Evaluation of the MPS Predictive Policing Trial (June 2015)

Prepared for:

PS Ross Gibbings, MPS; Det. Chief Supt. Richard List, Thames Valley Police; DS Colin McLoughlin, MPS; Insp Andre Ramsey, MPS.

Prepared by:

Bryant, R*., Azhar, H., Blackburn, B., & Falade, M. , School of Law, Criminal Justice & Computing, Canterbury Christ Church University

This report is a **redacted** version of an evaluation of an MPS trial of predictive policing in London; it includes the aspects of both predictive accuracy and operational implementation. The redactions relate to passages containing information which might enable the identification of commercial products and services (supplied free of charge to the MPS) but subject to non-disclosure agreements between all parties participating within this predictive policing research.

The report represents the views of the authors' and not necessarily those of Canterbury Christ Church University or the Metropolitan Police Service.

CONTENTS

* correspondence: robin.bryant@canterbury.ac.uk

1   Introduction

This report is an evaluation of a trial of predictive policing by the MPS in the Greater London Area. Since 2013, MPS Territorial Policing, Capability and Support (TP C & S) has been trialling a number of predictive policing initiatives in the capital. By late 2014, three commercial predictive policing products and one 'in house' (MPS analyst) crime forecasting model ('MBR') were being used to support operational decision-making in most London boroughs.   The commercial products were each allocated two Borough Operational Command Units (BOCU) in which to operate. The MBR (in principle at least) operated in the remaining 26 of the 32 MPS BOCUs.

The methods used to conduct this evaluation were as follows.

1.1   Literature review

A review of publications, research, media coverage relating to 'predictive policing', crime forecasting and measures of predictive accuracy.

1.2   Deskbound research

Analysis of other research conducted into predictive policing, summarising evaluations, analysing the spatial and temporal distribution of crime in a number of localities, analysing dosage rates, comparisons with Kent Police evaluations (unpublished), the Police Service of the Netherlands (unpublished) and analysing stakeholder survey results collected by MSc postgraduate students and by the MPS.

1.3   Assessment of predictive accuracy

As part of the evaluation we assessed the hit rates and predictive accuracy indices of the crime forecasting products in two towns outside London: Reading and Slough. This involved the writing of Matlab code to identify the number of successful forecasts, the use of ArcGIS software to analyse the spatial and temporal distribution of crime in Reading and Slough and the analysis of crime data provided by Thames Valley Police.

1.4   MSc research

We have summarised the findings of six MSc dissertations and a single MSc dissertation undertaken by a university researcher during 2014. All of the MSc dissertations discussed, inter alia, five aspects of predictive policing in the MPS: theoretical background, predictive accuracy, operational implementation, 'patrol dosage' and user opinion.

2   Overview of Predictive Policing

In this section of the evaluation we describe the underpinning theory associated with predictive policing, together with a discussion of predictive policing as an operational police tactic.

2.1   'Predictive policing'

Before discussing the main theories surrounding predictive policing it is important to discuss what exactly is meant by 'predictive policing' and how it might differ from the more established field of 'crime forecasting'. It is clear that mainstream media and some police forces have interpreted 'predictive policing' as meaning the application of crime forecasting to accurately and precisely 'predict' crime using 'scientific' (mathematical) algorithms performed using computing technology. Ratcliffe (2014, p.4) however, describes predictive policing as the '[…] use of historical data to create a spatiotemporal forecast of areas of criminality or crime hot spots that will be the basis for police resource allocation decisions […]'. It is this coupling of operational decision making with crime forecasting that forms the basis of any claim of novelty for 'predictive policing'.

The crime forecasting aspects of predictive policing tend to be applied only to those crimes whose occurrence exhibits some mathematical relationship with physical space and particularly public places such as town centres (for example, crimes of acquisition).   As far as we are aware, there are no predictive policing algorithms that forecast crimes with few correlations with public spaces (such as corporate fraud). Further, predictive policing, as a commercial product, is often predicated on the assumption that deploying police resources (particularly visible patrol) to relatively small geographical areas at times that are forecast to have high likelihoods of experiencing crime will have a dissuasive effect, but at the same time not lead to significant displacement of crime to nearby areas. Certainly many of the companies involved in selling predictive policing to police forces claim net reductions in crime that have resulted from the use of their algorithms and predictions, a typical assertion being a ''21% drop in violent crime, a 28% decrease in property crime, a 50% drop in residential burglaries and a 34% decrease in vehicle theft as compared to the same period last year' (PredPol, 2015a). There is little discussion by the companies concerned that deploying police resources to areas that are forecasted to experience crime might instead result in an *increase* in detection and sanction rates and, somewhat ironically, lead to an increase in recorded crime.

2.2   Background

Many of the theories and concepts underpinning much current predictive policing date back a number of decades, and include routine activity theory, patterns of spatial and temporal victimisation and distance decay effects. The genuinely 'new' theories employed are also reasonably well known in other disciplines, or within the private sector where they are mainly used to 'map' consumer behaviour. The algorithms employed by predictive policing products may look complex but they are almost always based on theories that are relatively simple to understand.

The application of both 'new' and 'old' theories to forecast crime are also not new per se, but the popularity of (or at least, interest in) these techniques with police forces has surged in the past decade, in part prompted by high profile uses by the LAPD and others (Perry et al 2013).

One of the reasons for this interest is that these new techniques offer the possibility of identifying non-obvious mathematical associations. These relationships are often counter-intuitive and difficult to spot without the assistance of statistical analyses or considerable experience. The growth in NORA (non-obvious relationship analysis) is not limited to the forecast of crime; the US chain Wal-Mart famously predicted that they should increase their stocks of pop-tarts, bottled water and duct-

tape when major weather events are predicted, as demand would increase considerably for these products (Katrandjian, 2011). This is a good example of the kind of insight these new techniques can provide, with the change in demand for bottled water and duct tape seeming obvious to someone familiar with shopping patterns, but the increase in demand for pop-tarts being a relationship that could only be revealed through statistical analysis of shopping behaviour. A well-known example of NORA being used by a police force was carried out in Richmond Virginia, where it detected a relationship between complaints about gunfire on New Year's Eve. Simple statistical analysis showed that the bulk of these calls were in a short period of time around NYE and in a small number of blocks. By redistributing their officers in this short time window Richmond police were able to reduce overtime costs, reduce complaints and increase weapons seizures (Sklansky, 2011).

The past plays a significant part in all predictive policing, acting as a 'prologue' to the present and allowing analysts to draw conclusions based on previous events. Merkin's maxim of assuming that present trends will continue lies at the heart of predictive policing, with trends in location being used to predict the geo-spatial location of future crime, and the trends in time and date allowing analysts to predict the temporal locations (Perry et al., 2013). These trends are only as strong as the data used for the analysis; data that is incomplete, short or censored is of little use for making accurate predictions. When considering the data used within predictive policing, it is worth considering that more data is not always going to yield the best results. The most useful is the data which is most applicable to the problem. Consider an analyst predicting crime in Lambeth: the locations of every crime reported in the past thirty years would probably be of little use for predicting the locations of current offending, as crime location trends will probably have changed a number of times over the thirty year period. But the crime data for the past thirty days would most certainly be relevant.

2.3   Rational Choice Theory and Routine Activity Theory

From a criminological perspective the major theoretical backbone to predictive policing involves the rational behaviour of an offender and the effect routine activities have on the offending process. These are expressed in theory as Routine Activity Theory (RAT) and Rational Choice Theory (RCT).

2.3.1   Rational Choice Theory (RCT)

Rational Choice Theory is an overarching theoretical perspective on offending that places most emphasis on the choices made by the offender. It is argued that the offender will follow the utilitarian path of risk/reward when considering whether or not to commit an offence, with offences that are seen as being low risk and high reward being most attractive. This makes it easier to make accurate predictions about behaviour as a statistically significant number of offenders will offend only when the cost/benefit ratio is in their favour (Newburn, 2013).

2.3.2   Routine Activity Theory (RAT)

Developed by Cohen and Felson (1979), RAT argues that there are three main components of a criminal act: a motivated offender, a suitable target and the absence of a capable guardian. The convergence of these three components within the same space and time means that crime is more likely to occur, although crucially, not certain to happen. An understanding of the role these three factors play for a potential crime incident can help with formulating plans to disrupt this relationship. Within predictive policing the aim is to utilise predictions to send capable guardians (police officers) to locations where a potential offender and a suitable target are likely to be present (Perry et al., 2013). RAT also proposes that the offender is likely to encounter a suitable target as a result of routine activities such as travelling to and from work, or going to the local shop. It is argued that strengthening the guardians (through a visible police presence) within these areas can help

dissuade the offender from future offending (Johnson et al., 2014). There is some limited empirical evidence to suggest that the location of a capable guardian does not even need to coincide precisely with the offender/target interaction, as the presence of patrolling police officers in an area is claimed to have a crime reduction effect of around 2 hours (Koper, 1995).

2.4 Techniques used for forecasting crime

A multitude of techniques are used in crime forecasting, in terms of both where it will happen (the location) and when (the time). These techniques range from the simple to the highly complex, with increased complexity not necessarily providing 'better' results. Some of the key techniques are explained below. It is important to note that these are not mutually exclusive: indeed, a number of commercial predictive policing products incorporate aspects from one or more of the various techniques.

2.4.1 Hot Spot mapping

This involves using spatial and temporal data about previous crimes in a locality to predict the risk of future crime. The area studied will be sub-divided into smaller areas using existing boundaries (such as postcodes or officer beats), or a standardised grid of identically sized areas (either population or geographically identical). The past offending data is then mapped onto this grid and a count taken of the number of offences in each box. The size of the grid boxes is important because larger boxes are less useful in operational terms, and smaller boxes are more likely to 'miss' many crimes. The number of offences within each box will indicate the risk of crime in that area, with those of high risk seen as being 'hot' and those of low risk being 'cold'. The risk levels can then be colour-coded and super-imposed on a map. This visual representation has been shown to assist officers in understanding the power of predictive methods and increases staff engagement (Perry et al., 2013).

It is also possible to represent the spread of risk across the map as ellipses, with the most closely clustered crimes centred in the middle of the ellipse. This technique was used by an early piece of crime forecasting software: STAC (Spatial and Temporal Analysis of Crime) which uses 'standard deviational ellipses' to show the risk associated with clusters of crime (Chainey et al., 2008). The method can produce potentially misleading results if the cluster of crimes is distributed over a small geographic area because large areas of low risk being can be included within an ellipse (Perry et al., 2013).

2.4.2 Heuristic methods

Heuristic methods are similar to the 'cognitive shortcuts' that form part of our everyday working practices, and are based on experience and inherited forms of reasoning, and the drawing of inference. Analysts use their knowledge of the crime and geography of their area to help generate their forecasts. These are arguably the simplest methods available to an analyst, and range from manual identification of hot spots by eye to simple statistical methods. These relatively straightforward techniques have been found to be extremely useful when attempting to uncover actionable hot spots, as the analyst will probably have in depth knowledge of the areas they are covering, the likely offenders and offences committed there in the past.

However, it is worth noting that there are problems associated with this method, as familiarity with the areas under consideration is required (ideally native experience and/or working knowledge) combined with experience of successful hot spot identification. In addition these types of methods tend to reduce the impact of repeat victimisation on predictions, because a single geographical location can only be represented once, so will only be counted once (Perry et al., 2013).

2.4.3   Regression methods

These techniques allow an analyst (or an automated system) to attempt to forecast offending based on the relationship between a dependent variable (normally the incidence of crime) and other independent ('explanatory') variables. Whilst the existence of a mathematical relationship does not necessarily imply that one variable causes the other, it can still yield interesting predictive results. For example a regression model aimed at forecasting the number of future burglaries will probably include the number of prior burglaries as well as other variables such as: the number of vandalism complaints, the population density, or the number of unoccupied homes in the area. This can provide the analyst with insights such as showing the mathematical relationships between different crimes, or the association between population and crime.

Whilst the results for a regression analysis appear likely to be useful, their accuracy is dependent on the reliability and size of the dataset used. Data which is short, incomplete or volatile (subject to dramatic changes) is unlikely to produce a good regression analysis. The analyst needs to be aware of the quality of the data at all times and cannot blindly use any data provided (Perry et al., 2013).

2.4.4   Data Mining

Data mining is the practice of searching through large quantities of data and attempting to identify patterns and trends which can then be applied to create forecasts. The methods used in data mining are varied and technically complex, so it is best to consider them a suite of tools that can be employed to extract information from large datasets, and applied where appropriate. There are two main types of data mining employed, each producing different types of output for police organisations:

•   Classification methods, which produce statistical chances of offences occurring within geographical locations e.g. there is a 70% chance of a burglary occurring within the neighbourhood within the next week.

•   Clustering methods, which allow the analyst to determine areas that share similar characteristics with other areas that are already at high risk of crime. This allows the identification of potential future hot spots and can indicate locations where forces may wish to target resources.

The complexity of these methods should not be seen as a harbinger of accuracy and precision; in practice they have not always produced more accurate results than simpler techniques (albeit that these methods are still somewhat in their infancy). There are also concerns that analysts cannot fully evaluate the predictions as some products provide only the predictions (referred to as 'black box' models), and sometimes this is also on the grounds of commercial confidentiality. This will mean that the analyst will need to trust in the accuracy and precision of the model without the capacity to check beyond noting whether a particular prediction was accurate or not. Again these approaches to crime forecasting are heavily affected by the quality of the data used, as poor quality data is very likely to yield inaccurate results for the analyst (Perry et al., 2013).

2.4.5   Single and dual Kernel Density Estimation (KDE)

'Kernel Density Estimation' (KDE) is a technique where the influence of a crime event is spread over an area (of time and space) using a mathematical function referred to in the literature as a 'Kernel'. A single KDE uses just the crime events themselves whereas a dual KDE will utilise another variable (such as population density) to help create the measure of the spread. This method allows an analyst to show that for an area close to an incident the underlying risk of crime is potentially high rather than just simply showing the original geo-spatial location of the incident. KDE is therefore seen as

being one of the better current methods for visualising crime data (Chainey et al., 2008), but Eck et al. (2005) note that the 'visual lure' of KDE has perhaps discouraged a thorough evaluation of its value.

2.4.6   Near-repeat methods

Near-repeat methods (NRM) operate under the principle that the majority of future crimes (at least of particular types) will occur close to recent instances of crime, in both time and space. This is similar to RAT (see 2.3.2 above) with offenders repeatedly targeting areas with weak guardians, and to RCT (see 2.3.3 above) with a low-security area or many potential victims affecting the offender's perception of the cost/benefit ratio. Consequently areas that currently suffer high levels of crime are predicted to be likely to suffer high levels of crime in the near future. It is also argued that the effect of a crime stretches beyond the effect on the victim; that the risk of crime increases for a short distance around a crime event and for a short period of time afterwards. This is supported by research with Mohler et al. (2011) which found that in 2001-05 in the San Fernando Valley more than 100 burglaries took place within 3 hours and 200m of a prior burglary. This effect is most strongly felt with burglaries. The predictive use of NRM for other crimes tends to be lower (Perry et al., 2013).

This method can be used in a simple heuristic manner in that areas where a crime incident has just occurred can be marked as being at higher risk of a similar crime. It can also be used in more complex ways, for example to create a mathematical model, such as ProMap or Mohler's 'earthquake modelling' algorithm.

2.4.7   Risk Terrain Analysis (RTA) and Risk Terrain Modelling

Risk Terrain Analysis (RTA) attempts to build a picture of future offending by determining the risk associated with locations within the assessed area. The forecasts are produced by a Risk Terrain Model and are based upon the distance between locations and particular geographical traits (such as pubs) that are deemed to increase the risk of crime. This method is similar to hot spotting in that it measures the risk of crime in a given area; however, RTA bases this measure of risk on geographical and other environmental traits rather than just on past offending.

To determine the risk, a statistical analysis of the distance between geographical traits and certain crime incidents is normally undertaken. The traits that have a strong statistical relationship to crime are then counted in each area and a colour is applied to the map relating to the risk. Unlike hot spotting (with which it shares a visual similarity), RTA gives the inherent risk posed by the geographical traits present within an area. This means that the analyst is able to not only determine areas that are of high crime risk now, but could be at risk of high levels of offending in the future. This can also be used to judge whether displacement of offending is likely to occur, should a hot spot be targeted for action (Perry et al., 2013).

The Amsterdam police have apparently successfully used Risk Terrain Analysis in crime forecasting. The police analysts first use a neural network to identify the top 3% risk areas in Amsterdam, and subsequently a logistic regression classifier to determine the times when those areas are most at risk (the classifiers are chosen based on their good performance in the regression, and not because of any criminological considerations). A team of analysts examine the maps and determine one or more areas (typically circles with radius 1 km) that seem to be most at risk, and when. Patrols are sent there at the relevant times, and they retrieve local knowledge and use this to write a deployment advice document (personal communication, D. Willems, Netherlands Police).

2.4.8   Mixed methods

Many of the commercially available products for predictive policing utilise one or more of the methods described above. Often this involves combining hotspot mapping with near repeat analysis and Risk Terrain Modelling. The Venn diagram reproduced below in Figure 1 illustrates the ways in which the three approaches are often combined and the reasons for doing so.

Figure 1 Combining hotspot mapping, near repeat analysis and Risk Terrain Modelling (Caplan et al., 2013, p.19)

The relative contributions of each approach to achieving forecasting accuracy and precision will probably vary according to the type of environment (urban, commercial etc.), temporal factors (season, time of day) and crime type (theft from person, theft from motor vehicle). A set of data is usually divided up into a 'training set' and a 'testing set'. The training set is used to set the parameters of the algorithm(s) involved whilst the testing set is used to assess how well the algorithm forecasts crime.

2.5   Predictive policing as an operational response

Although some of the underpinning philosophy of 'predictive policing' harks back to the original function of the police under Peel's 1829 principle of 'the prevention of crime' (Reiner 2010), it also resonates strongly in a time of austerity for the police service. In principle, by utilising predictive techniques, police forces will be able to allocate resources with greater effectiveness and efficiency, so that crime incidents can be prevented. Good and focussed forecasting techniques potentially allow greater flexibility for police force resource allocation, and the development of innovative approaches to crime prevention. This section of the report will look at the role of predictive policing within policing, and present some of the main benefits and challenges that this approach presents.

2.5.1 Types of intervention

When utilising predictive policing, and assuming a relative high level of both precision and accuracy (see section 4 below), it is possible for police forces to use the results in a number of ways. Perry et al. (2013) identify three key interventions that could be developed through predictive policing (our interpretation):

•   Generic interventions: using the predictions to allocate more resources (usually in the form of more uniformed officers or PCSOs) to areas identified as being at high risk of crime. This can vary from simply placing officers in a given area to placing officers in the given area at a specific time.

•   Crime specific interventions: tailoring resources in a given area to match the expected frequency of crime forecasted.

•   Problem specific interventions: introducing resources that would 'fix' a problem area. This could mean introducing special measures (such as 'exclusion areas'), no-drinking zones or the introduction of more street lighting.

Whilst these interventions are not exhaustive they are a useful indictor of the possible outcomes of predictive policing. It is important to note that response is not limited to unformed police patrol. Nor are the interventions mutually exclusive; forces can choose to apply any, and all of these, to a given geographical area, at any time. We could also add intelligence-led approaches to counter a forecast increase in crime to the list drawn up in 2013 by Perry et al. (see section 6 below).

2.5.2 Uses of predictive policing

Individual police forces will no doubt determine the particular form that predictive policing will assume, if any, for their organisation. It seems likely, for example that larger forces would be able to use the outputs for different goals within the same organisation. However, algorithms developed for one police force, locality and set of crime types are very likely to need refining for a different location and crime type, or if the police force employs a different policing model.

The Sacramento police department implemented predictive policing as part of an initiative to reduce the number of 'Part 1' crimes (Homicides, aggravated assaults etc.) within 42 hotspots across the city. They utilised the work of Koper (1995, see below) to see if regular patrols lasting 13-15 minutes every two hours in these hot spots would reduce the number of offences. The 42 hot spots were divided into those being treated normally and those receiving the regular patrols. Following a 90 day trial the rates of offending in both groups were compared with the previous year's figures. It was claimed that the group with regular patrols had a 25% reduction in offending compared to the previous year, whereas Part 1 offences in the untreated area rose by 27.3% compared with the previous year. The calls also decreased in the treated area by 7.7% compared to a rise of 10.9% in the untreated areas (Oulette, 2012).

In this example Sacramento police department demonstrate situational awareness in their use of the predictions. They used the predictions not just as a measure of where to place officers, but as a method of introducing real changes to the way they police these areas. Perry et al. (2013) argue that the absence of such situational awareness is the main barrier to effective uses of predictive policing, with police forces often content to merely be shown where to place officers, with little understanding of why they are doing it, or what use it will be. They outline four key resources that are required in order for predictive policing to be successful:

• Top level support within the police force.

• Resources dedicated to the task.

• Enthusiastic and interested staff at all levels.

• Good working relationships between analysts and other staff.

The surveys conducted as part of the evaluation of predictive policing in London attempted to measure how far some of these resources were in place in the MPS (see section 5.3 below).

2.5.3   Potential pitfalls and misconceptions

In order to effectively implement predictive policing it is also important to note some of the main pitfalls and misconceptions surrounding the use of these techniques. Many of the problems are caused by misconceptions about the role of predictive policing and what it can bring to a police force. It is clear that in the popular meaning of the phrase, the term 'predictive policing' is a misnomer. Rather than producing predictions in the usual sense of the word, the algorithms forecast crime with associated likelihoods (probabilities). It is crucial to understand that these 'predictions' are therefore not really predictions in the usual sense of the word (as in predicting that it will rain tomorrow) but are instead a mathematical representation of the risk of crime within a specified area (Goode, 2011). Technically, the confusion arises because whereas the number of crimes that occur is a discrete variable taking integer values (no crime, one crime, two crimes), probabilities are continuous values ranging between 0 and 1 (or if expressed as percentages, between 0 and 100%). To 'predict' that a burglary will occur with a probability of 0.46 on 14th February 2017 between the hours of 0700 and 1500 within a 250m by 250m square in Hillingdon means that the crime is not likely to occur on that occasion (on the balance of probabilities). However, if the algorithm is performing as intended it means that a crime would have occurred on 46 occasions in 100 such circumstances. Unfortunately our instinctive reasoning as humans is such that we make judgements based on the first few outcomes (successes and failures) rather than over the longer time frame. Unless this is made clear, it could be argued that some of the commercial companies of 'predictive policing' solutions are becoming victims of their own publicity, as expectations amongst police officers will be high (based on the hype), and many will inevitably be disappointed with the outcomes.

Another misconception is that the amount of data or the cost/complexity of a particular model will correspond with the accuracy of the forecasts. As we have already noted, some (relatively) simple techniques yield compelling results when used in the right circumstances (see the Richmond VA example), and highly complex techniques have only been found to be marginally better at predicting future risk (Perry et al., 2013). Similarly, it is important to recognise that the majority of the implementation work will still need to be undertaken at an operational level; determining responses, resources and time management will not be set by the forecasts. The forecasts are merely a guide

to aid in these decisions, and in isolation will not result in crime reduction. Focusing wholly on 'predictive accuracy' at the expense of tactical utility will result in inefficient use of the forecasts.

3   Crime forecasting products trialled by the MPS

As part of the evaluation of predictive policing conducted for the MPS we examined the predictive accuracy of four products which employ algorithms that attempt to forecast crime. Three of these products are commercial products, the fourth is an experimental model devised by analysts from the MPS.
The crime forecasting models are used to produce 'predictive boxes' (rectangular geographical areas) in advance of an MPS shift or shifts. All four products have in common the use of data concerning crime times and locations both in the previous 24 hours but also data from the last three years. However, they differ in what other additional data they utilise beyond this (if any). In the case of the external companies involved, MPS data (in the form of Excel csv files) is automatically sent to the servers of the company concerned, with the appropriate security measures adopted. Forms of data 'cleansing' and preparation are likely to be undertaken by the companies at this stage.

All products have a crime forecasting phase, when the algorithms are used to generate 'predictions' (probability forecasts). Although running the algorithm requires computer processing power it is probably not a case of 'big data' analysis for most of the four models (the data sets are simply too small to warrant the description) and the calculations probably only take a short time to perform. The next phase will be to display the forecasts in a way deemed suitable for police operational use – in the case of all four products considered this means rectangles displayed on a map.   At various points the performance of the model will be reviewed by its owner (e.g. in terms of the number of successful hits in a previous set of predictions) and adjustments could be made to the values of some of the variables in the algorithm.

The following section of the report provides a fuller description of the MPS MBR predictive modelling.

As the participation of the three commercial products in this research was secured on the basis of a non-disclosure agreement prohibiting their identification, no description of their respective workings will be provided.

3.1   Metropolitan Police Service (MPS) algorithm ('MBR')

The MPS predictive crime initiative was trialled in early 2013. The test boroughs were provided with daily maps that highlighted at risk areas for seven crime types to be integrated into their daily patrols. The crimes covered by the maps are criminal damage; domestic burglary; robbery; theft from the person; theft of and theft from a motor vehicle; and violence with injury. During the trial period, feedback led to two further versions being created to reduce the clutter and increase the clarity of the maps by reducing the number of at risk areas highlighted, better enabling the dispersion of resources. The current version 3, which highlights ten at risk areas, was rolled out to 26 out of the 32 boroughs in the MPS catchment areas by the end of 2013. To maintain consistency across the boroughs each map produced uses the 250m by 250m grid squares.

Whilst there are similarities to the process of kernel density estimation (KDE) mapping (see section 2 above), the MPS predictive maps differ from KDEs through the addition of temporal weighting. The weighting is based on the near-repeat boost explanation, meaning that greater weight is given to crimes that occurred more recently than to those in the more distant past. The MPS predictive maps are produced using the following seven-step algorithm:

1. First a weighting of either 1, 2 or 3 is given to the each of the crimes that occur within the past 21 days of the 'index day' (the date for which the map is being created). Any crime

between 1 – 7 days receives a weight of 3, for 8 – 14 days the weighting is 2 and for 15 – 21 days the weight is 1 (effectively unweighted).

2. Next, three buffer areas are created for each crime, again with a 3, 2, 1 weighting as follows: for a 50m radius the weight is 3, for a 125m radius the weight is 2 and for a 250m radius the weight is 1.

3. A grid with 250m squares is created for the prediction area, with each buffer being intersected by a grid or number of grids.

4. The buffer-grid intersection is then defined as the area of each buffer that intersects with the grid square.

5. A proportional weight for the intersect is gained by dividing the spatio-temporal weighting of the whole buffer by the buffer-grid intersection ratio, this is performed for each intersected buffer section.

6. The proportional weightings of each intersect within a grid square is then summed up to create an overall grid square weighting.

7. Finally each grid square is ranked based on the weights within a borough and the top ten and top 20 grid squares are highlighted.

4.   Assessment of Predictive Accuracy

In order to inform an understanding of the predictive accuracy of the four products a quasi-experimental trial was conducted. The four providers were asked to provide separate forecasts for the towns of Reading and Slough for the four MOPAC crime types on a thrice-daily police shift basis. This was a particularly demanding challenge, given the requirement by the MPS to forecast specific crime types within a short eight hour (shift) time period.

The intention of the trial was not to determine which model was 'best' – putting aside what this actually means (see the discussion concerning accuracy and precision below) – but to instead assess the potential of forecasting crime to support operational policing. For this reason we disaggregated the data to analyse predictive accuracy by crime type; we also took into account the size of predictive boxes and we attempted to provide some speculative explanation for the results achieved.

Crime forecasts were provided by the three companies and the MPS analysts (in advance of each police shift) for Reading and Slough, for each of four MOPAC crime types, and by date and shift. These predictions were compared after the event against actual recorded crime data provided by Thames Valley Police (for a description of the method used see 4.6.3 below). It is important to note that we made no attempt to assess the geocoding accuracy of the Thames Valley Police data, which was assumed to meet Ratcliffe's 85% acceptability level. Geocoding errors can occur in a number of ways but principally through an incorrect recording by police officers at the outset of the location of the crime. Recent research conducted by West Midlands Police and the University of Salford (Harrell, 2015) found that only 31% of all robberies were allocated to the correct geographical location with a mean positional error of 193 m (larger than the average size of rectangle used by all the products in their predictions). It has been known for some time that major positional errors can frequently occur and as Hart and Zandbergen (2012) note '[…] geocoding quality research clearly demonstrates that errors in geocoding can be very substantial […]' (p.15).

4.1   Definitions of the four MOPAC crime types

Police crime recording is governed by the Home Office Counting Rules (HOCR) and the National Crime Recording Standard (NCRS) and the National Standard for Incident Recording (NSIR). The MOPAC 7 crime definitions are based on the HOCR categories. Predictive accuracy for the purposes of the trial in Reading and Slough was measured in respect of 'residential burglary', 'robbery of personal property', 'theft from person' and 'theft from a motor vehicle' (subsets of four 'MOPAC 7' crime categories). The definitions of these crime types are shown in Table 3 below.

Crime type	Subset of MOPAC 7 definition
Residential burglary	Burglary is the theft, or attempted theft, from a building/premises where access is not authorised. Damage to a building/premises that appears to have been caused by a person attempting to enter to commit a burglary, is also counted as burglary. (The MOPAC category includes burglaries of both residential properties and non-residential properties (such as commercial property, sheds, and outbuildings).
Robbery of personal property	Theft involving the use of force, or the threat that force would be used. This category includes robberies of both personal and business property.
Theft from person	Theft of property from a person that is not accompanied with force or the threat or fear of force (such as pick pocketing). This will include theft of property that is being worn, carried, physically attached in some way to the victim, or contained in an article of clothing being worn by the victim.
Theft from a motor vehicle	Theft of property located inside a motor vehicle

Table 3 Definitions of four MOPAC crime types (Source: MOPAC, 2012)

The data used was that as originally recorded by Thames Valley Police on entry to their database for the day on which predictions were made. TVP apply the National Crime Recording System (NCRS). This means that a crime classified as one type might be subsequently reclassified as another type on further investigation or supervisor intervention.

4.2   Times at which crimes occurred

The time at which a crime occurs can often be the subject of some uncertainty and this is particularly the case for some non-interpersonal crimes such as theft from a vehicle. For example, a person might park his or her car at 8am and return to the vehicle at 6pm to find that the car has been broken into and a SatNav stolen. In the absence of any other information this crime would be recorded as occurring between 8am (0800) and 6pm (1800), a period of time that begins in one shift (Early) but ends in the next (Late). The mid-point of the period of time in this case would be 1pm (1300), that is within the early shift. If the mid-point is used as a proxy for the time in this way then it gives rise to a single unique date and shift. In this example, a predictive policing algorithm model

that correctly predicts a crime of theft from motor vehicle as occurring within the Early shift would be considered as a success (e.g. awarded a score of 1) but if instead it had predicted the late shift it would be deemed to have failed (e.g. scored a 0). Although this is intuitively appealing (crimes either occur or don't occur within a shift) it does imply a level of temporal certainty which is not actually the case.

In practice a number of possibilities can arise when attempting to decide how to 'allocate' a crime to a police shift: the time at which the crime occurred is known with some certainty; the start and end points of the crime fall within the same shift or the start point is in one shift but the end point occurs in either the next consecutive shift or a later shift.

In terms of the TVP MOPAC crime data for the period of the trial, Table 4 below describes the proportion of crimes that fell within a single shift or more than one shift.

MOPAC Crime Type	1 shift	2 shifts	3 shifts	4 or more shifts
Robbery of Personal Property	91%	0%	4%	4%
Domestic Burglary	45%	33%	7%	15%
Theft from the Person	87%	4%	2%	7%
Theft from Vehicle	36%	20%	29%	15%

Table 4 Proportions of reported crimes whose time period fell in one or more shifts

As can be seen, the proportion of domestic burglaries and theft from vehicles that span two or more shifts is particularly high.

An alternative to using the midpoint to decide on the success or otherwise of a prediction is to use a more 'aoristic' approach which treats the time interval in which a crime occurred as a continuous rather than discrete variable. There is some discussion in the literature concerning the best way of doing this e.g. using random allocation, the Monte Carlo method and so on. A common aoristic approach is to use a linear distribution of crime likelihood over the period between start and ends point. For example, consider the circumstances where crime is recorded as occurring between 22.30 on 2 February 2015 (t_{min}) and 08.30 on 3 February 2015 (t_{max}). The mean time is 03.30 on 3 February i.e. during the Night shift that started at 23.00 on 2 February. The aoristic alternative is to assign 'fractional' crimes to shifts spanning the range of times between t_{min} and t_{max}. This spans the late shift on 2 February 2015; the night shift from 2 to 3 February 2015 and the early shift 3 February 2015. Table 5 below shows how the crime would be proportionally assigned to shifts.

Shift	Proportion of crime
Late shift 2 February	0.05
Night shift 2 February 2015 to 3 February 2015	0.80
Early shift 3 February 2015	0.15

Table 5 Allocation of crimes using an aoristic method

As part of the evaluation we calculated the predictive accuracy of the algorithms trialled using the alternative linear aoristic method in addition to the midpoint method, comparing the two outcomes.

4.3   Crime in the towns of Reading and Slough

The towns of Reading and Slough in Berkshire were chosen as the geographical locations for the predictive accuracy trial. This choice was largely a pragmatic one: predictive policing in the MPS greater London area was already underway which meant that London boroughs could not be used; Thames Valley Police were considering introducing predictive policing as part of their operational policing response but had not as yet made any significant changes and the towns were considered to be 'representative' of typical London boroughs.

4.3.1   Crime in Reading

The town (and unitary authority area) of Reading is situated approximately 60 km west of central London. It has an area of approximately 4039 hectares (authors' calculation) and a population of approximately 159,000 (based on the mid-2013 ONS estimate). The population density distribution of Reading is shown in Figure 4 below. The average population density is 39.37 persons per ha.
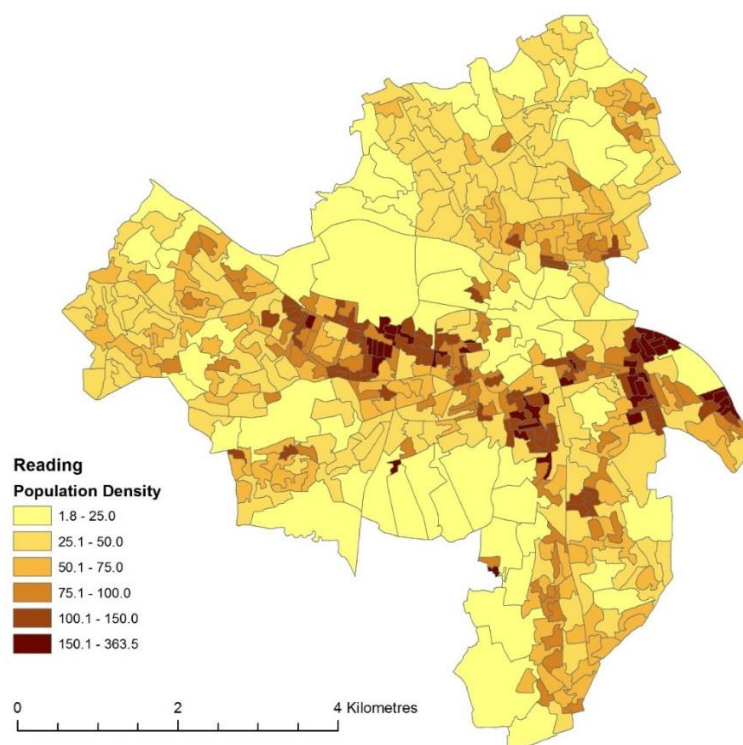
Figure 4 Population density of the town of Reading

As can be seen from Figure 4, the population of Reading is largely situated in the central third of the area. According to the 2011 census, 74.8% of the population were described as White (65.3% White British), 9.1% as South Asian, 6.7% as Black, 3.9% Mixed Race, 4.5% as Chinese and 0.9% as other ethnic group (ONS, 2011).

In terms of recorded crime, towns and cities are normally compared to others within a 'Most Similar Group' (MSG). These are groups of local areas that have been found to be the 'most similar to each

other using statistical methods, based on demographic, economic and social characteristics which relate to crime' (HMIC, 2013). Seven variables are used to determine the MSG ; percentage of ACORN neighbourhoods (e.g. 'Hard Pressed' neighbourhoods), percentage of terraced households, Output Area density (a population density measure), percentage of overcrowded households, percentage of single parent households, population sparsity and long-termed unemployed per worker (ibid). The MSG for Reading includes Bristol, Southampton, Oxford, Northampton, Portsmouth, Brighton & Hove, Hounslow, Hillingdon, Slough, Ealing, Watford, Exeter, Eastbourne and Worthing. During the period of the predictive accuracy trial, the recorded crime 'mix' of Reading was closest to that of Brighton & Hove and Hounslow (see Figure 5 below)
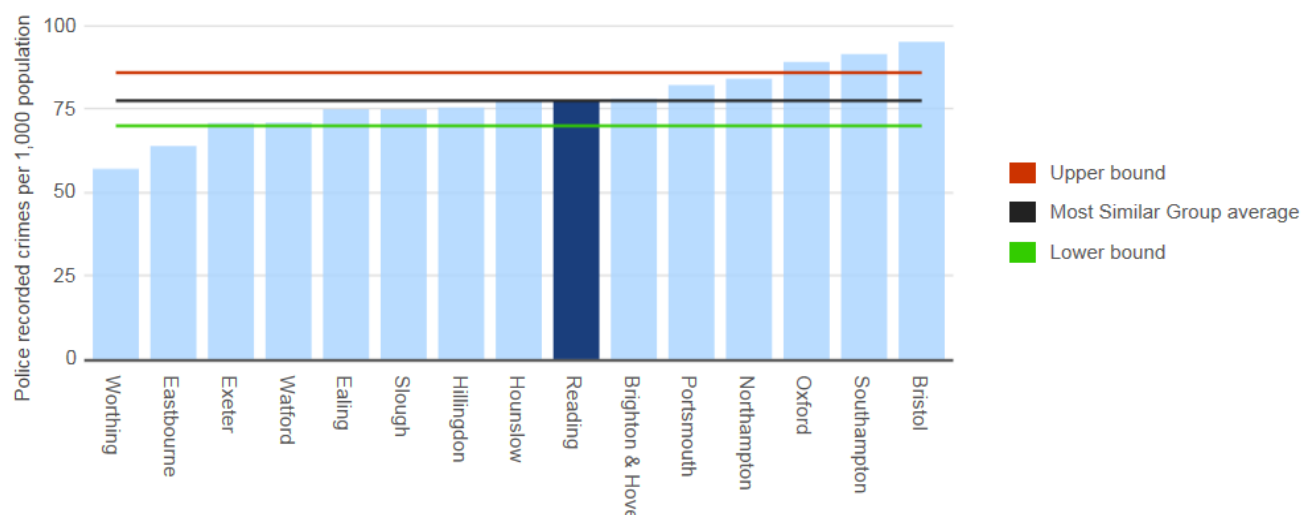
Figure 5 Crime in Reading compared with MSG (source: police. data.uk)

In terms of particular categories of crime types, Reading was close to the average for the MSG other than in two categories, one of which was 'theft from the person' (see Figure 6 below). The other category was 'other crime' (which includes forgery, perjury and numerous other miscellaneous crimes).
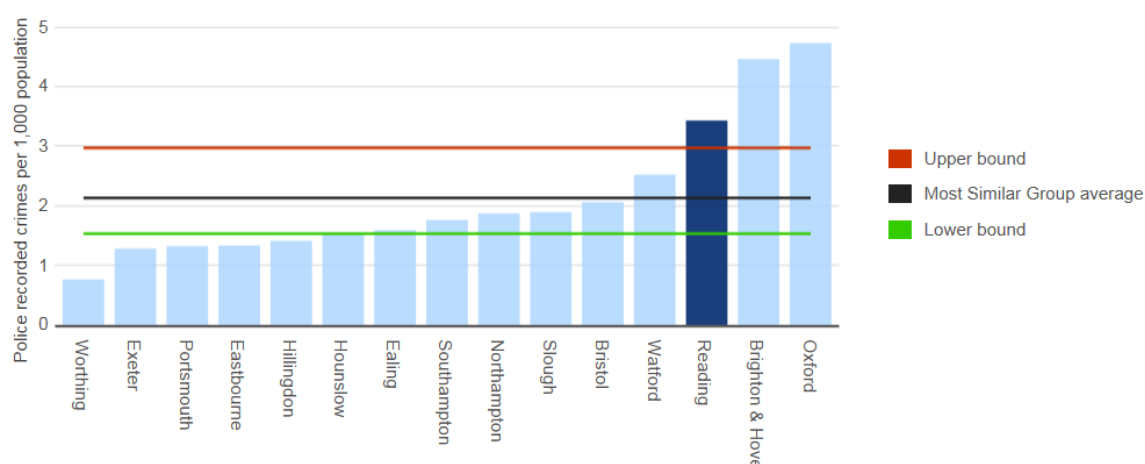

Figure 6 The crime of 'theft from the person' in Reading compared with MSG (source: police.data.uk)

Thames Valley Police provided data on the MOPAC crimes that were the subject of the trial (see below) for a three year period up until December 2014 (this data was also made available to the

companies and MPS). Figure 7 below shows the spatial distribution of all the four MOPAC crime types over this period , using a Kernel Density Estimation (KDE) approach to identify the recorded crime location 'hotspots'. The use of KDE has certain advantages over alternative methods such as aggregation of point data and choropleth mapping as it produces a smooth map in which the density at every location reflects the number of crimes in the surrounding area (Gorr and Olligschlaeger, 1998).
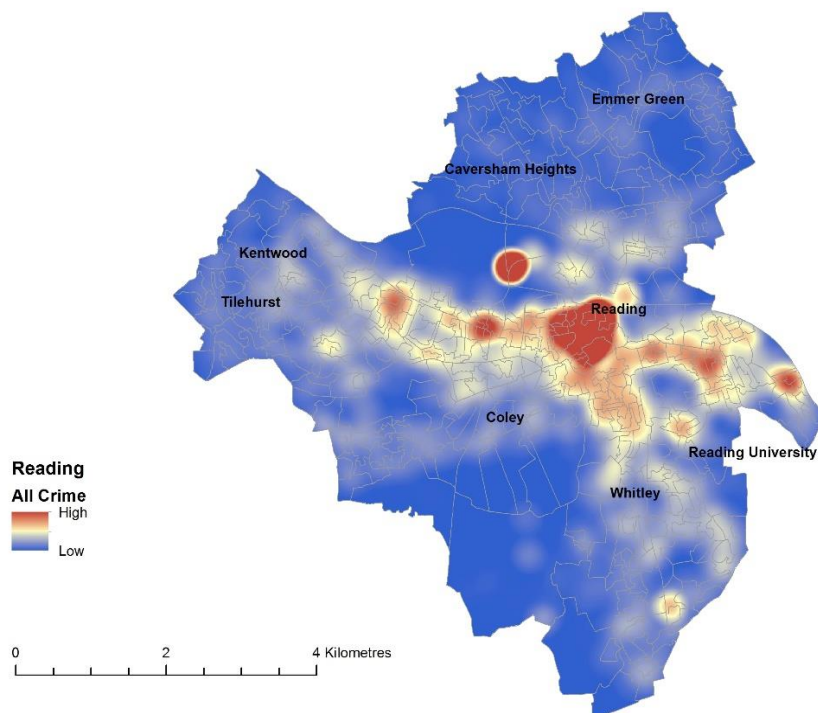

Figure 7 KDE spatial distribution of total of four MOPAC crime types in Reading, three years data

The following figures illustrate the spatial distribution of each of the four TVP crime types closest to the MOPAC crime types ('domestic burglary', 'robbery', 'theft from person' and 'theft from vehicle') for the three year period.
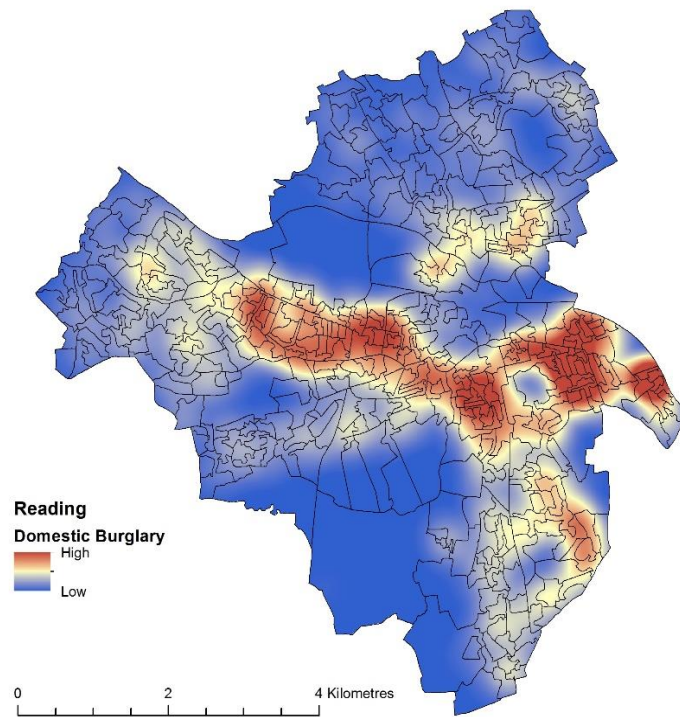
Figure 8 KDE spatial distribution of 'domestic burglary' in Reading, three years data
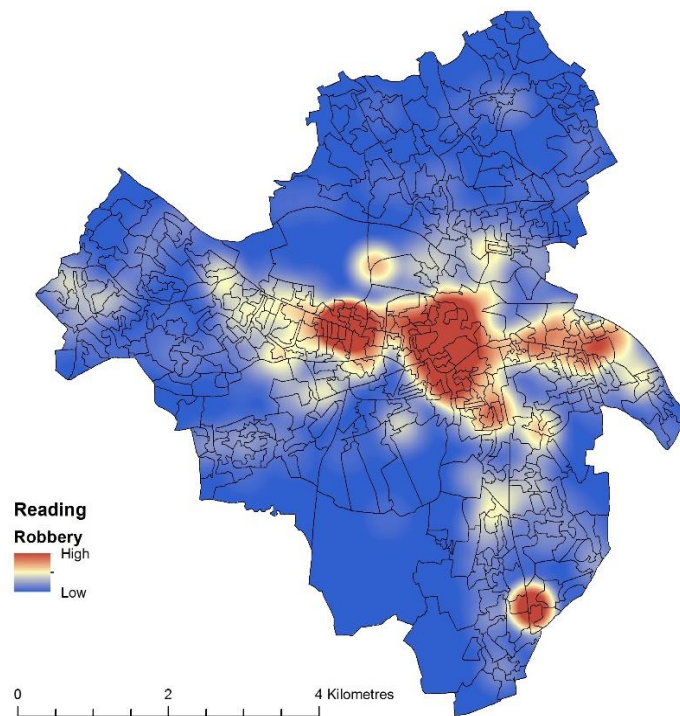

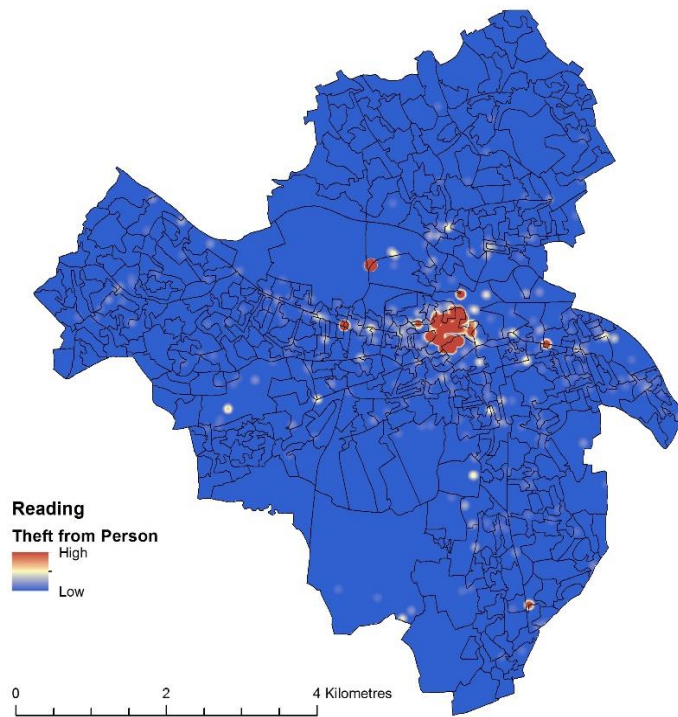Figure 9 KDE spatial distribution of 'robbery' in Reading, three years data

Figure 10 KDE spatial distribution of 'theft from person' in Reading, three years data
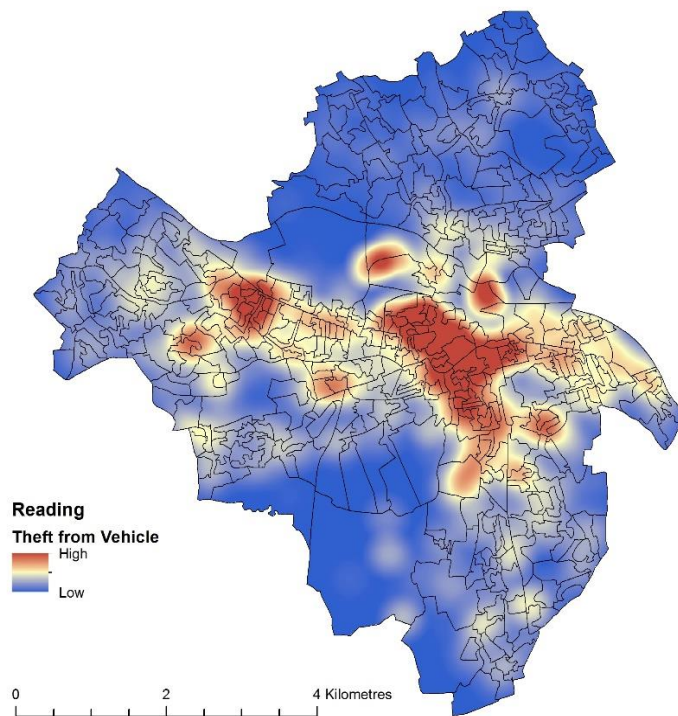

Figure 11 KDE spatial distribution of 'theft from vehicle' in Reading, three years data

As can be seen, 'theft from person' appears to exhibit a much more geographically concentrated density of offending than the other three crime types. To test this further we conducted further analysis of the spatial distribution of crime in Reading. There are a number of indicators of spatial association used in crime analysis, with one of the most common in crime analysis being the 'Local Moran's I' (or 'LISA'). The measure is particularly good at identifying both local spatial clusters and spatial outliers. Figure 12 below shows the results
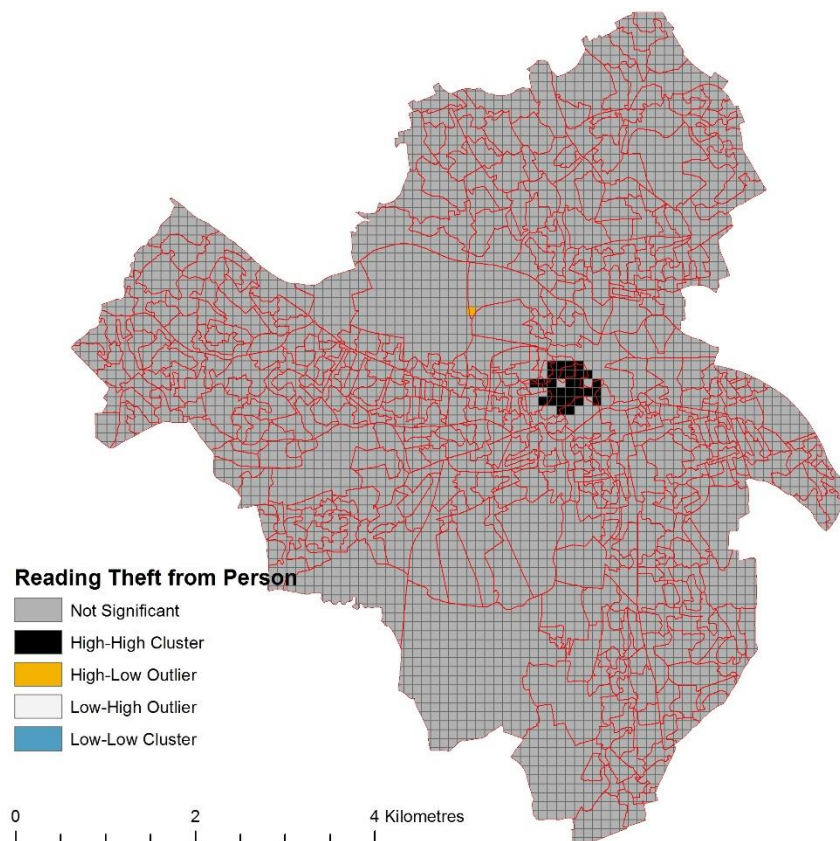

Figure 12 Visual representation of high theft from person crime areas in Reading using Local Moran's I, three years data

As can be seen from Figure 12, there are no significant outliers and high 'theft from person' areas in Reading adjoin other high 'theft from person' areas and are concentrated in the centre of the town of Reading. This in marked contrast to the 'LISA' map for Slough for 'theft from motor' vehicle (Figure 20 below) where there are numerous dispersed clusters of crime.

4.3.2   MOPAC crimes in Reading during the trial period

Table 6 below shows the numbers of crimes that were recorded by TVP during the period of the trial (12 January 2015 until 5 April 2015 inclusive).

Town	MOPAC Crime	Totals	Shift	Sub-totals
READING	Burglary	77	Early Late Night	26 31 20
	Robbery	23	Early Late Night	1 17 5
	Theft from Person	81	Early Late Night	27 21 33
	Theft from a motor vehicle	112	Early Late Night	20 21 71
	Total	**293**		

Table 6 Crimes recorded by TVP in the period 12 January 2015 until 5 April 2015 inclusive

As can be seen from Table 6, in some categories (such as robbery Early and Night shift) only a very small absolute number of crimes were recorded. This makes the interpretation of percentages particularly problematic.

4.3.3 Crime in Slough

The town of Slough is somewhat closer to London (about 30 km west of central London) and also demographically closer to many London boroughs (when compared with Reading). The area of Slough is 3255 hectares (authors' calculations) with a population of approximately 143,000 (ONS mid 2013 estimate). The population density distribution of Slough is shown in Figure 13 below. The average population density is 43.93 persons per ha (somewhat higher than Reading).

Figure 13 Population density of the town of Slough

In terms of demography, the 2011 census 45.7% of the population of Slough was White, 3.4% of mixed race (1.2% White and Black Caribbean, 0.4% White and Black African, 1.0% White and Asian, 0.8% Other Mixed), 39.7% Asian (17.7% Pakistani, 15.6% Indian, 0.4% Bangladeshi, 0.6% Chinese, 5.4% Other Asian), 8.6% Black (5.4% African, 2.2% Caribbean, 1.0% Other Black), 0.7% Arab and 1.9% of other ethnic heritage.

In terms of recorded crime, the MSG for Slough includes Bristol, Southampton, Northampton, Portsmouth, Harlow, Luton, Reading, Hounslow, Croydon, Hillingdon, Ealing, Plymouth, Stevenage and Enfield. During the period of the predictive accuracy trial, the recorded crime 'mix' of Slough was closest to that of Hillingdon and Ealing (see Figure 14 below)
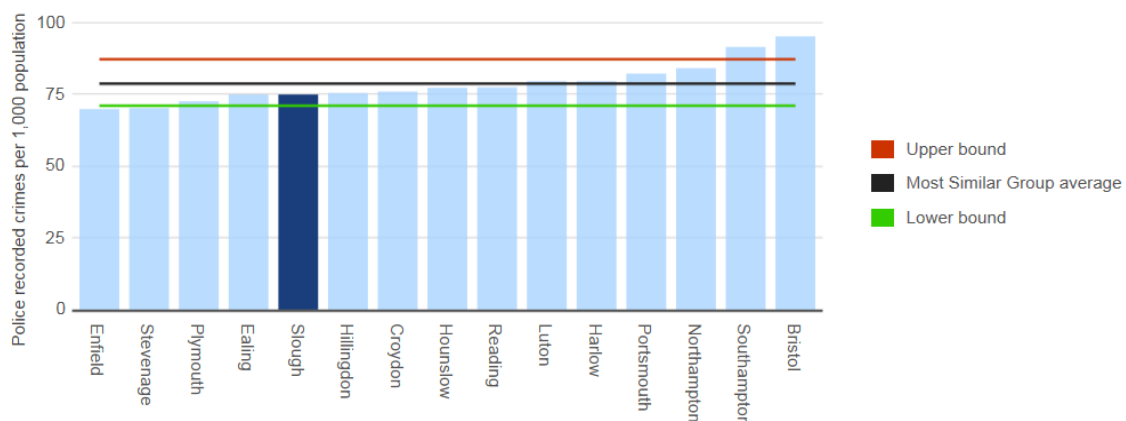
Figure 14 Crime in Slough compared with MSG (source: police. data.uk)

In all crime categories Slough was close to the average for the MSG.

As noted earlier, Thames Valley Police provided data on the MOPAC crimes that were the subject of the trial for a three year period up until December 2014. Figure 15 below shows the spatial distribution of all the four MOPAC crime types over this period in Slough , using a Kernel Density Estimation (KDE) approach to identify the recorded crime location 'hotspots'
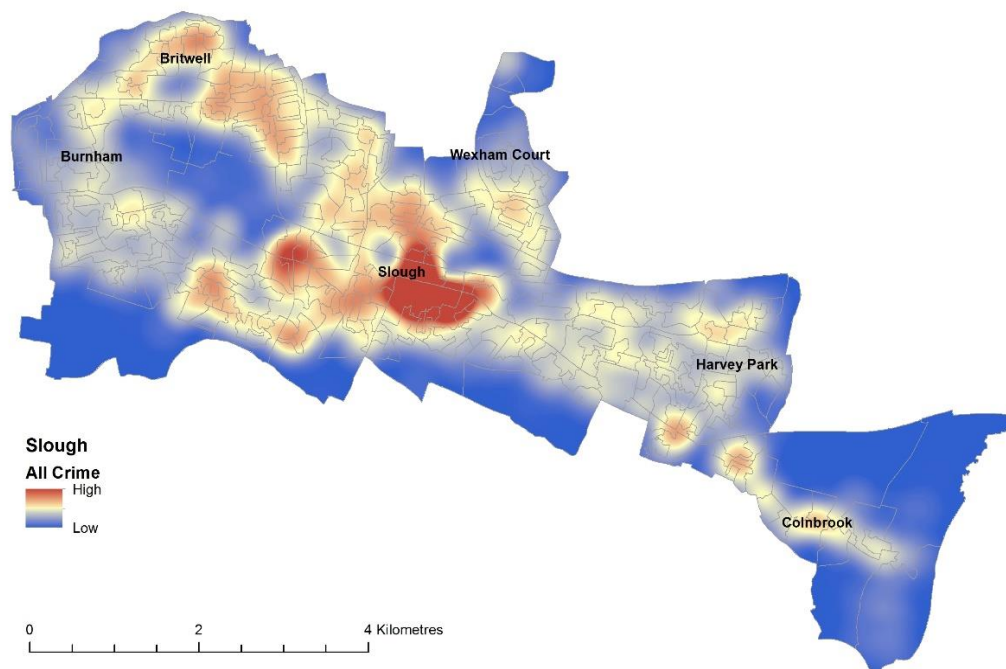

Figure 15 KDE spatial distribution of total of four MOPAC crime types in the town of Slough, three years data

The following figures illustrate the spatial distribution of each of the four TVP crime types closest to the MOPAC crime types ('domestic burglary', 'robbery', 'theft from person' and 'theft from vehicle') for the three year period in Slough. As can be seen (and as was the case with Reading, above), 'theft from person' appears to exhibit a much more geographically concentrated density of offending than the other three crime types.
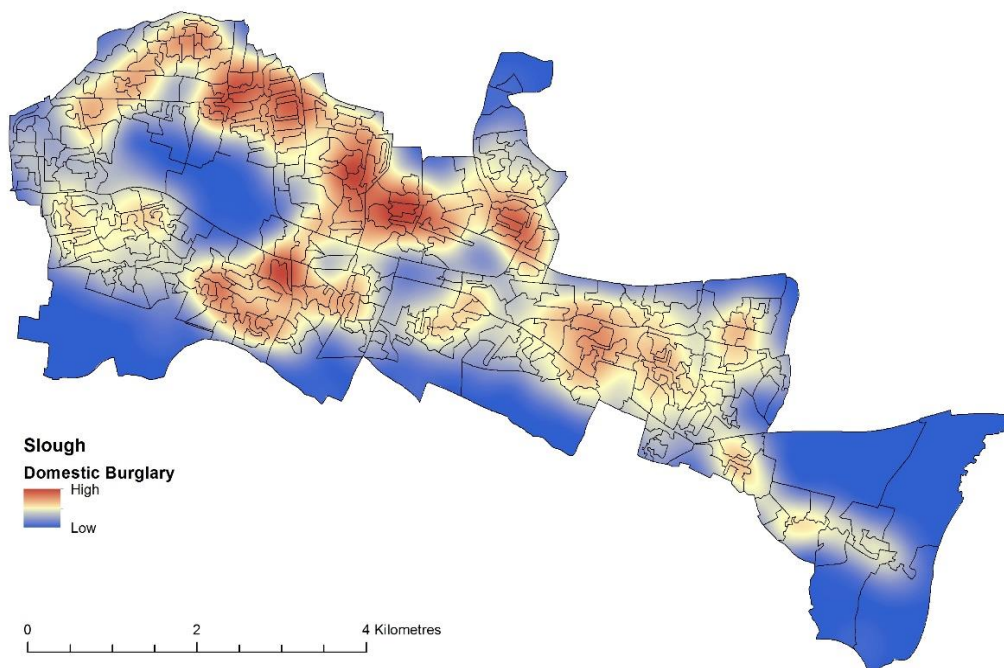
Figure 16 KDE spatial distribution of 'domestic burglary' in Slough, three years data

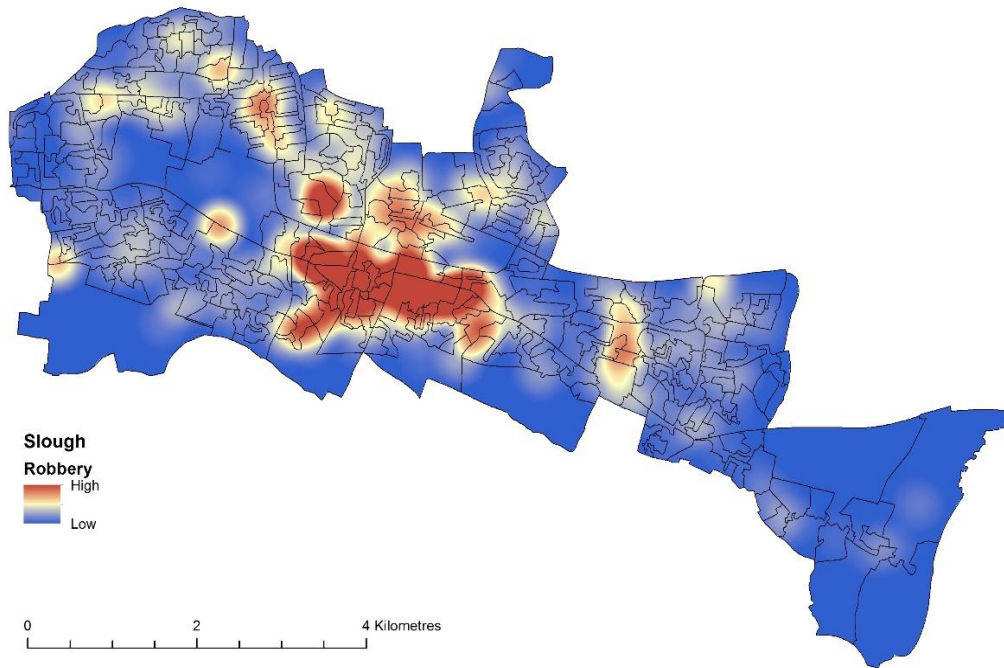Figure 17 KDE spatial distribution of 'robbery' in Slough, three years data
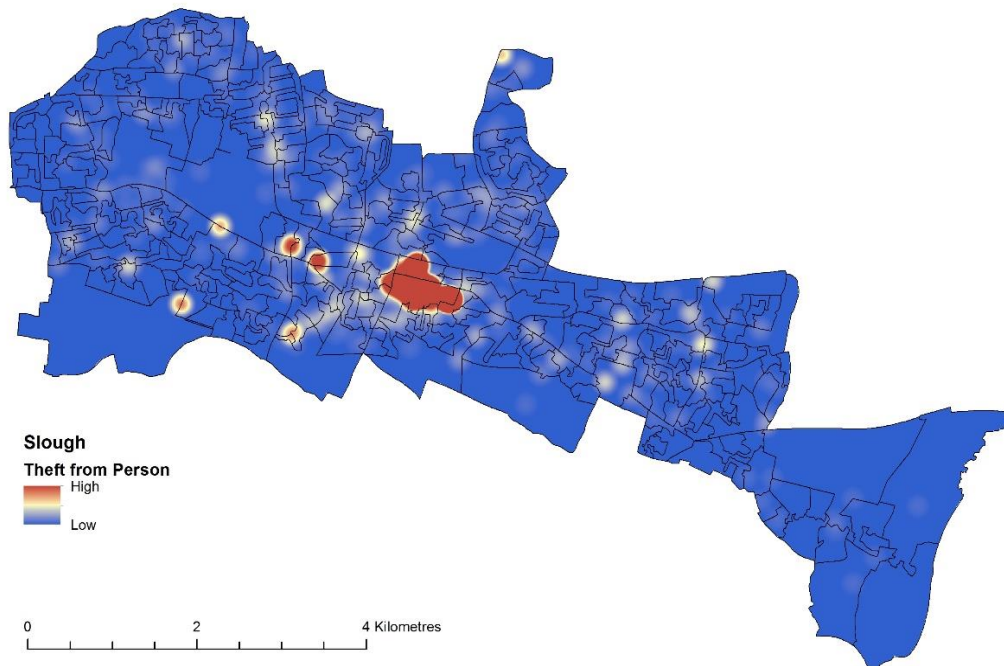
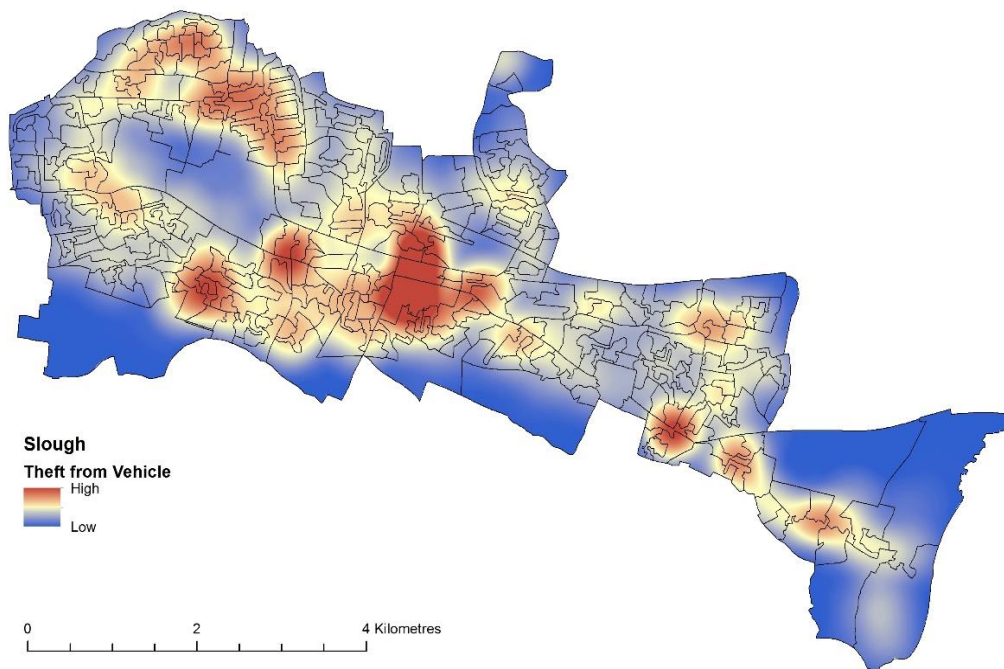Figure 18 KDE spatial distribution of 'theft from person' in Slough, three years data

Figure 19 KDE spatial distribution of 'theft from vehicle' in Slough, three years data

'Theft from motor vehicle' in Slough appears particularly dispersed. To test this the Local Moran's I values were calculated and Figure 20 below shows the result.
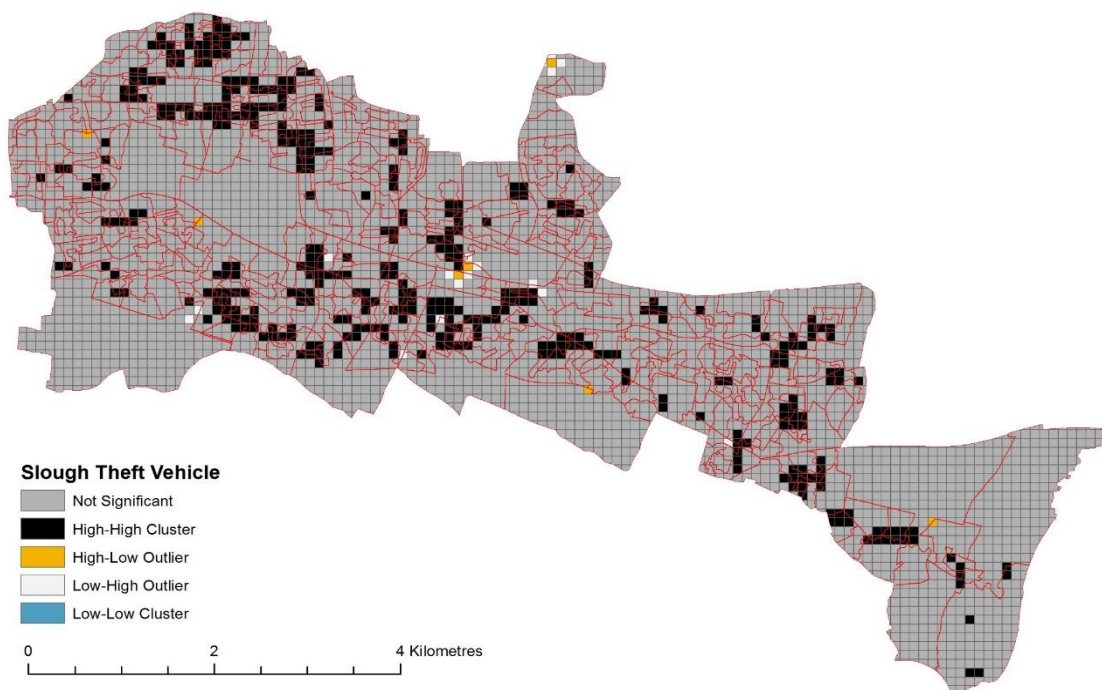
Figure 20 Visual representation of high 'theft from motor vehicle' areas in Slough using Local Moran's I, three years data

Figure 20 contrasts markedly with Figure 12 above (which shows 'theft from person' crimes in Reading), and these two Figures represent the two extremes in the trial: most clustered and least dispersed being 'theft from person in Reading' and least clustered, most dispersed being 'theft from motor vehicle' in Slough.

4.3.4   MOPAC crimes in Slough during the period of the trial

Table 7 below shows the numbers of crimes were recorded by TVP during the period of the trial (12 January 2015 until 5 April 2015 inclusive) in the town of Slough.

Town	MOPAC Crime	Totals	Shift	Sub-totals
			Early	64
	Burglary	168	Late	67
			Night	37
			Early	7
	Robbery	24	Late	11
			Night	6
SLOUGH			Early	33
	Theft from Person	76	Late	33
			Night	10
			Early	30
	Theft from a motor vehicle	187	Late	46
			Night	111
	Total	**455**		

Table 7 numbers of crimes recorded by TVP from 12 January 2015 until 5 April 2015 inclusive

4.4 Comparing Crime in Reading and Slough

As Tables 6 and 7 suggest, the absolute number of MOPAC recorded crimes in Reading and Slough were significantly different, with Slough experiencing approximately 60% more crimes during the trial period. In terms of any differences in crime type proportions between the two towns a four by two contingency table was constructed and tested for statistically significant difference. This is shown in Table 8 below (values in parenthesises are expected values under the assumption of independence).

MOPAC Crime	Reading	Slough	TOTALS
Burglary	77 (95.97)	168 (149.03)	**245**
Robbery	23 (18.41)	24 (28.59)	**47**
Theft from Person	81(61.50)	76 (95.50)	**157**
Theft from a motor vehicle	112 (117.12)	187 (181.88)	**299**
TOTALS	**293**	**455**	**748**

Table 8 Contingency table showing actual and predicted values (in parenthesis)

The chi-square statistic is 18.5791 and the p-value is 0.000334 and hence the result is significant at $p < 0.05$. The biggest chi-squared value occurs with 'theft from the person' which is significantly higher (as a proportion) in Reading when compared with Slough.

4.5 Summary of Reading and Slough

Table 9 below summarise the main differences between the towns of Reading and Slough discussed in the earlier sections of the evaluation.

Town	Geography	Demography	MSG comparators	Crime distribution	Differences in crime numbers (during trial period)
Reading	Larger area, lower population density	More closely resembles non-London towns in SE England	Brighton & Hove Hounslow	About average for its MSG for three MOPAC crime types. Above average rates of 'theft from person'. Theft from person crimes highly clustered on town centre.	Significantly higher numbers of 'theft from person' when compared with Slough

Slough	Smaller area, higher population density	More closely resembles London boroughs	Ealing Hillingdon	About average for its MSG for four MOPAC crime types.	No significant differences

Table 9 Summary of differences between the towns of Reading and Slough

4.6 Predictive Accuracy Trial

Please note that for the purposes of this part of the report the four crime forecasting products have been anonymised and referred to as 'A', 'B', 'C' and 'D'. No importance should be attached to the alphabetical order of the letters used.

It is important to note that results given below are illustrative only due to a number of limitations, principally owing to the small geographical area used (the towns of Reading and Slough), the short duration of the trial (c. 3 months), the lack of randomisation in selection, the limited number of crime types studied (four) and data problems experienced during the trial (including missing and late data). A more scientifically-based trial would, for example, where possible independently identify times and locations of recorded crimes (see below). In more general terms there are the usual and unavoidable issues that beset all evaluations which use reported and recorded crime data. In 2002 it was estimated that for all crimes occurring, 47 per cent are reported, 27 per cent are recorded, 5 per cent are cleared up, and 2 per cent result in conviction (Wright, 2002). Finally, there is also the more general issue of the lack of previous research to guide an objective testing of predictive accuracy, or indeed As Hart and Zandbergen (2012, p. 6) note ' 'there has been surprisingly little comparative research on [the] strengths and weaknesses' of hotspot mapping.

4.6.1   Defining 'Predictive Accuracy'

Although the phrase 'predictive accuracy' is often used in the literature surrounding predictive policing it is rarely defined. The word 'predictive' (read also 'prediction') in this context generally appears to mean forecasting a particular **outcome** (a crime event) at a **location** (a pre-defined geographical area) within a specific **time period** (between a beginning and end time).   However, the word 'accuracy' within the term 'predictive accuracy' is often employed with more ambiguity. it is perhaps best thought of as a combination of 'correctness' and precision. Figure 21 below illustrates two 'successful' forecasts (X) and (Y), where the red star (representing a crime location) is within the boundary of the rectangle of prediction. Although both are correct forecasts it is also clear that prediction (X) is more precise as it used a smaller area to capture the crime.
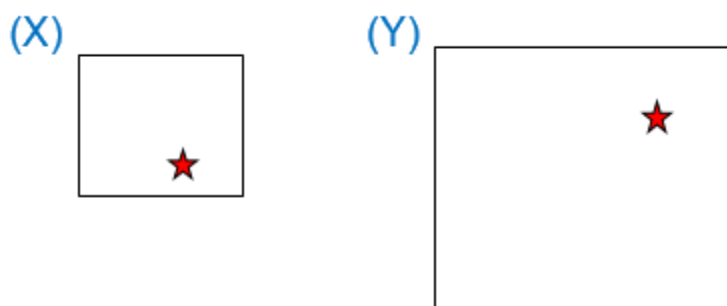

Figure 21 (X) and (Y) are both hits (correct predictions), but (X) is more precise

4.6.2 Measuring predictive accuracy

There are a number of measures of predictive accuracy used within the literature, with two of the most common being the 'hit rate' and the 'Predictive Accuracy Index' (PAI). These are described in the sections below. Alternatives include the 'Recapture Rate Index' (RRI) and the use of 'Accuracy Concentration Curves'. A number of authorities claim that accuracy is best measured by the PAI, whereas RRI is more appropriate for measuring precision (Levine, 2008) but this distinction appears not to be widely adopted. Hit rates and PAIs were calculated in this study as they allow for a limited comparison with similar academic research and data shared by other forces.

4.6.2.1 'Hit rate'

The 'hit rate' is the simplest measure of 'predictive accuracy' (although better thought of instead as a measure of 'correctness') and is simply the number of crimes successfully forecast as a proportion of the total number of crimes that actually occurred within a given time period, expressed as a percentage. The formula is particularly simple, and can be expressed as:

$$\left(\frac{n}{N}\right) \times 100$$

 (Where n is the number of crimes successfully forecast and N the total number of crimes).

For example, if a model successfully predicts 54 crimes of theft from motor vehicles in Reading from a total 1286 of such crimes in a certain period of time then the hit rate is calculated as (54/1286)x100 i.e. 4.2% (to one decimal place). Note that the value of the hit rate varies between 0% and 100%.

The hit rate has the advantages of being easy to understand, intuitive and (provided other parameters are kept constant) allows for a rapid comparison between the predictive accuracy of two or more forecasting techniques. However, it also has a number of disadvantages, the biggest being that it is highly influenced by the size of the forecasting area(s). Clearly if for example, the area of prediction were to be the whole of Reading we would inevitably capture all crime and the hit rate would be 100%, irrespective of the 'correctness' of our forecast.

4.6.2.2 Predictive Accuracy Index

The Predictive Accuracy Index (PAI) as a measure for evaluating predictive accuracy (in the context of 'hot spots'), originally proposed by Chainey et al. (2008) in response to the shortcomings of the hit rate.   The PAI is defined as follows:

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

(Where n is the number of crimes successfully forecast; N the total number of crimes; a the total area of predictions and A the overall area)

In our example, if (in a certain period of time) a model successfully forecasts 54 crimes of theft from motor vehicle in Reading (area 40.39 km^2) from a total of 1286 of such crimes and by using a total area of predictions of 0.314 km^2 then the PAI is [4.2/((0.314/40.39)x100)] = 5.4 (to one decimal place).

Note that the PAI is an index and hence has no units. It can take any value of 0 and over. As illustrated, the PAI takes into account the area under prediction as well as the hit rate and as Van Patten et al. (2009, p. 10) note 'The PAI provides an objective criterion with which to evaluate the accuracy of a hot spot from either measured or predicted crimes'.

The PAI is simply an example of a more general measure used in forecasting, but using crime data. Perhaps a more appropriate label than 'PAI' would be 'Forecast Precision Index' (Swain, 2012) although PAI has now assumed popular usage.

Although the PAI is undoubtedly a 'fairer' instrument of comparison than the hit rate, it should be noted that there have been significant criticisms of the PAI as a measure of forecast accuracy and precision for crime (e.g. Levine, 2008). A major limitation of the PAI is that (like the hit rate) it does not take into account temporal aspects. For example, different predictive policing products might well be calibrated to produce predictions for a particular time period (a shift) and a comparison with another model producing predictions for a longer time period (a day) using the PAI can be misleading.

A second problem with the PAI (not recognised in the literature but uncovered by the authors of this evaluation) is that it can appear artificially high in circumstances where part of the denominator (a/A) is very small in numerical value (as a consequence of A being large) and the crime morphology of the area is such that A can be chosen to be large in size, but not sufficiently large that incorporates more significant crime hotspots. (This 'lack' of more than one crime hotspot of significant magnitude may be as a consequence of population distribution; that is an abundance of 'empty' space outside a single urban area). In these circumstances the PAI is not a 'fair' measure of predictive accuracy and precision, and this phenomenon could potentially be exploited by companies to artificially boost the apparent predictive accuracy of their crime forecasting algorithms. Figure 22 below illustrates how, in the case of highly clustered crime with a forecast based on the same total area 'a' (the sum of the areas of prediction, for example rectangular boxes) the PAI may be much larger if the enclosing area ('A') is chosen to be as large as possible but not to include further significant hotspots.

Figure 22 The same forecast and same total predictive area ('a') will tend to give rise to a higher PAI (right hand side) as the overall area ('A') is much larger on the right hand side, but does not include additional crime hotspots of the high magnitude

4.6.3   Overview of method adopted to assess predictive accuracy

Figure 23 below provides an overview of the method used to assess predictive accuracy of the four products.
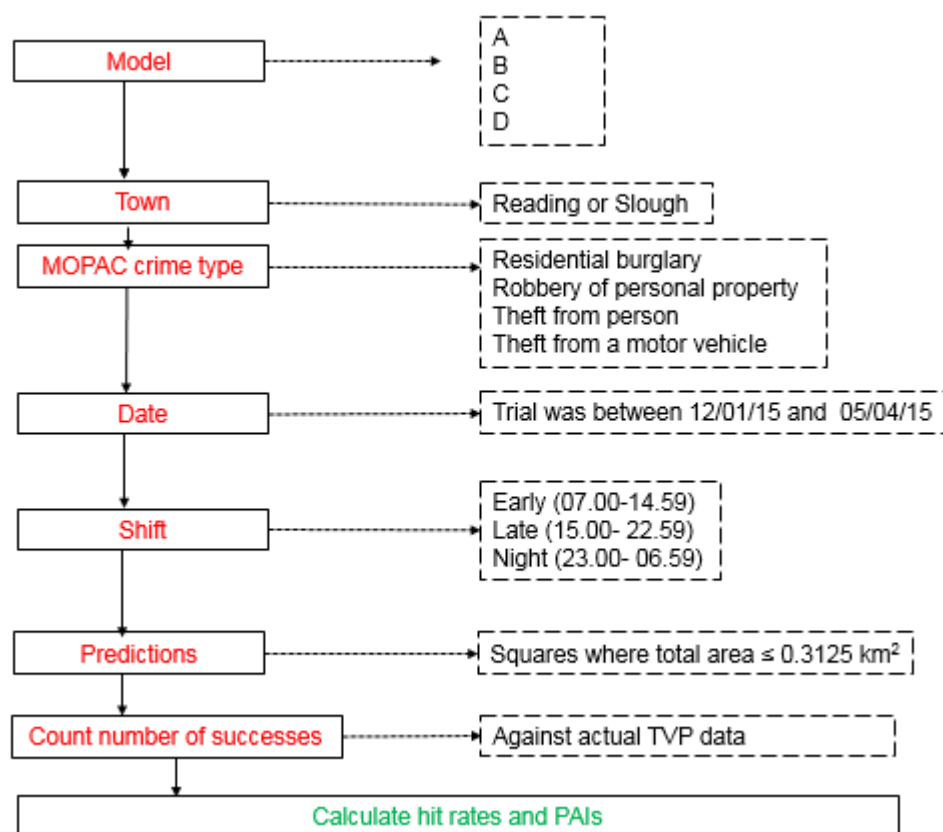

Figure 23 Overview of stages of analysis of predictive accuracy

4.6.4 Stages of predictive accuracy assessment

Table 10 below provides further details concerning each stage of the analysis. In some cases there are references to further information.

Stage of analysis	Notes
Model selection	The products have been anonymised for the purposes of this part of the evaluation and referred to as A, B, C and D. Five years of crime data from Thames Valley Police (TVP) was provided to each product in order to build and calibrate their models.
Town selection	Defined as the area covered by : Reading Borough Council (the policing areas of Mapledurham and Thames; Peppard and Caversham; Kentwood and Tilehurst; Abbey with Battle; Southcote / Norcot; Minster and Katesgrove; Redlands with Park; Whitley and Church) with a small 'buffer zone' close to the perimeter; or Slough Borough Council (the policing areas of Britwell / Haymill; Farnham /

	Baylis / Stoke; Wexham Lea / Central; Cippenham; Chalvey / Upton / Town; Langley / Kedermister; Colnbrook / Poyle / Foxborough) with a small 'buffer zone' close to the perimeter
MOPAC crime type selection	Defined in 4.1, Table 3 above
Date	The trial commenced on 12 January 2015 and 5 April 2015 (inclusive), with a short period before this to calibrate systems. One vendor began the trial later than 12 January 2015 but the hit rate and PAIs were calculated accordingly.
Shift	Three periods during the day: Early (0700-14.59) Late (1500-2259) and Night (2300-0659). Some products decided to produce predictions for a 24 hour period rather than for each shift. In this case to increase comparability, the single forecast was repeated for each shift.
Prediction	In consultation with the MPS, the maximum total area for prediction for each town was set at 312,500 m^2 (0.3125 km^2). This was largely for operational reasons as it judged to be a realistic size of area to be policed in any given shift. Most products chose to use between 5 and 14 rectangles (two dimensional 'boxes'), for example 14 boxes at approximately 150m by 150m. (See Table 11 below).
	Geolocation system used by three of the products was WGS84 ('latitude and longitude'), the other using six figure Ordnance Survey 'Eastings and Northings' (OSGB36). All geolocation data was converted into WGS84 format using both MatLab software and Ordnance Survey batch conversion tools (Ordnance Survey, 2015).
	Ranking although provided by some products was not used (as the absolute numbers of crimes were too small to warrant this). The ranking systems used by products was different with some using probability and others simple ordinal ranking. The problem with the latter approach is that it implies probabilities between each ranking are equal and this is certainly not likely to be the case.
Count number of successes	TVP geolocation data was in UTM 'Eastings' and 'Nothings' format (common to police forces) but converted into WGS84 format using MatLab software (to 3 d.p.) and using an Ordnance Survey (2015) conversion tool (to 11 d. p.). These results were then compared and the more accurate of the results used.
	The number of successes per town, MOPAC crime type, date and shift were calculated for each of the four products (see below). Note that hits were crime type specific, for example a crime of 'residential burglary' was counted if it occurred within the predicted box for 'residential burglary' (but not for one of the remaining three crime types).

Table 10 Description of stages of analysis of predictive accuracy

4.6.5   Model choice of rectangular boxes for predictions

Each vendor was set the challenge of producing forecasts that were less than a given total area of 0.3125 km^2 for each of the two towns (this was considered important for police operational reasons)

but were otherwise given freedom to vary the size and number of rectangles produced. Table 11 below provides a summary of each vendor's decisions. In some cases the total area exceeded the permitted maximum, but by a very small percentage and this was taken into account in the calculation of the PAIs.

Model	Average no. of rectangles per day, shift and crime type	Notes	Average length and width of rectangle	Average area of rectangle
A	12	5 with highest probability also generated	152.1m by 147.3m	22404 m^2
B	5	20 with likelihoods generated	250m by 250m	62500 m^2
C	5	Variability in size of rectangle	259.95m by 249.87m	64954 m^2
D	14	Variability in size of rectangle	152.1m by 147.3m	22404 m^2

Table 11 Description of rectangular 'prediction boxes' used by crime forecasting products

4.7   Counting successful 'hits'

In principle, counting the number of successful predictions is simple: if during a shift one or more recorded MOPAC crimes occur within a predicted square then each counts as a single 'hit' or success. However, in practice there are a number of other decisions required. The general approach adopted in this evaluation was to 'err on the side' of the products where doubt existed. So for example, if a crime occurred exactly on the perimeter of a rectangle this was counted as a success. Figure 24 below gives examples of successful predictions.
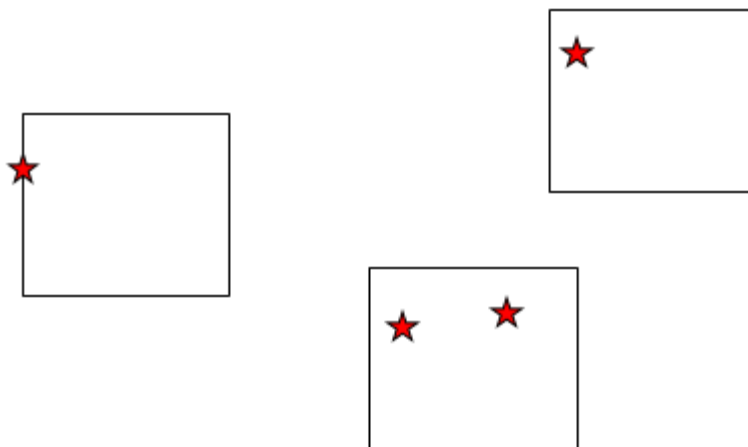

Figure 24 Examples of successful predictions

In a number of cases a model successfully predicted the same crime but with two or more overlapping rectangles (see Figure 25 below for an example). In this case the hit counted once only.
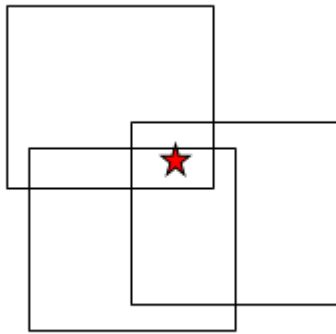

Figure 25 The same crime predicted by a model with three overlapping rectangles

4.8 Hit Rates and PAI Results

In this section of the evaluation we examine and discuss the results of the predictive accuracy trial in Reading and Slough.

4.8.1 Overall predictive accuracy results for Reading and Slough

Using the mid-point method of crime shift allocation hit rates for Reading varied between 8.7% and 17.7% with associated PAIs between 11.3 and 23.0. Marascuilo's Post Hoc Multiple proportion comparison was used to check for statistically significant differences in hit rates between the four products. No significant differences were found other than between the two 'extremes' (the lowest and highest hit rates).

Slough hit rates were much lower (between 4.0% and 10.1%) as were the PAIs (between 4.8 and 10.5). In terms of statistical significance, hit rates of two of the products in Slough were about the same, with two other products performing better in relative comparison. However, evidence for this pattern was not clearly reinforced by the observed PAIs.

Using the aoristic method of crime shift allocation the hit rates for all of the products were marginally higher for Reading (between 9.4% and 18.5%), as were the associated PAIs (12.1 to 23.8). In Slough the hit rates were also marginally higher for three of the four products (between 4.8% and 9.3%) as were the PAIs (from 4.9 to 9.7).

4.8.3 Results per crime type per town

The results were further disaggregated to examine the outcomes at the level of the four MOPAC crime types and within each town. However, in some cases the absolute numbers involved are very small and a difference of only a few hits could have made a significant difference to the outcome. The use of proportions (e.g. in the form of percentages) is particularly problematic in these circumstances as they become very sensitive to small changes.

Manual examination of the results showed that all products in Reading show a relatively consistent order of predictive accuracy by crime type:

- Burglary – 'very low' to 'low' predictive accuracy (hit rates of 0 – 5%)
- Theft from motor vehicle – 'low' predictive accuracy (hit rates of 1-10%)
- Robbery – 'low' to 'medium' predictive accuracy (hit rates of 0-20%)
- Theft from person – 'medium' to 'good' predictive accuracy (hit rates of 13- 54%)

Predictive accuracy in both regions was fairly inconsistent when ordering by crime type. However theft from person routinely outperformed other crimes in both areas, usually by about 3-5 times.

4.8.4 Discussion of results

An obvious form of comparison is with hit rates and PAIs found with comparable studies elsewhere. However, this is highly problematic and potentially misleading for a number of reasons. Firstly, there are problems with the PAI itself being dependent on decisions concerning the areas of measurement (see 4.6.2.2 above). Secondly, there are obviously differences in crime recording rules between countries and the definitions of crime will vary. Thirdly, the studies themselves will have significant variations in the methods adopted and will be of variable scientific quality. Most other studies do not examine forecast products that use relatively small rectangular boxes as the spatial unit of prediction. However, perhaps the biggest problem of all is that (as noted earlier) neither the hit rate nor the PAI take into account the time period for which predictions have been made, and in the case of the trials in Reading and Slough the use of 8 hour shifts probably put major demands on the forecasting ability of the products. Having said this, taken as a whole studies elsewhere are useful guides to the kind of hit rates and PAIs that are 'normal'. Table 12 below summarises the results found elsewhere, either in the literature, through correspondence or through the authors' calculations.

Reference	Location(s)	Duration of trial	Crime Types	Crime forecasting methods tested	Highest Hit rates found	Highest PAIs found	Notes
Politie (Amsterdam Police) (2015) (unpublished)	Amsterdam, Netherlands	Ongoing	Domestic burglary 'Mugging' Robbery	Risk Terrain Modelling with neural networking	C 41% (for 'mugging')	c. 14 (for 'mugging')	Forecasting method was developed 'in house'. Squares are 125m by 125m. PAI estimated but uncertain.

Kent Police (2015) (unpublished)	Kent, UK	Commenced 21/07/14 (ongoing)	16 offence types (including burglary, public order, robbery	Commercial product	19.2% (for shoplifting)	To be calculated	Predictions for two 12 hour 'shifts' per day (0700 to 1900; 1900 to 0700). PAIs calculated from data provided by Kent Police.
MSc dissertations (2014) (unpublished)	London boroughs	Variable, from two weeks to two years	Residential burglary Theft of motor vehicle Theft from motor vehicle Robbery Violence with injury Theft from the person Criminal damage	MBR Two commercial products	14.46% (theft from person)	21.91 (theft from person)	PAIs were calculated in London boroughs undertaking predictive policing trials. Some results unreliable.
Fan, S. (2014) *Using Spatial and Spatial-Temporal Predictive Accuracy Measures to Access (sic) the performance of the crime hotspot mapping methods*	Houston, Texas	12 months	Robbery Aggravated assault Burglary Larceny-theft Auto-theft	Risk-based thematic mapping Grid thematic mapping STAC (Spatial and Temporal Analysis of Crime) NNHC (Nearest Neighbour Hierarchical Cluster) Risk-adjusted NNHC KDE maps Local Moran's I Gi*	Larceny theft, ranging from 2.69% to 49.58%	Robbery, ranging from 0.09 to 54.39	Master's Thesis RRI also calculated. Unreliable results as insufficient detail provided
Drawve, G (2014*) A Metric Comparison of Predictive Hot Spot Techniques and RTM*	Little Rock, Arkansas	Three years	Robbery	STAC ellipse STAC convex Nnh ellipse Nnh convex KDE maps RTM	None given	An average of 77.473 (for KDE)	Paper claims that 'KDE excelled at short-term and long-term prediction'

Source	Location	Time period	Crime type	Method	Hit rate	PAI	Notes
Public Engines Inc (2014) *Predictive Analytics vs Hot Spotting* (unpublished)	Two cities (unspecified)	90 to 100 days	Not specified	'Predictive Analytics'	29.64%	4.01	Little information given on algorithms employed or method of measuring PAI (referred to as 'Observed Efficiency')
Drawve, G, Moak, S. & Berthelot, E. (2014) *Predictability of gun crimes: a comparison of hot spot and risk terrain modelling techniques*	Little Rock, Arkansas	Two years	Gun crime	Nnh RTM	7.5% (our calculation from data given in paper)	21.05 (for Nnh method)	Limited to gun crime only (which also included 'non-shooting' incidents)
Hart, T. & Zandbergen, P. (2013) *Kernel density estimation and hotspot mapping*	Arlington, Texas	12 months	Aggravated assault Robbery Commercial burglary Motor vehicle theft	KDE maps	Average of all crimes 41.52% Highest hit rate for robbery but value not cited.	Average of all crimes, 5.31. Highest PAI for robbery but value not cited.	Intention was to examine the effects on predictive accuracy of user-defined parameters within KDE mapping
Turner, G., Brantingham, J. & Mohler, G. (2014) *Predictive Policing in Action in Atlanta* (unpublished)	Atlanta, Georgia	90 days	Not specified, but appears to include burglary, auto theft and robbery	'PredPol'	None given	24.0	Results identified within *The Police Chief* magazine

Hart, T. & Zandbergen, P. (2012) *Effects of Data Quality on Predictive Hotspot Mapping*	Arlington, Texas; Albuquerque, New Mexico; Charlotte-Mecklenburg, North Carolina; Las Vegas, Nevada; San Diego, California; Tampa, Florida.	12 months	Aggravated or simple assault; auto burglary; auto theft; burglary; drug offences; homicide; robbery	Grid-based thematic mapping Local Moran's I Gi* KDE maps NNHC STAC	Robbery 61.2% (using Local Moran's I with Grid)	Robbery, 808.4 (using NNHC)	The study also considered the quality of data and influence of cell size and other parameters on predictive accuracy. Values of PAI much larger than in any other study found.
Swain, A. (2012*) A comparison of Hotspot Mapping for Crime Prevention* (unpublished)	MPS data (Boroughs unspecified)	Unspecified but '3 temporal scales' used	Burglary Theft from Motor Vehicle Theft of Motor Vehicle Theft from Person	STAC, convex hulls STAC, ellipses NNH, convex hulls Grid (G*) KDE (Q) KDE (G*) Geog Boundary (Q)	Not given	Theft from person, c. 4 (using STAC convex hulls)	Separates out measures of accuracy and precision. PAI corresponds with FPI in this paper.
Tompson, L. & Townsley, M. (2009) *(Looking) Back to the Future: using space-time patterns to better predict the location of street crime*	Boroughs of Camden and Islington, London	12 months	'Street crime', namely robbery and theft from the person	KDE maps	'Street crime' (only crime type tested) 24.1%	'Street crime' (only crime type tested) 7.04	Main purpose of paper was to assess the temporal sensitivity of hotspot maps, not predictive accuracy
Van Patten, I, McKeldin-Coner, J. & Cox, D. (2009) *A Microspatial Analysis of Robbery: Prospective Hot Spotting in a Small City*	City of Roanoke, Virginia	3 years	Street robberies	LISA STAC, ellipses STAC, convex hulls NNH, ellipses NNH, convex hulls KDE	48.9% (STAC, convex hull)	43.63 (NNH, convex hull)	Study focused on varying prediction base and short term v. long term forecasting

| Chainey, S., Tompson, L. & Uhlig, S. (2008) *The Utility of Hotspot Mapping for Predicting Spatial Patterns of Crime* | Boroughs of Camden and Islington, London | 12 months | Residential burglary 'Street crime' (combining robbery of personal property and theft from person) Thefts from vehicles Thefts of vehicles | STAC (using ellipses of various sizes) Thematic mapping of output areas Grid thematic mapping KDE maps | 'Street crime' 20% | 'Street crime' 4.68 (for KDE) | Used 12 months retrospective data to generate hotspots and subsequent 12 months data to assess predictive accuracy |

Table 12 Hit rates and PAIs for studies undertaken between 2008 and 2015 inclusive.

The PAIs for the predictive products in Reading ranged between approximately 11 and 23 and as can be seen from Table 12, these values are within the expected range. PAIs for Slough were significantly lower, ranging between approximately 5 and 10.

In terms of the 'crime geography' of Reading and Slough we noted above that Reading's population is more 'concentrated' towards the centre whereas Slough appears more dispersed. In absolute terms, there were fewer of the four MOPAC crimes occurring in Reading (see Tables 6 and 7 above) and these were also more clustered (particularly in the case of 'theft from person' – see below and Table 8 above). Hence it might have simply been the case that the products had a somewhat easier task in Reading to identify a bigger proportion of fewer crimes that were concentrated within a relatively small geographical area.

In terms of each of the four MOPAC crimes in Reading it is probably no surprise that the particular order of 'predictability' was as found. The Reading results reflect earlier findings by the MSc students in the study conducted for the MPS in summer 2014. For example, one MSc student found in one London borough that the PAI for robbery was 10.93, burglary 3.27 and theft from motor vehicle 2.73. The Reading order of forecasting accuracy also reflect a number of findings elsewhere in the UK and beyond, which find that acquisitive crimes are strongly correlated with population, with particular spaces and times and hence prove to be among the 'easiest' to forecast. For example, 'shoplifting' tends to occur where there are more shops to steal from and pick-pocketing where there are people available in large numbers and in close proximity to one another and these kinds of opportunities are fixed in space and (in the case of Reading) available to motivated offenders in a small number of locations.

4.8.4.1 Theft from person in Reading

The MOPAC crime of 'theft from person' in Reading had consistently higher hit rates and PAIs for all four forecasting products.

The MOPAC crime 'theft from person' covers theft (including attempts) of item or items (e.g. handbag, wallet, cash, smartphone) directly from the victim, but without the use of physical force against the victim, or the threat of it. The crime can be subdivided into 'snatch thefts' (where there may be an element of force involved but this is just enough to snatch the property away) or 'stealth thefts' (e.g. pick-pocketing), where no force is used and the victim is unaware at the time of the crime. Nationally, stealth theft makes up around 70-80% of theft from the person incidents (Home Office, 2013). Research also suggests that offenders are prepared to travel relatively long distances

(up to 17 miles) to hot spot areas (shops, cafes, night time venues (pubs, clubs), public transport and public transport hubs) to commit 'theft from the person' (ibid.)

As part of the evaluation we examined in detail the predictions of one of the products for the crime of 'theft from person' in Reading and superimposed this on a satellite image of Reading (see Figure 26 below).
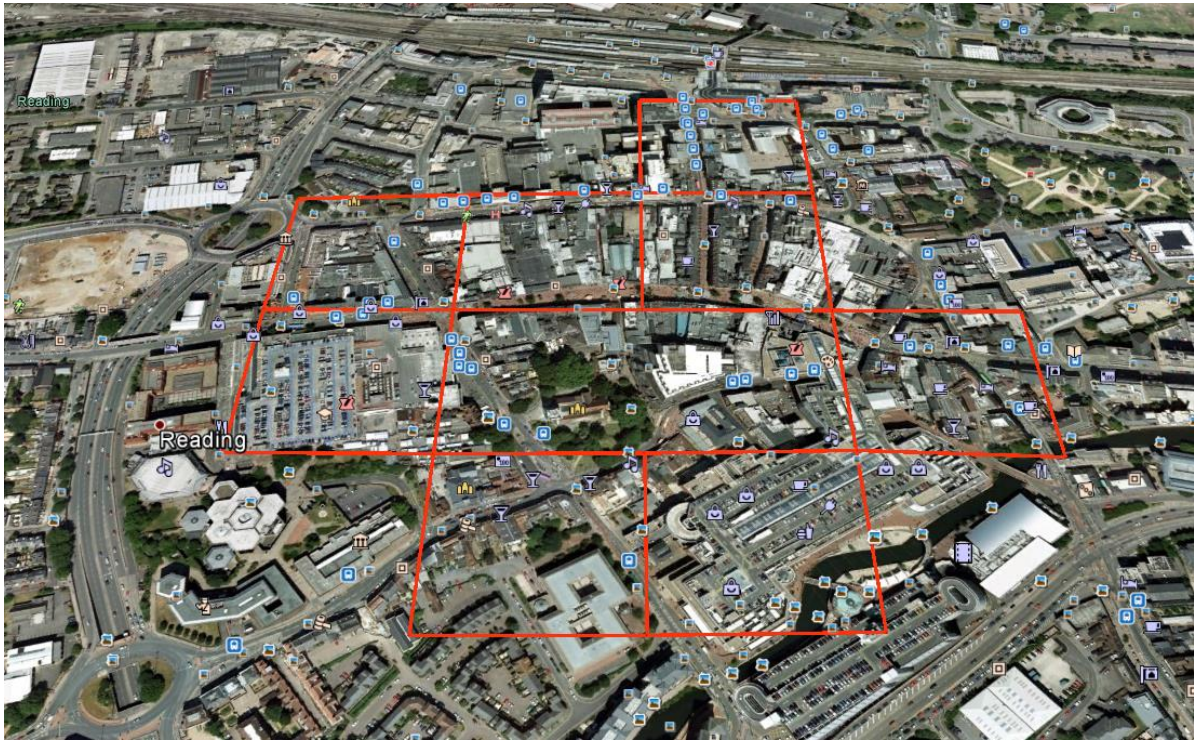

Figure 26 Model A predictions for theft from person in Reading (some rectangles superimposed), Image ©2015 Google.

It transpired that a large proportion of the successful predictions for 'theft from person' occurred entirely in the centre of Reading within one of the main commercial and shopping centres (an area including the Oracle Shopping Centre, which alone includes 90 retail stores, over 30 places to eat and drink and a 10 screen cinema). Further, just over half the hits recorded for the model for the crime of 'theft from person' during the trial were during the night shift (after 11pm).

In contrast, most of the products were least 'successful' in forecasting 'theft from vehicle' crimes in Slough, and as Figure 20 above illustrates, these were the least clustered and the most dispersed of all the MOPAC crimes studied within the trial.

5   Operational implementation

Only a limited evaluation of the operational implementation of predictive policing in London was possible as this part of the evaluation was conducted largely through secondary research.

In terms of operational implementation, we could perhaps summarise the rationale behind predictive policing as follows. Forecasting crime, sufficiently in advance of events, on the basis of relatively small geographical areas (often rectangular boxes) and within relatively short time periods (sometimes a shift) provides the basis to proactively deploy police resources to the areas at risk. The most common police operational response is to deploy police resources in the form of visible police patrol. Deployment is determined on the basis of shift and with the aim of providing an appropriate level of 'patrol dosage'. This strongly suggests that the underlying operational intent of predictive policing is one of effective dissuasion rather than detection – that is to prevent the crime occurring, rather than catching the offender in the act (otherwise non-regular visible patrol would be avoided and more intelligence-led techniques employed instead).

There is evidence to support this rationale in the literature but only in parts. In terms of shift deployment, Johnson et al. (2007) suggests that there is a good empirical case for crime forecasting being aligned to police operational shifts. However, in terms of supporting the assumption that visible policing (and even simply physical presence) has a deterrent effect on crime, the evidence is inconsistent. Whilst a number of studies have found that additional foot patrols have reduced personal robberies (Jones and Tilley, 2004) with many other crime types there is no conclusive evidence. It could be that it is not simply the presence of visible policing in a 'predictive box' that is most important for deterring crime, but the kind of action that police officers undertake when they are there (see Ratcliffe et al., 2011).

The use of rectangular boxes on a map is also perhaps not the most intuitively resonating means of conveying crime risks, in terms of either location or time.   It is highly unlikely that any police patrol will feel the need to 'robotically' deploy to all parts of an artificial square measuring 150m by 150m. Instead they are likely to fix on a building, parking bay, corner of two streets, shop, pub or other landmark or convenient location within the predicted boxes as the 'anchor' from which to police the area.   Research in evolutionary psychology and neurology suggests that modern people still intuitively think of their environment in ancestral ways rather than in terms of the geometry of urban rectangles and squares. It is also the case that some police officers, particularly those new to policing (and who might well live outside of the London borough that they police) will be unfamiliar with the local urban geography

5.2   'Patrol Dosage' Rates

The term 'dosage' is used by police forces as a shorthand for the proportion of time a visible police presence is physically located within a predictive box during a designated time interval (such as a shift). It is an inevitable fact of operational policing that the MPS is unable to deploy visible police resources to a predictive policing geographical area (a predicted square or circle) during the complete time period covered by a shift. Measuring dosage is therefore important for a number of reasons. If visible patrol does have a deterrent effect (see 5.1 above) then the effect might vary in size according to the time spent within a prediction area. The so-called 'Koper Curve' is often cited in support of this – research in the early 1990s in Minneapolis apparently found that ideal dosage was 10 to 15 minutes (Koper, 1995). In 2011 Sacramento Police Department deployed highly visible policing in crime hot spots for 12-16 minutes every two hours and apparently observed a 25%

decrease in crime over a three month period in these areas (COPS, 2012). However, there have been few other reliably conducted studies to verify the existence of the Koper curve.

For example, if dosage rates are sufficiently high then police visibility might have a dissuasive effect on certain crimes and hence reduce the hit rates (based on recorded crimes) and the associated PAI.

5.2.1 MPS predictive policing dosage

Most, but not all of the MSc students examined the phenomenon of 'dosage' during the summer of 2014. There were a number of attempts to fit mathematical models to dosage data patterns and one or two of these appeared largely successful (particularly those that employed Poisson regression). For example, there were statistically significant correlations between dosage rates and both location and time of day. (In terms of location, predicted areas closer to police stations and public amenities such as railway stations and dosage rates; in terms of time, differences between shifts). Another significant finding was the association between dosage rate and perceptions of 'traditional' crime hotspots. Put simply, those predicted areas that happened to coincide with already perceived hotspots were patrolled more often and for longer. Less certain were the findings on the mean and median of the dosage times but a very approximate estimate seems to put this at about 10 minutes for foot patrols and approximately 20 minutes for vehicle patrol per shift (8 hours). However, the standard deviations found were particularly high suggesting a great deal of variability in the dosage levels. Even less certain were the findings on any association between dosage rates and any subsequent impact on crime rates.

Further analysis of MPS dosage in the three boroughs of Hackney (which employs the MPS MBR model), Lewisham and Southwark was conducted by Oli Hutt from University College London in May 2015 (using data collected from October 2014 until January 2015). The results were mixed and showed both slight increases and decreases in crime depending on the time period utilised and the Borough examined, though overall the results support a slight, but statistically non-significant decrease in crime attributable to prediction guided patrolling (communication).

Kent Police have also monitored dosage rates since introducing a commercial crime forecasting model in the county. The mean dosage time for predictive boxes in Kent and Medway is 30 minutes (from a possible 12 hours), with a standard deviation of 42 minutes (21,372 observations from 1 August 2014 until 31 January 2015) (authors' calculation based on communication). As with the 'MSc' findings described above, this is a relatively high standard deviation suggesting a wide variation in dosage times in Kent. Further analysis is being undertaken on the Kent Police data.

5.3   Stakeholder surveys

Two surveys have been conducted within the MPS to assess stakeholder (principally user) opinion and experiences of employing predictive policing in the capital.

5.3.1   'MSc' stakeholder survey, summer 2014

As part of their dissertations, the MSc students conducted surveys of police officers and staff in a number of London boroughs on the 'operational implementation' of the predictive policing products being trialled. All of the dissertations surveyed police employee views of predictive policing, including understanding of intentions of predictive policing, implementation and ease of use. Sample methodologies and sizes varied significantly between studies making it difficult to draw generalisations (methods included questionnaires and semi-structured interviews). Nonetheless the following seemed to emerge as common themes:

•   There was a low operational take up of one of the predictive policing products (c. 16% of the maximum possible).

•   In some cases the operational use of the maps generated from predictive policing significantly declined over the period of study (possibly because changes in leadership had led to the perception of changing priorities).

•   A significant proportion of operational officers expressed the opinion that BOCU resources were insufficient to ensure satisfactory dosage rates (see 5.2 above).

•   Most of those surveyed found the underlying principles of operational predictive policing easy to understand, but were less certain concerning the theoretical underpinning.

•   Where relevant, users found all the mapping software employed (which used product predictions to produce geographic locations in either square or circle format) easy to use, although (on occasions) slow.

•   There was an overwhelming preference for maps which were easy to understand and which lent themselves to clear operational decisions.

•   Although views on predictive policing as an operational tactic were often positive (particularly amongst more senior managers) the opinion of a significant proportion (from 1/3rd to 2/3rds) of the police officers surveyed could be considered as 'sceptical'. These officers tended to be less senior, but more experienced and longer in service. There were a number of reasons given for the scepticism, but the most frequent observation was that either the boxes/circles tended to fall in geographical areas that the officers already perceived as likely to suffer higher levels of crime (and hence were no surprise) or, when they fell outside of the usual locations, few crimes were actually encountered. It should be noted that no objective evidence was offered to support these views (nor requested in the surveys).

•   A number of those interviewed expressed frustration at their perceived inability to give feedback on the success or otherwise of the predictions being supplied by products.

5.3.2   MPS stakeholder survey, Spring 2015

A further survey was undertaken of stakeholder views by the MPS in Spring 2015. The survey was conducted using an online tool and elicited c. 100 responses. As with the 'MSc' survey (see 5.3.1 above), caution needs to be applied in over-interpreting these results as the survey was 'self-selecting' and non-stratified and this can sometimes give rise to bias.

Responses were received from constable rank (57% of the sample), sergeant (28%), inspector (6%) and 'other' (9%). The majority (68%) of respondents had worked in their current post for more than one year (with 28% over five years). The MPS 'in-house' predictive policing crime maps had been used by approximately 44% of respondents, 56% had used the commercial products. As with the 'MSc' survey, a majority of respondents (55%) found the predictive crime mapping product easy to use but a significant minority (37%) were either neutral or found the products not easy to use (the remainder expressing a 'don't know' response). However, the data collected was aggregated in such a way that it is impossible to determine if the satisfaction rates varied significantly between products. Taken in conjunction with the 'MSc' survey (see above) the MPS survey also seems to indicate that the use of the maps is in decline, with less than one in four (23%) of the respondents using the maps on a shift basis and a majority (52%) either using them only 'sporadically' or not at

all. Those that do use the maps tend to use them to mostly 'inform street patrol routines' (66%) or to 'task further intelligence and analytical work' (24%).

The number of officers that are sceptical concerning the perceived accuracy of the mapping remains at about the same level as found with the 'MSc' surveys (in the MPS survey approximately 56% of the MPS survey respondents judged the predictions to be either 'no more accurate than methods or products used before' or 'inaccurate'). However, as with satisfaction rates quoted above, it is currently not possible to determine whether respondents varied in their views according to the predictive policing product they were utilising.

6   Observations and recommendations

Our observations and recommendations are divided into three sections: those that relate to the outcome of the predictive accuracy trial in Reading and Slough, those concerned with the MPS operational implementation of predictive policing and finally, those recommendations that would follow if the MPS decide to invest in a particular 'predictive policing' product or in-house solution.

6.1   Predictive Accuracy trial in Reading and Slough

- The products appeared to have achieved rates of predictive accuracy within the range found elsewhere in the literature, despite the demanding restriction of a relatively small total proportional geographic area in which to predict. However, considerable doubts remain over the reliability and validity of the predictive accuracy trial.
- There were differences in performance between the products in terms of hit rates and PAIs. However, there was little consistency in terms of statistically significant differences between products when Reading and Slough were compared.
- For all four products examined, the predictive accuracy results for Reading were higher than those for Slough. Most of the success in Reading was through forecasting theft from person crimes in the centre of Reading (within the main commercial and shopping locations).
- There is some limited evidence that most products produced their highest predictive accuracy results with crimes which are geographically highly clustered in a small number of locations.
- There is some very limited evidence that algorithms which are more responsive to the underlying morphology of urban locations provide greater predictive accuracy. However, further research is needed to test this.

6.2   Operational implementation

- According to the surveys conducted, operational take up of predictive policing within London boroughs remains patchy and there is some limited evidence that it might be declining.
- Most police officers and staff surveyed find the predictive policing maps easy to use.
- The majority of MPS police officers surveyed remain sceptical concerning the perceived accuracy of the predictive policing mapping.
- There is much spatial and temporal variation in 'dosage' rates and it is unclear what effect visible police presence has in deterring crime in areas forecast to experience crime.

6.3   Recommendations in the event of adoption of a crime forecasting model

Of particular importance in deciding whether to invest in commercial crime forecasting models is the cost-benefit calculations involved. Whilst increases in predictive accuracy achieved through the use of commercial products may be *statistically* significant the cost may not be financially justifiable. There are cost-benefit analyses to be conducted and as a result it might be concluded that the gains in marginal increases in the ability to forecast crime might be outweighed by the costs in both paying for the crime forecasts and the ongoing investment required in 'policing the predictions'.

In the event that the MPS invests in a 'predictive policing' product we would recommend the following.

- Allow the model(s) selected the best possible chance of forecasting crime in London. So, for example, do as much as possible to minimise geocoding errors throughout the system. The

MPS could also consider how it could widen the data available to products to help improve the crime forecasting. The Amsterdam police (see 2.4.7 above) utilise known offender locations and prison release data as part of the Risk Terrain Analysis that forms part of their approach.

- Consider whether it is possible to commission (and only pay for, if relevant) crime forecasting that is limited to localities and crime types that are likely to produce the higher predictive accuracy indices.

- Products should be asked to describe in more detail the underlying theory adopted and the assumptions made to develop their algorithms and not expect MPS officers and staff to rely on the output from models that simply provide (in effect) a series of geolocations and rank orders (albeit with relatively user-friendly GUIs). So called 'black box' approaches might be superficially attractive but knowledge transfer as well as good customer-client relationships are needed to develop successful predictive policing.

- Independent ongoing evaluation of the performance of the products should be conducted (rather than relying on the findings of the products themselves). This could include hit rates, PAIs (adjusted to take into account crime morphology) but also more sophisticated measures when longer periods of data become available. A particularly important piece of research would be to measure both **effectiveness** (e.g. crime reduction) and **efficiency** (e.g. operationally optimal box sizes and ranking). This evaluation might perhaps be undertaken by the MPS analytical capability that is already in place, using Randomised Control Trials and other means.

- Monitor and measure the operational implementation of the model. There are a number of often implicit assumptions that require testing 'in the field'. One assumption that seems widely shared in the MPS is that visible patrol has a deterrent effect on crime within a predictive box. Claims made by commercial companies for the success of predictive policing echo and implicitly encourage this assumption (e.g. PredPol, 2015a). The assumption that deterrence automatically follows could well be justified for some types of crime but will certainly not be true of all the MOPAC crime types. Similarly, the focus on dosage arises through a belief that the 'Koper curve' applies in these circumstances. This belief also needs testing in the context of London.

- Increase awareness amongst constables and others of what predictive policing is, but also what it is not. Crime forecasting produces probabilities, but these are not particularly intuitive and are generally not well understood. (There is a direct analogy here with weather forecasting).

- Consider responses other than, or in addition to, visible patrol as a response to the forecasts of crime generated by the model. For example, it could be that situational crime prevention is a better response to 'micro hot spots' or a more intelligence-led approach and (as Karn, 2013, p. 16 notes) 'particularly so in areas of high population turnover, where length of residence, social organisation and mutual trust, may be […] less'.

- Much of what is referred to in the media and within policing as 'predictive policing' is in reality crime forecasting conducted using mathematical algorithms. The predictive accuracy of these algorithms has yet to be fully scientifically assessed. The 'policing' aspects of 'predictive policing' appear to us to be underdeveloped and the MPS might reasonably expect and demand more of a product. For example, a commercial supplier could be asked to provide different sizes of predictive policing areas within the same shift prediction, and not simply produce rectangles of identical sizes. Forecasting could better reflect the reality that 'dosage' is very subject to rapidly changing circumstances within operational policing. Despite the advent of GPS and satellite navigation, predictions could also be given that

relate more to the lived experience of police officers and which describe locations more intuitively (e.g. close to Spar minimarket on the junction of Arvon Road and Drayton Park) rather than as a series of rectangles.

References

Caplan, J., Kennedy, L. & Piza, E. (2013) *Joint Utility of Event-Dependent and Environmental Crime Analysis Techniques for Violent Crime Forecasting* Crime & Delinquency 59(2), pp. 243-270

Chainey S, Thompson L and Uhlig S (2008) "The utility of hotspot mapping for predicting spatial patters of crime", *Security Journal*, 21 pp.4-28.

Cohen L and Felson M (1979) "Social change and crime rate trends: a routine activities approach" *American Sociological Review*, 44, 588-608

COPS (2012) *A Hot Spots Experiment: Sacramento Police Department* [Online] Available at: http://cops.usdoj.gov/html/dispatch/06-2012/hot-spots-and-sacramento-pd.asp

Drawve, G, Moak, S. & Berthelot, E. (2014) Predictability of gun crimes: a comparison of hot spot and risk terrain modelling techniques Drawve, G. (2014) *A Metric Comparison of Predictive Hot Spot Techniques and RTM* Justice Quarterly

Eck J, Chainey S, Cameron J, Leitner M and Wilson R (2005) "Mapping crime: understanding hot spots" USA: National Institute of Justice

Fan, S. (2014) *The Spatial-Temporal Prediction of Various Crime Types in Houston, TX Based on Hot-Spot Techniques* Master's Thesis, Louisiana State University

Goode, E (2011) "Sending the Police Before There's a Crime" New York Times 15th August 2011

Gorr, W. & . Olligschlaeger, A. (1998) Crime Hot Spot Forecasting: Modelling and Comparative Evaluation

Harrell, K. (2015) *The Predictive Accuracy of Hotspot Mapping of Robbery over Time and Space* [online] Available at: https://www.ucl.ac.uk/jdi/events/int-CIA-conf/icia-15/ICIAC15_POSTER_KHARRELL.pdf [Accessed 12 May 2015]

Hart, T. & Zandbergen, P. (2012) Effects of Data Quality on Predictive Hotspot Mapping Final Technical Report, National Institute of Justice, Available at: https://www.ncjrs.gov/pdffiles1/nij/grants/239861.pdf

Hart, T. & Zandbergen, P. (2012) *Effects of Data Quality on Predictive Hotspot Mapping*

Hart, T. & Zandbergen, P. (2013) *Kernel density estimation and hotspot mapping*

Heffner, J. (2014). Predictive Policing. *GeoConnexion International*, [online] (July/Aug), pp.20-23. Available at: http://www.geoconnexion.com/uploads/publication_pdfs/int-v13i7-predictive-policing.pdf [Accessed 1 Apr. 2015].

HMIC (2013) *Technical note: How Most Similar Groups are formed* [Online] Available at: https://www.justiceinspectorates.gov.uk/hmic/media/most-similar-groups-technical-note.pdf

Home Office (2013) Theft from the person Information pack for partners July 2013 [online] Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/282747/TheftFromPerson.pdf

Huet, E. (2015). *Server And Protect: Predictive Policing Firm PredPol Promises To Map Crime Before It Happens*. [online] Forbes. Available at: http://www.forbes.com/sites/ellenhuet/2015/02/11/predpol-predictive-policing/ [Accessed 29 Mar. 2015].

Johnson S, Guerette R and Bowers K (2014) "Crime displacement: What we know, what we don't know and what it means for crime reduction" *Journal of Experimental Criminology*, Vol 10, Issue 4, pp.549-571

Kahn, J. (2013) *Policing and Crime Reduction The evidence and its implications for practice* The Police Foundation [online] Available at: http://www.police-foundation.org.uk/uploads/catalogerfiles/policing-and-crime-reduction/police-foundation-police-effectiveness-report.pdf

Katarandjian, O. (2011) "Hurricane Irene: Pop-Tarts top list of hurricane purchases" ABC News 27th August

Koper, C. (1995) *Just enough police presence: Reducing crime and disorderly behavior by optimizing patrol time in crime hotspots*, Justice Quarterly, 12(4): 649-672.

Levine, N. (2008). The "hottest" part of a hotspot: Comments on "The utility of hotspot mapping for predicting spatial patterns of crime." Security Journal, 21, 295-302.

Mohler, G., Short, M., Brantingham, P., Schoenberg, F. and Tita, G. (2011). Self-Exciting Point Process Modeling of Crime. *Journal of the American Statistical Association*, Vol. 106, No. 493, pp.100-108.

MOPAC (2012) [Online] MOPAC Challenge Quarterly Performance Paper October 2012 Available at: https://www.london.gov.uk/sites/default/files/MOPAC%20Challenge%20Performance%20Paper_September%202012.pdf

Newburn, T (2013) *Criminology*, London: Routledge.

Ordnance Survey (2015) *Coordinate Batch Converter Files* [online] Available at: http://www.ordnancesurvey.co.uk/business-and-government/help-and-support/navigation-technology/os-net/batch-converter-file-formats.html (Accessed 15 June 2015).

ONS (2011), *Resident population estimates by ethnic group* (Excel spreadsheet) [online] Available at: http://www.ons.gov.uk/ons/rel/census/2011-census/key-statistics-for-local-authorities-in-england-and-wales/rft-table-ks201ew.xls (Accessed 12 May 2015)

Oulette, D (2012) "A Hot Spots Experiment: Sacramento Police Department", *Community Policing Dispatch*, Vol. 5, No. 6

Patten, I, McKeldin-Coner, J. & Cox, D. (2009) *A Microspatial Analysis of Robbery: Prospective Hot Spotting in a Small City*

Perry W, McInnis B, Price C, Smith S and Hollywood J (2013) "Predictive Policing: The role of crime forecasting in law enforcement operations" Cambridge: Rand

Ratcliffe, J. (2014) *What is the Future …of Predictive Policing?* Translational Criminology, George Mason University, Spring 2014, pp. 4 -5.

Ratcliffe, J., Taniguchi, T., Groff, E. & Wood, J. (2011) *The Philadelphia Foot Patrol Experiment: a Randomised Controlled Trial of Police Patrol Effectiveness in Violent Crime Hotspots* Criminology, Volume 49 , Number 3, pp. pp. 795-831.

Reiner R (2010) "The Politics of the Police" Oxford: Oxford University Press

Schoenberg, F. (n.d.). *Introduction to point processes*. [online] Stat.ucla.edu. Available at: http://www.stat.ucla.edu/~frederic/papers/encycpiece [Accessed 28 Mar. 2015].

Sklansky, D (2011) "The Persistent Pull of Police Professionalism" Washington DC: National Institute of Justice, Paper for the Harvard Executive Session on Policing and Public Safety

Swain (2012) A Comparison of Hotspot Mapping for Crime Prediction Available at: http://www.geos.ed.ac.uk/~gisteac/proceedingsonline/GISRUK2012/Papers/presentation-72.pdf

Tompson, L. & Townsley, M. (2009) (*Looking) Back to the Future: using space-time patterns to better predict the location of street crime*

Turner, G., Brantigham, J. & Mohler, G. (2014) Predictive Policing in Action in Atlanta, Georgia., Technology Talk, The Police Chief 81 (May 2014) 72-74, Available at: http://www.policechiefmagazine.org/magazine/index.cfm?fuseaction=display&article_id=3341&issue_id=52014

Van Patten, I. T., McKeldin-Coner, J., & Cox, D. (2009). A microspatial analysis of robbery: Prospective hot spotting in a small city. Crime Mapping, 1, 7-32.

Wright, A (2002), *Policing: An introduction to concepts and practice* (Cullompton: Willan).