Crime Mapping and the *CrimeStat* Program

Ned Levine

Ned Levine & Associates, Houston, TX

CrimeStat is a spatial statistics program used in crime mapping. The program inputs incident or point locations and outputs statistics that can be displayed graphically in a geographic information systems (GIS) program. Among the routines are those for summary spatial description, hot spot analysis, interpolation, space–time analysis, and journey-to-crime modeling. Version 3.0 has a crime travel demand module for analyzing travel patterns over a metropolitan area. The program and documentation are distributed by the National Institute of Justice.

Introduction

In this article, I will discuss the *CrimeStat* program and its potential uses for both crime mapping as well as other GIS applications.

Crime mapping

Modern law enforcement has a strong technology component involving forensics, incident reconstruction, assailant profiling, database analysis, and a wide range of specialized analytical components including crime mapping. Crime mapping is an important technical function that is part of modern police enforcement. Police analysts routinely map crime incidents in order to both detect general patterns of crime that can focus their enforcement and prevention efforts as well to identify and apprehend specific offenders who are committing crimes. Long known for the famous pin map, invented by the London Metropolitan Police Department in the 1820s, most large police departments in the United States and elsewhere routinely use geographic information systems (GIS) to map crime data as part of their strategic and tactical activities. The information gained from such analysis is used for a variety of applications from focusing deployment more specifically on hot spots to targeting crime prevention efforts on particular communities to tracking the behavior of a serial offender for whom the police intend to apprehend and, even, to mapping motor vehicle crash locations, another police function.
The MAPS unit
The CrimeStat program was developed by me under research grants from the Mapping and Analysis Program (MAPS) of the National Institute of Justice (NIJ). In the United States, the NIJ is the research agency of the U.S. Department of Justice (USDOJ) and is part of the Office of Justice Programs. In 1996, NIJ created a crime mapping unit to develop and promote criminal justice analytical tools using GIS technology. Originally called the Crime Mapping Research Center, the MAPS program has developed a collection of tools and applications on crime mapping (MAPS 2004). Among the tools developed or distributed by MAPS were a multi-jurisdictional spatial crime analysis package (RCAGIS), a mobile crime mapping tool for police cars (the Community Policing Beat Book), an ArcView crime analysis extension, a client–server version of a regional crime analysis program (RCAP-SDE), a school crime incident tool for documenting and mapping crimes occurring in schools (School COP), along with CrimeStat. The MAPS unit has also run an annual crime mapping conference, created formal and self-training courses on crime mapping, funded many research studies on crime mapping (e.g., LeBeau 1997; LaVigne and Wartell 1998, 2000; Harries 1999; Jefferis 1999; Rich 2001; Wartell and McEwen 2001; Stoe et al. 2003), and are developing a regional spatial data repository for crime and terrorist incidents occurring in the United States. More information about MAPS can be found on their Web site.1

The CrimeStat program
CrimeStat is a stand-alone Windows spatial statistics program for the analysis of crime incident locations that can interface with most desktop GIS programs. The purpose is to provide supplemental statistical tools to aid law enforcement agencies and criminal justice researchers in their crime mapping efforts. The NIJ is the sole distributor of CrimeStat and makes it available for free to analysts, researchers, educators, and students.2 The program is distributed along with a manual that describes each of the statistics and gives examples of their use.

The program is written in C++ and is multithreading. It was designed with large databases in mind as most metropolitan police departments work with large files. It will also take advantage of multiple processors in a computer which, for a large data set, will considerably cut down on calculation time. The program also includes Dynamic Data Exchange code to allow another program to call up CrimeStat and pass the data set name and variable parameters to it. Such a use was developed by the Criminal Division of the USDOJ in developing their Regional Crime Analysis GIS (RCAGIS).3 Several jurisdictions in the Baltimore metropolitan area use RCAGIS to link crime databases to a common interface and a set of analytical tools, including CrimeStat.

CrimeStat is being used by many police departments around the country as well as by criminal justice and other researchers. Four versions have been released. The first was released in November 1999 and an update version was released in
August 2000. Version 2.0 was released in August 2000. The latest version, *CrimeStat III*, was released in March 2005. From what we can tell, it has been used in many courses and has been a widely used research tool. The program has also been used by researchers from other fields than criminal justice including geography, epidemiology, forestry science, botany, and geology.

**Data input and output**

The program inputs incident locations (e.g., robbery locations, motor vehicle crash sites, residence locations of persons with a particular disease) in “.dbf,” “.shp,” “.dat,” or ASCII formats, as well as programs that follow the ODBC standard. It can use spherical or projected coordinates. It can also treat zones as pseudo-points (or points with intensities). Although there is a primary file that must be input, the program can allow secondary and other files to be used for comparison or specialized purposes. The user can also define a reference grid by the coordinates of the “corners” and the program will create it for specialized uses.

Once the data are input, the program calculates various spatial statistics. These vary from very simple descriptions to complicated spatial models involving the interaction of space and time. Many of the routines are written as graphical objects to *ArcView*, *ArcGis*, *MapInfo*, *Atlas GIS*, *Surfer* for Windows, and *ArcView Spatial Analyst*.

**Spatial statistics in CrimeStat**

The following gives some examples of the statistics that are in the program.

**Spatial description**

There are a number of statistics for describing the general properties of a distribution. These involve simple descriptions of overall pattern (global characteristics), descriptions of regional variation, and descriptions of small, concentrated clusters (hot spots). Among these are the mean center, center of minimum distance, standard deviational ellipse, and the directional mean (Ebdon 1988; LeBeau 1992).

These simple analogies to univariate statistics can be used to compare different types of distributions or to compare the same distribution for different time periods. As an example, Fig. 1 shows the standard deviational ellipses for burglaries in Precinct 12 in Baltimore County for June and July 1997. As seen, there is a spatial shift that occurred between June and July. As summer progresses, some vacationers occupy the communities along the Chesapeake Bay and the distribution of burglaries follows this pattern.

**Spatial autocorrelation**

A key concept in spatial statistics is that of *spatial autocorrelation* (Griffith 1987). There are various definitions of spatial autocorrelation but a simple one is that events are spatially arranged in a nonrandom manner, either more concentrated or, occasionally, more dispersed than would be expected on the basis of chance. There
are several well-known global measures of spatial autocorrelation—Moran’s $I$, Geary’s $C$, and the Moran correlogram—that are included in CrimeStat (Moran 1948; Geary 1954; Ebdon 1988). There are also several statistics that describe spatial autocorrelation through the properties of distances between incidents including nearest neighbor analysis (Clark and Evans 1954), linear nearest neighbor analysis, $K$-order nearest neighbor (Cressie 1991), and Ripley’s $K$ statistic (Ripley 1976, 1981). The testing of significance for Ripley’s $K$ is done through a Monte Carlo simulation that estimates approximate confidence intervals.

As an example, Fig. 2 below shows the Ripley’s $K$ distribution of motor vehicle crashes in Houston in 1998 and compares it to both an “envelope” from 100 random Monte Carlo simulations as well as the distribution of the 2000 population (measured by the centroids of census blocks). Ripley’s $K$ counts the cumulative number of other points within a circle of a certain radius placed over each point in the distribution. The count is made for multiple radii so that concentration can be compared at different scales. As seen, the distribution of vehicle crashes is highly concentrated (i.e., having a larger count within the search circle), more so than would be expected by the population distribution and certainly more so than would be expected under complete spatial randomness.

**Hot spot analysis**

An extreme form of spatial autocorrelation is a **hot spot**. While there is no absolute definition of a “hot spot,” it is often noted that incidents, particularly crime
incidents, tend to be concentrated in a limited number of locations (Block and Block 1995). Police officers, crime analysts, and researchers are very familiar with this concentration of events through the numerous calls for service from residents, the large number of crimes committed, as well as the sizeable number of arrests that are made in these areas. There is a large literature on high-crime areas so that the phenomenon is very well known (e.g., see Thrasher 1927; Shaw and McKay 1942; Newman 1972; Cohen and Felson 1979; Wilson and Kelling 1982; Bursik and Grasmick 1993; Bowers and Hirschfield 1999).

From an analytical perspective, tools that identify hot spots are very useful to police departments because they tend to focus their deployment and prevention resources on the areas that are most likely to generate incidents.\(^7\)

There are seven “hot spot” analysis routines in CrimeStat: the mode, the fuzzy mode, hierarchical nearest neighbor clustering (Everett 1974; D’andrade 1978), risk-adjusted nearest neighbor hierarchical clustering (Levine 2004a), the Spatial and Temporal Analysis of Crime routine (STAC; Block 1995), K-means clustering (Everett 1974; McBratney and deBruijswijk 1972), and the local Moran statistic (Anselin 1995).

To illustrate, Fig. 3 shows first- and second-order standard deviational ellipses of driving while intoxicated (DWI) crashes in central Houston from 1999 to 2001, using the nearest-neighbor hierarchical clustering routine. The first-order clusters are the grouping of incidents while the second order are the grouping of the first-order clusters. As seen, the incidents tend to occur in small clusters. Several of the
small clusters, in turn, are grouped into larger district clusters. Fig. 4 zooms into one of the clusters in the East End of Houston, a low-income community with many DWI crashes.

Using another example, Fig. 5 shows the clustering of street robberies in west Baltimore County using the STAC clustering algorithm. As seen, three of them fall along a major arterial in the county (State Highway 26); the robberies are concentrated at commercial strips along the arterial.

Because the hot spot tools are complex algorithms, statistical significance must be tested with a Monte Carlo simulation. The nearest-neighbor hierarchical clustering, the risk-adjusted nearest-neighbor hierarchical clustering, and the STAC routines each have a Monte Carlo simulation that allows the estimation of approximate confidence intervals or test thresholds for these statistics.

Of course, a hot spot routine only identifies a collection of points that are close together. It does not explain why they are together. For that, additional research and analysis is required. In the case of crime incident hot spots, the clustering could be due to a high concentration of potential victims (e.g., at a shopping mall), particular land uses that encourage crimes (e.g., an area with a concentration of bars and adult bookshops; Levine, Wachs, and Shirazi 1986), a common activity (e.g., a drug trade “center”), a location where many offenders live, or a neighborhood where a rash of incidents suddenly occur (e.g., vehicle thieves often hit a neighborhood for a short period of time). The hot spot could also be due to chance; in any distribution, a certain amount of clustering will occur by chance. That is why it is important to test any hot spot against a random distribution (through a Monte Carlo simulation,
for example) and to also examine several years of data to ensure that it is not transitory.

**Spatial modeling**
There are a number of tools in *CrimeStat* for spatial modeling. Typically, these extrapolate beyond the values in the data set, either in space or in time.

**Interpolation**
*Interpolation* involves extrapolating a density estimate from individual data points. A fine-mesh grid is placed over the study area, the distance from each grid cell to each data point is calculated, and an estimate of incident density for each grid cell is made using a mathematical function (a kernel) that relates the density to distance (Bailey and Gatrell 1995). *CrimeStat* uses five different mathematical functions to estimate the density and has two different applications of it—a single-variable kernel density estimation routine for producing a surface or contour estimate of the density of incidents (e.g., the density of burglaries) and a dual-variable kernel density estimation routine for comparing the density of incidents to the density of an underlying baseline (e.g., the density of burglaries relative to the density of households).

As an example, Fig. 6 shows a three-dimensional kernel density interpolation of 1990 motor vehicle crashes relative to 1990 population in Honolulu. The crash data came from the Honolulu Police Department while the population data were for census block groups. As seen, the interpolation of the crashes
shows an extremely concentrated pattern centering on Waikiki and downtown Honolulu. To a large extent, the concentration reflects that of population which is also highly concentrated. However, when the estimate of crashes is divided by the estimate of population within each grid cell, the result shows a more dispersed pattern. The population centers show a high crash risk in the center, but perimeter roads on the northern and western parts of the island also show a high risk.

Journey-to-crime analysis
An important analytical tool for police departments seeking to apprehend a serial offender is journey-to-crime analysis (sometimes known as geographic profiling). This is a criminal justice method for estimating the likely residence location of a serial offender given the distribution of incidents and a model for travel distance (Brantingham and Brantingham 1981; Canter and Gregory 1994; Rossmo 1995; Levine 2004b).

As an example, Fig. 7 shows the predicted residence location of an offender who committed 10 crimes between 1994 and 1996 in eastern Baltimore County. Nine of the committed offenses were larceny thefts, but one was an assault. The prediction is estimated from a travel demand function that is calibrated from a sample of 19,806 known larcenies. The calibration sample included the origin
location (usually the offender’s residence) and destination location (the crime location) from closed arrest records. As seen, the offenses were spread over an area of about 10 square miles. The journey-to-crime function estimates three areas of high likelihood for the offender, of which one is where the offender actually lived (house symbol).

**Space–time analysis**

There are several routines for analyzing clustering in time and in space. These include the Knox and Mantel indices, which examine the relationship between time and space, and the Correlated Walk Analysis module, which analyzes and predicts the behavior of a serial offender. The Knox and Mantel routines each have a Monte Carlo simulation to estimate confidence intervals around the calculated statistic.

The Correlated Walk Analysis includes separate regression routines for testing the significance of various lags for time, direction, and distance. Based on an analysis of repetitive behavior in time, direction, or distance, a guess can be made about where and when the next event will take place. Fig. 8 shows the sequence of six offenses committed by a single individual between 1993 and 1997. The offenses included four residential burglaries and two residential robberies. Two of the locations were burglarized twice in the sequence. The map shows the predicted next event (the seventh event) from a Correlated Walk Analysis of the sequence, and the actual location where the next crime was committed (a residential burglary). As seen, the prediction was reasonably close in distance (error of 0.77 miles) and in time (error of 1.9 days).

Figure 6. Honolulu crash risk: 1990.
Version 3.0 includes a crime travel demand module for modeling criminal travel behavior over an entire metropolitan area. It is an application of travel demand modeling used widely in transportation planning (Ortuzar and Willumsen 2001). There are four separate stages. In the first, predictive models of crimes occurring in a series of zones (crime destinations) and originating in a series of zones (crime origins) are estimated using either a Poisson or ordinary least-squares regression. In the second stage, a gravity-type model is fit to the predicted origins and destinations to yield a model of crime trips from each origin zone to each destination zone. The calibrated model can be compared with an actual distribution of crime trips, usually obtained from police arrest records.

In the third stage, the predicted crime trips are separated into different travel modes (e.g., walking, biking, driving, transit). The aim is to examine possible strategies used by offenders in targeting their victims. In the fourth, and final, stage, the predicted crime trips by travel mode are assigned to particular routes, which can fall along a street network or a transit network. The A* shortest path algorithm is used to estimate the likely travel route taken (Nilsson 1980; Sedgewick 2002). The cost of travel along the network can be estimated using distance, travel time, or a generalized cost.

As an example, Fig. 9 shows the major crime zone-to-zone trip links (all types) for Baltimore County from 1993 to 1997. The trips were calculated from the point location of incidents and assigned to 325 destination traffic analysis zones within Baltimore County.
Baltimore County and to 532 origin traffic analysis zones in both Baltimore County and the City of Baltimore. Each trip link is displayed as a line from the centroid of the origin zone to the centroid of the destination zone with the thickness of the line being proportional to the number of crimes on the link. All of the major links are crime “trips” to shopping malls. There are multiple origin locations, although some produce more crime trips to the malls than others.

As another example of the crime travel demand module, Fig. 10 shows the total number of vehicle theft trips traveling on each major roadway link to Baltimore County from both Baltimore County and the City of Baltimore. Each segment count is obtained by summing the number of trips from each origin zone to each destination zone after assigning it to a probable route using the A* algorithm. Travel is weighted by travel time so that the routes indicate those with the shortest travel time. As seen, there is a substantial amount of travel on the Baltimore Beltway (I-695). Even though travel on that roadway is more circuitous than more direct routes, it is faster because it is a freeway. In general, travel time is a much better predictor of travel behavior than distance (Ortuzar and Willumsen 2001).

**Options**
There are also several miscellaneous options in CrimeStat that make the program easier to use. Parameters can be saved and reloaded, tab colors can be changed, and Monte Carlo simulation data can be output.
CrimeStat is accompanied by sample data sets and a manual that gives the background behind the statistics with many examples. The manual includes examples contributed by researchers from many different fields. As mentioned, the software and documentation are available for free from the NIJ (see footnote 2).

**Future plans**

In the next version of CrimeStat, we plan to include several spatial regression routines including nonlinear models, add more options to the crime travel demand model, incorporate Bayesian modeling techniques, and integrate the SatScan hot spot routine which examines space–time clustering (Kulldorff 1997). Also, we will update the interface and improve the integration of the program with other GIS and statistical applications.

**Conclusion**

The integration of GIS into law enforcement has been an important technological breakthrough for crime analysts and criminal justice researchers. The technology has allowed police departments to monitor crime and other incidents in a much more visual manner than was previously possible. GIS is almost universally used within large, medium, and even small police departments. Nevertheless, the sudden availability of large amounts of data has created problems in processing the information for these departments. Over the course of a year, a large police
The department will process hundreds of thousands of incidents so that visual maps by themselves are insufficient for monitoring the levels of crime in a jurisdiction. The amount of information that is now documented by a crime mapping GIS system is enormous.

Hence, there is a strong need for statistical and other analytical tools that can summarize and assess the important trends in the data. *CrimeStat* is one of the tools that was developed to allow this processing to occur. It is clearly not the only tool that conducts spatial analysis of crime incidents (see the MAPS Web site for more information). But, it is an important tool that been used by crime analysts, criminal justice researchers, and even researchers from other fields. Over the years, we have expanded the program to incorporate more needs as these have been articulated by users and law enforcement personnel in general. After three versions, the program has developed way beyond the original conception of it in 1996 (Levine 1996).

In some ways, it is an exploration as new needs are emerging all the time and we run to keep up with these trends. In this sense, spatial statistics is, in itself, an emerging field. While the statistics impose a certain structure on the data, examining some aspects of spatial relations and not others, the statistics themselves are being created by the emerging needs of users. The researchers need to keep abreast of the explorations of the analysts, while the reverse is also true. It is this symbiosis that produces the creative endeavor that we call spatial analysis.

---

*Figure 10.* Auto theft trips by road segment.
Notes

2. The program is available at: http://www.ojp.usdoj.gov/nij/maps or http://www.icpsr.umich.edu/crimestat
3. The RCAGIS product is described at: http://www.icpsr.umich.edu/NACJD/RCAGIS
4. Version 1.0 had over 2500 unique IP downloads, version 1.1 had more than 5000 unique IP downloads, and version 2.0 had more than 7000. Many of the chapters have been downloaded more than 20,000 times. It is among the top downloads at the ICPSR site (http://www.icpsr.umich.edu/access/quick-data.html). The program was also recognized in a Vice Presidential National Partnership for Reinventing Government award, as part of its contribution to the RCAGIS project (see footnote 2).
5. I would like to thank the Baltimore County Police Department (BCPD) and, in particular, Phil Canter for providing information on crime incidents in their jurisdiction.
6. The information is courtesy of the Houston-Galveston Area Council. More information can be found at http://www.h-gac.com/safety
7. Because incidents tend to cluster in a limited number of locations, they pose some difficult statistical problems for modeling the behavior. Ordinary least squares cannot be used as it will usually underpredict incidents at the hot spot locations and overpredict incidents at most other locations. Even the use of highly skewed Poisson and negative binomial distributions may not solve the problem. “Hot spots” are typically caused by factors that are unique to the location and which would not be measured at other locations.
8. The broader study can be found in Levine, Kim, and Nitz (1995) and Kim and Levine (1996).

References


