

Chapter 5

Exploring Spatial Patterns of Crime Using Non-hierarchical Cluster Analysis

Alan T. Murray and Tony H. Grubestic

Abstract Exploratory spatial data analysis (ESDA) is a useful approach for detecting patterns of criminal activity. ESDA includes a number of quantitative techniques and statistical methods that are helpful for identifying significant clusters of crime, commonly referred to as hot spots. Perhaps the most popular hot spot detection methods, both in research and practice, are based on tests of spatial autocorrelation and kernel density. Non-hierarchical clustering methods, such as k-means, are less used in many contexts. There is a perception that these approaches are less definitive. This chapter reviews non-hierarchical cluster analysis for crime hot spot detection. We detail alternative non-hierarchical approaches for spatial clustering that can incorporate both event attributes and neighborhood characteristics (i.e., spatial lag) as a modeling parameter. Analysis of violent crime in the city of Lima, Ohio is presented to illustrate this for hot spot detection. We conclude with a discussion of practical considerations in identifying hot spots.

Keywords Clustering • Hot spots • Spatial patterns

A.T. Murray (✉)

GeoDa Center for Geospatial Analysis and Computation, School of Geographical Sciences and Urban Planning, Arizona State University,

Tempe, AZ 85287-5302, USA

e-mail: atmurray@asu.edu

T.H. Grubestic

Geographic Information Systems and Spatial Analysis Laboratory,
College of Information Science and Technology, Drexel University,

Philadelphia, PA 19104, USA

e-mail: grubestic@drexel.edu

5.1 Introduction

Cluster detection and hot spot mapping in criminology, geography and related socio-economic planning sciences has evolved significantly over the past decade (Eck et al. 2005; Chainey et al. 2008). While many of the most basic approaches remain popular, such as spatial autocorrelation, spatial ellipses, kernel density estimation and spatial scan statistics (Wang 2005; Eck et al. 2005; Kent and Leitner 2007; Chainey et al. 2008; Rogerson and Yamada 2009; Anselin et al. 2009), advanced approaches now include fuzzy clustering (Grubestic 2006), spatio-temporal modeling of crime (Ratcliffe 2002; Grubestic and Mack 2008; Leitner et al. 2011), geospatial visual analytics (Anselin and Kochinsky 2010), and agent-based simulation (Eck and Liu 2008). Further, the emergence of proactive policing, predictive hot spotting and crime forecasting strategies suggests a growing need for objective spatial pattern detection methods to establish a better understanding of the distributions and morphologies crime (Cohen et al. 2004; Gorr et al. 2003; Johnson and Bowers 2004; Wu and Grubestic 2010).

Broadly defined, a crime hot spot represents a grouping of incidents that are spatially and/or temporally clustered (Harries 1999; Eck et al. 2005; Grubestic 2006). The genesis of crime hot spots is often linked to environmental factors (Brantingham and Brantingham 1981), social disorganization (Shaw and McKay 1942; Sampson and Groves 1989; Morenoff et al. 2001) and opportunity (Cohen and Felson 1979). Regardless of the underlying factors that fuel the emergence of hot spots, law enforcement agencies recognize the importance (and benefits) of detection and intervention in these problematic areas (Harries 1999; Braga 2001; Ratcliffe 2004). However, the ability to identify hot spots is highly dependent on the capability to detect patterns, and this requires the selection of appropriate techniques for carrying out hot spot analyses. Such pattern detection is typically viewed as exploratory spatial data analysis (ESDA) (Murray and Estivill-Castro 1998; Anselin 1998; Wu and Grubestic 2010), but can be confirmatory in some contexts.

At the intersection of ESDA, GIS, and crime analysis is the use of ESDA for identifying significant patterns of criminal activity (Harries 1999; Anselin et al. 2000; Murray et al. 2001). Again, while local indicators of spatial association (Messner et al. 1999; Anselin et al. 2000, 2009) and kernel density mapping (McLafferty et al. 2000) are popular approaches for identifying hot spots, alternative techniques such as cluster analysis are less utilized in practice. Grubestic (2006) notes that there are three major problems associated with applying cluster analysis for crime hot spot detection:

1. The choice between non-hierarchical and hierarchical methods can be confusing¹;
2. There are problems regarding the manner in which some techniques treat geographic space (e.g., spatial bias);
3. There is relatively little guidance for determining the appropriate number of clusters in a study area.

¹ A discussion of hierarchical and non-hierarchical methods can be found in Hartigan (1975), Everitt (1993) and Kaufman and Rousseeuw (2005), among others. Non-hierarchical, or partitioning, approaches identify a pre-specified number of clusters, k , such that each object is a member of exactly one cluster, where membership similarity is optimized. In contrast, hierarchical methods

While these challenges can be daunting, non-hierarchical cluster analysis is potentially useful for finding crime hot spots, reflected by its inclusion in the National Institute of Justice sponsored and supported crime analysis tool, CrimeStat (Levine 2010).

The non-hierarchical technique implemented in CrimeStat (version 3.3) is the k -means approach proposed by Fisher (1958). The k -means technique is based upon multivariate analysis of variance in the evaluation of homogeneity among entities (Estivill-Castro and Murray 2000). Specifically, the scatter matrix of similarity between entities may be evaluated by its trace (Aldenderfer and Blashfield 1984), and homogeneity is then measured for a grouping of events using the sum of squares loss function (Rousseeuw and Leroy 1987). The benefits of using k -means lie in its ability to handle extremely large numbers of observations and still generate clusters relatively quickly, although this is contingent on the number of iterations selected for the routine.

Other non-hierarchical clustering approaches have been developed and utilized. Some are detailed in Kaufman and Rousseeuw (2005). In the context of geographic applications, a review of approaches is given in Murray and Estivill-Castro (1998), Murray (2000a, b) and Grubestic (2006). Clearly, if one is intent on identifying crime hot spots that are strongly related in some predefined sense (e.g. crime type), then multiple non-hierarchical clustering techniques may be useful. This is a subtle but important point. If an analyst is able to choose from a suite of alternative clustering approaches, a clearer picture of the spatial morphology of crime may emerge. However, it is also possible that the selection of an inappropriate technique may skew the identification and interpretation of crime hot spots, minimizing the usefulness of the approach. This is particularly true where non-hierarchical approaches are concerned because many analysts may not be aware of the biases and inaccuracies associated with a particular approach. Simply put, all clustering methods are not equivalent. Unfortunately, the overall body of research focusing on the subtle differences in the use and application of non-hierarchical techniques for geographic applications is rather limited (Murray 1999, 2000a; Murray and Grubestic 2002; Grubestic 2006). Empirical results suggest that substantial variation exists in the structure and quality of clusters, depending on the approach.

The purpose of this chapter is to review clustering approaches for identifying spatial patterns of crime, focusing on the basic tenets of crime mapping and analysis from a geographic perspective. This is followed by an examination of the statistical foundations of non-hierarchical cluster analysis, highlighting the strengths and weaknesses of the most widely utilized approaches. Section 5.4 introduces alternative approaches for non-hierarchical cluster analysis that incorporate additional geographic context through the use of spatial lags. Application results examine violent crime in Lima, Ohio. We conclude with a brief discussion and final remarks.

build clusters based on agglomeration (e.g., begin with n clusters and merge the two most similar groups to get $n-1$ clusters) or division (e.g., begin with one cluster and divide it into two most similar clusters), creating a decomposition hierarchy of clusters ranging from n to 1.

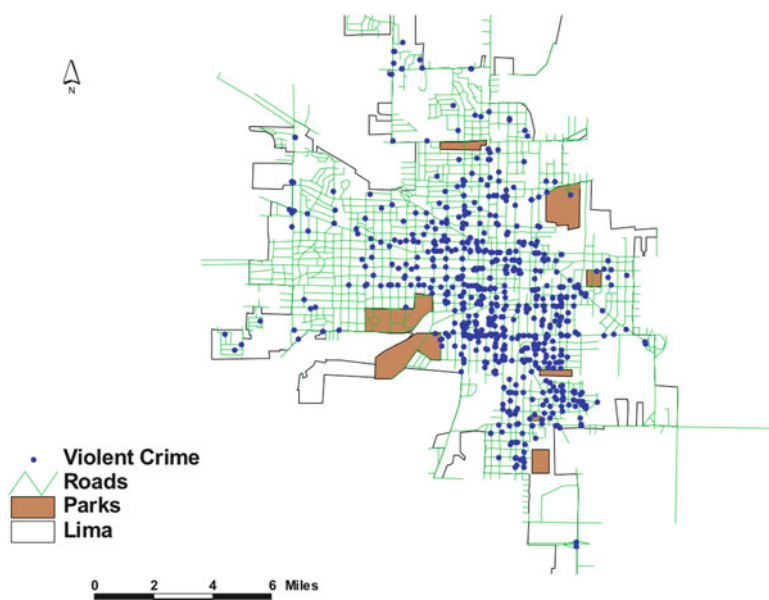


Fig. 5.1 Violent Crime in Lima, Ohio

5.2 Spatial Patterns of Crime

Identifying significant geographic relationships in the occurrence of criminal activity is, perhaps, the most fundamental component of crime mapping and analysis. Of course, the process is complicated by a vast array of techniques and methods available to analysts. In many instances, the first step in developing a better understanding of crime distributions and their contributing factors is to generate a map. This might involve plotting incident locations, differentiating them by crime type and adding topographic information for additional spatial context. For example, Fig. 5.1 illustrates 848 violent crimes (homicide, rape, robbery and assault) in the city of Lima, Ohio.² Alternatively, if the crime information is only recorded at a more aggregate level, such as census block groups, then a choropleth map of total crime or crime rates for a geographic area can be created. At this level of geographic detail, broader patterns of neighborhood distress and spatial inequity may become apparent. For instance, Fig. 5.2 depicts violent crime rates in Lima using a choropleth display of block group crime rates per 1,000 people. Ignoring the overlaid ellipses for the moment, this display emphasizes differences in the attribute of interest using seven unique classes. As with any choropleth display, the goal is to

² Lima, Ohio is a city of approximately 38,000 people and is located about 70 miles north of Dayton on the Interstate 75 corridor.

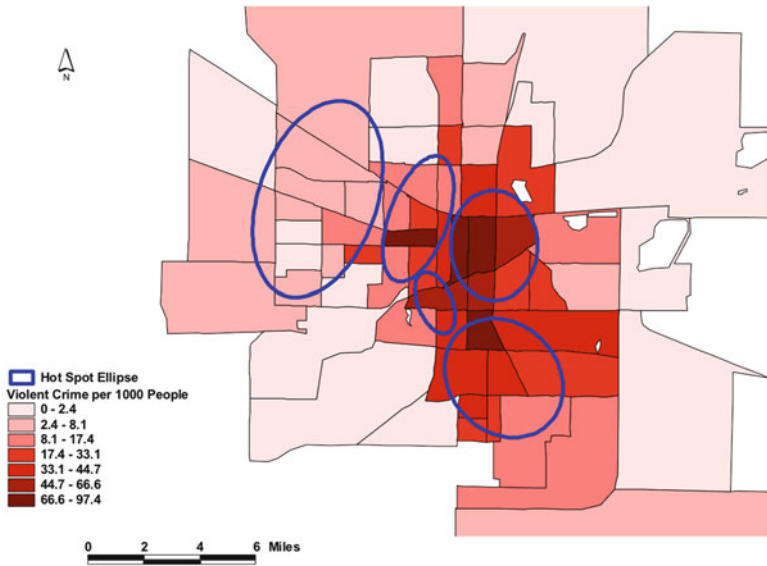


Fig. 5.2 Crime rates by block group in Lima

effectively show spatial variation in the variable's distribution. Creation of a traditional choropleth map involves deciding where to establish the class break/cutoff values (Dent 1999; Murray and Shyy 2000). In Fig. 5.2, class breaks of 2.4, 8.1, 17.4, 33.1, 44.7 and 66.6 (shown in the legend) are used, derived using the natural breaks options in ArcGIS. This classification helps communicate how violent crime rates vary spatially in Lima, but does so in a much different way than the point map displayed in Fig. 5.1.

Perhaps the most intriguing aspect of crime mapping and analysis is the subtle methodological overlap of choropleth mapping approaches, non-hierarchical cluster analysis and hot spot detection techniques. Choropleth mapping is an area of cartography and GIS that has received considerable interest over the past 50 years (Murray and Shyy 2000; Armstrong et al. 2003; Xiao and Armstrong 2005; Cromley and Cromley 2009). Numerous choropleth mapping approaches have been developed, most of which are accessible and readily available in commercial GIS and cartography software. As noted, the display shown in Fig. 5.2 was generated using the natural breaks option in ArcGIS (version 10.3), an approach that is also available in TransCad, MapInfo, Maptitude and many other GIS packages. Natural breaks is widely considered the standard/default choropleth mapping method. In brief, the natural breaks approach attempts to minimize the sum of variance in created classes (Dent 1999). This is identical to the goal of non-hierarchical clustering, such as k-means, a sum of squares approach.

By analyzing either Fig. 5.1 or Fig. 5.2, analysts could make inferences about the spatial distribution, and perhaps the potential impact, of violent crime in Lima. Clearly,

the intent of crime analysis is that such displays are helpful for understanding crime trends and patterns so that appropriate law enforcement action can be prescribed.

The next step would typically involve assessment of spatial autocorrelation, at least for aggregate crime rates such as the block groups in Fig. 5.2, as this would help confirm whether clustering is occurring. Packages like as CrimeStat, GeoDa (Anselin et al. 2006) and ArcGIS allow analysts to derive such measures. In this instance, we find that Moran's I is 0.710 with a standard normal z-value of 11.43 ($p=0$), indicating spatial clustering of violent crime in Lima. Unfortunately, global metrics do not pinpoint where this clustering is taking place. As a result, if an analyst is interested in determining where hot spots exist, additional analysis is necessary. In many cases, local spatial statistics and non-hierarchical clustering approaches are advocated for identifying and assessing potential hot spots (Anselin 1995; Harries 1999; Messner et al. 1999; Levine 2006; Ratcliffe 2005). These approaches are typically coupled with standard deviation ellipses in an effort to represent the co-variation within a cluster group about the major and minor axes.

The ellipses associated with the k -means generated clusters using CrimeStat (version 3.3) are also shown in Fig. 5.2. Fundamentally, this shows the integration of non-spatial *and* spatial grouping processes. The ellipses represent the spatial grouping of the associated areas, whereas the choropleth classes reflect attribute (violent crime rate) variation. Furthermore, it is worth reiterating that the ellipses were generated in CrimeStat from spatial clusters identified using a k -means heuristic, although alternative options for summarizing distributions are also available. As noted previously, this is all the more interesting because the natural breaks choropleth classes shown in Fig. 5.2 are also identified using equivalent criteria.

There are a number of questions arising from this brief review on spatial aspects of crime hot spot detection. Is the sum of squares clustering approach and its most popular solution technique (k -means) viable for spatial data? If not, why? Are there feasible alternatives to these approaches that can either complement or improve upon the results generated through traditional solution techniques? In an effort to address these questions, the next section outlines the fundamental nature of non-hierarchical clustering, with a focus on the sum of squares approach.

5.3 Statistical Clustering

As noted previously, cluster analysis is a popular approach for developing classification systems and taxonomies. A simple search on the Social Sciences Citation Index reveals that nearly 130,098 entries have referenced "cluster analysis" since 1980, equating to approximately 6,195 per year (1980–2011). In crime analysis, as in other problem domains, the sum of squares variance minimization approach continues to be the dominant non-hierarchical partitioning technique (Levine 2010). In fact, most commercial statistical packages, including SPSS,

S-Plus, SAS, Stata and NCSS, provide capabilities for carrying out cluster analysis using the sum of squares approach (Murray and Grubestic 2002; Grubestic 2006). Consider the following notation:

$$\begin{aligned}
 i &= \text{index of entities;} \\
 k &= \text{index of clusters;} \\
 p &= \text{total number of clusters;} \\
 f_i &= \text{attribute measure;} \\
 d_{ik} &= \text{measure of proximity between entity } i \text{ and cluster } k; \\
 z_{ik} &= \begin{cases} 1 & \text{if entity } i \text{ is in cluster } k \\ 0 & \text{otherwise.} \end{cases}
 \end{aligned}$$

Where crime analysis is concerned, entities correspond to the location of a crime(s). The variable f_i indicates the number of crimes occurring at a particular location i . If there is a need to attribute a measure of importance to particular crime types (e.g. severity), it is possible to extend the specification of f_i to reflect such differentiation.³ The sum of squares approach is as follows:

Sum of Squares Clustering Model (SSCM)

$$\text{Minimize } \sum_i \sum_{k=1}^p f_i d_{ik}^2 z_{ik} \quad (5.1)$$

Subject to:

$$\sum_{k=1}^p z_{ik} = 1 \quad \forall i \quad (5.2)$$

$$z_{ik} = (0,1) \quad \forall i,k \quad (5.3)$$

The objective (5.1) of the SSCM is to minimize the total weighted squared difference in cluster group membership. This is equivalent to minimizing the within group sum of squares (Hartigan 1975; Kaufman and Rousseeuw 2005). Constraint (5.2) ensures that each entity is assigned to a group and Constraint (5.3) imposes integer restrictions on the decision variables.

The formulation of the sum of squares clustering model illustrates that this is an optimization problem. The overall goal of the SSCM is to identify the best, or optimal, partition of entities. One approach for solving the SSCM is the k -means heuristic developed by Fisher (1958) and MacQueen (1967), when Euclidean

³ Details on multivariate integration for such purposes may be found in Kaufman and Rousseeuw (2005) as well as other clustering texts.

distance is considered. In vector quantization, this heuristic is also known as the generalized Lloyd algorithm (Estivill-Castro and Murray 2000). This optimization problem is recognized as being inherently difficult to solve optimally, so the application of heuristic techniques such as the k -means approach are considered a good option for obtaining a solution. The k -means heuristic has four main steps (Murray and Grubestic 2002):

1. generate p initial clusters
2. compute the center of each cluster
3. assign each entity to its closest cluster
4. if groupings have changed in step 3, return to step 2. If not, a local optima has been found.

A notable feature of the SSCM is that the center of each grouping is a centroid, reflecting the squared Euclidean proximity measure in the objective (5.1). In addition, the k -means heuristic is a popular approach for solving the SSCM for a number of reasons. First, it is statistically grounded and widely available in most commercial statistical software packages (Murray and Grubestic 2002). Second, it has the ability to handle relatively large data sets (Huang 1998). Third, it converges quickly to find a local optima (Murray and Grubestic 2002).

While these advantages are certainly appealing and have contributed to its widespread application, including the NIJ supported CrimeStat software package, there are questions pertaining to the appropriateness of the SSCM when applied to geographic data (Murray and Grubestic 2002; Grubestic 2006). Although many of the biases inherent in the SSCM are widely noted (see Murray and Estivill-Castro 1998; Kaufman and Rousseeuw 2005; among others), the SSCM continues to be relied upon in geographic and non-geographic inquiry.

What is wrong with the sum of squares approach, particularly with respect to the spatial analysis of urban crime? One major issue is the sub-optimality associated with the use of the k -means heuristic in solving the SSCM. Often, implementation of this heuristic provides analysts a solution based on one instance. In order for the k -means heuristic to be effective for solving the SSCM, it must be re-started hundreds or thousands of times (depending on problem size), using a different initial clustering in step 1 for each instance.⁴ Standard practice, however, has been to use only one initial starting configuration. The result is that the identified cluster solutions are likely sub-optimal, which means that they may be of limited use for inferential analysis and policy making. The extent to which sub-optimality was an issue was examined in Murray and Grubestic (2002), who found that non-optimal solutions were generally identified using major statistical packages such as SPSS, S-Plus and SAS. In some instances, SSCM solutions were found to deviate more than 30% from the optimal solution, which means that subsequent analysis is being

⁴ It is well known that any single application of the k -means heuristic is susceptible to becoming trapped in a local optima (Estivill-Castro and Murray 2000; Murray and Grubestic 2002), which prohibits the approach from identifying an optimal solution.

conducted on clusters that are not most similar. Further, limited testing of CrimeStat found instances where the identified solutions deviated more than 72% from the optimal solution.⁵

A second and more significant problem with the SSCM is that spatial clusters are biased by outliers. Although this bias is discussed by Kaufman and Rousseeuw (2005) and others, Murray and Grubestic (2002) demonstrated the influence of this bias using spatial information rather than non-spatial data. The SSCM is biased because of the use of the squared Euclidean distance measure in objective (5.1). The result in application is that outliers, or more distant events from others, have greater influence on the structure of the identified clusters, effectively distorting potential hot spots. One option is to identify and remove outliers using the approaches detailed in Messner et al. (1999) and Grubestic (2006). Alternatively, it may be preferable to utilize a modeling approach that does *not* spatially bias clusters.

Though not an issue with the SSCM generally, Murray and Grubestic (2002) note that most software packages do not provide the capability to include a f_i value in objective (5.1), rather this is assumed to equal 1.⁶ Given this, it makes sense that statistical packages like CrimeStat would attempt to summarize k -means generated clustering results using standard deviation ellipses, because the clusters are identified on the basis of space alone.

Finally, the SSCM does not explicitly address attribute similarity, but rather focuses on spatial proximity. Integration of the choropleth display with the ellipses in Fig. 5.2 is an interesting approach for examining spatial and non-spatial patterning in this regard, but lacks direct examination of both issues. Murray and Shyy (2000) present a clustering based approach for choropleth mapping that considers attribute and spatial similarity simultaneously. Murray (2000b) details a spatial lag approach to integrate attribute and spatial proximity.

5.4 Spatial Lag in Cluster Analysis

Geographic analysis using spatial statistical techniques has been significantly enhanced when more is known about what is taking place near a particular entity of interest. The reason this has been the case is that the assumption of independence

⁵ Analysis was carried out using 114 crime events in a neighborhood located in Akron, Ohio. The number of clusters obtained ranged from 4 to 11 ($p=4-11$). A separation value of 4 was utilized in the application of the k -means solution technique in CrimeStat for each value of p and the identified solution compared with the “optimal” solution using the approach reported in Murray and Grubestic (2002). For this range of clusters, the average sub-optimality of CrimeStat solutions was 38.28% (min=12.01%; max=72.19%). It should be noted that one can alter the separation distance in CrimeStat, in essence representing a pseudo-restart of the heuristic. Unfortunately, it is not possible to compare or assess cluster solution quality.

⁶ An assumed value of $f_i = 1$ implies the occurrence of a single event, rather than reflecting the aggregate summary of areas like police beats, census blocks or alternative areal units.

between entities in statistical testing is known to be problematic for spatial data as the existence of spatial autocorrelation can alter significance levels and reduce interpretative capabilities (Griffith and Amrhein 1997). One approach for dealing with spatial autocorrelation involves the use of a spatial lag. A spatial lag represents an averaging process of an entity’s neighbors. In most cases, neighbors represent other entities or areas next to a particularly entity. As a point of reference, consider the following notation:

- i (and j)= index of entities;
- l_i = spatial lag for entity i ;
- Ω_i = spatial neighbors of entity i

Neighbors are often defined as those entities sharing a common border or point and do not include the entity itself.⁷ Using this notation, the spatial lag for entity i may be defined as follows:

$$I_i = \frac{\sum_{j \in \Omega_i} f_j}{|\Omega_i|} \tag{5.4}$$

Spatial lag enables one to summarize what is taking place in a neighborhood around a particular area. For example, one can compute the average number of crime events occurring in neighborhoods that are adjacent to a neighborhood of interest. This is an indirect spatial proximity metric. The integration of both space and attribute values is relatively straightforward:

$$\delta_{ik} = \sqrt{(f_i - \bar{f}_k)^2 + (l_i - \bar{l}_k)^2} \tag{5.5}$$

where \bar{f}_k represents the average attribute value for cluster k and \bar{l}_k indicates the average lag value for cluster k . With this, (5.5) represents an integration of attribute similarity with an indirect spatial proximity metric. Murray (2000b) introduced an alternative clustering model based on this:

Spatial Lag Cluster Model – Center 1 (SLCM-C1)

$$\text{Minimize } \sum_i \sum_{k=1}^p \delta_{ik} z_{ik} \tag{5.6}$$

Subject to:
 (5.2)–(5.3)

Although the constraints for this model are the same as those in the SSCM, the objective of SLCM-C1 is much different. Objective (5.6) minimizes the total

⁷It is also common to view neighbors as being within a specified distance of a given location.

dissimilarity in selected clusters. This differs in three ways from objective (5.1) for the SSCM. First, there is no attribute (f_i) weighting. Second, there is no explicit representation of distance in (5.6) as there is in (5.1). Finally, the similarity measure, δ_{ik} , is not squared in (5.6), whereas it is in (5.1). The implication of this is that the cluster centers in SLCM-C1 are *not* centroids, in contrast to the SSCM. This general representational distinction is a subtle but exceptionally important point. Simply put, by avoiding the use of a centroid in (5.6), the biasing influence of outliers in the SLCM-C1 is minimized. That said, there are tradeoffs with this type of formulation; namely, solving the SLCM-C1 remains challenging due to its implied non-linear form. As a result, the alternating heuristic has generally been relied upon for solving the SLCM-C1 (Murray 2000b).

Clearly, one drawback of the SLCM-C1 is the inability to alter the importance of either attribute or spatial lag influence in the identification of clusters. The SLCM-C1 treats attribute and lag with equal importance. However, this may not necessarily be appropriate for exploratory analysis. For example, one might want to investigate the clusters associated with maximizing attribute similarity only (somewhat equivalent to classes created in choropleth maps using the natural breaks approach). Alternatively, one might wish to view the clusters where lag similarity is optimized. Given these two extremes, it is also possible that one might want to examine the clusters associated with slightly more importance on attribute similarity than lag – or something else in between. Unfortunately, it is not possible to structure the relative importance of variables using the SLCM-C1. In an effort to provide more flexibility, Murray (2000b) presented a modified interpretation of similarity:

$$a_{ik} = |f_i - \bar{f}_k| \quad (5.7)$$

$$s_{ik} = |l_i - \bar{l}_k| \quad (5.8)$$

Essentially, these measures track the similarity structured in (5.5), but do so separately. With this modified representation, it is now possible to alter how much significance the individual components have in structuring clusters. Incorporating them independently into a non-hierarchical clustering model may be accomplished by assigning weights to both attributes and lag:

w_a = weight for attribute similarity
 w_s = weight for spatial lag similarity

Murray (2000b) derived a variant of the SLCM as follows:

Spatial Lag Clustering Model – Center 2 (SLCM-C2)

$$\text{Minimize } w_a \sum_i \sum_{k=1}^p a_{ik} z_{ik} + w_s \sum_i \sum_{k=1}^p s_{ik} z_{ik} \quad (5.9)$$

Subject to:

(5.2)-(5.3)

Objective (5.9) of the SLCM-C2 maximizes the total weighted attribute similarity and maximizes the total weighted spatial lag similarity in selected clusters. In this revised form, (5.9) is now a multi-objective optimization problem that may be used to identify a range of non-dominated clustering solutions (Cohon 1978), each potentially valuable in identifying crime hot spots. Unfortunately, the SLCM-C2 remains a difficult optimization problem to solve optimally, so a heuristic is necessary (Murray 2000b).

Finally, it is also possible to view the above lag models from the more traditional median perspective. Murray (2000b) proposed a multi-objective median based clustering model incorporating spatial lag. Using a median based approach, similarity may be defined as follows:

$$\hat{a}_{ik} = |f_i - \bar{f}_k| \quad (5.10)$$

$$\hat{s}_{ik} = |l_i - \bar{l}_k| \quad (5.11)$$

where j is the index of potential medians (same as the index i). This approach enables similarity to be defined a priori between entities, rather than being a function of identified clusters. In order to present the median clustering model, additional decision variables must first be defined:

$$x_j = \begin{cases} 1 & \text{if median } j \text{ is selected to facilitate cluster creation} \\ 0 & \text{otherwise} \end{cases}$$

$$y_{ij} = \begin{cases} 1 & \text{if entity } i \text{ is in cluster } j \\ 0 & \text{otherwise.} \end{cases}$$

With the above notation, it is possible to structure a median-based non-hierarchical clustering model with objectives for maximizing both attribute and spatial lag homogeneity.

Spatial Lag Clustering Model – Median (SLCM-M)

$$\text{Minimize } w_a \sum_i \sum_j \hat{a}_{ij} y_{ij} + w_s \sum_i \sum_j \hat{s}_{ij} y_{ij} \quad (5.12)$$

Subject to:

$$\sum_j y_{ij} = 1 \quad \forall i \quad (5.13)$$

$$\sum_j x_j = p \quad (5.14)$$

$$y_{ij} \leq x_j \quad \forall i, j \quad (5.15)$$

$$y_{ij} = (0,1) \quad \forall i, j \quad (5.16)$$

$$x_j = (0,1) \quad \forall j$$

Objective (5.12) of the SLCM-M minimizes the total weighted attribute dissimilarity and minimizes the total weighted spatial lag dissimilarity in selected clusters. This is equivalent to what is structured in objective (5.9) in the SLCM-C2. Constraint (5.13) ensures that each entity is included in a cluster. Constraints (5.14) and (5.15) require that only p clusters be generated. Constraints (5.16) impose integer restrictions on decision variables.

One of the most appealing features of the SLCM-M is that it is an integer programming problem that can be solved optimally for small and medium sized applications using commercial software and/or specialized techniques. This is a major departure from previously discussed models that rely on heuristic solution techniques and have the potential for getting “stuck” in a local optima. In addition, the multi-objective nature of this clustering model enables a number of things to be addressed. One important feature is that it simultaneously integrates both attribute similarity, as is done in choropleth mapping, and spatial proximity, as is done using standard deviational ellipses (along with the use of a k -means clustering heuristic). As with the other spatial lag models (SLCM-C1 and SLCM-C2), the SLCM-M avoids spatial bias inherent in the SSCM, but remains a within group variance minimization approach. One final feature is that the SLCM-M allows for non-dominated clustering solutions to be identified, an essential characteristic for ESDA and critically important for comparing alternative hot spot solutions.

5.5 Cluster Model Application for Hot Spot Detection

In an effort to illustrate the power and flexibility of the SLCM-M for exploratory analysis, the 62 block groups and violent crime rates for Lima, Ohio displayed in Fig. 5.2 will be used for analysis. Reported SLCM-M results are optimal to within 0.1% and the time required to solve associated problems was less than 1 s on an Intel Xeon quad core computer (2.27 GHz) with 8 gigabytes of RAM.

The first step in this exploratory analysis is deciding what number of clusters will be evaluated. Next, the associated non-inferior tradeoff curve must be generated using trial-and-error or techniques detailed in Cohon (1978). Considering that previous analyses in this chapter examined seven classes in Fig. 5.2, seven clusters will be evaluated using the SLCM-M. Figure 5.3 displays one non-dominated clustering

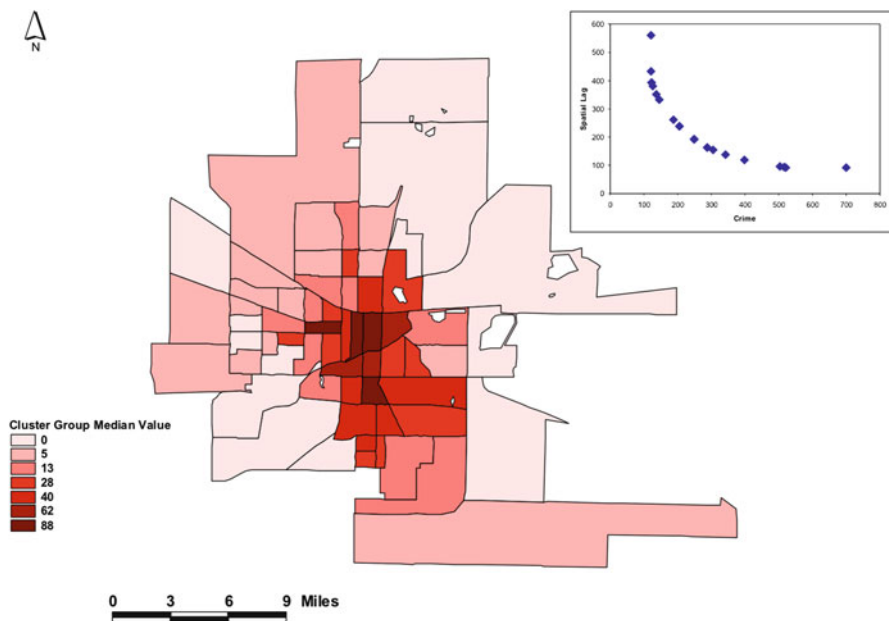


Fig. 5.3 Structured clusters using the SLCM-M ($w_a=1$, $w_s=0.01$)

solution using weights of $w_a=1$ and $w_s=0.01$.⁸ In addition, Fig. 5.3 also shows the non-inferior tradeoff curve for the range of possible solutions that may be identified by varying the weights of importance for attribute similarity and lag similarity. Thus, plotted in this tradeoff curve is the total dissimilarity of violent crime against the total dissimilarity of spatial lag for the range of identified clustering solutions. The highlighted tradeoff point (*) corresponds to the displayed clustering solution. As a result, each point on the non-inferior tradeoff curve has an associated unique spatial clustering that may be analyzed and evaluated. For example, Fig. 5.4 depicts another tradeoff solution for weights of $w_a=1$ and $w_s=0.7$, which not only represents another point on the tradeoff curve but also has a unique corresponding spatial clustering pattern. Other tradeoff solutions could be shown as well. Comparing Figs. 5.3 and 5.4 (as well as 2), one can see subtle cluster changes as the influence of spatial lag is increased. The significance of this is that different spatial patterns emerge, patterns which may be more suggestive of underlying social and environmental characteristics or conditions for a region.

⁸The legend in this case does not have the same interval interpretation as that shown in Figure 5.2. Rather than depicting interval ranges, only the median group value is shown. Once spatial lag importance is increased, it is unlikely that groups will have non-overlapping values characteristic of choropleth maps. This point is discussed in Murray and Shy (2000).

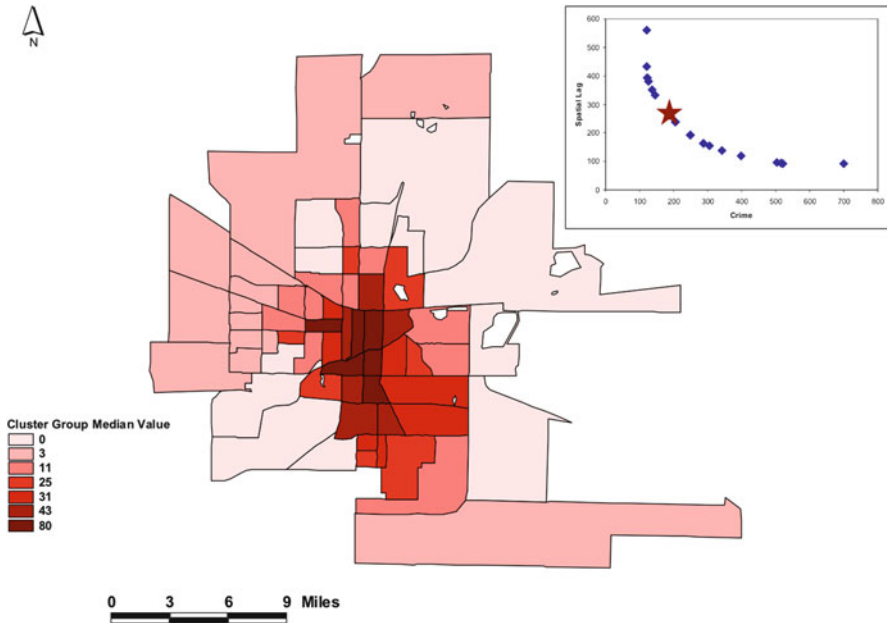


Fig. 5.4 SLCM-M clusters increasing spatial lag importance ($w_a = 1$, $w_s = 0.7$)

All of the figures suggest that there is a relative concentration of violent crime in the downtown area (center) of Lima. The highest crime rate areas in Fig. 5.3 correspond to lower income neighborhoods in the city. Further, these areas also have high minority concentrations, high unemployment, and a high percentage of households headed by single women. Thus, the choropleth displays (Figs. 5.2, 5.3, and 5.4) do a particularly good job highlighting higher violent crime rate areas and track well with the socio-economic factors likely to be influencing violent crime in Lima. Interestingly, as the weight for spatial lag is increased, the depicted geographic variation is less significant.

5.6 Discussion and Conclusion

The above analysis is insightful in many ways. There is a clear indication that downtown Lima represents one or more clusters in Figs. 5.1, 5.2, 5.3, and 5.4. However, point based displays (Fig. 5.1) are difficult to assess in a relative manner, ignoring background rates and activity. Ellipses (Fig. 5.2) are misleading, failing to adequately identify or delineate hot spot cluster. Figure 5.4, on the other hand, shows that there are actually spatial spillover effects that constitute a corridor area that is a hot spot (darkest units). This provides definitive instruction on where to allocate resources and personnel in order to combat violent crime in Lima.

There are a number of important issues associated with the detailed methods, and non-hierarchical clustering in particular. One important application issue remains identifying the appropriate number of clusters. There is actually little theoretical guidance for selecting the number of clusters to generate. In choropleth mapping, Dent (1999) suggests that 4–6 classes (clusters) should be selected (see also Harries 1999 as well with respect to crime analysis). Cromley (1995), also in the context of choropleth display, discusses the “elbow” in the curve approach. This is consistent with the rule of thumb well established in cluster analysis (Everitt 1993) as well as the economic interpretation found in location modeling (ReVelle 1987). However, this is less than definitive and certainly subjective, not unlike the criticisms of simple choropleth mapping and visual inspection (Messner et al. 1999). In the statistical literature additional methods for detecting the appropriate number of clusters have been proposed (Gordon 1996; Lozano et al. 1996; Podani 1996; Milligan and Cooper 1985). It is not clear, however, whether these alternatives might be useful in the analysis of crime. As a result, an important area for continued future research is exploring the applicability of these techniques for guiding users in the specification of the number of clusters to find.

Although there is significant flexibility and exploratory capabilities offered in the multi-objective structure and weighting in the SLCM, it does present a potential difficulty when carrying out analysis. Specifically, there is currently no theoretical basis for opting for a particular set of weights responsible for producing an associated non-dominated solution. In multi-objective modeling, the entire set of non-dominated solutions is considered potentially valuable (Cohon 1978). So, an analyst faces the question of addressing which ones are significant. This depends on external interpretation of the set of identified non-dominated solutions. It is unclear whether technical or theoretical approaches will be able to establish practical guidelines for analysts in the evaluation of alternative weightings.

One of the distinguishing features of non-hierarchical clustering is that of mutual exclusivity. In other words, entities are partitioned so that all of them are members of a cluster, but no two clusters share a common entity. As a result, the implication is that all of the identified clusters are significant. However, this is not well suited for hot spot detection in crime analysis. Rather, in hot spot detection it is recognized that crime events do and will happen, but it is when they localize and/or concentrate in some manner that a sub-area becomes a significant concern. This alternative interpretation of produced partitions leaves analysts to infer cluster significance using their own judgment. Given that hot spots represent areas in need of attention, this is obviously problematic. Potential approaches for addressing this issue may be found in the work of Arnold (1979) and Milligan and Mahajan (1980), which suggests Monte Carlo tests for examining partition validity and significance.

Aside from the detection of crime hot spots, the delineation of activity clusters does have a broader use. Clusters and their associated center locations may be important for finding criminals. In particular, the center of a cluster may correspond to where a perpetrator of certain crimes lives/works or where the next crime event may occur (LeBeau 1987; Rossmo 2000). Thus, the nature of the cluster (grouping of entities) and its subsequent interpretation (location of centers) is very spatial.

This suggests similarities with location modeling approaches, such as those discussed in ReVelle (1987) and Murray and Estivill-Castro (1998). More research is needed to establish the significance of cluster centers for this purpose as well as what interpretation of the “center” is most appropriate (e.g. mean, median).

A final point in the application of clustering models is the influence of scale variation. As an example, do clustering approaches produce equivalent results when using point based information as opposed to the use of area based aggregations of point information? In spatial analysis this line of inquiry is referred to as the modifiable areal unit problem (MAUP). Openshaw and Taylor (1981) note the possibility that analytical results may be altered by varying scale or modifying the boundaries of the reporting units. Criminology research has long been aware of scale and aggregation issues, and their implications in analysis (Parker 1985). Often times crime event locations are not made accessible for detailed analysis, making this concern a non-issue. However, when individual locations do exist, it is reasonable that clustering using these events be carried out. Another aspect of this issue is that a hot spot may exist in different ways and at different levels of spatial scale, as noted in Harries (1999) and Eck et al. (2005). At the individual crime incident level, hot spots may run along a particular street segment or route, rather than being circular (centered on a point) or elliptical. In such cases utilizing clustering models as currently specified may be problematic. Recent research has begun to deal with these spatial patterning issues (Yamada and Thill 2007; Shiode and Shiode 2009). Research examining scale and unit definition differences as well as patterning in clustering analysis is much needed.

This chapter has examined the statistical orientation of non-hierarchical clustering for assessing patterns of crime. Extensions and new approaches for this assessment were also reviewed and introduced. The use of spatial lag was shown to be an interesting way to incorporate geographic relationships and likely represents a promising avenue for relating non-hierarchical clustering to local spatial statistics. There are clearly unique and challenging aspects to the use of non-hierarchical clustering for identifying patterns of crime. Research examining these issues is necessary if clustering is to be effective tool in the exploratory analysis of crime activity.

Acknowledgement This material is based upon work supported by the National Science Foundation under grants SES-1154316 and SES-1154324. Any opinions, findings, conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

- Aldenderfer M, Blashfield R (1984) Cluster analysis. Sage Publications, Beverly Hills
- Anselin L (1995) Local indicators of spatial association-LISA. *Geogr Anal* 27:93–115
- Anselin L (1998) Exploratory spatial data analysis in a geocomputational environment. In: Longley PA, Brooks SM, McDonnell R, Macmillan B (eds) *Geocomputation, a primer*. Wiley, New York

- Anselin L, Cohen J, Cook D, Gorr W, Tita G (2000) Spatial analysis of crime. In: Criminal justice volume 4. Measurement and analysis of crime and justice. National Institute of Justice, Washington, DC
- Anselin L, Syabri I, Kho Y (2006) Geoda: an introduction to spatial data analysis. *Geogr Anal* 38:5–22
- Anselin L, Meyer WD, Whalley LA, Savoie MJ (2009) Actionable cultural understanding for support to tactical operations (ACUSTO). US army corps of engineers. ERDC/CERL TR-09-13
- Anselin L, Rey SJ, Kochinsky J (2010) Flexible geospatial visual analytics and simulation technologies to enhance criminal justice decision support systems. *Crime Mapp* 3(1).
- Armstrong MP, Xiao N, Bennett DA (2003) Using genetic algorithms to create multicriteria class intervals for Choropleth maps. *Ann Assoc Am Geogr* 93(3):595–623.
- Arnold FJ (1979) A test for clusters. *J Mark Res* 16:545–551
- Braga AA (2001) The effects of hot spots policing on crime. *Ann Am Acad Polit Soc Sci* 578(1):104–125
- Brantingham PJ, Brantingham PL (1981) Environmental criminology. Waveland Press, Prospect Heights
- Chainey S, Tompson L, Uhlig S (2008) The utility of hotspot mapping for predicting spatial patterns of crime. *Secur J* 21:4–28
- Cohen LE, Felson M (1979) Social change and crime rate trends: a routine activity approach. *Am Sociol Rev* 44(4):588–608
- Cohen MA, Rust RT, Steen S, Tidd ST (2004) Willingness-to-pay for crime control programs. *Criminology* 42(1):89–110
- Cohon J (1978) Multiobjective programming and planning. Academic, New York
- Cromley R (1995) Classed versus unclassed choropleth maps: a question of how many classes. *Cartographic* 32(4):15–27
- Cromley RG, Cromley EK (2009) Choropleth map legend design for visualizing community health disparities. *Int J Health Geogr* 8:52
- Dent B (1999) Cartography: thematic map design, 5th edn. WCB/McGraw-Hill, Boston
- Eck JE, Chainey S, Cameron JG, Leitner M, Wilson RE (2005) Mapping crime: understanding hot spots. U.S. Department of Justice. <https://www.ncjrs.gov/pdffiles1/nij/209393.pdf>
- Eck J, Liu L (2008) Varieties of artificial crime analysis: purpose, structure, and evidence in crime simulations. In: Liu L, Eck J (eds) Artificial crime analysis systems: using computer simulations and geographic information systems. Information Science Reference, Hershey, pp 413–432
- Estivill-Castro V, Murray A (2000) Hybrid optimization for clustering in data mining. CLAIO 2000, on CD-ROM. IMSIO, Mexico
- Everitt B (1993) Cluster analysis. Halsted, New York
- Fisher W (1958) On grouping for maximum homogeneity. *J Am Stat Assoc* 53:789–798
- Gordon AD (1996) How many clusters? An investigation of five procedures for detecting nested cluster structure. In: Hayashi C, Ohsumi N, Yajima K, Tanaka Y, Bock H, Baba Y (eds) Data science, classification, and related methods. Springer, Tokyo, pp 109–116
- Gorr W, Olligschlaeger A, Thompson Y (2003) Short-term forecasting of crime. *Int J Forecast* 19(4):579–594
- Griffith D, Amrhein C (1997) Multivariate statistical analysis for geographers. Prentice-Hall, Upper Saddle River
- Grubestic TH (2006) On the application of fuzzy clustering for crime hot spot detection. *J Quant Criminol* 22(1):77–105
- Grubestic TH, Mack EA (2008) Spatio-temporal interaction of urban crime. *J Quant Criminol* 24(3):285–306
- Harries K (1999) Mapping crime: principle and practice. National Institute of Justice, Washington, DC
- Hartigan J (1975) Clustering algorithms. Wiley, New York
- Huang Z (1998) Extensions to the k-means algorithm for clustering large data sets with categorical values. *Data Min Knowl Discov* 2:283–304.

- Johnson SD, Bowers KJ (2004) The burglary as clue to the future: the beginnings of prospective hot-spotting. *Eur J Criminol* 1(2):237–255
- Kaufman L, Rousseeuw PJ (2005) Finding groups in data. Wiley, New York.
- Kent J, Leitner M (2007) Efficacy of standard deviational ellipses in the application of criminal geographic profiling. *J Investig Psychol Offender Profil* 4:147–165
- LeBeau JL (1987) The methods and measures of centrography and the spatial dynamics of rape. *J Quant Criminol* 3:125–141
- Leitner M, Barnett M, Kent J, Barnett T (2011) The impact of Hurricane Katrina on reported crimes in Louisiana: a spatial and temporal analysis. *Prof Geogr* 63(2):244–261
- Levine N (2010) CrimeStat: A spatial statistics program for the analysis of crime incident locations, version 3.3. Ned Levine & Associates/National Institute of Justice, Washington, DC
- Levine N (2006) Crime mapping and the crimestat program. *Geogr Anal* 38(1):41–56.
- Lozano JA, Larranaga P, Grana M (1996) Partitional cluster analysis with genetic algorithms: searching for the number of clusters. In: Hayashi C, Ohsumi N, Yajima K, Tanaka Y, Bock H, Baba Y (eds) Data science, classification, and related methods. Springer, Tokyo, pp 117–124
- MacQueen J (1967) Some methods for classification and analysis of multivariate observations. In: Le Cam L, Neyman J (eds) Proceedings of the fifth Berkeley symposium on mathematical statistics and probability, vol I. University of California Press, Berkeley.
- McLafferty S, Williamson D, McGuire PG (2000) Identifying crime hot spots using kernel smoothing. In: Goldsmith V, McGuire PG, Mollenkopf JH, Ross TA (eds) Analyzing crime patterns: frontiers of practice. Sage, Thousand Oaks, pp 77–85.
- Messner SF, Anselin L, Baller RD, Hawkins DF, Deane J, Tolnay SE (1999) The spatial patterning of county homicide rates: an application of exploratory spatial data analysis. *J Quant Criminol* 15:423–450
- Milligan GW, Cooper MC (1985) An examination of procedures for determining the number of clusters in a data set. *Psychometrika* 50:159–179
- Milligan GW, Mahajan V (1980) A note on procedures for testing the quality of a clustering of a set of objects. *Decis Sci* 11:669–677
- Morenoff JD, Sampson RJ, Raudenbush SW (2001) Neighborhood inequality, collective efficacy, and the spatial dynamics of urban violence. *Criminology* 39:517–559
- Murray A (1999) Spatial analysis using clustering methods: evaluating the use of central point and median approaches. *J Geogr Syst* 1:367–383
- Murray A (2000a) Spatial characteristics and comparisons of interaction and median clustering models. *Geogr Anal* 32:1–19
- Murray A (2000b) Spatially lagged choropleth display. In: Forer P, Yeh A, He J (eds) In: Proceedings of 9th international symposium on spatial data handling, 1a40–49. International Geographical Union, Beijing
- Murray A, Estivill-Castro V (1998) Cluster discovery techniques for exploratory spatial data analysis. *Int J Geogr Inf Sci* 12:431–443
- Murray AT, Grubestic TH (2002) Identifying non-hierarchical spatial clusters. *Int J Ind Eng* 9:86–95.
- Murray A, Shyy T (2000) Integrating attribute and space characteristics in choropleth display and spatial data mining. *Int J Geogr Inf Sci* 14:649–667
- Murray A, McGuffog I, Western J, Mullins P (2001) Exploratory spatial data analysis techniques for examining urban crime. *Br J Criminol* 41:309–329
- Openshaw S, Taylor P (1981) The modifiable areal unit problem. In: Wrigley N, Bennett R (eds) Quantitative geography: a British view. Routledge and Kegan Paul, London, pp 60–69
- Parker RN (1985) Aggregation, ratio variables, and measurement problems in criminological research. *J Quant Criminol* 1:269–280
- Podani J (1996) Explanatory variables in classifications and the detection of the optimum number of clusters. In: Hayashi C, Ohsumi N, Yajima K, Tanaka Y, Bock H, Baba Y (eds) Data science, classification, and related methods. Springer, Tokyo, pp 125–132
- Ratcliffe JH (2002) Aoristic signatures and the spatio-temporal analysis of high volume crime patterns. *J Quant Criminol* 18(1):23–43

- Ratcliffe JH (2004) The hotspot matrix: a framework for the spatio-temporal targeting of crime reduction. *Police Pract Res* 5(1):5–23
- Ratcliffe JH (2005) Detecting spatial movement of intra-region crime patterns over time. *J Quant Criminol* 21(1):103–123
- ReVelle C (1987) Urban public facility location. In: Mills E (ed) *Handbook of regional and urban economics*. Elsevier Science, New York
- Rogerson P, Yamada I (2009) *Statistical detection and surveillance of geographic clusters*. CRC Press, New York
- Rossmo DK (2000) *Geographic profiling*. CRC Press, Boca Raton
- Rousseeuw P, Leroy A (1987) *Robust regression and outlier detection*. Wiley, New York
- Sampson RJ, Groves WB (1989) Community structure and crime: testing social-disorganization theory. *Am J Sociol* 94:774–802
- Shaw CR, McKay HD (1942) *Juvenile delinquency and urban areas: a study of rates of delinquency in relation to differential characteristics of local communities in American cities*. University of Chicago Press, Chicago
- Shiode S, Shiode N (2009) Detection of multi-scale clusters in network space. *Int J Geogr Inf Sci* 23(1):75–92
- Wang F (2005) *Geographic information systems and crime analysis*. Idea Group Publishing, Hershey
- Wu X, Grubestic TH (2010) Identifying irregularly shaped crime hot-spots using a multiobjective evolutionary algorithm. *J Geogr Syst* 12:409–433
- Xiao N, Armstrong MP (2005) ChoroWare: a software toolkit for Choropleth map classification. *Geogr Anal* 38(1):102–121.
- Yamada I, Thill J-C (2007) Local indicators of network-constrained clusters in spatial point patterns. *Geogr Anal* 39(3):268–292.