16 Tools in the Spatial Analysis of Offenses: Evidence from Scandinavian Cities

Vania Ceccato

CONTENTS

16.1 Introduction 267
16.2 Preparing a Data Set for Offense Analysis 269
    16.2.1 Quality of Data 269
    16.2.2 Type of Technique and Application's Goal 270
16.3 Techniques for Detection of Spatial Concentrations of Offenses 272
    16.3.1 Nearest Neighbor Hierarchical Clustering Technique 272
    16.3.2 K-Means Portioning Clustering Technique 275
    16.3.3 Hot Spots of Offense: Exploring Time Scale with Area-Based Data 276
16.4 Toward Explanations of Offense Patterns: Modeling Vandalism in Malmö, Sweden 278
16.5 Final Considerations 284
Acknowledgments 285
References 285

16.1 INTRODUCTION

Crime events are far from being random phenomena. They tend to occur in particular geographical areas in a city; they may occur at certain hours of the day and even in association with specific demographical, land use, and socioeconomic aspects of the population. As Hirschfield et al. [1] argue, the discovery of these patterns and regularities through crime analysis is the first step to more finely targeting resources to fight crime and formulate preventive strategies.

Recent literature suggests that geographical information systems (GIS) and spatial statistics can be used for urban planners in the toolbox designed to help in defining measures toward crime reduction in urban areas. Analysis of crime has been facilitated by the use of GIS in combination with spatial statistical techniques that are capable of handling spatially referenced offense data and integrating many types of data onto a common spatial framework, which together opens up new possibilities for better intervention practices in local planning [2–4]. Issues on crime data acquisition and data quality are examples of the remaining challenges for crime analysis and mapping. In a planning context, there is still a constant search of adequacy between choosing "appropriate tools" in accordance to the application goals, either for short or long-term decision-making.

After decades searching for answers in tables and pin maps, police officers and planners now aim at having more robust indicators of urban criminogenic conditions of the city. There is no single way to identify the dynamics of an offense, its spatial distribution over time and space, or the conditions that underlie its occurrence. Neither is the use of statistical packages and GIS the only way for providing a satisfactory basis for decision-making. What we argue here, however, is that this set of tools can, together with experts' knowledge and experience, contribute to a better understanding of the processes that are taking place in the city and provide support for short- and long-term strategies in local planning.

This chapter examines the potential of GIS in combination with spatial statistics in an exploratory analysis of urban geography of offenses in two Scandinavian cities. The term exploratory analysis implies here the use of techniques for detection of patterns in data (clusters) as well as statistical modeling. Techniques such as K-means portioning and Kulldorff's scan test are used to provide a simplified representation of where significant statistical concentrations of offenses occur across the city, while regression models are applied to explain such clusters. Three cluster techniques are applied to data on pickpocketing in Copenhagen, the capital of Denmark. This is followed by an attempt to explain patterns of vandalism using demographic, socioeconomic, and land use covariates in Malmö, the third largest Swedish city. The chapter concludes with a discussion of the strengths and limitations of these techniques for local planning.

These two Scandinavian cities were chosen for the following reasons. First, data availability was a decisive factor. In both Sweden and Denmark, the local police authorities systematically record offense data at a very detailed level of time and space. Second, Copenhagen and Malmö are part of the so-called Öresund region that includes southern parts of Sweden and the northern region of Denmark. The Öresund Bridge, which opened in July 2000, is the first fixed link between Sweden and Denmark, which replaces the boat traffic between Copenhagen and Malmö. Nordic and Baltic regions play an important role in international organized crime [5], and therefore it is reasonable to expect that a new transport link could potentially affect regional and local patterns of criminal activities in the region. For an extensive discussion of trends in offense patterns, see Ceccato and Haining [6]. Third, Malmö and Copenhagen have been targets of several governmental initiatives at all levels aimed at decreasing segregation and improving the quality of life of the citizens, including safety. In Malmö, for instance, since 1997, the State and the municipality have defined a strategy to meet the local needs of the so-called "problem areas." With regard to offense, most of these initiatives (e.g., *Storstadsatsningen*) are of a preventive character focusing on long-term structural changes. Fourteen development centers work to promote employment, which is believed to be crucial to decrease social exclusion, [7] and consequently discourage individuals from becoming offenders. Improvement of physical environment — which is an issue of controversy — is an example of a short-term intervention to improve urban quality and promote safety.

The structure of this chapter is as follows. Section 16.2 presents some guidelines on offense analysis using GIS and spatial statistics techniques. We discuss issues regarding data quality and the process of choosing the most suitable technique in relation to the application's goal. Three examples of techniques for detection of spatial concentrations of offenses are discussed in Section 16.3. A table summarizes the advantages and limitations of each technique from the user's point of view. In Section 16.4, the use of regression models to explain patterns of offenses using zone data is presented. Section 16.5 summarizes with a discussion about the potential of these techniques for planners to monitor, intervene, and strategically plan safety issues at neighborhood and municipal levels.

16.2 PREPARING A DATA SET FOR OFFENSE ANALYSIS

There are two important aspects to be aware of when working with crime analyses at urban level. The first relates to the quality of the data set, while the second concerns issues of selecting the most suitable technique in relation to the application's goals.

16.2.1 Quality of Data

Data reliability is an important issue when mapping offenses. Underreporting is a known cause of lack of reliability in databases of offenses. According to Bowers and Hirschfield [8], levels of reporting tend to vary with the type of offense that has been committed and its seriousness. The British offense survey has shown, for instance, that burglary and theft of vehicle is far more likely to be reported than many other types of offenses, such as vandalism and domestic violence. There are also indications that the offense reporting level may be underestimated in areas where people think that it is not worthwhile reporting them, for example, in deprived areas with low social capital [9,10].

There are other problems of data quality that take place during the process of recording offenses. This can be caused by the lack of information about the event from the victim him/herself (not knowing exactly where the offense took place) or by the police officer failing to record the event properly (missing record on the exact location/time of the event). It may be the case that the police officer failed when entering the data on the system, at the first attempt, and the record becomes duplicated as soon he or she makes the second attempt. This may create extra cases in those particular locations, and that, if not identified in advance, may contribute to "false hot spots."

Many of these problems of data quality relate to the lack of systematization of procedures when recording an offense. Despite the fact that there are conventions for recording offenses in many countries, including Scandinavia, differences still occur in practice. For instance, an assessment of the Swedish offense database over

Skåne (Southern Sweden) has shown that large municipalities often have better offense records than the small ones. It is not rare that, in small towns, victims or the police officers only make an estimate of the offense's location (such as "in the surroundings of Jonsson's bakery," "Emma Larsson's garage," "in front of surgery medical center," "30 meters from F-bank," or "local bus stop at G-school"), which contributes to poor-quality records.

AU: Are all of these examples needed? I think 3 would be sufficient – this many distract from the point.

Similar reasoning explains the problems found when recording the exact time that an offense took place. This happens, for instance, when someone burglarizes a house while the owner is away. When there is no available information about the time that the event occurred, a range in hours is often the common practice (e.g., 12:00–16:00). However, for analysis purposes, this limits any space–time trend assessment. Aoristic models are used to interpolate time-related missing data (see, for instance, [11,12]).

Another issue related to offense data is related to the geocoding process. Geocoding is the process of matching records in two databases: the offense address database (without map position information) and the reference street map or any other "address dictionary" (with known map position information). The quality of the geocoding process depends very much on the quality of the offense records, the quality of the address dictionary, and the chosen method for geocoding. In cases where the matching of the exact offense location is not possible, a common practice is to choose a near location (such as midpoint of street) or the polygon centroid of a region (e.g., district polygon). This practice creates the so-called "dumping sites" for records [8], which is believed to generate false offense concentrations and consequently, a poor basis for any type of planning intervention. Ceccato and Snickars [13] estimate, for instance, that about 25% of all offenses committed in a Swedish neighborhood were attached to the polygons' centroids of the local commercial area instead of their "real locations."

Cases also exist where the offense site is unidentifiable, for example when it took place between A and B, on a bus, train, airplane, or through the Internet. Evidence from the Swedish database shows that it is unclear in some cases if the address, when reported, is actually the one where the individual was victimized. This may create "false hotspots," since a high number of records may be reported in airports, bus stations, and railway stations. One good way to identify these false hot spots is to try to check for long-term patterns. In case of false hot spots, they may disappear over time, since changes in the way the offense is reported may change, and this affects the choice for "dumping sites."

16.2.2 Type of Technique and Application's Goal

Another key issue when dealing with crime mapping refers to the process of identifying the most appropriate technique in accordance with what the user wants to achieve. According to Craglia et al. [2], crime mapping has essentially three main areas of application: dispatching, community policing, and offense analysis and resource planning (Table 16.1). Each tends to operate at different geographical scales, involving different actors (e.g., police officers, planners, community experts), and has different requirements in terms of data quality and currency and analytical

TABLE 16.1
Typology of Offense Mapping Applications

Application	Data	Geography	Function
Dispatching	Second/minutes	x,y co-ordinate	Visualization
Community policing	Hours/days	Neighborhood/district	Mapping/some analysis
Offense analysis and resource planning	Weeks/months/years	City	Analysis/modeling

Source: From Craglia, M., Haining, R., and Wiles, P., *Urban Stud.,* 37, 712, 2000. With permission.

capabilities. Its application also differs, as suggested by Haining [14, pp. 37–38], in terms of time horizons, from the short-term tactical to long-term strategic deployment of resources. Tactical deployment of resources is often focused on a very narrow and specific set of objectives (e.g., sudden upsurge in street robberies), requiring rapid data collection followed by relevant data processing, perhaps a hot spot analysis to identify the areas in need. Strategic deployment of resources is based on long-term data series and on analyses of the underlying factors (e.g., demographic, socioeconomic, land use) that might help to characterize and explain crime patterns.

On a daily basis, police officers may make use of precise data to dispatch patrol cars to the scene of crime. Community policing requires user-friendly systems for officers to enter the location and characteristics of a reported offense. According to Rich [15], officers in the United States have become producers of maps rather than simply users. Using information collection and automated mapping, officers are able to "walk and use" the system, which requires a limited expertise. This may require a need for rapid data processing, perhaps by using pin maps or cluster detection techniques, for hot spot detection. This is fundamental to identifying areas that need attention and prioritizing resources. There are examples where this may involve other actors such as planners, experts, neighborhood volunteers, and residents (see, for instance, [13]).

An assessment of urban criminogenic conditions could combine different tools and types of data, from simple monitoring of frequency diagrams to more sophisticated techniques. Ratcliffe and McCullagh [11] exemplify, for instance, how the combination of hotspot analysis within a GIS environment, a hot spot perception survey of police officers, and small focus groups can be used to assess the dissemination information on high-risk offense areas. Another way is by using techniques that allow different types of data to calculate risks of being a victim of crime across the city. Tracking changes over time is also a very important issue, not only in monitoring how the offense risks vary over time but also because these changes should potentially affect the police actions and security measures on short- and long-term interventions. A map showing that residential burglary is concentrated in neighborhoods A and B during the day and in C and D during the night may be crucial for avoiding intervention in the "wrong neighborhood" and, consequently, waste of resources. Maps showing changes over time are important to highlight not only how an offense prevention program has reduced offenses in the target area but also whether or not the program has been successful in avoiding offense displacement to the surrounding areas.

Crime is often regarded as the tip of the iceberg of other long-term social problems. The role of long-term strategic planning often involves the coordinated work of the police with public sector agencies and other local actors. In this context, offense mapping goes beyond the detection of patterns and tries to explain why certain areas have a high risk for offense through modeling or combination of several data sets and techniques. Bowers and Hirschfield [8, p.5] state "mapping can be used to make useful inferences about the underlying processes that are causing particular types of offense cluster to form … it can be used as evidence of the likely presence of a particular process." In the next section, three examples of cluster detection techniques in offense mapping are compared, and this is followed by a discussion of their potential and limitations in the context of local planning.

16.3 TECHNIQUES FOR DETECTION OF SPATIAL CONCENTRATIONS OF OFFENSES

Areas with high spatial concentrations of offenses are often generated by dominance of certain types of land uses in the city (such as a concentration of pubs, restaurants, tourism-related places) but also by the relationship between activities and places (e.g., pickpocketing in central urban areas or drug selling points near schools and clubs for young people). There are many different statistical techniques designed to identify spatial concentrations of an event [16–18]. Cluster statistical techniques "aim at grouping cases together into relatively coherent clusters" [1]. This section presents three different methods for cluster detection, using point and area-based data, and reviews issues and challenges associated with such techniques.

16.3.1 Nearest Neighbor Hierarchical Clustering Technique

The nearest neighbor hierarchical (NNH) clustering technique identifies groups of incidents that are spatially close. It clusters points together on the basis of a criterion. The clustering is repeated until either all points are grouped into a single cluster or else the cluster criterion fails [17]. For this example, CrimeStat® was utilized. This is a spatial statistics program for the analysis of offense incident locations, developed by Ned Levine & Associates under grants from the National Institute of Justice [1] (available at http://www.icpsr.umich.edu/NACJD/offencestat.html#SOFTWARE). Because this package is used in many police departments in United States as well as by criminal justice and other research institutes, it has been decided to assess its applicability in the Scandinavian context. For the purpose of this case study, we focus on pickpocketing over Copenhagen, as it is among the twenty most common offenses in the period 2000 and 2001. The data set used in this analysis was extracted from the Copenhagen Policy Authority's database on offenses from 2000 to 2001. As Figure 16.1 shows, pickpocketing is very concentrated in inner city areas of Copenhagen, excluding the municipality of Frederiksberg.

AU: URL not found. Use http://www.icpsr.umich.edu/NACJD/crimestat.html#SOFTWARE instead?

The CrimeStat NNH technique uses a nearest neighbor method that defines a threshold distance and compares the threshold to the distances for all pairs of points.

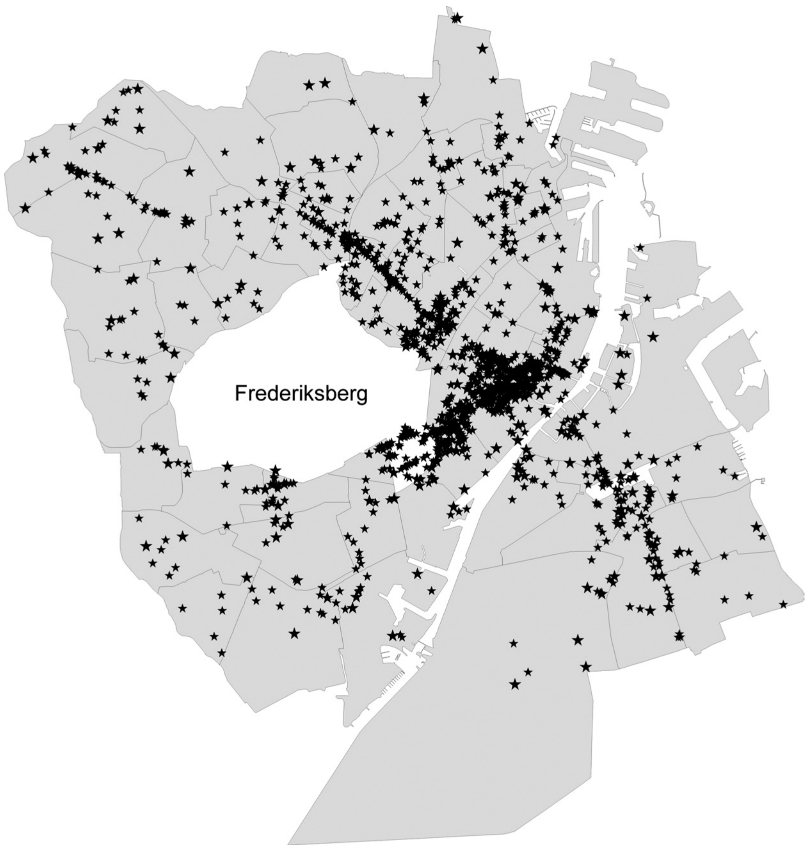


FIGURE 16.1 Pin map of pickpocketing over Copenhagen city, 2000 to 2001. Frederiksberg does not belong to Copenhagen municipality.

Only points that are closer to one or more other points than the threshold distance are selected for clustering. This threshold distance is a probability level for selecting any two points (a pair) on the basis of a chance distribution. In this first criterion, we have chosen the default value for the threshold distance (with probability of 0.5), which means that if the data were spatially random, approximately 50% of the pairs will be closer than this distance. However the number of clusters is dependent on the threshold distance and the minimum number of points in each cluster. In this case, changing the threshold distance from the default 0.5 to the minimum value (the likelihood of obtaining a pair by chance would be 0.001%), does not affect the number of clusters. The second criterion is the minimum number of points that should be included in any cluster. Since we wanted to detect clusters that would reveal vulnerable microenvironments in the city, we used the default of 10 as the minimum cluster size.

Figure 16.2 illustrates the first- and second-order clusters of pickpocketing over Copenhagen using two minimum cluster sizes. Decreasing the number of points per cluster from 10 to 5 increases the number of cluster found from 14 (13 first order,

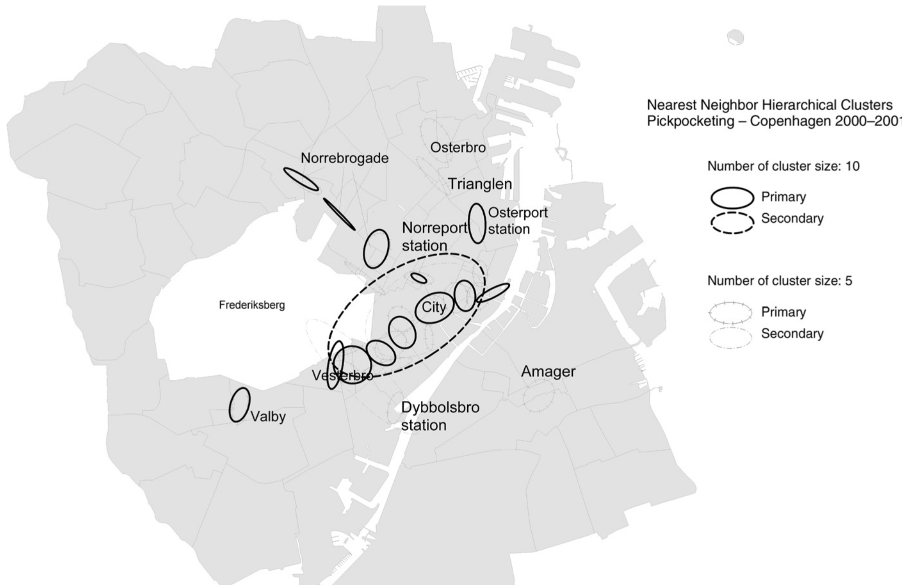


FIGURE 16.2 Nearest neighbor hierarchical (NNH) clusters of pickpocketing over Copenhagen varying cluster size.

1 second order) to 22 (20 first order, 2 second order). Therefore, the user's knowledge of the study area in making sense of the NNH outputs is fundamental in choosing these criteria. The third criterion is the output size of the clusters (in the form of a deviational ellipse) that the user specifies in terms of standard deviations. We used the default value of one standard deviational ellipse "since other values might create an exaggerated view of the clusters" [18].

As these findings illustrate, the NNH technique can identify in detail geographical environments where pickpocketing is concentrated — an important piece of information for short- and long-term intervention. What is evident in this pattern is how clusters of pickpocketing follow main streets (e.g., Norrebrogade), stations (e.g., Osterport station, Norreport station), and local centers (e.g., Trianglen, Österbro), most of which are concentrated in the inner city areas of Copenhagen. This is confirmed by the form and location of the second-order cluster. These places are mostly constituted by either transport links (such as main streets) or transport nodes (such as train stations) — public places that lack "capable guardians" despite being crowded places. Travelers who could in theory be considered as informal guardians may in practice be ineffective. Most people have no sense of ownership in places like train stations and often do not want to get involved, either to intervene during the act or later as witnesses. The same reasoning could be applied to main streets, with a large flow of people passing through during the day. Therefore, transport links and transport nodes of any kind are typical examples of poorly guarded places and highly attractive to motivated offenders for committing pickpocketing.

Another advantage of the NNH technique is that it can be applied to an entire data set (e.g., from the neighborhood level to the county level), which facilitates comparisons between different areas and different cluster levels over time. This means that there is no need to break down the database into different levels. This technique allows the user to detect clusters at several levels: from microurban environments, through first-order clusters (e.g., street corners), to a more comprehensive view of the whole city or county by checking second-, third-, and higher-order clusters.

However, the NNH fails to detect clusters based on attributes other than the observations' location (e.g., offense). NNH technique detects areas where a lot of pickpocketing is committed, regardless of underlying distribution of population or characteristics of land use, for instance. Moreover, as it is suggested, the size of the grouping area is dependent on the sample size, which does not provide a consistent definition of a hot spot area, since a cluster should be dependent on environment and not on the sample size. Finally, the total number of clusters is dependent on the minimum size cluster, which in its turn varies across users, their experience, and knowledge to make sense of the results from each output.

16.3.2 K-Means Portioning Clustering Technique

The K-means clustering technique is a portioning procedure where the data are grouped into K groups defined by the user. The routine searches to find the best positioning of the K centers and then assigns each point to the center that is nearest. Unlike NNH technique, all observations are assigned to clusters, and therefore there is no hierarchy in the procedure, which creates clusters at one level only [17,18]. Since we already knew from the results of NNH technique that there were 13 first-order clusters when using the default values for all criteria, we decided to set as 13 the total K groups in CrimeStat, so we could check whether the clusters had the same geography as the ones produced by the NNH technique.

When many clusters look concentrated in a geographical area such as in Copenhagen city center, a smaller separation is suggested. This will tend to subdivide more concentrated clusters, reducing the distance of each point from the cluster center. In the case of pickpocketing in Copenhagen, the difference in the geography is visually detected in the results when the default value 4 is decreased to 2 (Figure 16.3). As with the NNH method, the K-means clusters also vary, depending on the user's knowledge and experience in identifying the "right number of clusters" or the "right separation" for a specific area. There is always a risk of setting either too many clusters, which will result in clusters that don't really exist, or too few, which will lead to poor differentiation among environments that are different in nature. For an extensive discussion of this issue, see Murray and Grubesic [19].

Compared with the NNH clusters, the K-means clusters are generally larger in size, especially those in the peripheral parts of the city (see, for instance, the clusters in Valby, Amagerbrogade, and Emdrup). At least three new locations appear as clusters in the K-means map that were not clustered in the NNH method, regardless of the variation in the criteria. Both NNH and K-means techniques are better seen as exploratory tools for refining clusters of high values. This implies that if police officers or planners have previous knowledge of where there should be "cluster of high values" then these techniques could be used to verify if these hot spots actually

FIGURE 16.3 K-Means clusters of pickpocketing over Copenhagen, varying the cluster's center.

correspond to their perception. In many cases, these outputs generate new questions for the user when "unexpected" hot spots are identified. Therefore, the user's experience is crucial to differentiate "possible real clusters" from the ones that are a product of the statistical procedure or of the data quality. As Craglia et al. [2, p. 716] suggest, these techniques are useful for short-term intervention, since they provide a good basis of where offenses are concentrated. However, "from a strategic point of view, there is often a need to go beyond the pure count of events," which these techniques are unable to provide. In the next section, Kulldorff's scan text is discussed, using the pickpocketing data set in polygon-based format.

16.3.3 Hot Spots of Offense: Exploring Time Scale with Area-Based Data

In this section, we assess the Kulldorff's scan test to detect clusters of pickpocketing over Copenhagen in relation to its nighttime and daytime population. First, we split the pickpocketing data set in 24 slices, corresponding to 24 hours of the day to check the variations in the geography of this offense over time. Then, we discuss the reasons why choosing the "right" denominator when detecting patterns of clusters over time is an important issue for planning purposes.

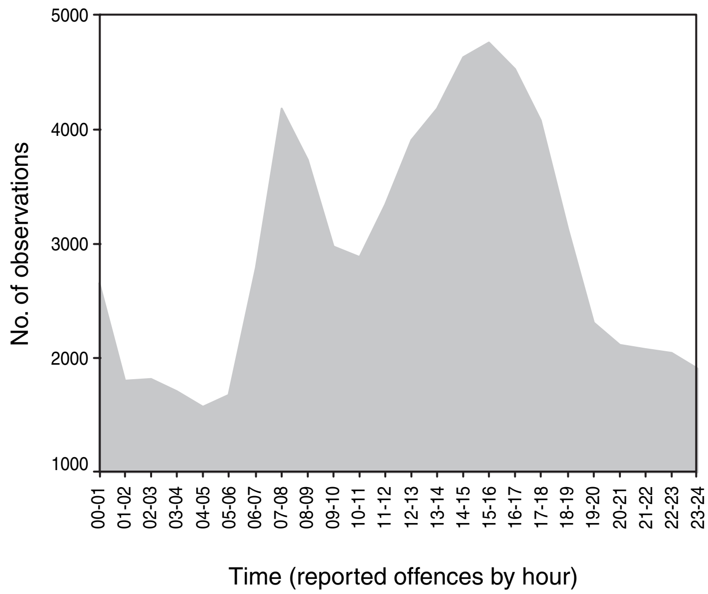
In order to detect geographical clusters of pickpocketing for the 24 hours of the day, Kulldorff's scan test was used [20] in the data set of Copenhagen. This software has a number of techniques routinely used in spatial epidemiology but could be used for virtually any application searching for measures of relative risk. Kulldorff's scan test has a rigorous inference theory for identifying statistically significant clusters

[21]. The tests use the Poisson version of the scan test, where the number of events in any area is assumed to be Poisson distributed. This adjusts for heterogeneity in the background population. The spatial scan statistic imposes a circular window on the map, which is in turn centered on each of several possible centroids positioned throughout the study area. For each centroid, the radius of the window varies continuously in size up to a maximum window size that includes 50% of the spatial units. The circular window is flexible both in location and size and is moved across the maps to search out the most significant clusters irrespective of size. The spatial scan statistic uses a large number of distinct geographical circles, with different sets of neighboring polygons within them, each being a possible candidate for a cluster. Each spatial unit is represented by a centroid, which determines, for any given window, whether the spatial unit is inside or outside the window. The likelihood function is maximized over all windows, identifying the window that constitutes the most likely cluster (that is to say, the cluster that is least likely to have occurred by chance). Its distribution under the null hypothesis and its corresponding p-value is obtained by repeating the same analytic exercise on a large number of replications of the data set generated under the null hypothesis, in a Monte Carlo simulation (for more detailed information on the spatial scan test, see [20,22]).

Clusters of pickpocketing vary in size depending on the time of day and the denominator with which they are compared. Figure 16.4 shows the frequency of pickpocketing by (a) hours of day in Copenhagen, as well as two examples of clusters resulting from Kulldorff's scan test, standardized by (b) daytime and nighttime population and (c) total population. The first slice was chosen because it refers to the lowest frequency counts of pocket picking, between 7:00 and 8:00 in the morning. The second slice refers to one of the peak hours for this offense, between 2:00 and 3:00 in the afternoon. This example shows that clusters of pickpocketing are unchanged in size when allowing for variation in total population in each area.

However, if we assume that pickpocketing is a function of the daily variation of total people who work and live in a certain area, then the denominator with which the offenses are compared should also reflect this change. Nighttime population was used as denominator for the sample between 7:00 and 8:00 in the morning, while daytime population was used for the afternoon sample. As Figure 16.4 illustrates, the afternoon cluster was susceptible to change in the denominator and, as a consequence, the cluster became larger than the one standardized by the total population. Despite the fact there has been no large geographical change in the cluster location in the case of pickpocketing (no other clusters appeared, for instance), this may have many practical implications since decisions can be taken based on inaccurate information. For police tactical work, this means that patrols may be sent to a high-risk area that is smaller than it should be. In cities where there is a great variability in space–time clusters, targeting resources to the "right area" may be a difficult task for planners working with strategic distribution of resources.

Lack of demographic and land use data used to create the ratios often limits crime analysis over time. When they are available, they may not be appropriate to all types of offenses. For pickpocketing, robbery, and other violent crimes, the standardization is commonly performed using population totals, and less commonly, day- and nighttime population. However, for offenses such as vandalism, car-related



a

FIGURE 16.4 Cluster of pickpocketing in Copenhagen: (a) Frequency of pickpocketing by hours of day. (b) Standardized by night and day time population. (c) Standardized by total population.

thefts, and residential burglary, there is no agreement on which denominator should be used for the standardization when the time dimension is taken into account. For car theft, for instance, the total number of cars in each zone might be a good indicator but such data is rarely available by day and night time periods. Table 16.2 summarizes the main advantages and disadvantages of each cluster technique.

16.4 TOWARD EXPLANATIONS OF OFFENSE PATTERNS: MODELING VANDALISM IN MALMÖ, SWEDEN

A common question among experts and those involved in security issues in strategic local planning is to what extent the relative risk of being a victim of crime varies across the city and why the risk is greater in certain areas than in others. In this section, we suggest a set of procedures that first provide a notion of risk variability for vandalism across the city, followed by an attempt to explain such distribution, linking vandalism to socioeconomic and land use covariates. The Malmö data set used in this analysis was extracted from Skåne Policy Authority's database on offenses in 2001. In the Malmö case, vandalism involves "offenses on physical targets (e.g., causing damage to cars, walls, buildings, including graffiti) and disturbance (e.g., by starting a fire)."

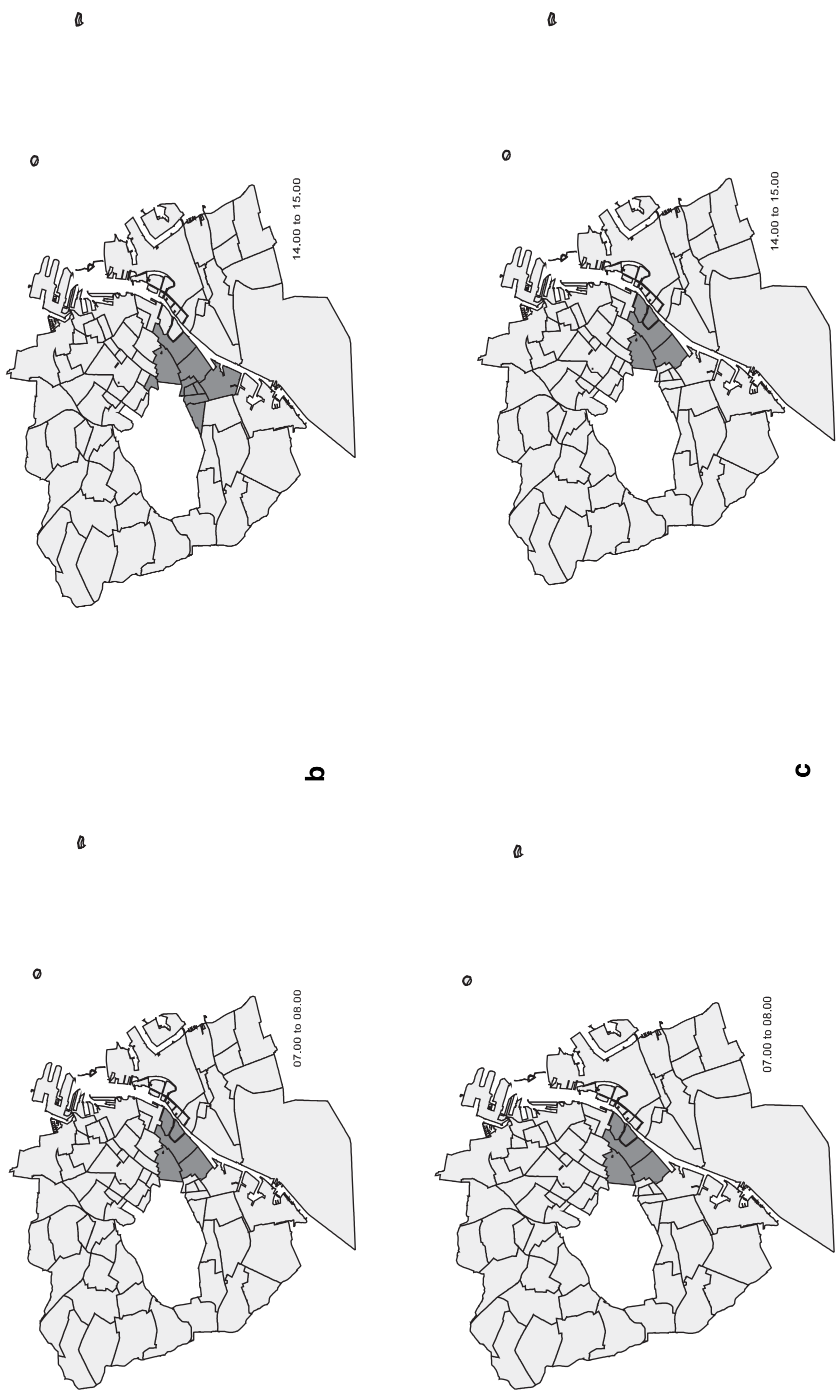


FIGURE 16.4 (continued)

TABLE 16.2
Cluster Detection Techniques: Advantages and Disadvantages

Technique Type	Definition	Advantages	Disadvantages
Hierarchical technique: Nearest neighbor hierarchical clustering	NNH clusters observations together on the basis of a threshold distance between their locations Number of clusters is highly dependent on the threshold distance and the minimum number of points in each cluster chosen by the user	Identification of small scale clusters (street, local commercial centers, etc.) Multilevel clustering detection (e.g., street, district, city, county levels) that allows different policing strategies	Detects clusters based only on observations (e.g., offenses, x-y co-ordinates); attributes associated with the observation are not incorporated in the routine Hot spots are dependent on number of observations Minimum size cluster may vary across users
Partioning technique K-Means portioning clustering	K-means divides the observations into a predetermined number of clusters chosen by the user Separation value may also have influence on the pattern	User defines the number of clusters Output is suitable for larger geographical areas	User defines the number of clusters *a priori;* therefore it is highly dependent on user's knowledge and experience There is no hierarchy in the routine
Risk-based technique Kulldorff's scan test	Kulldorff's scan test detects clusters based on an underlying denominator variable, such as population, number of households.	Has a rigorous inference theory for identifying statistically significant clusters (Haining and Cliff 2003) Standardization provides a robust measure of how offenses vary over an area Does not depend on polygon size Provides a general view of concentration patterns	Difficulty in including detailed environmental data (x,y data when working with polygon data, for instance) Using polygon data, the results may be too crude given the high degree of heterogeneity within zones and across them.

AU: Ref. [14]? [21]?

Figure 16.5 shows the map of relative risk for vandalism in Malmö, using a polygon framework. This defines the expected number of acts of vandalism for each polygon, under the assumption that vandalism occurs randomly across the city. The standardized vandalism rate (SVR) for polygon i is given by:

$$SVR(i) = [O(i) / E(i)] \times 100 \qquad (16.1)$$

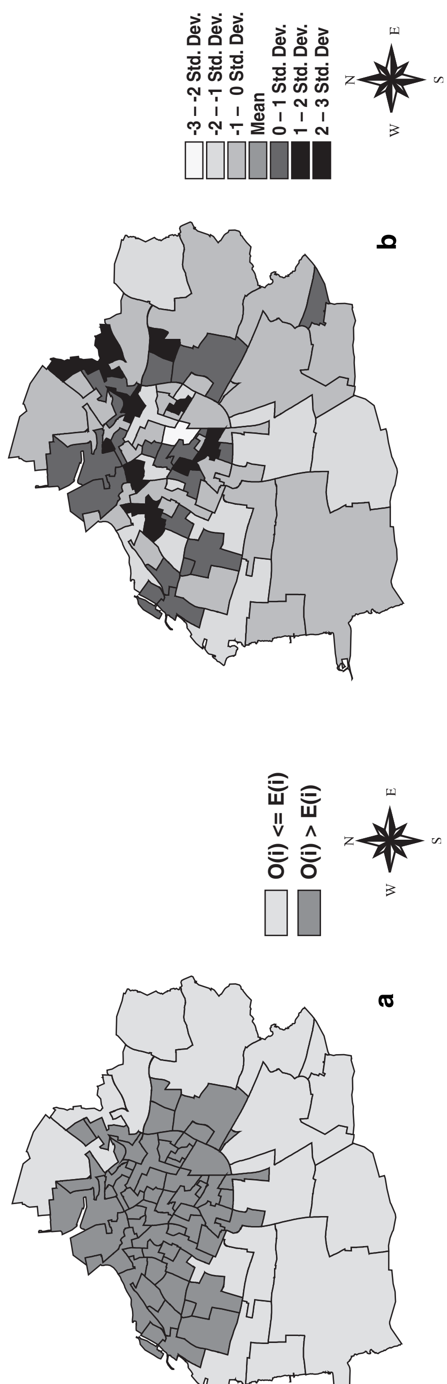


FIGURE 16.5 (a) Standardized vandalism rates. (b) Spatial pattern of residuals — ordinary least square regression model.

where O(i) is the observed number of cases of vandalism, and E(i) is the expected number of cases of vandalism. An average vandalism rate for Malmö was obtained by dividing the total number of offenses by the total size of the chosen denominator. For vandalism, the best denominator suggested is the area of the polygon (see [23]). Since the polygon area varies greatly from the Malmö city center to the periphery, the total population in each polygon was used in this study (minimum population size was 998, the maximum was 7836, the mean was 3748, and the standard deviation was 1617). For each polygon i, this average rate is multiplied by the size of the chosen denominator in polygon i to yield E(i). The observed number of incidences of vandalism in each polygon is later divided by the expected number and then multiplied by 100. Any polygon with an SVR greater than 100 has a vandalism rate greater than would be predicted on the basis of its area (Figure 16.5(a)).

As suggested by Craglia et al. [2] and Haining [14] this statistic may not be the best estimate of the relative risk across the region, since counts for areas with relatively small numbers of population will be sensitive to small errors in reporting offenses and sensitive to random errors in the occurrence of offenses. Rates computed on areas with small numbers of population will therefore be less robust than those computed on large numbers of population. However, the SVR still provides a local measure of crime concentration in relation to the population in the area. Not surprisingly, most of the more vulnerable areas for vandalism in Malmö are concentrated in the central areas of the city — where a mixture of land use determines daily activities and the vulnerability of the area for vandalism. For long-term intervention, more knowledge would be needed on the underlying processes that are causing this particular type of offense to cluster in the central areas and vary greatly in other parts of the city.

In order to try to explain the relationship between vandalism rates and differences in demography, socioeconomic, and land use composition of the city, the ordinary least square linear regression model was fitted. Earlier research has emphasized the relationship between vandalism, social disorganization risk factors [24,25], and low guardianship. Individuals living in areas with high rates of disorder or crime tend to lose their sense of commitment to the neighborhood. Individuals living in problem areas may refrain from local social life, and this breaks down formal and informal social control and involvement at the neighborhood level [26–28]. This in turn leads to more crime and disorder. Although social capital seems to affect crime in general, its effect may depend on crime and neighborhood type. Levels of crime are significantly higher than expected in disadvantaged areas with low levels of social capital [29]. Based on this existent literature, a set of explanatory variables were drawn, as the proportion of:

- Population younger than age 18 (X_1)
- Population with (at least) one parent born abroad (X_2)
- Population born abroad (X_3)
- Population moving into the area (X_4)
- Population moving out the area (X_5)
- Privately owned single family houses (X_6)
- Average income per household (X_7)

- Local leisure associations by population (X_8)
- “Neighborhood Watch Schemes” by population (X_9)
- Bus stops by population (X_{10})
- Central area — dummy (X_{11})
- Commercial areas — dummy (X_{12})

Since the set of SVR values show a highly skewed distribution, the raw SVR was transformed using its square root transformation to produce a data set that is more nearly normal. The regression analysis and the creation of the lagged variables was implemented in SpaceStat 1.91 [30], because the software has regression modeling capabilities that are appropriate for spatial analysis. SpaceStat provides several statistics measuring the fit of the model, including diagnostic tests, such as tests for multicollinearity among independent variables and tests on model residuals (normality, heteroscedasticity, and spatial autocorrelation). In order to perform tests for spatial autocorrelation in the residuals, a binary weight matrix was used for accounting for the spatial arrangement of the data.

Model results show that the land use variables were statistically significant in explaining the pattern of relative risk in Malmö. The presence of bus stops seems to deter vandalism, while the dynamic of inner city areas (especially the commercial area) is responsible for high vandalism rates. However, we should be careful in drawing conclusions from these results, since the model shows spatial dependence on the residuals (Moran’s I is significant), which is a violation of classic regression assumptions (Table 16.3). Martin [31] suggests that, if significant spatial dependence is found on the residuals, this indicates that some source of variation has been omitted from the model or that the functional form of the model is not correct. As Figure 16.5(b) shows, areas with higher vandalism rates than predicted by the model tend to occur in groups. One way to account for this spatial dependence on residuals is

TABLE 16.3
Results of Classic and Spatial Lag Models — SVR

Classic Model — Ordinary Least Square Estimation	**Spatial Lag Model — Maximum Likelihood Estimation**
Y = Square Root of the Standardised Vandalism Rate	Y = Square Root of the Standardised Vandalism Rate
Y = 15.72+(-5.18)X10**+ (6.87)X11**+(4.78)X12*	Y = 0.65W_y+6.48+(-3.64)X10**+ (3.35)X11**
(–4.91) (4.47) (1.92)	(7.01) (4.34) (–4.43) (2.67)
(t-values in brackets)	(z-values in brackets)
* significant at the 5% level	* significant at the 5% level
**significant at the 1% level	**significant at the 1% level
R^2 X 100 = 46.5% R2 (adjusted) X 100 = 44.07%	R^2 X 100 = 59.1 % R2 (adjusted) X 100 = 67.7%
Log Likelihood –194.87	Log Likelihood –181.36
Akaike Information Criterion 397.75	Akaike Information Criterion 370.73
Normality of errors — Jarque-Bera 1.49 Prob 0.51	Heteroskedasticity — Breusch Pagan 2.32 Prob 0.31
Multicollinearity condition 2.74	Spatial Lag dependence (likelihood ratio test) 30.84 Prob 0.00
Heteroskedasticity — Breusch Pagan 7.76 Prob 0.06	
Moran’s I (error) 0.29 Prob 0.00	Lagrange Multiplier test (error) 0.20 Prob 0.65

by applying a spatial lag model. A spatial lag model treats spatial dependence as a spatial diffusion process. Thus, in this particular case, the spatial lag model tests the explanation that vandalism rates are spatially autocorrelated, because offenses from areas with high vandalism rates spill over into adjacent areas through people's spatial interaction. Patterns of offending do not recognize district boundaries, and motivated offenders do not only operate in their own districts. Therefore, the spatial lag model was used in recognition that the effects on vandalism ratios, of high or low levels might extend beyond the boundaries of the particular spatial unit. The spatial lag variable (Y_w) is automatically computed in SpaceStat as the average of vandalism rates in adjacent polygons (Table 16.3).

By comparing the fit of the spatial lag model with the classic model, we realize that the overall fit of the model was substantially improved by the inclusion of the spatial lag variable (Table 16.3). Now approximately 60% of the variation in the vandalism rate is accounted by two variables in the model, mostly by the presence of bus stops and "being in the inner city areas." Martin [31] suggests that the inclusion of the lag variable improves the model fit at the same time that it reduces the magnitude of the effects of each independent variable. The most extreme case is the variable X_{12} (dummy for commercial areas) that becomes no longer significant in the spatial lag model. The model results also show that the problem with heteroscedasticity has been reduced, while the spatial dependence on the residuals is no longer significant.

High rates of vandalism are found in central areas, especially in less guarded places, possibly with fewer bus stops. These findings are quite general but are indicative of the processes that generate vandalism in central areas. In the case of Malmö, the presence of people in public places, around bus stops in central areas, seems to deter vandalism. These results may produce insights that may be helpful in long-term strategies, in pointing the way in terms of how resources should be targeted, particularly to vandalism in central areas.

16.5 FINAL CONSIDERATIONS

This chapter has illustrated how GIS and a set of spatial statistics techniques can be used to detect and aid in explaining the geography of selected offenses in two Scandinavian cities. A discussion of the results of three cluster techniques applied to pickpocketing data is presented. This is followed by the attempt to explain patterns of vandalism using demographic, socioeconomic, and land use covariates. These case studies also show how results are highly dependent on the employed criteria and/or data chosen for each technique.

The cases of pickpocketing and vandalism provide evidence of the complexity of factors underlying the geography of crime in urban areas. Although in Copenhagen transport nodes, such as train stations, seem to create just the "right" criminogenic conditions for pickpocketing, in Malmö, bus stops and the dynamics of their surroundings seem to deter vandalism, particularly in central areas. Despite the limitations in comparing offense data between countries [32], future research should focus on identifying trends of offenses in urban areas across countries and search for local factors that may be responsible for particular patterns of crime.

Much work still remains to be done in solving problems of data quality. There is a general consensus that underreporting and lack of systematization of procedures when recording and geocoding offenses are still the common problems when mapping and assessing the geography of crime. There is also a need to make users (e.g., police officers, experts, planners) aware of what can (or cannot) be done using spatial statistics and GIS. The combination of techniques and different data sources in GIS provides an approach for providing a better knowledge base for decision-making. This chapter illustrates the importance of employing a variety of techniques and experimenting different criteria in each method to crosscheck the spatial patterns that seem to exist at the first attempt. The user's knowledge and experience is therefore important in the process of making sense of different outputs.

ACKNOWLEDGMENTS

This research was undertaken while Vania Ceccato was a visiting fellow in the Department of Geography at the University of Cambridge, England. The support of the Marie Curie Fellowship Scheme (Grant reference HPMF-CT-2001-01307) and STINT — The Swedish Foundation for International Cooperation in Research and Higher Education (Dnr PD2001-1045) are gratefully acknowledged by the author. The author would also like to express her thanks to the municipality of Malmö, Länsförsäkringar Skåne, the Skåne Police Authority, and Copenhagen Police Authority for providing the data set used in this analysis.

REFERENCES

1. Hirschfield, A., Brown, P., and Todd, P., GIS and the analysis pf spatially-referenced crime data: experiences in Merseyside, UK. *Int. J. Geogr. Inf. Syst.,* 9, 191–210, 1995.
2. Craglia, M., Haining, R., and Wiles, P., A comparative evaluation of approaches to urban crime patterns, *Urban Stud.,* 37, 711–729, 2000.
3. Messner, S.F., Anselin, L., Baller, R., Hawkins, D.F., and Tolnay, S.E., The spatial patterns of homicide rates: an application of exploratory data analysis. *J. Quant. Criminol.,* 15, 423–450, 1999.
4. Ceccato, V., Haining, R., and Signoretta, P., Exploring offence statistics in Stockholm city using spatial analysis tools, *Ann. Assoc. Am. Geogr.,* 92, 29–51, 2002.
5. Galeotti, M., Cross-border crime in the former Soviet Union. International Boundaries Research Unit (IBRU) Boundary and territory briefing, Durham, UK, 1995.
6. Ceccato, V. and Haining R., Crime in border regions: The Scandinavian case of Öresund, 1998–2001. *Ann. Assoc. Am. Geogr.*, in press.
7. Berg, M., En tyckande bomb: invandring och enklaver. *Veckans affärer (special),* 28–30, 2003.
8. Bowers, K.J. and Hirschfield, A., Introduction, in *Mapping and Analysing Crime Data: Lessons from Research and Practice,* Hirschfield., A., and Bowers, K.J., Taylor and Francis, London, 2001.
9. Farr, M., How does crime reporting, crime levels and crime mix vary by type of neighbourhood? A national perspective, presented at Understanding crime and the neighbourhood: concepts, evidence and police design, Bristol, 2003.

AU: Update?

10. Ceccato, V. and Haining, R., Collective resources and vandalism: evidence from a Swedish city, *Urban Stud.*, submitted. AU: Update?
11. Rattcliffe, J.H. and McCullagh, M.J., Aoristic crime analysis. *Int. J. Geogr.Inf. Sci.,* 12, 751–764, 1998.
12. Rattcliffe, J.H. and McCullagh, M.J., Crime, repeat victimisation and GIS, in *Mapping and Analysing Crime Data: Lessons from Research and Practice,* Hirschfield., A., and Bowers, K.J., Eds., Taylor and Francis, London, 2001.
13. Ceccato, V. and Snickars, F., Adapting GIS technology to needs of local planning. *Environ. Plann. B Plann. Design,* 27, 923–937, 2000.
14. Haining, R., *Spatial Data Analysis: Theory and Practice,* Cambridge University Press, Cambridge, 2003.
15. Rich, T., *The use of computerised mapping in crime control and prevention programs.* NIJ Research in Action, 1995, (available from http://www.ncjrs.org/txtfiles/riamap.txt).
16. Everitt, B., *Cluster Analysis,* John Wiley & Sons, New York, 1974.
17. Levine, N., "Hot Spot" Analysis I. in *Crime Stat: Software Guide,* Levine, N., Ed., 2002.
18. Levine, N., "Hot spot" Analysis II. in *Crime Stat: Software Guide,* Levine, N., Ed., 2002.
19. Murray, A.T. and Grubesic, T.H., Identifying non-hierarchical spatial clusters. *Int. J. Ind. Eng.,* 9, 86–95, 2002.
20. Kulldorff. M.A., A spatial scan statistic. Commun. Stat. Theory Methods, 26, 1481–1496, 1997.
21. Haining, R. and Cliff, A., Using a Scan statistic to map the incidence of an infectious disease: measles in the USA 1962–1995. *Proceeding of GEOMED Conference, 2001, Inserm, Paris.*
22. Hjalmars, U., Kulldorff, M., Wahlquist, Y. and Lannering, B., Increased incidence rates but no space-time clustering of childhood malignant brain tumors in Sweden. *Cancer,* 1995.

AU: Date, etc.? (Cancer, 85, 2077-2090, 1999?)

23. Wikström, P.O., *Urban Crime, Criminals, and Victims: the Swedish Experience in an Anglo-American Comparative Perspective.* Springer-Verlag, Stockholm, 1991.
24. Park, R.E. and Burgess, E.W., *The City,* University of Chicago Press, Chicago, 1925.
25. Shaw, C.R. and McKay, H.D., *Juvenile Delinquency and Urban Areas.* University of Chicago, Chicago, Press, 1942.
26. Skogan, W.G., *Disorder and Decline: Crime and the Spiral of Decay in American Neighbourhoods,* Free Press, New York, 1990.
27. Perkins, D.D., Meeks, J.W., and Taylor, R.B., The physical environment of street blocks and resident perceptions of crime and disorder: implications for theory and measurement, *J. Envir. Psychol.,* 12, 21–34, 1992.
28. Kelling, G.L. and Coles, C., *Fixing Broken Windows: Restoring Order and Reducing Crime in Our Communities*, Free Press, New York 1996.
29. Sampson, R.J. and Groves, W.B., Community structures and crime: testing social disorganization theory, *Am. J. Sociol.*, 97, 774–802, 1989.
30. Anselin, L., SpaceStat tutorial a book for using SpaceStat in the analysis of spatial data, West Virginia University, Morgantown, 1992.
31. Martin, D., Spatial patterns in residential burglary: assessing the effect of neighbourhood social capital, *J. Contemp. Criminal Justice,* 18, 32–146, 2002.
32. Aebi, M., Barclay, G., Jehle, J., and Killias, M., *European Sourcebook of Crime and Criminal Justice Statistics: Key Findings:* Council of Europe. 1999.