J Geograph Syst (1999) 1:385–398

Journal of Geographical Systems

Hotbeds of crime and the search for spatial accuracy

J.H. Ratcliffe, M.J. McCullagh

School of Geography, University of Nottingham, University Park Nottingham NG7 2RD, UK
(e-mail: michael.mccullagh@nottingham.ac.uk)

Received: 9 October 1998 / Accepted: 22 September 1999

Abstract. One of the most important aspects of spatial crime analysis is the identification of hotspots: areas of the highest crime concentration. This paper advances a methodology for hotspot detection based on a global moving window approach combined with the use of local statistics to define the hotspot limit. This technique generates hotspots that both follow the urban morphology of the crime distribution and ensures their spatial segregation. The hypothesis that police officers can construct an accurate perception of crime distribution from exposure to daily policing practices is used to demonstrate an application in the use of hotspot analysis. Significant regions generated from recorded crime data are compared with perceived local hotspots catalogued from surveys with police officers. Results from this study show two discrete types of hotspot, here termed hotpoints and hotbeds. The morphology of these crime hotpoints and hotbeds is discussed and possible causes documented.

Key words: Crime, hotspots, local statistics, police, mapping, GIS

JEL classification: C88, C12, K42, R14

1 Introduction

This research note examines the hypothesis that police officers can construct an accurate perception of crime distribution by comparing crime hotspots generated from the recorded crime data with the perceived local hotspots catalogued from surveys with police officers. A distinction is made in the UK between incidents and recorded crime, where an incident is a request for police assistance but may not involve criminal activity whereas recorded crime may lead to criminal prosecution. Hotspots are aggregations of the raw crime data, designed to identify the sites of highest incident concentration. Like all aggregation techniques they can be susceptible to the Modifiable Areal Unit Problem (MAUP). The MAUP is a potential source of error that can affect spatial studies which utilise aggregate data sources (Bailey and Gatrell 1995). Spatial objects such as enumeration districts or police beat boundaries are examples of the type of aggregating zones used to show results of some spatial phenomena. Regular, often square, grids are also common, though polygons have been used in other studies of crime distribution (Hirschfield et al. 1997). These zones are often arbitrary in nature and different areal units can be just as meaningful in displaying the same base level data, though will do so with different graphical results. It is this variation in acceptable areal solutions that generates the term 'modifiable.' This problem has been addressed in the area of spatial crime analysis, where; 'the areal units (zonal objects) used in many geographical studies are arbitrary, modifiable, and subject to the whims and fancies of whoever is doing, or did, the aggregating.' (Openshaw 1984).

Recognition of the MAUP has resulted in a number of proposed solutions. One of the most influential contributions to the field of spatial pattern analysis in the last twenty years has been the Mark 1 Geographical Analysis Machine (GAM). The GAM was designed to be an automated process by which point data could be automatically searched with various algorithms to detect clusters (Openshaw and Charlton 1987). It also included the population at risk in its analysis. GAM was originally introduced to investigate the number of leukaemia cases around power stations in the north-east of England in the early 1980's, and a number of advances in spatial point analysis have come from the field of medical epidemiology (Besag and Newell, 1991; Gatrell et al. 1996). To overcome the MAUP these techniques employ a 'moving window' approach (Bailey and Gatrell 1995).

The 'moving window' technique applies a moveable sub-region (usually a circle) over the entire study area to measure dependence in subsets of the study area and is particularly suited to crime hotspot detection. A two-dimensional grid lattice that covers the entire study area with a rectangular grid of intersecting lines is defined and at each grid intersection circles are placed over the study area. Points falling within the circle are retrieved from the data to compute a spatial pattern test statistic. The use of a moving window where windows overlap can help defeat much of the MAUP. It is also a local method in a crime analysis environment where global methods are difficult to justify. Given the short distance 'journey-to-crime' limitations of many offenders local methods are more desirable than global methods that attempt to relate crime locations right across the study area and can interpolate values into areas void of criminal activity.

2 Deriving data for the current study

Hotspots were generated from Nottinghamshire Constabulary recorded crime data for the period April 1996 to April 1997. The study area covered three force sub-divisions: West Bridgford, the Meadows, and Clifton. Two crime types were chosen for examination: residential burglary and motor vehicle crime.

A moving window analysis was used in a program that creates a grid lattice over the data and generates a grid size relative to the chosen circle radius, such that the overlap is 0.8r, comparable to that used in the original GAM.

Within each circle generated by the program, some algorithm is necessary to calculate a value for the circle. The simplest solution is to count the number of crimes occurring within the search area. This program uses a quartic kernel estimation (from Bailey and Gatrell 1995):

$$\hat{\lambda}_\tau(s) = \sum_{di \leq \tau} \frac{3}{\pi\tau^2}\left(1 - \frac{d_i^2}{\tau^2}\right)^2 \tag{1}$$

where s represents the centre of the search circle, τ the bandwidth and d is the distance of each point (i) within the bandwidth from the centre of the search area. The calculation of the intensity $\lambda_\tau(s)$ is therefore the summation of the intensity of those values which have a smaller distance from s than d, where the intensity is inversely weighted, with values at the centre weighted by $3/\pi\tau^2$, dropping smoothly to a value of zero at the maximum distance τ. The three-dimensional function scans within the search circle and not only detects points within the search region but measures their influence and calculates their contribution to the intensity of the search relative to their proximity to the centre of the search circle. The closer an event is to the centre of the circle, the greater its contribution to the intensity reading. The output from this process can be viewed as a raster image. Figure 1 is an example showing the intensity of residential burglaries in the Meadows area of Nottingham over the period April 1996 to April 1997. The shape of the crime distribution follows the geography of two large council estates located in the south-west of the sub-division. After discussion with officers from the local station it transpired that areas (a) and (b) are favourite targets for a local family of burglars who live between the two estates. Figure 1 also shows that the moving window approach does create a crime intensity surface which is not restricted to geometric shapes such as standard deviational ellipses [the basis of the often used STAC program (Illinois CJIA 1996)], though choice of aggregation classes will effect the graphical display. The output is a continuous function and does not attempt to provide a cut-off point to indicate the edge of a hotspot.

3 Local indicators of spatial association

Exploratory spatial data analysis (ESDA) techniques are available which can identify the spatial association and autocorrelation in georeferenced data sets. It is now recognised that while global spatial autocorrelation may provide a limited set of association results, local patterns of influence can also be used to describe the spatial independence of data (Getis and Ord 1996). Local statistical methods have been developed which can assess the spatial association of a variable within a specified distance of a single point. This class of statistic has been rapidly growing in acceptance (Unwin 1996), and are referred to by the term LISA – Local Indicators of Spatial Association (Anselin 1995). These types of statistics differ from procedures which are applied globally, such as Moran's I and Geary's c which are tested on the complete study area under examination. LISA statistics are applied locally, and are particularly suited to identifying the existence of local spatial clustering around an individual location, or 'hotspots' (Anselin 1995; Getis and Ord 1996).

Although all LISA statistics assess the local association amongst the data, the G_i and G_i^* family of statistics evaluate association by measuring additive qualities, and in doing so can compare local averages to global averages. In this manner they are ideal for defining hotspots and, more importantly, placing

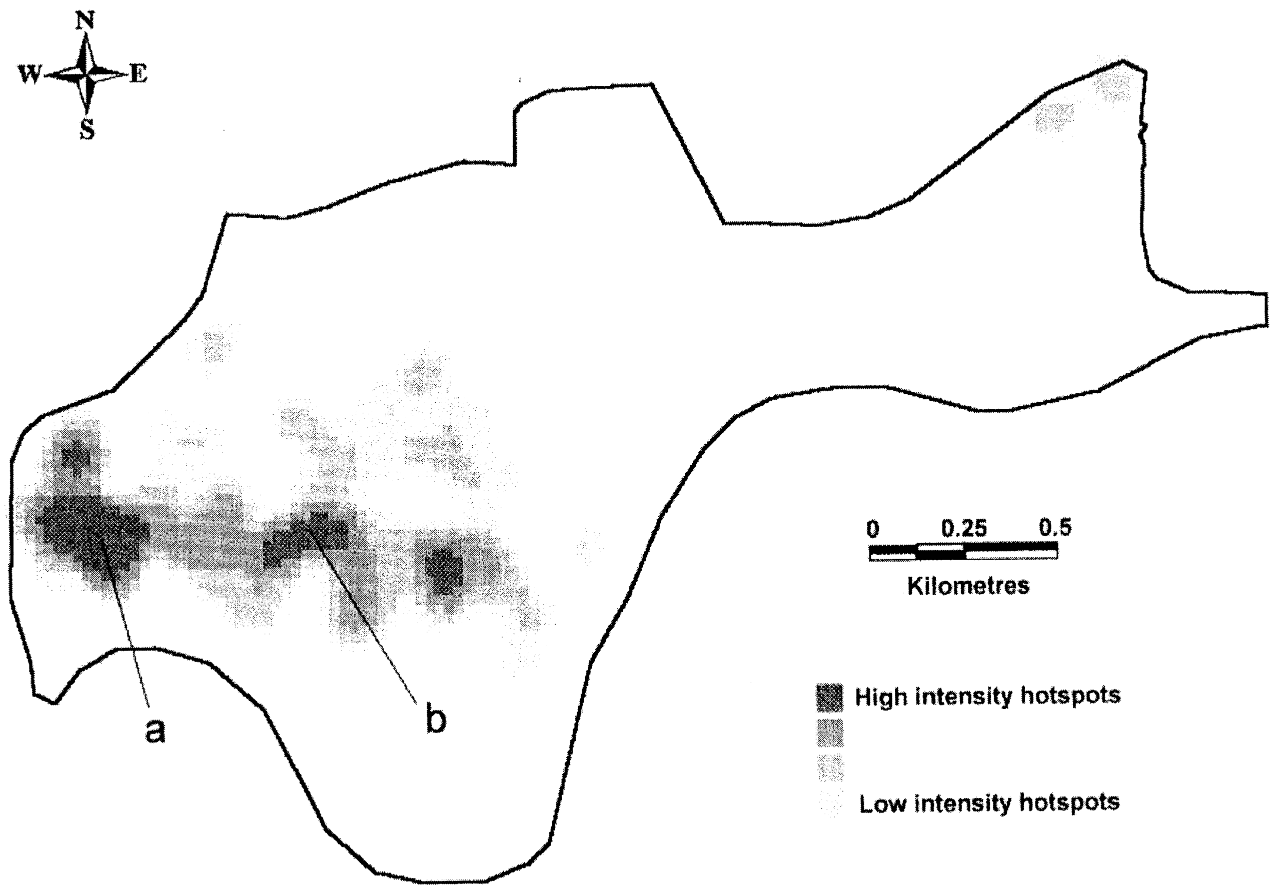

Fig. 1. Raster image of residential burglary intensity for Meadows sub-division. Regions (*a*) and (*b*) indicate two of the highest risk areas

a spatial limit on those hotspots. The statistics G_i and G_i^*, introduced by Getis and Ord (1992) for the study of local patterns in spatial data, were extended and re-written in 1995 to redefine G_i as a standard variate and to allow for non-binary classifications of the distance d (Ord and Getis 1995). The simplified equation from the 1992 paper is given here:

$$G_i(d) = \frac{\sum_j w_{ij}(d)x_j}{\sum_j x_j}, \quad j \neq i \tag{2}$$

where the G_i statistic is calculated for all values of x within d of point i, in a study area containing j values. A weighting w is applied; a binary classification dependent on whether the point is within d of i.

The $G_i^*(d)$ is the same statistic but includes the value at the test location. When the test is performed using both the G_i and G_i^* statistic, there is a problem of overlap. This is caused when selected sites of i are close together, as is usually the case when the test is performed on every location in the study area (as can be easily done with a computer program). It is also a factor when the choices of distance d are large. It is understandable therefore that sites which are located near each other will display similar values of the spatial statistic, as the search areas of each site i will overlap to some degree. This overlap means that nearby values of i will have correlating values of j, and the greater the overlap, the greater the correlation between the spatial statistic. This correlation between local statistics' values means that they lack independence and it is necessary to use a procedure to validate the results. Ord and Getis (1995) suggest a Bonferroni test and produced a table of standard measures for a variety of percentiles for the largest G_i or G_i^* of n values, which are used here to estimate probability.

In this study, the denominator has been modified to include only values greater than zero. The nature of urban crime is that much of a study area can escape the effects of crime due to simple geography. If a database of residential burglary is constructed, for example, parklands and the central business district of a city would see a negligible or zero crime level. It is extremely difficult to isolate residential only areas in these types of studies, and it is more practical to widen the study to include the whole region, including the no crime areas. This does however have the effect of reducing the value of the denominator and increasing the test statistic for every region. Including only those sites that have a value isolates the locations where crime occurs. This seems intuitively more sensible as hotspots should only reasonably stand out against a background of other crime areas, and not against an artificial low value background. This modification is reflected in the equation (Ord and Getis 1995):

$$G_i^*(d) = \frac{\sum_j w_{ij}(d)x_j - W_i^*\bar{x}^*}{s^*\{[(nS_{1i}^*) - W_i^{*2}]/(n-1)\}^{1/2}}, \quad \text{for all } j,\ x_j \neq 0 \tag{3}$$

where $w_{ij}(d)$ is a spatial weights vector with values for all cells j within distance d of target cell i, W_i^* is the sum of the weights, S_{1i}^* is the sum of squared weights, and s^* is the standard deviation of the data in the cells.

4 Survey data

Crime hotspots generated from recorded crime data were compared with the perceived crime hotspots from a survey of 65 Nottinghamshire police officers. The police officers were surveyed across three sub-divisional stations of Nottinghamshire Constabulary, UK. Each individual was presented with a number of maps of their patrol area and asked to indicate on these maps the locations of the crime hotspots for the two types of high volume crime: motor vehicle crime and residential burglary. Multiple entries were permitted, as were nil returns if the officer felt that there were no areas of concentrated activity. The officers were also asked to outline the hotspot they felt was the most important, and this shape was used as the basis for the hotspot sizes.

The survey was piloted at the Meadows station in February 1998 and extended to Clifton and West Bridgford officers in early March 1998. The Meadows station has 29 officers assigned to it, of which 26 were surveyed. One caveat of the research is the difference between the survey time period and the crime data analysed. The respondents were asked to identify hotspots of crime based on the period from the end of August 1997 to the end of February 1998. Data for this period was unavailable so this study has therefore used crime data from the most recent period available, April 1996 to April 1997. A year's worth of data was used in an attempt to iron out any short term data fluctuations and identify the long term hotspots as consistent problem areas, areas which it was hoped that the officers would identify in the survey.

5 Results

As discussed earlier, the choice of bandwidth in this study has been determined by the size of hotspots interpreted from the survey of the local police officers. Selecting a suitable distance for the G_i and G_i^* statistic is another consideration. Figure 2 shows the effect of increasing bandwidth for calculating the G_i^* statistic on apparent residential burglary and motor vehicle crime densities in the Meadows (April 1996–April 1997). Although the area of significant hotspot activity increases as the search radius is enlarged, the number of crimes per unit area decreases and this has the effect of diluting the intensity of the hotspots.

As can be seen in Fig. 2 there is no discernible natural break and the choice of a suitable threshold should be based on a combination factors. Firstly, the crossing point of the two curves could be used as a good indicator of the position of maximum information transfer. Both residential and vehicle crime show this crossing point lies in search radii of between 4 and 5 cells. Converting this to metres means that the search radii should lie in the range 100 m to 150 m to satisfy both types of crime. Secondly, the search limit should include sufficient cells to avoid problems of the skewed distributions identified in Getis and Ord (1996). They claim that when the number of neighbours is large, approximate normality of the statistics can be assured, and suggest that a conservative choice of d would be such that the number of neighbours, j, is at least 30. However, they also say that 'when n is small, as few as eight neighbours could be used without serious inferential error unless the underlying distribution is very skewed.' (p. 265). A search limit of 3 times the basic cell size would produce $j = 29$ for a G_i^* statistic. Thirdly, consideration must

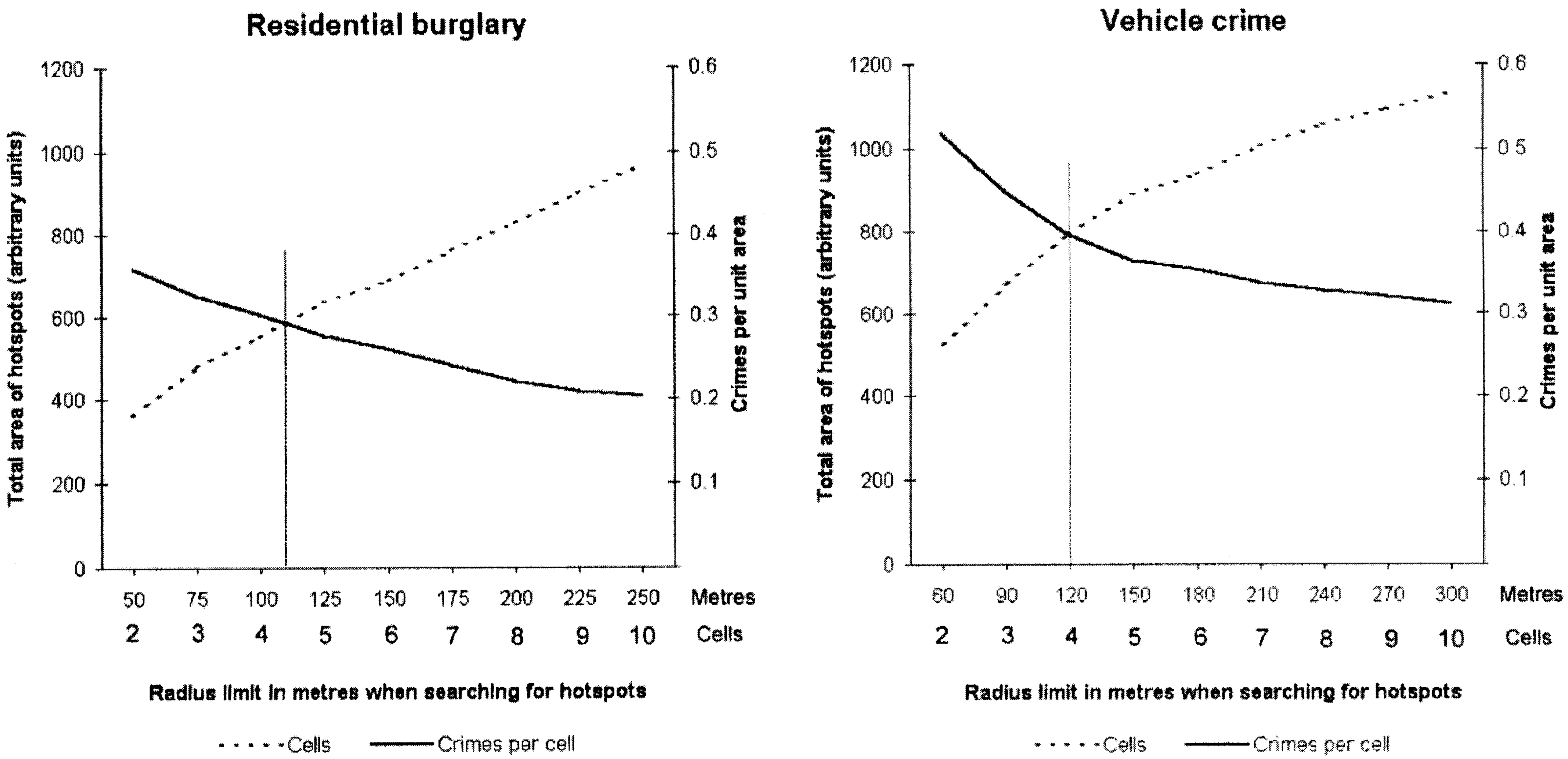

Fig. 2. Meadows crime hotspots (April 1996 to April 1997). Although the number of crimes represented in significant raster cells increases, the ratio of crimes to cells decreases, diluting the intensity of the hotspots

Table 1. Results from the hotspot analysis and police perception survey of three Nottinghamshire sub-divisions

Sub-division	Crime type	Generated hotspots	Police estimates	Correct guesses	% correct
Clifton	Residential burglary	6	38	25	65.8
	Motor vehicle crime	19	27	11	40.7
Meadows	Residential burglary	1	56	51	91.1
	Motor vehicle crime	4	55	33	60.0
West Bridgford	Residential burglary	1	53	33	62.3
	Motor vehicle crime	2	47	23	48.9

Table 1 shows that the level of accuracy for police estimates of hotspot locations is above 60% for residential burglary in all three areas, while the awareness of vehicle crime only just reaches 60% in one area (the Meadows) and does not reach 50% in the other locations.

be given to the size of the circle used during the formation of the original intensity surface. The cell size was set to 0.2 of the radius of the intensity integrating search circle, giving a size of 25 m and 30 m for burglary and vehicle crime respectively, as in Fig. 2. A search radius for hotspots similar to that chosen originally for intensity surface mapping provides minimum data distortion.

Given the crossing point argument discussed above, the requirement for a conservatively large choice of j, and the desire to minimise data distortion, this study used a limit of 5 times the cell size. Thus for residential burglary the limit was set to 125 m, and for vehicle crime to 150 m. This allowed j to be 81, which was also the same distance chosen for the original circle radii.

The hotspot estimate locations from the survey of police officers were digitised and compared with the computer generated hotspots using the 5 cell bandwidths and a G_i^* statistic which accepted locations with a high confidence level ($p < 0.001$). With a chosen fixed bandwidth it was possible to produce a simple binary classification and categorise the study areas into "hot" and "not" regions. Two maps of the Meadows sub-division demonstrate the comparison. Figure 3 shows the crime incidents as black dots and the crime hotspots as grey shading. Figure 3(a) shows the single hotspot and incidents for residential burglary, and Fig. 3(b) the four hotspots and incidents for vehicle crime. Of the 289 residential burglaries, 175 are located within a hotspot (61%). This percentage within a hotspot is similar to that of vehicle crime (323 of 504 crimes in a hotspot – 64%). It can be seen that vehicle crime is more generally distributed across the sub-division, while the residential burglaries are concentrated in the South and West of the sub-division, an area with a high proportion of council (social) housing.

The results from the three surveyed subdivisions can be seen in Table 1 and Fig. 4 shows, for the Meadows as an example, the centre points of the police guesses of hotspot location and the hotspots derived directly from the crime database using LISA methods. The Meadows is a comparatively poor redeveloped inner city area, while Clifton and West Bridgford are respectively suburban social housing and private residences. In the Meadows, of the 56 police residential burglary hotspot guesses, 91% (51) are within the hotspot region, while only 60% (33) of the 55 vehicle crime guesses are within the delimited region. This shows that sub-division burglary in the Meadows is con-

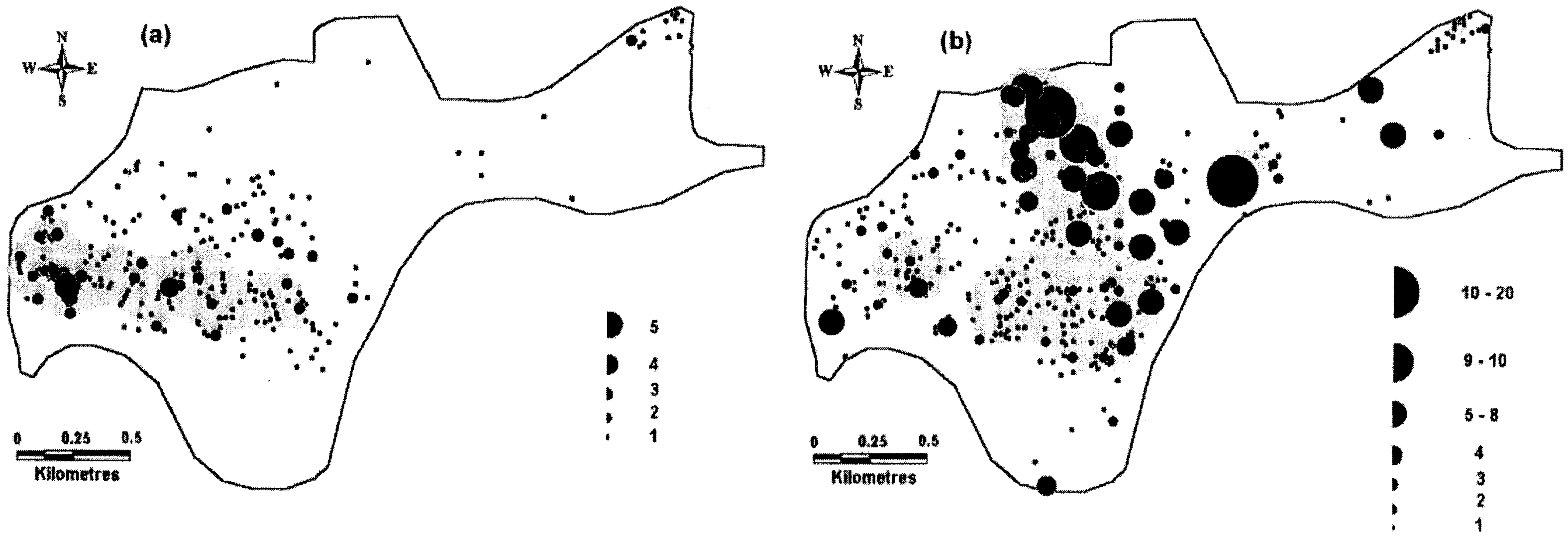

Fig. 3a,b. Meadows police sub-division crimes (*black dots*) and hotspots (*grey regions*) for residential burglary (**a**) and vehicle crime (**b**)

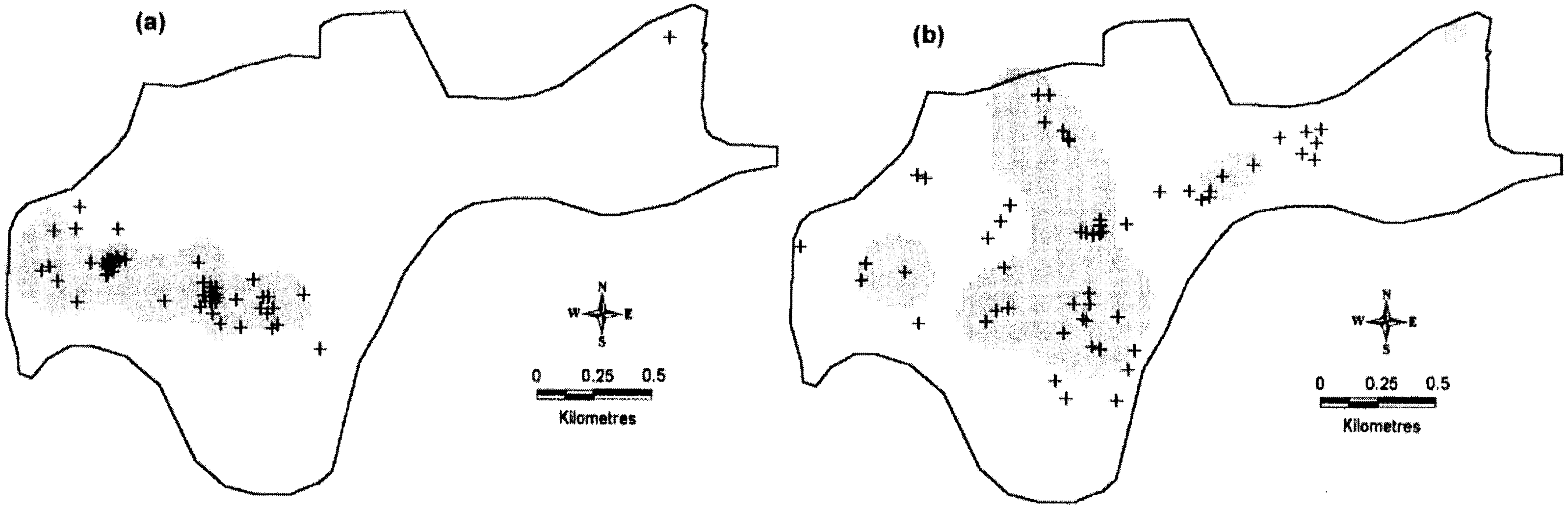

Fig. 4a,b. Meadows police officers' estimates of hotspot location (*black crosses*) and hotspots (*grey regions*) for residential burglary (**a**) and vehicle crime (**b**) centrated in a small area and the local operational officers are familiar with the crime problem and its location. The motor vehicle crime distribution is spread more widely across the sub-division and the officers are less familiar with the important vehicle crime hotspots. This correlates highly with the results in the other two sub-divisions, shown in Table 1, which indicate that knowledge of vehicle crime hotspots is less accurate than for residential burglary crime patterns.

One possible explanation for the officers' more inaccurate location of motor vehicle crime is that burglary has a higher priority, and motor vehicle crime, although recorded, is not pursued as actively. In addition the far greater number of motor vehicle crimes can generate information overload with officers who find it easier to recall the locations of fewer and discrete burglaries.

6 Characteristics of hotspots

There are a number of interesting features apparent in the hotspots generated by the process described above. The graphs in Fig. 2 show the number of crimes per cell for each potential search limit. It is also possible to calculate a value for crimes per cell in each individual hotspot for a predetermined search limit. In this manner a hierarchy of hotspot intensity can be constructed and each of the hotspot regions can be graded in a scale of importance. The police and crime prevention authorities can then be made aware of both the statistically significant crime hotspots in the local area, and the severity of each of the crime sites. An ability to determine a strategic ranking of hotspots permits the allocation of crime prevention resources in a logical and prioritised way and can lend scientific support to the process of resource allocation.

Another interesting feature of hotspot areas is that some remain almost fixed in size (hotpoints) growing little as the search limit for the LISA statistic is increased, while others expand, change shape, and can coagulate with neighbouring hotspots in an amorphous fashion to produce larger hotspots of indefinite but cohesive shape. These effects can be seen particularly well in the Clifton sub-division residential burglary data shown in Fig. 5. The hotspots identified as *a* are hotpoints which were often found to be caused by repeat victimisation of the same location. This has the effect of increasing the number of 'hits' at one single location. The lack of criminal activity in the surrounding area causes the area of the 'hotpoint' to remain stationary as the search limit increases. The lack of events in the vicinity may be caused by a number of factors. It might be that the premises are very large (such as a block of flats) and all crimes recorded at that location are geocoded with the same grid reference. It is also possible that one location is particularly vulnerable to crime and is a magnet to criminal activity. Either of these reasons would generate a large number of incidents at one location to the exclusion of surrounding areas.

In comparison the areas denoted by *b* in Fig. 5 show hotspot areas expanding as the search limit increases. There is an interesting difference in the two large hotspots in that the left hand (largest) hotspot has grown proportionately and smoothly, but the right hand hotspot has grown by aggregation from at least two individual hotspots. This demonstrates a differing functional morphology for the two and we would suggest that the right hand blob con-

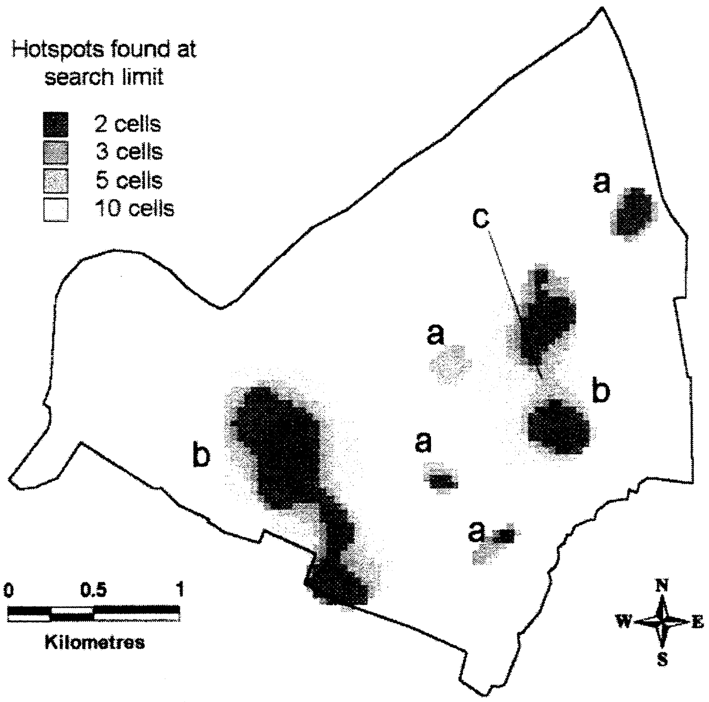

Fig. 5. Residential burglary hotspots for Clifton police sub-division (April 1996 to April 1997)

stitutes a group of hotspots sharing, at the different search limit, the same "hotbed". This hotbed of crime is noticeable in housing estates where crime is endemic and incidents are recorded at a large number of premises but not on a smoothly increasing areal basis. The increase in search limit has the effect of including more incidents and increasing the area of the hotbed. It can often be increased to the extent that two smaller hotbeds combine into a larger hotbed as the search limit increases. Area *c* in Fig. 5 demonstrates this by showing the joining of two separate hotspots when the five-cell search limit is reached. The different morphology of hot spots, points and beds can be identified from the graphical output of the generation procedure. These distinctions are important because they indicate different types of criminal activity and may elicit a different policing response.

7 Conclusion

This paper has addressed the issues of crime hotspot mapping and the use of hotspots as a means of assessing an aspect of the utility of introducing crime mapping systems into police forces both in the UK and abroad. It has been concerned solely with the use of such systems at the grass roots level for identifying local community level intelligence and problem identification, not for response evaluation. A definitive framework for this limited application of crime mapping has still to be formulated and the full scope of possible applications is not yet realised yet alone specified. Other applications abound but have not been considered here.

The Getis and Ord G_i and G_i^* statistics were shown to be effective in delineating hotspots within complex crime patterns, particularly when a simple binary classification into significant and insignificant regions was sought to compare with those delimited using police perception alone.

One feature explored in this paper has been the examination of the potential of mapping hotspots to estimate the requirements and suitability of a prospective mapping system. An example assessed the different levels of knowledge of officers in one sub-division of Nottinghamshire Constabulary, and provided an assessment of whether hotspot mapping might improve on that knowledge base. The results indicated a different perception of the important areas for residential burglary and vehicle crime between man and machine. This finding suggests significant implications for the dissemination of divisional crime intelligence, where concentration can be made in briefing officers on those criminal activities where a mismatch between knowledge and actuality occurs.

This type of work begins with the accurate identification of crime hotspots. The methodology suggested here is the use of a moving window approach to build up a raster grid of crime intensity and then to use LISA statistics to identify the local hotspots from the global background noise. This technique produces hotspots that are statistically significant, spatially precise, and follow the urban morphology of the underlying crime distribution. Visualisation of the results identifies two broad categories of hotspot: hotpoints and hotbeds. Hotpoints are characterised by a number of incidents at a single remote location, while hotbeds are formed by the amalgamation of a high number of crimes occurring at discrete locations close to each other. Both the use of this process for assessing police perception of crime distribution, and the recognition of different types of hotspot have important implications for crime prevention and the formation of local policing strategy. Identification of hotpoints and hotbeds may necessitate different crime prevention approaches. Numerous incidents at a hotpoint might demand an intensive crime prevention effort to make the target premises (in the case of burglary) more secure (known as target hardening), while the existence of a crime hotbed might require an increase in police patrols or targeting of known suspicious individuals.

References

Anselin L (1995) Local indicators of spatial association – LISA. *Geographical Analysis* 27:93–115

Bailey TC, Gatrell AC (1995) *Interactive spatial data analysis*. Longman

Besag J, Newell J (1991) The detection of clusters in rare diseases. *Journal of the Royal Statistical Society A* 154:143–155

Gatrell AC, Bailey TC, Diggle PJ, Rowlingson BS (1996) Spatial point pattern analysis and its application in geographical epidemiology. *Transactions, Institute of British Geographers*, NS 21:256–274

Getis A, Ord JK (1992) The analysis of spatial association by use of distance statistics. *Geographical Analysis* 24:189–206

Getis A, Ord JK (1996) Local spatial statistics: an overview. In: Longley P, Batty M (ed) *Spatial analysis: modelling in a GIS environment*. GeoInformation International, London pp. 374

Hirschfield A, Yarwood D, Bowers K (1997) Crime pattern analysis, spatial targeting and GIS: The development of new approaches for use in evaluating community safety initiatives. In: Evans-Mudie (ed), *Crime and health data analysis using GIS*, SCGISA Conference Proceedings, Sheffield

Illinois, Criminal Justice Information Authority (CJIA) (1996) *Spatial and temporal analysis of crime program (STAC)*. Chicago

Openshaw S (1984) The modifiable areal unit problem. *Concepts and Techniques in Modern Geography 38:41*

Openshaw S, Charlton M (1987) A Mark 1 Geographical Analysis Machine for the automated analysis of point data sets. *International Journal of Geographical Information Systems* 1:335–358

Ord JK, Getis A (1995) Local Spatial Autocorrelation Statistics: Distributional issues and an application. *Geographical Analysis* 27:286–306

Unwin DJ (1996) GIS, spatial analysis and spatial statistics. *Progress in Human Geography* 20:540–441