crimes become. This same pattern is present within the two distance measures, and is most readily apparent in the centroid to crime distance depicted along the x-axis. There is a very pronounced 'elbow' effect for 0 and 1 activity space inclusion as the centroid to crime distance increases. This elbow effect nearly disappears when a pair falls within two activity spaces; implying in such cases the fact that a crime falls within multiple spaces requires greater distances to discount the pair as being linked. The top-left most plot can be used to illustrate that the two distance measures are capturing a different subset of behaviour. The top-left most plot suggests that the linked probability between two crimes, where the inter-crime distance is zero or close to zero, and that the centroid-to-crime distance is very long (as would happen when the crimes are situated along the edge of an activity space) is very low. This helps to explain why straight crime-to-crime distance is itself not a fully sufficient special predictor for crime linkage status. Finally, as expected, pairs that are closest in all measures (days elapsed, crime-to-crime distance, centroid-to-crime distance) then the predicted probabilities are highest (see: lower-left plot).

It is well understood that models will achieve lower likelihood ratios with fewer variables than models with more variables; however there is a question as to whether the differences are statistically significant. The likelihood ratio test is a chi-squared test statistic with degrees of freedom equal to number of constrained parameters – i.e. variables not shared between models. Because each subsequent model represents an increasingly complex model, only tests between sequential tests are reported in Table 9. Results indicate that each model represents a significant improvement over its simpler version; culminating in the "Spline, AS" model once again performing best.

Table 9: Model Likelihood Ratio Significance Tests

	LR 1	**LR 2**	**Likelihood Ratio**	**d.f.**	**Sig**
M1 - M2	15838.2	17274.24	2872.08	6	p < 0.001
M2 - M3	17274.24	18247.5	1946.52	5	p < 0.001

Next, model fit was assessed using calibration plots. The "calibrate" function from the "rms" R package was used to construct bootstrap estimates of the calibration curve (Harrell, 2001). For each plot, an "ideal" diagonal reference line is provided as well as the actual calibration curve. Deviations from ideal represent under or over-fitting of predicted values. Results indicate that the simplest model – "No Spline, No AS" – presents the worst calibration curve with obvious deviations. The calibration curves for the remaining two models are essentially identical indicating good fit with only slight deviations from ideal.

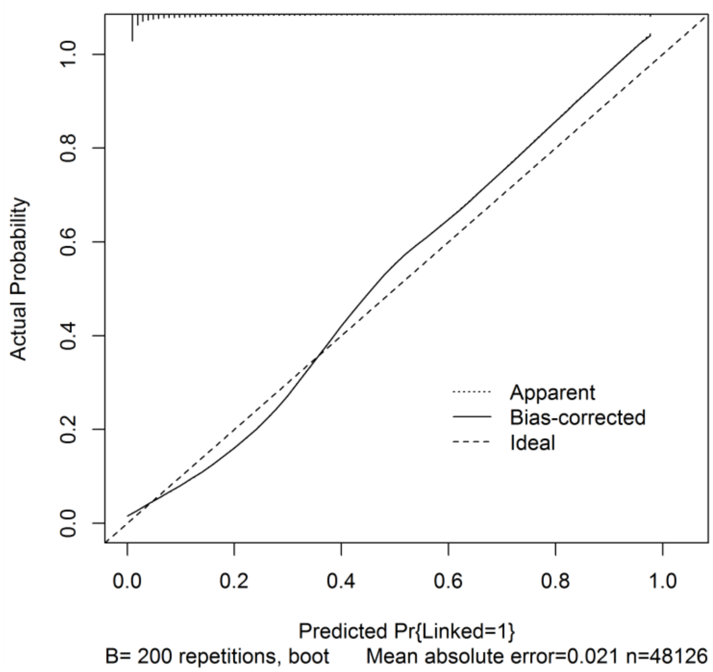


Figure 35: No Spline, No AS Model Calibration Plot

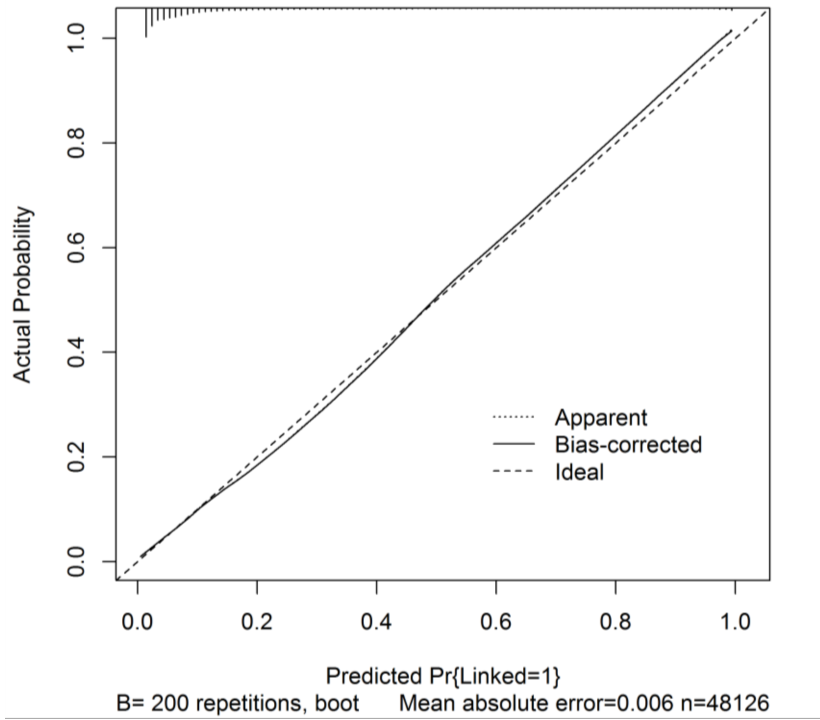


Figure 36: Spline, No AS Model Calibration Plot

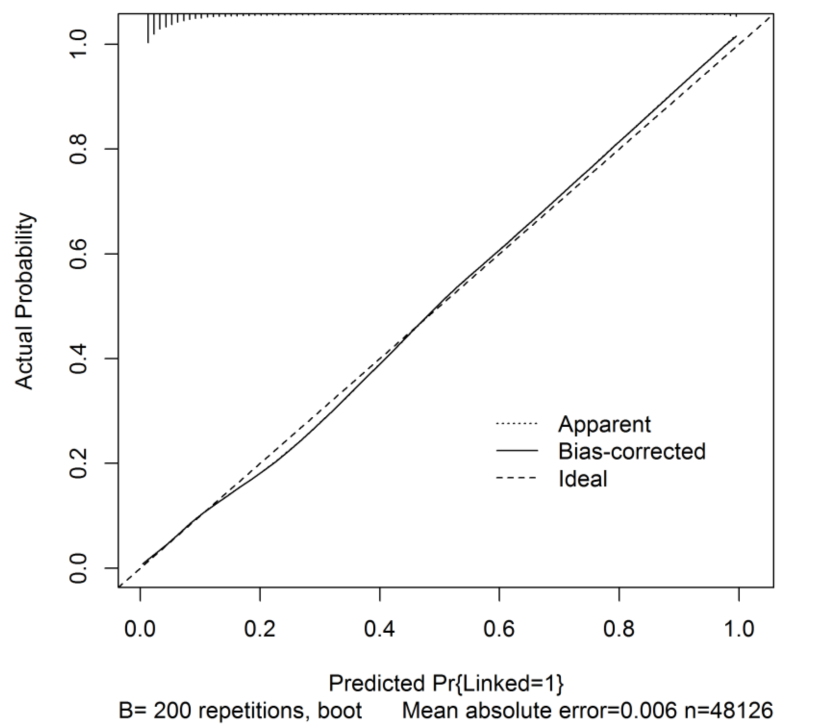


Figure 37: Spline, AS Model Calibration Plot

Next, classification performance of each of the three models was assessed on each of the three data sets: training, test and validation. It should be noted that transforming a continuous output from logistic regression (or any model) into a binary classification necessarily results in significant information loss. Despite this limitation, this process is presented here as a means to estimate the possible practical utility in making actionable decisions regarding linked crime. For each fit model, the ROC curve was constructed (see:), from which a decision classification threshold was calculated using Youden's index (Youden, 1950). This is consistent with past research (M. J. Tonkin, 2012). Then for each data set, each crime pair was classified as either "Linked" if their linked likelihood was greater than the calculated decision threshold, or "Unlinked" otherwise. This resulted in a 2x2 confusion table whose cell counts corresponded to the number of true positives (TP), false positives (FP), true negatives (TN) and false positives (FP) for each model on each data set. These confusion tables can be seen in , page 264.

The measures outlined in Table 7 were then calculated for each of these nine confusion tables. Results of these various measures of classification performance are reported in Table 10. Of the five reported metrics, the MCC provides the most general assessment of overall classification performance and indicated that for all data sets the "Spline, AS" model performed the "best". Given the extremely unequal group sizes, the accuracy measure is generally misleading, but is reported here as it demonstrates that the "Spline, No AS" models is the most conservative in making linked classifications. It should be noted that the MCC and precision measures indicate very poor performance on the full validation set regardless of the model used.

Table 10: Confusion Matrix Performance Metrics by Linkage Model by Data Set

Data Set	Model Name	Sensitivity	Specificity	Precision	Accuracy	MCC
Train	No Spline, No AS	0.757	0.850	0.503	0.835	0.522
	Spline, No AS	0.749	0.864	0.525	0.845	0.537
	Spline, AS	0.784	0.848	0.508	0.837	0.539
Test	No Spline, No AS	0.605	0.868	0.478	0.824	0.431
	Spline, No AS	0.626	0.873	0.497	0.832	0.457
	Spline, AS	0.709	0.849	0.485	0.826	0.485
Validation	No Spline, No AS	0.605	0.869	0.005	0.869	0.044
	Spline, No AS	0.626	0.876	0.005	0.875	0.048
	Spline, AS	0.709	0.850	0.005	0.850	0.049

4.6.3 Discussion

Three logistic regression models were fit to the data in order to assess the predictive power of (1) Criminal activity space measures and (2) the impact of allowing non-linear variation in continuous variables. The three fit models thus far presented correspond to the baseline model most commonly used throughout other crime linkage research which involves no activity space based measures and no splines. As the goal of the current study was to investigate the predictive utility of activity space based measures, and because no previous crime linkage studies have thus far introduced splined predictors, a model that included splined predictors without activity space based measures was necessary to adequately assess the impact of said activity space measures.

The results paint a very clear picture: splining predictors provides a direct and appreciable gain in predictive power to the logistic regression models. This is apparent from the likelihood ratio gain (see Table 9) which also indicates that the inclusion of criminal activity space measures can improve these models even further. In terms of relative impact, splining just inter-crime distance, the most commonly collected linkage feature, provides a larger relative gain than does the inclusion of the two criminal activity space measures (LR gain of 2872.08 vs 1946.52). At face value, these results indicate that reasonable improvements to linkage models such as these at no additional data acquisition cost, which would not necessarily be the case if one were to include the activity space based measures outlined.

An extensive array of validation tests were performed to assess the degree to which the models described were over fit or not as well as determine each models' validity. Results from the bootstrapped calibration curves indicate that over fitting is not readily apparent. Model validity was further provided by the similar performances reported for each model between the training and test data sets. The test data was also unique relative to other studies of crime linkage in that it was constructed prospectively; that is all crimes for which linkage assessment was performed were the most recent crimes within the data set for each offender. Thus the high apparent performance observed on the training set implies that the modelling methodology would behave similarly in applications on "real world" data on newly occurring crime.

The specific performance of the three models was presented in two ways: (1) as the area under the ROC curve which is a general performance index of a predictive variable on a diagnostic test. The AUC has become the most widely reported statistic within the crime linkage literature as well as being extensively used in medical diagnostics research (Cook, 2007; Woodhams & Bennell, 2015). AUC results from the ROC analysis were in agreement with the likelihood ratio tests: specifically "Spline, AS" model – the mode incorporating both splining and activity space based measures, was observed to have the highest AUC values. Furthermore, the “Spline, No AS” model reported higher AUC values than the baseline “No Spline, No AS” model. This further suggests that both splining continuous variables as well as adding criminal activity space measures increase predictive performance. The AUC, however, does not itself provide any actionable information regarding specific classification performance. For example, an AUC of 0.9 – which would be very “high” – means that if you take a positive sample, in this case a linked crime pair, then it will have a higher score than an unlinked sample 90% of the time. Models with higher AUC values are better at rank ordering predictions such that true positives in general have higher scores than true positives, but if one is interested in making a specific decision about the classification of a given crime pair the AUC is of no practical use.

The second way model performance was assessed addresses the problem outlined previously. This was to evaluate specific classification performance for a given decision threshold. Crime pairs were separated into a linked or unlinked group according whether their linked probability met a specific decision threshold. This threshold was determined statistically using Youden’s index. Dichotomizing the logistic regression outputs into linked or unlinked classifications in many ways oversimplifies the results. For example, crime pairs on either side of a threshold end up being classified differently, even though their predicted probability is very similar. Despite this limitation, the classification results presented in both and Table 10 provide a number of important insights.

First, there is no single measure that fully captures the “full picture” of classification performance from a confusion table. This is evidenced by the spread observed the five provided measures. Accuracy in particular has been used in some studies as evidence of classification performance which is a highly misleading practice. To understand why, one need only look at the accuracy comparisons between the training set and validation set. Due to the sheer number of possible crime combinations and the relative rarity of linked crimes, there is an extreme difference in group sizes between linked crimes and unlinked crimes. In such cases, accuracy measures are not very informative as one could achieve very high accuracy measures simply by assigning all crime pairs to the majority class: unlinked. This is the primary reason as to why accuracy scores seem to increase on the full validation data set compared to the test set in Table 10.

Furthermore, it can be seen that the sensitivity of the tests does not change between the test set and the validation set; this is expected because both data sets contain the same linked pairs. The difference arises in the specificity which assess how well the test is identifying true negatives; increases in sensitivity scores implies that as more cases are added to the sample a larger proportion of them are obviously negative. The obvious problem that arises then, is that not all of these new crime pairs are obviously negative, and it is this ambiguous cases that get identified as false positives. The impact of this can be seen in the precipitous drop in precision between the test set and the validation set.

Precision, or the positive predictive value, assess how accurate the test is given a positive prediction. For both training and test, precision was around 50%; that is given a positive prediction, that positive prediction would be a true positive only about 50% of the time. As more crime pairs were added, given the imperfect specificity, the number of false positives necessarily increases. This inevitability is reflected in the lowered precision score on the validation set. The MCC sees a similar decrease as precision for many of the same reasons. However, the MCC does indicate that for all three data sets, the “Spline, AS” model does perform the “best” overall. Keep in mind, however that these specific classification performance metrics are themselves dependant on what the classification decision threshold is set to. Ideally this decision threshold would be set with explicit knowledge of the trade-offs between false positives and false negatives. Without more specific information to determine which of the two outcomes is most important to minimize, we are left with the presented "best guess".

The classification results imply that the presented tests are better at predicting if two crimes are *not linked* rather than if two crimes *are* linked. At best, positive predictions are suggested to be correct only 50% of the time, compared to the ~99.9% accuracy of negative predictions. This is unsurprising given the extremely rare occurrence rate of linked crime given the methodology. Recall that the initial training set was constructed such that there were 5,946,076 unique crime pair combinations, and of these only 8,021 crime pairs were actual linked pairs. This represents a ratio of approximately 740 unlinked crime pairs for every one linked crime pair within the sample. The employed methodology utilized a down-sampling procedure to randomly sample from the unlinked crimes to create a data set of a more manageable size where the ratio of unlinked to linked was 5 to 1. This procedure was similar in spirit to those used in other studies (M. Tonkin et al., 2011).

The obvious limitation of this procedure, and procedures like it, is that the actual incidence rate is lost. This results in models whose outputs are optimistic especially in instances were threshold based decisions are required (Freeman, Moisen, & Frescino, 2012). The AUC, however, is not itself sensitive to differences in group sizes in the same way some of the threshold based measures are. However, this results in a situation where one can report models with moderate to high AUC scores but observe poor predictive performance on one of the classes – the linked class in this case.

This discussion is important because while most crime linkage research has focused on assessing predictive ability of various measures in general, issues such as proper decision threshold setting are of central importance in moving such research forward into more practical – not purely academic – applications. However, given that the central question was concerned with assessing the predictive gains associated with activity space measures, it seems generally supported given the present data that such measures do offer some improvement.

Originally, the activity space measures were developed to act as “tie breaker” in linkage cases where two candidate crimes were nearly equidistance away from the same crime of interest. It was hypothesized that if the crime of interest was a location within a given offender’s space and not another’s, then it would be more likely that the offender whose space covers the crime of interest committed said crime. The various metrics reported here tend to support this interpretation, with the model containing activity space based measures outperforming the two alternatives in terms of both AUC and threshold based classification via the MCC.

None of the reported models offered much predictive utility in identifying linked crimes on the full validation set. This is most attributable to the large increase in candidate crimes being evaluated and subsequently the likelihood that any of those crimes would resemble a linked crime pair necessarily increased. Thus it would seem that the present methodology is limited in its applicability to threshold based linkage tasks.

A key criticism of the current set of models is their lack of behavioural information outside of location choice as well as the modelling technique used. It is possible that the predictive shortcomings observed here may be reduced by adopting a more flexible modelling strategy that allows for greater investigation of variable interactions and as of yet ignored behavioural data. Thus the next study attempts to address some of these shortcomings by introducing a new modelling technique as well as extensive behavioural information.

4.7 Study – Do Machine Learning Models Improve Crime Linkage Performance? (RQ 2.2)

The first study of this chapter provided an in depth analysis of the benefits of adding non-linear splines and criminal activity space measures in assessing crime linkage status. The primary goal of said study was to assess whether these two features contributed to an apparent increase in linkage ability and by extension serve as validation of the underlying assumptions of activity spaces. However, the previous study represents only one modelling technique's interpretation. Furthermore, the previous study's use of logistic regression resulted in a number of important limitations: first, due to the diverse nature and mixture of crime types it was not possible to explicitly model offender behaviours as has been done in previous crime linkage research. It is possible that in the absence of such information the apparent strength of spatial predictors may be inflated.

Secondly, logistic regression, as with all regression modelling strategies, requires potential variable interactions to be explicitly defined. In introducing a large number of behaviour domains, the number of potential interactions increases dramatically, and it is not always obvious how such variables should be expected to interact. Logistic regression alos assumes a linear relationship between the independent variables and the log odds of the dependent variable; an assumption that is not always sound. This relates to the behavioural actions limitation mentioned previously: as it is not obvious that the presence of a burglary specific action and its impact on the likelihood of that crime being linked to a violent assault would be linear. While the presence or absence of a behavioural domain in one crime may be significant to actions observed in another unrelated crime type, modelling this relationship within logistic regression would have been overly complex.

Machine learning is a branch of computer science and statistics that is focused on the development of algorithms which enable computers to perform difficult automated tasks (Bishop, 2016). Examples of ‘non-trivial’ tasks include: ‘mastering’ game of Go (Silver et al., 2017), facial recognition (Teller & Veloso, 1995), and life sciences research (Chen, Elenee Argentinis, & Weber, 2016). A common thread between these various examples is the non-deterministic nature of the task at hand: it was no simple matter to simply program a computer to be good at Go, for example, given the sheer size of the decision domain space. Instead, algorithms have been developed whereby the computer is left to make the determination about what data is important and how different variables should interact. This is of particular value in domains where interactions are exceedingly complex and unknown such as in non-deterministic games and health sciences (for example). It is argued here that crime linkage – a subset of human behaviour research – is also an example of a sufficiently complex domain space. This is supported empirically within the data made available for study which contains over 400 specific behaviour domains – far too many to evaluate and specify relationships for by-hand.

Thus the present study seeks to address these limitations while also providing cross validation of past research via the use of machine learning algorithms. If the relationships observed in study 1 are stable and representative of true behavioural relationships within linked crime, then one would expect similar results even when employing a different modelling technique. Ideally the modelling technique used would address the outlined limitations while also contributing additional information to understanding the underlying relationships of the variables of interest. Of the various machine learning algorithms available, random forests are often described as the ‘simplest’ and most robust in the face of noisy data thus being a safe place to start (Boulesteix, Janitza, Kruppa, & König, 2012).

4.7.1 Random Forests

Random forest models are a form of machine learning, which itself is an applied form of statistics and computing. Notionally, random forests build many individually weak ‘decisions trees’ that can be used for prediction. These weak predictive models are used via committee vote to make predictions on data (Deming & Morgan, 1993). Within machine learning, this approach has been termed as “ensemble” learning, stemming from the use of a collection of many models to make predictions as opposed to just one. The primary motivation behind ensemble approaches is that by aggregating predictions from many un-correlated sources, the likelihood of making an accurate prediction increases.

Consider the following example: say there exist 5 different models; model one is correct 60% of the time, model two 45% of the time, and three through five are all correct 25% of the time. None of these five models are terribly accurate on their own, but they all capture some amount of pertinent information about the outcome of interest. Now assume that all five models agree on a positive prediction for a given observation. In such a case, there would be a 91% chance that the prediction is correct: $1 - (1 - .6)(1 - .45)(1 - .25)(1 - .25)(1 - .25) = 0.91$. Or put another way: in this example there is only a 9% chance that all five models are simultaneously incorrect. Importantly this example is only true if all five example models are completely uncorrelated which is an unrealistic assumption. Regardless, this example illustrates the principle underlying ensemble methods: one can achieve more accurate results in aggregate than using any single model.

The random forest approach uses many de-correlated decision trees to construct the underlying ensemble of classification models. Classification trees have seen some use within the crime linkage literature in the past, but random forests as suggested here have not (Tonkin, 2012). Notionally, classification trees are straight forward in their creation and are very similar to the clustering technique described in section 3.4.1. The data begins at the "root" from which a variable is selected to split the data. This decision threshold is selected such that it maximizes the separation between the classes or outcomes of interest. This splitting process is repeated until either the classes (linked versus unlinked this case) are fully separated or the tree reaches a maximum "depth" – which is the number of splitting decisions within a given tree. Each distinct decision pathway is termed a "branch" with terminal nodes being designated as "leafs". Figure 38 provides a visual example of one depth two classification tree. Specifying the number of trees to "grow" – i.e. models to fit – and their maximal depth are important considerations when creating random forest models.

Random forest models have a number of useful properties advantageous to the goals at hand: first they are robust in the face many noise variables (Deming & Morgan, 1993). This is important as the signal to noise ratio of the hundreds of behavioural variables is not known in advance. Second, variable interactions do not have to be pre-specified. Subsequent decisions within a branch of a given decision tree simulate variable interactions. Given a sufficient number of trees of a sufficient depth, one can be reasonably confident that all meaningful variable interactions would be captured. Finally, and most importantly, the data splitting process provides a useful metric to assess the importance of any given variable in making classification decisions. For example, if one fit an ensemble of 100 decision trees of depth 10, then there are 1000 possible decision points. By keeping track of the number of times a given variable is used to make a decision for each of the decision nodes, variables can be scored according to their importance.

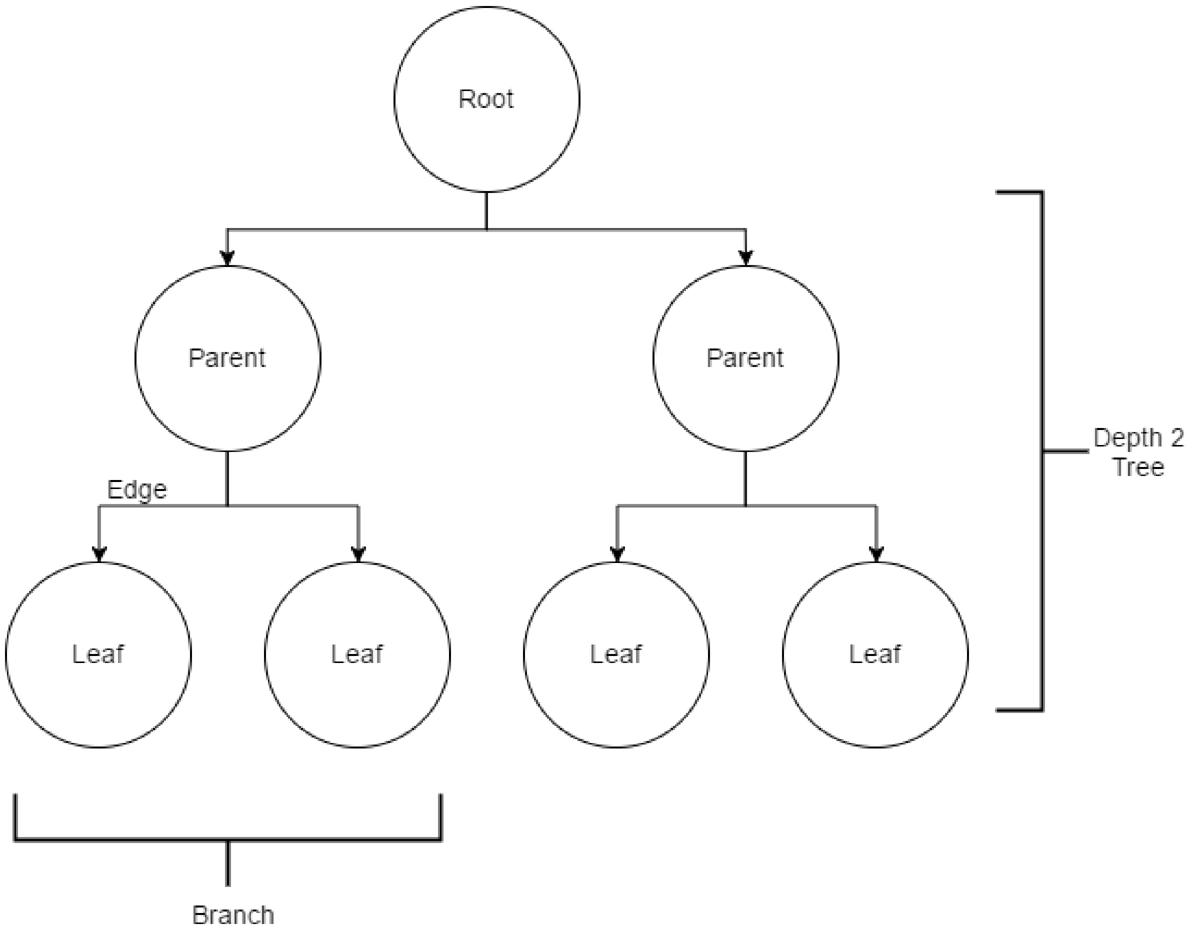


Figure 38: Depth 2 Classification Tree Example

Thus the present study has three primary objectives:

1. Replicate the linking methodology presented in study 1 with a different modelling strategy
2. Determine the relative importance of included variables in determining linkage outcomes

4.7.2 *Method*

The three training data sets constructed in study 1 were re-used in the present study with only one difference: the behavioural M.O. fields were added back into the data set. Similarly to study 1, two different random forest models were created to compare the relative impact of the various predictors. In this case, the interest was on the specific M.O. behaviours, as it has been suggested that behavioural information may be of use in linkage tasks.

The first model acted as the baseline model which included only the same predictors as used in the best performing model from study 1: crime to crime distance, days elapsed, linked crime pair probability, same location type, minimum centroid to crime distance and total activity space inclusion. This also allows for direct comparison between the logistic regression models from study 1 and the random forest models from the present study. The second random forest model included some 400 behavioural variables (roughly 200 new variables for each crime in each crime pair), with each behavioural variable corresponding to the presence of a single distinct action: for example possession of a firearm. If these specific behavioural actions correspond to an appreciable pattern between linked and unlinked crime, then one would expect the more complex behavioural model to outperform the simpler model.

The two random forest models were created using H2o, an open source machine learning platform, and the R H2o interface package (The H2o.ai team, 2017). For both models, the maximum depth of any given tree was capped at 300, and the number of tree models to create was capped at 200. To reduce the potential for over fitting, the minimum number of observations allowed in a leaf was set to 10. When trees are “grown” in random forest models, variables are selected randomly – with replacement – at each decision node. This ensures that the same variables are not constantly selected. The sampling procedure used was the default suggested value by H2o: namely at each decision node $\frac{p}{3}$ variables were selected, were p is the total number of predictor variables. Example code for model specification can be seen in .

Model performance was evaluated similarly to study 1: for each model the ROC curve was created for both the training and test data sets. Binary classification was conducted on all three data sets similar to study 1. Decision thresholds were calculated using the F1 score as opposed to Youden's index, however. This was due to the specific implementation of the H2o package within R. The F1 score is the harmonic mean of precision (positive predictive value) and recall (sensitivity within psychology) (Powers, 2011). The F1 score is given by:

$$F1 = 2 * \frac{precision * recall}{precision + recall}$$

The factor of two ensures that the F1 value takes a score of 1 when both precision and recall are both also 1, and the final F1 score has a range of 0 to 1. Thus the classification threshold used is the threshold that maximizes the F1 score. Binary classification results using the F1 score were recorded for all three training sets.

Finally, the relative importance of the various variables was assessed. Previously this process was described as assessing the frequency with which a given variable was selected for non-terminal decision nodes. However this was a simplification; the specific variable importance algorithm employed by H2o can be found in equation 45 from (Friedman, 1999). It is reproduced here for convenience: for a collection of decision trees $\{T_m\}_1^M$:

$$\hat{I}_j^2 = \frac{1}{M}\sum_{m=1}^{M} \hat{I}_j^2(T_m)$$

Where $\hat{I}_j^2$ is the improvement in squared error associated with a given variable over M decision trees. In other words, the variable importance provides an estimate on the relative gain, as shown in the reduction in squared error, of the random forest model for each included predictor. Large differences in importance magnitude would imply some variables influence the predicted outcome more than others and thus are of greater predictive value. Given the results from the first study of this chapter, as well as discussions in past research, it is highly likely that the various spatial measures will outperform other measures here.

4.7.3 Results

Two random forest models were fit to the training data. The first of these two models corresponded to the "simple" model which only included the same predictors as those outlined in the previous study. The second "full" model incorporated further information in the form of over 200 binary coded behavioural variables. For both models, the trees were grown to maximum depth (300) with the minimum number of observations per terminal node set to 10. Distribution metrics surrounding the tree depth as well as the number of terminal nodes (leaves) is provided in Table 11. Table 12 provides the r^2, AUC and MSE (mean squared error) for both random forest models on both the training and validation data sets.

The results indicate that while the full model achieves higher apparent performance on the training set, the simpler model outperforms the full model on the test data set. This is apparent in all three measures. Due to the manner in which random forest models are "grown", it is not unusual for performance metrics on the training data to be vastly better than those on validation data. The more complex model failing to outperform the simpler model, however, indicates that the behavioural data used in the full model is either: not predictive of linked crime or that it is improperly specified. If the behavioural data are in effect acting as noise, then it would be expected that the full model would perform worse as it would invariable have to choose from a pool of only noisy variables to grow some subset of the predictive trees.

Table 11: Random Forest Model Descriptions

Model	# Trees	Min Depth	Max Depth	Mean Depth	Min Leaves	Max Leaves	Mean Leaves
Simple	200	23	35	27.99	1230	1363	1307.53
Full	200	33	58	42.9	778	1000	889.7

Table 12: Random Forest Model Metrics on Training vs Test Data

Model	Data	r^2	AUC	MSE
Simple	Training	0.502	0.916	0.069
	Test	0.367	0.862	0.088
Full	Training	0.515	0.937	0.067
	Test	0.314	0.844	0.095

The relative impact, or importance, of the various predictors upon model performance can be directly observed from the variable importance plot(s). Figure 39 and Figure 40 show the relative importance of the top 10 (in the case of the simple model all predictors are shown) predictors. Variable importance is a relative measure which compares the gains associated with choosing a given variable over the others. Results indicate that for both models, the proximity measures (minimum centroid to crime distance, and inter-crime distance) were the most often selected predicted variables which provided the most gains. The full model only saw one MO behaviour land in the top 10 predictors, but at a level that was so low it is unlikely to be a significant predictor.

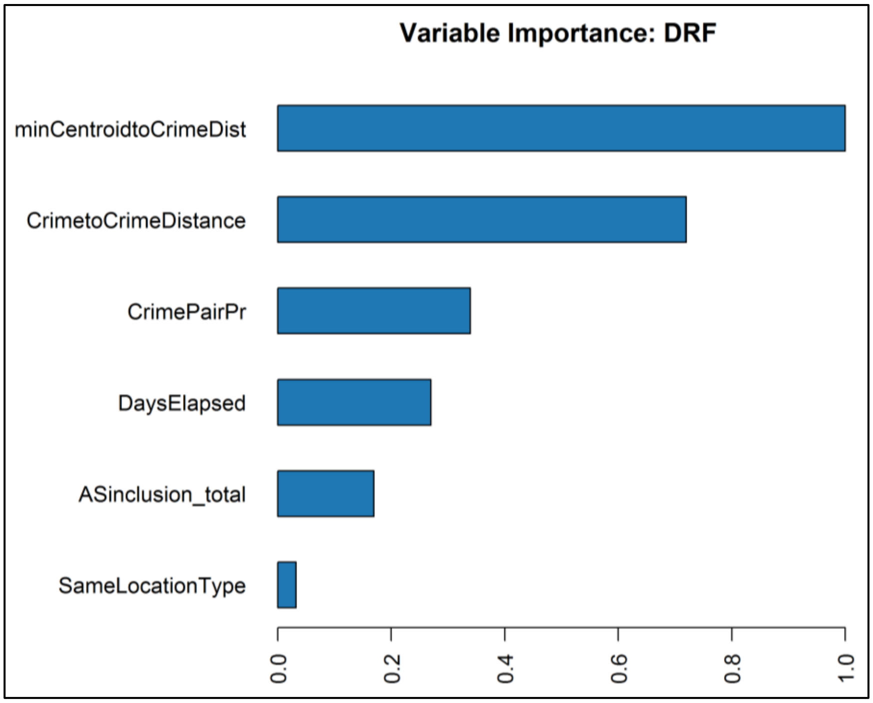


Figure 39: Simple Random Forest Variable Importance Plot

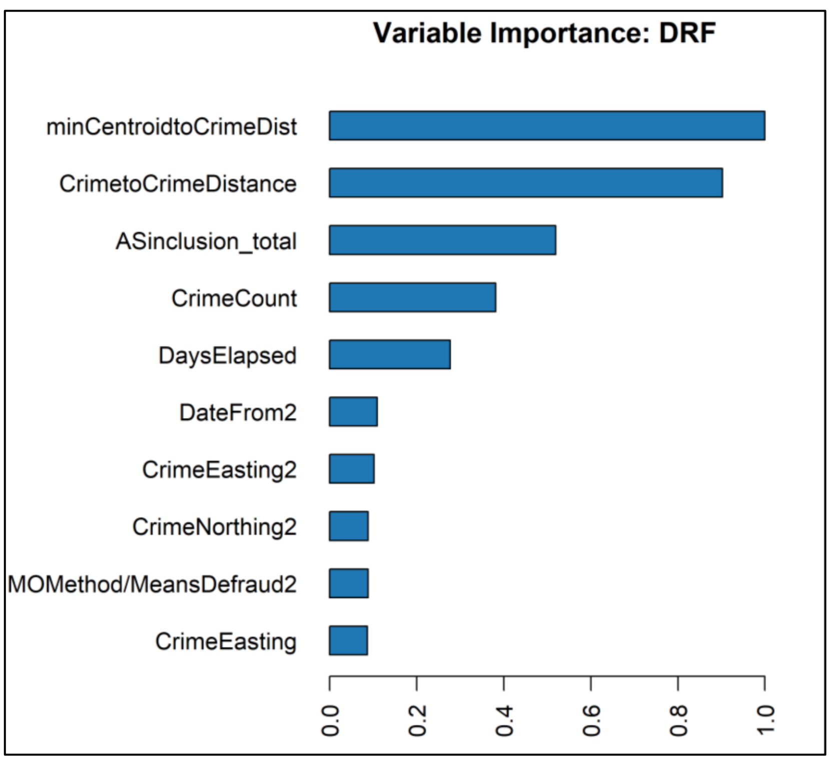


Figure 40: Full Random Forest Variable Importance Plot

The above figures largely agree with the results from the previous study using logistic regression. For the current data and research design, spatial measures far outstrip other predictors in terms of importance. Importantly, centroid-to-crime distance again outperforms straight crime-to-crime distance; a result that is not surprising given the results of the previous study. Interestingly, the relative importance of the activity space inclusion variable increased on the 'full' random forest model surpassing even temporal proximity. Crime eastings and northings also crept into the top 10 predictors for the full model which suggests that the spatial relationships between crimes could be further enhanced by incorporating specific location information. So far the spatial relationships between crimes have been represented almost entirely abstractly. The linked status between crimes has been dependent on how close two crimes are to each other regardless of where in physical space those crimes actually occurred. The simpler models introduced here and in the previous study used a 'same location type' variable in an attempt to provide some locational context for the crime relationships, but based on the results shown here it is possible that the abstract representation is overall too simple. This is left as an area of investigation for further study.

Finally, Table 13 provides the same binary classification metrics as reported in study 1. These metrics were calculated using the resulting confusion table from applying the decision threshold that maximized the F1 statistic. The binary classification results indicate a similar pattern to that indicated by Table 12: namely the Full model has better apparent performance on the training data, but worse overall performance on the test. As in study 1, the precision and MCC values both indicate that as the sample size grows, both models' performance drops to negligible levels.

Table 13: Random Forest Binary Classification Performance Metrics

Data Set	Model	Sensitivity	Specificity	Precision	Accuracy	MCC
Train	Simple	0.807	0.934	0.711	0.913	0.706
	Full	0.783	0.950	0.756	0.922	0.723
Test	Simple	0.594	0.926	0.618	0.871	0.529
	Full	0.556	0.937	0.637	0.873	0.520
Validation	Simple	0.594	0.926	0.008	0.925	0.063
	Full	0.556	0.935	0.008	0.934	0.063

4.7.4 Discussion

The present study sought to assess to what degree the impact of activity space based measures has on the linked versus unlinked decision on a crime linkage task via an as of yet untested modelling strategy employing random forests. Study 1 presented a similar research question, but instead chose to employ the commonly used logistic regression modelling strategy, and while several improvements over conventional approaches were discussed, there yet remained a question as to whether specific behavioural information would be more useful than general spatial trends in making linked status decisions. It was suggested that the apparent strength of the spatial predictors was potentially a result of the lack of more predictive information, such as behaviours. Thus the present study presented two distinct random forest models: the first a random forest model represented the "standard" approach, employing spatial and temporal predictors only. The second model built upon the first by including a large collection of specific behavioural information.

The resulting performance of the two presented models on a validation set indicated that while both models seemed to perform well on the training data, there was a considerable difference in predictive ability on the test set. The model that incorporated specific behavioural information performed less well than its simpler alternative. This result appears counter-intuitive as one would expect models with access to larger amounts of relevant information to generally perform better. However, if a large percentage of the behavioural variables were “noisy” – that is having little association with the linked status of crime, then it would be both possible and expected for the larger model to perform poorly.

Random forest models are grown iteratively such that at each node, or decision point, a small sub-selection of variables is pulled from all available predictors. While this ensures that the same variables are not selected over and over, if a large proportion of the available predictors are noise, then the likelihood that only noisy variables are available for selection increases. Thus it is possible the behavioural domains provided no useful information, which is partially reflected within the tree size metrics. While the full model on average created larger trees – as indicated by the larger average depth, it consistently produced fewer distinct groups, as indicated by the smaller average number of terminal nodes (# of leafs). This outcome makes sense if one assumes the model has to effectively “waste time” evaluating useless predictors; a problem the simple model did not share. Thus it would seem reasonable to conclude that given the current methodology, the binary coded behaviour variables did not contribute meaningfully to the efficacy of linked status prediction.

Despite contending with a large number of noisy variables, the full model did agree with the simple model in terms of the observed variable importance rankings. These rankings indicated that the specific propinquity measures were most effective at identifying linked crime patterns given the available data. This result is in agreement with study 1 and serves to validate the apparent efficacy of activity space based measures. Furthermore, the simple model indicated similar non-linear interactions within the continuous variables as demonstrated in study 1, as demonstrated by the larger average depth of the trees employed than in the full model. While it is not possible to manually inspect each tree's full formulation within the full random forest model, it is reasonable to conclude that the majority of decision points within the simple model were created from splitting the continuous variables as otherwise such large tree depths would not be necessary. This observation implies a certain degree of validity to the splining approach from study 1.

Finally, the random forest models presented here generally indicated a level of performance beyond what was achieved with the logistic regression models in study 1. This is most clearly demonstrated via the classification metrics. Both random forest models achieved MCC rates higher than the best performing logistic regression model from study 1. In the case of the simple random forest model, this implies that the random forest modelling strategy is a more powerful test of linked crime status. Unfortunately, as was the case in study 1, the precision and MCC rates fell to negligible levels as the sample size increased. A true direct comparison of the two modelling techniques on strict binary classification is not possible however, as the decision thresholds were set using two different criterion (Youden's Index for logistic regression, F1 for random forests). Thus while it would appear random forests may be a more powerful modelling strategy, this interpretation is only preliminary and further study and replication is warranted.

4.8 Conclusion

The present study explored the utility of the activity space construct as it has thus far been defined in an example application of crime linkage analysis. Crime linkage analysis was chosen as there was ample empirical evidence to suggest that offenders were following general offending patterns consistent with activity spaces, such as their crimes' propinquity with each other. Fundamentally, it was argued that if offenders followed a consistent targeting strategy then their spatial behaviour could be meaningfully modelled into an activity space. Furthermore, it was noted that past studies of crime linkage analysis generally ignored individual morphology patterns. This omission was directly assessed by incorporating activity space based measures, which in turn resulted in more accurate crime linkage models and thus better performance on crime to crime linkage tasks.

The first study of this chapter indicated that the use of activity space could meaningfully strengthen traditional crime linkage models and deliver more accurate results. The second study of this chapter built upon this observation by assessing whether the strength of the spatial predictors was a result of their relative merit, or merely the consequence of being the only available predictors. Furthermore, the second study sought to assess whether a different modelling strategy could improve upon the commonly used logistic regression approach. While the specific behaviour information was not found to be useful, the random forest models validated the efficacy of the activity space based measures. The random forest models were also demonstrated to potentially be more powerful in their general performance compared to their logistic regression alternative. Finally, as both models' performance was similar to each other as well as both models treating the various predictors similarly in terms of importance, the predictive effects can be understood as well behaved.

Despite these findings, both the first and second studies' results of this chapterhighlight a fundamental flaw within crime linkage research which makes it difficult to assess the true gain of any of the proposed methodological improvements. Namely, crime linkage studies focus primarily upon the identification of linked crimes specifically and ignore the individual. This is a result of the commonly employed validation procedure of evaluating the predictive performance of the linked versus unlinked class of specific crime pairs. While this formulation made sense as an initial starting point, it seems curious that fifteen years after the introduction of the general approach by Bennell and Canter that no significant improvements have been made.

The results from the binary classification on the full validation data shows that regardless of the predictors employed or the predictive model constructed, identifying specific linked crime pairs in a dragnet fashion results in outputs where researchers and practitioners cannot have any confidence in positive predictions. While the specific goal of this chapter was to assess the contribution activity space based measure could make within a tradition crime linkage research design, the flaws observed here when using past research methods makes it difficult to make any specific general conclusions.

Activity spaces are ideographic measures that evaluate a specific individual's spatial involvement irrespective of any other individual. As such, activity space measures are tailored to evaluating specific individuals. Crime linkage analysis, on the other hand, has been primarily focused on describing aggregate patterns of serial offending *in general*. While this body of research has outlined a number of important observations, the foremost of which being the continued support for the propinquity of crime, it remains a method geared towards evaluating *crimes* and not *offenders*. Thus in order to more appropriately assess the practical utility of activity spaces, the example use case should focus on individuals rather than specific crimes.

Recall that one of the fundamental assumptions made within crime linkage research is that offenders are consistent and distinctive. The natural conclusion from this proposition is that crimes within an offender's series should show higher levels of consistency than with crimes from other offender's series. Indeed this is the very hypothesis that is directly tested within crime linkage research. From this argument, however, arises a natural extension whereby the results from a predictive model about the linked status of individual crimes could then be aggregated by offender. If the output from the predictive model is an assessment of the general probability of any two crimes being linked, then by aggregating by offender one would expect higher predicted probabilities on average for an offender's entire series who committed the crime than for series of other offenders. Exploring this possibility is the focus of the next chapter.

Chapter 5

Activity Space, Crime Linkage and Offender Prioritization

5.1 Introduction

Chapter 4 introduced crime linkage analysis and outlined how including criminal activity spaces into such models could provide more accurate results over models lacking such information. While this alone constitutes an advancement in the use of behavioural geography in crime linkage tasks, as a whole the crime linkage literature fails to articulate just how such processes are intended to be of use to practitioners. Commonly used measures such as the AUC, or even straight classification tables, may provide an indication as to the apparent level of performance of the predictive model in general, but they fail to capture the pertinent information of interest to practitioners: namely how often "does it work?"

To illustrate the above, how would a practitioner operationalize predictive crime linkage models as they are typically presented within the literature? Common practice is to now report the AUC as a general indication of model performance, but this only provides a general indication that higher scores are 'better'. If practitioners required a certain degree of performance - for example requiring that some percentage of true linked crimes should appear in the top n records of a ranked list – the AUC cannot be used to answer such a question. The crime linkage research has presented the problem as one of classification of crime-pairs (see Chapter 4 for more details), however there has been no systematic review of whether this formulation translates into efficiency or clearance rate gains in practice. Furthermore, there exists the question as to whether criminal activity space, as described thus far in this thesis, can provide any tangible benefit to practitioners. Chapter 4 demonstrated that activity space information meaningfully contributed to model performance given established crime linkage methodologies, but as was outlined there are a number of problems with the binary classification approach that makes it difficult to fairly assess what practical contribution activity spaces – or even linkage models in general – can make to practitioners. Given that crime linkage analysis has been the primary lens through which this thesis has evaluated the predictive utility of criminal activity spaces, exploring the subtle issues surrounding established performance metrics in that sub-field is not only of interest but is necessary.

This chapter focuses on addressing research question set 3: can an offender prioritization system be developed such that likely responsible offenders for a given crime can be identified based on their prior criminal history? This chapter first introduces the false-positive paradox; a statistical paradox that further motivates a change in perspective in regards to crime linkage analysis. From there a candidate prioritization approach is presented and subsequently tested. As there has been little formal research into this area, the method presented here should be considered preliminary. As such an exploration into a number of free parameters – geographic bias, model bias, and aggregation bias – are explored in detail. The ultimate goal is to present a 'baseline' process which may be further validated and/or challenged by future works.

5.2 The False Positive Paradox

Any attempt to operationalize crime linkage methods such that they could be directly invoked by practitioners must first contend with what has been named the "False Positive Paradox". The false positive paradox is the unfortunate, and often counter-intuitive result of attempting to test for overly rare occurrences within a population (Rheinfurth & Howell, 1998). To illustrate what the false positive paradox is, consider the following example: assume there exists a crime linkage model that is 99% specific and 99% sensitive. This would mean that for a given set of crime pairs, 99% of the true linked pairs would be identified as linked, and 99% of the crime pairs that are not linked would be correctly identified as unlinked. At face value, this would be an exceptional test for crime linkage. Consider a sample of 10000 crime pairs, where the base-rate for linked crimes is 1%; there would be 100 linked crime pairs, and 9900 unlinked crime pairs. Now if the previously described linkage test was applied to this sample, the test would identify 99 of the 100 linked cases correctly, and 9801 of the 9900 unlinked cases correctly; however this results in 99 false positives (9900-9801) which is the same number as true positives. Thus in this example, given a 99% sensitive and specific test, a crime pair that was predicted to be linked would only be actually linked 50% of the time. Table 14 summarises these example hits and misses in a confusion table format similar to the confusion table results reported in and Appendix .

Table 14: False Positive Paradox Example 1

	Hit	Miss
True Positive	99	1
True Negative	9801	99

Acknowledgment of the false positive paradox is of particular importance in crime linkage analysis due in part to how such studies have typically constructed their samples. The growth rate for a given set of n crimes into crime pairs is given by the total number of unique combinations:

$$\frac{n^2 - n}{2}$$

Where *n* is the number of individual crimes within the dataset. This is equivalent to an *n Choose 2* combination problem and results in an exponential growth in the number of crime pairs that are generated from a given starting sample of crimes. From these, only an exceptionally small number of pairs will be linked pairs. The exact number of which can be expressed as the sum of all linked pairs committed for all offenders within the sample. Thus for i offenders:

$$\sum_{k=1}^{i} \frac{x_k^2 - x_k}{2}$$

Where x_k is the number of crimes (x) committed by offender *k* within the sample, and i is the number of distinct offenders represented. These two equations let us approximate the base rate of linked crimes within a sample of n crimes by taking the ratio:

$$\frac{Linked\ Crime\ Pairs}{Total\ Crime\ Pairs} = \frac{\sum_{n=1}^{i} \frac{x_n^2 - x_n}{2}}{\frac{n^2 - n}{2}}$$

From the present data sample used throughout this thesis of 3,449 crimes committed by 997 distinct offenders, this results in a base rate of linked crime of approximately 0.0009287%, rounded to 0.001%.

Applying this base rate to the earlier example of 10,000 crime pairs yields the following distribution: 10 true linked crime pairs, 9990 unlinked pairs. Of these, 9 of the true linked crime pairs and 9,890 unlinked crime pairs would be correctly classified; that leaves 1 false negative and 100 false positives. Thus in this example the number of false positives is 10x as many as the true positives, and this discrepancy is only exacerbated by increasing the number of pairs for consideration despite the apparent strength of the test. Table 15 summarises the example classification distribution.

Table 15: False Positive Paradox Example 2

	Hit	Miss
True Positive	9	1
True Negative	9890	100

This serves to outline the extreme difficulty faced in any diagnostic or predictive task oriented around rare occurrences. Crime linkage tasks are no exception. This exercise also serves to illustrate the weakness with relying solely on the AUC, described in Chapter 4 to assess predictive ability in a binary classification task. To briefly summarise: the AUC is a rank order measure that describes the ability of a test to rank positive cases higher than negative cases; to wit if the 9 positive cases described above rank higher than the majority of the 9890 negative cases then the test will have a high AUC. This does not, however, show the tremendous number of false positives that the test also generates in a classification context as would be the case in an applied setting. As pointed out by Lee Rainbow (Behavioural Investigative Analyst, UK) (Woodhams & Bennell, 2015, p. 173), there exists a need to clarify what exactly the expected use case any crime linkage research is intended to assist, and to fully consider if and how the research methods used can be used by practitioners, and to readily understand the limitations associated with such methods.

5.3 Operationalized Crime Linkage Analysis

Within crime linkage research, there has been two distinct use cases described: the first is comparative case analysis (CCA), and the second is crime linkage analysis (CLA). Comparative case analysis is the proactive search for similar crimes to a given index crime. Research into CCA is (usually) oriented around identifying patterns from solved crimes, and much of the crime linkage literature has focused on this task (Woodhams & Bennell, 2015). Crime linkage analysis, conversely, is often defined as a much more targeted task which seeks to determine if a given small set of crimes is committed by the same offender. As outlined by Rainbow (Woodhams & Bennell, 2015, p. 175): “In reality, such analysis is performed on a predetermined set of offenses provided by the investigating officer, who has reason to believe they may be linked. Such analysis does not involve the proactive searching of datasets for further offenses, but is restricted to the prescribed cases identified by the investigation.” Figure 41 and Figure 42 show a visual representation of these two processes.

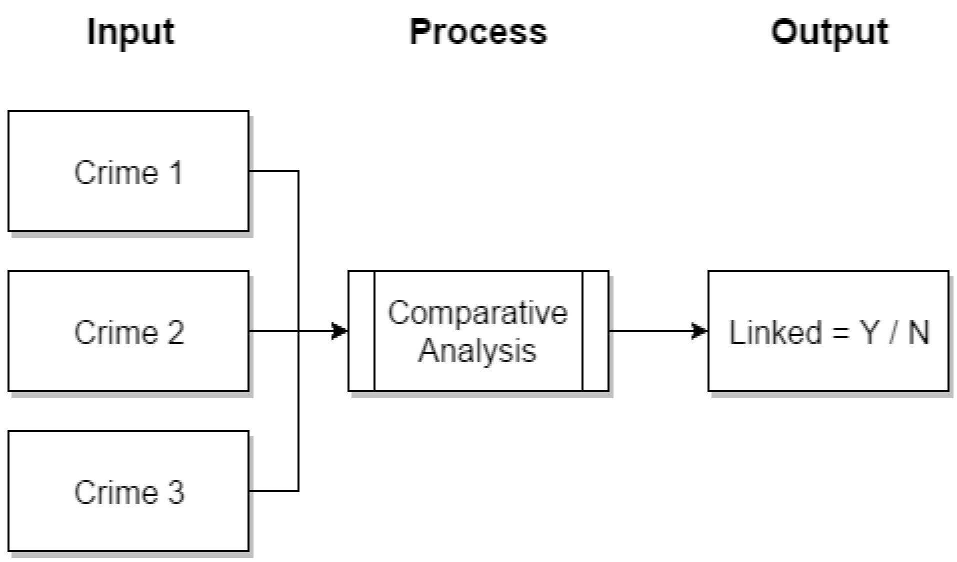


Figure 41: CLA Process Diagram

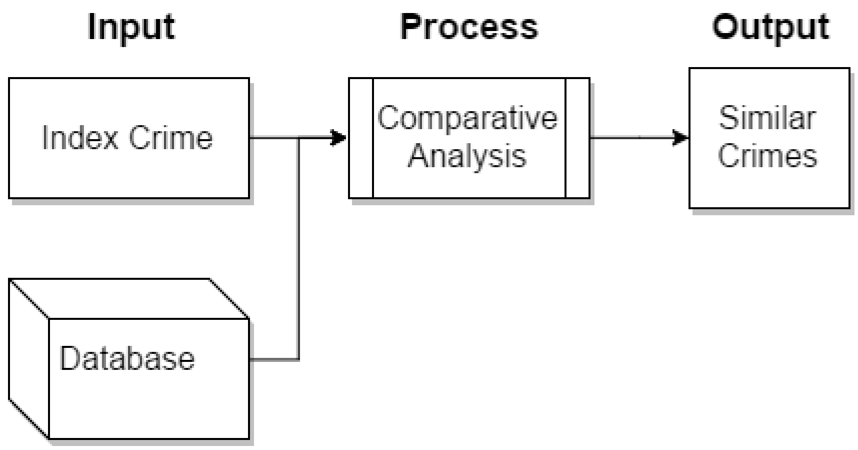


Figure 42: CCA Process Diagram

The specific differences between these two processes are rather subtle. If one assumes the comparative analysis being performed is the same, then the only appreciable difference is in the scope of the inputs. The outputs, while they appear dissimilar are logically connected. Again, if the comparative process is the same, then any crimes that would be pulled from the database in CCA as a "similar" crime would be identified as a "linked" crime in CLA if both that crime and the index crime had been presented as inputs to a CLA task.

Assuming the comparative method is the same may not always be appropriate however. Comparative methods within the linkage literature can be broadly classified as one of two methods: Statistical or Clinical. The linking process described in Chapter 4 is statistical in nature and, as one might expect, involves reducing the linking features into principle codified variables. Conversely, the clinical approach attempts to generate a more holistic picture of the offense, and seeks to build understanding of *why* that particular offence was committed. The obvious trade-off between these two approaches is one of cost; the clinical approach necessitates extensive review of the crime set by experts, such as Behavioural Investigative Analysis (BIAs) in the case of the UK, and thus is inherently limited in how extensive such reviews can be. The algorithmic approach of statistical comparison methods do not share these same limitations, but come at the cost of a loss of “behavioural richness” and detail (Woodhams & Bennell, 2015, p. 180).

Furthermore, in the case of clinical classification, there exists the question surrounding the accuracy of human judges in making such linking decisions. Several studies have investigated this issue and have found that while police professionals are fairly successful in making accurate linkage predictions (Bennell et al., 2010; Santtila, Korpela, & Häkkänen, 2004; M. J. Tonkin, 2012)students who were trained to use heuristics based upon past statistical linkage findings also performed well (Bennell et al., 2010; M. J. Tonkin, 2012), and finally that automated linkage algorithms ultimately performed as well or better than their human counter-parts (Bennell et al., 2010; Santtila et al., 2004; M. J. Tonkin, 2012). While these studies are in no way definitive proof of the efficacy of statistical models compared to human experts, they do imply that the pursuit of an automated system to handle the linkage task is not without merit.

In order to determine the efficacy of derived linkage models, a specific use-case has to be defined. Ideally, such a case would be applicable to the processes of CLA described above. Recall that in Chapter 4, the models were assessed according to their derived AUC – which is a performance metric that summarises how well the model identifies linked crime pairs from unlinked crime pairs. In effect, this – as well as the classification contingency tables (see: and Appendix) – summarises how well the model performs at CCA. As highlighted by section 5.2, the false positive paradox illustrates why this is insufficient for practical applications.

One possible solution to this problem may lie in the observation that CLA is concerned with identifying if two crimes committed by the same offender, and that this process is easily extended to identify which offender committed the index crime by assessing the “likely” crimes that are the resulted output from CCA. Thus by chaining CCA and CLA as they have been described above, a foundation for an offender prioritization system whose output is a ranked list of offenders rather than crimes can be established. This circumnavigates the false-positive paradox classification problem by shifting the focus of the process away from a strict linked / unlinked classification task to one that provides a relativistic assessment of known offenders. Figure 43 shows a possible implementation of such a process.

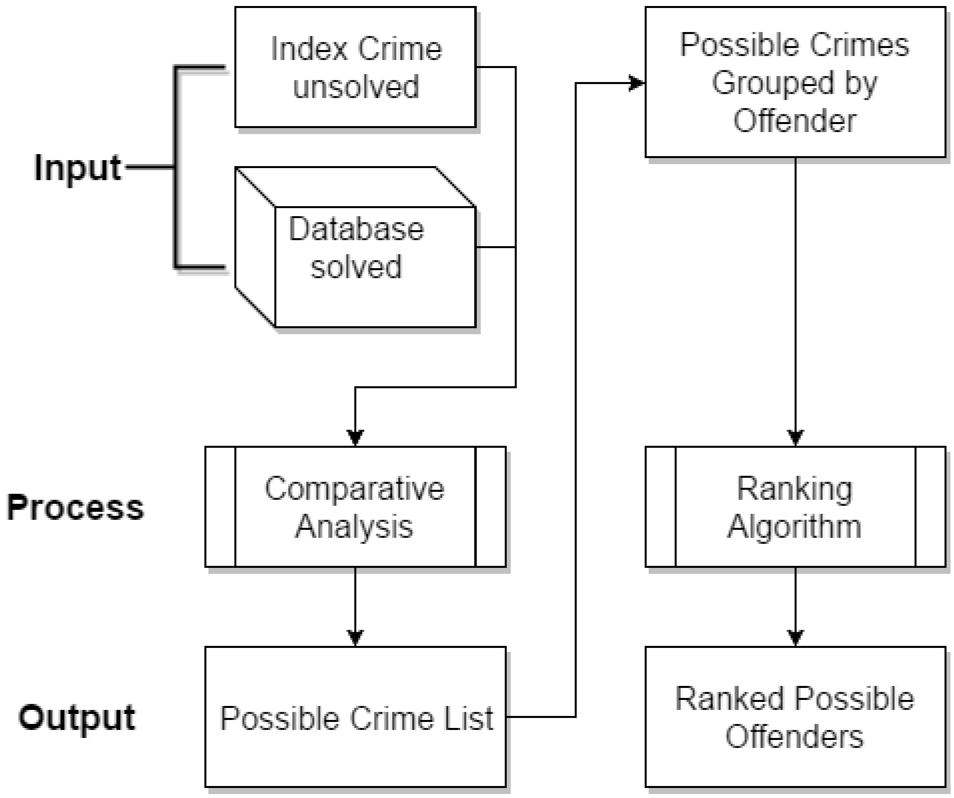


Figure 43: Offender Prioritization Diagram

The proposed process expands upon the "traditional" approach to crime linkage analysis by focusing on possible offenders rather than linked / unlinked judgements on individual crimes. Part of the challenge associated with the false positive paradox was that for any given index crime, the same offender could be successfully linked (or falsely linked) multiple times if they had committed multiple offenses. The proposed method ensures that for a given index crime, any given offender can only be proposed once. This process orientates the question away from the linked status of individual crimes to one that is interested in identifying possible offenders. Furthermore, such an approach may be particularly beneficial for assessing the practical utility of a linkage model for assessing offender's whose criminal history is diverse and not easily assessed by M.O. behavioural similarity as has been the traditional approach in the linkage literature.

5.4 Quantifying Prioritization

Offender prioritization as a process has thus far been a very under-developed area of research. Much like crime linkage research, the few studies that have attempted to study the feasibility of an offender prioritization system have done so via specific crimes only. Snook et al (2006) applied the propinquity argument directly to a set of armed robberies and ranked offenders based on their home location's proximity to the crime in question. Additional filters were applied regarding each offender's criminal history; i.e. whether or not they had committed the same type of crime in the past or not for example. Their results indicated some level of success with 65% of the responsible offenders falling within the top 10% of ranked suspects. Research by Goodwill and Alison (2006) further suggests that spatial measures result in the most reliable rankings of possible suspects. While neither of these studies employed any rigorous modelling, they do provide a foundation from which to build a practitioner focused assessment scheme.

In general, efficacy can be conceptualized by how successfully a given model is able to prioritize the responsible offender from the total pool of possible offenders, as was done by Snook et al. (2006). For example, consider for a given target crime there exists 100 possible offenders within a database. If, after the prioritization process is complete, the actual offender responsible is the 50th most likely offender from the pool of 100, then the process is offering no practical benefit due to the limited number of available investigative man-hours. However, if that responsible offender is the 3rd most likely offender, then it is clear the model is providing objectively useful results. Rather than having to potentially sift through all 100 possible suspects, the algorithms have instead delivered a list of three; a much more reasonable number to investigate given limited resources.

A significant limitation of this approach, however, is that it implicitly assumes that the index crime was committed by an offender that is known to the police. If, for example, the index crime a practitioner would like to generate a list for has been committed by a new offender (not previously known to the police), then any attempt to prioritize other known offenders is a meaningless endeavour. However, given that there is evidence to suggest that the vast majority of crimes are committed by a small collection of highly prolific offenders (Wolfgang, 1973; Wolfgang et al., 1983), this assumption is argued to be a reasonable one.

5.5 Method

5.5.1 Data

The data used in this study is the same input / validation split set constructed and described in Chapter 4. To recap: the data set consisted of 3,449 solved crimes for model fitting and 997 index crimes for validation. These index crimes are the final crime in the series of the 997 unique serial offenders present within the data. Linked predictions will be carried out using the derived models discussed in Chapter 4. Specifically, these models are: 1) No Spline, No Activity Space model; 2) Spline, No Activity Space model; 3) Spline, Activity Space model. The proposed methodology (discussed below) involves applying the discussed models to the full validation set which consists of all combinations between the validation crimes and the crimes used in model fit (n = 997 * 3449 = 5,946,076 crime pairs for analysis).

5.6 Study – Developing an Offender Prioritization Process

The offender prioritization process involved aggregating the linked likelihood output for each model for each crime pair according to their corresponding offender. Specifically, each crime pair was constructed in the form: [Solved Crime A – Index Crime B], where the solved crime was pulled from the police database sample, and the index crime was the example unsolved crime for which offenders were to be prioritized. Each of these pairs was assigned a linked likelihood score according to a derived model from Chapter 4. These pairs were then aggregated according to the offender who committed [Solved Crime A], for the index crime to create a proportional score of the likelihood a given offender committed the given index crime. This proportional score was calculated as the averaged linked likelihood of each crime in a given offender's series to the index crime. Offenders were then ranked, from highest to lowest score, for the given index crime. After all offenders were ranked according to their linked likelihood, the rank of the responsible offender was recorded. This process was repeated for each of the 997 index crimes for each of the three derived models from Chapter 4. Figure 44 illustrates this prioritization process for a given index crime in more detail.

Model performance was evaluated as the proportion of index crimes for which the responsible offender was prioritized into the top five possible offenders. The chosen cut-off of the top five was selected after informal discussion with practitioners who advised the top five as an adequate "starting point" for these analysis. This cut-off acts as a representation of the limited resources for which investigators can devote to any one crime and is thus subject to change depending on circumstance. The proportion of index crimes for which the prioritization process successfully ranked the responsible offender in the top five was compared across all three models for all 997 index crimes.

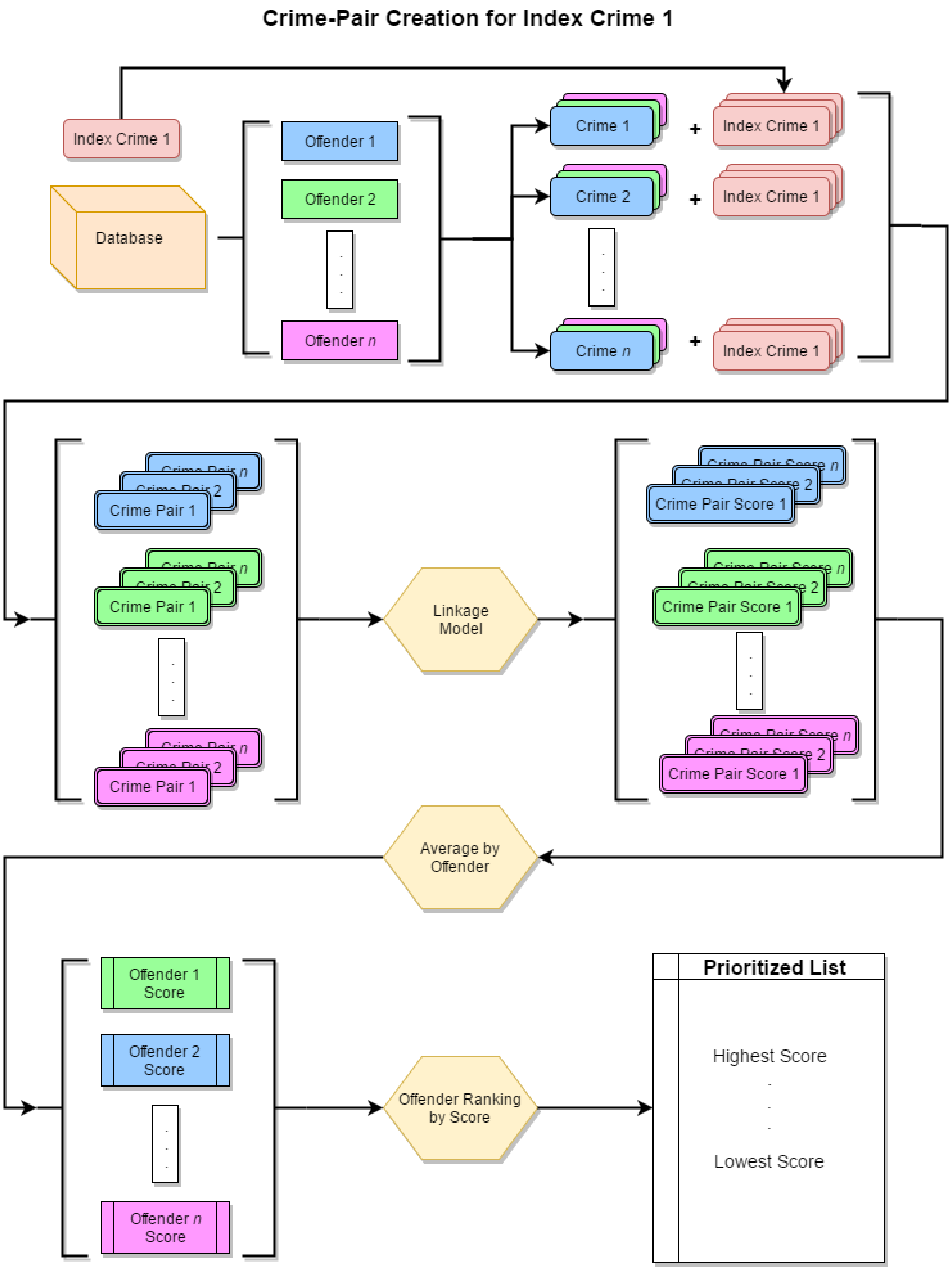


Figure 44: Offender Prioritization Process

5.6.1 Results - Study 1 Offender Prioritization Process

The efficacy of the three logistic regression models discussed in Chapter 4 was assessed utilizing the offender prioritization procedure described in section 5.6 on a sample of 997 index crimes. The rank distribution achieved by the prioritization process for the actual linked offenders for these 997 index crimes is shown in Figure 45, Figure 46, and Figure 47. The distribution charts are heavily skewed towards the lower ranks (rank 1 to 25). This indicates the prioritization process is operating in the intended direction. The higher the frequency of occurrence of the linked offender being assigned a low rank (between 1 and 25 for example) indicates that the model is correctly identifying the actual linked offender for their index crime within the sample.

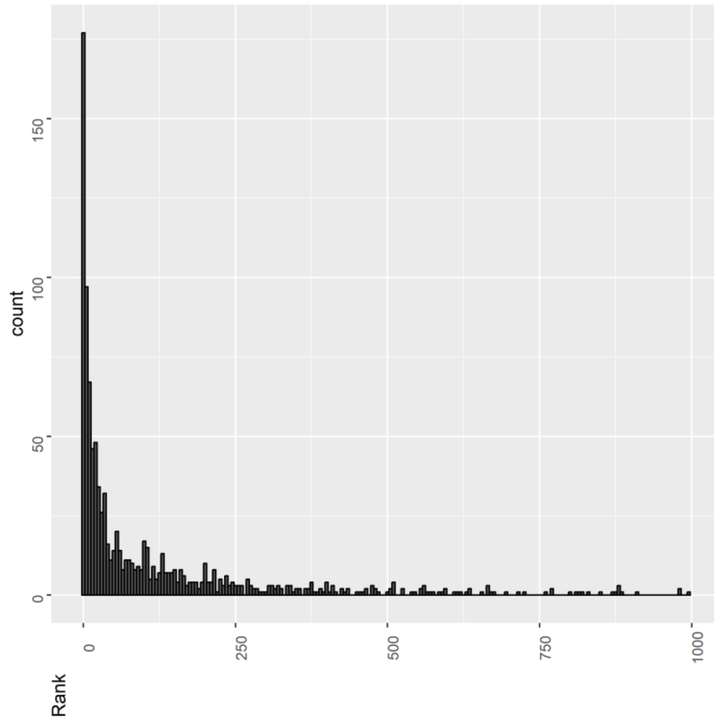


Figure 45: Model One – Rank Distribution of Linked Offenders across 997 Index Crimes (binwidth = 5 Ranks)

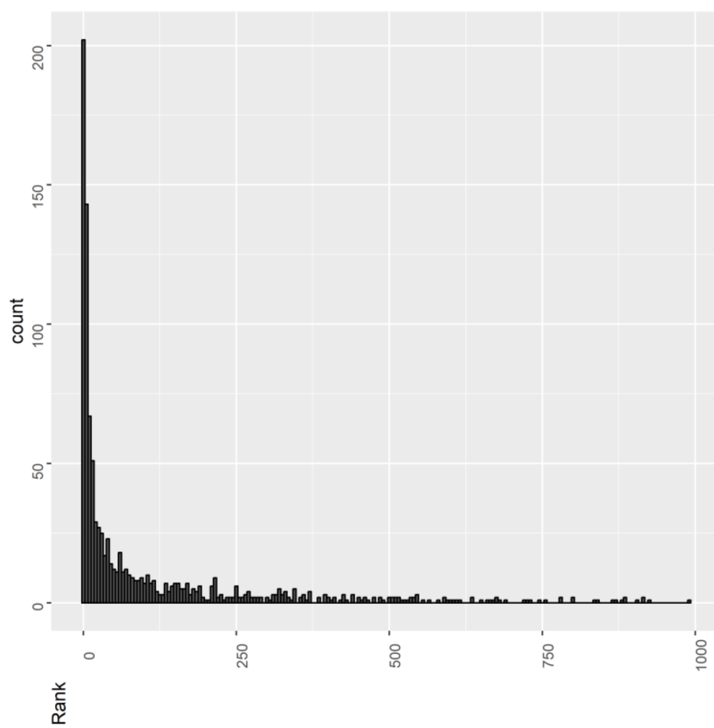


Figure 46: Model Two – Rank Distribution of Linked Offenders across 997 Index Crimes (binwidth = 5 Ranks)

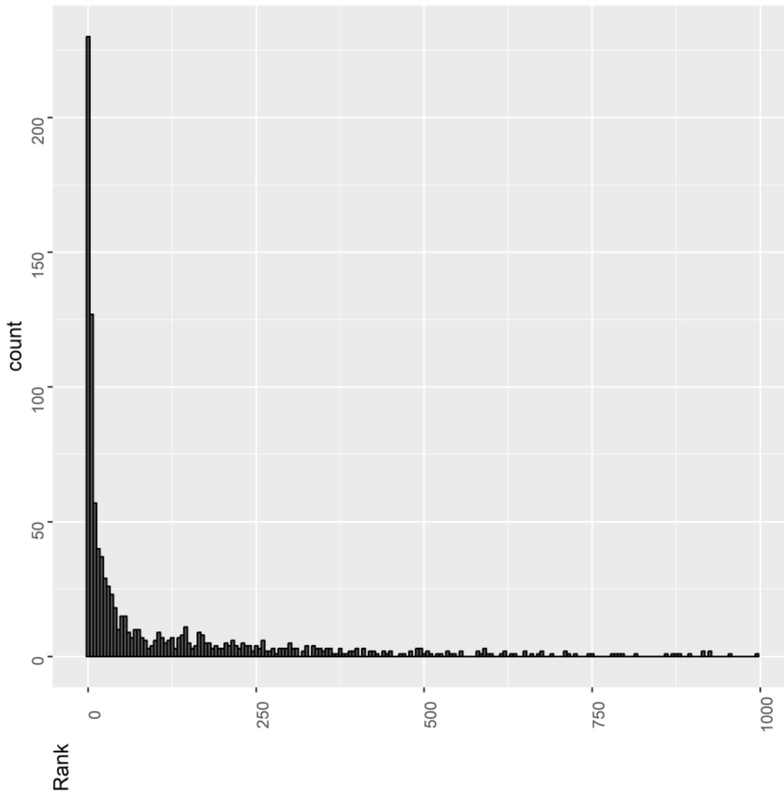


Figure 47: Model Three - Rank Distribution of Linked Offenders across 997 Index Crimes (binwidth = 5 Ranks)

These distribution figures indicate that of the three models, the most “complex” activity space model (model 3) performs the best (most Rank 1 to Rank 25 hits) under the proposed prioritisation scheme. The splined, no activity space model (model 2) performs the next best, and the no spline no activity space model (model 1) performs the worst. However, all three models performed well above chance levels.

Table 16 summarises several dispersion indices surrounding the prioritized rank of linked offenders. Table 16 largely confirms the overall picture provided by the Rank distribution figures: with the exception of the achieved median rank, model 3 performed the best at prioritizing linked offenders for the observed index crime sample.

Table 16: Prioritization Rank Distribution Descriptive Statistics

Model #	Min Rank	1st Quartile	Median Rank	Mean rank	3rd Quartile	Max Rank
1	1	6	33	113.4	139	996
2	1	4	24	108.2	132	991
3	1	3	23	110	147	993

Finally, the distribution surrounding the top five ranks was investigated. Recall the goal was to assess whether such a prioritization process as the one proposed could prioritize offenders into the top five possible suspects. While Table 16 indicates that all three models are able to do this to a certain extent – seen from their 1st quartile Rank, the distribution of only the top 10 ranks was also investigated and these results are summarised in Figure 48, Figure 49, and Figure 50. The distribution pattern remains relatively stable for the top 10 distribution charts when compared with the overall charts above, which implies the ranking algorithm is stable across the entire spectrum of ranks.

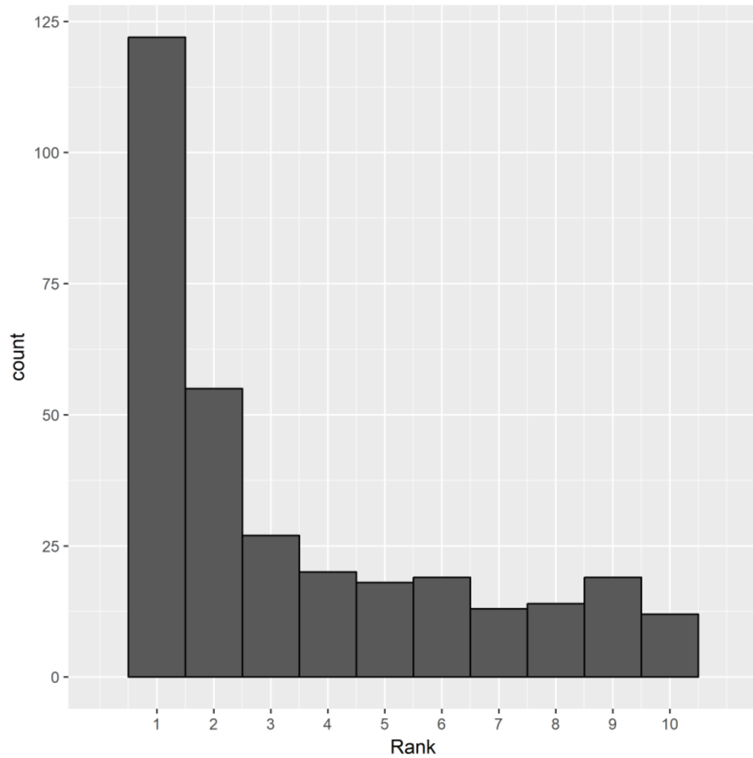


Figure 48: Model One Offender Prioritization Top 10 Rank Distribution

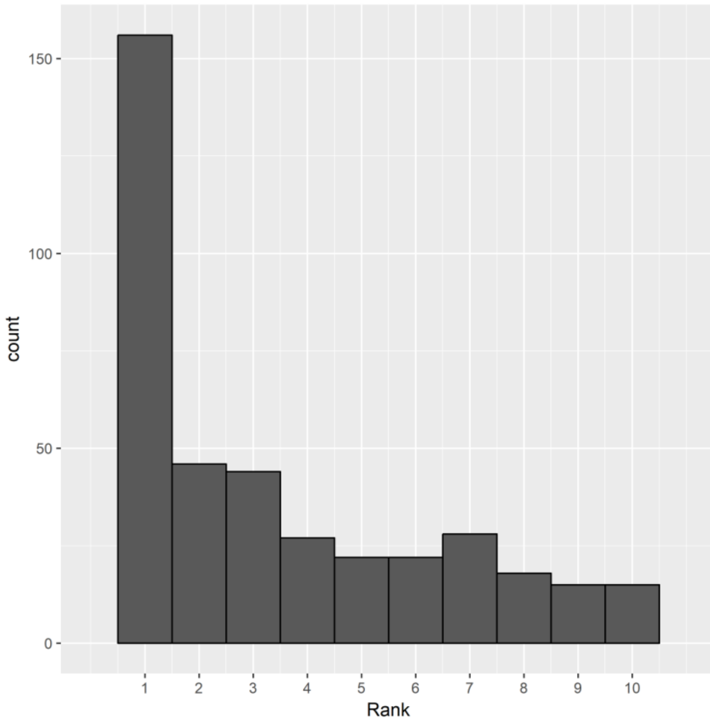


Figure 49: Model Two Offender Prioritization Top 10 Rank Distribution

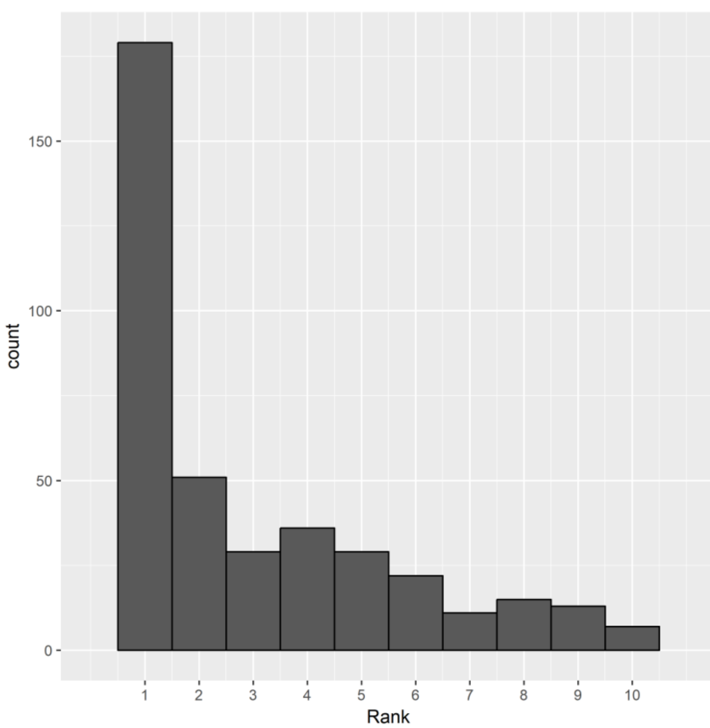


Figure 50: Model Three - Offender Prioritization Top 10 Rank Distribution

Finally, Table 17 summarises the observed and expected chance hit rates for the prioritization process for the observed data. The results indicate that the prioritization process, while far from “perfect”, is able to prioritize offenders well above expected chance levels for all three models. Model 3, however, once again performed “the best” with the highest raw hit rate for classifying linked offenders into the top five of possible suspects from the sample.

Table 17: Offender Prioritization Top 5 Hit Rate across All Three Models for 997 Index Crimes

Model #	Top 5 Count	% Hit of Index Crimes	% Hit by Chance
1	242	24.3%	0.501505%
2	295	29.6%	0.501505%
3	324	32.5%	0.501505%

Note: A “hit” is described as occurring when the actual linked offender was ranked in ranks 1-5 according to the prioritization process for their respective index crime.

These results clearly indicate that the performance gain from introducing spline fits for non-linear terms as well as introducing activity space predictors can be translated into practical real world gains, despite the very modest impact these factors had on the observed AUC performance value from model fitting.

5.7 Study – Validating the Proposed Prioritization Process via Bootstrapping

In order to address generalizability concerns, the prioritization process outlined in section 5.6 was subjected to a bootstrapping validation procedure. A bootstrapping procedure was chosen for its ability to examine the stability of observed relationships within a data sample over a large number of iterations. Bootstrap validation involves simulating many samples from a single starting sample by re-sampling with replacement from the parent sample (Harrell, 2001). For example, if the starting sample was n = 100 cases large, then a bootstrapped sample would be a sample where cases are randomly selected from the original 100 (with replacement) until a new sample of n = 100 cases is created. The entirety of the analysis procedure is then applied to this new bootstrapped sample and performance metrics are recorded. This entire resample, analysis and assessment procedure is iterated many times, after which confidence intervals and mean levels can be directly estimated from aggregate measures over all iterations. If the regression models are detecting non spurious relationships, then their performance on the prioritization task should remain stable over a large number of bootstrap iterations.

The data used is the same input / validation sets employed throughout this thesis. Because the goals is to assess prioritization performance on simulated "unsolved" index crimes, the validation set upon which the final prioritization process is performed was not changed in any way. The bootstrapping procedure was employed to simulate many linkage models using resamples of the input dataset. From the starting 10.9 million crime pairs present within the input dataset, a case-controlled bootstrapping procedure was employed. During each resample iteration, a full set of linked crime pairs was randomly selected (n = 12,626) with replacement, and five times as many randomly selected (with replacement) unlinked pairs was also selected. This resulted in a unique case-controlled bootstrap sample of 75,756 crime pairs for each iteration. This case-control procedure was necessary due to the extremely uneven distribution of the test cases (linked vs unlinked). If the sample was not case controlled, for example, it would be highly unlikely that any of the bootstrapped samples would contain any linked crimes at all. While case-controlling the two classes does ensure a mix of linked to unlinked pairs, it comes at the cost of losing base-rate occurrence information. Because of this limitation, the probability outputs from the linkage models do not correspond to true probability weights and should not be used or interpreted as such. The prioritization process outlined in section 5.6 uses the probability outputs as a proportional score; the higher the score the higher the relative likelihood that given offender committed the given index crime *relative to the other offenders within the sample*.

Next, three linkage models were fit to the bootstrapped dataset using logistic regression. This is the same procedure used in Chapter 4. The specific predictors used are also unchanged. To review, these models were: 1) No Spline, No Activity Space; 2) Spline, No Activity Space; 3) Spline, Activity Space. The specific predictors used in each of these three models is summarised in Table 18. The AUC, predictor coefficients and pseudo-R^2 statistics were recorded for each model after each bootstrap iteration.

Table 18: Linkage Model(s) Predictor Summary

Model Number	Model Name	Predictors
1	No Spline, No Activity Space	• ICD • Days Elapsed • Same Location Type (1 / 0) • Jaccard Index of Crime Types
2	Spline, No Activity Space	• 5 knot splined ICD • 5 knot splined Days Elapsed • Same Location Type (1 / 0) • Crime Pair Base Rate
3	Spline, Activity Space	• 5 knot splined ICD • 5 knot splined Centroid to Crime Distance • 5 knot splined Days Elapsed • Same Location Type (1 / 0) • Crime Pair Base Rate • Activity Space Inclusion (0 - 4)

These specific models were chosen so that 1) the impact of splining non-linear predictors – something that is not yet standard practice within the crime linkage literature – can be assessed and 2) the impact of the inclusion of Activity Space predictors can be assessed. If neither splining the non-linear predictors nor adding an activity space predictor significantly deviates in performance from the "simplest" model, then it can be safely concluded that neither provides any tangible gain in assessing crime linkage. This would also serve as evidence to suggest that offenders do *not* behave in geographically local ways as has been commonly suggested.

For each of the three models, the prioritization process was conducted as described in section 5.6. For each index crime within the validation set, the rank of the responsible offender was recorded. From these, a final sum of the number of offenders who were ranked in the top 5 possible offenders whom actually committed the given index crime was calculated. The entirety of this process was replicated 200 times. More iterations were not technologically feasible. Figure 51 summarises the bootstrapping procedure used here. Analysis was performed utilizing R.

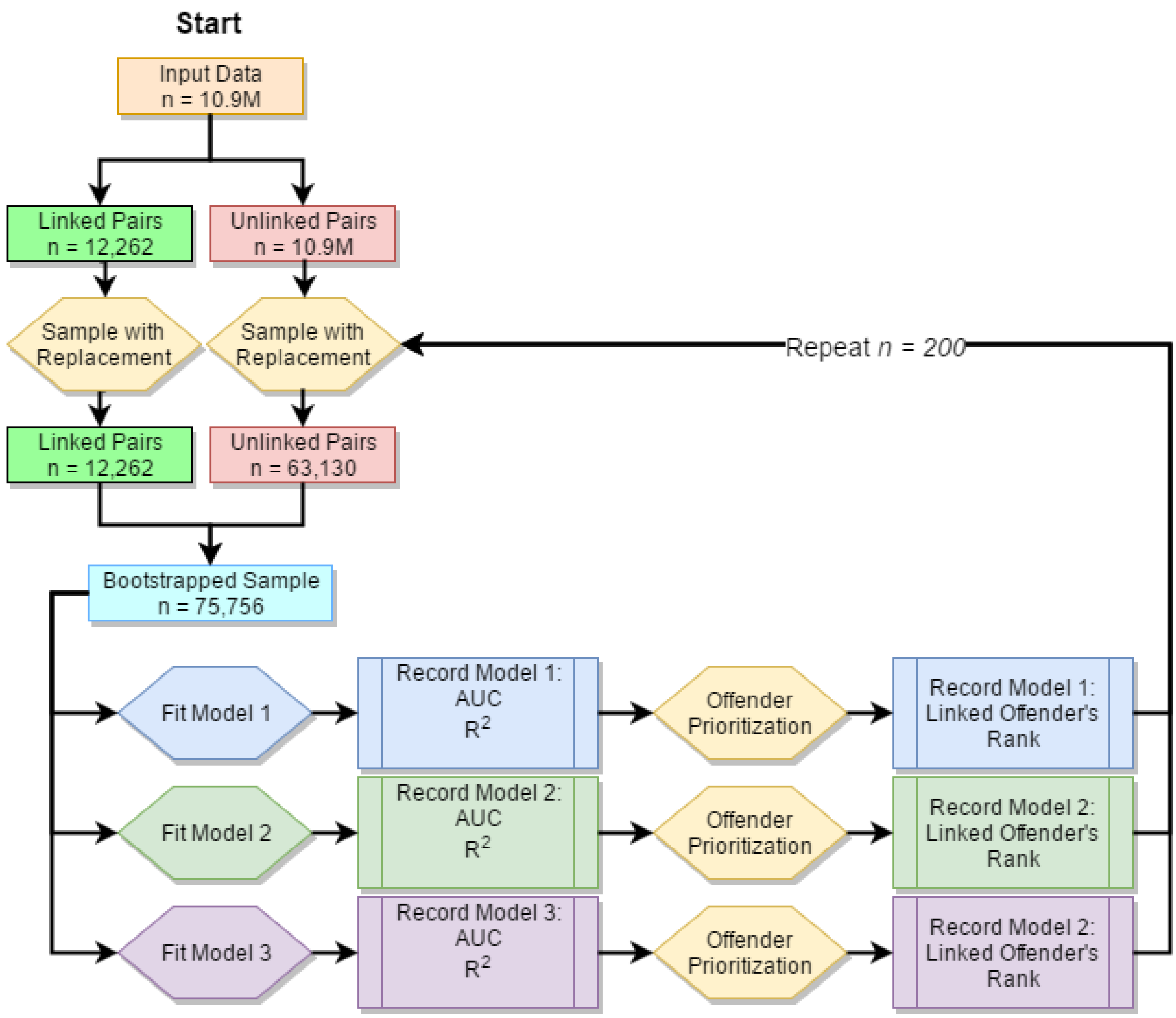


Figure 51: Offender Prioritization Bootstrapping Procedure

Final performance was assessed as the average sum of offenders who were correctly prioritized into the top 5 from a possible pool of 997 offenders, across all 997 index crimes for all 200 bootstrap iterations for all three models. The model with the highest prioritization rate into the top five was considered to be the best performing model. 95% confidence intervals surrounding the average count per rank was calculated for all three models.

Finally, there was a question regarding the aggregation procedure used in the previous study. Recall that for each crime in an offender's series, the prioritization procedure aggregated the scores by taking an average. However, it was not clear whether this was the most effective approach. Thus the current study introduces and compares three procedures: (1) Random sampling, (2) mean, and (3) max. The Random procedure does its name implies: for each offender's series a random crime pair's linked probability was selected to represent the series. The Mean method is equivalent to the procedure outlined in the previous study, and finally the Max method takes the maximum observed linked probability within a series as the value for the entire series. The random method allows for the examination of what would happen if crimes were treated as independent events with no relationship to one another and is analogous to how previous research into crime linkage analysis has been conducted. The other two measures allow for the evaluation of how consistent offenders are across a crime series. If, for example, offenders are generally very stable in how they offend, then one would expect little variation between the three methods. However, if there is substantial variation, then the random method should clearly perform the worst and the Max method may be the 'best' as it would be the most sensitive to novel information as the approach would be able to link offenders to an index crime due to a single large score even if the majority of their crime history is very dissimilar to said index crime.

5.7.1 Results - Study 2 Bootstrapping Validation of Prioritization Method

The bootstrapping procedure described in section 5.6.1 was employed to assess reproducibility of 1) the linkage models described in Chapter 4 and 2) the apparent prioritization performance described in section 5.6. The frequency with which each of the three linkage models (described in section 5.6.1) were able to successfully prioritize a linked offender into the top five of possible suspects for an index crime (n = 997) was recorded. This process was repeated for 200 bootstrap iterations, for each of the three proposed prioritization methods and was recorded. These results can be seen in Figure 52. Table 19 provides summary statistics for hit count of each model across all 200 iterations. Of note, for all three models the ‘Max’ method had the highest mean hit count (M1 mean = 273.1, 95% CI = 272.46 – 273.73; M2 mean = 357.51, 95% CI = 357.07 – 357.94; M3 mean = 362.22, 95% CI = 361.73 – 362.71).

Table 19: Bootstrap Performance Metrics

Model	Method	N	Mean	Std.Dev	Error	Lower CI	Upper CI	2.50%	97.50%
M1:	Max	200	273.10	4.61	0.64	272.46	273.73	265.00	284.03
M1:	Mean	200	239.70	3.13	0.43	239.26	240.13	235.00	246.03
M1:	Random	200	218.17	8.88	1.23	216.94	219.40	199.93	236.05
M2:	Max	200	357.51	3.17	0.44	357.07	357.94	352.00	363.03
M2:	Mean	200	296.40	3.12	0.43	295.96	296.83	289.98	301.03
M2:	Random	200	268.15	9.77	1.35	266.80	269.50	248.98	285.00
M3:	Max	200	362.22	3.57	0.49	361.73	362.71	356.00	371.00
M3:	Mean	200	312.49	3.08	0.43	312.06	312.91	307.00	318.00
M3:	Random	200	284.72	9.19	1.27	283.45	285.99	265.00	301.00

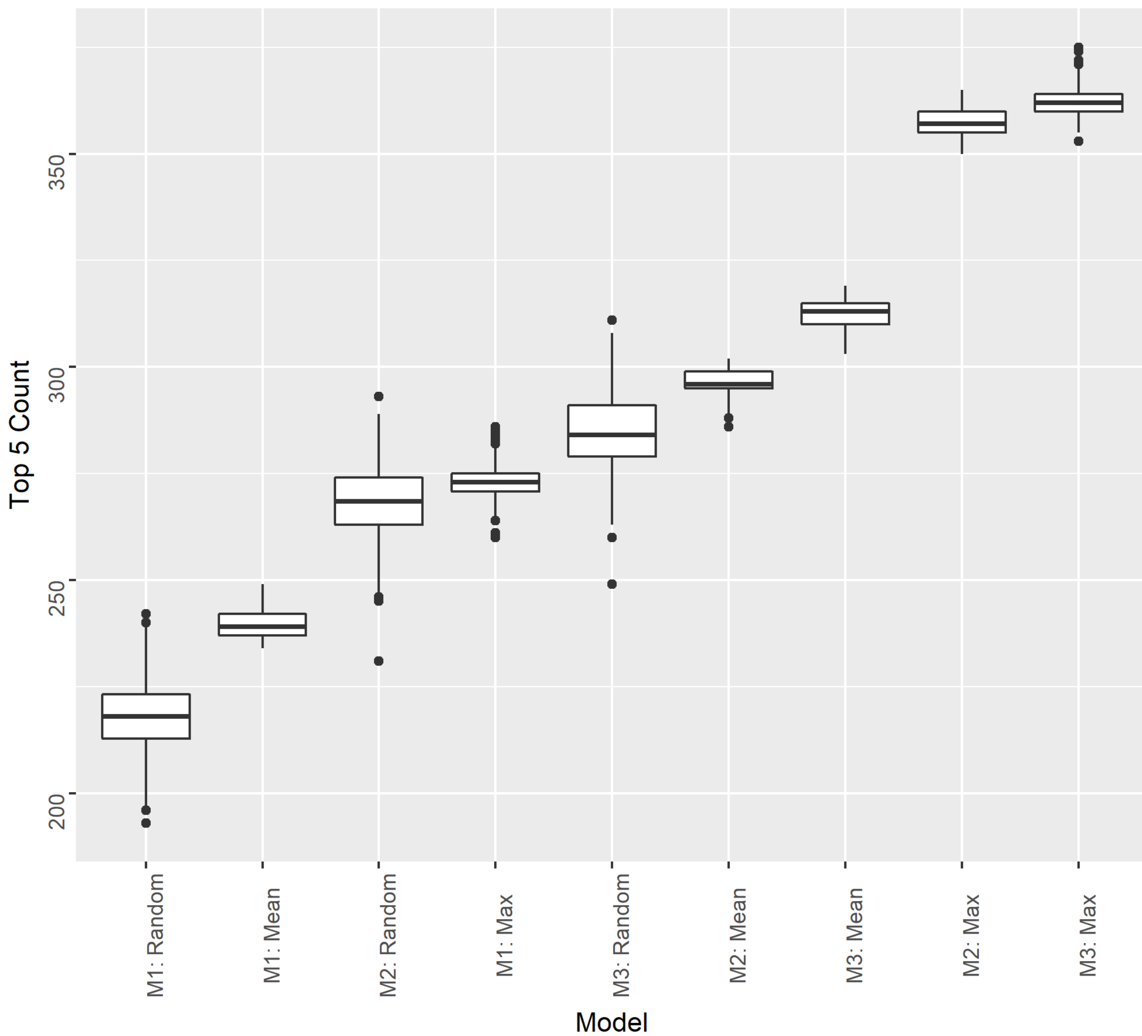


Figure 52: Prioritization Top 5 Hit Rate Distribution for Three Linkage Models for 997 index crimes across 200 Iterations for three Prioritization Methods

While the results reported in Figure 52 give a generalized overview of the prioritization capability of the three models thus far proposed, the generalized approach raised the question as to how consistently the three models were performing from iteration to iteration and method to method. The bootstrapping procedure allowed for the comparison of how consistently each model was able to prioritize for any given index crime by summing the total number of iterations for which a given model successfully prioritized the linked offender for a given index crime. Thus, the individual index crimes for which the three models successfully identified the linked offender were compared in order to assess the relative gain, in index crimes, each model represents in a practical context.

These comparisons are summarised in Table 20. Of the 997 index crimes, at least one of the three models was able to identify the linked offender at least once over the 200 bootstrap iterations for 438 index crimes. All three models successfully prioritized the linked offender for 202 index crimes across all 200 iterations (from line 4 of Table 20). This subgroup of index crimes was effectively the easiest to prioritize as the offenders who committed them could be readily identified regardless of the linkage model employed. Lines 5-7 from Table 20 highlight the number of index crimes for which any two models were successful but the third model failed to detect. It is not clear if there is any significant pattern present within the two-way model comparisons – more in-depth post-hoc analysis would be required – but it is at least clear that the competing models offer differing levels of performance for different subsets of crimes within the validation set. Finally, the last rows show the number of index crimes for which each individual model successfully identified the linked offender and the other two models did not. Once again, it is shown that the activity space model offers the highest hit rate of index crimes.

Table 20: Prioritization Hit Rate over 200 Iterations Three-way Model Comparisons

M1	M2	M3	200 Iterations	< 200 Iterations	Total Hits
1	-	-	226	34	260
-	1	-	267	57	324
-	-	1	286	53	339
1	1	1	200	2	202
0	1	1	51	12	63
1	0	1	5	3	8
1	1	0	7	3	10
1	0	0	14	26	40
0	1	0	9	40	49
0	0	1	30	36	66

Note. 1 / 0 denotes the presence of the model. For example, row four shows a “1” for all three models, which corresponds to a three-way comparison of the index crimes for which all three models were successful.

Finally, the index crimes for which each model successfully prioritized the linked offender and the other two competing models did not was mapped. This was done to investigate whether any given model was over-fit to a given geographic sub-area. If the models were relatively unbiased to these sub-areas, then the resulting visual spread should be roughly even throughout the observed sample space. Figure 53 shows the geographic location of the 202 index crimes for which all three models successfully identified the linked offender for all 200 iterations. It can be seen that the spatial distribution of these crimes is relatively even across the four major sub-divisions. Figure 54 shows the geographic location of the 40 index crimes for which Model One prioritized the linked offender; crime locations are colour-coded according to the number of iterations for which successful prioritization was observed (higher is better). The results for Models Two and Three can be seen in Figure 54 and Figure 55 respectively. These results suggest that the activity space based model may be somewhat spatially biased as the distribution of its' exclusive hits is not as uniform in appearance as the other two. More formal testing may be warranted in future research.

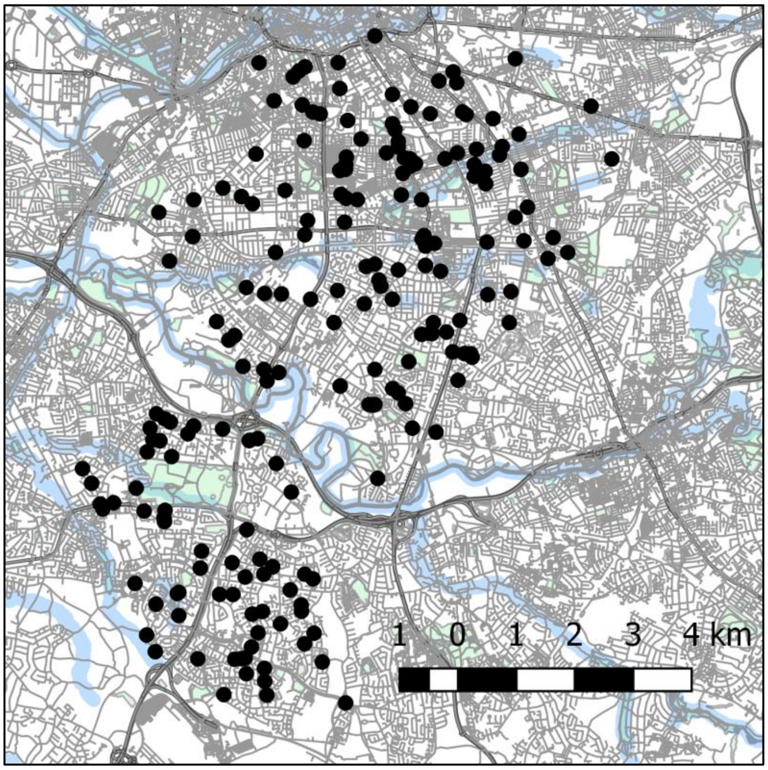


Figure 53: Index Crimes All Models Hit 100% of Iterations (n = 225 Index Crimes)

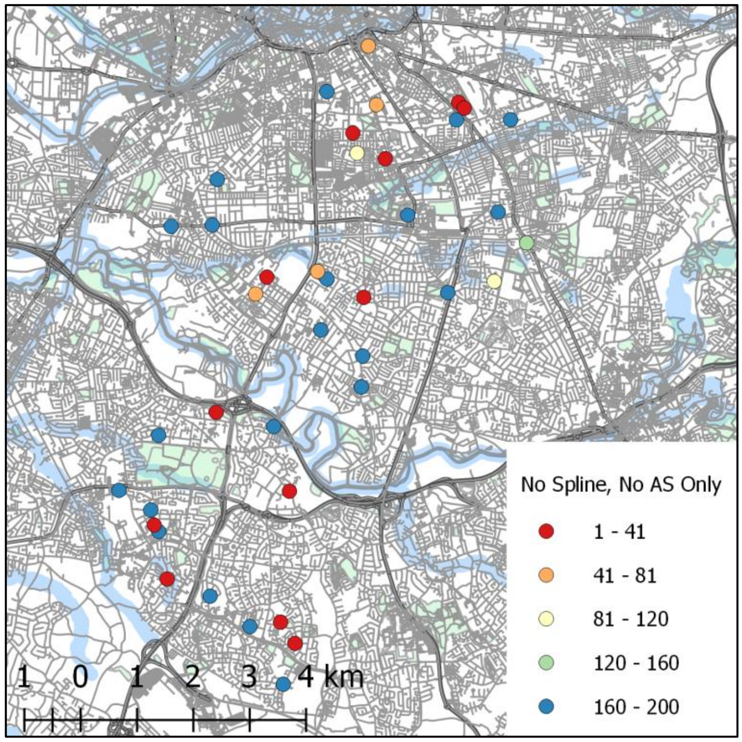


Figure 54: Model One - Exclusive Hits Categorized by Iteration Hit Count (n = 40 Index Crimes)

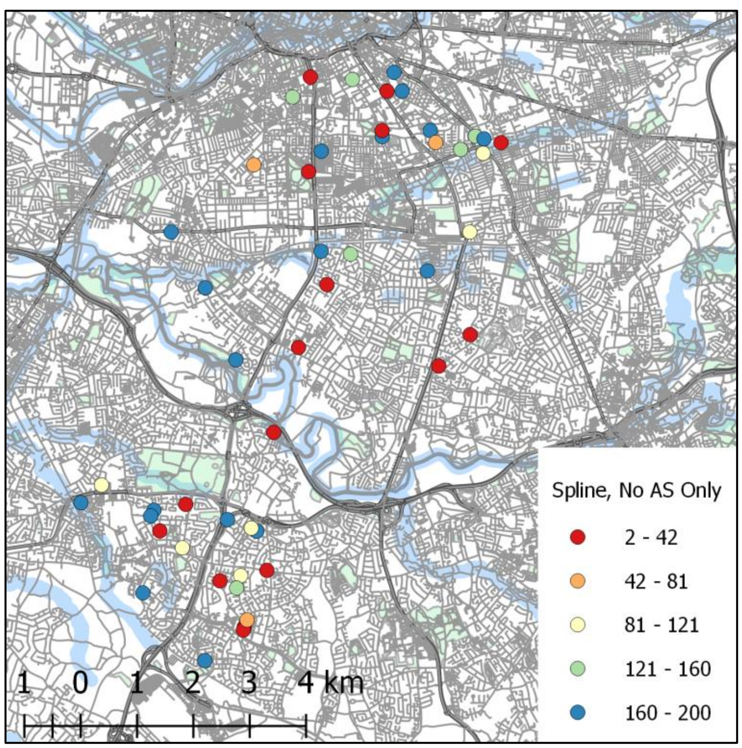


Figure 55: Model Two - Exclusive Hits Categorized by Iteration Hit Count (n = 49 Index Crimes)

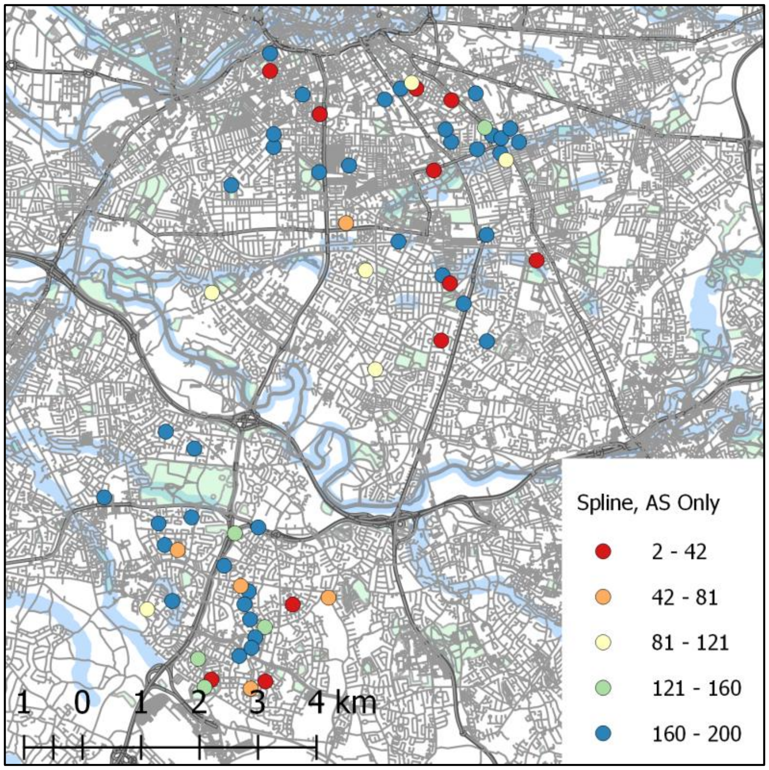


Figure 56: Model Three - Exclusive Hits Categorized by Iteration Hit Count (n = 66 Index

5.8 Discussion

The primary focus of this chapter was to address apparent limitations surfaced in Chapter 3; namely the false positive paradox and the resulting low precision validation tests. The present chapter proposed that one possible solution to these problems was a shift in perspective away from crime to crime linkage outcomes to offender to crime linkage outcomes. The first study presented a suspect prioritization process which centred on aggregating crime-to-crime linkage predictions by offender in order to assess how likely their crime series included test index crimes. The second study explored how consistently the various models performed over a set of bootstrapped samples and presented performance metrics assessing the unique contribution a given model made to prioritization performance.

Results from the first study of this chapter indicated that while the outlined method produced prioritized lists which contained the actual offender within the top five ranks far above chance levels, overall the “success” rate ranged from 24% to 32% of the sample of index crimes. In other words for the given sample, the proposed prioritization method was successful in about a third of the cases. While this figure may appear low in absolute terms, given the myriad list of known and unknown intervening variables, it is argued here that this number is very reasonable. It is difficult to quantify whether the observed success rate is “good” or “bad” relative to competing methods as there are few to no studies investigating a similar process to which to compare.

Regardless of the immediate practical application discussion, the first study of this chapter’s results are in line with the fundamental assumptions put forth in crime linkage theory: namely offenders do in fact appear to offend in uniquely consistent ways. The left skewed rank charts strongly imply that offenders’ past crimes could be used to predict whether they were responsible for committing their future crimes. If the underlying linking algorithm were producing predictions no better than chance, then one would expect a flat distribution where the mean and median rank were equivalent to approximately $\frac{n}{2}$ where n is the number of index crimes tested (n = 997).

Given the variety of crime types present within the validation set, as well as the variety present in criminal histories of the repeat offenders within the sample both in terms of raw number as well as duration there was an as of yet unanswered question regarding whether different offenders were “easier” or “harder” to identify linked crimes and subsequently prioritize for. While this potential bias has been present in all studies thus far investigated, it had not been directly addressed. Rather than attempt to identify all contributing factors that would influence whether a particular offender could be prioritized, the second study of this chapter investigated this question more abstractly.

Each of the three models initially presented in Chapter 3 were subjected to a bootstrapping validation procedure. This had several benefits: first it provided a within-model metric by which the consistency of the relationships identified in Chapter 3 could be assessed. If the covariates were resulting in spurious relationships, then one would expect model performance to return to chance levels when evaluated over the bootstrapped aggregate. This did not happen which further suggests the observed performance can be attributed to underlying behavioural / decision making processes. Next, it also provided an opportunity for between-model comparisons. By quantifying how many unique offenders each model was able to successfully prioritize on, the relative "gain" of each model could be approximated. This in turn could provide insight into how sensitive each model was to information the other models were not.

The vast majority of successfully prioritized offenders were 'found' equally well across all three models; implying an underlying behavioural process that is relative straight forward to detect. This general assertion is at odds with the fact that each model was also able to uniquely prioritize a subset of offenders successfully. One potential interpretation is a behavioural hierarchy where most offence targeting behaviour closely follows general patterns of propinquity with little need for further in-depth analysis to identify. Targeting behaviour not belonging in such a group may require more advanced identification methods that go beyond simple propinquity arguments. This thesis has proposed criminal activity spaces as a means to understand this second class of behaviour. Further analysis would be necessary to begin to fully understand these observed differences and whether they are indicative of some larger behavioural hierarchy.

The present chapter also explored the impact of different aggregation techniques upon the final prioritization performance of the various models. Most unexpectedly, the ‘max’ method consistently performed the best. Recall that the prioritization process involved creating link scores from an index crime to all other crimes committed by a given offender or offenders. The set of scores unique to each offender was then aggregated to create an offender level score which was then used to rank offenders relative to each other in terms of likelihood of being linked with the index crime. Conventional crime linkage theory posits that offenders are relatively consistent within their own crime series and so it was expected that the linked offender would be more consistently linked to their own index crime than those of other offenders. This would have been reflected via the ‘mean’ method outperforming other options, however this was not observed.

The implications of the ‘max’ method outperforming the ‘mean’ method include calling into question the consistency argument traditionally given as a justification for crime linkage analysis. In order for the max method to have performed as well as it has here, then offenders would have had to have committed at least one crime that was a-typical of their crime series in general (otherwise the linkage scores would have all been similar and thus had a similar mean score) while simultaneously being more similar to the index crime than any other crime committed by other offenders. This in and of itself is not problematic for the greater crime linkage literature or even the criminal activity spaces hypothesis as the ‘max’ method is still a measure of consistency in the form of past behaviour being indicative of future behaviour. The difference, however, is that the max method is more sensitive to sudden changes or variances in behaviour that the mean method is not. Such variations are also explicitly accounted for within the activity spaces construct in the form of activity clusters. Crimes that would score highly on the ‘max’ method but low on the ‘mean’ method may in fact consistently be crimes belonging to as-of-yet un-described clusters for a given offender. Again further research is required to fully explore and answer such questions.

The final validation inspection involved assessing whether the developed models were unintentionally regionally biased in some way. This was, rather informally, assessed by plotting the uniquely identified crimes for each model on a map. The rational for doing so was that if the models were in fact spatially biased (sensitive to some unidentified local characteristic) then one would expect to see some degree of regional clustering; either across all points or at least across a particular subset of points. The preliminary, and somewhat informal, initial investigation undertaken here did not appear to suggest such bias was present; for all models the points (regardless of how often they were identified) were largely evenly distributed across the sample space.

Chapter 6

Conclusions and Closing Remarks

6.1 Introduction

This thesis has focused on establishing the theoretical foundations which justify the use of a general ‘criminal activity space’ construct for describing individual criminal mobility. Such a construct, it has been argued, is a sensible way to model criminal spatial behaviour given the complex interactions between an individuals’ specific domain knowledge as well as the surrounding criminal opportunity ‘backcloth’ of their immediate environment. Crime linkage analysis was presented as a practical test-bed for exploring the validity and impact of such a measure. Finally, an offender prioritization system was presented which expanded upon traditional crime linkage methods to surface likely *individuals* over likely events. These broad goals were explored under three overarching research aims each composed of several targeted research questions. For sake of convenience these aims and research questions are restated here:

Aim 1: To explore and develop an empirical activity space measure for offender criminal activity.

Research Question Set 1:

1. Can diverse individual crime patterns be described via a generalized ‘activity space’?
2. Are such activity spaces predictive of future criminal spatial activity?

Chapter 3 focused on exploring how the criminal activity space could be measured in such a way as to capture meaningful individual variations. In response to research set 1 question

1, an algorithm was presented which would generate circular activity space geometries based on spatial clustering and outlier detection methods. Several properties of the resulting spaces were explored including their general size distribution and frequency with which multiple clusters would manifest. Finally, it was shown that spaces constructed in this way had some predictive ability for future crime by demonstrating that future crimes were contained within spaces constructed from past crimes more readily than an offender's home location.

Aim 2: Examine the validity of the activity space construct by exploring the predictive gains afforded by the inclusion of activity space based measures into crime linkage analysis tasks and models.

Research Question Set 2:

1. Can the inclusion of activity space measures improve established models of crime linkage analysis?
2. Can the findings from (1) be replicated using alternative modelling methodologies?

Chapter 4 explored the construct validity of the awareness space model presented in Chapter 1 by integrating empirical activity space measures into established crime linkage methodologies. It was hypothesized that if the activity space measures were representative of a general decision making process which guides where offenders may commit their crimes, then models containing activity space based measures would be more predictive than models without such measures. Question set 2, question 1 was empirically tested by developing three separate logistic regression models; with one model being a replication of past work acting as the control and two additional models incorporating additional predictors. The results of several validation tests indicated that models containing activity space based measures were more predictive as measured by AUC scores, model maximum likelihood comparisons and straight validation set classification performance metrics.

Question 2 of set 2 was explored by replicating the results from the logistic regression models with a different machine learning algorithm in random forests. Inspection of model performance metrics indicated that theoretically the random forest models performed similarly or slightly better than the logistic regression models. Straight classification performance further confirmed similar performance between the approaches. Finally, the variable importance measures given by the fit random forest models coincided with the regression coefficients provided by the logistic regression models. The random forest model also allowed for testing of a number of binary behaviour fields present within the data. These variables were not found to be significantly predictive, however there are considerable limitations in both the data quality and applicability of the domain fields that makes these results non-conclusive.

Aim 3: Assess the stability of the activity space based approach of crime linkage by exploring the possibility of a 'suspect prioritization' system which surfaces likely individuals linked to a crime event rather than simply other individual events.

Research Question Set 3:

3.5. Can offenders be meaningfully prioritized – i.e. ranked – for a given candidate set such that the responsible offender appears near the top of an automatically generated list?

3.6. Does the inclusion of activity space measures improve rank performance?

3.7. Does the aggregation method significantly impact rank performance?

3.8. Do other systematic factors, such as the underlying geography, disproportionately impact ranking results?

Chapter 5 explored construct validity of the awareness space model by testing a related but distinct domain of linkage analysis via a prioritization framework. Chapter 4 demonstrated that activity space measures contributed positively to model outcomes in a classical linkage task – that is crime to crime linkage classification. One could explain this line of reasoning to asking how individuals could be linked to crimes. In such a framework it would make sense for offenders whose activity spaces were more consistently linked to a given crime in aggregate. In order to test this hypothesis – that linked offenders would more consistently offender in and around their crimes than non-linked offenders, a prioritization system was developed which would rank offenders according to aggregate linked scores based on their criminal history.

Results from the first study in chapter 5 showed that there was an obvious skew to the rank distribution of linked offenders such that their rank position in a prioritized list was consistently skewed low. In other words, linked offenders were consistently ranked into the top quantiles of all possible offenders. Thus for research question 3.5 this thesis found evidence that offenders could be meaningfully prioritized for future crimes using their past criminal history as predictors.

Research question 3.6 asked whether such a prioritization system was improved or not by including offender activity spaces. Here again an array of predictive models was tested given this new success criterion which indicated that again models with activity space based measures generally outperformed models without such measures further solidifying the validity of the measure. Thus this thesis found evidence to suggest that, at least for a crime-linkage based prioritization system, offender activity spaces were a meaningful predictor at differentiating a pool of likely candidates for unsolved crimes.

Finally, Chapter 5 explored a number of validation tests based on the ranking outcomes. Specifically research questions 3.7 and 3.8 were targeted at identifying if any part of the proposed methods had an obvious systematic bias. A bootstrapping validation procedure was employed to explore how robust the various methods explored throughout the current would be as well as to determine what aggregation method was the most reliable. Bootstrapping results showed that the models were fairly consistent with their initially reported behaviour and that aggregating offenders according to their maximum linked score on any one crime yielded the best results by a considerable margin. This finding was not expected and is particularly intriguing for what it may indicate about overall spatial behaviour. Further research into individual criminal spatial consistency is highly suggested.

6.2 The Criminal Activity Space Construct

A key consideration of this research was defining what exactly a 'criminal activity space' should be defined and how it could be empirically measured. Past studies have struggled with such a definition and ultimately the one used here was provided as a starting point. The specifics of the measure are often determined by many factors not directly under a researchers control; the most obvious of which being data quality and availability. Crime data in particular is no different and often subject to greater limitations due to the temporal uncertainty associated with many crime events. Because of this uncertainty, certain compromises had to be made which directly impact the final activity space measure.

In the research presented here, activity spaces were described in terms of simple circular geometry. While circles can be used to account for a certain degree of temporal uncertainty - as shown by Miller (2005) – they can be overly simplistic. In order to construct more representative activity spaces two additional steps were outlined: 1) clustering and 2) outlier detection. Direct observation revealed that a singular surface – regardless of shape – was not going be a sufficiently descriptive geometry to capture the apparent variation not only between individuals, but also within individuals. Thus by 'clustering' proximal points together into distinct groups, more robust geometries could be derived. This achieved two goals: first it allowed for more specific activity spaces and second it increased the relative density of observed activity within said spaces. This second point is particularly important as a general risk that is introduced when attempting to describe activity with broad geometric surfaces is one of scale. It is trivially easy to create a single surface that encapsulates the entirety of an individual spatial activity (i.e. one giant circle), however such trivial solutions often have low point density. This implies that there are significant portions of the indicated space for which there is insufficient evidence to suggest the individual has ever been active there. This tension between spatial precision and point density necessitated the final step of outlier detection. By identifying points well removed - i.e. outliers - from the main cluster, the resulting activity space estimate was shown to both have a higher point density while also not sacrificing predictive power on future crime locations.

Finally, activity spaces as described here are completely 'blind' to specific journey-to-crime paths. That is, the activity spaces defined throughout the current research make no attempt to resolve what paths, if any, would result in the observed spatial clusters of activity. Furthermore, an individual's home location was not evaluated in any way with regard to their proposed activity spaces. These two points represent a potentially significant source of bias in the activity space measures. Despite these limitations, however, activity spaces were shown to consistently have a significant predictive relationship with future crime locations. This was demonstrated both by the frequency with which future crimes fell within an offender's activity space as well as by the results of the various linkage analyses conducted throughout this thesis.

6.3 Activity Space in Crime Linkage Analysis

Significant portions of this thesis focused on the crime linkage task - the process by which two crimes are determined to be perpetrated by the same individual or not. It was argued that this problem lent itself well to investigating the legitimacy of the activity space measure due to the fundamental assumptions outlined for the feasibility of crime linkage to be possible. To recap these were: 1) Offender consistency and 2) Offender distinctiveness. It was noted early on that activity spaces most likely follow similar rules; that is that the activity space of one person would represent their environmental behaviour over time as well as being unique to that individual. Factors such as an individual's routine activities as well as their individual awareness space of opportunity would ultimately give shape to their unique activity space. If activity spaces reflect such processes, then one would expect linkage tasks to be improved by incorporating such measures into the predictive models. Generally speaking, this was found to be the case.

Perhaps one of the most important findings arising from this research is related to the general implications arising from the activity space assumption. That is, that activity spaces represent environmental activity *in general*. Put another way, activity spaces are action independent; the specific activity that brought an individual to an area is largely irrelevant - what matters is the fact that the individual has a history in a given area at all. This implies, for example, that traditional approaches to crime linkage analysis which focus on specific crime types are needlessly narrow. By limiting ones sample to only robberies, for example, one is biasing their data to only individuals that commit many such offences. However real world data are filled with examples where individuals do not specialize to such an extreme where they only commit one type of crime. Individuals may exhibit a strong preference for certain types of crime, but it is rare that they would only ever commit one type exclusively. Activity space measures offer an avenue by which such variability can be accounted for in a consistent way.

From a methodological standpoint, the current research also addressed several outstanding issues from the literature. The majority of these revolved around appropriate models and model validation. Past studies have relied heavily upon the AUC to assess model quality; higher AUC values were taken to mean that, in general, models would yield more accurate predictions. However, due to the manner in which samples are typically constructed for linkage studies - such AUC interpretations do not always yield intuitive results. For a given set of crime events, the number of linked events across that sample is usually observed to be disproportionately small. This results in problems when attempting to construct predictive models on this rare outcome. Thus various sampling procedures are typically employed, such as negative case sub-sampling, in order to yield more balanced training sets. This in turn, yields models that have some predictive power in both directions. In other words, they can predict both positive and negative cases to some extent.

To recap, the standard decision workflow, as described in the literature by both researchers and practitioners, involves getting some predictive score from a linkage model which then has to be interpreted. The problem with the AUC in such cases is that it provides only an aggregate view of how likely higher such scores belong to the positive class than the negative class. Typically this is seen as a benefit: one is able to evaluate all possible cut-offs and in so doing get a general sense of the predictive ability of a model. Given the sampling procedure mentioned previously, however, the end result here is models that at first glance look pretty 'good' but would utterly fail in practical contexts. Models can have high AUC scores on validation sets of novel data but still provide little to no predictive power on positive predictions on sample sets that have not been down-sampled in a similar way as the initial training and test sets were. This more stringent criterion was demonstrated in this thesis to simulate what practitioners would see if they attempted to employ such models in applied settings. Obviously this weakness demonstrates a significant limitation with either: the final models, underlying methodology / assumptions; or both. The suspect prioritization portion of this thesis was presented as a possible solution to this observed problem.

6.4 Offender Prioritization

This thesis presented a so-called 'offender prioritization system' as a natural evolution to traditional crime linkage methods. The motivation was simple: if activity spaces were demonstrated to improve individual crime linkage predictions for offenders based on their offending history, then this implies that the linkage task could be improved if the focus of linking was changed away from crime-to-crime links towards offender-to-crime links. This is quite obviously a different linking scenario, and not all linkage applications are interchangeable. However, much of the linkage research justifies itself, at least in part, with the fact that a majority of crime is committed by a small number of prolific offenders. It is not a stretch to imagine that once a set of crimes has been determined to have been conducted by the same individual then practitioners might then be interested in determining a pool of suspects whom they might want to investigate for said group of crimes. Prior to the current research, however, there had been little formal research done within this domain within the literature. Furthermore, by aggregating linkage predictions to the offender level, it was believed that many of the precision problems identified in the previous crime linkage sections could be addressed.

The suspect prioritization process was built upon many of the same assumptions surrounding crime-to-crime linkage; namely that offenders would be consistent and distinct. If this is the case, it was hypothesized that a set of crimes committed by the same offender would have higher likelihood scores from a crime-to-crime linkage process than a set of crimes committed by different offenders. Furthermore, if this was in the fact the case, then offenders could be ranked based on an aggregate score of the linked likelihood of their previous crimes to some candidate crime. In this way, the focus is no longer whether a certain crime pair rises above an arbitrary decision threshold, but rather if any given offender's score was higher or lower than a pool of competing alternative offenders. The only remaining challenge was deciding at which point the process could be considered 'successful' - namely how far down a prioritized list was the actual offender allowed to appear before it was considered a 'miss'. Fortunately this decision is more intuitive and straight-forward to evaluate. A cut-off of the top five was used in the current research after discussions with practitioners around what a reasonable number of candidates would be to investigate.

The results indicated that such an approach held considerable promise: across a sample of nearly 1,000 crimes committed by an equal number of individual offenders, approximately 30% of offenders fell within the top five of offenders for their respective crime. Given the extremely low precision scores reported in the crime-to-crime linkage sections of this thesis, this number was very encouraging.

There were a number of free parameters present within the prioritization process, including: 1) linkage model used and 2) aggregation technique employed which warranted further testing. As expected given the results leading up to the initial study of the prioritization process, linkage models employing criminal activity spaces were demonstrated to yield the 'best' results. "Best" here meaning that that particular model surfaced the largest number of both total successfully prioritized offenders as well as the largest proportion of uniquely identified individuals. Note, however, that it was shown that a large proportion of the individuals identified successfully at all were shared across the various models.

The second free parameter was the aggregation technique employed to condense the relationship matrix of an offenders set of historical crimes to a candidate crime to one single score. Several aggregate methods were compared: mean, random and max. It was believed that the mean would produce the most stable and accurate results because, again, if offenders are truly consistent and distinctive then the linked likelihood scores were believed to be consistent across the entirety of their criminal career. The results, however, did not follow this line of thinking. Instead, the max aggregation method was observed to result in the most successful prioritizations. While the implications of such a finding are far from clear given the current set of studies, it is reasonable to conclude that at a minimum the consistency argument used in crime linkage analysis may be overly simplistic. If an offender's behaviour followed strict notions of consistency, then there should have been little to no difference in the tested aggregation methods.

There are a number of plausible explanations for the max method's superior performance. One possibility stems from the variable nature of behaviour in general as well as individual learning and experience. In fact, variable experience and familiarity levels are predicted in the awareness / activity space model: namely that over time an individual's awareness space becomes more refined and expansive. This in turn opens up knowledge of new areas of opportunity. Presumably there exists a point in time whereby an individual would venture into a new area for the first time. Such instances would be far removed from the historic locations of crime for that individual. Recall that such points would manifest as outlier points in the current activity space model. Further activity around this point would logically be far removed from previous activity nodes and thus have a lower average likelihood score with the exception of the outlier point. Thus the max method, given the particular combination of linkage methodology, activity space formulation, and prioritization methodology employed could be a plausible metric for capturing such emerging locations. The max method would also generally yield similar or greater values in instances of repeat victimization; an edge case while not explicitly accounted for in the current research is important to keep in mind.

Given the results of the various spatial and linking models discussed above, it would seem fair to say that offender spatial behaviour is a complicated issue that has thus far been overly simplified within the literature. Past crime linkage studies typically evaluated linked crimes based only on the similarity between two crimes directly. These studies largely ignored the context in which those crimes exist: namely their local environments and the histories of the offenders who may have committed them. This thesis introduced the activity space construct as a means to introduce additional context into the linking decision: namely where have individual offenders committed crimes before, and do such locations relate to their future crime locations. The results reported here suggest that the spatial constraints faced by offenders are as important, or possible even more important, than the specific actions or crimes taken at such locations. This has both theoretical and practical implications, the most fundamental of which being that offenders and their criminal histories should be considered much more holistically than they have been in the past. This includes being more generalist in data collection for academics, and not over emphasizing offender ‘types’ by practitioners.

6.5 Limitations

The studies presented as part of this thesis were all based on a sizeable data set of real-world crime data provided by the Greater Manchester Police (GMP). As outlined in Chapter 2, however, this datasets was not without problems. These included standard data quality issues often observed when working with real-world data such as: incorrect / inconsistent geo-coding / missing data. Such issues are often expected, but it is worth noting that in the case of home locations in particular, the data was deemed unreliable enough to forgo any formal analysis of offender home locations; this is a potentially significant limitation as this thesis is subsequently unable to make any comments on the relationship between where offenders live and where they subsequently commit their crimes.

A related data-quality issue concerns the inconsistent time representation present within the data. Because the data was aggregated together over time form a series of ‘data-dumps’, there are gaps in incident rates that do not appear to be due to seasonal fluctuations in crime rates. As such, there is a possibility that the data is not fully representative of repeat offenders, as it is likely that actual repeat offenders had to be excluded because their repeat crimes were not captured. Limitations such as this are not new to the present thesis; studies involving criminal data often acknowledge that the total of reported crime is only a small fraction of all criminal activity taking place, but the problem faced by the current thesis is compounded by the added layer of additional systematically missing data.

Finally, the current thesis is limited by the offending window present for study. Because activity space was presented as a measure of holistic offender activity, it works ‘best’ with more data rather than less (see: Chapter 2). The data available for use as part of this thesis, however, only covered a number of years. Arguably, the models built upon this data may only be descriptive of a certain subset of offenders: namely those offenders who offend repeatedly over a short number of years. This limitation is also why the current thesis was unable to investigate changes in activity space over time in more depth; the observation window was simply too small.

There are several significant limitations arising from the methodology used. While data availability has been a constant limiting theme throughout this work, it is particularly impactful in the prioritization method for several reasons. First, limited information regarding any particular individuals' past crimes influences the activity space estimate for said offender - this has been well established up to this point. Data quality is also a potential factor in the max aggregation's performance. Incomplete criminal history data could result in more disjointed clusters which in turn would limit the apparent consistency of an offender; not because they are in fact random in their offending patterns but because we have an incomplete picture of said activity. Unfortunately it is not possible to know what proportion of crimes are truly missed or omitted for a given offender, and this limitation is faced by practitioners on a daily basis.

Finally, the prioritization method described here is biased in a systematic way because only offenders with multiple crimes in their criminal history were included. While this was done intentionally so that the predictive ability of the methods and models could be evaluated, it still results in a more 'ideal' scenario than is available in practice. Because offenders were only included who had three or more crimes on record, it is not clear if the models would remain predictive for offenders who first cross that threshold. Furthermore, the performance metrics cited are only true when being evaluated on crimes for which the responsible offender is present in the data. Put another way, if such a prioritization method was employed by practitioners and used on crimes which were committed by first time offenders then the process is doomed to failure. However, it is not clear that any alternative data driven model of past behaviour would perform any better in such a situation.

6.6 Future Directions

While many of the concepts presented in this thesis are not in and of themselves 'new', the specific combination of elements presented here has not been attempted before. Given this fact, as well as many of the unique challenged associated with criminal data, there are a number of outstanding questions that remain unaddressed. These can be loosely categorized based on the research focuses of the chapters used throughout this thesis including: 1) activity spaces, 2) crime linkage analysis and 3) suspect prioritization.

6.6.1 Activity Spaces

The activity space measure employed in this thesis was used as a starting point. Future investigation into how best to represent such spaces could include not just developing more refined geometry measures, but also investigating whether path based estimations are more or less appropriate. If more accurate information regarding travel and home locations were available then it may be possible to develop highly detailed life-paths in lieu of the geometric spaces used here.

Furthermore, while only crime locations were employed in the present study, it is reasonable to assume that non-criminal activity would be equally predictive of possible crime activity. The dubious nature of personal privacy and consent, however, makes evaluating the legitimacy of a 'life-path' activity space model tenuous at best.

Further studies and developments in the activity space literature would be greatly aided by the development of a sort of ‘spatial-index’. Such a ‘spatial-index’ could be a measure used to relate the how often a given activity space contains subsequent crimes (the ‘hit’ rate) and how much area the activity space actually covers. This idea was presented somewhat informally within this thesis as notions of activity space ‘density’, but could be further developed. Geographic offender profiling, for example, typically measure a given model’s performance by evaluating how much of the total ‘search area’ investigators would have to traverse before the offender’s home location was located. Currently, there is no direct way in which to compare two competing activity space measures to each other.

Finally, more targeted research at the relationship between criminal histories and activity spaces may be revealing.

This thesis has only scratched the surface of investigating how and why activity spaces develop over time. In fact, there is only the most basic of support highlighting the fact that spaces do appear to change over time; the body of work presented here can make no solid claims as to exactly how or why such changes occur.

6.6.2 Crime Linkage Analysis

Crime linkage analysis was used in this work to provide a lens through which to evaluate the predictive ability of activity spaces, but in so doing has highlighted a need for more general approaches for crossing crime types in the linkage task. Specifically, one of the unanswered questions from the work presented here is how to mediate observed behaviours across disparate crime types. At the surface level, it is not obvious how best to mediate the behaviours employed during a violent assault and a covert dwelling burglary, for example. The data made available for the current study was not up to the task of addressing this problem and thus it remains a ripe topic for more targeted future research. Finally, results from the full behavioural random forest model suggested that additional special measures – including absolute geo locations – could further improve linkage accuracy. Further study could focus on the question of whether the abstract spatial proximity measure is sufficient for linkage prediction, or if more specific locational context would be of benefit.

6.6.3 Offender Prioritization

First and foremost, replication is by far the most important future goal for the prioritization process. Replication would provide insight into whether the methods described in this thesis are valid or simply an artefact of the data made available for study. Beyond that, some of the validation work suggested that different linkage models were able to identify different subsets of offenders. While it is not surprising that the results of different models are themselves different, it does beg the question as to whether there are notional groups of offenders whose behavioural pattern would be more readily identified by a more specific model. General crime linkage analysis treats all crime as equal and the underlying process which gives rise to linked crime as similar across the pool of offenders. Further developing the prioritization process may give insight into whether such a generic process exists or if researchers and practitioners would be better served by developing more targeted methods.

Due to data limitations within the sample available for study here, relationships between an offender's home location and their crimes was not investigated. However,the geographic offender profiling (GOP) literature has established that anchors such as the home can have direct relationships to where offenders commit crimes. Furthermore, GOP methods often produce outputs very similar to those of the prioritization process proposed as part of this thesis. Where the current study used solved crimes of repeat offenders to produce a prioritized list of offenders, GOP instead looks at a list of unsolved crimes and identifies how lives in close proximity to the cluster of unsolved crimes. Future research could focus on merging these two approaches: rather than simply ask which offender is most likely responsible for one target crime, for example, future research could generalize the prioritization process out to a set of crimes. In the same vein, future studies could re-introduce offender home location and identify if it plays a key role in crime linkage and offender prioritization or not.

Finally, further research surrounding the aggregation method is warranted; not just to identify which aggregation method is 'best', but to also establish if the aggregation method in general is appropriate. The aggregation method was developed here in response to a limitation in the crime-to-crime linkage methodology. There is a possibility, however, that a prioritization method that is not contingent on aggregated crime-to-crime linked likelihoods could be developed.

6.7 Closing Remarks

This thesis has built upon past work of crime linkage analysis and activity space constructs to present an offender-based prioritization system. Offender-level activity spaces served as the foundation from which many of the individual-level arguments flowed including: using individual-level calibrated activity spaces of past crime to inform decisions regarding suspects for unsolved crimes; the acknowledgment that where crimes occur can be just as or more important than what crime is actually committed. The results presented here challenge the status quo of using fixed crime types as established within the linkage literature. In addition, the methods and results presented here have expanded established ideas of activity space into the criminal domain. The culmination of this work in the proposed prioritization system offers a unique twist on established linkage methods that has direct applications in both practical and academic contexts.

References

Akers, R. L. (1985). *Deviant behavior: a social learning approach*. Wadsworth Pub. Co.

Atkinson, R. C., & Shiffrin, R. M. (1968). Human Memory: A Proposed System and its Control Processes1. In K. W. S. and J. T. Spence (Ed.), *Psychology of Learning and Motivation* (Vol. 2, pp. 89–195). Retrieved from http://www.sciencedirect.com/science/article/pii/S0079742108604223

Bandura, A. (1986). *Social foundations of thought and action: a social cognitive theory*. Prentice-Hall.

Baumgartner, K., Ferrari, S., & Palermo, G. (2008). Constructing Bayesian networks for criminal profiling from limited data. *Knowledge-Based Systems*, *21*(7), 563–572. https://doi.org/10.1016/j.knosys.2008.03.019

Bennell, C. (2005). Improving police decision making: general principles and practical applications of receiver operating characteristic analysis. *Applied Cognitive Psychology*, *19*(9), 1157–1175. https://doi.org/10.1002/acp.1152

Bennell, C., Bloomfield, S., Snook, B., Taylor, P., & Barnes, C. (2010). Linkage analysis in cases of serial burglary: comparing the performance of university students, police professionals, and a logistic regression model. *Psychology, Crime & Law*, *16*(6), 507–524. https://doi.org/10.1080/10683160902971030

Bennell, C., & Canter, David. V. (2002). Linking commercial burglaries by modus operandi: tests using regression and ROC analysis. *Science & Justice*, *42*(3), 153–164. https://doi.org/10.1016/S1355-0306(02)71820-0

Bennell, C., & Jones, N. J. (2005). Between a ROC and a Hard Place: A Method for Linking Serial Burglaries by Modus Operandi. *Journal of Investigative Psychology & Offender Profiling*, *2*(1), 23–41. https://doi.org/10.1002/jip.21

Bennell, C., Jones, N. J., & Melnyk, T. (2009). Addressing problems with traditional crime linking methods using receiver operating characteristic analysis. *Legal and Criminological Psychology*, *14*(2), 293–310. https://doi.org/10.1348/135532508X349336

Bennell, C., Mugford, R., Ellingwood, H., & Woodhams, J. (2014). Linking Crimes Using Behavioural Clues: Current Levels of Linking Accuracy and Strategies for Moving Forward: Linking crimes using behavioural clues. *Journal of Investigative Psychology and Offender Profiling*, *11*(1), 29–56. https://doi.org/10.1002/jip.1395

Bennell, C., & Woodhams, J. (2012). Behavioural Linking of Crimes: Behavioural linking. *Journal of Investigative Psychology and Offender Profiling*, *9*(3), 199–200. https://doi.org/10.1002/jip.1379

Bernasco, W. (2010). A sentimental journey to crime: Effects of residential history on crime location choice. *Criminology*, *48*(2), 389–416.

Bishop, C. M. (2016). *Pattern Recognition and Machine Learning* (Softcover reprint of the original 1st edition 2006 (corrected at 8th printing 2009)). New York, NY: Springer New York.

Boulesteix, A.-L., Janitza, S., Kruppa, J., & König, I. R. (2012). Overview of random forest methodology and practical guidance with emphasis on computational biology and bioinformatics. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *2*(6), 493–507.

Box, G. E. P., & Cox, D. R. (1964). An Analysis of Transformations. *Journal of the Royal Statistical Society. Series B (Methodological)*, *26*(2), 211–252.

Brantingham, & Brantingham. (1995). Criminality of place. *European Journal on Criminal Policy and Research*, *3*(3), 5–26.

Brantingham, P. J., Brantingham, P. L., & others. (1981). *Environmental criminology*. Sage Publications Beverly Hills, CA.

Brantingham, P. L., & Brantingham, P. J. (1993). Nodes, paths and edges: Considerations on the complexity of crime and the physical environment. *Journal of Environmental Psychology*, *13*(1), 3–28.

Breunig, M. M., Kriegel, H.-P., Ng, R. T., & Sander, J. (2000). LOF: identifying density-based local outliers. *ACM Sigmod Record*, *29*, 93–104. Retrieved from http://dl.acm.org/citation.cfm?id=335388

Brown, L. A., & Moore, E. G. (1970). The Intra-Urban Migration Process: A Perspective. *Geografiska Annaler. Series B, Human Geography*, *52*(1), 1–13. https://doi.org/10.2307/490436

Buliung, R. N., & Kanaroglou, P. S. (2006). A GIS toolkit for exploring geographies of household activity/travel behavior. *Journal of Transport Geography*, *14*(1), 35–51. https://doi.org/10.1016/j.jtrangeo.2004.10.008

Canter, D., Hammond, L., Youngs, D., & Juszczak, P. (2013). The Efficacy of Ideographic Models for Geographical Offender Profiling. *Journal of Quantitative Criminology*, *29*(3), 423–446. https://doi.org/10.1007/s10940-012-9186-6

Canter, D., & Larkin, P. (1993). The environmental range of serial rapists. *Journal of Environmental Psychology*, *13*(1), 63–69.

Canter, D. V. (1977). *The psychology of place*. London: Architectural Press.

Canter, D. V. (1995). *Criminal shadows: inside the mind of the serial killer*. London: HarperCollins.

Canter, D. V. (2000). Offender profiling and criminal differentiation. *Legal and Criminological Psychology*, *5*(1), 23–46.

Canter, & Hammond. (2006). A comparison of the efficacy of different decay functions in geographical profiling for a sample of US serial killers. *Journal of Investigative Psychology and Offender Profiling*, *3*(2), 91–103. https://doi.org/10.1002/jip.45

Canter, & Youngs. (2009). *Investigative psychology: offender profiling and the analysis of criminal action*. Chichester ; Hoboken, NJ: Wiley, a John Wiley & Sons, Ltd., Publication.

Chaix, B., Kestens, Y., Perchoux, C., Karusisi, N., Merlo, J., & Labadi, K. (2012). An Interactive Mapping Tool to Assess Individual Mobility Patterns in Neighborhood Studies. *American Journal of Preventive Medicine*, *43*(4), 440–450. https://doi.org/10.1016/j.amepre.2012.06.026

Chen, Y., Elenee Argentinis, J., & Weber, G. (2016). IBM Watson: How Cognitive Computing Can Be Applied to Big Data Challenges in Life Sciences Research. *Clinical Therapeutics*, *38*(4), 688–701. https://doi.org/10.1016/j.clinthera.2015.12.001

Chidananda Gowda, K., & Krishna, G. (1978). Agglomerative clustering using the concept of mutual nearest neighbourhood. *Pattern Recognition*, *10*(2), 105–112. https://doi.org/10.1016/0031-3203(78)90018-3

Clarke, R. V., & Cornish, D. B. (1985). Modeling Offenders’ Decisions: A Framework for Research and Policy. *Crime and Justice*, *6*, 147–185.

Clarke, R. V. G., & Felson, M. (1993). *Routine Activity and Rational Choice*. Transaction Publishers.

Cohen, L. E., & Felson, M. (1979). Social Change and Crime Rate Trends: A Routine Activity Approach. *American Sociological Review*, *44*(4), 588. https://doi.org/10.2307/2094589

Cook, N. R. (2007). Use and Misuse of the Receiver Operating Characteristic Curve in Risk Prediction. *Circulation*, *115*(7), 928–935. https://doi.org/10.1161/CIRCULATIONAHA.106.672402

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*(4), 281–302. https://doi.org/10.1037/h0040957

Deming, W. E., & Morgan, S. L. (1993). *The Elements of Statistical Learning* (Second, Vol. 11). Retrieved from http://www.newbooks-services.de/MediaFiles/Texts/0/9780387848570_Intro_001.pdf

Dhillon, I. S., Mallela, S., & Kumar, R. (2003). A divisive information-theoretic feature clustering algorithm for text classification. *Journal of Machine Learning Research*, *3*(Mar), 1265–1287.

Dijst, M. (1999). Two-earner families and their action spaces: A case study of two dutch communities. *GeoJournal*, *48*(3), 195. https://doi.org/10.1023/A:1007031809319

Emeno, K., & Bennell, C. (2013). The effectiveness of calibrated versus default distance decay functions for geographic profiling: a preliminary examination of crime type. *Psychology, Crime & Law*, *19*(3), 215–232. https://doi.org/10.1080/1068316X.2011.621426

Exploring Crime Analysis: Readings on Essential Skills. (2004). BookSurge.

Falk, Ö., Wallinius, M., Lundström, S., Frisell, T., Anckarsäter, H., & Kerekes, N. (2013). The 1 % of the population accountable for 63 % of all violent crime convictions. *Social*

Psychiatry and Psychiatric Epidemiology, 49(4), 559–571. https://doi.org/10.1007/s00127-013-0783-y

Farrington, D. P., Coid, J. W., Harnett, L., Jolliffe, D., Soteriou, N., Turner, R., & West, D. J. (2006). *Findings 281*. Retrieved from http://www.crim.cam.ac.uk/people/academic_research/david_farrington/hofind281.pdf

Fitterer, J., Nelson, T. A., & Nathoo, F. (2015). Predictive crime mapping. *Police Practice and Research, 16*(2), 121–135. https://doi.org/10.1080/15614263.2014.972618

Frank, R., Andresen, M. A., & Brantingham, P. L. (2012). Criminal directionality and the structure of urban form. *Journal of Environmental Psychology, 32*(1), 37–42. https://doi.org/10.1016/j.jenvp.2011.09.004

Frank, R., Andresen, M. A., & Brantingham, P. L. (2013). Visualizing the directional bias in property crime incidents for five Canadian municipalities. *The Canadian Geographer / Le Géographe Canadien, 57*(1), 31–42. https://doi.org/10.1111/j.1541-0064.2012.00450.x

Freeman, E. A., Moisen, G. G., & Frescino, T. S. (2012). Evaluating effectiveness of down-sampling for stratified designs and unbalanced prevalence in Random Forest models of tree species distributions in Nevada. *Ecological Modelling*, *233,* 1–10. https://doi.org/10.1016/j.ecolmodel.2012.03.007

Friedman, J. (1999, February 24). *Greedy Function Approximation: A Gradient Boosting Machine*. Retrieved from https://statweb.stanford.edu/~jhf/ftp/trebst.pdf

Golledge, R. G., & Stimson, R. J. (1997). *Spatial behavior: a geographic perspective*. New York: Guilford Press.

Goodwill, A. M., & Alison, L. J. (2006). The development of a filter model for prioritising suspects in burglary offences. *Psychology, Crime & Law*, *12*(4), 395–416. https://doi.org/10.1080/10683160500056945

Greenberg Raanan, M., & Shoval, N. (2014). Mental maps compared to actual spatial behavior using GPS data: A new method for investigating segregation in cities. *Cities*, *36*, 28–40. https://doi.org/10.1016/j.cities.2013.09.003

Gwinn, S. L., Bruce, C. W., Hick, S. R., Cooper, J. P., & International Association of Crime Analysts. (2008). *Exploring crime analysis: readings on essential skills*. Overland Park, Kan.: International Association of Crime Analysts.

Hägerstraand, T. (1970). What About People in Regional Science? *Papers in Regional Science*, *24*(1), 7–24. https://doi.org/10.1111/j.1435-5597.1970.tb01464.x

Harrell, F. E. (2001). *Regression Modeling Strategies*. Retrieved from http://link.springer.com/10.1007/978-1-4757-3462-1

Hartley, T., & Burgess, N. (2002). Models of spatial cognition. *Encyclopaedia of Cognitive Science, London: Palgrave MacMillan*. Retrieved from http://www.icn.ucl.ac.uk/nburgess/papers/hartleyecs.pdf

Hawkins, D. M. (1980). *Identification of Outliers.* Dordrecht: Springer Netherlands.

Hazelwood, R. R., & Warren, J. I. (2003). Linkage analysis: Modus operandi, ritual, and signature in serial sexual crime. *Aggression and Violent Behavior*, *8*(6), 587–598. https://doi.org/10.1016/S1359-1789(02)00106-4

Hazewinkel, M. (Ed.). (1988). *Encyclopaedia of mathematics: an updated and annotated translation of the Soviet "Mathematical encyclopaedia."* Dordrecht ; Boston : Norwell, MA, U.S.A: Reidel ; Sold and distributed in the U.S.A. and Canada by Kluwer Academic Publishers.

Heft, H. (2013). Environment, cognition, and culture: Reconsidering the cognitive map. *Journal of Environmental Psychology*, *33*, 14–25. https://doi.org/10.1016/j.jenvp.2012.09.002

Hermer, L., & Spelke, E. (1996). Modularity and development: the case of spatial reorientation. *Cognition*, *61*(3), 195–232. https://doi.org/10.1016/S0010-0277(96)00714-7

Horton, F. E., & Reynolds, D. R. (1971). Effects of Urban Spatial Structure on Individual Behavior. *Economic Geography*, *47*(1), 36. https://doi.org/10.2307/143224

Johnston, R. J. (1972). Activity Spaces and Residential Preferences: Some Tests of the Hypothesis of Sectoral Mental Maps. *Economic Geography*, *48*(2), 199. https://doi.org/10.2307/142876

Kanungo, T., Mount, D. M., Netanyahu, N. S., Piatko, C. D., Silverman, R., & Wu, A. Y. (2002). An efficient k-means clustering algorithm: analysis and implementation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *24*(7), 881–892. https://doi.org/10.1109/TPAMI.2002.1017616

Kassambara, A., & Mundt, F. (2016). factoextra: Extract and Visualize the Results of Multivariate Data Analyses (Version R package version 1.0.3). Retrieved from https://CRAN.R-project.org/package=factoextra

Kent, J., Leitner, M., & 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), 181–200. https://doi.org/10.1016/j.compenvurbsys.2004.10.002

Kershaw, C., Nicholas, S., & Walker, A. (2008). *Crime in England and Wales 2007/08: findings from the Critish Crime Survey and police recorded crime.* London: Home Office.

Kestens, Y., Lebel, A., Daniel, M., Thériault, M., & Pampalon, R. (2010). Using experienced activity spaces to measure foodscape exposure. *Health & Place*, *16*(6), 1094–1103. https://doi.org/10.1016/j.healthplace.2010.06.016

Kovács, F., Legány, C., & Babos, A. (2005). Cluster validity measurement techniques. *6th International Symposium of Hungarian Researchers on Computational Intelligence*. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.100.2848&rep=rep1&type=pdf

Kwan, M.-P. (1998). Space-Time and Integral Measures of Individual Accessibility: A Comparative Analysis Using a Point-based Framework. *Geographical Analysis*, *30*(3), 191–216. https://doi.org/10.1111/j.1538-4632.1998.tb00396.x

Learmonth, A. E., Nadel, L., & Newcombe, N. S. (2002). Children's Use of Landmarks: Implications for Modularity Theory. *Psychological Science*, *13*(4), 337–341. https://doi.org/10.1111/j.0956-7976.2002.00461.x

Lee, J. H., Davis, A. W., Yoon, S. Y., & Goulias, K. G. (2016). Activity space estimation with longitudinal observations of social media data. *Transportation*, *43*(6), 955–977. https://doi.org/10.1007/s11116-016-9719-1

Lewin, K. (1951). *Field theory in social science: selected theoretical papers (Edited by Dorwin Cartwright.)*. Oxford, England: Harpers.

Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of*

Experimental Social Psychology, *49*(4), 764–766. https://doi.org/10.1016/j.jesp.2013.03.013

Lundrigan, S., Czarnomski, S., & Wilson, M. (2010). Spatial and environmental consistency in serial sexual assault. *Journal of Investigative Psychology and Offender Profiling*, *7*(1), 15–30. https://doi.org/10.1002/jip.100

Lynch, K. (1984). *Good City Form*. MIT Press.

Miller, H. J. (2005). A Measurement Theory for Time Geography. *Geographical Analysis*, *37*(1), 17–45. https://doi.org/10.1111/j.1538-4632.2005.00575.x

Mischel, W., & Shoda, Y. (2000). *A cognitive-affective system theory of personality: Reconceptualizing situations, dispositions, dynamics, and invariance in personality structure.* Retrieved from http://search.proquest.com.libaccess.hud.ac.uk/docview/1290960263?pq-origsite=summon

O'Leary, M. (2011). Modeling criminal distance decay. *Cityscape*, 161–198.

Parthasarathi, P., Hochmair, H., & Levinson, D. (2015). Street network structure and household activity spaces. *Urban Studies*, *52*(6), 1090–1112. https://doi.org/10.1177/0042098014537956

Paulsen, D. (2007). Improving Geographic Profiling through Commuter/Marauder Prediction. *Police Practice and Research*, *8*(4), 347–357. https://doi.org/10.1080/15614260701615045

Perchoux, C., Chaix, B., Cummins, S., & Kestens, Y. (2013). Conceptualization and measurement of environmental exposure in epidemiology: Accounting for activity space related to daily mobility. *Health & Place*, *21*, 86–93. https://doi.org/10.1016/j.healthplace.2013.01.005

Piaget, J., Inhelder, B., & Piaget, J. (1997). *The child's conception of space*. London ; New York: Routledge.

Powers, D. M. (2011). *Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation*. Retrieved from http://dspace2.flinders.edu.au/xmlui/handle/2328/27165

Ratcliffe, J. H. (2004). Crime Mapping and the Training Needs of Law Enforcement. *European Journal on Criminal Policy and Research*, *10*(1), 65–83. https://doi.org/10.1023/B:CRIM.0000037550.40559.1c

Ren, F. (2016). *Activity Space*. Retrieved from http://www.oxfordbibliographies.com/display/id/obo-9780199874002-0137

Rheinfurth, M. H., & Howell, L. W. (1998). *Probability and statistics in aerospace engineering*. Retrieved from http://ntrs.nasa.gov/search.jsp?R=19980045313

Rossmo, D. K. (2000). *Geographic profiling*. Boca Raton, Fla: CRC Press.

Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, *20*, 53–65.

Sampson, R. J., & Raudenbush, S. W. (1999). Systematic Social Observation of Public Spaces: A New Look at Disorder in Urban Neighborhoods. *American Journal of Sociology*, *105*(3), 603–651. https://doi.org/10.1086/210356

Santtila, P., Korpela, S., & Häkkänen, H. (2004). Expertise and Decision- Making in the Linking of Car Crime Series. *Psychology, Crime & Law*, *10*(2), 97–112. https://doi.org/10.1080/10683160210000030559

Sherman, J. E., Spencer, J., Preisser, J. S., Gesler, W. M., & Arcury, T. A. (2005). A suite of methods for representing activity space in a healthcare accessibility study. *International Journal of Health Geographics, 4,* 24. https://doi.org/10.1186/1476-072X-4-24

Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., … Hassabis, D. (2017). Mastering the game of Go without human knowledge. *Nature*, *550*(7676), 354–359. https://doi.org/10.1038/nature24270

Slater, C., Woodhams, J., & Hamilton-Giachritsis, C. (2015). Testing the Assumptions of Crime Linkage with Stranger Sex Offenses: A More Ecologically-Valid Study. *Journal of Police and Criminal Psychology*, *30*(4), 261–273. https://doi.org/10.1007/s11896-014-9160-3

Snook, B., Cullen, R. M., Mokros, A., & Harbort, S. (2005). Serial murderers' spatial decisions: factors that influence crime location choice. *Journal of Investigative Psychology and Offender Profiling*, *2*(3), 147–164. https://doi.org/10.1002/jip.35

Snook, B., Wright, M., House, J. C., & Alison, L. J. (2006). Searching for a Needle in a Needle Stack: Combining Criminal Careers and Journey- to- Crime Research for Criminal Suspect Prioritization. *Police Practice and Research*, *7*(3), 217–230. https://doi.org/10.1080/15614260500432972

Snow, J. (1855). *On the mode of communication of cholera*. John Churchill.

Sutherland, E. H., Cressey, D. R., & Luckenbill, D. F. (1992). *Principles of Criminology*. Rowman & Littlefield.

Swets, J. (1988). Measuring the accuracy of diagnostic systems. *Science*, *240*(4857), 1285–1293. https://doi.org/10.1126/science.3287615

Teller, A., & Veloso, M. (1995). Algorithm evolution for face recognition: what makes a picture difficult. *Proceedings of 1995 IEEE International Conference on Evolutionary Computation*, *2*, 608–613 vol.2. https://doi.org/10.1109/ICEC.1995.487453

The H2o.ai team. (2017). h2o: R Interface for H2O (Version R package version 3.10.3.6). Retrieved from https://CRAN.R-project.org/package=h2o

Tolman, E. C. (1948). Cognitive maps in rats and men. *Psychological Review*, *55*(4), 189–208. https://doi.org/10.1037/h0061626

Tonkin, M. J. (2012). *Behavioural case linkage: Generalisability, ecological validity, and methodology* (University of Leicester). Retrieved from https://lra.le.ac.uk/handle/2381/27620

Tonkin, M., Woodhams, J., Bull, R., Bond, J. W., & Palmer, E. J. (2011). Linking Different Types of Crime Using Geographical and Temporal Proximity. *Criminal Justice and Behavior*, *38*(11), 1069–1088. https://doi.org/10.1177/0093854811418599

Tonkin, Matthew, Grant, T., & 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), 59–77. https://doi.org/10.1002/jip.74

Tonkin, Matthew, & Woodhams, J. (2015). The feasibility of using crime scene behaviour to detect versatile serial offenders: An empirical test of behavioural consistency, distinctiveness, and discrimination accuracy. *Legal and Criminological Psychology*, n/a-n/a. https://doi.org/10.1111/lcrp.12085

Townsley, M., & Sidebottom, A. (2010). All offenders are equal, but some are more equal than others: Variation in journeys to crime between offenders. *Criminology*, *48*(3), 897–917.

Twyman, A. D., Newcombe, N. S., & Gould, T. J. (2009). Of mice (Mus musculus) and toddlers (Homo sapiens): Evidence for species-general spatial reorientation. *Journal of Comparative Psychology*, *123*(3), 342–345. https://doi.org/10.1037/a0015400

Van Daele, S., & Vander Beken, T. (2011a). Out of Sight, Out of Mind? Awareness Space and Mobile Offenders. *European Journal of Crime, Criminal Law & Criminal Justice*, *19*(2), 125–137. https://doi.org/10.1163/157181711X566326

Van Daele, S., & Vander Beken, T. (2011b). Outbound offending: The journey to crime and crime sprees. *Journal of Environmental Psychology*, *31*(1), 70–78. https://doi.org/10.1016/j.jenvp.2010.09.002

Wang, F. (2012). Why police and policing need GIS: an overview. *Annals of GIS*, *18*(3), 159–171. https://doi.org/10.1080/19475683.2012.691900

Ward, J. H. (1963). Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association*, *58*(301), 236–244. https://doi.org/10.2307/2282967

Winter, J. M., Lemeire, J., Meganck, S., Geboers, J., Rossi, G., & Mokros, A. (2013). Comparing the Predictive Accuracy of Case Linkage Methods in Serious Sexual Assaults. *Journal of Investigative Psychology and Offender Profiling*, *10*(1), 28–56. https://doi.org/10.1002/jip.1372

Wolfgang, M. E. (1973). Crime in a Birth Cohort. *Proceedings of the American Philosophical Society*, *117*(5), 404–411.

Wolfgang, M. E., Filio, R., & Tracey, P. (1983). *The 1945 and 1958 Birth Cohorts: A Comparison of the Prevelence, Incidence, and Severity of Deliquent Behavior*. Retrieved from https://www.ncjrs.gov/pdffiles1/Digitization/86680NCJRS.pdf

Wong, D. W. S., & Shaw, S.-L. (2011). Measuring segregation: an activity space approach. *Journal of Geographical Systems*, *13*(2), 127–145. https://doi.org/10.1007/s10109-010-0112-x

Woodhams, J., & Bennell, C. (2015). *Crime Linkage Theory, Research, and Practice*. CRC Press.

Woodhams, J., Hollin, C. R., & Bull, R. (2007). The psychology of linking crimes: A review of the evidence. *Legal and Criminological Psychology*, *12*(2), 233–249. https://doi.org/10.1348/135532506X118631

Youden, W. J. (1950). Index for rating diagnostic tests. *Cancer*, *3*(1), 32–35. https://doi.org/10.1002/1097-0142(1950)3:1<32::AID-CNCR2820030106>3.0.CO;2-3

Zipf, G. (1949). Human behaviour and the principle of least-effort. Cambridge MA edn. *Reading: Addison-Wesley*.

Appendix A - Code Snippets

Figure 57: R Code: Unsupervised Clustering Algorithm

```
AS_iteration = AS_iteration%>%
            group_by(SRN, sIndex)%>%
            do(cluster_check(.))

cluster_check = function(data){
  k = round(sqrt(nrow(data)))
  x = data
  x$rowID = seq.int(nrow(x))
  x=as.data.frame(x)
  while(k>=1){
    if(k==1){
      clusters=x
      clusters$sil_width=1
      clusters$cluster=1
      clusters$neighbor=1
      return(clusters)
    }
    else{
      cluster=eclust(x[,c("CrimeEasting", "CrimeNorthing")],
                    "hclust",
                    k=k,
                    graph=FALSE)
      t=fviz_silhouette(cluster)
      t=t$data
      t=as.data.frame(t)
      clusters=merge(x,t,by.x="rowID",by.y="name")
      clusters$cluster=as.numeric(clusters$cluster)
      if(all(clusters$sil_width>0)){
        return(clusters)
      }
      k=k-1
    }
  }
}
```

Figure 58: R Code: LOF Outlier Detection Algorithm

```
AS_iteration = AS_iteration%>%group_by(SRN,sIndex,cluster)%>%
               do(LOF_offender(.))

LOF_offender=function(data){
 xy = data%>%
      group_by(CrimeEasting,CrimeNorthing)%>%
      sample_n(1)%>%
      select(CrimeEasting,CrimeNorthing)
  k = round(sqrt(nrow(xy)))
  unq = unique(xy)
  xy = as.matrix(xy)
  if(k==1|nrow(unq)/nrow(xy)<=0.5){
    lof=data
    lof$lof=1
  }
  else{
    lof = lofactor(xy, k)
    lof = cbind(xy, lof)
    lof = as.data.frame(lof)
    lof = inner_join(data,lof)
  }
  return(lof)
}
```

Figure 59: R Code: Activity Space Geometry Estimation Algorithm

```
centroids=AS_iteration%>%
filter(lof<1.5)%>%
ungroup()%>%
group_by(SRN, cluster,sIndex)%>%
   summarise(
     Midx=mean(CrimeEasting),
     Midy=mean(CrimeNorthing))

maxDist=AS_iteration%>%
filter(lof<1.5)%>%
inner_join(centroids,by=c(
      "SRN"="SRN", "cluster"="cluster", "sIndex"="sIndex")
      )
maxDist=maxDist%>%
  mutate(dist=sqrt((Midx-CrimeEasting)^2+(Midy-CrimeNorthing)^2))%>%
  group_by(SRN,cluster,sIndex)%>%
  summarise(
    Midx=mean(Midx),
    Midy=mean(Midy),
    MaxDist=max(dist))
```

Figure 60: R Code: Random Forest Model Specification

```
library(h2o)
h2o.init(nthreads = 10)

train.h2o = as.h2o(train,destination_frame = "train.h2o")
test.h2o = as.h2o(test,destination_frame = "test.h2o")

test.h2o$Linked = as.factor(test.h2o$Linked)
train.h2o$Linked = as.factor(train.h2o$Linked)
test.h2o = h2o.assign(test.h2o,key = "test.h2o")
train.h2o = h2o.assign(train.h2o,key = "train.h2o")

vfinal.h2o = as.h2o(vfinal,destination_frame = "vfinal.h2o")
vfinal.h2o$Linked = as.factor(vfinal.h2o$Linked)
vfinal.h2o = h2o.assign(vfinal.h2o,key = "vfinal.h2o")

#Select predictors of choice#
#Model 1#
X1 = c("CrimetoCrimeDistance","minCentroidtoCrimeDist","DaysElapsed","CrimePairPr",
"SameLocationType","ASinclusion_total")

#Model 2#
X2=c(4:7,10,13:14,28,33,38,39,40:52,54,59:359,362:407,412:423,426:708,710:755)

rforest = h2o.randomForest(
  x = X1,
  y = "Linked",
  model_id = "rforest",
  training_frame = train.h2o,
  mtries = -1,
  min_rows = 10,
  max_depth = 300,
  ntrees = 200,
  nfolds = 10,
  validation_frame = test.h2o)
```

Appendix B - Exploratory Graphs

Linkage Model Predictor Estimates

Figure 61: Quartile Distribution of Linked Status by Predictor(s)

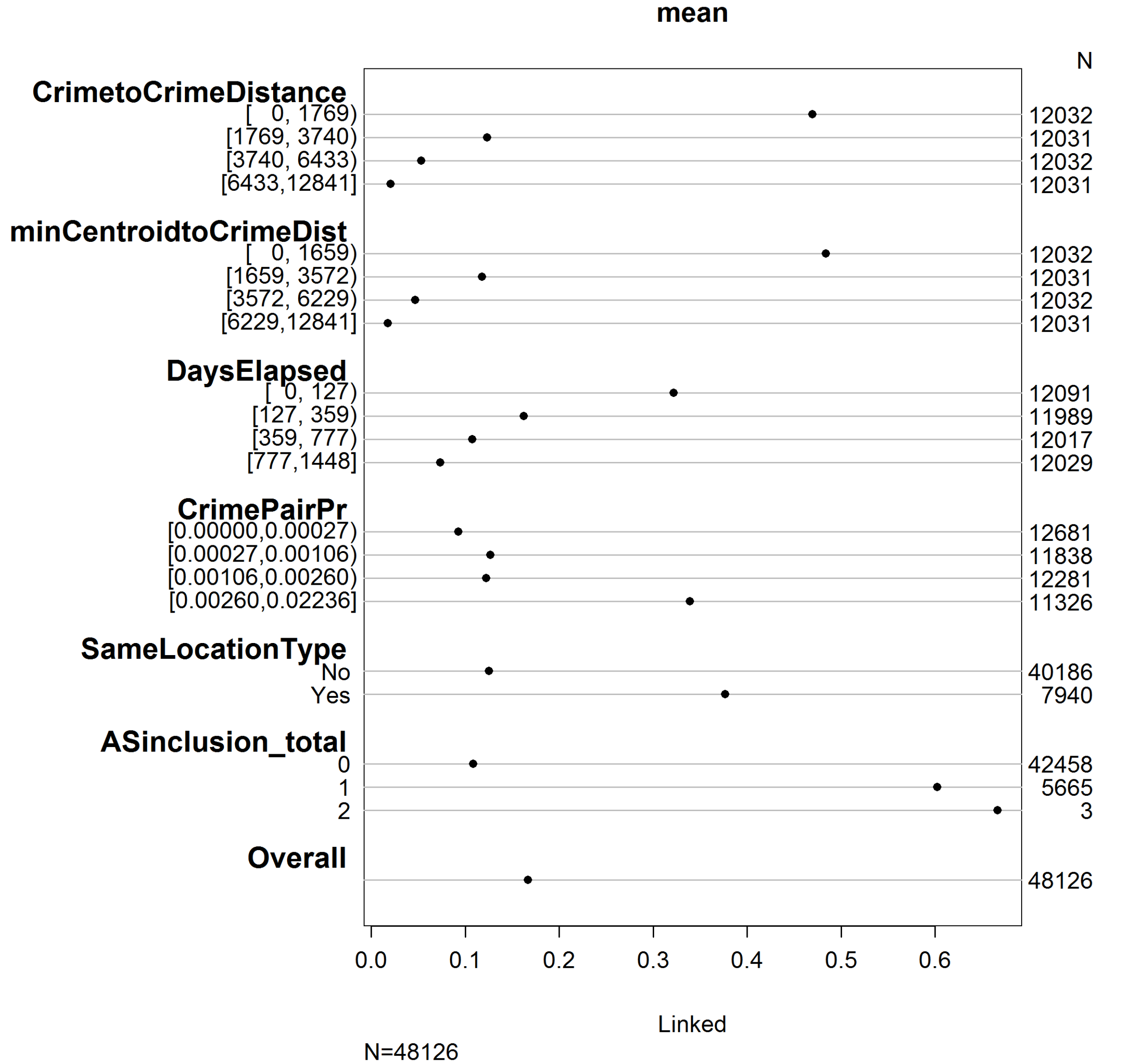


Figure 62: Non-Parametric Inspection of Linked Probability Change by Predictor

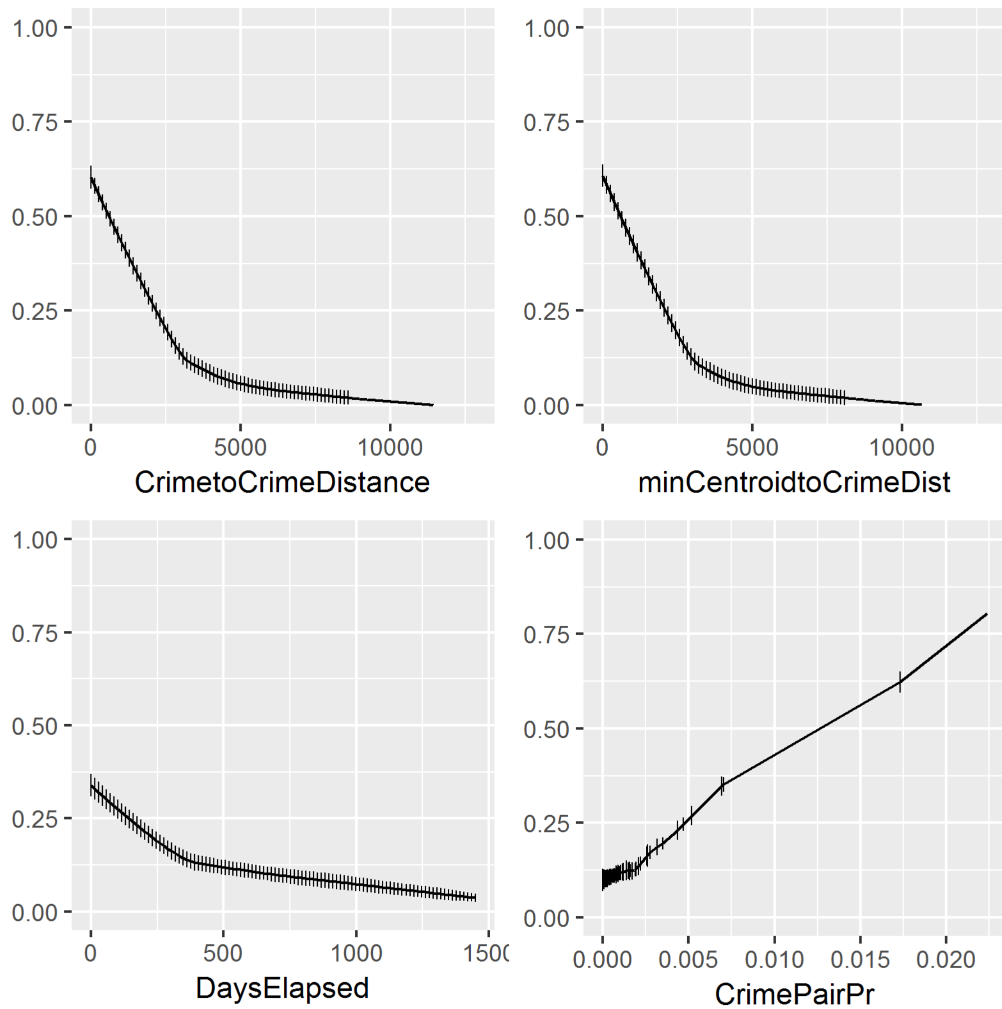
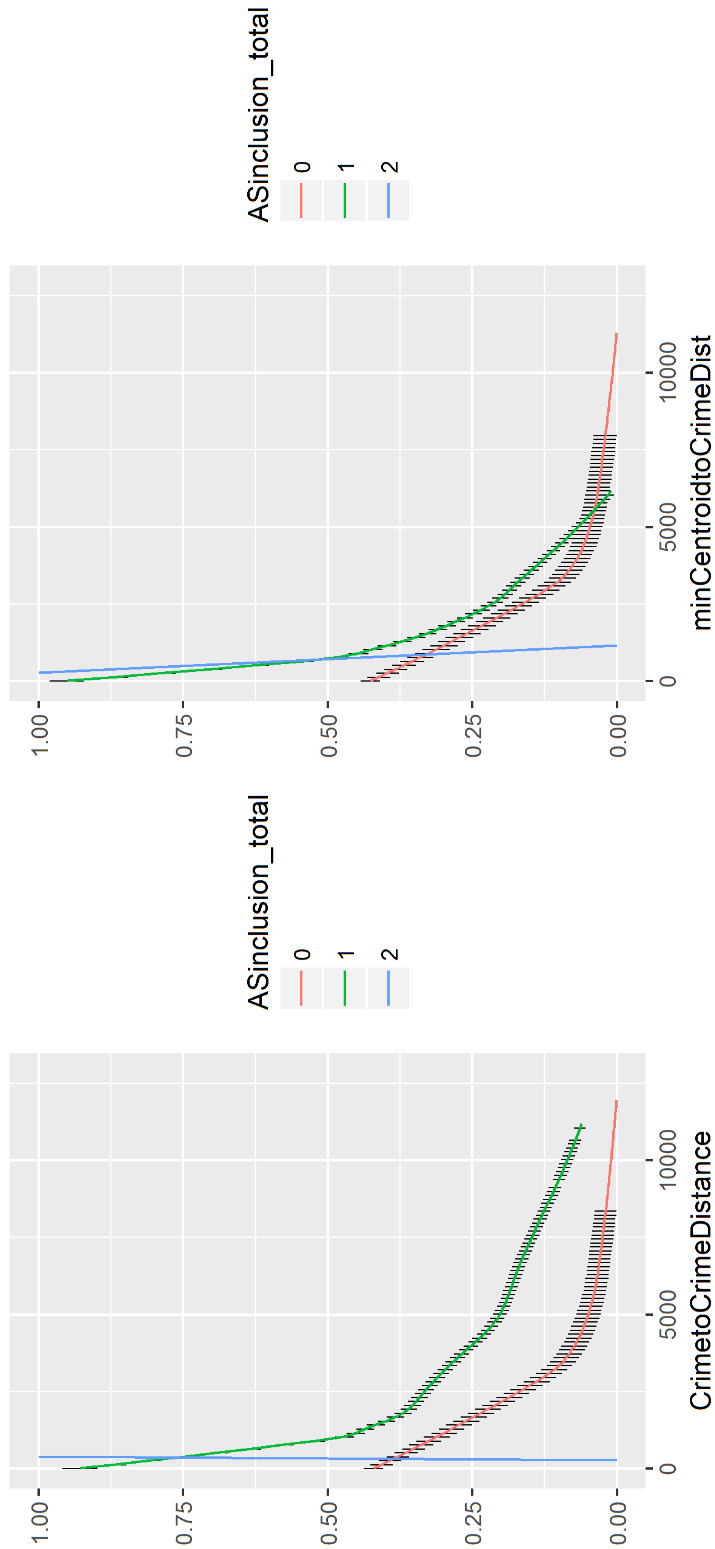


Figure 63: Predictor Interaction Estimates from Non-Parametric Test



Appendix C - Model Fit Diagnostic Graphs

Figure 64: M1 - ROC Curve

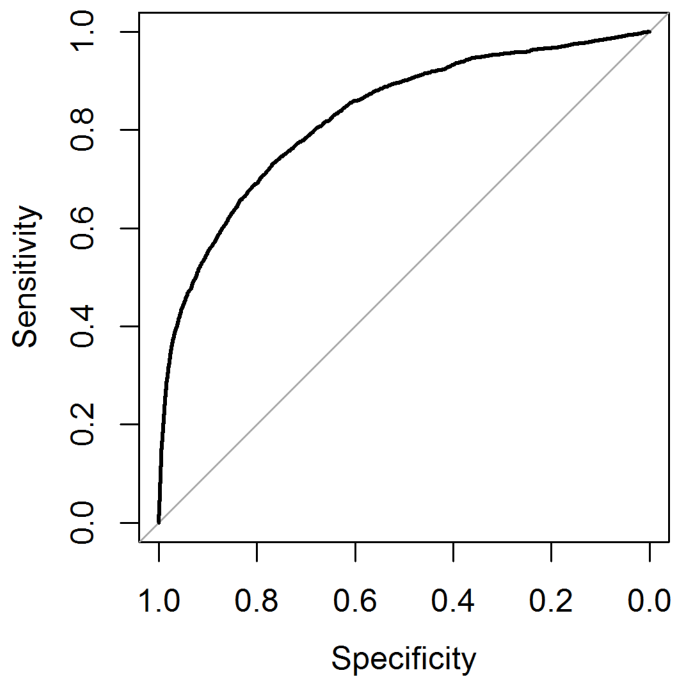


Figure 65: M2 - ROC Curve

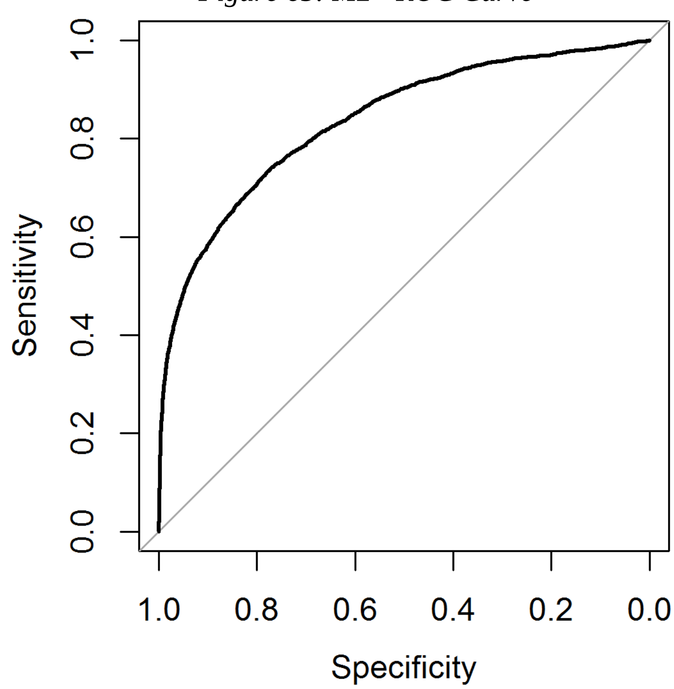


Figure 66: M3 - ROC Curve

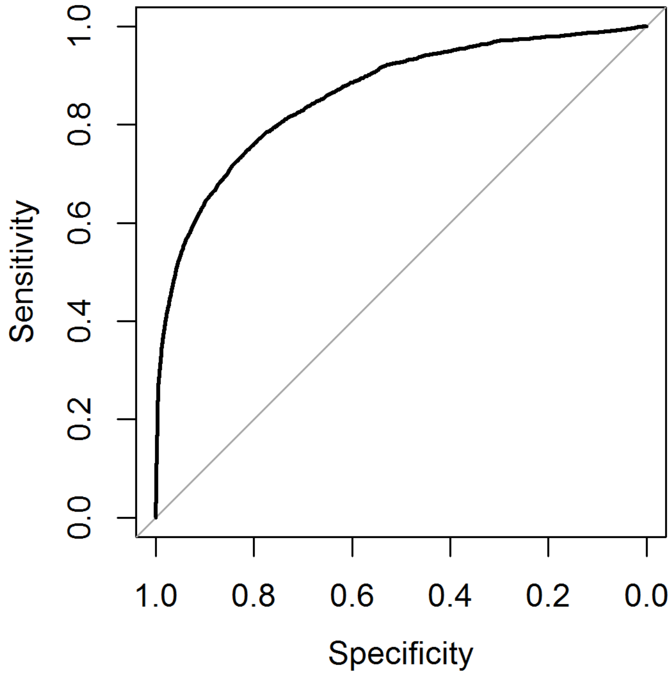


Figure 67: M1 – Train Data Score by Class Density Distribution with Youden's Index Decision Threshold

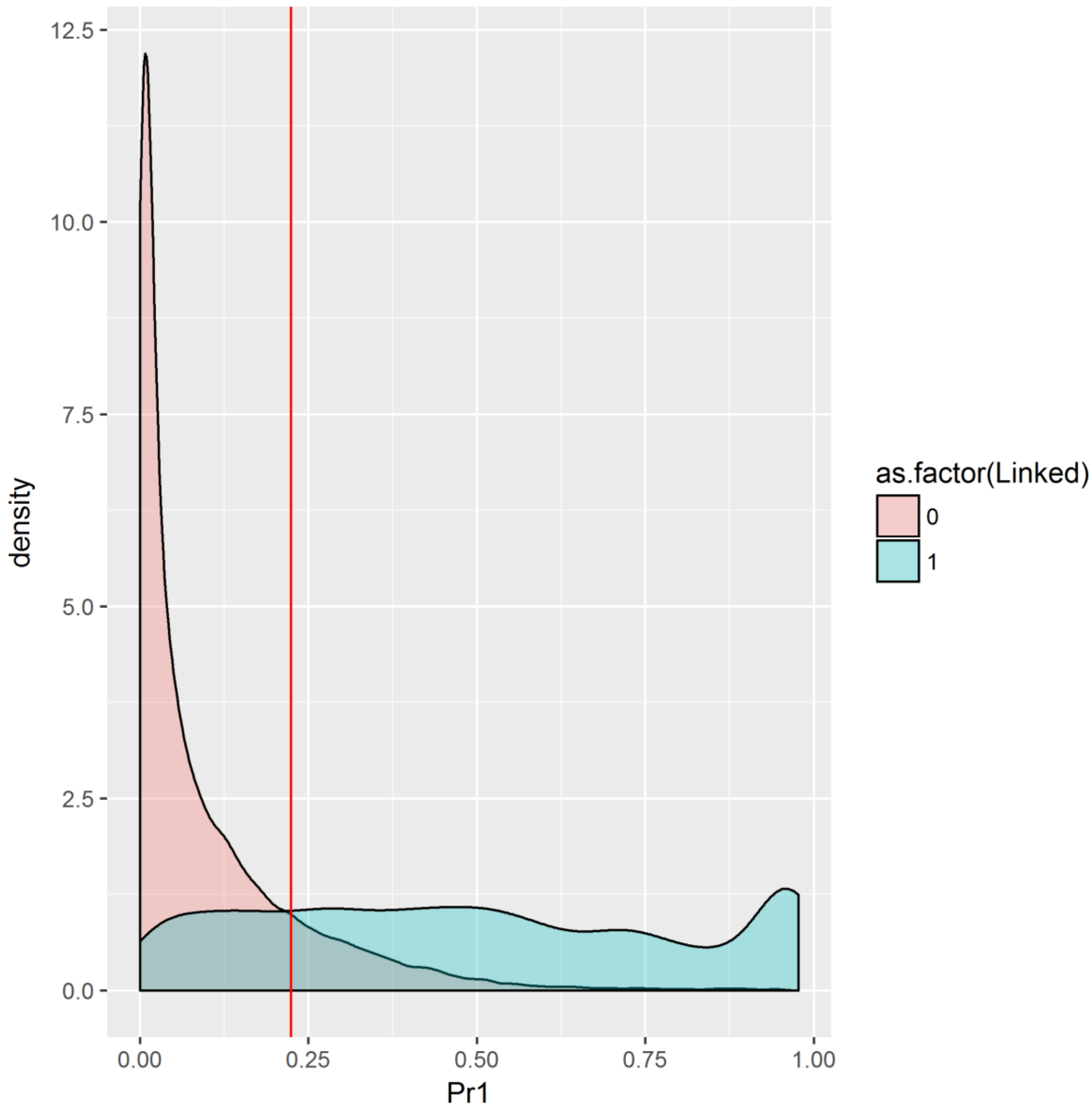


Figure 68: M1 - Test Data Score by Class Density Distribution with Youden's Index Decision Threshold

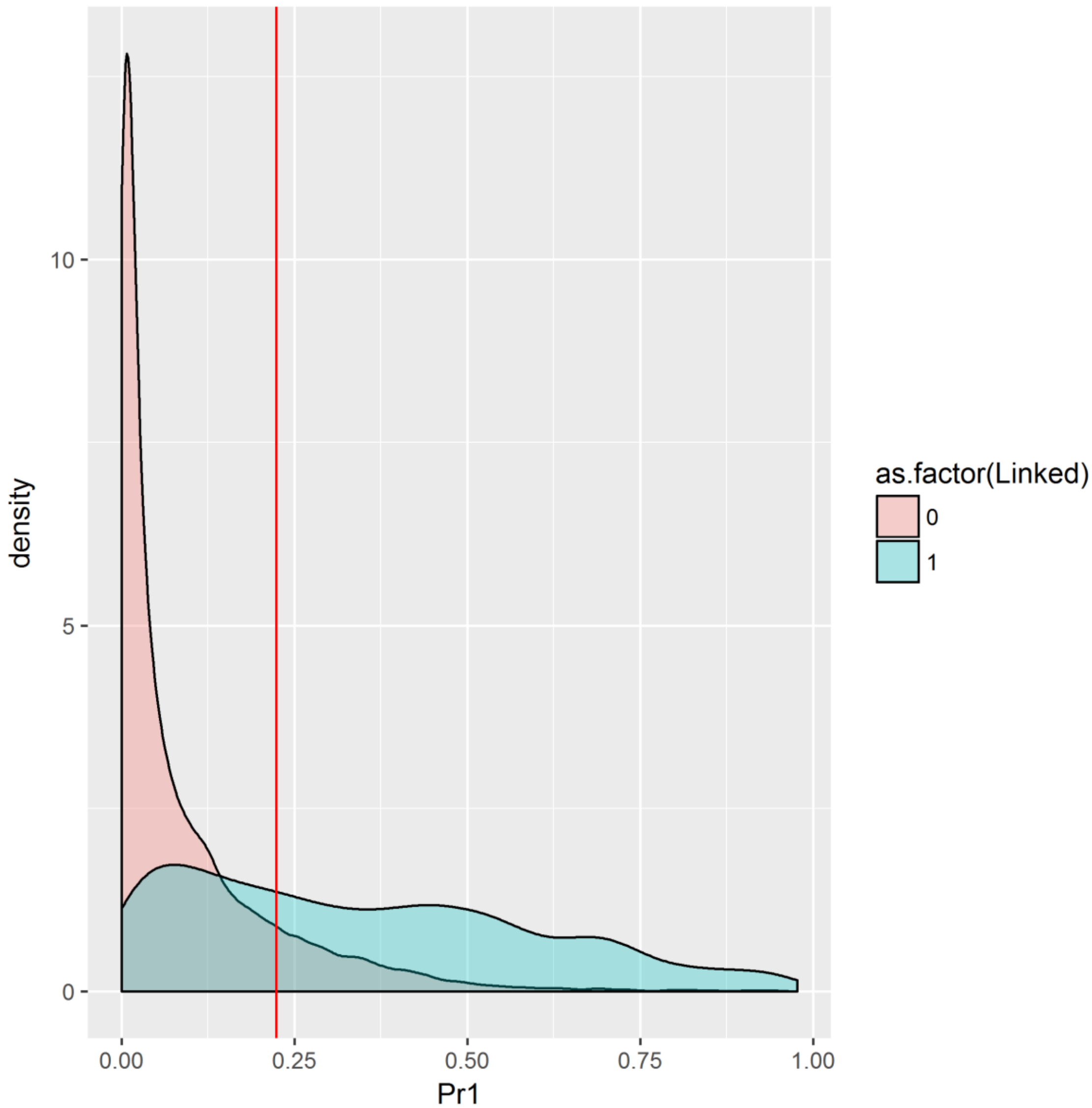


Figure 69: M1 - Validation Data Score by Class Density Distribution with Youden's Index Decision Threshold

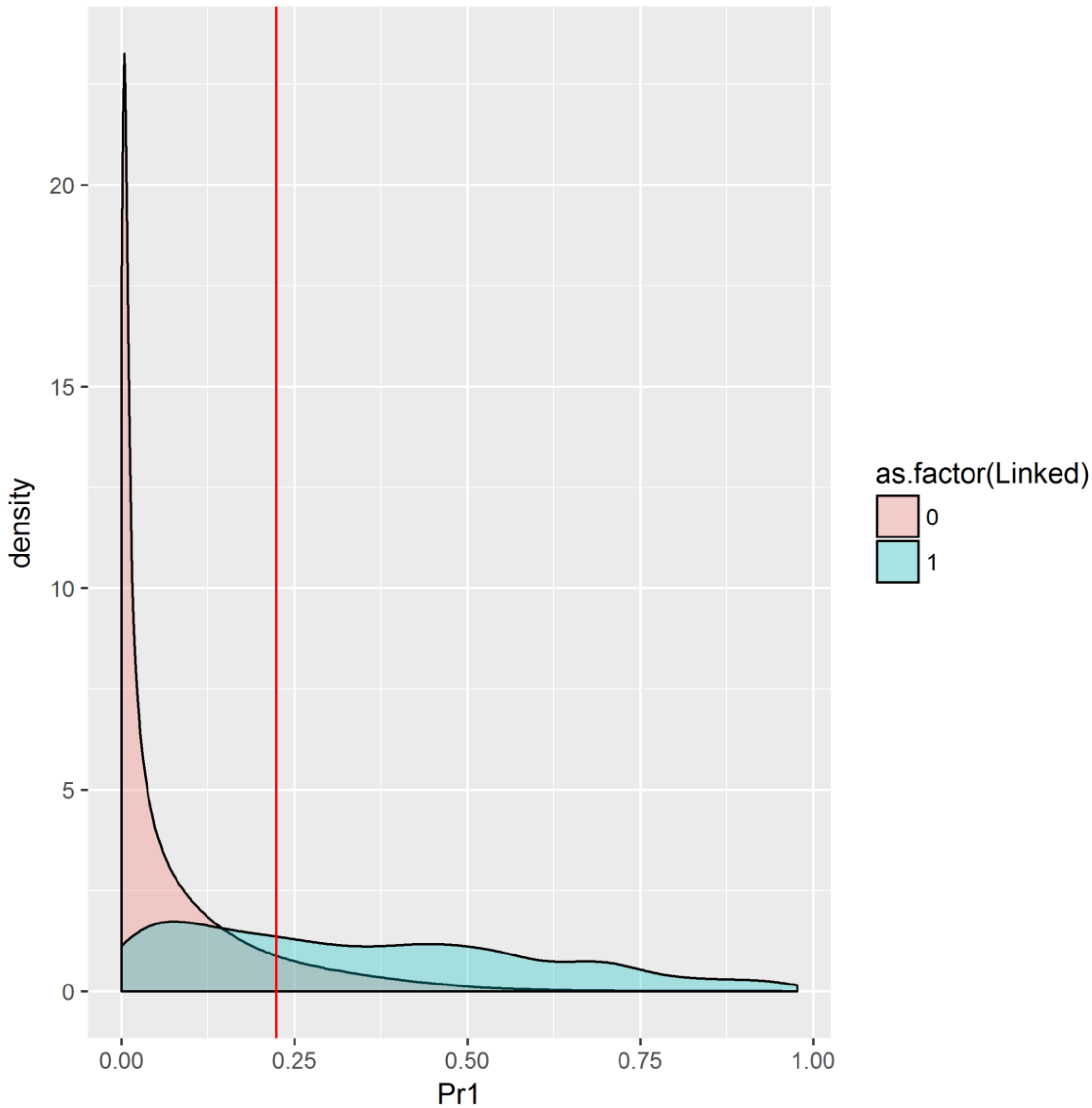


Figure 70: M2 - Train Data Score by Class Density Distribution with Youden's Index Decision Threshold

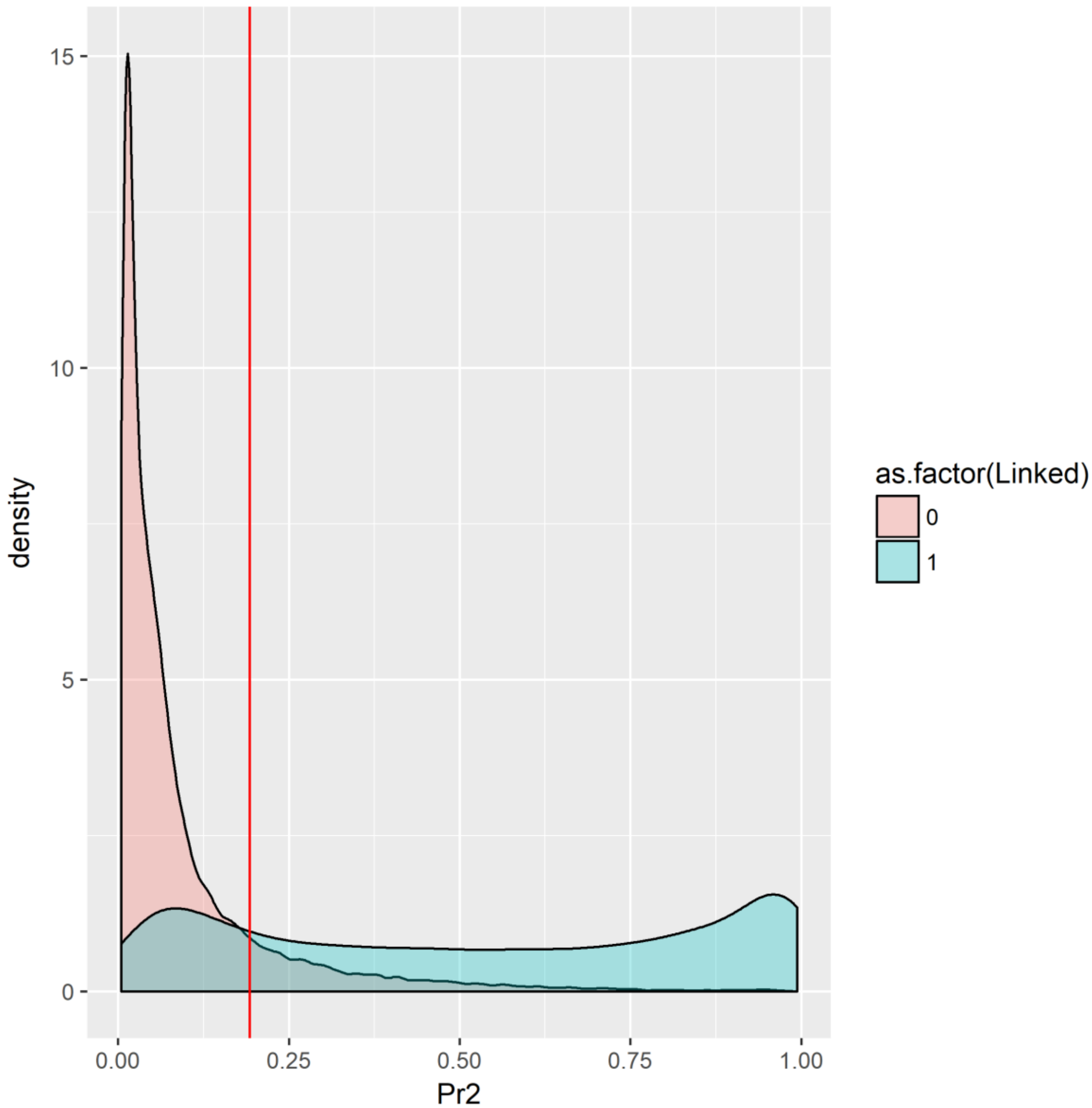


Figure 71: M2 - Test Data Score by Class Density Distribution with Youden's Index Decision Threshold

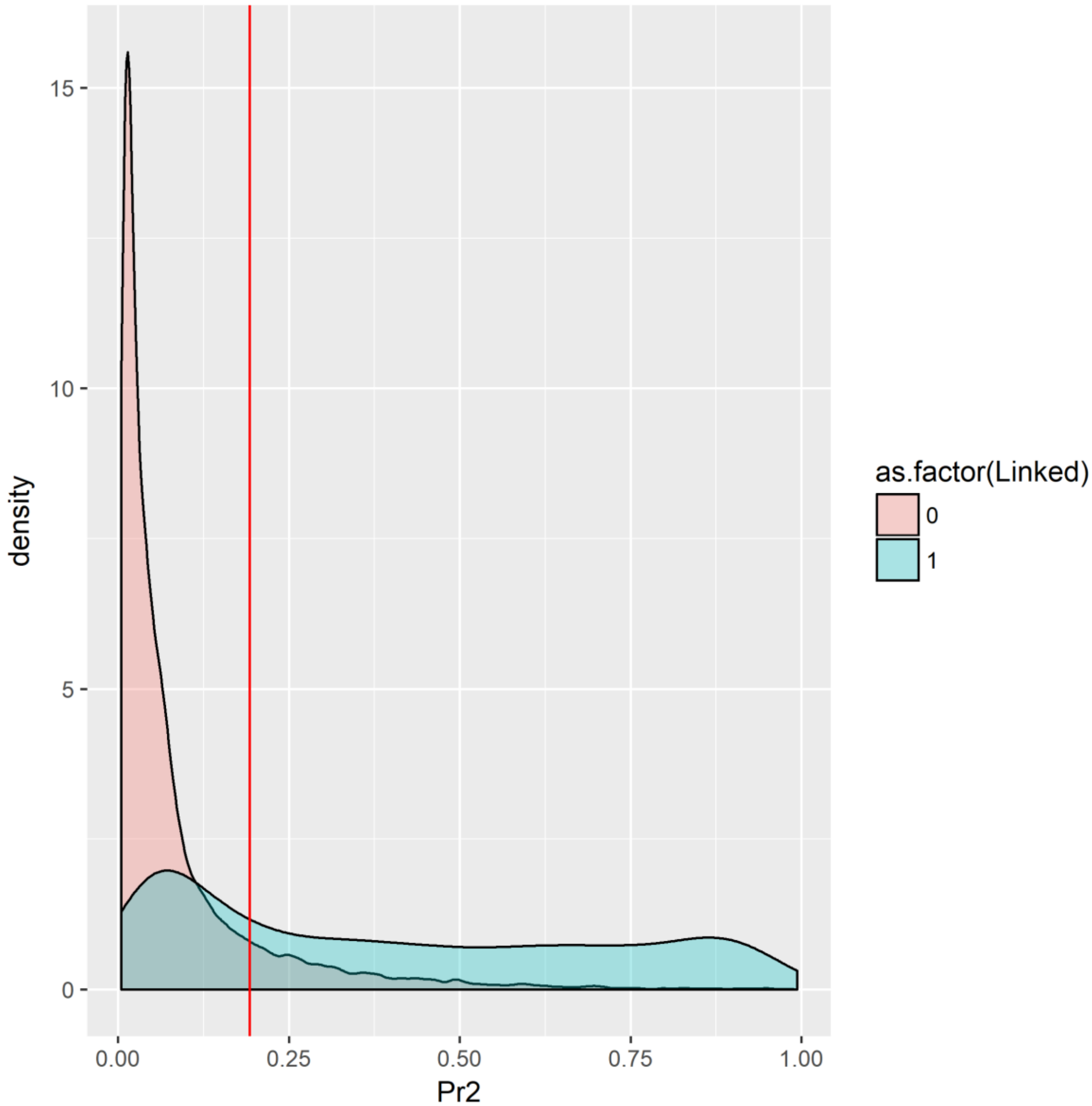


Figure 72: M2 - Validation Data Score by Class Density Distribution with Youden's Index Decision Threshold

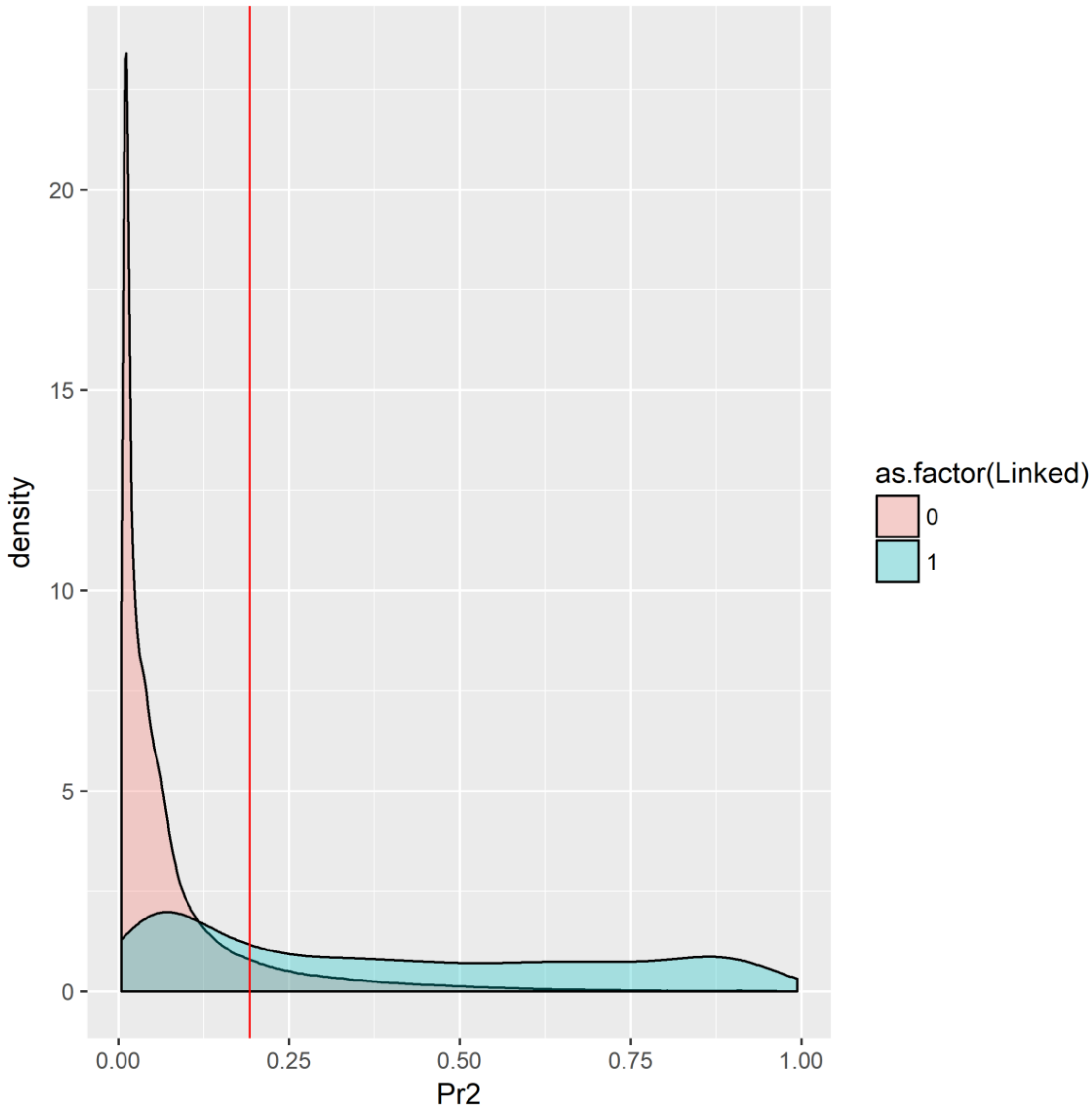


Figure 73: M3 - Train Data Score by Class Density Distribution with Youden's Index Decision Threshold

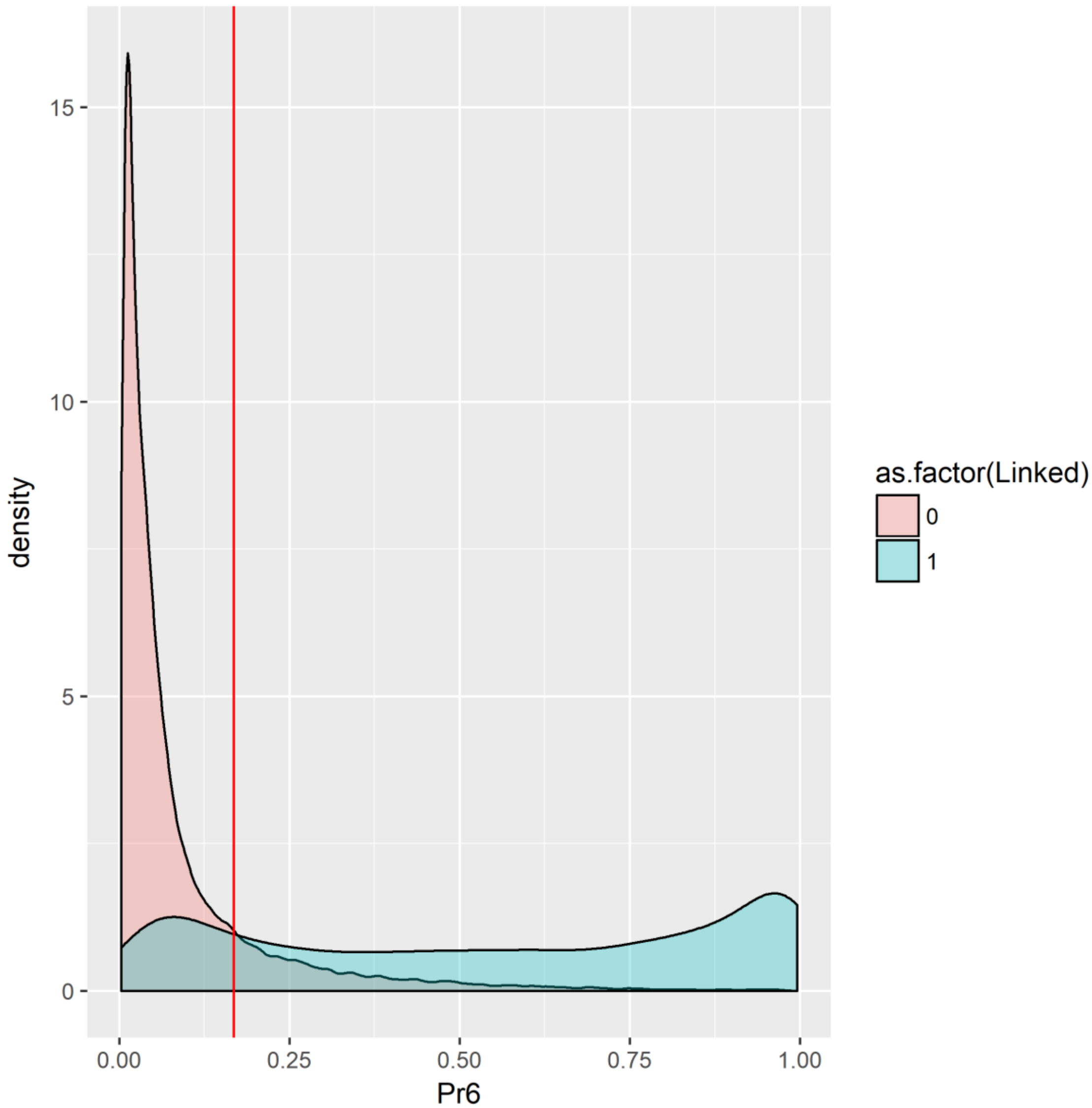


Figure 74: M3 - Test Data Score by Class Density Distribution with Youden's Index Decision Threshold

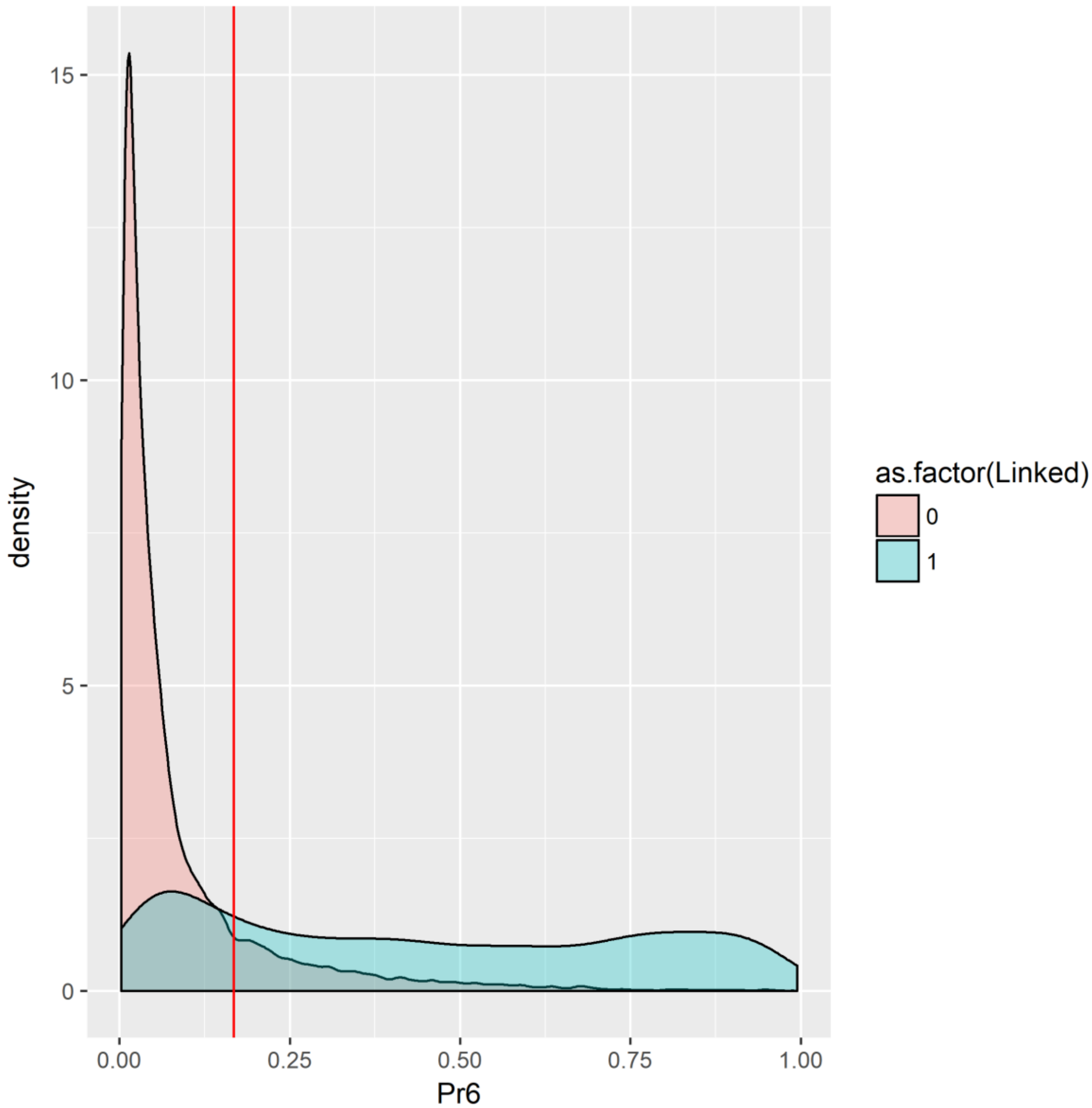


Figure 75: M3 - Validation Data Score by Class Density Distribution with Youden's Index Decision Threshold

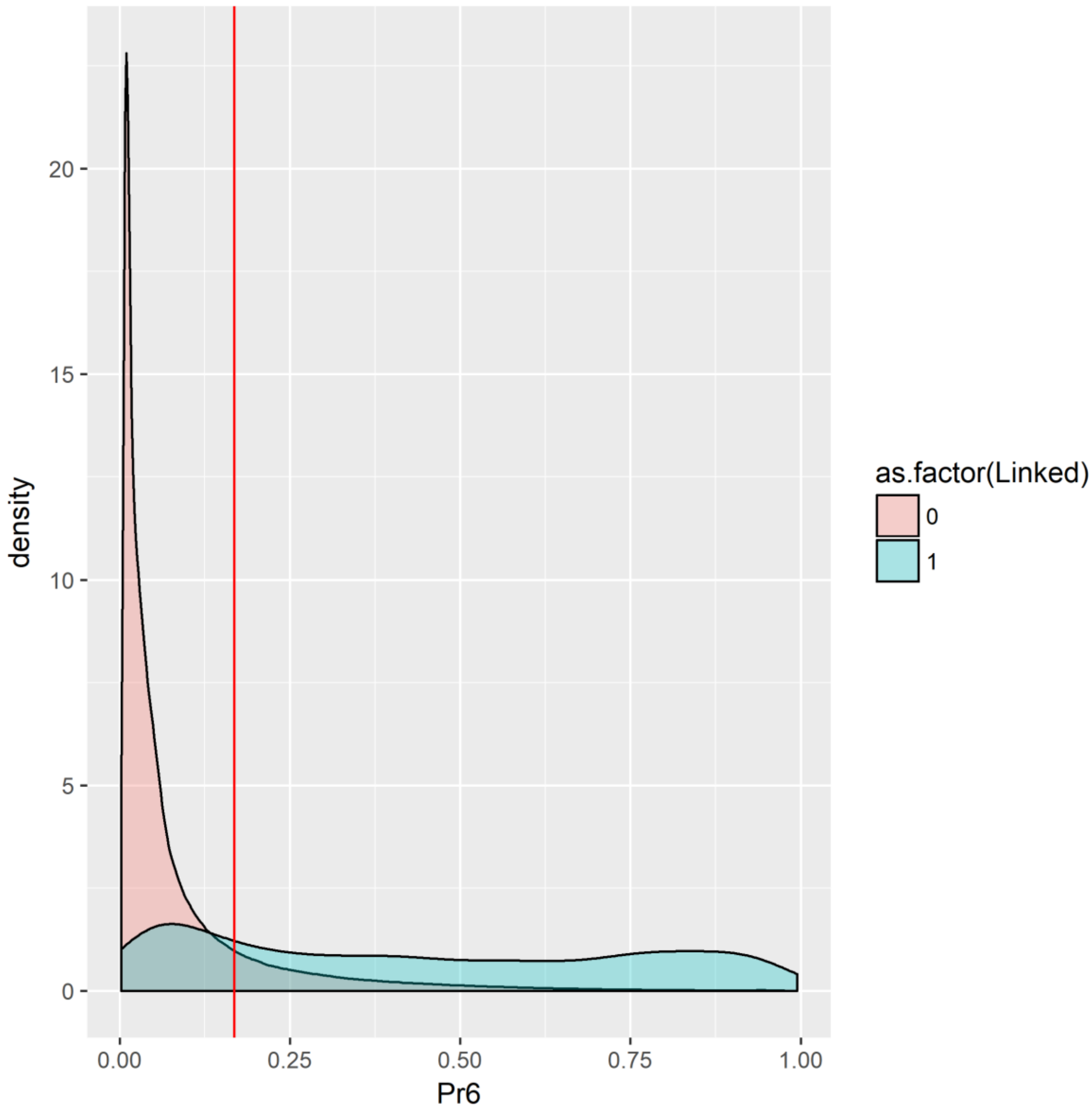


Table 21: Logistic Regression Model(s) Binary Classification Confusion Tables

			Linked Prediction	
			False	True
M1 - Train	Linked Actual	False	34099	6006
		True	1949	6072
M2 - Train	Linked Actual	False	34668	5437
		True	2015	6006
M3 - Train	Linked Actual	False	34018	6087
		True	1736	6285
M1 - Test	Linked Actual	False	14964	2281
		True	1364	2085
M2 - Test	Linked Actual	False	15063	2182
		True	1291	2158
M3 - Test	Linked Actual	False	14645	2600
		True	1002	2447
M1 - Validate	Linked Actual	False	2984913	450291
		True	1364	2085
M2 - Validate	Linked Actual	False	3007800	427404
		True	1291	2158
M3 - Validate	Linked Actual	False	2919245	515959
		True	1002	2447

Appendix E

Table 22: Random Forest Model(s) Binary Classification Confusion Table

Model	Data Set			Linked Prediction	
				0	1
Simple Model	Train	Linked Actual	0	37476	2629
			1	1546	6475
Full Model	Train	Linked Actual	0	38082	2023
			1	1740	6281
Simple Model	Test	Linked Actual	0	15976	1269
			1	1399	2050
Full Model	Test	Linked Actual	0	16151	1094
			1	1533	1916
Simple Model	Validation	Linked Actual	0	3179672	255532
			1	1399	2050
Full Model	Validation	Linked Actual	0	3210355	224849
			1	1533	1916

Appendix E - Crime Group to Subgroup Dictionary

Table 23: Crime Sub-Classifications by Crime Group

Crime Sub-Classification	Crime Group
Aggravated burglary dwell	Aggravated Burglary Dwell
Burglary in a Dwelling and Attempted Burglary in a Dwelling	Burglary Dwelling
Burglary in a dwelling	Burglary Dwelling
Distraction burglary in a dwelling	Burglary Dwelling
Burglary in a Building other than a Dwelling and Attempted Burglary in a Building other than a Dwelling	Burglary Non-dwelling
Burglary in a building other than a dwelling	Burglary Non-dwelling
Arson Not Endangering Life	Criminal Damage
Arson endangering life	Criminal Damage
Arson not endangering life	Criminal Damage
Criminal Damage Endangering Life	Criminal Damage
Criminal Damage to a Building other than a Dwelling	Criminal Damage
Criminal Damage to a Dwelling	Criminal Damage
Criminal Damage to a building other than a dwellin	Criminal Damage
Criminal Damage to a dwelling	Criminal Damage
Criminal Damage to a vehicle	Criminal Damage
Other Criminal Damage	Criminal Damage
R or R aggrv Crim Dam to a building not a dwelling	Criminal Damage
Threats to commit crimina	Criminal Damage
Forgery re drugs	Drug Offences
Possession of Controlled Drugs (Cannabis)	Drug Offences
Possession of Controlled Drugs (excluding Cannabis)	Drug Offences
Possession of controlled drugs (Cannabis)	Drug Offences
Possession of controlled drugs (excl. Cannabis)	Drug Offences
Trafficking in Controlled Drugs	Drug Offences
Trafficking in controlled drugs	Drug Offences
Forgery - other	Forgery
Cheque and credit card fraud	Fraud
Fraud by abuse of position	Fraud
Fraud by false representation: other frauds	Fraud
Fraud forgery etc associated with vehicle or driver records	Fraud
Making or supplying articles for use in fraud	Fraud
Other Frauds	Fraud
Possession of articles for use in fraud	Fraud
Absconding from Lawful Custody	Other Offences

Absconding from custody	Other Offences
Abstracting electricity	Other Offences
Attempt pervert course ju	Other Offences
Blackmail	Other Offences
Cruelty/Neglect of childr	Other Offences
Dangerous Driving	Other Offences
False Statements	Other Offences
Going Equipped for Stealing etc.	Other Offences
Going equipped for steali	Other Offences
Immoral procurement	Other Offences
Incest	Other Offences
Kidnapping	Other Offences
Obscence Publications etc	Other Offences
Other	Other Offences
Other Notifiable Crime2	Other Offences
Other Offences against the State & Public Order	Other Offences
Perverting the Course of Justice	Other Offences
Possession of Article with Blade or Point	Other Offences
Possession of article with blade or point	Other Offences
Possession of firearms offences	Other Offences
Possession of items to endanger life	Other Offences
Possession of weapons	Other Offences
Proceeds of crime	Other Offences
Sex Offender failed to Register	Other Offences
Stealing by employee	Other Offences
Trade Descriptions etc	Other Offences
Use of substance or object to endanger life	Other Offences
Vehicle interference and tampering	Other Offences
Violent Disorder	Other Offences
Robbery of Personal Property	Robbery
Robbery of business property	Robbery
Robbery of personal property	Robbery
Abuse of children through prostitution / pornography	Sexual Offences
Exposure and Voyeurism	Sexual Offences
Exposure and voyeurism	Sexual Offences
Incest	Sexual Offences
Other miscellaneous sexual offences	Sexual Offences
Rape of a female aged 16 and over	Sexual Offences
Rape of a female child under 13	Sexual Offences
Rape of a female child under 16	Sexual Offences
Rape of a male aged 16 and over	Sexual Offences

Rape of a male child under 13	Sexual Offences
Sexual activity involving child under 16	Sexual Offences
Sexual activity with a person with mental disorder	Sexual Offences
Sexual assault on a female aged 13 and over	Sexual Offences
Sexual assault on a female child under 13	Sexual Offences
Sexual assault on a male aged 13 and over	Sexual Offences
Sexual assault on a male child under 13	Sexual Offences
Sexual grooming	Sexual Offences
Trafficking for sexual exploitation	Sexual Offences
USI with girl under 13	Sexual Offences
Aggravated vehicle taking	Theft and Handling Stolen Goods
Handling	Theft and Handling Stolen Goods
Handling Stolen Goods	Theft and Handling Stolen Goods
Misc. Thefts	Theft and Handling Stolen Goods
Other Theft	Theft and Handling Stolen Goods
Shoplifting	Theft and Handling Stolen Goods
Steal mail bags	Theft and Handling Stolen Goods
Theft from M/Meters	Theft and Handling Stolen Goods
Theft from a vehicle	Theft and Handling Stolen Goods
Theft from motor vehicle	Theft and Handling Stolen Goods
Theft from person	Theft and Handling Stolen Goods
Theft from the Person	Theft and Handling Stolen Goods
Theft in dwelling	Theft and Handling Stolen Goods
Theft of motor vehicle	Theft and Handling Stolen Goods
Theft of pedal cycle	Theft and Handling Stolen Goods
Theft or Unauthorised Taking of a Pedal Cycle	Theft and Handling Stolen Goods
Actually bodily harm and other injury	Violence Against the Person
Affray	Violence Against the Person
Assault without injury on a constable	Violence Against the Person
Common assault	Violence Against the Person
Harassment	Violence Against the Person
Inflicting grievous bodily harm without intent	Violence Against the Person
Possession of Other Weapons	Violence Against the Person
Possession of other weapons	Violence Against the Person
Public Fear_ Alarm or Distress	Violence Against the Person
Public fear alarm or distress	Violence Against the Person
Rac. or relig. aggrv Criminal Damage to a vehicle	Violence Against the Person
Rac. or relig. aggrv Other Criminal Damage	Violence Against the Person
Racially and Religiously aggravated common assault	Violence Against the Person
Racially or Religiously Aggravated Public Fear_ Alarm or Distress	Violence Against the Person

Racially or religiously aggravated actual bodily harm and other injury	Violence Against the Person
Racially or religiously aggravated harassment	Violence Against the Person
Racially or religiously aggravated public fear alarm or distress	Violence Against the Person
Abduction of Children	Violence Against the Person (Serious)
Attempted Murder	Violence Against the Person (Serious)
Causing death by careless or inconsiderate driving	Violence Against the Person (Serious)
Causing death by driving: unlicensed drivers etc.	Violence Against the Person (Serious)
Infanticide	Violence Against the Person (Serious)
Kidnap/Hijack	Violence Against the Person (Serious)
Manslaughter	Violence Against the Person (Serious)
Murder	Violence Against the Person (Serious)
Possession of firearms with intent	Violence Against the Person (Serious)
Threats to Kill	Violence Against the Person (Serious)
Threats to kill	Violence Against the Person (Serious)
Wounding or carrying out an act endangering life	Violence Against the Person (Serious)