BSGLg, 60, 2013, 7-19

# EXAMINING THE INFLUENCE OF CELL SIZE AND BANDWIDTH SIZE ON KERNEL DENSITY ESTIMATION CRIME HOTSPOT MAPS FOR PREDICTING SPATIAL PATTERNS OF CRIME

Spencer CHAINEY

**Abstract**

Hotspot mapping is a popular technique used for helping to target police patrols and other crime reduction initiatives. There are a number of spatial analysis techniques that can be used for identifying hotspots, but the most popular in recent years is kernel density estimation (KDE). KDE is popular because of the visually appealing way it represents the spatial distribution of crime, and because it is considered to be the most accurate of the commonly used hotspot mapping techniques. To produce KDE outputs, the researcher is required to enter values for two main parameters: the cell size and bandwidth size. To date little research has been conducted on the influence these parameters have on KDE hotspot mapping output, and none has been conducted on the influence these parameter settings have on a hotspot map's central purpose – to identify where crime may occur in the future. We fill this gap with this research by conducting a number of experiments using different cell size and bandwidth values with crime data on residential burglary and violent assaults. We show that cell size has little influence on KDE crime hotspot maps for predicting spatial patterns of crime, but bandwidth size does have an influence. We conclude by discussing how the findings from this research can help inform police practitioners and researchers make better use of KDE for targeting policing and crime prevention initiatives.

**Keywords:**

hotspot analysis, kernel density estimation, crime prediction, cell size, bandwidth, burglary, violent crime

***Résumé***

*La cartographie des points chauds (hotspots) est une technique populaire pour orienter les patrouilles de police et assister d'autres initiatives visant à la réduction de la criminalité. Il existe un certain nombre de techniques d'analyse spatiale qui peuvent être utilisées pour identifier les points chauds, mais la plus populaire au cours des dernières années est l'estimation à noyau de densité (Kernel Density Estimation – KDE). KDE est très populaire en raison de la manière visuellement attrayante dont elle représente la distribution spatiale de la criminalité, et parce que la méthode est considérée comme la plus précise parmi les techniques de cartographie des points chauds couramment utilisées. Pour produire des résultats avec KDE, le chercheur est tenu de fixer les valeurs de deux paramètres principaux : la taille des cellules et la taille de la fenêtre de convolution. A ce jour, peu de recherches ont été menées sur l'influence qu'ont ces paramètres sur l'interprétation finale d'une carte des points chauds – à savoir, identifier où la criminalité peut se produire dans l'avenir. Nous comblons cette lacune avec cette recherche, en effectuant un certain nombre d'expériences en utilisant différentes tailles de cellules et de valeurs de fenêtre, avec des données de la criminalité sur les cambriolages résidentiels et les agressions violentes. Nous montrons que la taille des cellules a peu d'influence sur les cartes de points chauds de criminalité issues de KDE pour prédire la répartition spatiale de la criminalité, mais par contre la taille de la fenêtre a une influence. Nous concluons en discutant de la manière dont les résultats de cette recherche peuvent aider à informer les praticiens de la police et assister les chercheurs dans une meilleure utilisation de KDE permettant de mieux cibler les initiatives de prévention du crime et de maintien de l'ordre.*

***Mots-clés***

*Analyse des points chauds, estimateur à noyau de densité, prédiction de criminalité, taille de cellule, taille de fenêtre, cambriolage, agression violente*

## I. INTRODUCTION

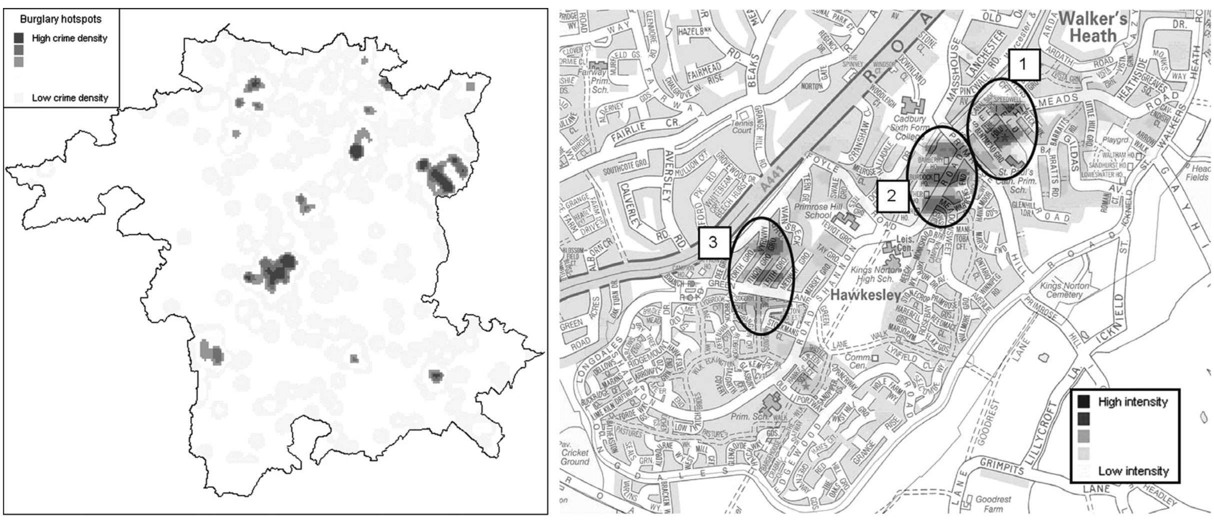
The mapping of hotspots of crime has become common practice in police agencies across the world. A hotspot is defined as being an area of high concentration of crime relative to the distribution of crime across the entire study area (Home Office, 2005; Chainey and Ratcliffe, 2005; Sherman, 2009). In these terms, hotspots can exist at different geographic scales of interest, whether it is at the city level for exploring localities where crime is highest, or at a local residential housing estate level, identifying particular streets or clusters of buildings where crime is seen to highly concentrate. Hotspot analysis has been applied to many forms of crime: from the analysis of gang-related murders in Belo Horizonte, Brazil (Beato, 2008), violent crime in Philadelphia (Ratcliffe et al., 2011), residential burglary, street robbery and vehicle crime in London (Eck et al., 2005), and street assaults in Melbourne, Australia (Mashford, 2008).

The use of hotspot mapping has also helped initiate the concept of hotspot policing – the targeting of police patrol strategies to crime hotspots in an effort to reduce the high volume of crime that is committed at these locations (see Braga, 2007 and Ratcliffe et al. 2011 for examples). Hotspot maps are also routine outputs that feed into *Compstat* style meetings (for a description of *Compstat* and examples see Chainey and Ratcliffe, 2005 and Home Office, 2005) and the intelligence production process of the UK's National Intelligence Model (NPIA, 2010). Hotspot mapping has therefore become a ubiquitous application in contemporary policing.

There are many spatial analysis techniques that can be applied to produce hotspot maps of crime. These include the use of spatial ellipses as shown by Block and Block (2000) when analysing hotspots around rapid transit stations in Chicago, applying a thematic (or choropleth) mapping approach to geographic administrative units as illustrated by Ratcliffe and McCullagh (2001) in their analysis of burglary across a study area's census zones, and grid thematic mapping as used by LeBeau (2001) to map patterns of emergency calls and violent offences in North Carolina. However, it is the use of kernel density estimation that in recent times has become the technique of choice by police practitioners and researchers (Chainey et al., 2008a), and as illustrated by examples of hotspot maps presented at the 2012 International Crime and Intelligence Analysis Conference (Figure 1).

Kernel density estimation (KDE) is also considered to be the most accurate of these *common* hotspot mapping techniques. This was illustrated by Chainey et al. (2008a) in a study that compared the hotspot mapping outputs generated using spatial ellipses, thematic mapping of census areas, grid thematic mapping and KDE for their ability to predict spatial patterns of crime. That is, based on the principle that hotspot mapping is used as a basic form of crime prediction – it uses data on past incidents to determine where crime may occur in the future - they showed that KDE outputs consistently produced better prediction results in comparison to the other techniques.

Like many spatial analysis techniques KDE requires the researcher to determine the values to enter for certain technical parameters in order to produce mapping




(a) (b)

**Figure 1.** Examples of kernel density hotspot maps of crime, as presented at the 2012 International Crime and Intelligence Analysis Conference (a) hotspot map of burglary for the county of Worcestershire (UK) and (b) hotspot map of criminal damage in Hawkesley, Birmingham (UK).

output. These two main parameters are the value for the cell size (sometimes referred to as the resolution) and the value for the bandwidth (often also referred to as the search radius). An alternative method to specifying a fixed bandwidth is the adaptive KDE approach where the bandwidth varies based on a user determined number of neighbours to include in the kernel density calculation. This adaptive kernel approach is rarely used by crime mapping practitioners (Chainey et al., 2008), hence the focus of this research was towards the more commonly used fixed kernel bandwidth approach. There is currently very little guidance on cell size and bandwidth size selection for the practical application of KDE hotspot mapping in policing, with the researcher either giving little thought to these values and their influence, settling for the default values determined by their KDE software application, or drawing from their own particular whims, fancies or experience (Eck et al., 2005; Chainey and Ratcliffe, 2005).

In this paper we aim to better inform practitioners and researchers by examining the influence that cell size and bandwidth size have on KDE hotspot mapping outputs. We follow as a guide the methodology used by Chainey et al. (2008a) that compared the spatial prediction measures of different hotspot mapping techniques, by comparing the influence that different cell size and bandwidth size values have on the spatial prediction abilities of KDE hotspot mapping outputs.

Section 2 describes in further detail the kernel density estimation function and how cell size and bandwidth values fit mathematically into its formulation. Section 3 describes the methodology, with results (section 4), a discussion and conclusions then following.

## II. KERNEL DENSITY ESTIMATION

The spatial application of kernel density estimation emerged as a popular technique in spatial epidemiology to assist the study of disease patterns (for an early example see Bithell, 1990). Similar to disease, crime incidents are most usually geographically referenced as points. The kernel density estimation function is applied to these points to obtain a smooth surface estimate representing the density of the point distribution. In mathematical terms, KDE is expressed as:

$$f(x,y) = \frac{1}{nh^2}\sum_{i=1}^{n} k\left(\frac{d_i}{h}\right) \qquad (1)$$

Where *f(x,y)* is the density value at location *(x,y)*, *n* is the number of incidents/points, *h* is the bandwidth, $d_i$ is the geographical distance between incident *i* and location *(x, y)* and *k* is a density function, known as the kernel. *k* can take many forms although the results between different functions produce very similar density values (Bailey and Gatrell, 1995). A common choice for *k* is the quartic function (Bailey and Gatrell, 1995; Ratcliffe, 2002; Levine, 2004).

Evaluation of the components of the KDE equation show that the density value for each location is affected by the number of points, their spatial distribution, and the bandwidth. For the purpose of generating a hotspot map of crime for a single study area, using data for a particular retrospective snapshot of previous incidents, the number of crime incidents across the area would remain the same (and hence not influence changes in the density estimate), the spatial distribution of the crime incidents is static, therefore it is the bandwidth that will influence different values of *f* at each *x,y* location. Each *x,y* location is represented spatially as a grid cell (the coordinates referring to the centroid of that cell), with the calculated density value *f* attributed to each cell. The cell size chosen by the researcher can vary, resulting in many calculations of *f* if the cell size is small or much fewer calculations if the cell size is large. Whilst cell size is not an input to the KDE equation, the representation of these density values for areas of different size will be subject to the Modifiable Areal Unit Problem (Openshaw, 1984) – different size cells may produce different results of the spatial KDE distribution of crime.

There is currently very little guidance on the choice of cell size a researcher should select and no research that we are aware of that investigates comprehensively the impact it can have on a crime hotspots central aim – to accurately identify areas where there have been high concentrations of crime, using the hotspot mapping output to determine where policing resources should then be targeted. The little guidance that is offered is by Chainey and Ratcliffe (2005) who recommend that a suitable KDE cell size to choose for crime hotspot mapping is to divide the shorter side of the study area's minimum bounding rectangle (MBR) by 150. Whilst simple to calculate and used to determine the default cell size in the Hotspot Detective MapInfo add-on (Ratcliffe, 2002), this approach has not been rigorously evaluated.

The choice of bandwidth size value for crime researchers to select is similarly uninformed. Whilst there are several bandwidth size optimisation routines such as the Mean Integrated Square Error (Fotheringham et al., 2000; Bowman and Azzelini, 1997), Akaike Correlation Coefficient and the Cross Validation method (Silverman, 1986; Brunsdon, 1995; Fotheringham et al., 2002), these tend to produce large bandwidths and are considered unsuitable for the purposes of exploring local spatial patterns of the density distribution of crime (Uhlig, 2005). Bailey and Gatrell suggest a value derived from calculating $h = 0.68n^{-0.2}$ as a *rough choice* (Bailey and Gatrell, 1995, 86) for the bandwidth (where $n$ is the number of observed events across the study area), but again experimentation of this approach tends to produce bandwidth values that are much larger than those used by crime researchers in practice (Uhlig, 2005). Chainey (2011) recommends a good starting bandwidth is to measure the shorter side of the study area's MBR, divide by 150, and multiply this value by 5. Whilst simple to calculate, the choice of this bandwidth size has not been evaluated, but is common applied - Hotspot Detective for MapInfo uses a very similar procedure for calculating bandwidth default values (Ratcliffe, 2002). Many others suggest an approach of experimenting with different sizes of bandwidth (Bailey and Gatrell, 1995; Eck et al., 2005; Chainey and Ratcliffe, 2005). Whilst this encourages the researcher to explore their data under different bandwidth conditions it often leaves the researcher choosing the mapping output that *looks the best* (Chainey and Ratcliffe, 2005, 159), rather than being more scientifically informed on the influence that bandwidth size selection may have on the hotspot map's central purpose – to accurately assist the targeting of police interventions by helping determine where crime is likely to occur in the future.

## III. METHODOLOGY

Kernel density hotspot maps were created using MapInfo Professional version 10.5 and the MapInfo add-on programme Hotspot Detective (Ratcliffe, 2002). The study area chosen was the district of Newcastle-upon-Tyne in North East England (Figure 2). Newcastle is one of England's largest ten cities and therefore includes many of the urban geographical features and amenities that one would expect in a typical city. This includes a vibrant shopping and entertainment area in the centre of the city, a large number of economic and commerce functions, a mainline train station, a metro system, and two large universities. The district also includes rural areas towards the north. The district population was 292,000 at the time of the 2011 Census of England and Wales.

Geocoded crime point data was provided by Northumbria Police for a one year period (1st October 2009 to 30th September 2010). Burglary to a residential dwelling and violence with injury were the two subsets of data that were chosen for analysis. Two types of crime

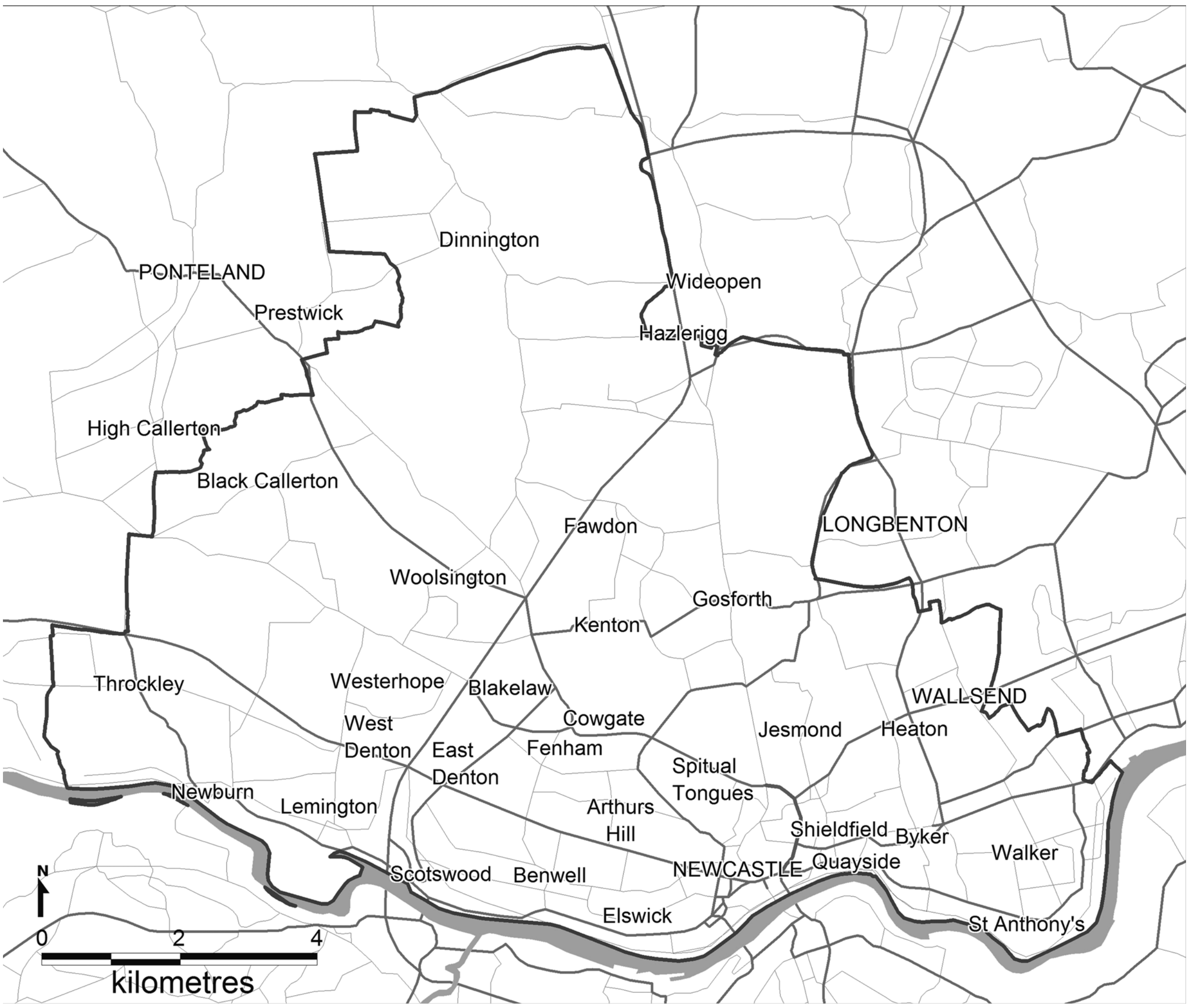



**Figure 2.** Newcastle-upon-Tyne study area

were selected to explore consistencies between the results. Previous research into the spatial prediction of hotspot maps has used residential burglary (Chainey et al., 2008a), but to date no study has used violent crime data. The analysis of residential burglary would therefore enable comparisons with previous research, with the analysis of violent crime offering a new perspective of spatial prediction patterns using KDE hotspot maps. These crime types were also chosen because they are groupings that are regularly analysed by police and crime reduction practitioners - therefore the implications of the research would be of practical interest. Table 1 lists the number of crimes in each sample crime type dataset. The two sets of geocoded crime data were validated using a methodology for geocoding accuracy analysis as reported in Chainey and Ratcliffe (2005, 61-63). This revealed the crime data to be more than 95% accurate to the street address level and fit for purpose for this research.

| **Crime type** | **Number of incidents** (1st October 2009 to 30th September 2010) |
|---|---|
| Residential burglary | 1304 |
| Assaults with injury | 1838 |

**Table 1.** The number of crime incidents for each crime type in Newcastle-upon-Tyne

In following the methodology used by Chainey et al. (2008a), a suitable date had to be chosen within the data time period as the day on which retrospective data were selected to generate hotspot maps against which *future* events could be compared. For simplicity, the 1st April 2010 was selected in order to maximise the use of 6 months of retrospective data for generating KDE hotspot maps, and to use the complete set of 6 months of data after this date for measuring the hotspot maps' abilities for predicting future events. In their analysis that compared two different measurement dates (1st January and 13th March), Chainey et al. (2008a) found no difference in their results. We were therefore confident that the selection of the 1st April 2010 would offer a measurement date that generated representative results.

The retrospective time data was sliced into six time periods and used as input data to generate KDE hotspot maps. This meant that rather than using just one retrospective time period (e.g. the three months prior to the measurement date) which may generate an anomalous result, the use of a number of retrospective time periods would form a more reliable basis on which to draw conclusions. Retrospective input data was sliced into the time periods shown in Table 2a, for each crime type. This concept of using different slices of data as the input data was also followed through to the analysis against measurement data. Six time periods of measurement data were used. This meant that rather than using just one measurement data period for the research (e.g. the three months after the measurement date), the use of a number of measurement data time periods would generate results from which more reliable conclusions could be made. Measurement data was sliced into the time periods shown in Table 2b. This meant that KDE hotspot maps that were generated for each period of input data would be measured for their ability to predict spatial patterns of crime, when the prediction period was the next month, the next two months, and to the next six months.

In their study that compared common hotspot techniques, Chainey et al. (2008a) introduced the Prediction Accuracy Index (PAI). The index was devised as a simple method to allow comparisons between different types of hotspot maps. The index considers the hit rate value (the proportion of crime that occurs within the areas where crimes were predicted to occur i.e. the hotspots) against the size of the areas where crimes were predicted to occur (i.e. the areas determined as hotspots), relative to the size of the study area. The PAI is calculated by dividing the hit rate percentage by the

| **Time periods of data used to create KDE hotspot maps** | | | | | |
|---|---|---|---|---|---|
| 1 month | 2 months | 3 months | 4 months | 5 months | 6 months |
| 01 March 2010 - 31 March 2010 | 01 February 2010 - 31 March 2010 | 01 January 2010 - 31 March 2010 | 01 December 2009 - 31 March 2010 | 01 November 2009 - 31 March 2010 | 01 October 2009 - 31 March 2010 |

a

| **Time periods of data used to measure the spatial prediction abilities of KDE hotspot maps** | | | | | |
|---|---|---|---|---|---|
| 1 month | 2 months | 3 months | 4 months | 5 months | 6 months |
| 01 April 2010 - 30 April 2010 | 01 April 2010 - 30 May 2010 | 01 April 2010 – 31 June 2010 | 01 April 2010 - 31 July 2010 | 01 April 2010 - 31 August 2010 | 01 April 2010 - 30 September 2010 |

b

**Table 2.** (a) The temporal slices of input data for generating hotspot maps, for a measurement date of the 1st April 2010 and (b) the temporal slices of measurement data for calculating the ability of KDE hotspot maps to predict spatial patterns of crime

area percentage (the area of the hotspots in relation to the whole study area (see Equation 2)).

$$\frac{\left(\frac{n}{N}\right) \times 100}{\left(\frac{a}{A}\right) \times 100} = \frac{Hit\ Rate}{Area\ Percentage} = Prediction\ Accuracy\ Index \quad (2)$$

*n:* number of crimes in areas where crimes are predicted to occur (e.g. hotspots)
*N:* number of crimes in study area
*a:* area (e.g. km$^2$) of areas where crimes are predicted to occur (e.g. area of hotspots)
*A:* area (e.g. km$^2$) of study area

For example, if 25% of future crime events took place in 50% of the study area, the PAI value would equal 0.5; if 20% of future crime events took place in 10% of the area, the PAI would equal 2. Therefore, the higher the PAI, the better the hotspot map for predicting spatial patterns of crime.

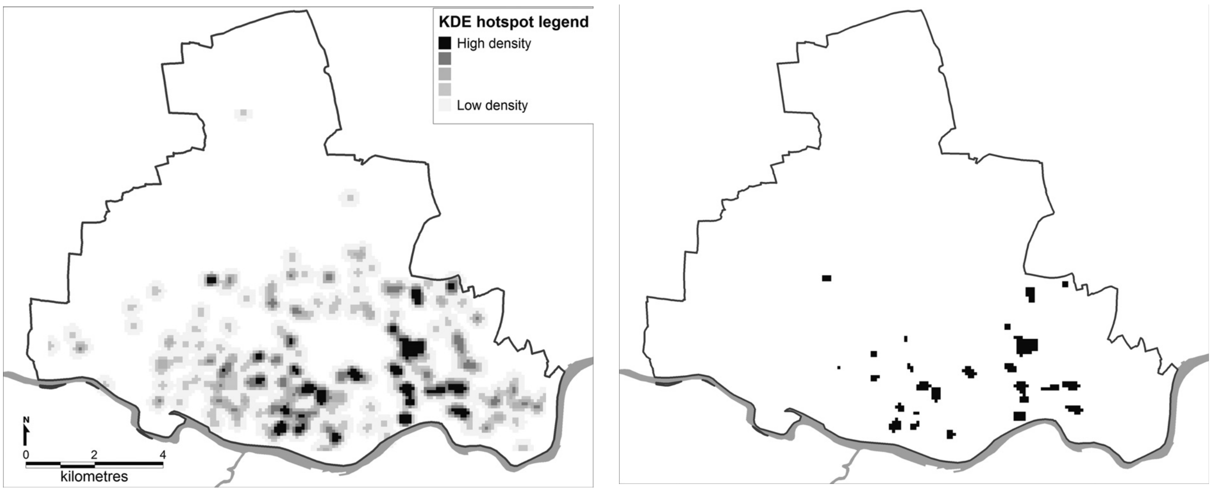
Since the PAI was introduced, other approaches for measuring the predictive abilities of mapping output have been developed. Perhaps the most rigorous of these is proposed by Johnson et al. (2009). The problem with a single measure such as the PAI is that it only offers a comparison between one hit rate and one defined hotspot area, and no comparison against chance expectation. Johnson et al. (2009) proposed the use of an accuracy concentration curve. This is generated by plotting the percentage of crimes that have been accurately predicted (i.e. the hit rate) against the incremental risk ordered percentage of the study area i.e. comparing the number of *future* crimes in 1% of the study area, with this 1% area containing the highest KDE values; comparing the number of *future* crimes in the areas containing the top 2% of KDE values in the study area; comparing the number of *future* crimes in the areas containing the top 3% of KDE values in the study area ..., to comparing the number of *future* crimes in the areas containing 100% of the study area.

However, in Johnson et al.' study (2009) they only compared results between mapping techniques for one input data period (two months) and one output data period (seven days). Calculating an accuracy concentration curve and comparing it against a Monte Carlo simulated result (produced after running at least 19 simulations in order to use a 0.05 level of significance) is practical for comparing one set of data input and output for two different techniques (i.e. two experiments). In our study that uses six different input datasets, and six different output datasets for eight different bandwidth settings and eight different cell size settings (more details on bandwidth and cell size settings are described below), for two types of crime (therefore involving 1152 experiments), and generating 19 Monte Carlo simulations for each experiment, this approach is not practical nor proportionate to the aims of this research – to explore differences between cell size value and bandwidth value, for the same study area, using the same hotspot mapping technique. The use of the PAI has since been discussed further by Pezzuchi (2008), Levine (2008) and Chainey et al. (2008b; 2008c), with researchers concluding it to be a useful measure for comparing multiple hotspot mapping outputs. This has included minimising chance expectation by using the mean PAI results and observing the variation in the standard deviation generated from the many experiments.

Eight cell size values were chosen for comparison: 30 m, 60 m, 90 m, 120 m, 150 m, 180 m, 210 m and 240 m. A value that is often used for the cell size (as referred to in section 2) is the result from measuring the shortest side of the minimum bounding rectangle of the study area, and dividing this distance by 150. Although the choice of 150 is rather arbitrary, in practice it provides a useful starting measure and is the procedure that is used to calculate cell size in the popular Hotspot Detective for MapInfo software (Ratcliffe, 2002). This gave the value of 89.6 (rounded upto 90 m). We therefore felt it useful to generate results for this measure in comparison to other cell size values, using multiples of 30 m in our cell size experiments. For each cell size experiment, the bandwidth was controlled to a single size: a bandwidth of 450 m was used, as per the guidance described in section 2.

Eight bandwidth size values were chosen for comparison: 100 m, 200 m, 300 m, 400 m, 500 m, 600 m, 700 m and 800 m. If we had followed the recommendations of Chainey (2011) (i.e. five times the cell size) this would have suggested a bandwidth value of 450 m. Rather than use multiples of 150 m, we decided to use multiples of 100 m in order to explore the influence of a small bandwidth (100 m), to help more simply present results, but still enable a comparison between the outputs generated between 400 m and 500 m as an indication of the effectiveness of this rather crude approach for determining bandwidth size. For each bandwidth size experiment, the cell size was controlled to a single size: a cell size of 90 m was used, as per the guidance described in section 2.

To identify if predicted spatial patterns of crime generated by KDE hotspot maps under different cell size and bandwidth settings differed, Prediction Accuracy Index measures were aggregated and averaged for the periods of input data and for the periods of measurement data. This meant that the PAI measures could be compared, with any differences being explained in relation to the cell size and bandwidth size rather than different periods of input and measurement data. This approach was applied separately to the two crime datasets: residential burglary and assault with injury. The standard deviation and coefficient of variation of the PAI for




(a) KDE hotspot map (b) Top thematic class of KDE hotspot map

**Figure 3.** Hotspots were determined by selecting the top thematic class calculated using five classes and the default values generated from applying the quantile thematic range method in MapInfo

each crime type across the eight different cell size values and eight bandwidth values were also calculated.

During the data time period (1st October 2009 to 30th September 2010) there could have been police operations and crime reduction initiatives that had an impact on crime levels, plus there could have been an impact from seasonal influences. For this study, because the focus was on comparing KDE parameter settings against the same data, any changes in crime patterns would have the same impact on different cell size and bandwidth size parameter entries and would not affect the ability to examine results and draw conclusions on the analyses.

A final parameter to consider for KDE hotspot map generation is a threshold value for determining which areas are *hot*. For purposes of research comparison, we followed the methodology used by Chainey et al. (2008a). This involved using five thematic classes and default values generated from using the quantile thematic classification method in MapInfo. *Hot* was then determined by the top thematic class (Figure 3).

## IV. RESULTS

### A. The influence of cell size on KDE hotspot maps for predicting where crime may occur

Table 3 shows the PAI results for residential burglary and assaults with injury for different cell sizes. The PAI results for residential burglary varied between 6.6 for a cell size of 240 m to 7.1 for 30 m and 60 m cell sizes. The PAI results for assaults with injury were much higher than those for residential burglary, but again showed only a small amount of relative variation from 59.9 for a cell size of 210 m to 68.5 for a cell size of 60 m. These results suggest that although PAI values decrease with increases in cell size, this difference is marginal. These results are also shown in Figure 4.

| | **Residential burglary** | | | **Assaults with injury** | | |
|---|---|---|---|---|---|---|
| **Cell size (m)** | **PAI** | **SD** | **CV** | **PAI** | **SD** | **CV** |
| **30** | 7.1 | 0.60 | 0.08 | 68.4 | 3.06 | 0.04 |
| **60** | 7.1 | 0.66 | 0.09 | 68.5 | 2.96 | 0.04 |
| **90** | 6.7 | 0.53 | 0.08 | 65.1 | 2.66 | 0.04 |
| **120** | 6.9 | 0.57 | 0.08 | 64.5 | 2.50 | 0.04 |
| **150** | 6.7 | 0.53 | 0.08 | 63.0 | 3.17 | 0.05 |
| **180** | 6.8 | 0.64 | 0.10 | 64.3 | 2.95 | 0.05 |
| **210** | 6.7 | 0.46 | 0.07 | 59.9 | 2.39 | 0.04 |
| **240** | 6.6 | 0.48 | 0.07 | 60.2 | 2.85 | 0.05 |

**Table 3.** KDE hotspot map PAI, standard deviation (SD) and coefficient of variation (CV) results for residential burglary and assaults with injury for different cell sizes

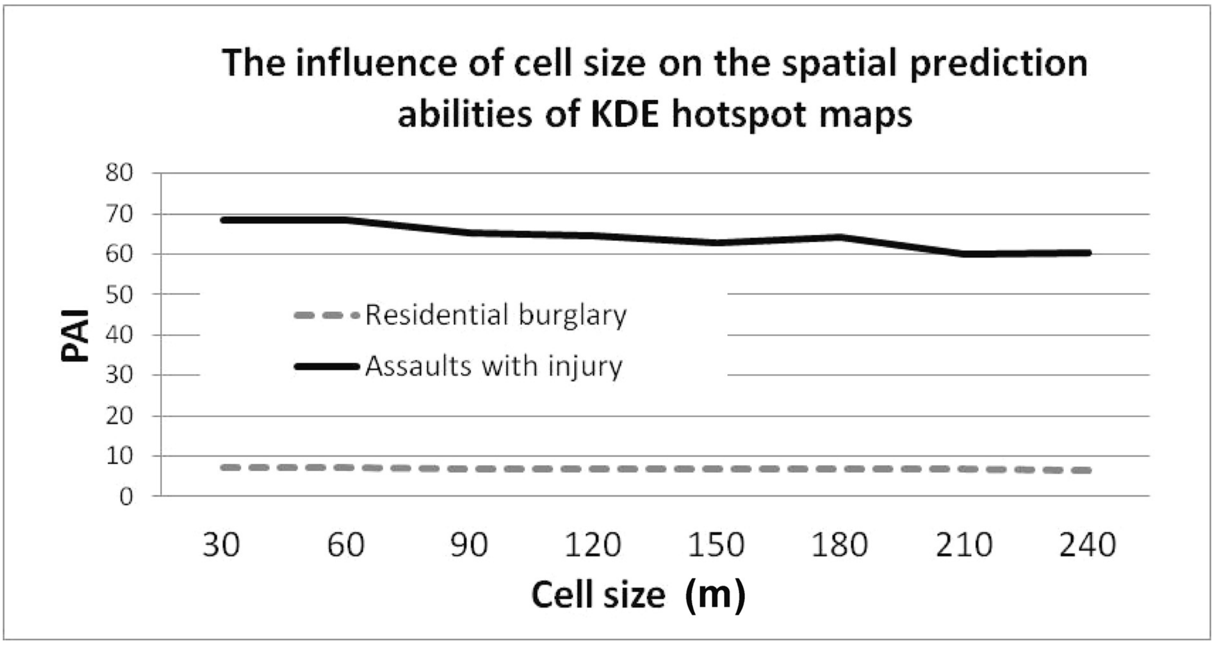



**Figure 4.** The influence of cell size on KDE hotspot map PAI values for residential burglary and assaults with injury

There was little statistical variation in the results for each cell size, as indicated by the low coefficient of variation (CV) values, and little difference in the CV values between cell sizes.

The similarity in results for different cell sizes is further illustrated by the difference in the number of crimes that maps of different cell sizes predict in KDE generated hotspot areas (Table 4). When the KDE hotspot areas were controlled to identify 1% of the total study area (i.e. the 1% of areas with the highest KDE values), generated from 3 months of input data using cell sizes of 30 m and 240 m to predict where crimes would occur in the next 3 months, very similar results were produced: for residential burglary, KDE outputs generated using a 30 m cell size predicted 29 crimes, in comparison to 28 crimes using a cell size of 240 m; for assaults with injury, KDE outputs generated using a 30 m cell size predicted 158 crimes, in comparison to 153 crimes using a cell size of 240 m. That is, as the spatial resolution of the KDE hotspot map begins to degrade, the ability of the map to predict where crime occurs in the future reduces only slightly.

**A. The influence of bandwidth size on KDE hotspot maps for predicting where crime may occur**

Table 5 shows the PAI results for residential burglary and assaults with injury for different bandwidth sizes. The PAI results for residential burglary varied between 5.6 for bandwidth sizes of 700 m to 13.1 for 100 m bandwidth sizes. The PAI results for assaults with injury were much higher than those for residential burglary, but also showed large variation from 42.9 for bandwidth sizes of 800 m to 142.8 for bandwidth sizes of 100 m. These results suggest that as bandwidth size increases, the power of the KDE hotspot map to pre-

| Crime type and cell size (m) | Crimes committed April – June 2010 | Number of crimes in hotspots (1% of area) | Percentage of crimes in hotspots |
|---|---|---|---|
| **Residential burglary: 30 m** | 329 | 29 | 8.8% |
| **Residential burglary: 240 m** | 329 | 28 | 8.5% |
| **Assaults with injury: 30 m** | 459 | 158 | 34.3% |
| **Assaults with injury: 240 m** | 459 | 153 | 33.3% |

**Table 4.** Crimes predicted using kernel density estimation outputs of difference cell sizes for residential burglary and assaults with injury, based on using three months of input crime data (January – March 2010) and 3 months of measurement data (April – June 2010). The area determined as *hot* was controlled to cover 1% of the study area's total area.

| | Residential burglary | | | Assaults with injury | | |
|---|---|---|---|---|---|---|
| **Bandwidth size (m)** | **PAI** | **SD** | **CV** | **PAI** | **SD** | **CV** |
| **100** | 13.1 | 2.8 | 0.22 | 142.8 | 11.53 | 0.08 |
| **200** | 11.1 | 1.3 | 0.12 | 91.7 | 4.65 | 0.05 |
| **300** | 8.7 | 1.0 | 0.12 | 79.4 | 3.03 | 0.04 |
| **400** | 7.1 | 0.7 | 0.10 | 68.3 | 3.54 | 0.05 |
| **500** | 6.5 | 0.6 | 0.09 | 60.2 | 2.60 | 0.04 |
| **600** | 5.9 | 0.6 | 0.11 | 54.3 | 2.52 | 0.05 |
| **700** | 5.6 | 0.6 | 0.11 | 48.6 | 2.23 | 0.05 |
| **800** | 5.7 | 0.5 | 0.09 | 42.9 | 1.98 | 0.05 |

**Table 5.** KDE hotspot map PAI, standard deviation (SD) and coefficient of variation (CV) values for residential burglary and assaults with injury for different bandwidth sizes

dict spatial patterns of crime degrades. These results are also shown in Figure 5. With the exception of residential burglary KDE hotspot maps generated using a bandwidth of 100 m, there was little statistical variation in the results for each bandwidth size and little difference in the CV values between cell sizes.

The difference in results for different bandwidth sizes is further illustrated by the difference in the number of crimes that maps of different bandwidth sizes predict in hotspots generated using KDE (Table 6). To illustrate this (and to allow for easier comparisons with future research) we controlled the KDE hotspot areas to identify only the top 1% of density values (i.e. the 1% of areas with the highest KDE values), generated from 3 months of input data using bandwidth sizes of 100 m and 800 m to predict where crimes would occur in the next 3 months. For residential burglary, KDE outputs generated using a 100 m bandwidth size predicted 35 crimes (i.e. 11% of all burglaries in just 1% of the study area), in comparison to 22 crimes using a bandwidth size of 800 m; for assaults with injury, KDE outputs generated using a 100 m bandwidth size predicted 166 crimes (i.e. 36% of all violent assaults in 1% of the study area), in comparison to 137 crimes using a bandwidth size of 800 m. That is, as the smoothing of the KDE hotspot map increases, the ability of the map to predict where crime occurs degrades. These results also illustrate the

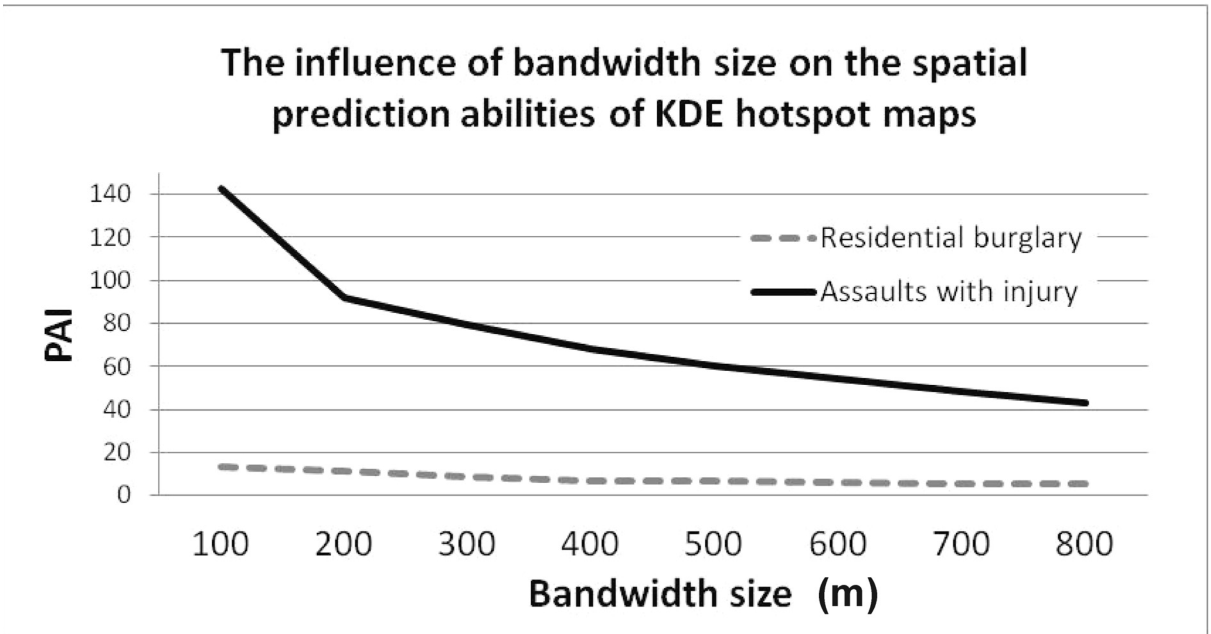



**Figure 5.** The influence of bandwidth (m) size on KDE hotspot map PAI values for residential burglary and assaults with injury.

| Crime type and bandwidth size (m) | Crimes committed April – June 2010 | Number of crimes in hotspots (1% of area) | Percentage of crimes in hotspots |
|---|---|---|---|
| **Residential burglary: 100 m** | 329 | 35 | 10.6% |
| **Residential burglary: 800 m** | 329 | 22 | 6.7% |
| **Assaults with injury: 100 m** | 459 | 166 | 36.2% |
| **Assaults with injury: 800 m** | 459 | 137 | 29.8% |

**Table 6.** Crimes predicted using kernel density estimation outputs of difference bandwidth sizes for residential burglary and assaults with injury, based on using three months of input crime data (January – March 2010) and 3 months of measurement data (April – June 2010). The area determined as *hot* was controlled to cover 1% of the study area's total area.

proportion of crime that KDE hotspot maps can predict.

## V. DISCUSSION AND IMPLICATIONS

The findings from this research show that KDE hotspot maps generated using different cell sizes have little impact on the mapping outputs ability to predict spatial patterns of crime, but that different bandwidth sizes do have an impact. Cell size mainly impacts on the visual appeal of the KDE mapping output, with higher resolutions producing maps that avoid the *blocky* pixilation of outputs generated using larger cell sizes. For example, the maps shown in Figure 6 are equally as good as each other for predicting where crime may occur in the future, but Figure 6a is the more preferable output due to its better visual appeal. While smaller cell sizes require greater computer processing due to the larger number of calculations that are required, in our experiments this extra length of processing was not a significant impairment.

Bandwidth size does though affect the ability of KDE hotspot maps to predict spatial patterns of crime. For example, the maps shown in Figure 7 were generated using the same period of input data but have very different PAI values. That is, the smaller the bandwidth, the better the KDE map is at predicting spatial patterns of crime.

The research has also shown the large variation that exists between the ability to predict different types of crime using KDE. This was initially shown by Chainey et al. (2008a), with street robbery KDE maps generating higher PAI values than KDE hotspot maps of residential burglary, and vehicle crime. The PAI results for residential burglary in this study of crime in Newcastle-upon-Tyne are higher than those found by Chainey et al. (2008a) for residential burglary in London, indicating differences between areas. However, it is the high PAI values generated for violent assaults that offer new insights into the spatial prediction of KDE hotspot maps. This is reflected by the manner in which violent assaults cluster in comparison to burglary. Whilst burglary does concentrate spatially, these hotspots tend to be larger in number and more dispersed. This is most likely due to the wider (in spatial terms) opportunity for burglary, with residential properties spread geographi-

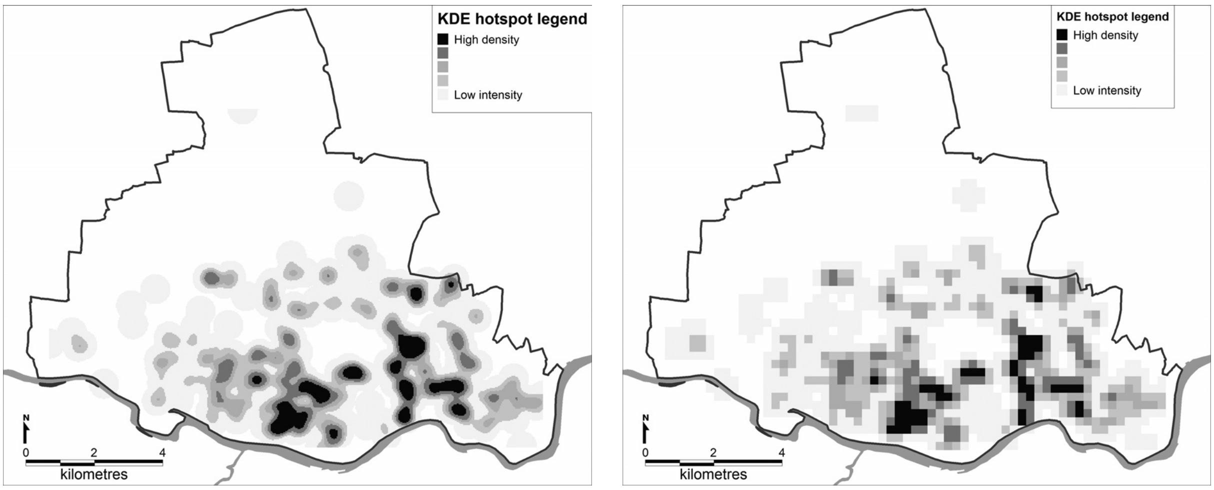



**Figure 6.** A comparison of KDE hotspot maps generated using the same bandwidth but with different cell sizes (a) 30 metres (PAI of 7.0) and (b) 240 metres (PAI of 7.0)

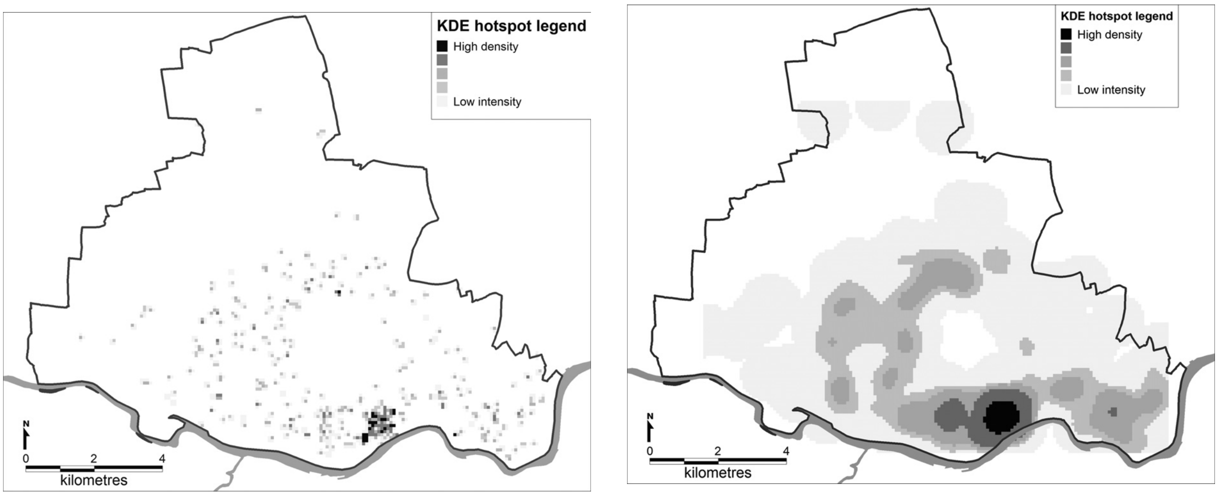
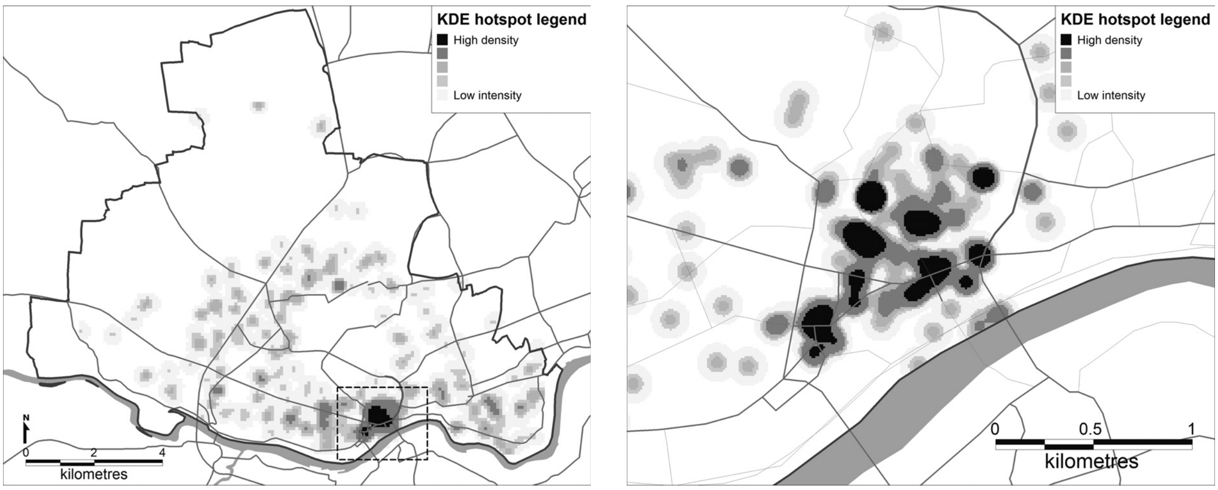



**Figure 7.** A comparison of KDE hotspot maps generated using the same cell size but with different bandwidth sizes (a) 100 metres (PAI of 119.3) and (b) 800 metres (PAI of 40.4)

cally across areas. Areas where violent assaults take place tend to be highly concentrated in areas that are associated with alcohol and the night-time economy (Maguire and Hopkins, 2003; Babor et al., 2003; Graham and Homel, 2008). Newcastle's night-time economy is heavily concentrated in the city centre, therefore the occasional violent interaction between people in this highly compact area heavily influences the spatial distribution of this type of crime. That is, the highly compact nature of the night-time economy has a direct impact on the highly compact, and predictable nature of where violent assaults are most concentrated.

The analysis of different cell sizes and bandwidths also offers practitioners the means to better qualify the default parameter values that are determined by Geographical Information System products such as ESRI's ArcGIS Spatial Analyst, Crime Analyst, and Hotspot Detective for MapInfo. Our results indicate that defaults for cell size such as those generated using Hotspot Detective (which involves dividing the shorter side of the MBR by 150) offer a useful starting point, but that reducing this value further will generate maps of greater visual appeal without affecting the maps ability to predict where crime is likely to occur in the future. However, bandwidth default values need further scrutiny by practitioners to ensure they are not too large and impair the purpose of the KDE hotspot mapping output. For example, the default Hotspot Detective KDE bandwidth size for three months of violent assaults data for Newcastle-upon-Tyne was 450 m – a bandwidth size that generated a PAI value of 60 compared to a PAI of 143 if a bandwidth of 100 m was used.

However, low bandwidth values produce KDE hotspot




(a) (b)

**Figure 8.** A procedure for creating precise and practical KDE hotspot maps for accurately assisting in the targeting of policing and crime reduction resources: (a) is a KDE hotspot map generated for a large area for identifying the key strategic areas for focus (bandwidth 300 m; cell size 90 m). Once a focus area is identified data for this area is selected, and a KDE hotspot map is generated using a smaller bandwidth (100 m) and cell size (10m).

maps that appear *spikey*, with many small areas identified as hotspots. In practice, this type of hotspot map is often considered unsuitable because it does not identify a small number of areas for strategic attention. Therefore, it is argued that a balance is required between KDE hotspot prediction accuracy, and output that is useful in practice. A way in which this can be overcome is to use a bandwidth size that is large enough to initially identify key hotspot areas for strategic attention, with these areas then being focused upon in more detail with a second hotspot map generated based on the distribution of crime in this focus area. Figure 8 shows an example of this – Figure 8a uses a bandwidth size of 300 m and cell size of 90 m to identify the main assaults hotspots in Newcastle-upon-Tyne. The main hotspot then becomes the area of attention, with a second KDE hotspot map generated for this area to more precisely identify the areas that are required for police attention. Figure 8b was generated using a bandwidth of 100 m and a cell size of 10m.

KDE is though not without its weaknesses. The procedure described above would fail to identify areas where there is a high and compact concentration of crime because larger bandwidths have the tendency to smooth these out over the area it generates density values for. An additional weakness is that the use of KDE requires the researcher to determine what is *hot* by deciding the value for the top thematic class. In this research we standardised this procedure by using the quantile thematic classification method in all experiments. However, most GIS software offer several options for the user to determine a thematic classification method preference, leading to subjectivity in hotspot mapping output. This calls for further research that identifies hotspot mapping methods that can overcome these KDE weaknesses.

## VI. CONCLUSION

Hotspot analysis is a basic form of crime prediction – using crime data from the past to predict where crime may occur in the future, with the outputs from hotspot mapping being used in practice for determining where police patrols and other crime prevention initiatives should be targeted. Kernel density estimation has become the most popular technique used in practice for identifying hotspots of crime.

Cell size and bandwidth size are the two main parameters that the user is required to enter in order to generate KDE hotspot mapping output. The findings from this research illustrate that cell size has little impact on a KDE hotspot map's ability to predict spatial patterns of crime, but that smaller sizes generate hotspot maps of greater visual appeal. Bandwidth size does though have an impact on a KDE hotspot output's ability to predict spatial patterns of crime, with the spatial prediction ability of the KDE hotspot map degrading as bandwidth size increases. To date, most users of KDE for hotspot mapping make use of default settings for cell size and bandwidth size, without qualifying these values. This research has helped to identify the influence these parameters have, and in so doing offer practitioners and researchers a more informed basis on which to qualify the values they should use for cell size and bandwidths for producing KDE hotspot maps.

## REFERENCES

Babor, T., Caetano, R., Casswell, S., Edwards, G., Giesbrecht, N., Graham, K., Grube, J., Gruenewald, P., Hill, L., Holder, H., Homel, R., Osterberg, E., Rehm, J., Room, R., Rossow, I. (2003). *Alcohol : No ordinary commodity. Research and public policy*. Oxford: Oxford University Press.

Bailey, T.C., and Gatrell, A.C. (1995) *Interactive Spatial Data Analysis*. Reading, Massachusetts: Addison-Wesley.

Beato, C. (2008). *Comprehendo e Avaliando: Projetos de Segurança Publica*. Belo Horizonte: Editora UFMG.

Bithell, J.F. (1990). An application of density estimation to geographical epidemiology. *Statistics in Medicine,* 9, 691-701.

Block, R. and Block, R.B. (2000) The Bronx and Chicago – Street Robbery and the Environs of Rapid Transit Stations. In Goldsmith, V., McGuire, P.G., Mollenkopf, J.H. and Ross, T.A. (Eds.) *Analysing Crime Patterns: Frontiers and Practice* (p. 137-152). Thousand Oaks (CA): Sage.

Bowman, A. and Azzelini, A. (1997). *Applied Smoothing Techniques for Data Analysis*. Oxford: Oxford University Press.

Braga, A. (2007). *Effects of Hot Spots Policing on Crime: A Campbell Collaboration Systematic Review.* http://www.aic.gov.au/campbellcj/reviews/titles.html

Brunsdon, C. (1995). Analysis of Univariate Census Data. In S. Openshaw (ed.) *Census Users Handbook* (p. 213-238). Cambridge: Geoinformation International.

Chainey, S.P. (2011). Identifying hotspots: an assessment of common techniques. Presentation at the International Crime and Intelligence Analysis Conference November 2011, Manchester, England.

Chainey, S.P. and Ratcliffe, J.H. (2005) *GIS and Crime Mapping*. London: Wiley.

Chainey,S.P., Tompson,L., Uhlig,S. (2008a). The utility of hotspot mapping for predicting spatial patterns of crime. *Security Journal*. 21, 1-2.

Chainey,S.P., Tompson,L., Uhlig,S. (2008b). Response to Pezzuchi. Security Journal, 21:4.

Chainey,S.P., Tompson,L., Uhlig,S. (2008c). Response to Levine. Security Journal, 21:4.

Eck, J.E., Chainey, S.P., Cameron, J.G., Leitner, M. and Wilson, R.E. (2005) *Mapping Crime: Understanding Hot Spots*. USA: National Institute of Justice.

Fotheringham, A. S., Brunsdon, C. and Charlton, M. (2000). *Quantitative Geography: Perspectives on Spatial Data Analysis.* London: Sage.

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

Graham, K. and Homel, R. (2008). *Raising the bar: preventing aggression in an around bars, clubs and pubs*. Cullompton: Willan Publishing.

Home Office (2005) *Crime Mapping: Improving Performance, A Good Practice Guide for Front Line Officers*. London: Home Office.

Johnson, S.D., Bowers, K.J., Birks, D.J., and Pease, K. (2009). Predictive mapping of crime by ProMap: accuracy, units of analysis, and the environmental backcloth. In D. Weisburd, W. Bernasco and G.J.N. Bruinsma eds., *Putting Crime in its Place* (p. 165–192). Dordrecht: Springer.

LeBeau, J.L. (2001) Mapping Out Hazardous Space for Police Work. In Bowers, K. and Hirschfield, A. *Mapping and Analysing Crime Data – Lessons from Research and Practice*. London: Taylor and Francis.

Levine, N. (2004) *CrimeStat III: A Spatial Statistics Program for the Analysis of Crime Incident Locations*. Houston (TX): Ned Levine and Associates. Washington (DC): National Institute of Justice.

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, 4.

Maguire, M. and Hopkins, M. (2003). Data analysis for problem-solving: alcohol and city centre violence. In Bullock, K. and Tilley, N. (eds) *Crime Reduction and Problem-Oriented Policing* (p. 24-38). Cullompton: Willan Publishing.

Mashford, T. (2008). Methods for implementing crime mapping within a large law enforcement agency; experiences from Victoria, Australia. In S.P. Chainey and L. Tompson, eds., Crime Mapping Case Studies: Practice and Research (p. 19-26). London: Wiley.

National Policing Improvement Agency (2010). The analysis of geographic information - workbook. Wyboston: NPIA.

Openshaw, S. (1984) *The Modifiable Areal Unit Problem*. Concepts and Techniques in Modern Geography 38. Norwich: Geobooks

Pezzuchi, G. (2008). A brief commentary on The Utility of Hotspot Mapping for Predicting Spatial Patterns of Crime. *Security Journal*, 21,4.

Ratcliffe, J., (2002) *HotSpot Detective 2.0 for MapInfo Professional 7.0*. Philadelphia: JHR Systems.

Ratcliffe, J. and McCullagh, M. (2001) Crime, Repeat Victimisation and GIS. In Bowers, K. and Hirschfield, A. *Mapping and Analysing Crime Data – Lessons from Research and Practice* (p. 61-92). London: Taylor and Francis.

Ratcliffe, JH, Taniguchi, T, Groff, E and Wood, J (2011) The Philadelphia Foot Patrol Experiment: A randomized controlled trial of police patrol effectiveness in violent crime hotspots. *Criminology*. 49(3): 795-831.

Sherman, L. (2009). Hot spots. In Wakefield, A. and Fleming, J. *The SAGE Dictionary of Policing*. London: Sage Publications.

Silverman, B.W. (1986). *Density Estimation for Statistics and Data Analysis*. London: Chapman and Hall.

Uhlig, S. (2005). *Examining the Prediction Accuracy of Crime Hotspots* (Unpublished Master Thesis in GIS, University College London).

*Coordonnées de l'auteur:*

Spencer CHAINEY
University College London,
Department of Security and Crime Science,
35 Tavistock Square, London, WC1E 9EZ.
England.
s.chainey@ucl.ac.uk