The author(s) shown below used Federal funds provided by the U.S. Department of Justice and prepared the following final report:

Document Title: CrimeStat II: Spatial Modeling, Part III

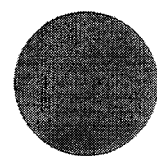Author(s):   Ned Levine and Associates

Document No.:   195774

Date Received: August 13, 2002

Award Number: 99-IJ-CX-0044

195774
Pt. III

CrimeStat II

Part III: Spatial Modeling

195774
Pt III

Chapter 8
Kernel Density Interpolation

In this chapter, we discuss tools aimed at interpolating incidents, using the kernel density approach. *Interpolation* is a technique for generalizing incident locations to an entire area. Whereas the spatial distribution and hot spot statistics provide statistical summaries for the data incidents themselves, interpolation techniques generalize those data incidents to the entire region. In particular, they provide *density* estimates for all parts of a region (i.e., at any location). The density estimate is an intensity variable, a Z-value, that is estimated at a particular location. Consequently, it can be displayed by either surface maps or contour maps that show the intensity at all locations.

There are many interpolation techniques, such as Kriging, trend surfaces, local regression models (e.g., Loess, splines), and Dirichlet tessellations (Anselin, 1992; Cleveland, Grosse and Shyu, 1993; Venables and Ripley, 1997). Most of these require a variable that is being estimated as a function of location. However, *kernel density estimation* is an interpolation technique that is appropriate for individual point locations (Silverman, 1986; Härdle, 1991; Bailey and Gatrell, 1995; Burt and Barber, 1996; Bowman and Azalini, 1997).

Kernel Density Estimation

Kernel density estimation involves placing a symmetrical surface over each point, evaluating the distance from the point to a reference location based on a mathematical function, and summing the value of all the surfaces for that reference location. This procedure is repeated for all reference locations. It is a technique that was developed in the late 1950s as an alternative method for estimating the density of a histogram (Rosenblatt, 1956; Whittle, 1958; Parzen, 1962). A histogram is a graphic representation of a frequency distribution. A continuous variable is divided into intervals of size, s (the interval or bin width), and the number of cases in each interval (bin) are counted and displayed as block diagrams. The histogram is assumed to represent a smooth, underlying distribution (a density function). However, in order to estimate a smooth density function from the histogram, traditionally researchers have linked adjacent variable intervals by connecting the midpoints of the intervals with a series of lines (Figure 8.1).

Unfortunately, doing this causes three statistical problems (Bowman and Azalini, 1997):

1.   Information is discarded because all cases within an interval are assigned to the midpoint. The wider the interval, the greater the information loss.

2.   The technique of connecting the midpoints leads to a discontinuous and not smooth density function even though the underlying density function is assumed to be smooth. To compensate for this, researchers will reduce the width of the interval. Thus, the density function becomes smoother with

301

Figure 8.1:
Constructing A Density Estimate From A Histogram
Method of Connecting Midpoints
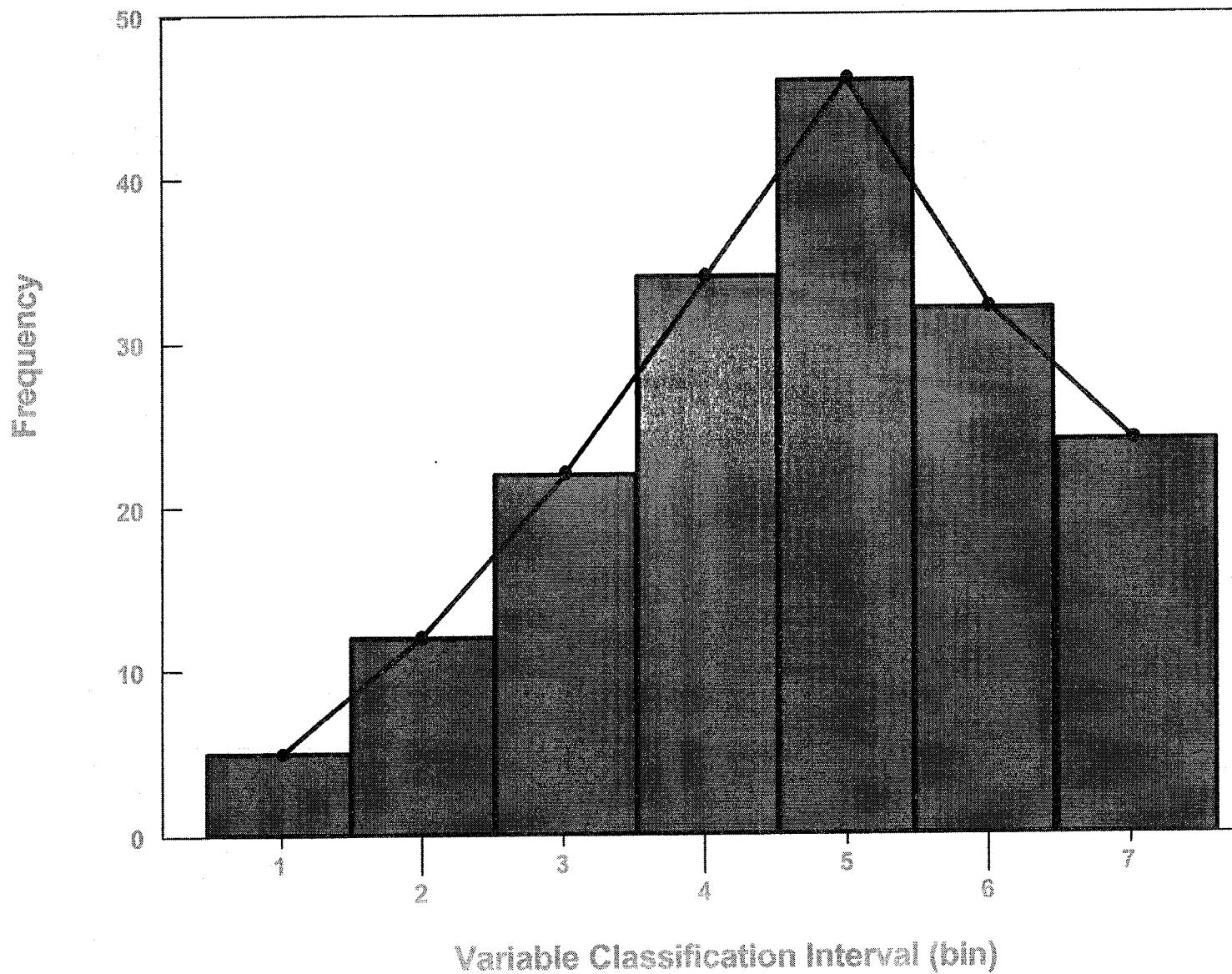
smaller interval widths, although still not very smooth. Further, there are limits to this technique as the sample size decreases when the bin width gets smaller, eventually becoming too small to produce reliable estimates.

3.   The technique is dependent on an arbitrarily defined interval size (bin width). By making the interval wider, the estimator becomes cruder and, conversely, by making the interval narrower, the estimator becomes finer. However, the underlying density distribution is assumed to be smooth and continuous and not dependent on the interval size of a histogram.

To handle this problem, Rosenblatt (1956), Whittle (1958) and Parzen (1962) developed the kernel density method in order to avoid the first two of these difficulties; the bin width issue still remains. What they did was to place a smooth *kernel function*, rather than a block, over each point and sum the functions for each location on the scale. Figure 8.2 illustrates the process with five point locations. As seen, over each location, a symmetrical kernel function is placed; by symmetrical is meant that is falls off with distance from each point at an equal rate in both directions around each point. In this case, it is a normal distribution, but other types of symmetrical distribution have been used. The underlying density distribution is estimated by summing the individual kernel functions at *all* locations to produce a smooth cumulative density function. Notice that the functions are summed at every point along the scale and not just at the point locations. The advantages of this are that, first, each point contributes equally to the density surface and, second, the resulting density function is continuous at all points along the scale.

The third problem mentioned above, interval size, still remains since the width of the kernel function can be varied. In the kernel density literature, this is called *bandwidth* and refers essentially to the width of the kernel. Figure 8.3 shows a kernel with a narrow bandwidth placed over the same five points while figure 8.4 shows a kernel with a wider bandwidth placed over the points. Clearly, the smoothness of the resulting density function is a consequence of the bandwidth size.

There are a number of different kernel functions that have been used, aside from the normal distribution, such as a triangular function (Burt and Barber, 1996) or a quartic function (Bailey and Gatrell, 1995). Figure 8.5 illustrates a quartic function. But the normal is the most commonly used (Kelsall and Diggle, 1995a).

The *normal distribution* function has the following functional form:

$$g(x_j) = \Sigma \left\{ \ [W_i * I_i] * \frac{1}{h^2 * 2\pi} * e^{-\left[\frac{d_{ij}^2}{2*h^2}\right]} \right\} \tag{8.1}$$

where d_{ij} is the distance between an incident location and any reference point in the region, h is the standard deviation of the normal distribution (the bandwidth), W_i is a weight at the point location and I_i is an intensity at the point location. This function extends to infinity in all directions and, thus, will be applied to any location in the region.

303

Figure 8.2:

Kernel Density Estimate
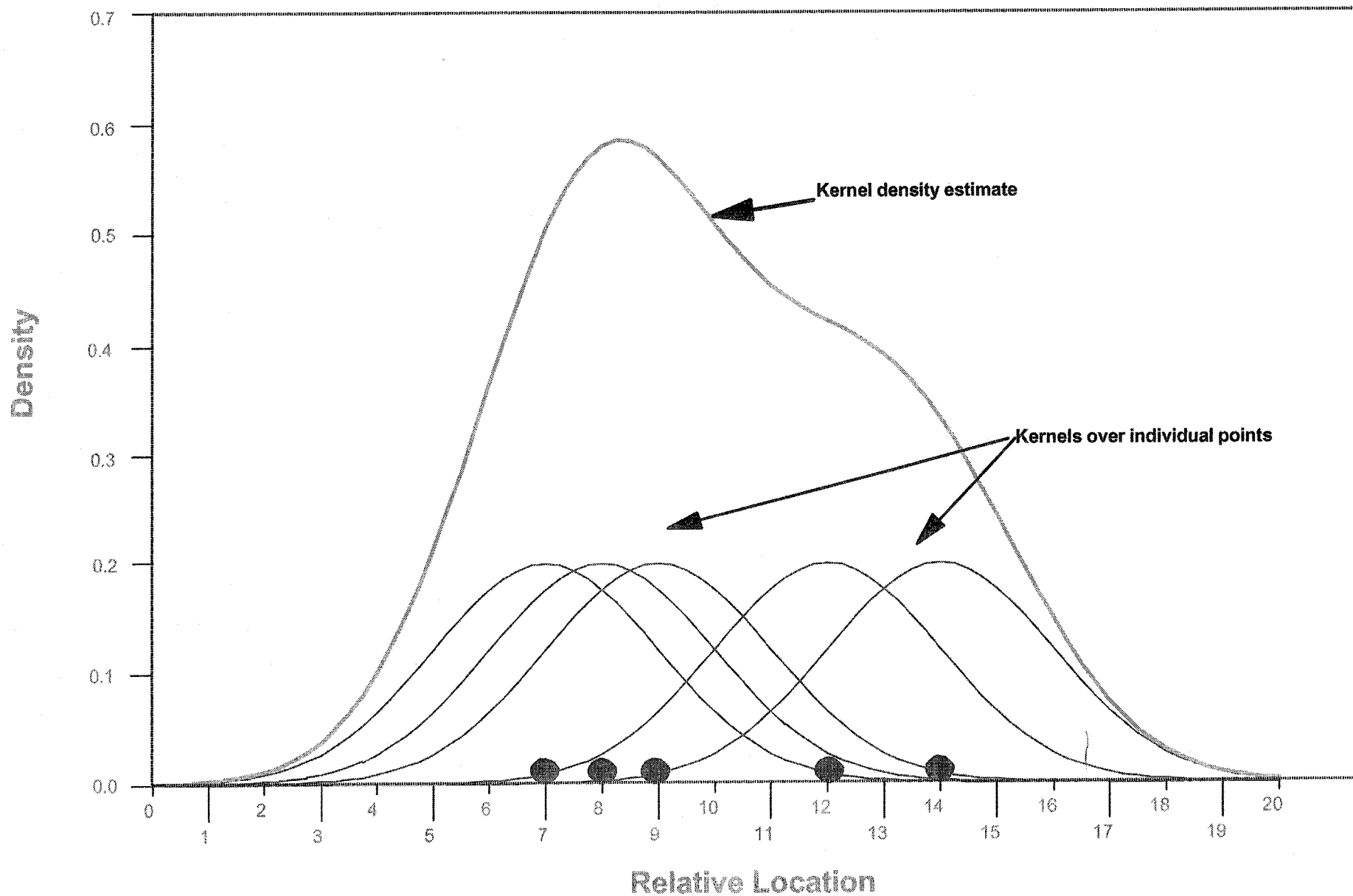## Summing of Normal Kernel Functions for 5 Points

Figure 8.3:

Kernel Density Estimate
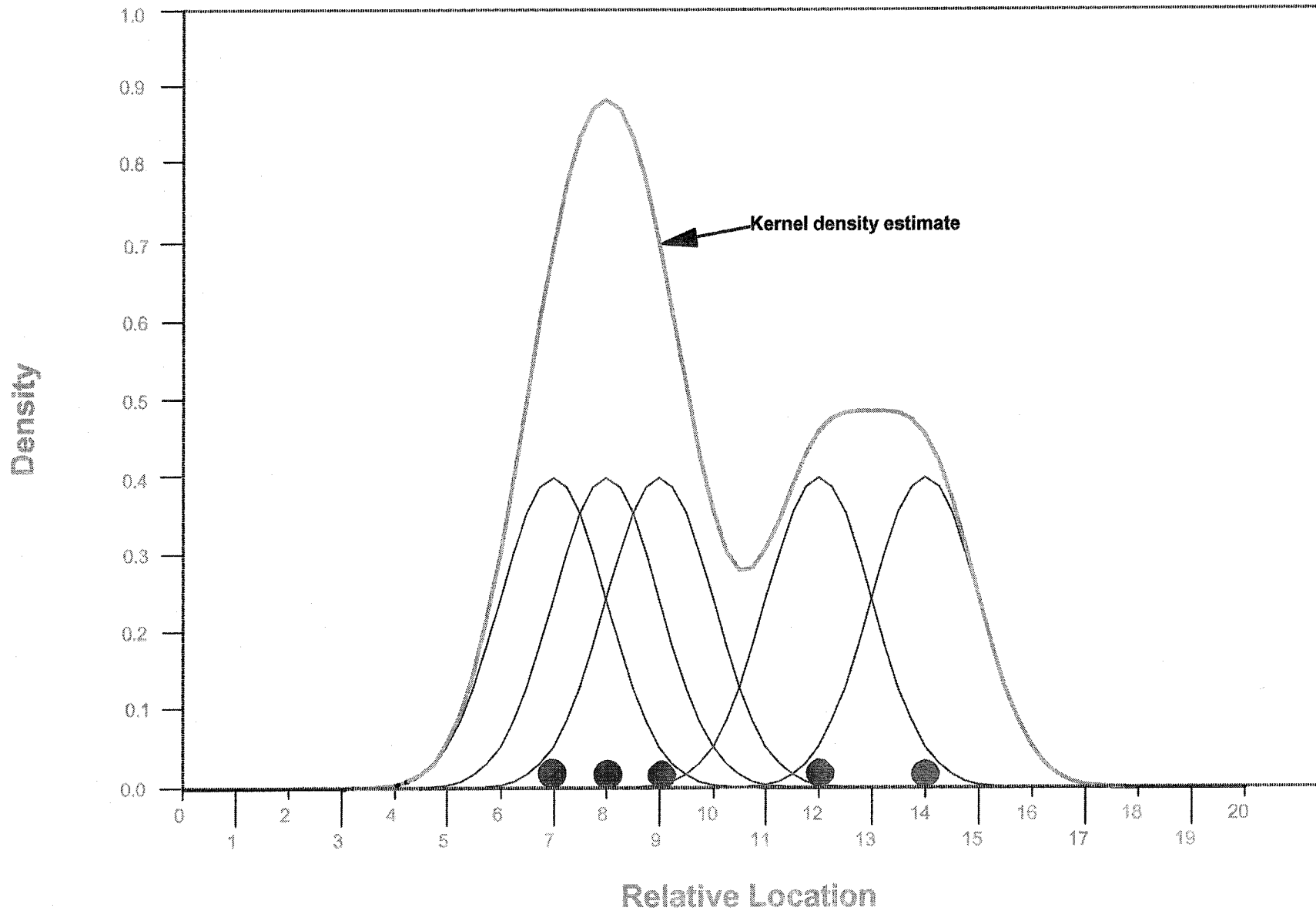## Smaller Bandwidth

Figure 8.4:
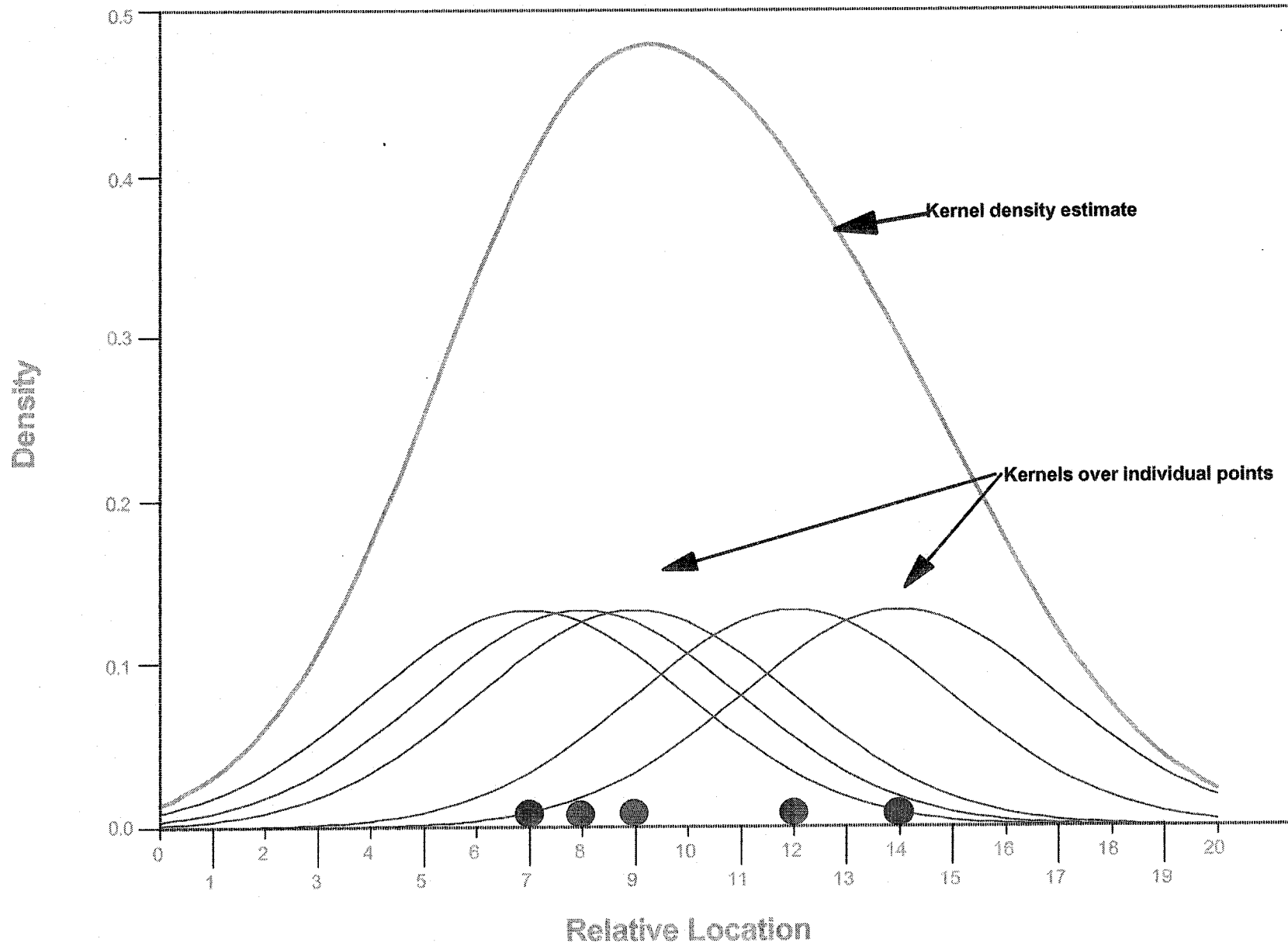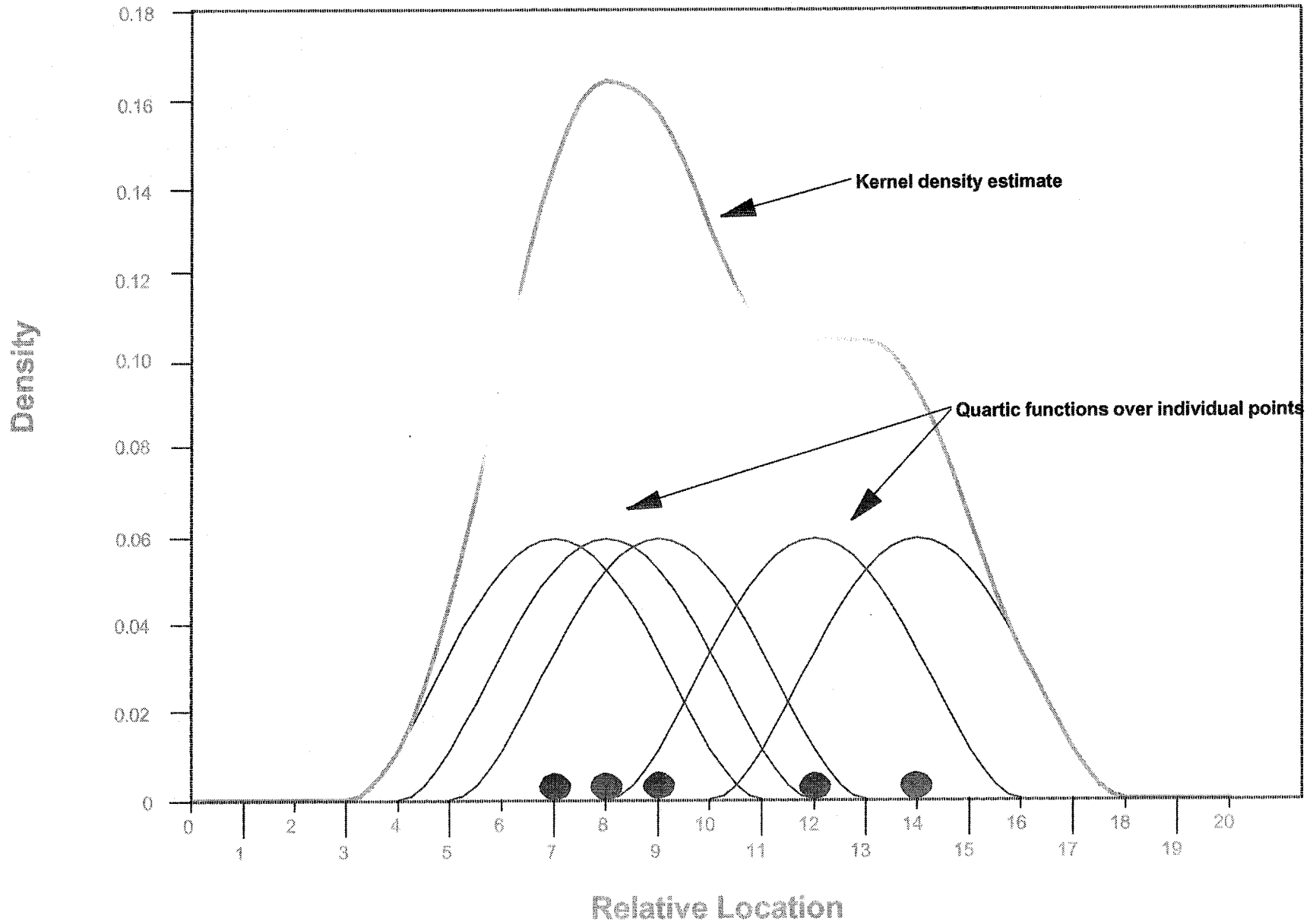# Kernel Density Estimate
Larger Bandwidth

Figure 8.5:

Kernel Density Estimate
Summing of Quartic Kernel Function

In *CrimeStat*, there are four alternative kernel functions that can be used, all of which have a circumscribed radius (unlike the normal distribution). The *quartic* function is applied to a limited area around each incident point defined by the radius, h. It falls off gradually with distance until the radius is reached. Its functional form is:

I. Outside the specified radius, h:

$$g(x_j) = 0 \tag{8.2}$$

II. Within the specified radius, h:

$$g(x_j) = \sum \left\{ \ [W_i * I_i] * [\frac{3}{h^2 * \pi}] * [1 - \frac{d_{ij}^2}{h^2}]^2 \ \right\} \tag{8.3}$$

where d_{ij} is the distance between an incident location and any reference point in the region, h is the radius of the search area (the bandwidth), W_i is a weight at the point location and I_i is an intensity at the point location.

The *triangular* (or conical) distribution falls off evenly with distance, in a linear relationship. Compared to the quartic function, it falls off more rapidly. It also has a circumscribed radius and is, therefore, applied to a limited area around each incident point, h. Its functional form is:

II Outside the specified radius, h:

$$g(x_j) = 0 \tag{8.4}$$

II. Within the specified radius, h:

$$g(x_j) = \sum [K - K/h] * d_{ij} \tag{8.5}$$

where K is a constant. In *CrimeStat*, the constant K is initially set to 0.25 and then re-scaled to ensure that either the densities or probabilities sum to their appropriate values (i.e., N for densities and 1.00 for probabilities).

The *negative exponential* (or peaked) distribution falls off very rapidly with distance up to the circumscribed radius. Its functional form is:

II Outside the specified radius, h:

$$g(x_j) = 0 \tag{8.6}$$

II. Within the specified radius, h:

$$g(x_j) = \sum A*e^{-K*d_{ij}} \tag{8.7}$$

308

where A is a constant and K is an exponent. In *CrimeStat*'s implementation, K is set to 3 while A is initially set to 1 and then re-scaled to ensure that either the densities or probabilities sum to their appropriate values (i.e., N for densities and 1.00 for probabilities).

Finally, the *uniform* distribution weights all points within the circle equally. Its functional form is:

II   Outside the specified radius, h:

$$g(x_j) = 0 \qquad\qquad (8.8)$$

II.   Within the specified radius, h:

$$g(x_j) = \Sigma K \qquad\qquad (8.9)$$

where K is a constant. Initially, K is set to 0.1 but then re-scaled to ensure that either the densities or probabilities sum to their appropriate values (i.e., N for densities and 1.00 for probabilities).

Weighting Effects

The user can select these five different kernel functions to interpolate the data to the grid cells. They produce subtle differences in the shape of the interpolated surface or contour. The normal distribution weighs all points in the study area, though near points are weighted more highly than distant points. The other four techniques use a circumscribed circle around the grid cell. The uniform distribution weighs all points within the circle equally. The quartic function weighs near points more than far points, but the fall off is gradual. The triangular function weighs near points more than far points within the circle, but the fall off is more rapid. Finally, the negative exponential weighs near point much more highly than far points within the circle.

The use of any of one of these depends on how much the user wants to weigh near points relative to far points. Using a kernel function which has a big difference in the weights of near versus far points (e.g., the negative exponential or the triangular) tends to produce finer variations within the surface than functions which are weight more evenly (e.g., the normal distribution, the quartic, or the uniform); these latter ones tend to *smooth* the distribution more.

However, Silverman (1986) has argued that it does not make that much difference as long as the kernel is symmetrical. There are also edge effects that can occur and there have been different proposed solutions to this problem (Venables and Ripley, 1997).

There have also been variations of the size of the of bandwidth with various formulas and criteria (Silverman, 1986; Härdle, 1991; Venables and Ripley, 1997).

309

Generally, bandwidth choice fall into either fixed or adaptive (variable) choices (Kelsall and Diggle, 1995a; Bailey and Gatrell, 1995). *CrimeStat* follows this distinction, which will be explained below.

The kernel function can be expanded to more than two dimensions (Härdle, 1991; Bailey and Gatrell, 1995; Burt and Barber, 1996; Bowman and Azalini, 1997). Figure 8.6 shows a three-dimensional normal distribution placed over each of five points with the resulting density surface being a sum of all five individual surfaces. Thus, the method is particularly appropriate for geographical data, such as crime incident locations. The method has also been developed to relate two or more variables together by applying a kernel estimate to each variable in turn and then dividing one by the other to produce a three-dimensional estimate of *risk* (Kelsall and Diggle, 1995a; Bowman and Azalini, 1997).

Significance testing of density estimates is more complicated. Current techniques tend to focus on simulating surfaces under spatially random assumptions (Bowman and Azaline, 1997; Kelsall and Diggle, 1995b). Because of the still experimental nature of the testing, *CrimeStat* does not include any testing of density estimates in this version.

CrimeStat Kernel Density Methods

CrimeStat has two interpolation techniques, both based on the kernel density technique. The first applies to a single variable, while the second to the relationship between two variables. Both routines have a number of options. Figure 8.7 shows the interpolation page in *CrimeStat*. Users indicate their choices by clicking on the tab and menu items. For either technique, it is necessary to have a reference file, which is usually a grid placed over the study region (see chapter 3). The reference file represents the region to which the kernel estimate will be generalized (figure 8.8).

Single Density Estimates

The single kernel density routine in *CrimeStat* is applied to a distribution of point locations, such as crime incidents. It can be used with either a primary file or a secondary file; the primary file is the default. For example, the primary file can be the location of motor vehicle thefts. The points can also have a weighting or an associated intensity variable (or both). For example, the points could represent the location of police stations while the weights (or intensities) represent the number of calls for service. Again, the user must be careful in having both a weighting variable and an intensity variable as the routine will use both variables in calculating densities; this could lead to double weighting.

Having defined the file on the primary (or secondary) file tabs, the user indicates the routine by checking the 'Single' box. Also, it is necessary to define a reference file, either an existing file or one generated by *CrimeStat* (see chapter 3). There are other parameters that must be defined.

310

Figure 8.6:

Kernel Density Surfaces
Summing of Normal Kernel Surfaces for 5 Points
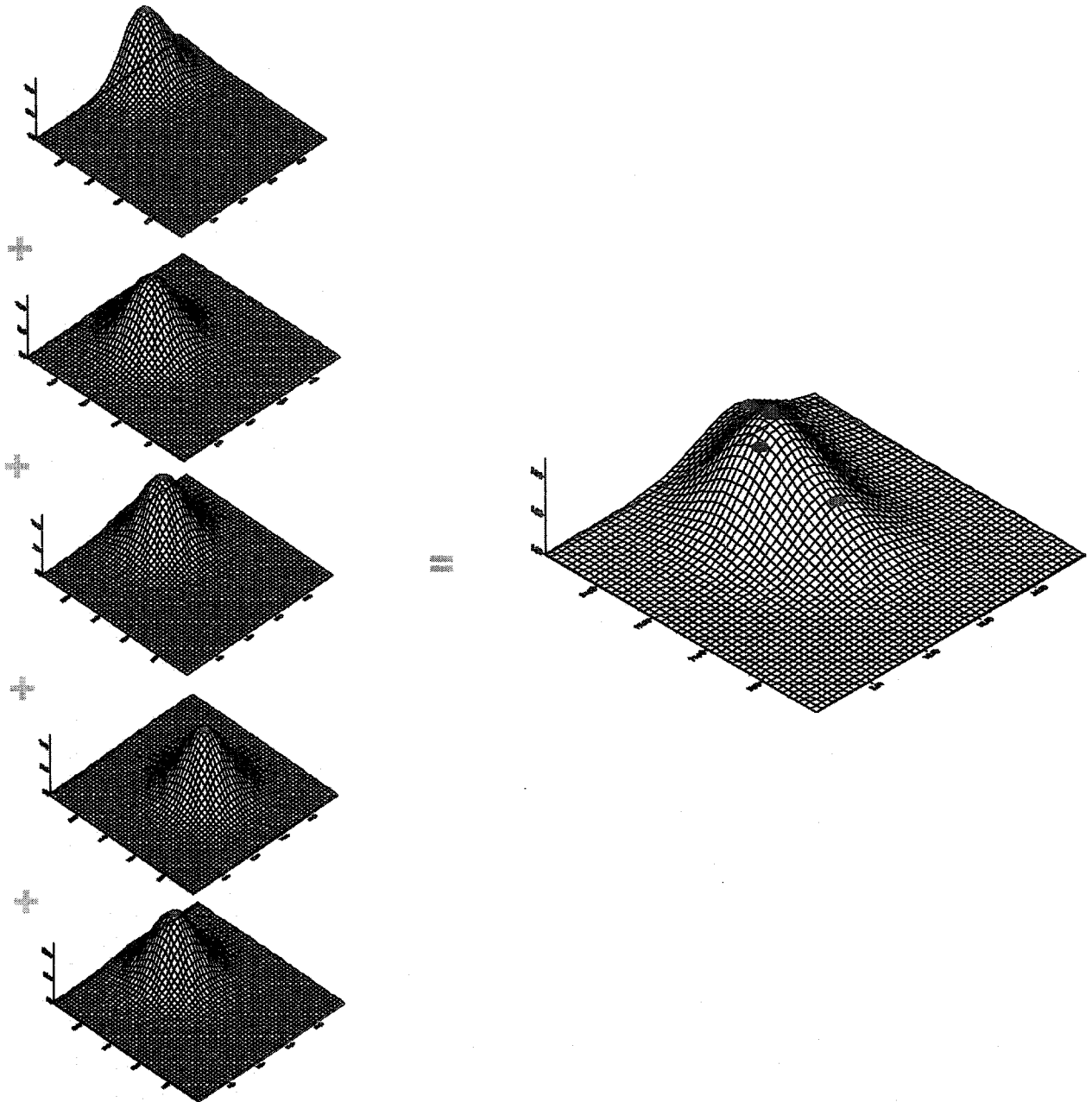
Figure 8.7: **Interpolation Screen**

Figure 8.8:

Grid Cell Structure for Baltimore Region
108 Width x 100 Height Grid Cells

Upper-right Coordinate

Baltimore County

City of Baltimore

Miles

0   2   4

Lower-left Coordinate
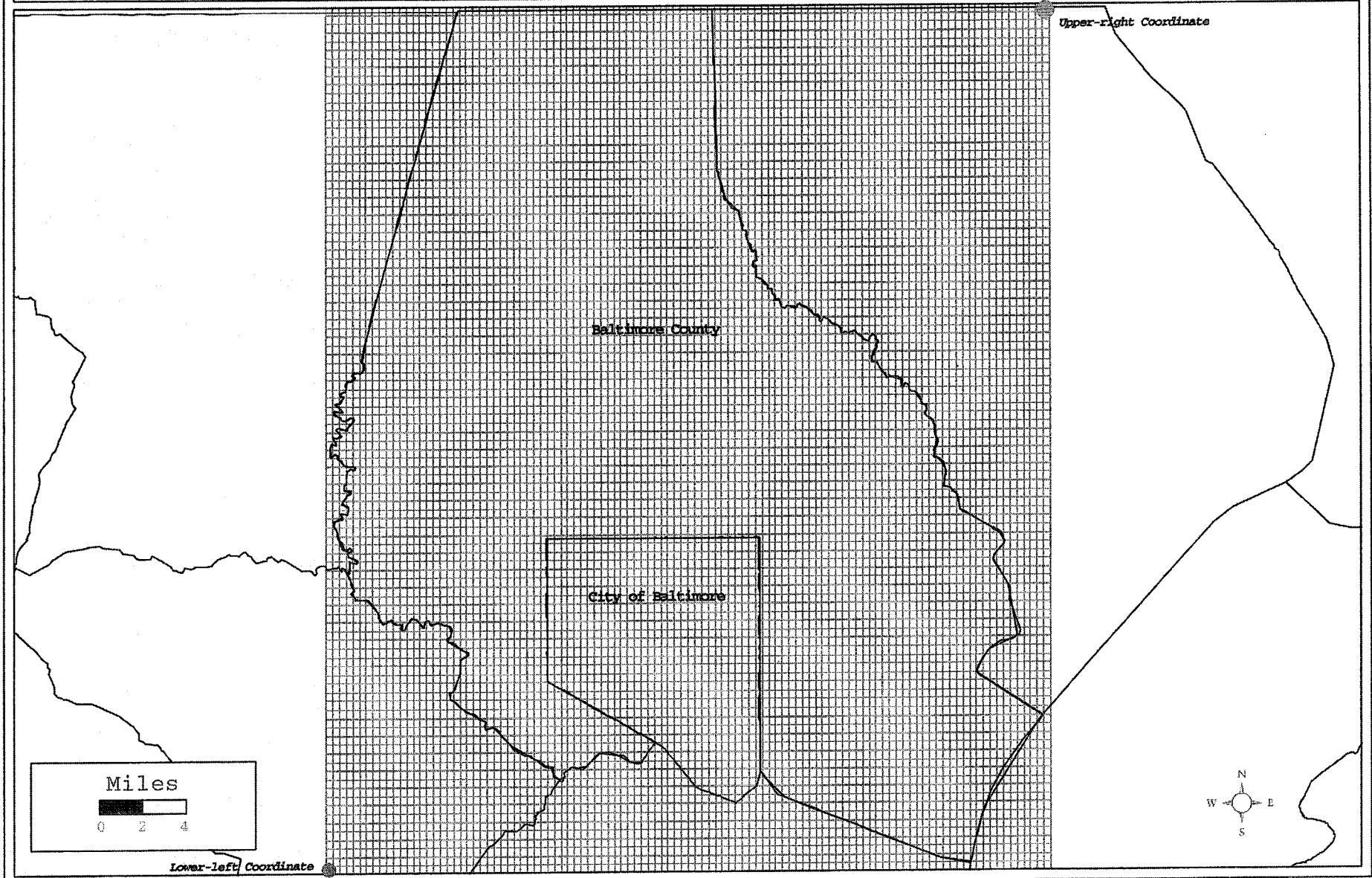
File to be Interpolated

The user must indicate whether the primary file or the secondary file (if used) is to be interpolated.

Method of Interpolation

The user must indicate the method of interpolation. Five types of kernel density estimators are used:

1.   Normal distribution (bell; default)
2.   Uniform (flat) distribution
3.   Quartic (spherical) distribution
4.   Triangular (conical) distribution
5.   Negative exponential (peaked) distribution

In our experience, there are advantages to each. The normal distribution produces an estimate over the entire region whereas the other four produce estimates only for the circumscribed bandwidth radius. If the distribution of points is sparse towards the outer parts of the region, then the four circumscribed functions will not produce estimates for those areas, whereas the normal will. Conversely, the normal distribution can cause some edge effects to occur (e.g., spikes at the edge of the reference grid), particularly if there are many points near one of the boundaries of the study area. The four circumscribed functions will produce less of a problem at the edges, although they still can produce some spikes. Within the four circumscribed functions, the uniform and quartic tend to smooth the data more whereas the triangular and negative exponential tend to emphasize 'peaks' and 'valleys'. The differences between these different kernel functions are small, however. The user should probably start with the default normal function and adjust accordingly to how the surface or contour looks.

Choice of Bandwidth

The user must indicate how bandwidths are to be defined. There are two types of bandwidth for the single kernel density routine, fixed interval or adaptive interval.

Fixed interval

With a fixed bandwidth, the user must specify the interval to be used and the units of measurement (squared miles, squared nautical miles, squared feet, squared kilometers, or squared meters). Depending on the type of kernel estimate used, this interval has a slightly different meaning. For the normal kernel function, the bandwidth is the standard deviation of the normal distribution. For the uniform, quartic, triangular, or negative exponential kernels, the bandwidth is the radius of the search area to be interpolated.

There are few guidelines for choosing a particular bandwidth other than by visual inspection (Venables and Ripley, 1997). Some have argued that the bandwidth be no larger

314

than the finest resolution that is desired and others have argued for a variation on random nearest neighbor distances (see Spencer Chainey application later in this chapter). Others have argued for particular sizes (Silverman, 1986; Härdle, 1991; Kadafar, 1996; Farewell, 1999; Talbot, Kulldorff, Forand, and Haley, 2000).[1] There does not seem to be consensus on this issue. Consequently, *CrimeStat* leaves the definition up to the user.

Typically, a narrower bandwidth interval will lead to a finer mesh density estimate with all the little peaks and valleys. A larger bandwidth interval, on the other hand, will lead to a smoother distribution and, therefore, less variability between areas. While smaller bandwidths show greater differentiation among areas (e.g., between 'hot spot' and 'low spot' zones), one has to keep in mind the statistical precision of the estimate. If the sample size is not very large, then a smaller bandwidth will lead to more imprecision in the estimates; the peaks and valleys may be nothing more than random variation. On the other hand, if the sample size is large, then a finer density estimate can be produced. In general, it is a good idea to experiment with different fixed intervals to see which results make the most sense.

Adaptive interval

An adaptive bandwidth adjusts the bandwidth interval so that a minimum number of points are found. This has the advantage of providing constant precision of the estimate over the entire region. Thus, in areas that have a high concentration of points, the bandwidth is narrow whereas in areas where the concentration of points is more sparse, the bandwidth will be larger. This is the default bandwidth choice in *CrimeStat* since we believe that consistency in statistical precision is paramount. The degree of precision is generally dependent on the sample size of the bandwidth interval. The default is a minimum of 100 points within the bandwidth radius. The user can make the estimate more fine grained by choosing a smaller number of points (e.g., 25) or more smooth by choosing a larger number of points (e.g., 200). Again, experimentation is necessary to see which results make the most sense.

Output Units

The user must indicate the measurement units for the density estimate in points per squared miles, squared nautical miles, squared feet, squared kilometers, or squared meters. The default is points per square mile.

Intensity or Weighting Variables

If an intensity or weighting variable is used, these boxes must be checked. Be careful about using both an intensity and a weighting variable to avoid 'double weighting'.

315

Density Calculations

The user must indicate the type of output for the density estimates. There are three types of calculation that can be conducted with the kernel density routine. The calculations are applied to each reference cell:

1. The kernel estimates can be calculated as *absolute density* estimates using formulas 8.1-8.9, depending on what type of kernel function is used. The estimates at each reference cell are re-scaled so that the sum of the densities over all reference grids equals the total number of incidents; this is the default value.

2. The kernel estimates can be calculated as *relative density* estimates. These divide the absolute densities by the area of the grid cell. It has the advantage of interpreting the density in terms that are familiar. Thus, instead of a density estimate represented by points per grid cell, the relative density will convert this to points per, say, square mile.

3. The densities can be converted into *probabilities* by dividing the density at any one cell by the total number of incidents.

Since the three types of calculation are directly interrelated, the output surface will not differ in its variability. The choice would depend on whether the calculations are used to estimate absolute densities, relative densities, or probabilities. For comparisons between different types of crime or between the same type of crime and different time periods, usually absolute densities are the unit of choice (i.e., incidents per grid cell). However, to express the output as a probability, that is, the likelihood that an incident would occur at any one location, then outputing the results as probabilities would make more sense. For display purposes, however, it makes no difference as both look the same.

Output Files

Finally, the results can be displayed in an output table or can be output into two formats: 1) Raster grid formats for display in a surface mapping program- *Surfer for Windows* '.dat' format (Golden Software, 1994) or *ArcView Spatial Analyst* 'asc' format (ESRI, 1998); or 2) Polygon grids in *ArcView* '.shp', *MapInfo* '.mif' or *Atlas*GIS* '.bna' formats.[2] However, all but *Surfer for Windows* require that the reference grid be created by *CrimeStat*.

Example 1: Kernel Density Estimate of Street Robberies

An example can illustrate the use of the single kernel density routine. Figure 8.9 shows a *Surfer for Windows* output of the 1180 street robberies for 1996 in Baltimore County. The reference grid was generated by *CrimeStat* and had 100 columns and 108 rows. Thus, the routine calculated the distance between each of the 10,800 reference cells and the 1180 robbery incident locations, evaluated the kernel function for each measured

316

Figure 8.9:

Baltimore County Robberies: 1996-97

Kernel Density Interpolation

Surface View
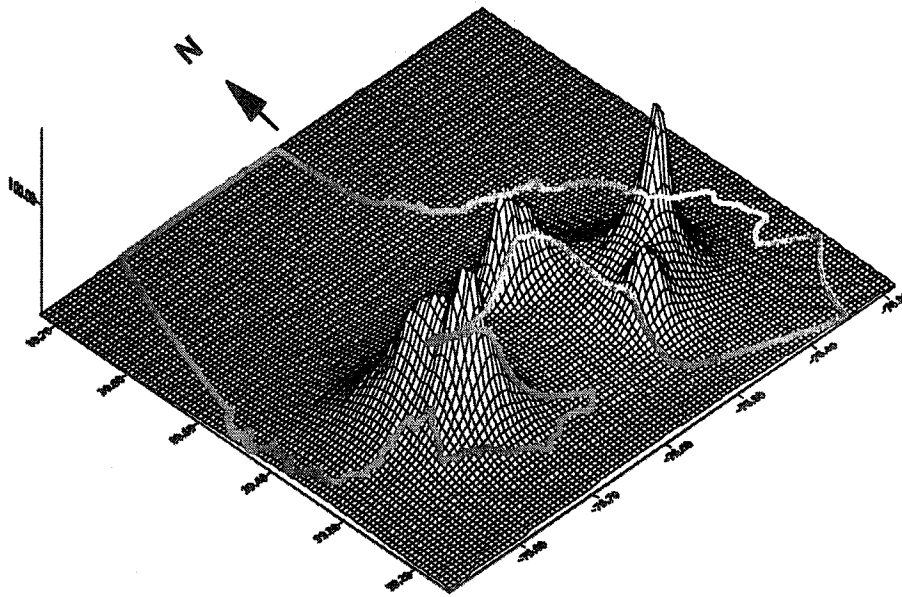
Contour View
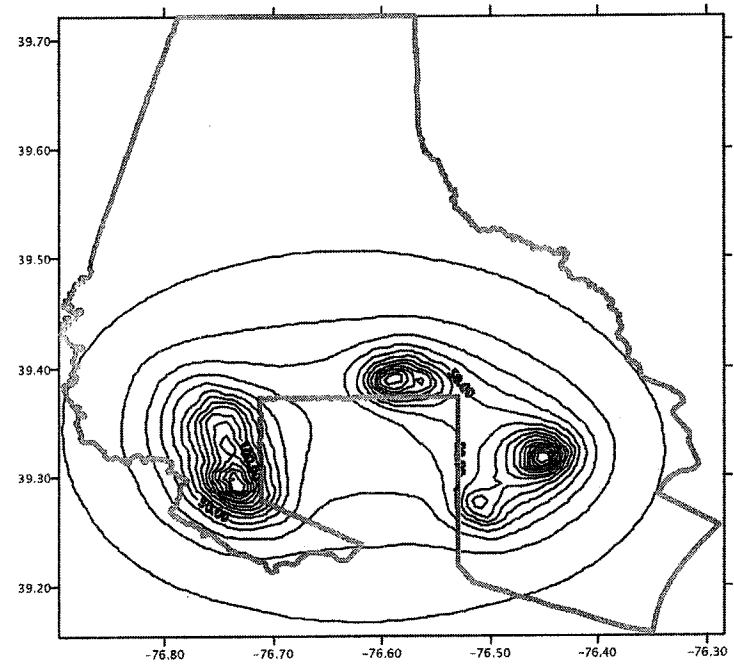
Ground Level View
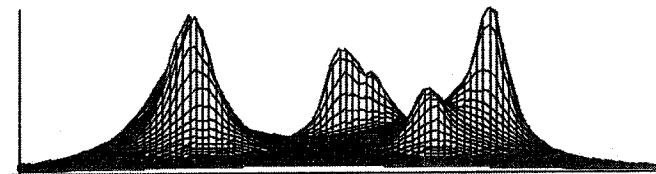
distance, and summed the results for each reference cell. The normal distribution kernel function was selected for the kernel estimator and an adaptive bandwidth with a minimum sample size of 100 was chosen as the parameters.

There are three views in the figure: 1) a map view showing the location of the incidents; 2) a surface view showing a three-dimensional interpolation of robbery density; and 3) a contour view showing contours of high robbery density. The surface and contour views provide different perspectives. The surface shows the peaks very clearly and the relative density of the peaks. As can be seen, the peak for robberies on the eastern part of the County is much higher than the two peaks in the central and western parts of the County. The contour view can show where these peaks are located; it is difficult to identify location clearly from a three-dimensional surface map. Highways and streets could be overlaid on top of the contour view to identify more precisely where these peaks are located.

Figure 8.10 shows an *ArcViewSpatial Analyst* map of the robbery density with the robbery incident locations overlaid on top of the density contours. Here, we can see quite clearly that there are three strong concentrations of incidents, one spreading over a distance of several miles on the west side, one on northern border between Baltimore City and Baltimore County, and one on the east side; there is also one smaller peak in the southeast corner of the County.

From a statistical perspective, the kernel estimate is a better 'hot spot' identifier than the cluster analysis routines discussed in chapter 6. Cluster routines group incidents into clusters and distinguish between incidents which belong to the cluster and those which do not belong. Depending on which mathematical algorithms are used, different clustering routines will return differing allocations of incidents to clusters. The kernel estimate, on the other hand, is a continuous surface; the densities are calculated at *all* locations; thus, the user can visually inspect the variability in density and decide what to call a 'hot spot' without having to define arbitrarily where to cut-off the 'hot spot' zone.

Going back to the *Surfer for Windows* output, figure 8.11 shows the effects of varying the bandwidth parameters. There are three fixed bandwidth intervals (0.5, 1, and 2 miles respectively) and there are two adaptive bandwidth intervals (a minimum of 25 and 100 points respectively). As can be seen, the fineness of the interpolation is affected by the bandwidth choice. For the three fixed intervals, an interval of 0.5 miles produces a finer mesh interpolation than an interval of 2 miles, which tends to 'oversmooth' the distribution. Perhaps, the intermediate interval of 1 mile gives the best balance between fineness and generality. For the two adaptive intervals, the minimum sample size of 25 gives some very specific peak locations whereas the adaptive interval with a minimum sample size of 100 gives a smoother distribution.

Which of these should be used as the *best* choice would depend on how much confidence the analyst has in the results. A key question is whether the 'peaks' are real or merely byproducts of small sample sizes. The best choice would be to produce an interpolation that fits the experience of the department and officers who travel an area.

318

Figure 8.10
Baltimore County Street Robberies: 1996
Kernel Density Estimate

Legend:
- Baltimore Beltway
- Major Road
- City of Baltimore
- Baltimore County

Robbery Density
- Low
- High

0   10   20 Miles
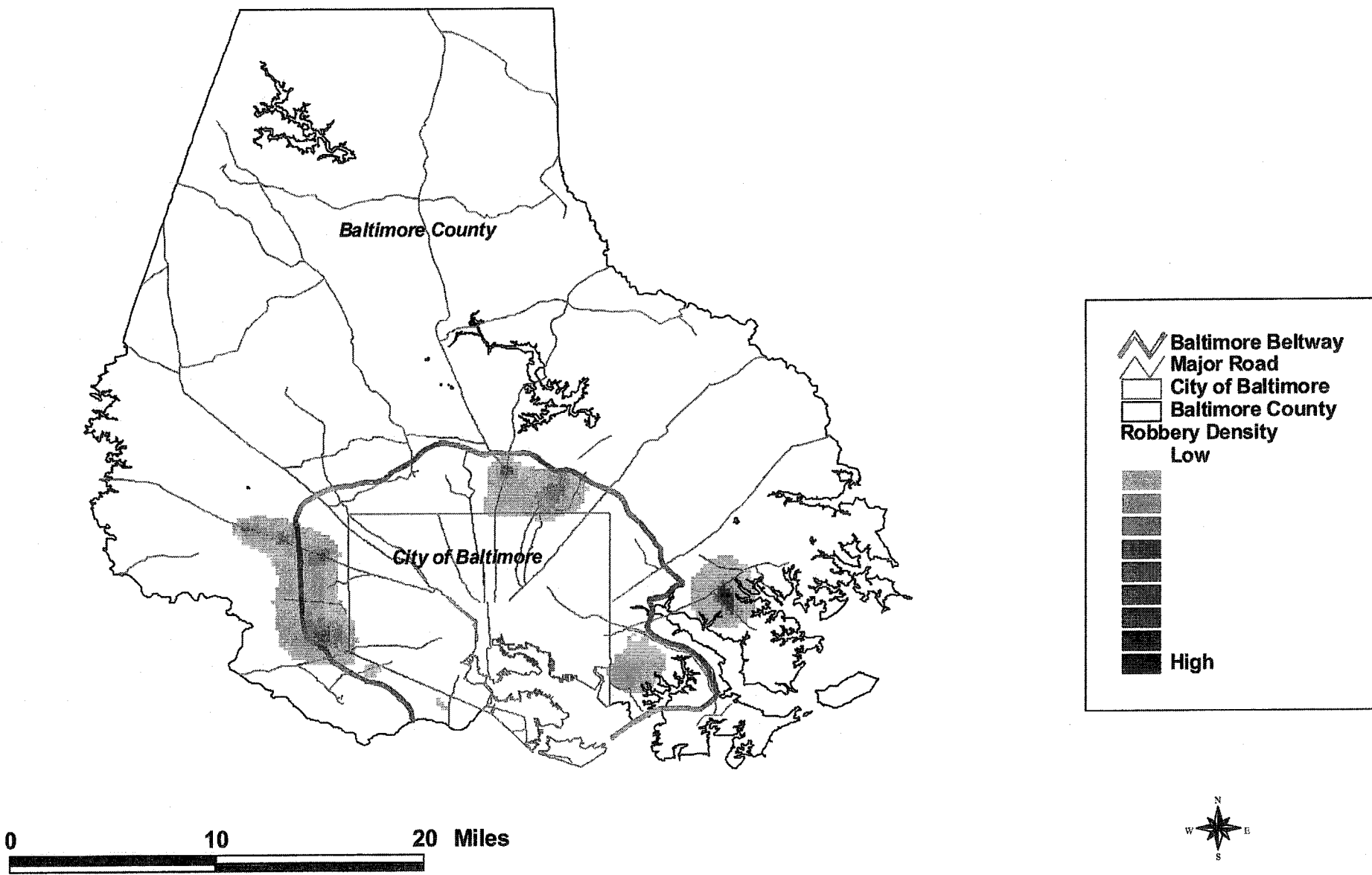
Figure 8.11:

Interpolation of Baltimore County Auto Thefts: 1996

Different Smoothing Parameters

Fixed/ h=0.5 mi

Fixed/ h=1.0 mi

Fixed/ h=2.0 mi

Adaptive/ n=50

Adaptive/ n=100
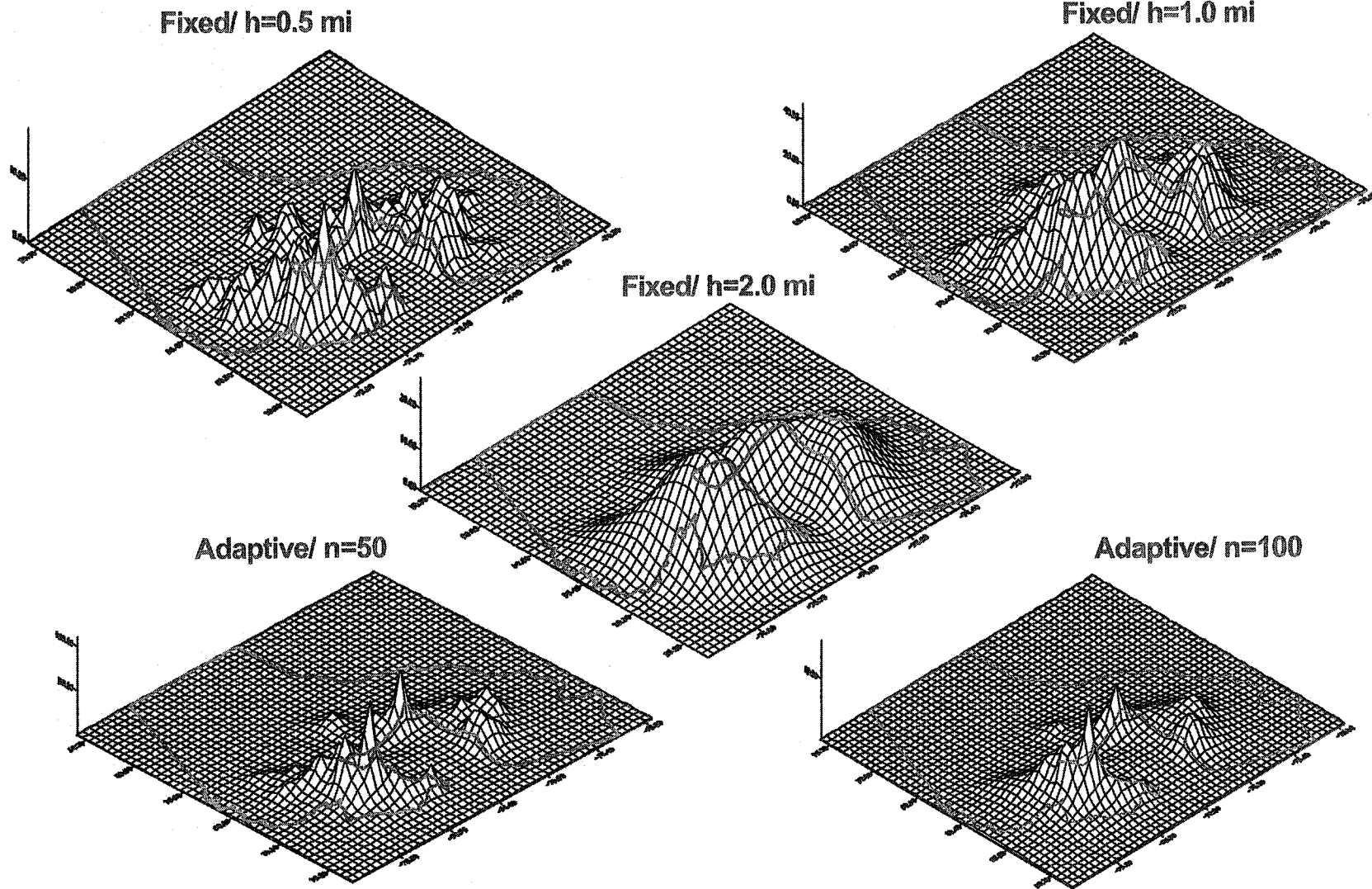
Kernel Density Interpolation to Estimate Sampling Bias in the Climatic Response of *Sphagnum* Spores in North America

Mike Sawada
Laboratory for Paleoclimatology and Climatology
University of Ottawa, Department of Geography, Canada

Sphagnum moss, the dominant species of bogs, thrives under certain ranges of temperature and precipitation. *Sphagnum* releases spores for reproduction and these are transported, often long distances, by wind and water. Thus, the presence of a spore in the fossil record may not indicate nearby *Sphagnum* plants. However, spores should be most numerous near *Sphagnum* plants. Over time, these spores and pollen from other plants accumulate in lake and bog sediments and leave a fossil record of vegetation history.

We wanted to use the amount of fossil *Sphagnum* spores in different parts of North America to infer past climates. To do so, we had to first show that *Sphagnum* spores are most abundant in climates where *Sphagnum* plants thrive and secondly, that this center of abundance is not biased sampling because of under sampling in parts of climate space. First, we developed a *Sphagnum* spore response surface showing the relative abundance of spores along the axes of temperature and precipitation (Fig. A).

CrimeStat was used in the second stage to develop a kernel density surface using a quartic kernel for 3007 sample sites within climate space (Fig. B). These were smoothed and visualized in *Surfer*. The surface showed that the intensity of points is higher in regions surrounding the response maximum. This gave us confidence that the *Sphagnum* response was real since other parts of climate space are well sampled but unlikely to produce high spore proportions. This fact allowed climate inferences to be made within the fossil record for past time periods using the amount of *Sphagnum* spores present.
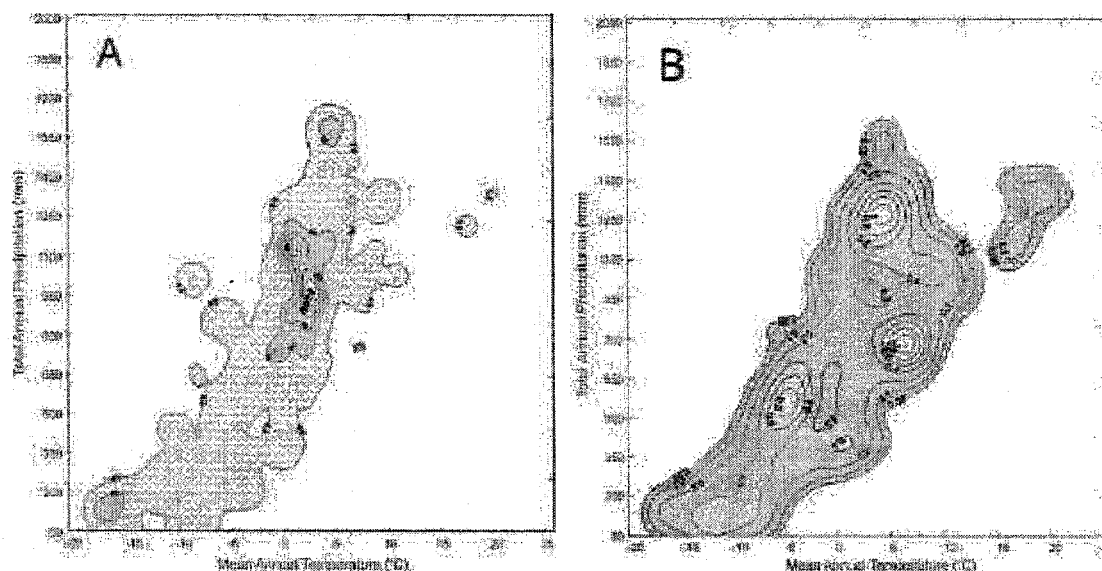
Figures modified from Gajewski, Viau, Sawada et al. 2001. Global Biogeochemical Cycles,

Describing Crime Spatial Patterns By Time of Day

Renato Assunção, Cláudio Beato, Bráulio Silva
CRISP, Universidade Federal de Minas Gerais, Brazil

We used the kernel density estimate to visualize time trends for crime occurrences on a typical weekday. We found markedly different spatial distributions depending on the time, with the amount of crime varying and the hot spots, identified by the ellipses, appearing in different places.

The analysis used 1114 weekday robberies from 1995 to 2000 in downtown Belo Horizonte. Breaking the data into hours, we used the normal kernel, a fixed bandwidth of 450 meters and outputted densities option (points per square unit of area). Note that the latter option could be useful if one is interested only in the hot spot locations, and not in the distribution during the day. To make the ellipses, we used the nearest neighbor hierarchical spatial clustering technique with a minimum of 35 incidents. We output the results to *MapInfo*, keeping the same scale for all maps. Four of them are shown below.

9:00 AM

1:00 PM

Legend
- High density
- Medium density
- Low density
- Very low density
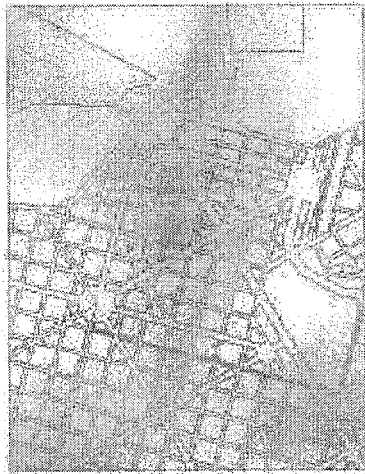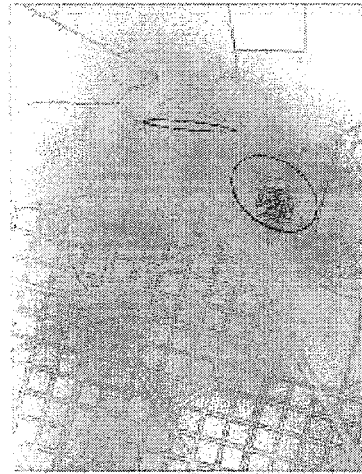
7:00 PM
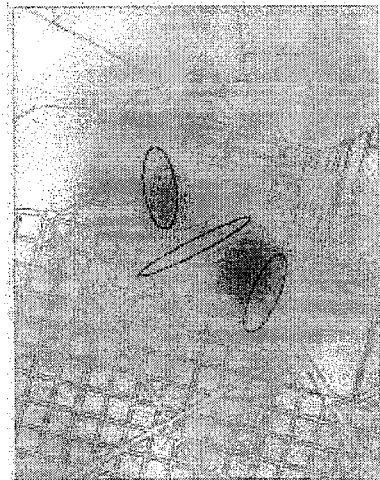
11:00 PM
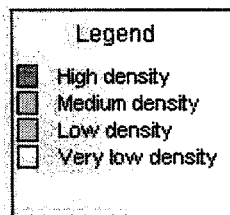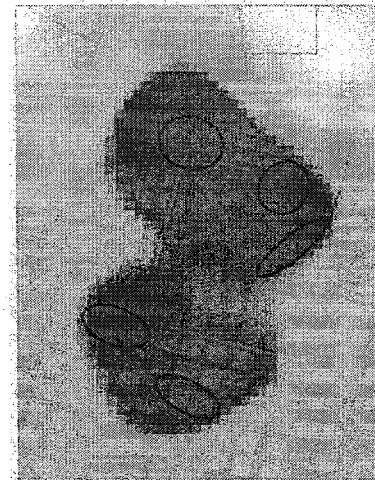
Using Kernel Density Smoothing and Linking to *ArcView*: Examples from London, England

Spencer Chainey
InfoTech Enterprises Europe
London, England

CrimeStat offers an effective method for creating kernel density surfaces. The example below uses residential burglary incidents in the London Borough of Croydon, England for the period June 1999 – May 2000 (N=3104). The single kernel routine was used to produce a kernel density surface representing the distribution of residential burglary.

The kernel function used was the quartic, which is favoured by most crime mappers as it applies added weight to crimes closer to the centre of the bandwidth. Rather than choosing an arbitary interval it is useful to use the mean nearest neighbour distance for different orders of K, which can be calculated by *CrimeStat* as part of a nearest neighbour analysis. For the Croydon data, an interval of 269 metres was chosen, which relates to a mean nearest neighbour distance at a K-order of 13. The output units were densities in square kilometres and was output to *ArcView*.

Kernel density estimation is a particularly useful method as it helps to precisely identify the location, spatial extent and intensity of crime hotspots. It is also visually attractive, so helping to invoke further enquiry and the reasoning behind why crime and disorder is concentrated. The density surface that is created can reflect the distribution of incidents against the natural geography of the area of interest, including representing the natural boundaries, such as reservoirs and lakes, or an alignment that follows a particular street in which there is a high concentration of offending. The method is also less subjective if clear guidelines are followed for the setting of parameters.

Residential burglary hotspots (by volume) in the London Borough of Croydon, England
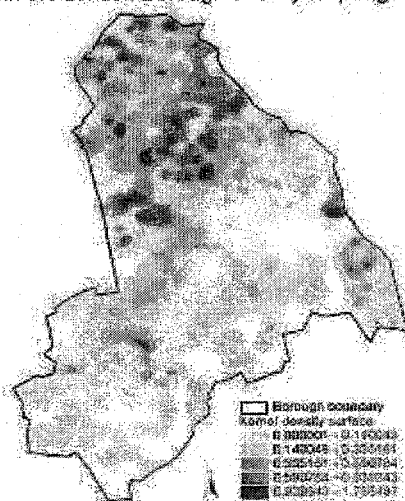
Again, experimentation and discussions with beat officers will be necessary to establish which bandwidth choice should be used in future interpolations.

Note in all five of the interpolations, there is some bias at the edges with the City of Baltimore (the three-sided area in the central southern part of the map). Since the primary file only included incidents for the County, the interpolation nevertheless has estimated some likelihood at the edges; these are *edge biases* and need to be ignored or removed with an ASCII editor.[3] Further, the wider the interval chosen, the more bias is produced at the edge.

Dual Kernel Estimates

The dual kernel density routine in *CrimeStat* is applied to *two* distributions of point locations. For example, the primary file could be the location of auto thefts while the secondary file could be the centroids of census tracts, with the population of the census tract being an intensity variable. The dual routine must be used with *both* a primary file *and* a secondary file. Also, it is necessary to define a reference file, either an existing file or one generated by *CrimeStat* (see chapter 3). Several parameters need to be defined.

File to be Interpolated

The user must indicate the order of the interpolation. The routine uses the language *first* file and *second* file in making the comparison (e.g., dividing the first file by the second; adding the first file to the second). The user must indicate which is the first file, the primary or the secondary. The default is that the primary file is the first file.

Method of Interpolation

The user must indicate the type of kernel estimator. As with the single kernel density routine, five types of kernel density estimators are used

4. Normal distribution (bell; default)
5. Uniform (flat) distribution
6. Quartic (spherical) distribution
7. Triangular (conical) distribution
8. Negative exponential (peaked) distribution

In our experience, there are advantages to each. The normal distribution produces an estimate over the entire region whereas the other four produce estimates only for the circumscribed bandwidth radius. If the distribution of points is sparse towards the outer parts of the region, then the four circumscribed functions will not produce estimates for those areas, whereas the normal will. Conversely, the normal distribution can cause some edge effects to occur (e.g., spikes at the edge of the reference grid), particularly if there are many points near one of the boundaries of the study area. The four circumscribed functions will produce less of a problem at the edges, although they still can produce some spikes. Within the four circumscribed functions, the uniform and quartic tend to smooth

324

the data more whereas the triangular and negative exponential tend to emphasize 'peaks' and 'valleys'. The differences between these different kernel functions are small, however. The user should probably start with the default normal function and adjust accordingly to how the surface or contour looks.

Choice of Bandwidth

The user must define the bandwidth parameter. There are three types of bandwidths for the single kernel density routine - fixed interval, variable interval, or adaptive interval.

Fixed interval

With a fixed bandwidth, the user must specify the interval to be used and the units of measurement (squared miles, squared nautical miles, squared feet, squared kilometers, or squared meters). Depending on the type of kernel estimate used, this interval has a slightly different meaning. For the normal kernel function, the bandwidth is the standard deviation of the normal distribution. For the uniform, quartic, triangular, or negative exponential kernels, the bandwidth is the radius of the search area to be interpolated. Since there are two files being compared, the fixed interval is applied both to the first file and the second file.

Variable interval

With a variable interval, each file (the first and the second) have different intervals. For both, the units of measurements must be specified (squared miles, squared nautical miles, squared feet, squared kilometers, or squared meters). There is a good reason why a user might want variable intervals. In comparing two kernel estimates, the most common comparison is to divide one by the other. However, if the density estimate for a particular cell in the denominator approaches zero, then the ratio will blow up and become a very large number. Visually, this will be seen as spikes in the distribution, the result, usually, of too few cases. In this case, the user might decide to smooth the denominator more than numerator in order to reduce these spikes. For example, the interval for the first file (the numerator) could be 1 mile whereas the interval for the second file (the denominator) could be 3 miles. Experimentation will be necessary to see whether this is warranted. But, in our experience, it frequently happens when either there are two few cases or there is an irregular boundary to the region with a number of incidents grouped at one of the edges.

Adaptive interval

An adaptive bandwidth adjusts the bandwidth interval so that a minimum number of points (sample size) is found. This sample size is applied to both the first file and the second file. It has the advantage of providing constant precision of the kernel estimate over the entire region. Thus, in areas that have a high concentration of points, the bandwidth is narrow whereas in areas where the concentration of points is more sparse, the bandwidth will be larger. This is the default bandwidth choice in *CrimeStat* since

325

consistency in statistical precision is important. The degree of precision is generally dependent on the sample size of the bandwidth interval. The default is a minimum of 100 points. The user can make the estimate finer by choosing a smaller number of points (e.g., 25) or smoother by choosing a larger number of points (e.g., 200).

Use kernel bandwidths that produce stable estimates

Note: with a duel kernel calculation, particularly the ratio of one variable to another, be careful about choosing a very small bandwidth. This could have the effect of creating spikes at the edges of the study area or in low population density areas. For example, in low population density areas, there will probably be fewer events than in more built-up area. For the denominator of a ratio estimate, an extremely low value could cause the ratio to be exaggerated (a 'spike') relative to neighboring grid cells. Using a larger bandwidth will produce a more stable average.

Output Units

The user must indicate the measurement units for the density estimate in points per squared miles, squared nautical miles, squared feet, squared kilometers, or squared meters.

Intensity or Weighting Variables

If an intensity or weighting variable is used for either the first file or the second file, these boxes must be checked. Be careful about using both an intensity and a weighting variable to avoid 'double weighting'.

Density Calculations

The user must indicate the type of density output. There are six types of density calculations that can be conducted with the dual kernel density routine. The calculations are applied to each reference cell:

1.   There is the *ratio of densities*, that is the first file divided by the second file. This is the default choice. For example, if the first file is the location of auto thefts incidents and the second file is the location of census tract centroids with the population assigned as an intensity variable, then ratio of densities would divide the kernel estimate for auto thefts by the kernel estimate for population and would be an estimate of auto thefts risk.

2.   There is also the *log ratio of densities*. This is the natural logarithm of the density ratio, that is

Log ratio of densities   =   Ln [$g(x_j) / g(y_j)$]   (8.10)

326

where $g(x_j)$ is the density estimate for the first file and $g(y_j)$ is the density estimate for the second file. For a variable that has a spatially skewed distribution, such that most reference cells have very low density estimates, but a few have very high density estimates, converting the ratio into a log function will tend to mute the spikes that occur. This measure has been used in studies of risk (Kelsall and Diggle, 1995b).

3. There is the *absolute difference in densities*, that is the first file minus the second file. This can be a useful output for examining differential effects. For example, by using the centroids of census block groups (see example 2 below) with the population of the census block group assigned as an intensity or weighting variable, there is a slight bias produced by the spatial arrangements of the block groups. The U. S. Census Bureau suggests that census units (e.g., census tracts, census block groups) be drawn so that there are approximately equal populations in each unit. Thus, block groups towards the center of the metropolitan area tend to be smaller because there is a higher population density at those locations. Thus, the spatial arrangement of the block groups will tend to produce a kernel estimate which has a higher value towards the center independent of the actual population of the block group; the bias is very small, less than 0.1%, but it does exist. A more precise estimate could be produced by subtracting the kernel estimate for the block group centroids *without* using population as the intensity variable from the kernel estimate for the block group centroids *with* population as the intensity variable. The resulting output could then be read back into *CrimeStat* and used as a more precise measure of population distribution. There are other uses of the difference function, such as subtracting the estimate for the population-at-risk from the incident distribution rather than taking the ratio or by calculating the net change in population between two censuses.

4. There is the *relative difference in densities*. Like the relative density in the single-kernel routine (discussed above), the relative difference in densities first standardizes the densities of each file by dividing by the grid cell area and then subtracts the secondary file relative density from the primary file relative density. This can be useful in calculating changes between two time periods, for example in calculating a change in relative density between two censuses or a change in the crime density between two time periods.

5. There is the *sum of the densities*, that is, the density estimate for the first file plus the density estimate for the second file. Again, this is applied to each reference cell at a time. A possible use of the sum operation is to combine two different density surfaces, for example the density of robberies plus the density of assaults;

6. Finally, there is the *relative sum of densities* between the primary file and the secondary file. The relative sum of densities first standardizes the

327

densities of each file by dividing by the grid cell area and then subtracts the secondary file relative density from the primary file relative density. This can be useful for identifying the total effects of two distributions. For example, the total impact of robberies and burglaries on an area can be estimated by taking the relative density of robberies and adding it to the relative density of burglaries. The result is the combined relative density of robberies and burglaries per unit area (e.g., robberies and burglaries per square mile).

Output Files

Finally, the user must specify the file formats for the output. The results can be output in three forms. First, the results are displayed in an output table. Second, the results can be output into two raster grid formats for display in a surface mapping program: *Surfer for Windows* format as a '.dat' file (Golden Software, 1994) and *ArcView Spatial Analyst* format as a '.asc' file (ESRI, 1998). Third, the results can be output as polygon grids into *ArcView* '.shp', *MapInfo* '.mif' and *Atlas*GIS* '.bna' format (see footnote 1). All but *Surfer for Windows* require that the reference grid be created by *CrimeStat*.

Example 2: Kernel Density Estimates of Vehicle Thefts Relative to Population

As an example of the use of the dual kernel density routine, the duel routine is applied in both the City of Baltimore and the County of Baltimore to 14,853 motor vehicle theft locations for 1996 relative to the 1990 population of census block groups. Again, a reference grid of 100 columns by 108 rows was generated by *CrimeStat*.

Figure 8.12 shows the resulting density estimate as a *Surfer for Windows* output; again, there is a map view, a surface view, and a contour view. The normal kernel function was used and an adaptive bandwidth of 100 points was selected. As seen, there is a very high concentration of auto theft incidents within the central part of the metropolitan area. The contour view suggest five or six peak areas that are close to each other.

Much of this concentration, however, is produced by high population density in the metropolitan center. Figure 8.13, for example, shows the kernel estimate for 1349 census block groups for both the City of Baltimore and the County of Baltimore with the 1990 population assigned as the intensity variable. Again, the normal kernel function was used with an adaptive bandwidth of 100 points being selected. The map shows three views: 1) a surface view; 2) a contour view; and 3) a ground level view looking directly north. The distribution of population is, of course, also highly concentrated in the metropolitan center with two peaks, quite close to each other with several smaller peaks.

When these two kernel estimates are compared using the dual kernel density routine, a more complicated picture emerges (figure 8.14). This routine has conducted three operations: 1) it calculated the distance between each of the 10,800 reference cells and the 14,853 auto theft locations, evaluated the kernel function for each measured

328

Figure 8.12:

Baltimore County Vehicle Thefts: 1996

Kernel Density Interpolation

Surface View

Contour View

Ground Level View
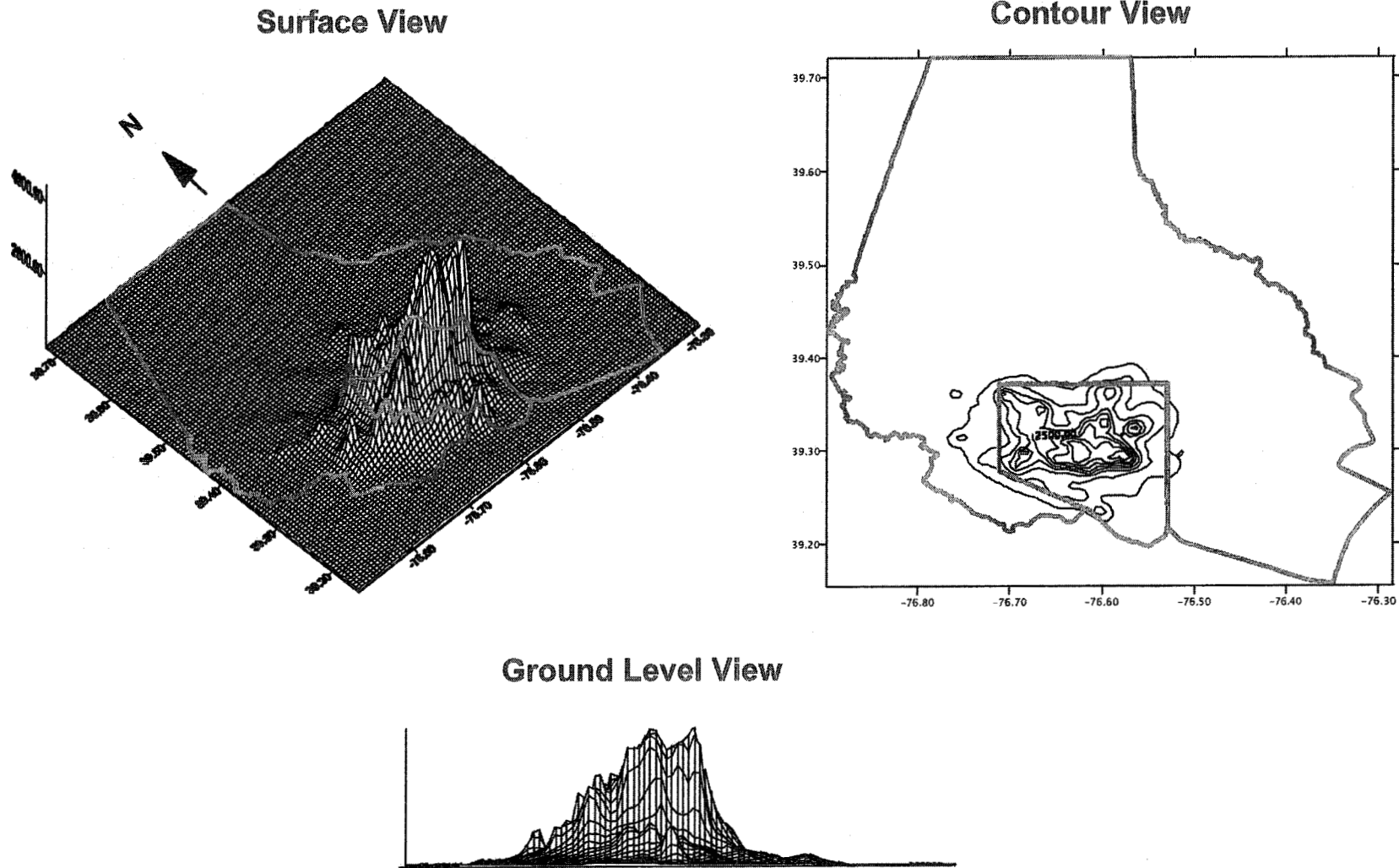
Figure 8.13:

Baltimore Metropolitan Population: 1990
Kernel Density Estimate of Block Group Population

Surface View

Contour View

Ground Level View
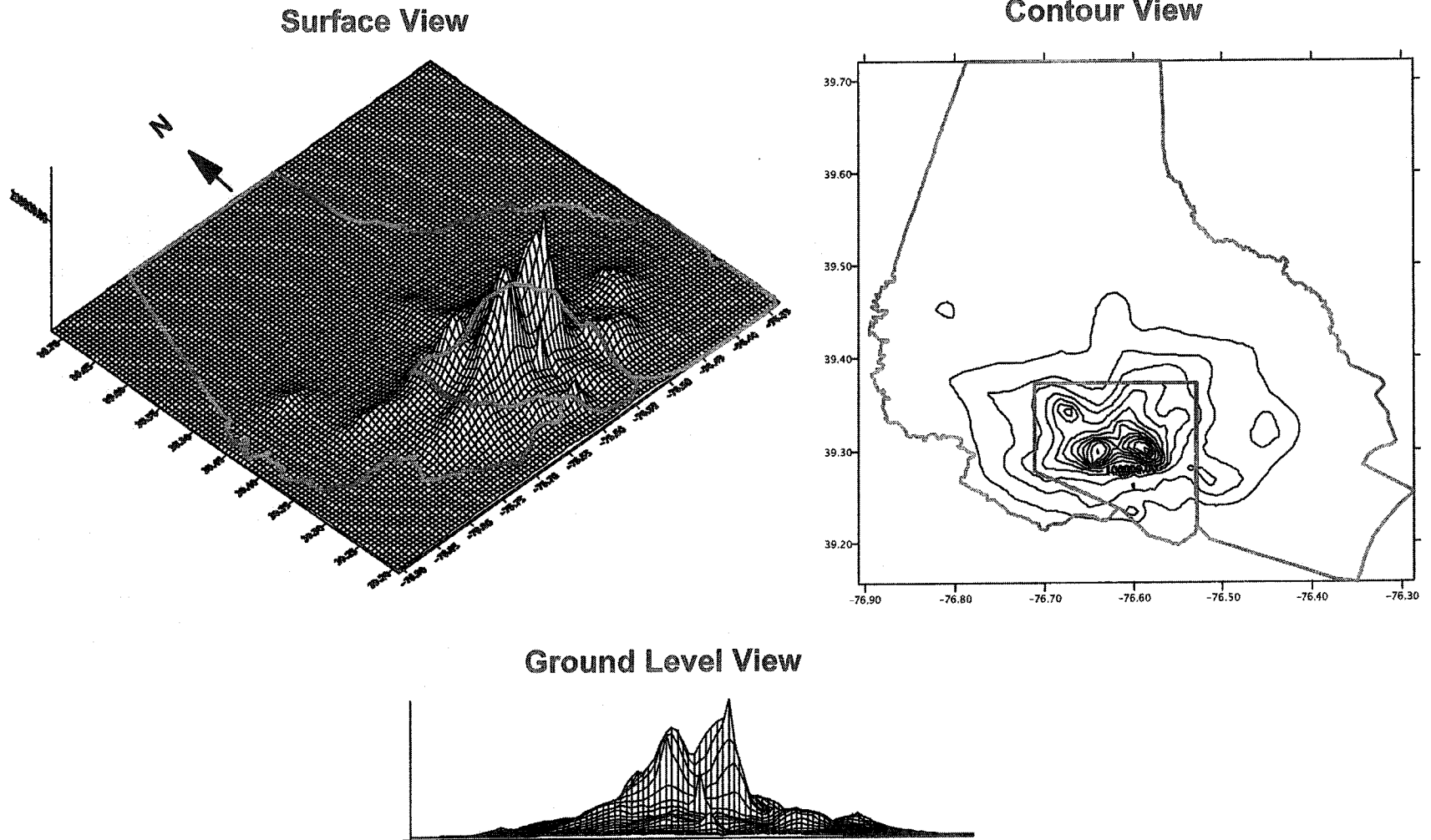
Figure 8.14:

Baltimore County Vehicle Theft Risk

Kernel Density Ratio of 1996 Vehicle Thefts to 1990 Population

Surface View

Contour View

Ground Level View
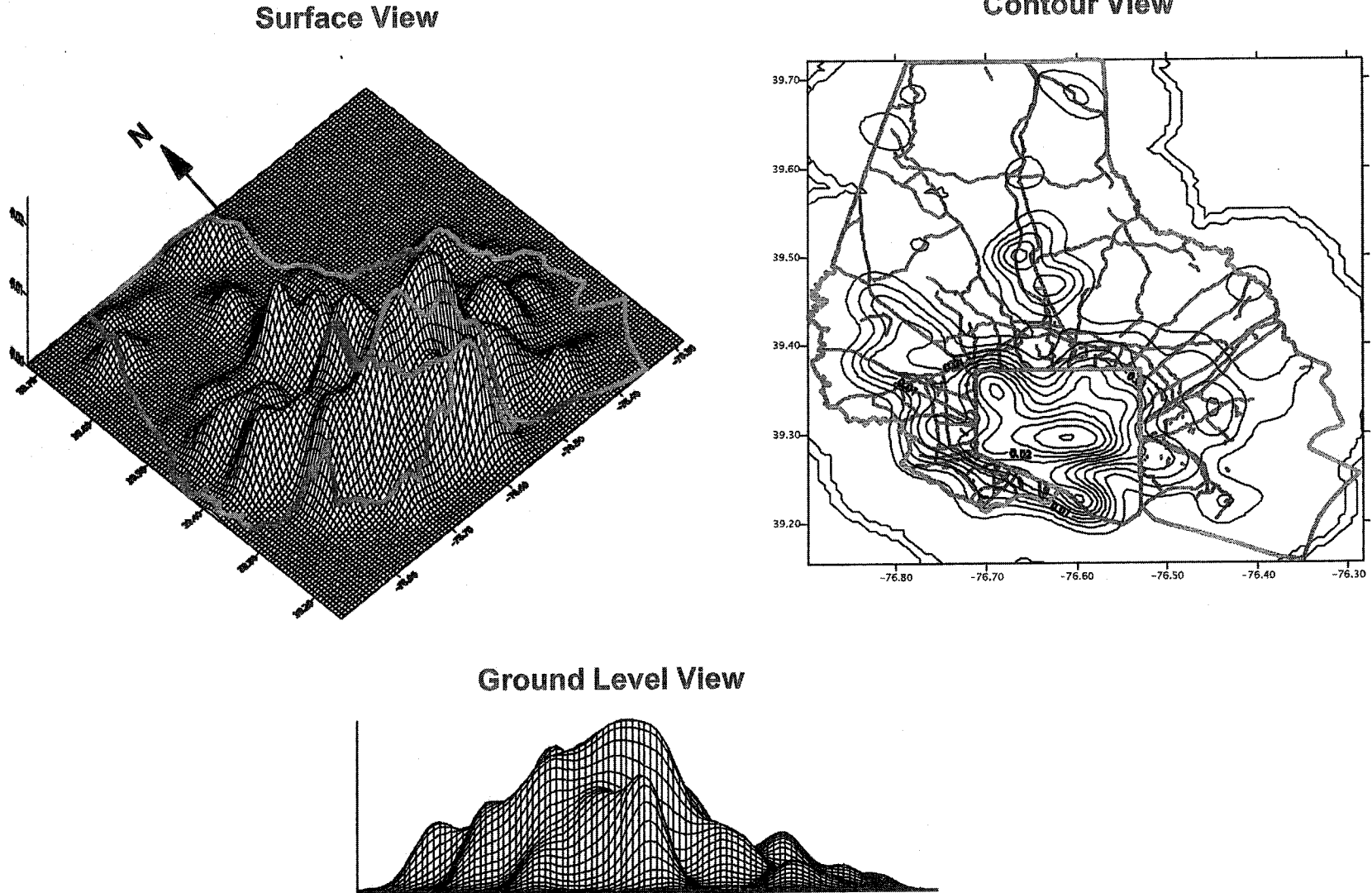
The Risk of Violent Incidents Relative to Population Density in Cologne Using the Dual Kernel Density Routine

Dietrich Oberwittler and Marc Wiesenhütter
Max Planck Institute for Foreign and International Criminal Law
Freiburg, Germany

When estimating the density of street crimes within a metropolitan area by interpolating crime incidents, the result is usually a very high concentration in the city center. However, there is also a very high concentration of people either living or pursuing their daily routine activities in these areas. The question emerges how likely is a criminal event when taking into account the number of people spending their time in these areas. The *CrimeStat* duel kernel density routine is able to estimate a ratio density surface of crime relative to the 'population at risk'.

In this example, data on 'calls to the police' for assault and battery from April 1999 to March 2000 (N=6363 calls) and population from Cologne were used. Exact information on the number of people spending their time in the city does not exist. Therefore, 1997 counts of passengers entering and leaving the public transport system at each of 550 stations and bus stops in the city was used as a proxy variable. The number of persons at each station or bus stop was assigned to adjacent census tracts and added to the resident population resulting in a crude measure of the 'population at risk'.

In the dual kernel routine, the density estimate of crime incidents is compared to the density estimate of the population at risk, defined by the centroids of census tracts with the number of persons as an intensity variable. We chose the normal method of interpolation and adaptive intervals with a minimum of five points. The adaptive bandwidth adjusts for the fact that there are fewer incidents and census tracts at the edges of the city, resulting in a relatively smoother density surface for the ratio. The results were output to *ArcView*.

The effect of adjusting the crime distribution for the underlying 'population at risk' becomes quite visible. Whereas the *concentration of crime* is highest in the city center (left map), the *crime risk* (right map) is in fact much higher in several more distant areas that are known for high concentrations of socially disadvantaged persons. Given the imperfect nature of the population data these results should be interpreted as a broad view on the distribution of crime risk that, nevertheless, has important policy implications.

Single kernel density of crime incidences
(assault & battery, Cologne 1999/2000)
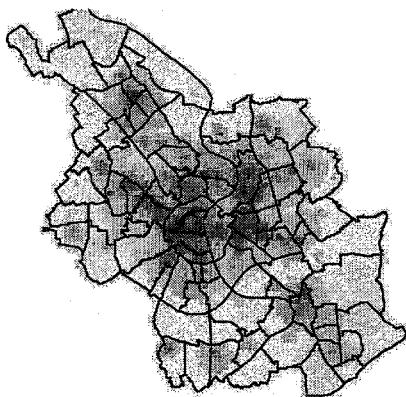
Dual kernel density of crime incidences
relative to population at risk
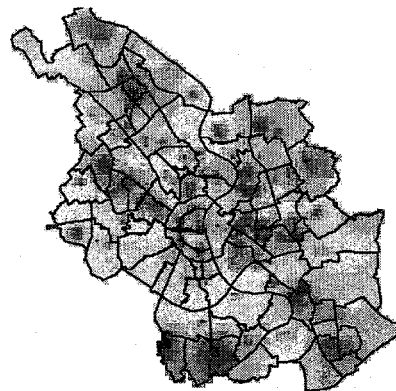
Kernel Density Interpolation of
Police Confrontations in
Buenos Aires Province, Argentina: 1999

Gastón Pezzuchi
Crime Analyst
Buenos Aires Province Police Force
Buenos Aires, Argentina

One of our first tryouts with the *CrimeStat* software involved the calculation of both single and dual kernel density interpolations using data on 1999 confrontations with the police within Buenos Aires Province, an area that covers 29 counties around the Federal Capital. The confrontations include mostly gun fights with the police but also other attacks (e.g., knives, rocks, sticks). In the last three years, there has been an increase in confrontations with the police. The single interpolation shows a density surface that gives a good picture of the ongoing level of violence while the duel interpolations shows a risk surface using the personnel deployment data; the latter are confrontations relative to the number of police deployed. Typically, police are allocated to areas according to crime rates.\

Example: Kernel Density Estimation (CrimeStat)

Single Interpolation - Density of Events | Dual Interpolation - Ratio of densities (Risk)

Events = Police shootings (aprox. 800)
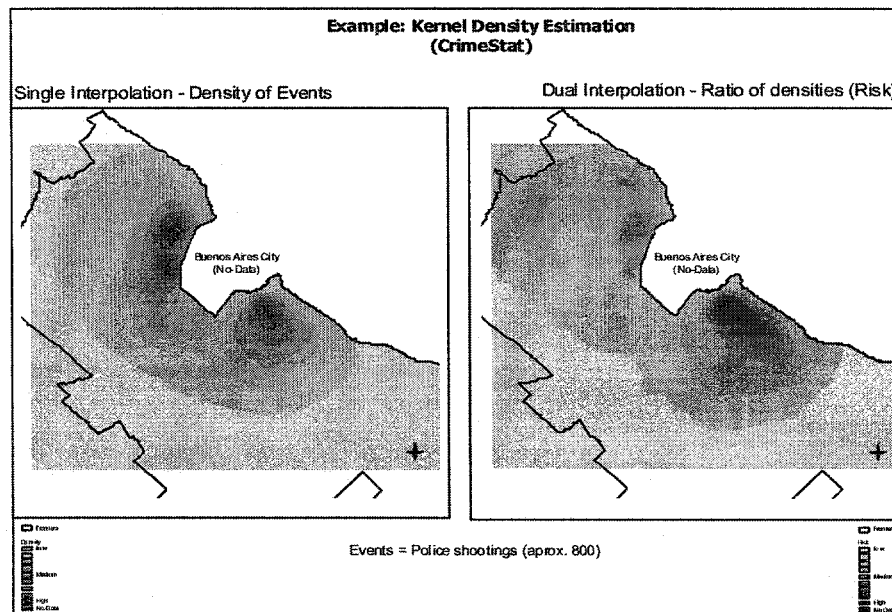
Both images are quite different, suggesting varying policing strategies. For example, though there are two well-defined hot spot areas in the Province (one in the north, the other in the south), the high levels of risk detected in the southern areas came as a complete surprise. The northern area has a higher crime rate than the southern area, hence a high police deployment. However, the level of confrontation are approximately equal between the two areas.

distance, and summed the results for each reference cell; 2) it calculated the distance between each of the 10,800 reference cells and the 1349 census block groups with population as an intensity variable, evaluated the kernel function for each intensity-weighted distance, and summed the results for each reference cell; and 3) divided the kernel density estimate for auto thefts by the kernel density estimate for population for each reference cell location.

While the concentration of motor vehicle thefts relative to population ("motor vehicle theft risk") is still high in the metropolitan center, there are bands of high risk that spread outward, particularly along major arterials. There are now many "hot spot" areas which have a high distribution of motor vehicle thefts relative to the residential population. We could, of course, refine this analysis further by taking, for example, employment as a baseline variable rather than population; employment is a better indicator for the daytime population distribution whereas the residential population is a better indicator for nighttime population distribution (Levine, Kim, and Nitz, 1995a; 1995b).

Example 3: Kernel Density Estimates and Risk-adjusted Clustering of Robberies Relative to Population

The final example shows how the duel kernel interpolation compares with the risk-adjusted nearest neighbor clustering, discussed in chapter 6. Figure 8.15 shows 7 first-order risk-adjusted clusters overlaid on the a duel kernel estimate of 1996 robberies relative to 1990 population.[4] As seen, there is a correspondence between the identified risk-adjusted clusters and the duel kernel interpolation of the ratio of robberies to population. For a broad regional perspective, the interpolation produces an adequate model of where there is a high robbery risk. At the neighborhood level, however, the risk-adjusted clusters are more specific and would be preferable for use by police in identifying high-risk locations.

The advantage of a dual kernel density interpolation routine is that two variables can be related together. By interpolating one variable to a reference grid and then interpolating a second variable to the same reference grid, the two variables have been interpolated to the same geographical units. The two interpolations can then be related, by dividing, subtracting, or summing. As has been mentioned throughout this manual, one of the problems with techniques that depend on the concentration of incidents is that they ignore the underlying population-at-risk. With the dual routine, however, we can start to examine the risk and not just the concentration.

Visually Presenting Kernel Estimates

Whether the single- or duel-kernel estimate is used, the result is a grid interpretation of the data. By scaling these values by color in a GIS program, a visualization of the data is obtained. Areas with higher densities can be shown in darker tones and those with lower densities can be shown in lighter tones; some people do the opposite with the high density areas being lighter.

334

Figure 8.15:
Risk-adjusted Robbery Clusters and Interpolated Robbery Risk
1996 Robberies Relative to 1990 Population

Baltimore County

City of Baltimore

Robbery locations
1st-order robbery risk clusters
City of Baltimore
Baltimore County
Robberies Per 1000 Population
Low

High

0   9   18   Miles
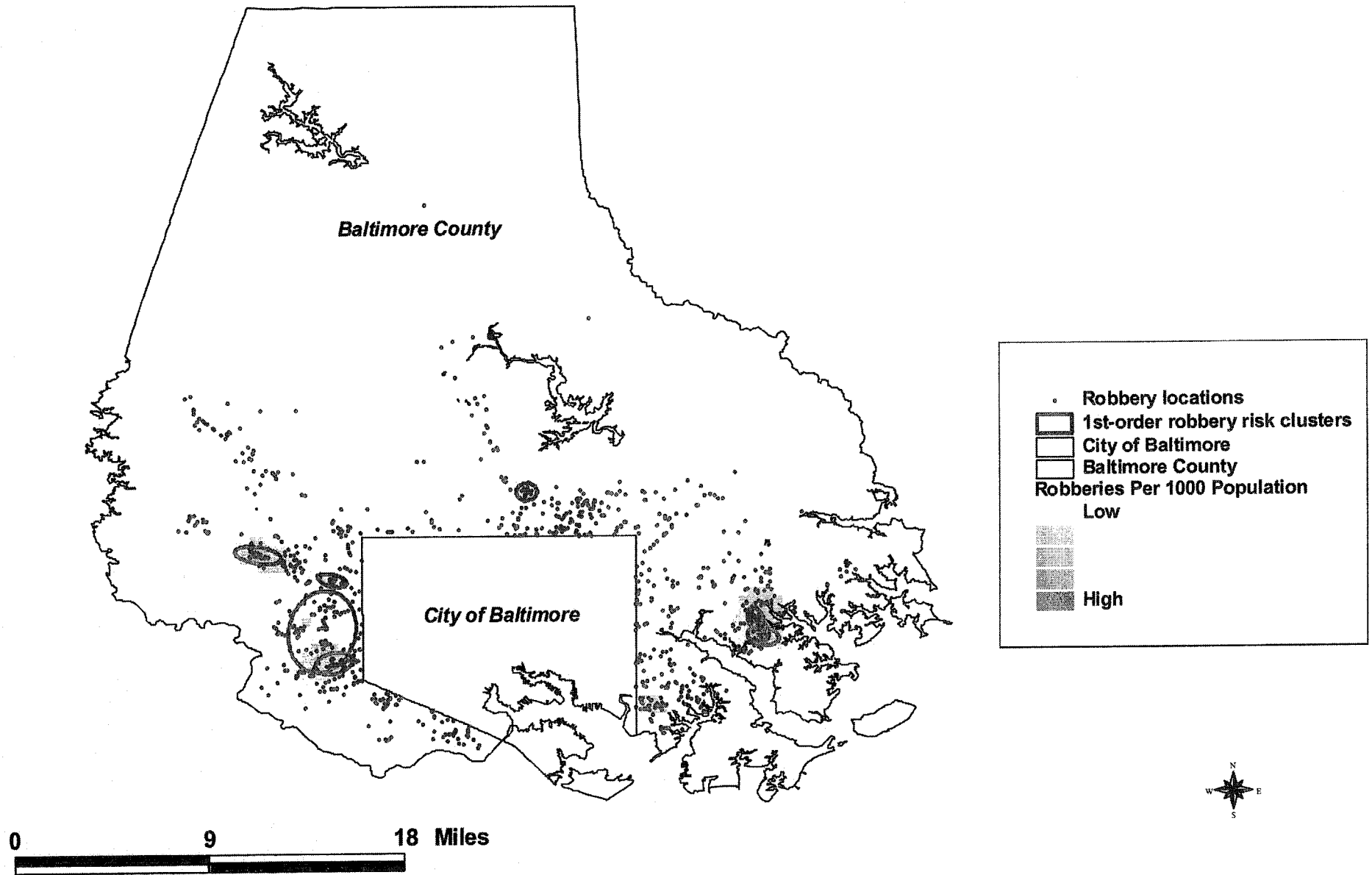
Evolution of the Urbanization Process in the Brazilian Amazonia

Silvana Amaral, Antônio Miguel V. Monteiro, Gilberto Câmara, José A. Quintanilha
INPE, Instituto Nacional de Pesquisas Espaciais, Brazil

The Brazilian Amazon rain forest is the world's largest contiguous area of tropical rain forest in the world. During the last three decades, the region has experienced the largest urban growth rates in Brazil, a process that has reorganized the network of human settlements in the region. We used the *CrimeStat* single and duel kernel density routines to visualize trends in urbanization from 1996 to 2000 in Amazonia. Two variables were used to measure urbanization: 1) the concentration of urban nuclei (city density); and 2) the ratio of urban to total population.

The concentration of cities was spatially associated withfederal roads in the eastern and southern portions, and along the Amazonas River in the middle of the region. Additionally, the surfaces of urban population show that city density is not always associated with large urban populations. From 1996 to 2000 city density increased in the western Amazonia (Pará state) at a greater rate than the growth of the urban population. In the southeastern part of the region (Rondônia state), there were many urban centers. But the ratio of urban to total population was small, indicating that they are predominately agricultural regions.

City density - 1996

Urban Pop/Total Pop-1996

City density - 2000

Urban Pop/Total Pop-2000
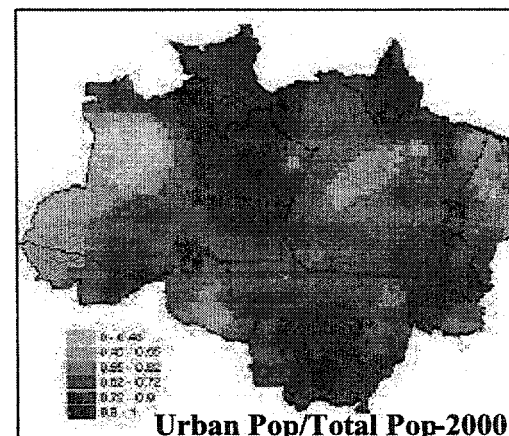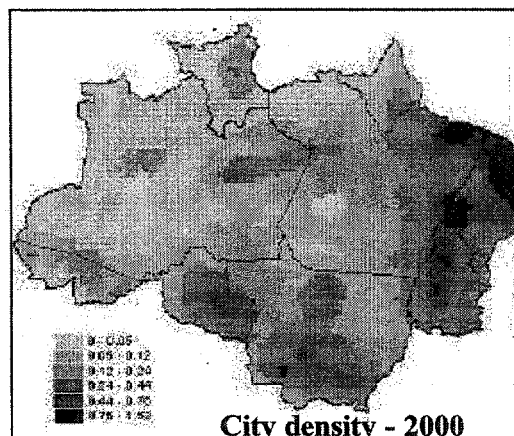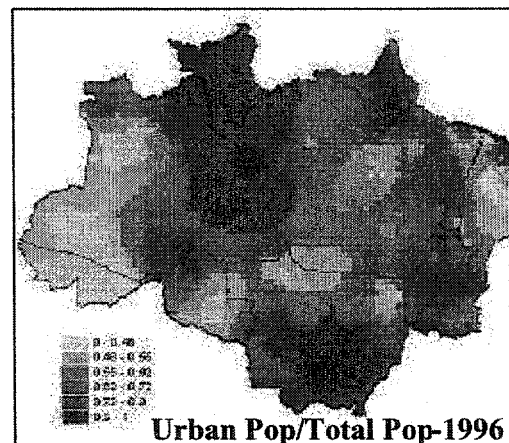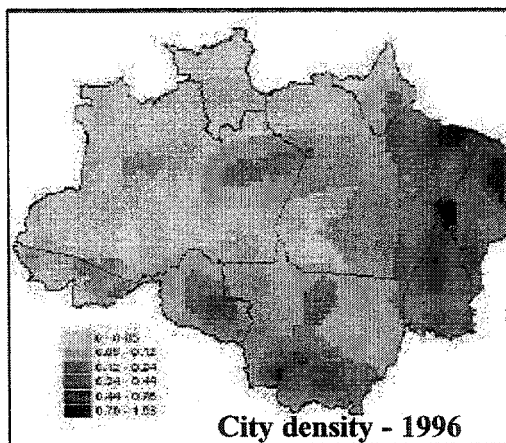
To make the visualization even more realistic, one could use a GIS program to cut out those grid cells that are outside the study area or are on water bodies. Before doing this, however, be sure to re-scale the estimated "Z" values so that they will sum to the total of the original grid. For example, if the original sample size was 1000, then the grid cells will sum to 1000 if the absolute density option is chosen. If, say, 20% of these cells are then removed to improve the visualization, then the grid cell Z values have to be re-scaled so that their sum will continue to be 1000. A simple way to do this is to, first, add up the Z values for the remaining cells and, second, multiply each grid cell Z by the ratio of the original sum to the reduced sum.

The visualization is useful for a broad, regional view. It is not particularly useful for micro analysis. The use of one of the cluster routines discussed in chapters 6 and 7 would be more appropriate for small area analysis.

Conclusion

Kernel density estimation is one of the 'modern' spatial statistical techniques. There is currently research on the use of this technique in both the statistical theory and in developing applications. For crime analysis, the technique represents a powerful way of conducting both 'hot spot' analysis as well as being able to link the 'hot spots' to an underlying population-at-risk. It can be used both for police deployment by targeting areas of high concentration of incidents as well as for prevention by targeting areas with high risk. It can also be used as a research tool for analyzing two or more distributions. More development of this approach can be expected in the next few years.

Endnotes to Chapter 8

1. There are differences in opinion about how wide a particular fixed bandwidth should be determined. The smoothing is done for a distribution of values, Z. If there are only unique points (and, hence, there is no Z value at a point), the distances between points can be substituted for Z. Thus, MeanD is the mean distance, sd(D) is the standard deviation of distance, and iqr(D) is the inter-quartile range of distances between points. These would be substituted for MeanZ, sd(Z), and iqr(Z) respectively

    Silverman (1986; 45-47; Härdle, 1991; Farewell, 1999) proposed a bandwidth, h, of:

    $$h = 1.06 * \min \left\{ sd(Z), \frac{iqr(Z)}{1.34} \right\} * N^{-1/5}$$

    where *min* is the minimum of the next two terms, *sd(Z)* is the standard deviation of the variable, Z, being interpolated, *iqr(Z)* is the inter-quartile range of Z, and N is the sample size.

    Bowman and Azzalini (1997; 31) defined a slightly different optimal bandwidth for a normal kernel.

    $$h = \left\{ \frac{4}{3N} \right\}^{1/5} * sd(Z)$$

    To avoid being influenced by outlier, they suggested using the median absolute deviation estimator for sd(Z)

    $$MAD(Z) = median \left\{ \frac{Z(i) - MedianZ}{0.6745} \right\}$$

    Scott (1992) suggested an upper bound on the normal kernel of

    $$h = 1.144 * sd(Z) * N^{-1/5}$$

    Bailey and Gatrell (1995, 85-87) offered a rough choice for the bandwidth of

    $$h = 0.68 * N^{-0.2}$$

    but suggested that the user could experiment with different bandwidths to explore the surface.

    On the other hand, the concept of an adaptive bandwidth is based more on sampling

338

theory (Bailey and Gatrell, 1995). By increasing the bandwidth until a fixed number of points are counted ensures that the level of precision is constant throughout the region. As with all sampling, the standard error of the estimate is a function of the sample size; a larger sample leads to smaller error. In general, if there was independent sampling, the 95% confidence interval of a bandwidth for a normal kernel could be approximated by

$$95\% \text{ C.I.} = \text{Mean}(Z) \ +/- \ 1.96 * \frac{.5}{N(h)^{1/2}} * sd(Z)$$

where $N(h)$ is the adaptive sample size (the number of points counted within the bandwidth for the adaptive kernel). This assumes that a point has an equal likelihood of falling within the bandwidth of one cell compared to an adjacent cell (i.e., it sits on the boundary of the bandwidth circle). The adaptive bandwidth criteria requires that the bandwidth be increased until it captures the specified number of points. On average, if there are N points in a region of area, A, and if the adaptive sample size is $N(p)$, then the average area required to capture $N(p)$ points is

$$A(p) = \frac{N(p) * A}{N}$$

and the average bandwidth, Mean(h), is

$$\text{Mean}(h) = \text{SQRT}[\frac{A(p)}{\pi}] = \text{SQRT}[\frac{N(p) * A}{N * \pi}]$$

Each of these provide different criteria for the bandwidth size with the adaptive being the most conservative. For example, for a standardized distribution with 1000 data points, a standardized mean of Z of 0 and a standardized standard deviation of 1, the Silverman criteria would produce a bandwidth of 0.2663; the Bowman and Azzalini criteria would produce a bandwidth of 0.2661; the Scott criteria would produce a bandwidth of 0.2874 and the Bailey and Gatrell criteria would produce a bandwidth of 0.1708. For the adaptive interval, if the required adaptive sample size is 25, then the average bandwidth would be approximately 0.3162 (this assumes that the area is a circle with a radius of 2 standardized standard deviations).

2. *CrimeStat* will output the geographical boundaries of the reference grid (a polygon grid) and will assign a third-variable (called *Z*) as the density estimate. Of the three polygon grid outputs, *ArcView* '.shp' files can be read directly into the program. For *MapInfo*, on the other hand, the output is in MapInfo Interchange Format (a '.mif' and a '.mid' file); the density estimate (also called *Z*) is assigned to

339

the '.mid' file. The files must be imported to convert it to a *MapInfo* '.tab' file. For *Atlas*GIS* '.bna' format, however, there are two files that are output - a '.bna' file which includes the boundaries of the polygon grid and a '.dbf' file which includes the grid cell names (called *gridcell*) and the density estimate (also called *Z*). The '.bna' file must be read in first and then the '.dbf' file must be read in and matched to the value of *gridcell*. For all three output formats, the values of *Z* can be shown as a thematic map but the ranges must be adjusted to illustrate the likely locations for the offender's residence (i.e., the default values in the GIS programs will not display the densities very well). On the other hand, the default interval values for *Surfer for Windows* and *ArcView Spatial Analyst* provide a reasonably good visualization of the densities.

3. All the *CrimeStat* outputs except for *ArcView* 'shp' files are in ASCII. There are usually 'edge effects' and values interpreted outside the actual geographical area. These can be removed with an ASCII editor by substituting '0' for the values at the edges or outside the study region. For 'shp' files, the values at the edges can be edited within the *ArcView* program. Another alternative is to 'cut out' the cells that are beyond the study area. Care must be taken, however, to not edit an output file too much otherwise it will bear little relationship to the calculated kernel estimate.

4. The risk-adjusted hierarchical clustering (Rnnh) method defined the largest search radius but a minimum of 25 points being required to be clustered. The kernel estimate for both the Rnnh and the duel-kernel routines used the normal distribution function with an adaptive bandwidth of 25 points.

340

Chapter 9
Journey to Crime Estimation

The *Journey to Crime* (Jtc) routine is a distance-based method which makes estimates about the likely residential location of a serial offender. It is an application of *location theory*, a framework for identifying optimal locations from a distribution of markets, supply characteristics, prices, and events. The following discussion gives some background to the technique. Those wishing to skip this part can go to page 13 for the specifics of the Jtc routine.

Location Theory

Location theory is concerned with one of the central issues in geography. This theory attempts to find an optimal location for any particular distribution of activities, population, or events over a region (Haggett, Cliff and Frey, 1977; Krueckeberg and Silvers, 1974; Stopher and Meyburg, 1975; Oppenheim, 1980, Ch. 4; Bossard, 1993). In classic location theory, economic resources were allocated in relation to idealized representations (Anselin and Madden, 1990). Thus, von Thünen (1826) analyzed the distribution of agricultural land as a function of the accessibility to a single population center (which would be more expensive towards the center), the value of the product produced (which would vary by crop), and transportation costs (which would be more expensive farther from the center). In order to maximize profit and minimize costs, a distribution of agricultural land uses (or crop areas) emerges flowing out from the population center as a series of concentric rings. Weber (1909) analyzed the distribution of industrial locations as a function of the volume of materials to be shipped, the distance that the goods had to be shipped, and the unit distance cost of shipping; consequently, industries become located in particular concentric zones around a central city. Burgess (1925) analyzed the distribution of urban land uses in Chicago and described concentric zones of both industrial and residential uses. Their theory formed the backdrop for early studies on the ecology of criminal behavior and gangs (Thrasher, 1927; Shaw, 1929).

In more modern use, the location of persons with a certain need or behavior (the 'demand' side) is identified on a spatial plane and places are selected as to maximize value and minimize travel costs. For example, for a consumer faced with two retail shops selling the same product, one being closer but more expensive while the other being farther but less expensive, the consumer has to trade off the value to be gained against the increased travel time required. In designing facilities or places of attraction (the 'supply' side), the distance between each possible facility location and the location of the relevant population is compared to the cost of locating near the facility. For example, given a distribution of consumers and their propensity to spend, such a theory attempts to locate the optimal placement of retail stores, or, given the distribution of patients, the theory attempts to locate the optimal placement of medical facilities.

341

Predicting Locations from a Distribution

One can also reverse the logic. Given the distribution of demand, the theory could be applied to estimate a central location from which travel distance or time is minimized. One of the earliest uses of this logic was that of John Snow, who was interested in the causes of cholera in the mid-19th century (Cliff and Haggett, 1988). He postulated the theory that water was the major vector transmitting the cholera bacteria. After investigating water sources in the London metropolitan area and concluding that there was a relationship between contaminated water and cholera cases, he was able to confirm his theory by an outbreak of cholera cases in the Soho district. By plotting the distribution of the cases and looking for water sources in the center of the distribution (essentially, the center of minimum distance; see chapter 4), he found a well on Broad Street that was, in fact, contaminated by seepage from nearby sewers. The well was closed and the epidemic in Soho receded. Incidently, in plotting the incidents on a map and looking for the center of the distribution, Snow applied the same logic that had been followed by the London Metropolitan Police Department who had developed the famous 'pin' map in the 1820s.

Theoretically, there is an optimal solution that minimizes the distance between demand and supply (Rushton, 1979). However, computationally, it is an almost impossible task to define, requiring the enumeration of every possible combination. Consequently in practice, approximate, though sub-optimal, solutions are obtained through a variety of methods (Everett, 1974, Ch. 4).

Travel Demand Modeling

A sub-set of location theory models the travel behavior of individuals. It actually is the converse. If location theory attempts to allocate places or sites in relation to both a supply-side and demand-side, travel demand theory attempts to model how individuals travel between places, given a particular constellation of them. One concept that has been frequently used for this purpose is that of the *gravity function*, an application of Newton's fundamental law of attraction (Oppenheim, 1980). In the original Newtonian formulation, the attraction, F, between two bodies of respective masses M_1 and M_2, separated by a distance D, will be equal to

$$F = g \, \frac{M_1 \, M_2}{D^2} \tag{9.1}$$

where g is a constant or scaling factor which ensures that the equation is balanced in terms of the measurement units (Oppenheim, 1980). As we all know, of course, g is the gravitational constant in the Newtonian formulation. The denominator of the equation, d^2, is known as the *distance decay* function and indicates that the attraction between the two bodies falls off as a function of their *squared* distance.

342

Social Applications of the Gravity Concept

The gravity model has been the basis of many applications to human societies and has been applied to social interactions since the 19th century. Ravenstein (1895) and Andersson (1897) applied the concept to the analysis of migration by arguing that the tendency to migrate between regions is inversely proportional to the squared distance between the regions. Reilly's 'law of retail gravitation' (1929) applied the Newtonian gravity model directly and suggested that retail travel between two centers would be proportional to the product of their populations and inversely proportional to the square of the distance separating them:

$$M_{ij} = K \frac{P_i P_j}{D_{ij}^2} \tag{9.2}$$

where M_{ij} is the interaction between centers i and j, P_i and P_j are the respective populations, D_{ij} is the distance between them raised to the second power and K is a balancing constant. In the model, the initial population, P_i, is called a *production* while the second population, P_j, is called an *attraction*.

Stewart (1950) and Zipf (1949) applied the concept to a wide variety of phenomena (migration, freight traffic, exchange of information) using a simplified form of the gravity equation

$$M_{ij} = K \frac{P_i P_j}{D_{ij}} \tag{9.3}$$

where the terms are as in equation 9.2 but the exponent of distance is only 1. In doing so, they basically linked location theory with travel behavior theory. Given a particular pattern of interaction for any type of goods, service or human activity, an optimal location of facilities should be solvable.

In the Stewart/Zipf framework, the two P's were both population sizes and, therefore, their sums had to be equal. However, in modern use, it's not necessary for the productions and attractions to be identical units (e.g., P_i could be population while P_j could be employment).

The total volume of productions (trips) from a single location, i, is estimated by summing over all destination locations, j:

$$M_i = K \ P_i \sum_j (P_j/D_{ij}) \tag{9.4}$$

Over time, the concept has been generalized and applied to many different types of travel behavior. For example, Huff (1963) applied the concept to retail trade between zones in an urban area using the general form of

343

$$A_{ij} = K \, \frac{S_j^{\beta}}{D_{ij}^{\lambda}} \tag{9.5}$$

where A_{ij} is the number of purchases in location j by residents of location i, S_j is the attractiveness of zone j (e.g., square footage of retail space), D_{ij} is the distance between zones i and j, β is the exponent of S_j, and λ is the exponent of distance (Bossard, 1993). $D_{ij}^{-\lambda}$ is sometimes called an *inverse distance* function. This is a *single constraint* model in that only the attractiveness of a commercial zone is constrained, that is the sum of all attractions for j must equal the total attraction in the region.

Again, it can be generalized to all zones by, first, estimating the total trips generated from one zone, i, to another zone, j,

$$M_{ij} = K \, \frac{H_i^{\rho} \, S_j^{\beta}}{D_{ij}^{\lambda}} \tag{9.6}$$

where M_{ij} is the interaction between two locations (or zones), H_i is productions of trips from location/zone i, S_j is the attractiveness of location/zone j, D_{ij} is the distance between zones i and j, β is the exponent of S_j, ρ is the exponent of H_i, λ is the exponent of distance, and K is a constant.

Second, the total number of trips generated by a location, i, to all destinations is obtained by summing over all destination locations, j:

$$M_i = K \, H_i^{\rho} \, \sum_j (S_j^{\beta}/D_{ij}^{\lambda}) \tag{9.7}$$

This differs from the traditional gravity function by allowing the exponents of the production from location i, the attraction from location j, and the distance between zones to vary. Typically, these exponents are calibrated on a known sample before being applied to a forecast sample and the locations are usually measured by zones. Thus, retailers in deciding on the location of a new store can use this type of model to choose a site location to optimize travel behavior of patrons; they will, typically, obtain data on actual shopping trips by customers and then calibrate the model on the data, estimating the exponents of attraction and distance. The model can then be used to predict future shopping trips if a facility is built at a particular location.

This type of function is called a *double constraint* model because the balancing constant, K, has to be constrained by the number of units in both the origin and destination locations; that is, the sum of P_i over all locations must be equal to the total number of productions while the sum of P_2 over all locations must be equal to the total number of attractions. Adjustments are usually required to have the sum of individual productions and attractions equal the totals (usually estimated independently).

344

The equation can be generalized to other types of trips and different metrics can be substituted for distance, such as time, effort, or cost (Isard, 1960). For example, for commuting trips, usually employment is used for attractions, frequently sub-divided into retail and non-retail employment. In addition, for productions, median household income or car ownership percentage is used as an additional production variable. Equation 9.7 can be generalized to include any type of production or attraction variable (9.8 and 9.9):

$$M_{ij} = K_1 \, P_i^{\rho} \, K_2 \, A_j^{\beta}/D_{ij}^{\,\lambda} \qquad (9.8)$$

$$M_i = K_1 \, P_i^{\rho} \, \Sigma \, (K_2 \, A_j^{\beta}/D_{ij}^{\,\lambda}) \qquad (9.9)$$

where M_{ij} is the number of trips produced by location i that travel to location j, P_i is either a single variable associated with trips produced from a zone or the cross-product of two or more variables associated with trips produced from a zone, A_j is either a single variable associated with trips attracted to a zone or the cross-product of two or more variables associated with trips attracted to a zone, D_{ij} is either the distance between two locations or another variable measuring travel effort (e.g., travel time, travel cost), ρ, β, and λ are exponents of the respective terms, K_1 is a constant associated with the productions to ensure that the sum of trips produced by all zones equals the total number of trips for the region (usually estimated independently), and K_2 is a constant associated with the attractions to ensure that the sum of trips attracted to all zones equals the total number of trips for the region. Without having two constants in the equation, usually conflicting estimates of K will be obtained by balancing the equation against productions or attractions. The summation over all destination locations, j (equation 9.9), produces the total number of trips from zone i.

Intervening Opportunities

Stouffer (1940) modified the simple gravity function by arguing that the attraction between two locations was a function not only of the characteristics of the relative attractions of two locations, but of intervening opportunities between the locations. His hypothesis "..assumes that there is no necessary relationship between mobility and distance... that the number of persons going a given distance is directly proportional to the number of opportunities at that distance and inversely proportional to the number of intervening opportunities"(Stouffer, 1940, p. 846). This model was used in the 1940s to explain interstate and intercounty migration (Bright and Thomas, 1941; Isbell, 1944; Isard, 1979). Using the gravity type formulation, we can write this as:

$$A_{ji} = K \; \frac{S_j^{\beta}}{\Sigma(S_k^{\xi}) \, D_{ij}^{\,\lambda}} \qquad (9.10)$$

where A_{ji} is the attraction of location j by residents of location i, S_j is the attractiveness of zone j, S_k is the attractiveness of all other locations that are *intermediate* in distance between locations i and j, D_{ij} is the distance between zones *i* and *j*, β is the exponent of S_j, ξ is the exponent of S_k, and λ is the exponent of distance. While the intervening

345

opportunities are implicit in equation 9.5 in the exponents, β and λ, and coefficient, K, equation 9.10 makes the intervening opportunities explicit. The importance of the concept is that the interaction between two locations becomes a complex function of the spatial environment of nearby areas and not just of the two locations.

Urban Transportation Modeling

This type of model is incorporated as a formal step in the urban transportation planning process, implemented by most regional planning organizations in the United States and elsewhere (Stopher and Meyburg, 1975; Krueckeberg and Silvers, 1974; Field and MacGregor, 1987). The step, called *trip distribution*, is linked to a five step model. First, data are obtained on travel behavior for a variety of trip purposes. This is usually done by sampling households and asking each member to keep a travel diary documenting all their trips over a two or three day period. Trips are aggregated by individuals and by households. Frequently, trips by different purposes are separated. Second, the volume of trips produced by and attracted to zones (called traffic analysis zones) is estimated, usually on the basis of the number of households in the zone and some indicator of income or private vehicle ownership. Third, trips produced by each zone are distributed to every other zone usually using a gravity-type function (equation 9.9). That is, the number of trips produced by each origin zone and ending in each destination zone is estimated by a gravity model. The distribution is based on trip productions, trip attractions, and travel 'resistance' (measured by travel distance or travel time). Fourth, zone-to-zone trips are allocated by mode of travel (car, bus, walking, etc); and, fifth, trips are assigned to particular routes by travel mode (i.e., bus trips follow different routes than private vehicle trips). The advantage of this process is that trips are allocated according to origins, destinations, distances (or travel times), modes of travel and routes. Since all zones are modeled simultaneously, all intermediate destinations (i.e., intervening opportunities) are incorporated into the model.

Alternative Distance Decay Functions

One of the problems with the traditional gravity formulation is in the measurement of travel resistance, either distance or time. For locations separated by sizeable distances in space, the gravity formulation can work properly. However, as the distance between locations decreases, the denominator approaches infinity. Consequently, an alternative expression for the interaction has been proposed which uses the negative exponential function (Hägerstrand, 1957; Wilson, 1970).

$$A_{ji} = S_j^{\beta} \, e^{(-\alpha D_{ij}^{\lambda})} \qquad\qquad (9.11)$$

where A_{ji} is the attraction of location j for residents of location i, S_j is the attractiveness of location j, D_{ij} is the distance between locations *i* and *j*, β is the exponent of S_j, *e* is the base of the natural logarithm (i.e., 2.7183...), and λ is an empirically derived exponent. Sometimes known as *entropy maximization*, the latter part of the equation is a negative exponential function which has a maximum value of 1 (i.e., $e^{-0} = 1$). This has the

346

advantage of making the equation more stable for interactions between locations that are close together. For example, Cliff and Haggett (1988) used a negative exponential gravity-type model to describe the diffusion of measles into the United States from Canada and Mexico. It has also been argued that the negative exponential function generally gives a better fit to urban travel patterns, particularly by automobile (Foot, 1981; Bossard, 1993).

Other functions have also be used to describe the distance decay - negative linear, normal distribution, lognormal distribution, quadratic, Pareto function, square root exponential, and so forth (Haggett and Arnold, 1965; Taylor, 1970; Eldridge and Jones, 1991). Later in the chapter, we will explore several different mathematical formulations for describing the distance decay. One, in fact, does not need to use a mathematical function at all, but could empirically describe the distance decay from a large data set and utilize the described values for predictions. The use of mathematical functions has evolved out of both the Newtonian tradition of gravity as well as various location theories which used the gravity function. A mathematical function makes sense under two conditions: 1) if travel is uniform in all directions; and 2) as an approximation if there is inadequate data from which to calibrate an empirical function. The first assumption is usually wrong since physical geography (i.e., oceans, rivers, mountains) as well as asymmetrical street networks make travel easier in some directions than others. As we shall see below, the distance decay is quite irregular for journey to crime trips and would be better described by an empirical, rather than mathematical function.

In short, there is a long history of research on both the location of places as well as the likelihood of interaction between these places, whether the interaction is freight movement, land prices or individual travel behavior. The gravity model and variations on it have been used to describe the interactions between these locations.

Travel Behavior of Criminals

Journey to Crime Trips

The application of travel behavior theory to crime has a sizeable history as well. The analysis of distance for journey to crime trips was applied in the 1930s by White (1932), who noted that property crime offenders generally traveled farther distances than offenders committing crimes against people, and by Lottier (1938), who analyzed the ratio of chain store burglaries to the number of chain stores by zone in Detroit. Turner (1969) analyzed delinquency behavior by a distance decay travel function showing how more crime trips tend to be close to the offender's home with the frequency dropping off with distance. Phillips (1980) is, apparently, the first to use the term *journey to crime* is describing the travel distances that offenders make though Harries (1980) noted that the average distance traveled has evolved by that time into an analogy with the journey to work statistic.

Rhodes and Conly (1981) expanded on the concept of a *criminal commute* and showed how robbery, burglary and rape patterns in the District of Columbia followed a distance decay pattern. LeBeau (1987a) analyzed travel distances of rape offenders in San

347

Diego by victim-offender relationships and by method of approach. Boggs (1965) applied the intervening opportunities model in analyzing the distribution of crimes by area in relation to the distribution of offenders. Other empirical descriptions of journey to crime distances and other travel behavior parameters have been studied by Blumin (1973), Curtis (1974), Repetto (1974), Pyle (1974), Capone and Nichols 1975), Rengert (1975), Smith (1976), LeBeau (1987b), and Canter and Larkin (1993). It has generally been accepted that property crime trips are longer than personal crime trips (LeBeau, 1987a), though exceptions have been noted (Turner, 1969). Also, it would be expected that average trip distances will vary by a number of factors: crime type; method of operation; time of day; and, even, the value of the property realized (Capone and Nichols, 1975).

Modeling the Offender Search Area

Conceptual work on the type of model have been made by Brantingham and Brantingham (1981) who analyzed the *geometry of crime* and conceptualized a criminal search area, a geographical area modified by the spatial distribution of potential offenders and potential targets, the awareness spaces of potential offenders, and the exchange of information between potential offenders. In this sense, their formulation is similar to that of Stouffer (1940), who described intervening opportunities, though their's is a behavioral framework. An important concept developed by the Brantingham's is that of decreased criminal activity near to an offender's home base, a sort of a safety area around their near neighborhood. Presumably, offenders, particularly those committing property crimes, go a little way from their home base so as to decrease the likelihood that they will get caught. This was noted by Turner (1969) in his study of delinquency in Philadelphia. Thus, the Brantingham's postulated that there would be a small safety area (or 'buffer' zone) of relatively little offender activity near to the offender's base location; beyond that zone, however, they postulated that the number of crime trips would decrease according to a distance decay model (the exact mathematical form was never specified, however).

Crime trips may not even begin at an offender's residence. Routine activity theory (Cohen and Felson, 1979; 1981) suggests that crime opportunities appear in the activities of everyday life. The routine patterns of work, shopping, and leisure affect the convergence in time and place of would be offenders, suitable targets, and absence of guardians. Many crimes may occur while an offender is traveling from one activity to another. Thus, modeling crime trips as if they are referenced relative to a residence is not necessarily going to lead to better prediction.

The mathematics of journey to crime has been modeled by Rengert (1981) using a modified general opportunities model:

$$P_{ij} = K U_i V_j f(D_{ij}) \tag{9.12}$$

where P_{ij} is the probability of an offender in location (or zone) i committing an offense at location j, U_i is a measure of the number of crime trips produced at location i (what Rengert called *emissiveness*), V_j is a measure of the number of crime targets (attractiveness) at location j, and $f(D_{ij})$ is an unspecified function of the cost or effort expended in traveling

348

from location i to location j (distance, time, cost). He did not try to operationalize either the production side or the attraction side. Nevertheless, conceptually, a crime trip would be expected to involve both elements as well as the cost of the trip.

In short, there has been a great deal of research on the travel behavior of criminals in committing acts as well as a number of statistical formulations.

Predicting the Location of Serial Offenders

The journey to crime formulation, as in equation 9.9, has been used to estimate the origin location of a serial offender based on the distribution of crime incidents. The logic is to plot the distribution of the incidents and then use a property of that distribution to estimate a likely origin location for the offender. Inspecting a pattern of crimes for a central location is an intuitive idea that police departments have used for a long time. The distribution of incidents describes an activity area by an offender, who lives somewhere in the center of the distribution. It is a *sample* from the offender's activity space. Using the Brantingham's terminology, there is a search area by an offender within which the crimes are committed; most likely, the offender also lives within the search area.

For example, Canter (1994) shows how the area defined by the distribution of the 'Jack the Ripper' murders in the east end of London in the 1880s included the key suspects in the case (though the case was never solved). Kind (1987) analyzed the incident locations of the 'Yorkshire Ripper' who committed thirteen murders and seven attempted murders in northeast England in the late 1970s and early 1980s. Kind applied two different geographical criteria to estimate the residential location of the offender. First, he estimated the center of minimum distance. Second, on the assumption that the locations of the murders and attempted murders that were committed late at night were closer to the offender's residence, he graphed the time of the offense on the Y axis against the month of the year (taken as a proxy for length of day) on the X axis and plotted a trend line through the data to account for seasonality. Both the center of minimum distance and the murders committed at a later time than the trend line pointed towards the Leeds/Bradford area, very close to where the offender actually lived (in Bradford).

Rossmo Model

Rossmo (1993; 1995) has adapted location theory, particularly travel behavior modeling, to serial offenders. In a series of papers (Rossmo, 1993a; 1993b; 1995; 1997) he outlined a mathematical approach to identifying the home base location of a serial offender, given the distribution of the incidents. The mathematics represent a formulation of the Brantingham and Brantingham (1981) search area model, discussed above in which the search behavior of an offender is seen as following a distance decay function with decreased activity near the offender's home base. He has produced examples showing how the model can be applied to serial offenders (Rossmo, 1993a; 1993b; 1997).

349

The model has four steps (what he called *criminal geographic targeting*):

1. First, a rectangular study area is defined that extends beyond the area of the incidents committed by the serial offender. The average distance between points is taken in both the Y and X direction. Half the Y inter-point distance is added to the maximum Y value and subtracted from the minimum Y value. Half the X inter-point distance is added to the maximum X value and subtracted from the minimum X value. These are based on projected coordinates; presumably, the directions would have to be adjusted if spherical coordinates were used. The rectangular study defines a grid from which columns and rows can be defined.

2. For each grid cell, the Manhattan distance to each incident location is taken (see chapter 3 for definition).

3. For each Manhattan distance from a grid cell to an incident location, MD_{ij}, one of two functions is evaluated:

    A. If the Manhattan distance, MD_{ij}, is less than a specified buffer zone radius, B, then

    $$P_{ij} = \prod_{c=1}^{T} \{k[(1-\phi)(B^{g-f}) / (2B - |x_i - x_c| + |y_i - y_c|)^g] \} \qquad (9.13)$$

    where P_{ij} is the resultant of offender interaction for grid cell, i; c is the incident number, summing to T; $\phi = 0$; k is an empirically determined constant; g is an empirically determined exponent; and f is an empirically determined exponent.

    The Greek letter, Π, is the product sign, indicating that the results for each grid cell-incident distance, MD_{ij}, are *multiplied* together across all incidents, c. This equation reduces to

    $$P_{ij} = \prod_{c=1}^{T} \{k(1-0)(B^{g-f}) / (2B - |x_i - x_c| + |y_i - y_c|)^g \} \qquad (9.14)$$

    $$P_{ij} = \prod_{c=1}^{T} \frac{KB^{g-f}}{(2B - |x_i - x_c| + |y_i - y_c|)^g} \qquad (9.15)$$

    Within the buffer region, the function is the ratio of a constant, k, times the radius of the buffer, B, raised to another constant (g-f),

350

divided by the difference between the diameter of the circle (2B) and the Manhattan distance, MD_{ij}, raised to a constant, g. This is a *non-linear* function.

B.   If the Manhattan distance, MD_{ij}, is greater than a specified buffer zone radius, B, then

$$P_{ij} = \prod_{c=1}^{T} \{k \ [\ \phi\ /\ (|\ x_i - x_c|\ +\ |\ y_i - y_c|\ )^f\ \} \qquad (9.16)$$

where P_{ij} is the resultant of offender interaction for grid cell, i, and incident location, j; c is the incident number, summing to T; $\phi = 1$; k is an empirically determined constant (the same as in equation 9.15 above); and f is an empirically determined exponent (the same as in equation 9.15 above).

Again, the Greek letter, Π, indicates that the results for each grid cell-incident distance, MD_{ij}, are multiplied together across all incidents, c. This equation reduces to

$$P_{ij} = \prod_{c=1}^{T} \{k\ [\ 1/(|\ x_i - x_c|\ +\ |\ y_i - y_c|\ )^f\ \} \qquad (9.17)$$

$$P_{ij} = \prod_{c=1}^{T} \left\{ \frac{k}{(|\ x_i - x_c|\ +\ |\ y_i - y_c|\ )^f} \right\} \qquad (9.18)$$

Outside of the buffer region, the function is a constant, k, divided by the distance, MD_{ij}, raised to an exponent, f. It is an inverse distance function and drops off rapidly with distance

4.   Finally, for each grid cell, i, the functions evaluated in step 3 above are summed over all incidents.

For both the 'within buffer zone' (near to home base) and 'outside buffer zone' (far from home base) functions, the coefficient, k, and exponents, f and g, are empirically determined. Though he doesn't discuss how these are calculated, they are presumably estimated from a sample of known offender locations where the distance to each incident is known (e.g., arrest records).

The result is a surface model indicating a likelihood of the offender residing at that location. He describes it as a probability surface, but it is actually a *density* surface. Since the probability of interaction between any one grid cell, i, and any one incident, j, cannot be greater than 1, the surface actually indicates the product of individual likelihoods that the

351

offender uses that location as the home base. To be an actual probability function, it would have to be re-scaled so that the sum of the grid cells was equal to 1.

The second function - 'outside the buffer zone' (equation 9.16) is a classic gravity function, similar to equation 9.5 except there is no attraction definition. It is the distance decay part of the gravity function. The first function, equation 9.13, is an increasing curvilinear function designed to model the area of decreased activity near the offender's home base.

Strengths and weaknesses of the Rossmo model

The Rossmo model has both strengths and weaknesses. First, the model has some theoretical basis utilizing the Brantingham and Brantingham (1981) framework for an offender search area as well as the mathematics of the gravity model and distinguishes two types of travel behavior - near to home and farther from home. Second, the model does represent a systematic approach towards identifying a likely home base location for an offender. By evaluating each grid cell in the study area, an independent estimate of the likelihood is obtained, which can then be integrated into a continuous surface with an interpolation graphics routine.

There are problems with the particular formulation, however. First, the exclusive use of Manhattan distances is questionable. Unless the study area has a street network that follows a uniform grid, measuring distances horizontally and vertically can lead to overestimation of travel distances; further, the more the layout differs from a north-south and east-west orientation, the greater the distortion. Since many urban areas do not have a uniform grid street layout, the method will necessarily lead to overestimation of travel distances in places where there are diagonal or irregular streets.[1]

Second, the use of a product term, Π, complicates the mathematics. That is, the technique evaluates the distance from a particular grid cell, i, to a particular incident location, j. It then *multiplies* this result by all other results. The process, if strictly applied, would be a compounding of probabilities with overestimation of the likelihood for grid cells close to incident locations and underestimation of the likelihood for grid cells farther away. In the description of the method, however, Rossmo actually mentions summing the terms. Thus, the substitution of a summation sign, Σ, for the product sign would help the mathematics.

A third problem is in the distance decay function (equation 9.16). The use of an inverse distance term has problems as the distance between the grid cell location, i, and the incident location, j, decreases. For some types of crimes, there will be little or no buffer zone around the offender's home base (e.g., rapes by acquaintances). Consequently, the buffer zone radius, B, would approach 0. However, this would cause the model to become unstable since the inverse distance term will approach infinity.

Fourth, the use of a mathematical function to describe the distance decay, while easy to define, probably oversimplifies actual travel behavior. A mathematical function to

352

describe distance decay is an approximation to actual travel behavior. It assumes that travel is equally likely in each direction, that travel distance is uniformly easy (or difficult) in each direction, and that, similarly, opportunities are uniformly distributed. For most urban areas, these conditions would not be true. Few cities form a perfect grid (Salt Lake City is, of course, an exception), though most cities have sections that are grided. Both physical geography limit travel in certain directions as does the historical street structure, which is often derived from earlier communities. A mathematical function does not consider this structure, but rather assumes that the 'impedance' in all directions is uniform.

This latter criticism, of course, would be true for all mathematical formulations of travel distance. There are corrections that can be made to adjust for this. For example, in the urban transportation planning system model, trip distribution between locations is estimated by a gravity model, but then the distributed trips are constrained by, first, the total number of trips in the region (estimated separately), second, by mode of travel (bus v. single driver v. drivers plus passengers v. walk, etc.), and, third, by the route structure upon which the trips are eventually assigned (Krueckeberg and Silvers, 1974; Stopher and Meyburg, 1975; Field and MacGregor, 1987). Calibration at all stages against known data sets ensures that the coefficients and exponents fit 'real world' data as closely as possible. It would take these types of modifications to make the travel distribution type of model postulated by Rossmo and others be more realistic.

Fifth, the model imposes mathematical rigidity on the data. While there are two different functions that could vary from place to place, the particular type of distance decay function might also vary. Specifying a strict form for the two equations limits the flexibility of applying the model to different types of crime or to places where the distance decay does not follow the form specified by Rossmo.

A final problem is that opportunities for committing crimes - the attractiveness of locations, are never measured. That is, there is no enumeration of the opportunities that would exist for an offender nor is there an attempt to measure the strength of this attraction. Instead, the search area is inferred strictly from the distribution of incidents. Because the distribution of offender opportunities would be expected to vary from place to place, the model would need to be re-calibrated at each location. In this sense, both the Canter model and my journey to crime model (both described below) also share this weakness. It is understandable in that victim/target opportunities are difficult to define *a priori* since they can be interpreted differently by individuals. Nevertheless, a more complete theory of journey to crime behavior would have to incorporate some measure of opportunities, a point that both Brantingham and Brantingham (1981) and Rengert (1981) have made.

Canter Model

Canter's group in Liverpool (Canter and Tagg, 1975; Canter and Larkin, 1993; Canter and Snook, 1999; Canter, Coffey and Huntley, 2000) have modified the distance

353

decay function for journey to crime trips by using a negative exponential term, instead of the inverse distance. Their *Dragnet* program uses the negative exponential function

$$Y = \alpha \; e^{(-\beta \; Dij/ \; P)} \tag{9.19}$$

where Y is the likelihood of an offender traveling a certain distance to commit a crime,, D_{ij} is the distance (from a home base location to an incident site), α is an arbitrary constant, β is the coefficient of the distance (and, hence, an exponent of e), P is a normalization constant, and e is the base of the natural logarithm. The model is similar to equation 9.11 except, like Rossmo, it does not include the attractiveness of the location.

Using the logic that most crimes are committed near the offender's home base, Canter, Coffey and Huntley (2000) use a five step process to estimate a search strategy:

1. The study area is defined by a rectangle that is 20% larger in area than that defined by the minimum and maximum X/Y points. A grid cell structure of 13, 300 cells is imposed over the rectangle. Each grid cell is a reference location, i.

2. A decay coefficient is selected. In equation 9.19, this would be the coefficient, β, for the distance term, D_{ij}, both of which are exponents of e. Unlike Rossmo, Canter uses a series of decay coefficients from 0.1 to 10 to estimate the sensitivity of the model. The equation indicates the likelihood with which any location is likely to be the home base of the offender based on one incident.

3. Because different offenders have different search areas, the measured distances for each cell are divided by a normalization coefficient, P, that adjusts all offenses to a comparable range. Canter uses two different types of normalization function: 1) mean inter-point distance between all offenses (across a group of offenders); and 2) the QRange, which is an index that takes into account asymmetry in the orientation of the incidents.

4. For each reference cell, i, the distance between each grid cell and each incident location is evaluated with the function and the standardized likelihoods are summed to yield an estimate of location potential.

5. A *search cost* index is defined by the proportion of the study area that has to be searched to find the offender. By calibrating the model against known cases, an estimate of search efficiency is obtained.

Additional modifications can be added to the functions to make them more flexible (Canter, Coffey and Huntley, 2000). For example, 'steps' are distances near to home where offenders are not likely to act while 'plateaus' are constant distances near to home where

354

there is the highest likelihood of acting. For example, Canter and Larkin (1993) found an area around serial offenders' homes of about 0.61 mile in radius within which they were less likely to commit crimes.

Canter and Snook (1999) provide estimates of the search cost (or efficiency) associated with various distance coefficients. For example, with the known home base locations of 32 burglars, a β of 1.0 yielded a mean search cost of 18.06%; that is, on average, only 18.06% of the study area had to be searched to find the location of 32 burglars in the calibration sample. Clearly, for some of them, a larger area had to be searched while for others a smaller area; the average was 18.06%. Conversely, the mean search cost index for 24 rapists was 21.10% and for 37 murderers 28.28%. They further explored the marginal increase in locating offenders by increasing the percentage of the study area that had to be searched. They found for their three samples (burglary, rape, homicide) that more than half the offenders could be located within 15% of the area searched.

The Canter model is different from the Rossmo model is that it suggests a search strategy by the police for a serial offender rather than a particular location. The strength of it is to indicate how narrow an area the police should concentrate on in order to optimize finding an offender. Clearly, in most cases, only a small area needs be searched.

Strengths and weaknesses of the Canter model

The model has both strengths and weaknesses. First, the model provides a search strategy for law enforcement. By examining what type of function best fits a certain type of crime, police can target their search efforts more efficiently. The model is relatively easy to implement and is practical. Second, the mathematical formulation is stable. Unlike the inverse distance function in the Rossmo model, equation 9.19 will not have problems associated with distances that are close to 0. Further, the model does provide a search strategy for identifying an offender. It is a useful tool for law enforcement officers, particularly as they frame a search for a serial offender.

There are also weaknesses to the model. First, it lacks a theoretical basis. Canter's research has provided a great deal in terms of understanding the activity spaces of serial offenders (Canter and Larkin, 1993; Canter and Gregory, 1994; Canter, 1994; Hodge and Canter, 2000). However, the empirical model used is strictly pragmatic. Second, mathematically, it imposes the negative exponential function without considering other distance decay models. In the *Dragnet* program, the decay function is a string of 20 numbers so that, in theory, any function can be explored. However, the default is a negative exponential. The negative exponential has been used in many travel behavior studies (Foot, 1981; Bossard, 1993), but it does not always produce the best fit. Later on, I'll show examples of travel behavior which show a distinctly non-monotonic function, even beyond a home base 'buffer zone'. While the model can be adapted to be more flexible by different exponents and including steps and plateaus, for example, it is still tied to the negative exponential form. Thus, the model might work in some locations, but may fail in others; a user can't easily adjust the model to make it fit new data.

355

Third, the coefficient of the negative exponential, α, is defined arbitrarily. In the *Dragnet* program, it is usually set as 0.5. While this ensures that the result never exceed 1.0 for any one incident, there is a limit on the location potential summation since the total potential is a function of the number of incidents (i.e., it will be higher for more incidents). Thus, the use of α ends up being arbitrary. It would have been better if the coefficient were calibrated against a known sample.

Fourth, and finally, also similar to the Rossmo model (and to my Jtc model below), criminal opportunities (or attractions) are never measured, but are inferred from the pattern of crime incidents. As a pragmatic tool for informing a police search, one could argue that this is not important. However, in a different location, the distance coefficient is liable to differ as is the search cost index. It would need to be re-calibrated each time.

Nevertheless, the Canter model is a useful tool for police department and can help shape a search strategy. It is different from the other location models in that it is not focused so much on the best prediction for a location of an offender (though the summation discussed above in step 4 can yield that) as it does in defining where the search should be optimized.

Geographic Profiling

Journey to crime estimation should be distinguished from *geographical profiling*. Geographical profiling involves understanding the geographical search pattern of criminals in relation to the spatial distribution of potential offenders and potential targets, the awareness spaces of potential offenders including the labeling of 'good' targets and crime areas, and the interchange of information between potential offenders who may modify their awareness space (Brantingham and Brantingham, 1981). According to Rossmo:

> "...Geographic profiling focuses on the probable spatial behaviour of the offender within the context of the locations of, and the spatial relationships between, the various crime sites. A psychological profile provides insights into an offender's likely motivation, behaviour and lifestyle, and is therefore directly connected to his/her spatial activity. Psychological and geographic profiles thus act in tandem to help investigators develop a picture of the person responsible for the crimes in question" (Rossmo, 1997).

In other words, geographic profiling is a framework for understanding how an offender traverses an area in searching for victims or targets; this, of necessity, involves understanding the social environment of an area, the way that the offender understands this environment (the 'cognitive map') as well as the offender's motives.

On the other hand, journey to crime estimation follows a much simpler logic involving the distance dimension of the spatial patterning of a criminal. It is a method aimed at estimating the distance that serial offenders will travel to commit a crime and, by

356

implication, the likely location from which they started their crime 'trip'. In short, it is a strictly statistical approach to estimating the residential whereabouts of an offender compared to understanding the dynamics of serial offenders.

It remains an empirical question whether a conceptual framework, such as geographic profiling, can predict better than a strictly statistical framework. Understanding of a phenomena, such as serial murders, serial rapists, and so forth, is an important research area. We seek more than just statistical prediction in building a knowledge base. However, it doesn't necessarily follow that understanding produces better predictions. In many areas of human activity, strictly statistical models are better in predicting than explanatory models. I will return to this point later in the section.

The *CrimeStat* Journey to Crime Routine

The journey to crime (Jtc) routine is a diagnostic designed to aid police departments in their investigations of serial offenders. The aim is to estimate the likelihood that a serial offender lives at any particular location. Using the location of incidents committed by the serial offender, the program makes statistical guesses at where the offender is liable to live, based on the similarity in travel patterns to a known sample of serial offenders for the same type of crime. The *Jtc* routine builds on the Rossmo (1993a; 1993b; 1995) framework, but extends its modeling capability.

1. A grid is overlaid on top of the study area. This grid can be either imported or can be generated by *CrimeStat* (see chapter 2). The grid represents the entire study area. Unlike Rossmo or Canter and Snook, there is no optimal study area. The technique will model that which is defined. Thus, the user has to select an area intelligently.

2. The routine calculates the distance between each incident location committed by a serial offender (or group of offenders working together) and each cell, defined by the centroid of the cell. Rossmo (1993a; 1995) used indirect (Manhattan) distances. However, this would be appropriate only when a city falls on a uniform grid. The *Jtc* routine allows both direct and indirect distances. In most cases, direct distances would be the most appropriate choice as a police department would normally locate origin and destination locations rather than particular routes that are taken (see below).

3. A distance decay function is applied to each grid cell-incident pair and sums the values over all incidents. The user has a choice whether to model the travel distance by a mathematical function or an empirically-derived function.

4. The resultant of the distance decay function for each grid cell-incident pair are summed over all incidents to produce a likelihood (or density) estimate for each grid cell.

357

5.   In both cases, the program outputs the two results: 1) the grid cell which has the peak likelihood estimate; and 2) the likelihood estimate for every cell. The latter output can be saved as a *Surfer® for Windows* 'dat', *ArcView Spatial Analyst©* 'asc', ASCII 'grd', *ArcView®* '.shp', *MapInfo®* '.mif', *Atlas*GIS™* '.bna' file or as an Ascii grid 'grd' file which can be read by many GIS packages (e.g., *ARC/INFO®*, *Vertical Mapper©*). These files can also be read by other GIS packages (e.g., *Maptitude*).

Figure 9.1 shows the logic of the routine and figure 9.2 shows the Journey to Crime (Jtc) screen. There are two parts to the routine. First, there is a calibration model which is used in the empirically-derived distance function. Second, there is the Journey to Crime (Jtc) model itself in which the user can select either the already-calibrated distance function or the mathematical function. The empirically-derived function is, by far, the easiest to use and is, consequently, the default choice in *CrimeStat*. The discussion of it is on p. 35. However, the mathematical function can be used if there is inadequate data to construct an empirical distance decay function or if a particular form is desired.

Distance Modeling Using Mathematical Functions

We'll start by illustrating the use of the mathematical functions because this has been the traditional way that distance decay has been examined. The *CrimeStat* Jtc routine allows the user to define distance decay by a mathematical function.

Probability Distance Functions

The user selects one of five probability density distributions to define a likelihood that the offender has traveled a particular distance to commit a crime. The advantage of having five functions, as opposed to only one, is that it provides more flexibility in describing travel behavior. The travel distance distribution followed will vary by crime type, time of day, method of operation, and numerous other variables. The five functions allow an approach that can simulate more accurately travel behavior under different conditions. Each of these has parameters that can be modified, allowing a very large number of possibilities for describing travel behavior of a criminal.

Figure 9.3 illustrates the five types.[2] Briefly, they are:

Linear

The simplest type of distance model is a linear function. This model postulates that the likelihood of committing a crime at any particular location declines by a constant amount with distance from the offender's home. It is highest near the offender's home but drops off by a constant amount for each unit of distance until it falls to zero. The form of the linear equation is

$$f(d_{ij}) = \quad A + B * d_{ij} \tag{9.20}$$

358

Figure 9.1: Journey to Crime Interpolation Routine

Primary file:
Crime locations

Distance decay function

Reference grid
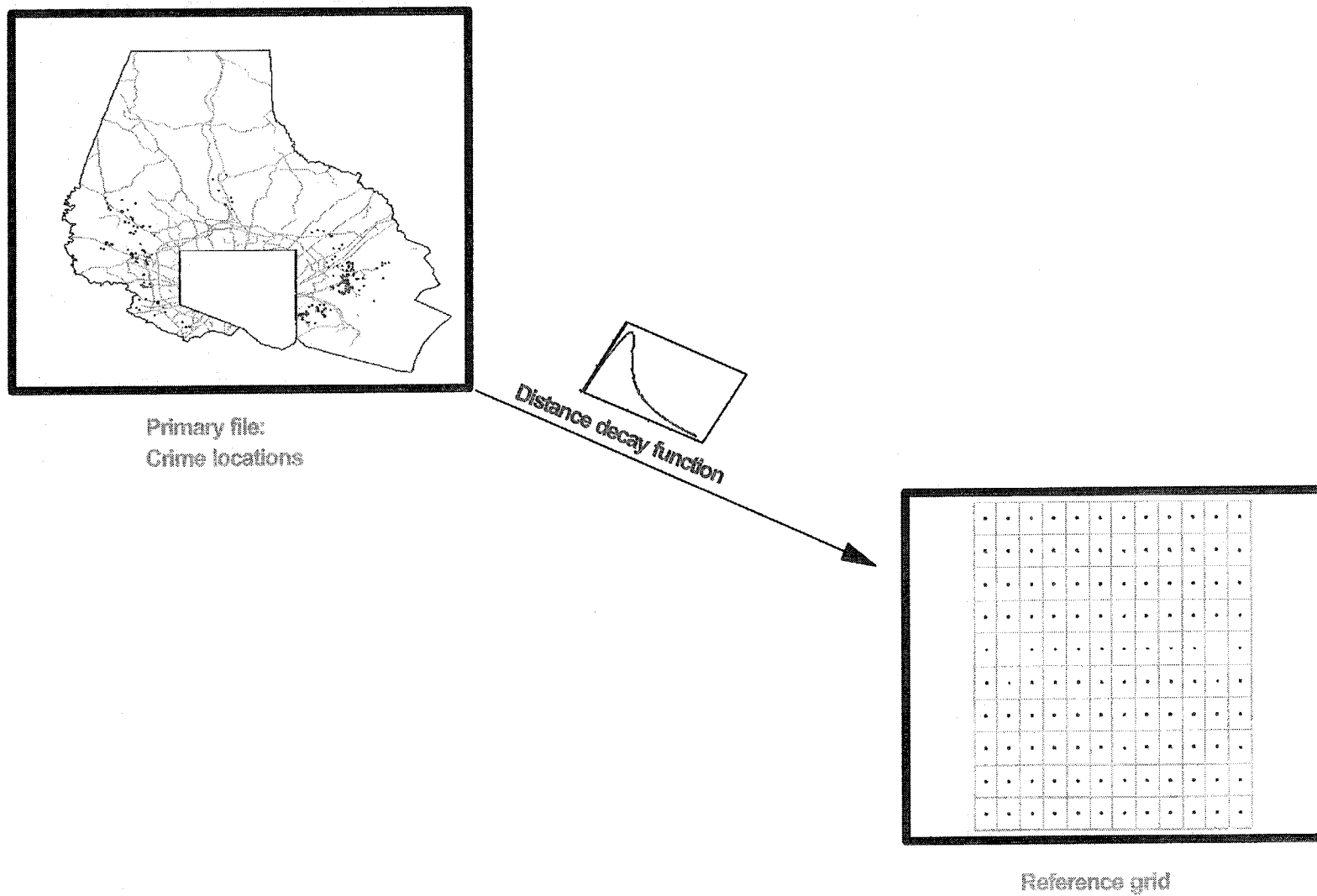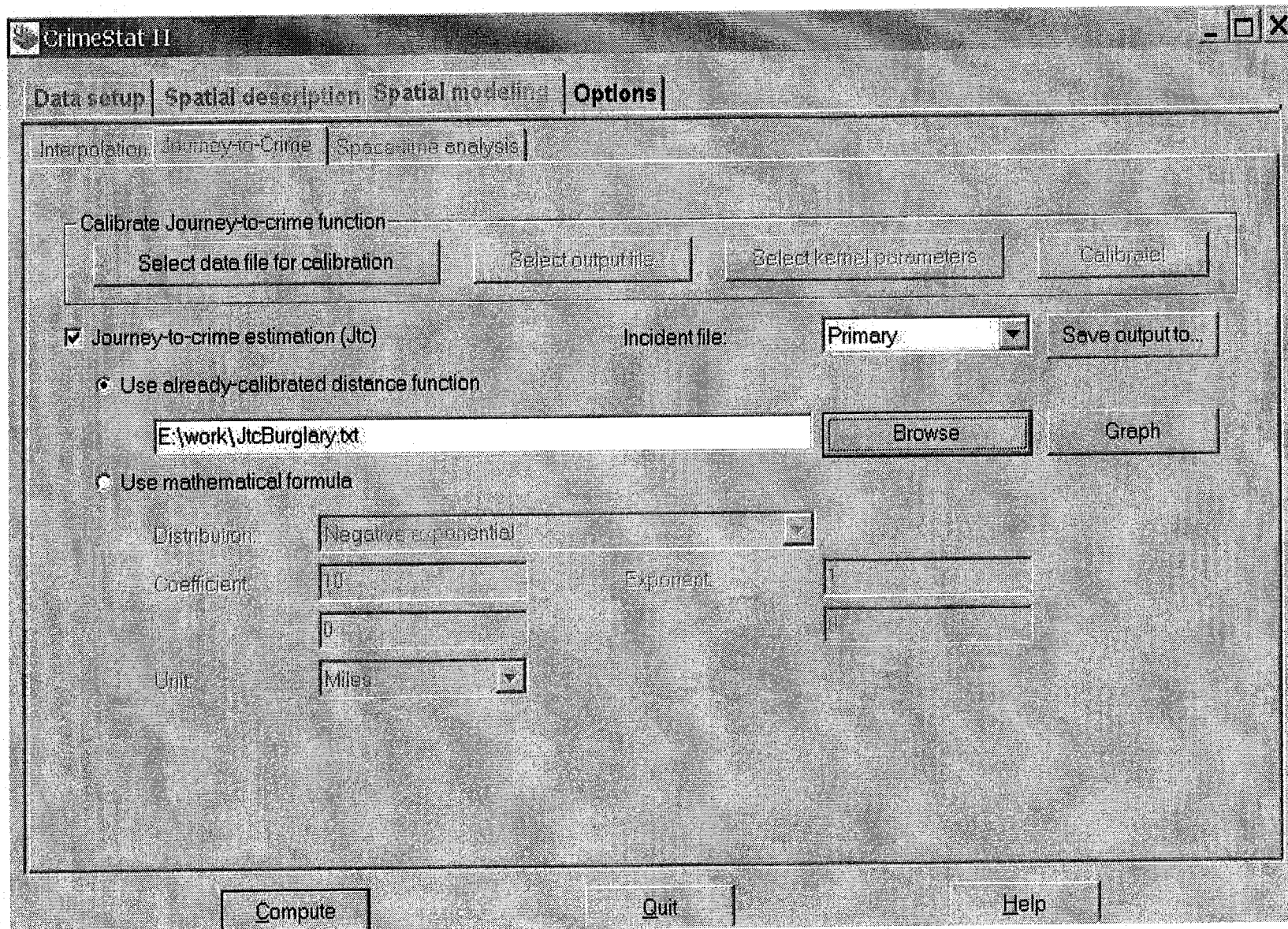
Figure 9.3:

Journey to Crime Travel Demand Functions
Five Mathematical Functions
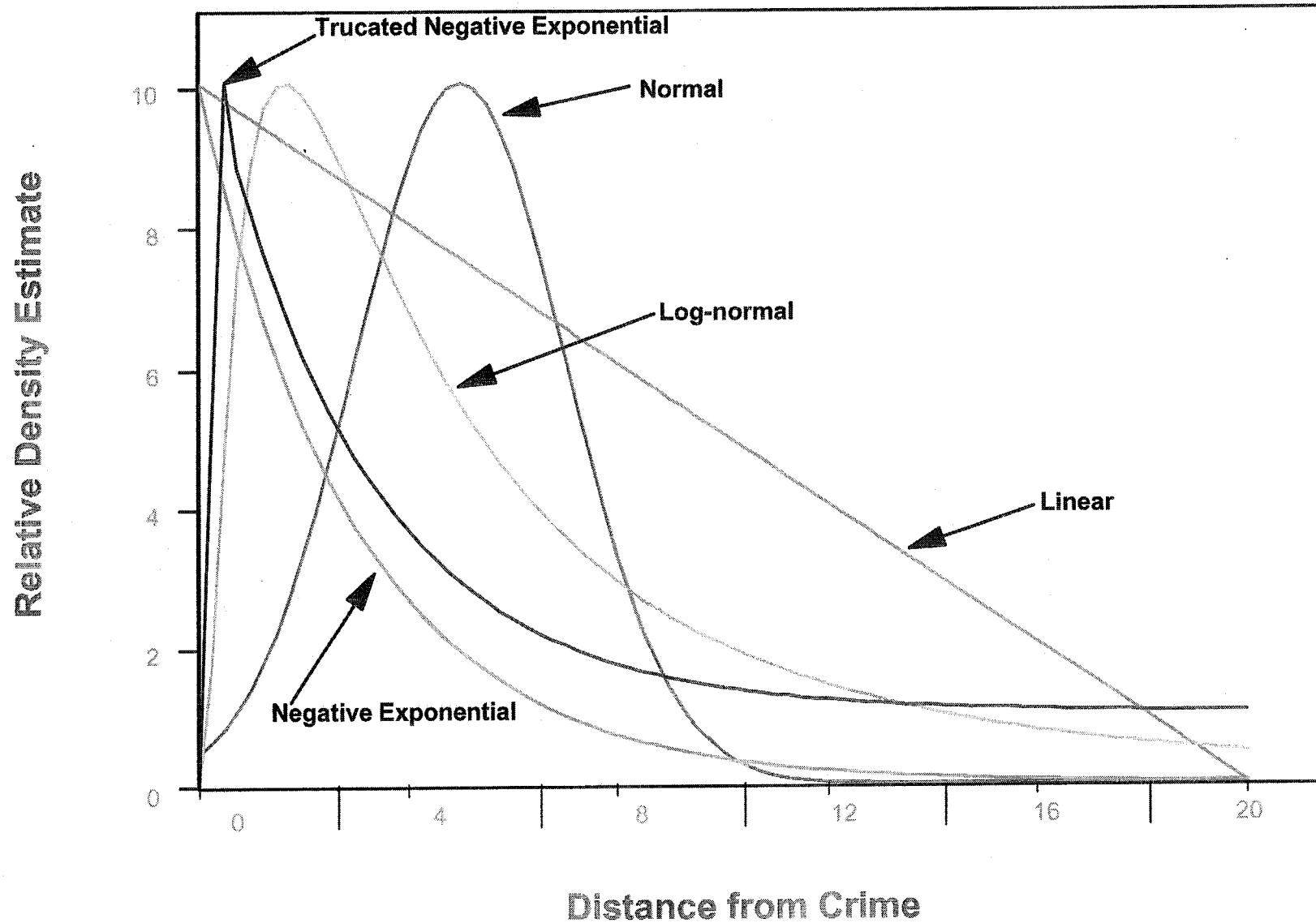
where $f(d_{ij})$ is the likelihood that the offender will commit a crime at a particular location, i, defined here as the center of a grid cell, d_{ij} is the distance between the offender's residence and location i, A is a slope coefficient which defines the fall off in distance, and B is a constant. It would be expected that the coefficient B would have a negative sign since the likelihood should decline with distance. The user must provide values for A and B. The default for A is 10 and for B is -1. This function assumes no buffer zone around the offender's residence. When the function reaches 0 (the X axis), the routine automatically substitutes a 0 for the function.

Negative Exponential

A slightly more complex function is the negative exponential. In this type of model, the likelihood is also highest near the offenders home and drops off with distance. However, the decline is at a constant *rate* of decline, thus dropping quickly near the offender's home until is approaches zero likelihood. The mathematical form of the negative exponential is

$$f(d_{ij}) = A*e^{-B*d_{ij}} \tag{9.21}$$

where $f(d_{ij})$ is the likelihood that the offender will commit a crime at a particular location, i, defined here as the center of a grid cell, d_{ij} is the distance between each reference location and each crime location, e is the base of the natural logarithm, A is the coefficient and B is an exponent of e. The user inputs values for A - the coefficient, and B - the exponent. The default for A is 10 and for B is 1. This function is similar to the Canter model (equation 9.19) except that the coefficient is calibrated. Also, like the linear function, it assumes no buffer zone around the offender's residence.

Normal

A normal distribution assumes the peak likelihood is at some optimal distance from the offender's home base. Thus, the function rises to that distance and then declines. The rate of increase prior to the optimal distance and the rate of decrease from that distance is symmetrical in both directions. The mathematical form is:

$$Z_{ij} = \frac{(d_{ij} - MeanD)}{S_d} \tag{9.22}$$

$$f(d_{ij}) = A * \frac{1}{S_d * SQRT(2\pi)} * e^{-0.5*Z_{ij}^2} \tag{9.23}$$

where $f(d_{ij})$ is the likelihood that the offender will commit a crime at a particular location, i (defined here as the center of a grid cell), d_{ij} is the distance between each reference location and each crime location, MeanD is the mean distance input by the user, S_d is the standard

362

deviation of distances, e is the base of the natural logarithm, and A is a coefficient. The user inputs values for MeanD, S_d, and A. The default values are 1 for each of these parameters.

By carefully scaling the parameters of the model, the normal distribution can be adapted to a distance decay function with an increasing likelihood for near distances and a decreasing likelihood for far distances. For example, by choosing a standard deviation greater than the mean (e.g., MeanD = $1, S_d = 2$), the distribution will be skewed to the left because the left tail of the normal distribution is not evaluated. The function becomes similar to the model postulated by Brantingham and Brantingham (1981) in that it is a single function which describes travel behavior..

Lognormal

The lognormal function is similar to the normal except it is more skewed, either to the left or to the right. It has the potential of showing a very rapid increase near the offender's home base with a more gradual decline from a location of peak likelihood (see Figure 9.3). It is also similar to the Brantingham and Brantingham (1981) model. The mathematical form of the function is:

$$f(d_{ij}) = A * \frac{1}{d^2_{ij} * S_d * SQRT(2\pi)} * e^{-[\ln(d^2_{ij}) - MeanD]^2 / 2 * s_d^2} \tag{9.24}$$

where $f(d_{ij})$ is the likelihood that the offender will commit a crime at a particular location, i, defined here as the center of a grid cell, d_{ij} is the distance between each reference location and each crime location, MeanD is the mean distance input by the user, S_d is the standard deviation of distances, e is the base of the natural logarithm, and A is a coefficient. The user inputs MeanD, S_d, and A. The default values are 1 for each of these parameters.

Truncated Negative Exponential

The truncated negative exponential is a joined function made up of two distinct mathematical functions - the linear and the negative exponential. For the near distance, a positive linear function is defined, starting at zero likelihood for distance 0 and increasing to d_p, a location of peak likelihood. Thereupon, the function follows a negative exponential, declining quickly with distance. The two mathematical functions making up this spline function are

Linear: $\quad f(d_{ij}) = 0 + B*d_{ij} = B*d_{ij} \qquad$ for $d_{ij} \geq 0, d_ij \leq d_p \qquad$ (9.25)

Negative
Exponential: $f(d_{ij}) = A*e^{-C*d_{ij}} \qquad$ for $X_i > d_p \qquad$ (9.26)

where d_{ij} is the distance from the home base, B is the slope of the linear function (default=+1) and for the negative exponential function A is a coefficient and C is an

363

exponent. Since the negative exponential only starts at a particular distance, $Maxd_{ij}$, A, is assumed to be the intercept if the Y-axis were transposed to that distance.

This function is the closest approximation to the Rossmo model (equations 9.13 and 9.16). However, it differs in several mathematical properties. First, the 'near home base' function is linear (equation 9.25), rather than a non-linear function (equation 9.13). It assumes a simple increase in travel likelihoods by distance from the home base, up to the edge of the safety zone.[3] Second, the distance decay part of the function (equation 9.26) is a negative exponential, rather than an inverse distance function (equation 9.13); consequently, it is more stable when distances are very close to zero (e.g., for a crime where there is no 'near home base' offset).

Calibrating an Appropriate Probability Distance Function

The mathematics are relatively straightforward. However, how does one know which distance function to use? The answer is to get some data and calibrate it. It is important to obtain data from a sample of known offenders where both their residence at the time they committed crimes as well as the crime locations are known. This is called the *calibration data set*. Each of the models are then tested against the calibration data set using an approach similar to that explained below. An error analysis is conducted to determine which of the models best fits the data. Finally, the 'best fit' model is used to estimate the likelihood that a particular serial offender lives at any one location. Though the process is tedious, once the parameters are calculated they can be used repeatedly for predictions.

Because every jurisdiction is unique in terms of travel patterns, it is important to calibrate the parameters for the particular jurisdiction. While there may be some similarities between cities (e.g., Eastern "centralized" cities v. Western "automobile" cities), there are always unique travel patterns defined by the population size, historical road pattern, and physical geography. Consequently, it is necessary to calibrate the parameters anew for each new city. Ideally, the sample should be a large enough so that a reliable estimate of the parameters can be obtained. Further, the analyst should check the errors in each of the models to ensure that the best choice is used for the Jtc routine. However, once it has been completed, the parameters can be re-used for many years and only periodically re-checked.

Data Set from Baltimore County

I'll illustrate with data from Baltimore County. The steps in calibrating the *Jtc* parameters were as follows:

1. 49,083 matched arrest and incident records from 1992 through 1997 were obtained in order to provide data on where the offender lived in relation to the crime location for which they were arrested.[4]

364

2.   The data set was checked to ensure that there were X and Y coordinates for both the arrested individual's residence location and the crime incident location for which the individual was being charged. The data were cleaned to eliminate duplicate records or entries for which either the offender's residence or the incident location were missing. The final data set had 41,424 records. There were many multiple records for the same offender since an individual can commit more than one crime. In fact, more than half the records involved individuals who were listed two or more times. The distribution of offenders by the number of offenses for which they were charged is seen in Table 9.1. As would be expected, a small proportion of individuals account for a sizeable proportion of crimes; approximately 30% of the offenders in the database accounted for 56% of the incidents.

3.   The data were imported into a spreadsheet, but a database program could equally have been used. For each record, the direct distance between the arrested individual's residence and the crime incident location was calculated. Chapter 2 presented the formulas for calculating direct distances between two locations and are repeated in endnote 5.[5]

4.   The records were sorted into sub-groups based on different types of crimes. For the Baltimore County example, eleven categories of crime incident were used. Table 9.2 presents the categories with their respective sample sizes. Of course, other sub-groups could have been identified. Each sub-group was saved as a separate file. The same records can be part of multiple files (e.g., a record could be included in the 'all robberies' file as well as in the 'commercial robberies' file). All records were included in the 'all crimes' file.

5.   For each type of crime, the file was grouped into distance intervals of 0.25 miles each. This involved two steps. First, the distance between the offender's residence and the crime location was sorted in ascending order. Second, a frequency distribution was conducted on the distances and grouped into 0.25 mile intervals (often called *bins*). The degree of precision in distance would depend on the size of the data set. For 41,426 records, quarter mile bins were appropriate.

6.   For each type of crime, a new file was created which included only the frequency distribution of the distances broken down into quarter mile distance intervals, d_i.

7.   In order to compare different types of crimes, which will have different frequency distributions, two new variables were created. First, the frequency in the interval was converted into the percentage of all crimes of in each interval by dividing the frequency by the total number of incidents, N, and multiplying by 100. Second, the distance interval was adjusted. Since the interval is a range with a starting distance and an ending distance but has been identified by spreadsheet program as the beginning distance only, a

365

Table 9.1

Number of Offenders and Offenses in Baltimore County: 1993-1997
Journey to Crime Database

Number of Offenses	Number of Individuals	Percent of Offenders	Number of Incidents	Percent of Incidents
1	18,174	70.0%	18,174	43.9%
2	4,443	17.1%	8,886	21.5%
3	1,651	6.4%	4,953	12.0%
4	764	2.9%	3,056	7.4%
5	388	1.5%	1,940	4.7%
6-10	482	1.9%	3,383	8.2%
11-15	61	0.2%	757	1.8%
16-20	10	<0.0%	175	0.4%
21-25	3	<0.0%	67	0.2%
26-30	0	<0.0%	0	0.0%
30+	1	<0.0%	33	<0.0%
	25,977		41,424	

Table 9.2

Baltimore County Files Used for Calibration

Crime Type	Sample Size
All crimes	41,426
Homicide	137
Rape	444
Assault	8,045
Robbery (all)	3,787
Commercial robbery	1,193
Bank robbery	176
Burglary	4,694
Motor vehicle theft	2,548
Larceny	19,806
Arson	338

366

small fraction, representing the midpoint of the interval, is added to the distance interval. In our case, since each interval is 0.25 miles wide, the adjustment is half of this, 0.125. Each new file, therefore, had four variables: the interval distance, the adjusted interval distance, the frequency of incidents within the interval (the number of cases falling into the interval), and the percentage of all crimes of that type within the interval.

8. Using a general statistical program, a series of regression equations was set up to model the frequency (or the percentage) as a function of distance. In this case, I used *Systat* (Systat, Inc, 1994), but other statistical packages could equally have been used. Again, because comparisons between different types of crimes were of interest, the percentage of crimes (by type) within an interval was used as the dependent variable (and was defined as a percentage, i.e., 11.51% was recorded as 11.51). Five equations testing each of the five models were set up.

Linear

For the linear function, the test was

$$Pct_i = A + Bd_i \tag{9.27}$$

where Pct_i is the percentage of all crimes of that type falling into interval i, d_i is the distance for interval i, A is the intercept, and B is the slope. A and B are estimated directly from the regression equation.

Negative Exponential

For the negative exponential function, the variables have to be transformed to estimate the parameters. The function is

$$Pct_i = A * e^{-B*d_i} \tag{9.28}$$

A new variable is defined which is the natural logarithm of the percentage of all crimes of that type falling into the interval, *ln(Pct_i)*. This term was then regressed against the distance interval, d_i.

$$\ln(Pct_i) = K - B*d_i \tag{9.29}$$

However, since the original equation has been transformed into a log function, B is the coefficient and A can be calculated directly from

$$\ln(Pct_i) = \ln(A) - B*d_i \tag{9.30}$$

$$A = e^K \tag{9.31}$$

367

If the percentage in any bin was 0 (i.e., $Pct_i = 0$), then a value of -16 was taken since the natural logarithm of 0 cannot be solved (it approximates -16 as the percentage approaches 0.0000001).

Normal

For the normal function, a more complex transformation must be used. The normal function in the model is

$$Pct_i = A * \frac{1}{S_d * SQRT(2\pi)} * e^{-0.5*Z_{ij}^2} \tag{9.32}$$

First, a standardized Z variable for the distance, d_i, is created

$$Z_i = \frac{(d_i - MeanD)}{S_d} \tag{9.33}$$

where MeanD is the mean distance and S_d is the standard deviation of distance. These are calculated from the original data file (*before* creating the file of frequency distributions). Second, a normal transformation of Z is constructed with

$$Normal(Z_i) = \frac{1}{S_d * SQRT(2\pi)} * e^{-0.5*Z_{ij}^2} \tag{9.34}$$

Finally, the normalized variable is regressed against the percentage of all crimes of that type falling into the interval, Pct_i with *no* constant

$$Pct_i = A * Normal(Z_i) \tag{9.35}$$

A is estimated by the regression coefficient.

Lognormal

For the lognormal function, another complex transformation must be done. The lognormal function for the percentage of all crimes of a type for a particular distance interval is

$$Pct_i = A * \frac{1}{d_{ij}^2 * S_d * SQRT(2\pi)} * e^{-[\ln(d_i^2) - MeanD]^2 / 2 * S_d^2} \tag{9.36}$$

The transformation can be created in steps. First, create L

$$L = \ln(d_i^2) \tag{9.37}$$

368

Second, create M

$$M = (1 - MeanD)^2 \tag{9.38}$$

Third, create O

$$O = \frac{m}{(2 * S_d^2)} \tag{9.39}$$

Fourth, create P by raising e to the Oth power.

$$P = e^O \tag{9.40}$$

Fifth, create the lognormal conversion, Lnormal

$$Lnormal(d_i) = A * \frac{1}{d^2_{ij} * S_d * SQRT(2\pi)} * P \tag{9.41}$$

Finally, the lognormal variable is regressed against the percentage of all crimes of that type falling into the interval, Pct_i with *no* constant

$$Pct_i = A * Lnormal(d_i) \tag{9.42}$$

A is estimated with the regression coefficient.

Truncated Negative Exponential

For the truncated negative exponential function, two models were set up. The first applied to the distance range from 0 to the distance at which the percentage (or frequency) is highest, $Maxd_i$. The second applied to all distances greater than this distance

Linear: $\quad Pct_i = A + Bd_i \quad\quad$ for $d_{ij} \geq 0$, $d_ij \leq Maxd_{ij}$ $\tag{9.43}$

Negative $\quad\quad\quad\quad -C*di$
Exponential: $Pct_i = A*e \quad\quad$ for $d_ij > Maxd_{ij}$ $\tag{9.44}$

To use this function, the user specifies the distance at which the peak likelihood occurs, d_p (the *peak distance*) and the value for that peak likelihood, P (the *peak likelihood*). For the negative exponential function, the user specifies the exponent, C.

369

In order to splice the two equations together (the spline), the *CrimeStat* truncated negative exponential routine starts the linear equation at the origin and ends it at the highest value. Thus,

$$A = 0 \tag{9.45}$$

$$B = P/d_p \tag{9.46}$$

where P is the peak likelihood and d_p is the peak distance.

The exponent, C, can be estimated by transforming the dependent variable, Pct_i, as in the negative exponential above (equation 9.28) and regressing the natural log of the percentage $(\ln(Pct_i)$ against the distance interval, d_i, *only* for those intervals that are greater than the peak distance. I have found that estimating the transformed equation with a coefficient, A in

$$Pct_i = A * e^{-C*d_i} \tag{9.47}$$

$$\ln(Pct_i) = Ln(A) - C*d_i \tag{9.48}$$

gives a better fit to the equation. However, the user need only input the exponent, C, in the Jtc routine as the coefficient, A, of the negative exponential is calculated internally to produce a distance value at which the peak likelihood occurs. The formula is:

$$A = e^{\ln(P) + C*(d_p - d_i)} \tag{9.49}$$

where P is the peak likelihood, d_p is the distance for the peak likelihood, C is an exponent (assumed to be positive) and d_i is the distance interval for the histogram.

9. Once the parameters for the five models have been estimated, they can be compared to see which one is best at predicting the travel behavior for a particular type of crime. It is to be expected that different types of crimes will have different optimal models and that the parameters will also vary.

Examples from Baltimore County

Let's illustrate with the Baltimore County data. Figure 9.4 shows the frequency distribution for all types of crime in Baltimore County. As can be seen, at the nearest distance interval (0 to 0.25 miles with an assigned 'adjusted' midpoint of 0.125 miles), about 6.9% of all crimes occur within a quarter mile of the offender's residence (it can be seen on the Y-axis). However, for the next interval (0.25 to 0.50 miles with an assigned midpoint of 0.375 miles), almost 10% of all crimes occur at that distance (9.8%). In

370

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

In order to splice the two equations together (the spline), the *CrimeStat* truncated negative exponential routine starts the linear equation at the origin and ends it at the highest value. Thus,

$$A = 0 \tag{9.45}$$

$$B = P/d_p \tag{9.46}$$

where P is the peak likelihood and d_p is the peak distance.

The exponent, C, can be estimated by transforming the dependent variable, Pct_i, as in the negative exponential above (equation 9.28) and regressing the natural log of the percentage $(\ln(Pct_i)$ against the distance interval, d_i, *only* for those intervals that are greater than the peak distance. I have found that estimating the transformed equation with a coefficient, A in

$$Pct_i = A * e^{-C*d_i} \tag{9.47}$$

$$\ln(Pct_i) = Ln(A) - C*d_i \tag{9.48}$$

gives a better fit to the equation. However, the user need only input the exponent, C, in the Jtc routine as the coefficient, A, of the negative exponential is calculated internally to produce a distance value at which the peak likelihood occurs. The formula is:

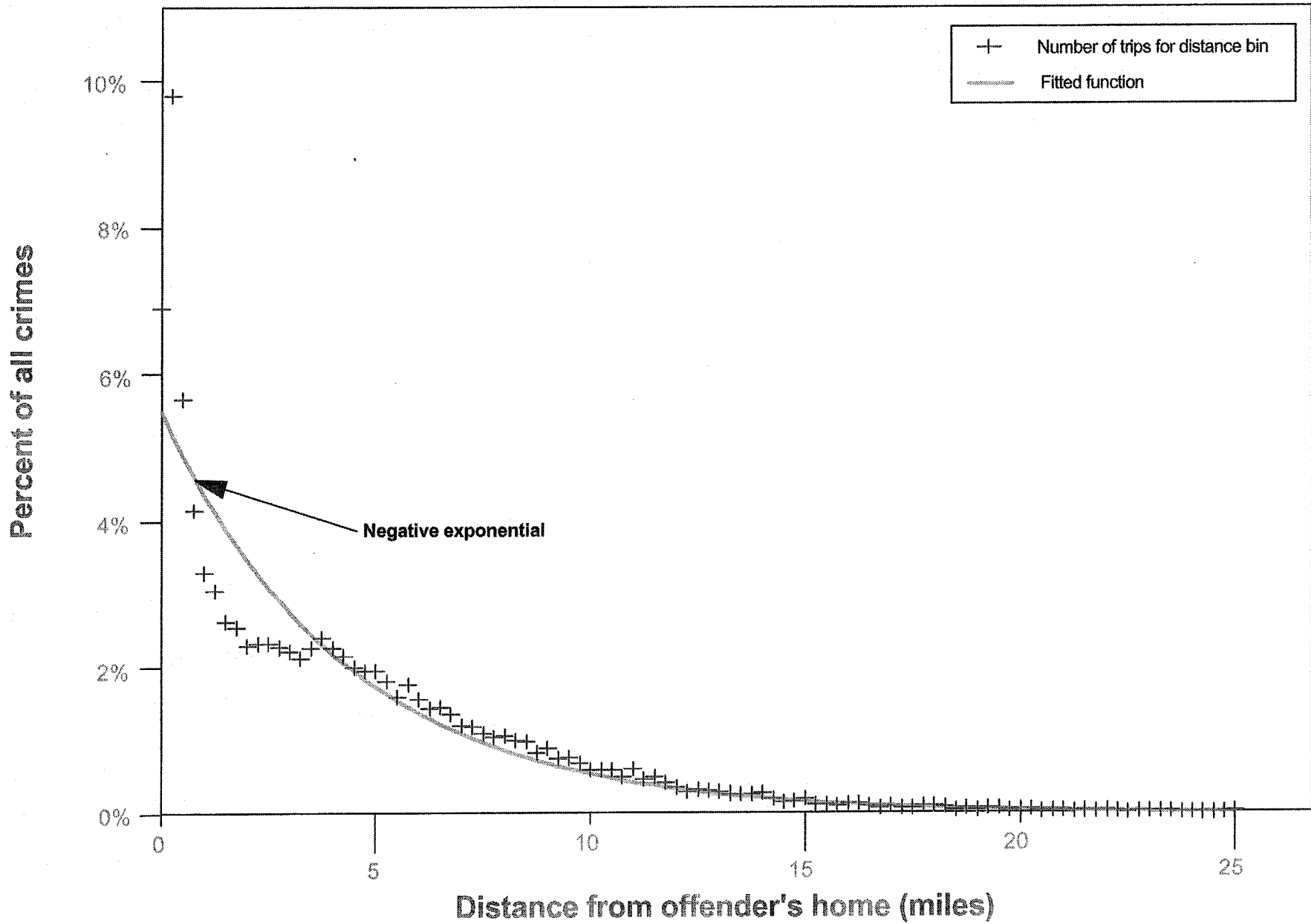$$A = e^{\ln(P) + C*(d_p - d_i)} \tag{9.49}$$

where P is the peak likelihood, d_p is the distance for the peak likelihood, C is an exponent (assumed to be positive) and d_i is the distance interval for the histogram.

9. Once the parameters for the five models have been estimated, they can be compared to see which one is best at predicting the travel behavior for a particular type of crime. It is to be expected that different types of crimes will have different optimal models and that the parameters will also vary.

Examples from Baltimore County

Let's illustrate with the Baltimore County data. Figure 9.4 shows the frequency distribution for all types of crime in Baltimore County. As can be seen, at the nearest distance interval (0 to 0.25 miles with an assigned 'adjusted' midpoint of 0.125 miles), about 6.9% of all crimes occur within a quarter mile of the offender's residence (it can be seen on the Y-axis). However, for the next interval (0.25 to 0.50 miles with an assigned midpoint of 0.375 miles), almost 10% of all crimes occur at that distance (9.8%). In

370

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice./boilerplate

Figure 9.4:
Journey to Crime Distances: All Crimes
Negative Exponential Distribution

subsequent intervals, however, the percentage decreases, a little less than 6% for 0.50 to 0.75 miles (with the midpoint being 0.625 miles), a little more than 4% for 0.75 to 1 mile (the midpoint is 0.875 miles), and so forth.

The best fitting statistical function was the negative exponential. The particular equation is

$$Pct_i = 5.575 * e^{-0.229*d_i} \qquad\qquad (9.50)$$

This is shown with the solid line. As can be seen, the fit is good for most of the distances, though it underestimates at close to zero distance and overestimates from about a half mile to about four miles. There is only slight evidence of decreased activity near to the location of the offender.

However, the distribution varies by type of crime. With the Baltimore County data, property crimes, in general, occur farther away than personal crimes. The truncated negative exponential generally fit property crimes better, lending support for the Brantingham and Brantingham (1981) framework for these types. For example, larceny offenders have a definite safety zone around their residence (figure 9.5). Fewer than 2% of larceny thefts occur within a quarter mile of the offender's residence. However, the percentage jumps to about 4.5% from a quarter mile to a half. The truncated negative exponential function fits the data reasonably well though it overestimates from about 1 to 3 miles and underestimates from about 4 to 12 miles.

Similarly, motor vehicle thefts show decreased activity near the offender's resident, though it is less pronounced than larceny theft. Figure 9.6 shows the distribution of motor vehicle thefts and the truncated negative exponential function which was fit to the data. As can be seen, the fit is reasonably good though it tends to underestimate middle range distances (approximately 3-12 miles).

Some types of crime, on the other hand, are very difficult to fit. Figure 9.7 shows the distribution of bank robberies. Partly because there were a limited number of cases (N=176) and partly because it's a complex pattern, the truncated negative exponential gave the best fit, but not a particularly good one. As can be seen, the linear ('near home') function underestimates some of the near distance likelihoods while the negative exponential drops off too quickly; in fact, to make this function even plausible, the regression was run only up to 21 miles (otherwise, it underestimated even more).

For some crimes, it was very difficult to fit any single function. Figure 9.8 shows the frequency distribution of 137 homicides with three functions being fitted to the data - the truncated negative exponential, the lognormal, and the normal. As can be seen each function fits only some of the data, but not all of it.

372

Figure 9.5:
Journey to Crime Distances: Larceny
Truncated Negative Exponential Function

Number of trips for distance bin

Fitted function

Truncated negative exponential

Percent of all larceny thefts

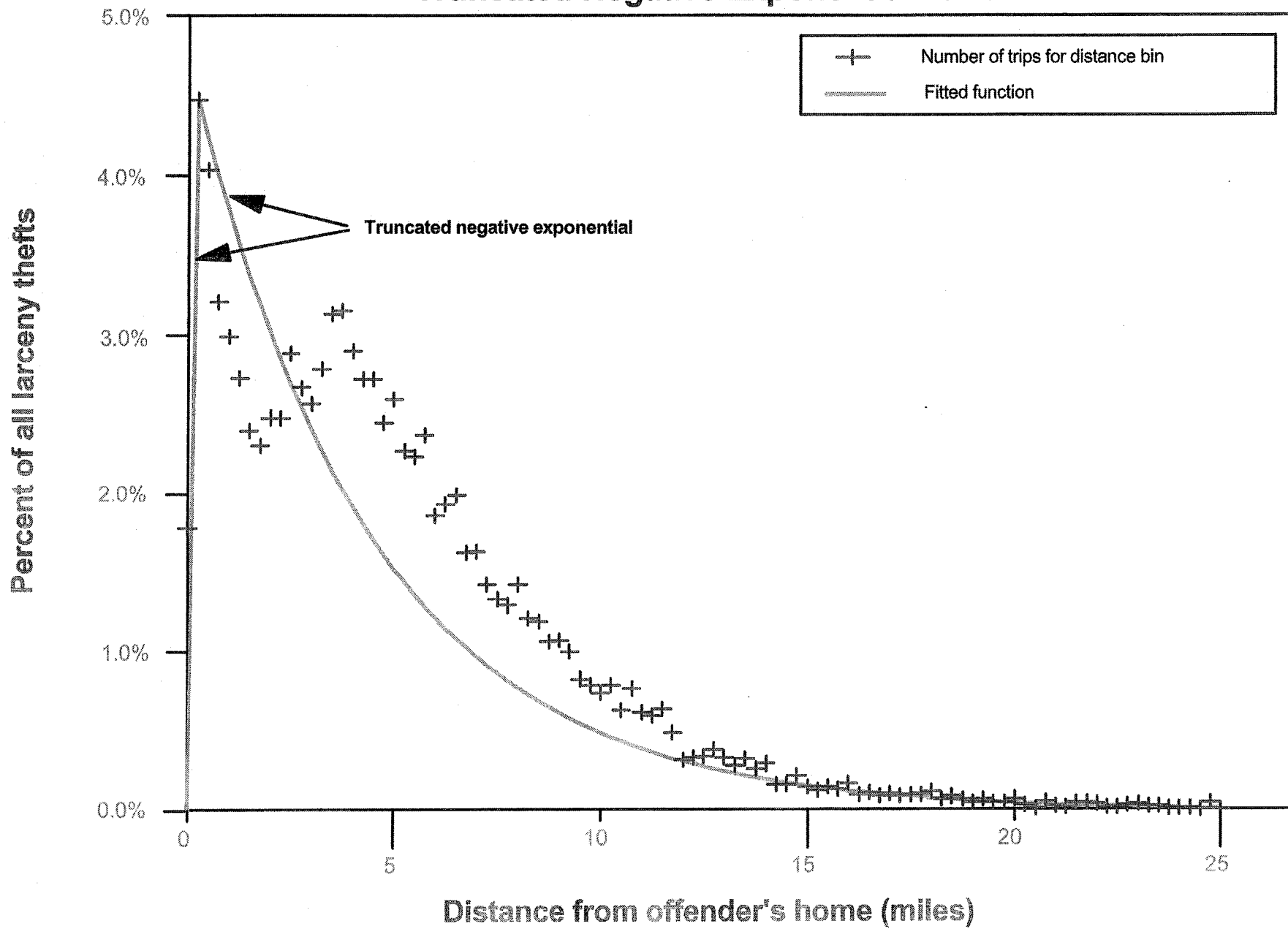**Distance from offender's home (miles)**

Figure 9.6:

Journey to Crime Distances: Vehicle Theft
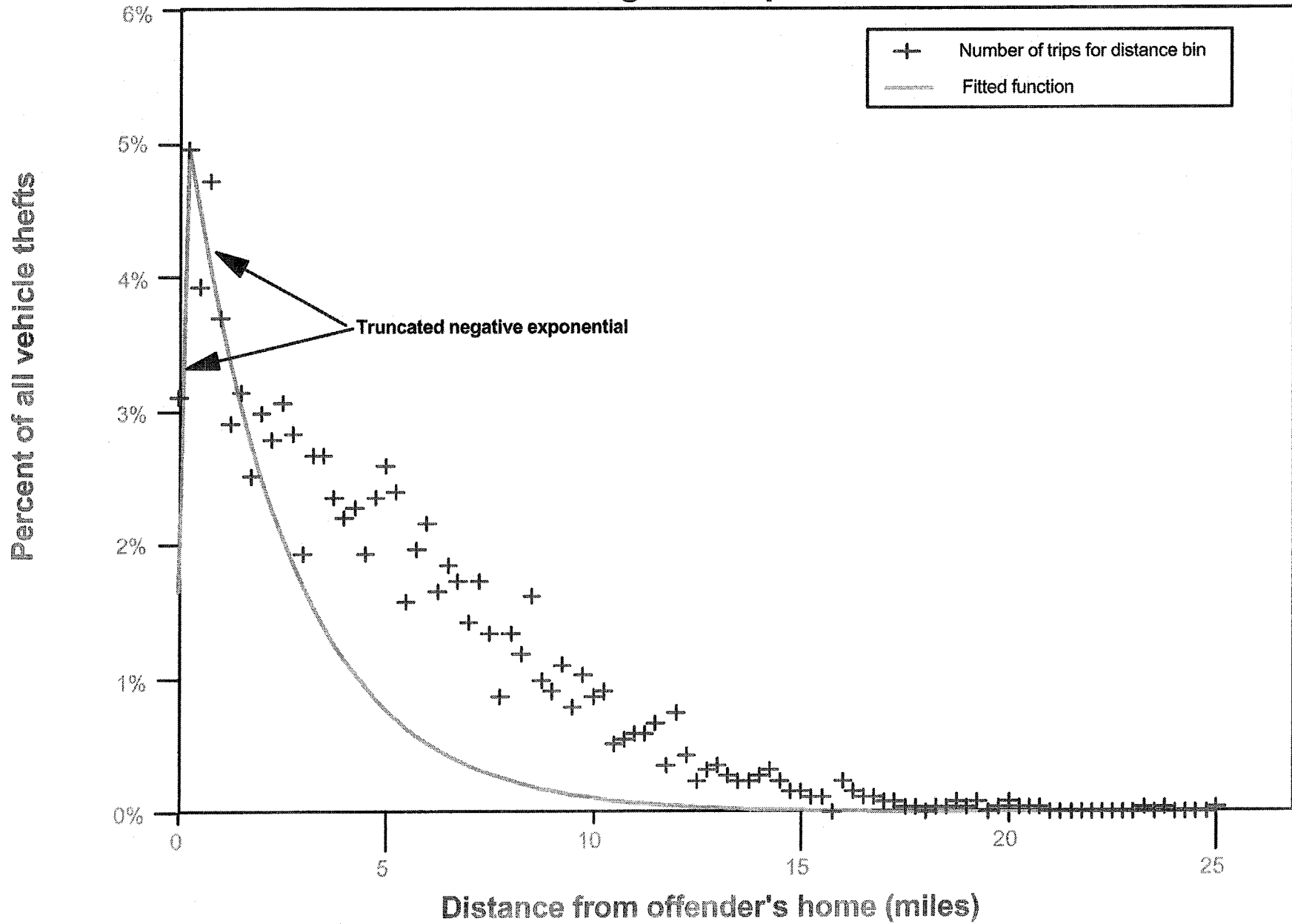## Truncated Negative Exponential Function

Figure 9.7:
Journey to Crime Distances: Bank Robbery
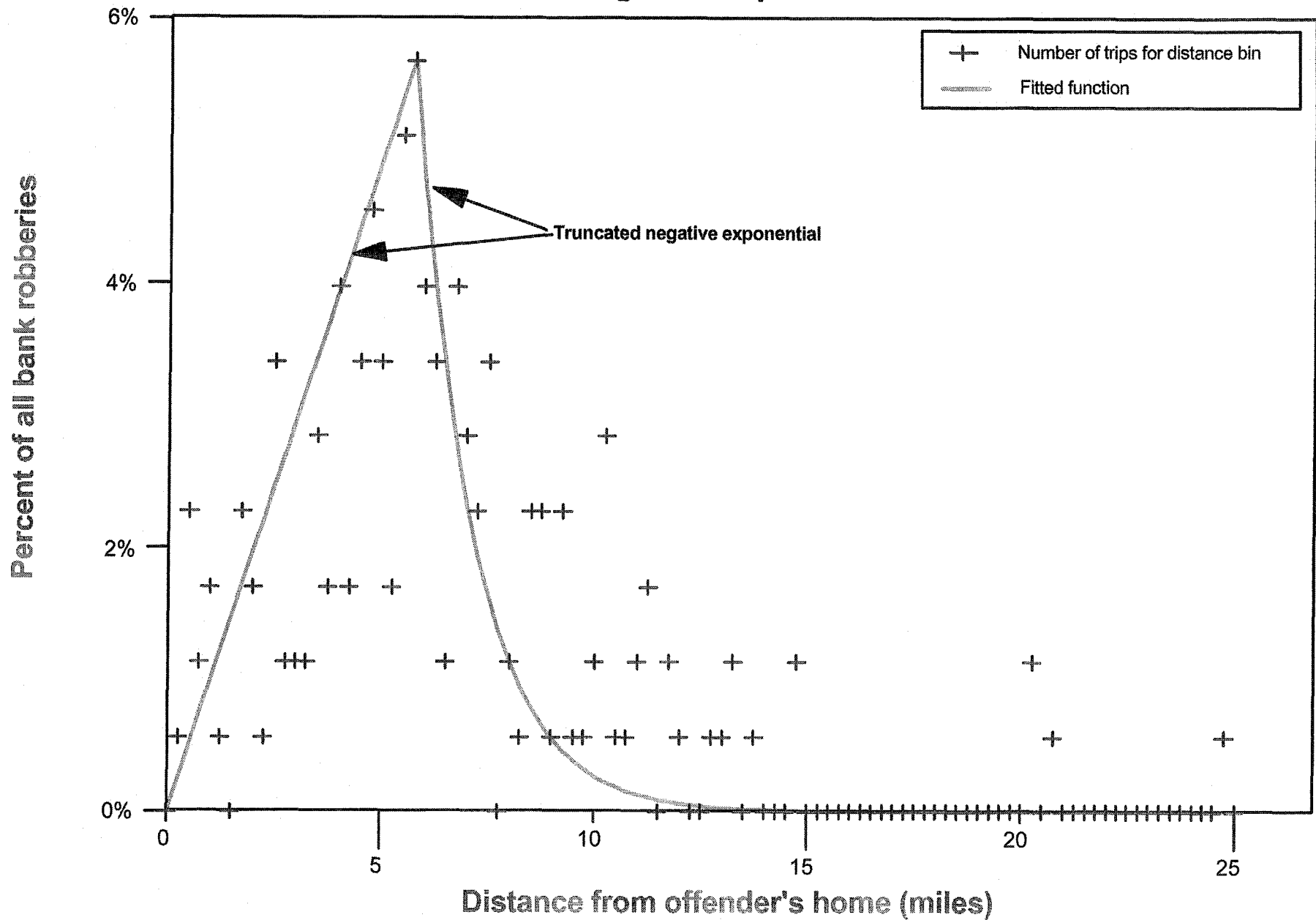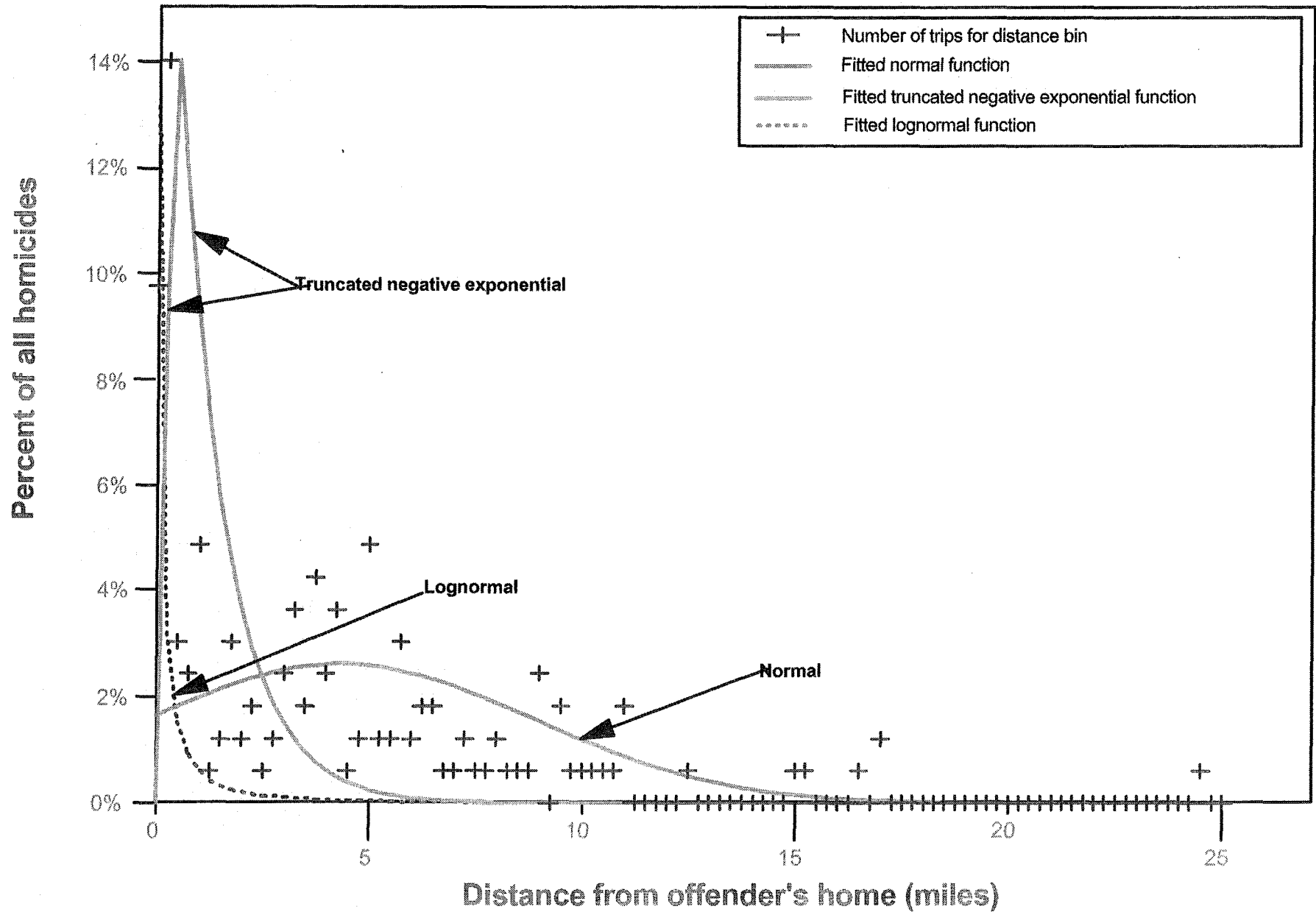## Truncated Negative Exponential Function

Figure 9.8:
Journey to Crime Distances: Homicide
Normal, Lognormal, and Truncated Negative Exponential Functions

Testing for Residual Errors in the Model

In short, the five mathematical functions allow a user to fit a variety of distance decay distributions. Each of the models will predict some parts of the distribution better than others. Consequently, it is important to conduct an error analysis to determine which model is 'best'. In an error analysis, the residual error is defined as

$$\text{Residual error} = Y_i - E(Y_i) \tag{9.51}$$

where Y_i is the observed (actual) likelihood for distance i and $E(Y_i)$ is the likelihood predicted by the model. If raw numbers of incidents are used, then the likelihoods are the number of incidents for a particular distance. If the number of incidents are converted into proportions (i.e., probabilities), then the likelihoods are the proportions of incidents for a particular distance.

The choice of 'best model' will depend on what part of the distribution is considered most important. Figure 9.9, for example, shows the residual errors on vehicle theft for the five fitted models. That is, each of the five models was fit to the proportion of vehicle thefts by distance intervals (as explained above). For each distance, the discrepancy between the actual percentage of vehicle thefts in that interval and the predicted percentage was calculated. If there was a perfect fit, then the discrepancy (or residual) was 0%. If the actual percentage was greater than the predicted (i.e., the model underestimated), then the residual was positive; if the actual was smaller than the predicted (i.e., the model overestimated), then the residual was negative.

As can be see in figure 9.9, the truncated negative exponential fit the data well from 0 to about 5 miles, but then became poorer than other models for longer distances. The negative exponential model was not as good as the truncated for distances up to about 5 miles, but was better for distances beyond that point. The normal distribution was good for distances from about 10 miles and farther. The lognormal was not particularly good for any distances other than at 0 miles, nor was the linear.

The degree of predictability varied by type of crime. For some types, particularly property crimes, the fit was reasonably good. I obtained R^2 in the order of 0.86 to 0.96 for burglary, robbery, assault, larceny, and auto theft. For other types of crime, particularly violent crimes, the fit was not very good with R^2 values in the order of 0.53 (rape), 0.41 (arson) and 0.30 (homicide). These R^2 values were for the entire distance range; for any particular distance, however, the predictability varied from very high to very low.

In modeling distance decay with a mathematical function, a user has to decide which part of the distribution is the most important as no simple mathematical function will normally fit all the data (even approximately). In these cases, I assumed that the near distances were more important (up to, say, 5 miles) and, therefore, selected the model which 'best' fit those distances (see table 9.2). However, it was not always clear which model was best, even with that limited criteria.

377

Figure 9.9:
Residual Error for Jtc Mathematical Models
Vehicle Theft
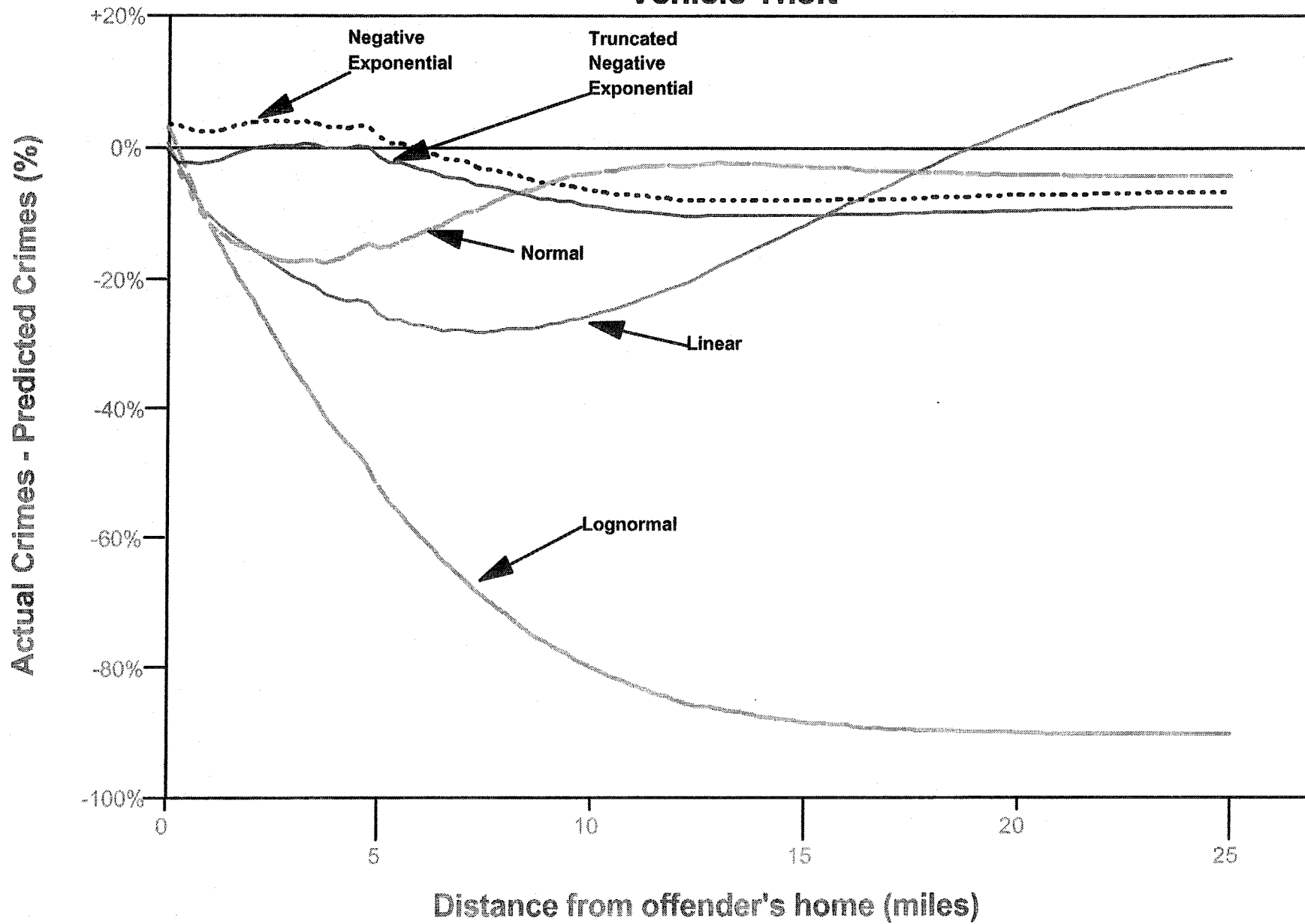
Using *CrimeStat* for Geographic Profiling

Brent Snook, Paul J. Taylor & Craig Bennell
Department of Psychology, University of Liverpool

A challenge for researchers providing investigative support is to use information about crime locations to prioritize geographic areas according to how likely they are to contain the offender's residence. One prescient solution to this problem uses *probability distance functions* to assign a likelihood value to the activity space around each crime location. A research goal is to identify the function that assigns the highest likelihood to the offender's actual residence, since this should prove more efficient in future investigations.

CrimeStat was used to test of the effectiveness of two functions for a sample of 68 German serial murder cases, using a measure known as *error distance*. The top figures below illustrate the two functions used and the bottom figures portray the corresponding effectiveness of the functions by plotting the percentage of the sample 'located' by error distance. A steeper effectiveness curve indicates that home locations were closer to the point of highest probability and that, consequently, the probability distance function was more efficient. In this particular test, no difference was found between the two functions in their ability to classify geographic areas.

Truncated Negative Exponential

Negative Exponential

Distance Decay Model

Cumulative Error Distance
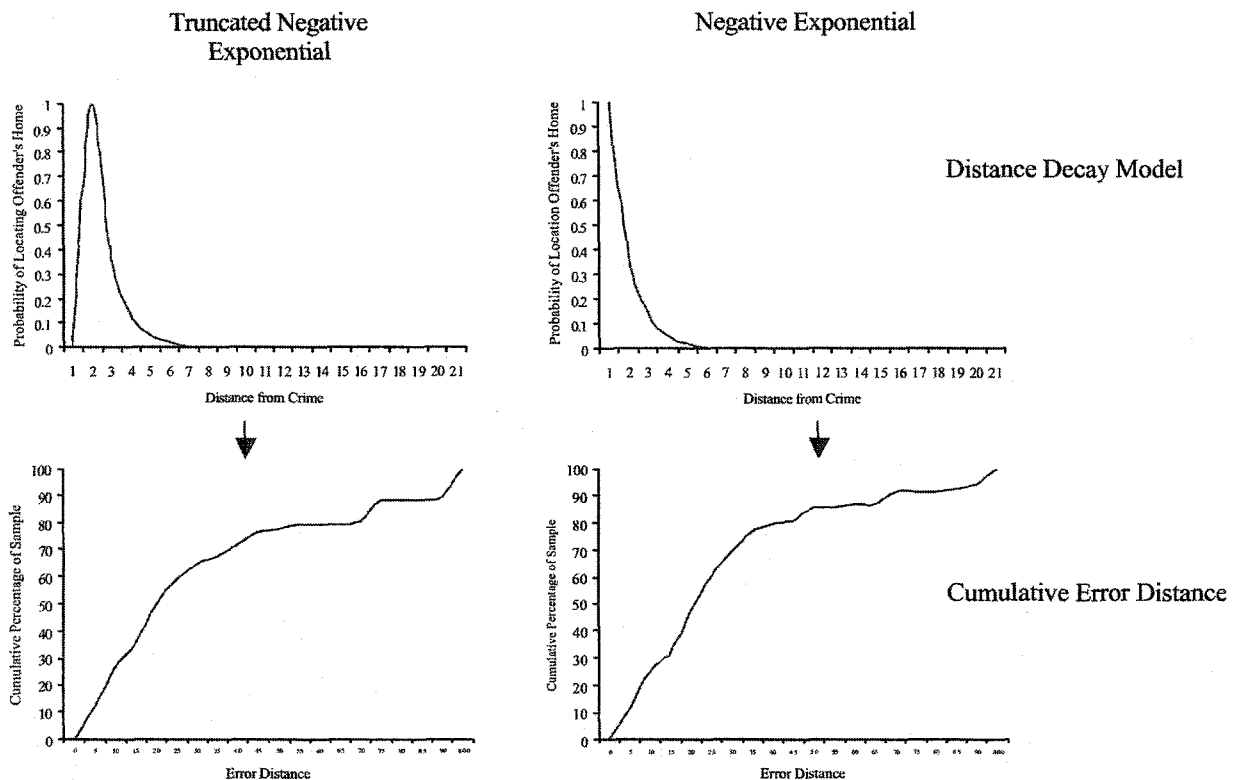
Original Article: Taylor, P.J., Bennell, C., & Snook B. (2002) *Problems of Classification in Investigative Psychology*. Proceedings of the 8th Conference of the International Federation of Classification Societies, Krakow, Poland

Problems with Mathematical Distance Decay Functions

There are several reasons that mathematical models of distance decay distributions, such as illustrated in the Jtc routine, do not fit data very well. First, as mentioned earlier, few cities have a completely symmetrical grid structure or even one that is approximately grid-like (there are exceptions, of course). Limitations of physical topography (mountains, oceans, rivers, lakes) as well as different historical development patterns makes travel asymmetrical around most locations.

Second, there is population density. Since most metropolitan areas have much higher intensity of land use in the center (i.e., more activities and facilities), travel tends to be directed towards higher land use intensity than away from them. For origin locations that are not directly in the center, travel is more likely to go towards the center than away from it.

This would be true of an offender as well. If the person were looking for either persons or property as 'targets', then the offender would be more likely to travel towards the metropolitan center than away from it. Since most metropolitan centers have street networks that were laid out much earlier, the street network tends to be irregular. Consequently, trips will vary by location within a metropolitan area. One would expect shorter trips by an offender living close to the metropolitan center than one living farther away; shorter trips for offenders living in more built-up areas than in lower density areas; shorter trips for offenders in mixed use neighborhoods than in strictly residential neighborhoods; and so forth. Thus, the distribution of trips of any sort (in our case, crime trips from a residential location to a crime location), will tend to follow an irregular, distance decay type of distribution. Simple mathematical models will not fit the data very well and will make many errors.

Third, the selection of a best mathematical function is partly dependent on the interval size used for the bins. In the above examples, an interval size of 0.25 miles was used to calculate the frequency distribution. With a different interval size (e.g., 0.5 miles), however, a slightly different distribution is obtained. This effects the mathematical function that is selected as well as the parameters that are estimated. For example, the issue of whether there is a safety zone near the offender's residence from which there is decreased activity or not is partly dependent on the interval size. With a small interval, the zone may be detected whereas with a slightly larger interval the subtle distinction in measured distances may be lost. On the other hand, having a smaller interval may lead to unreliable estimates since there may be few cases in the interval. Having a technique depend on the interval size makes it vulnerable to mis-specification.

Uses of Mathematical Distance Decay Functions

Does this mean that one should not use mathematical distance functions? I would argue that under most circumstances, a mathematical function will give less precision than an empirically-derived one (see below). However, there are two cases when a mathematical model would be appropriate. First, if there is either no data or insufficient

380

data to model the empirical travel distribution, the use of a mathematical model can serve as an approximation. If the user has a good sense of what the distribution looks like, then a mathematical model may be used to approximate the distribution. However, if a poorly defined function is selected, then the selected function may produce many errors.

A second case when mathematical models of distance decay would be appropriate is in theory development or application. Many models of travel behavior, for example, assume a simple distance decay type of function in order simplify the allocation of trips over a region. This is a common procedure in travel demand modeling where trips from each of many zones are assigned to every other zone using a gravity type of function (Stopher and Meyburg, 1975; Field and MacGregor, 1987). Even though the model produces errors because it assumes uniform travel behavior in all directions, the errors are corrected later in the modeling process by adjusting the coefficients for allocating trips to particular roads (traffic assignment). The model provides a simple device and the errors are corrected down the line. Still, I would argue that an empirically-derived distribution will produce fewer errors in allocation and, thus, require less adjustment later on. Errors can never help a model and its better to get it more correct initially to have to adjust it later on; the adjustment may be inadequate. Nevertheless, this is common practice in transportation planning.

The Journey to Crime Routine Using a Mathematical Formula

The Jtc routine which allows mathematical modeling is simple to use. Figure 9.10 illustrates how the user specifies a mathematical function. The routine requires the use of a grid which is defined on the reference file tab of the program (see chapter 3). Then, the user must specify the mathematical function and the parameters. In the figure, the truncated negative exponential is being defined. The user must input values for the peak likelihood, the peak distance, and the exponent (see equations 9.43 and 9.44 above). In the figure, since the serial offenses were a series of 18 robberies, the parameters for robbery have been entered into the program screen. The peak likelihood was 9.96% (entered as a whole number - i.e., 9.96); the distance at which this peak likelihood occurred was the second distance interval 0.25-0.50 miles (with a mid-point of 0.38 miles); and the estimated exponent was 0.177651. As mentioned above, the coefficient for the negative exponential part of the equation is estimated internally.

Table 9.3 gives the parameters for the 'best' models which fit the data for the 11 types of crime in Baltimore County. For several of these (e.g., bank robberies), two or more functions gave approximately equally good fits. Note that these parameters were estimated with the Baltimore County data. They will not fit any other jurisdiction. If a user wishes to apply this logic, then the parameters should be estimated anew from existing data. Nevertheless, once they have been calibrated, they can be used for predictions.

The routine can be output to *ArcView*, *MapInfo*, *Atlas*GIS*, *Surfer for Windows*, *Spatial Analyst*, and as an Ascii grid file which can be read by many other GIS packages. All but *Surfer for Windows* require that the reference grid be created by *CrimeStat*.

381

Figure 9.10: Jtc Mathematical Distance Decay Function

Table 9.3

**Journey to Crime Mathematical Models for Baltimore County
Parameter Estimates for Percentage Distribution**
(Sample Sizes in Parentheses)

ALL CRIMES

Negative Exponential:	Coefficient:	5.575107
	Exponent:	0.229466

HOMICIDE

Truncated Negative Exponential:	Peak likelihood	14.02%
	Peak distance	0.38 miles
	Exponent	0.064481

RAPE

Lognormal:	Mean	3.144959
	Standard Deviation	4.546872
	Coefficient	0.062791

ASSAULT

Truncated Negative Exponential:	Peak likelihood	27.40%
	Peak distance	0.38 miles
	Exponent	0.181738

ROBBERY

Truncated Negative Exponential:	Peak likelihood	9.96%
	Peak distance	0.38 miles
	Exponent	0.177651

COMMERCIAL ROBBERY

Truncated Negative Exponential:	Peak likelihood	4.9455%
	Peak distance	0.625 miles
	Exponent	0.151319

383

Table 9.3 (continued)

BANK ROBBERY

Truncated
Negative Exponential: Peak likelihood   9.96%
                               Peak distance   5.75 miles
                               Exponent   0.139536

BURGLARY

Truncated
Negative Exponential: Peak likelihood   20.55%
                               Peak distance   0.38 miles
                               Exponent   0.162907

AUTO THEFT

Truncated
Negative Exponential: Peak likelihood   4.81%
                               Peak distance   0.63 miles
                               Exponent   0.212508

LARCENY

Truncated
Negative Exponential: Peak likelihood   4.76%
                               Peak distance   0.38 miles
                               Exponent   0.193015

ARSON

Truncated
Negative Exponential: Peak likelihood   38.99%
                               Peak distance   0.38 miles
                               Exponent   0.093469

Distance Modeling Using an Empirically Determined Function

An alternative to mathematical modeling of distance decay is to empirically describe the journey to crime distribution and then use this empirical function to estimate the residence location. *CrimeStat* has a two-dimensional kernel density routine that can calibrate the distance function if provided data on trip origins and destinations. The logic of kernel density estimation was described in chapter 8, and won't be repeated here. Essentially, a symmetrical function (the 'kernel') is placed over each point in a distribution. The distribution is then referenced relative to a scale (an equally-spaced line for two-dimensional kernels and a grid for three-dimensional kernels) and the values for each kernel are summed at each reference location. See chapter 8 for details.

Calibrate Kernel Density Function

The *CrimeStat* calibration routine allows a user to describe the distance distribution for a sample of journey to crime trips. The requirements are that:

1.   The data set must have the coordinates of *both* an origin location and a destination location; and

2.   The records of all origin and destination locations have been populated with legitimate coordinate values (i.e., no unmatched records are allowed).

Data Set Definition

The steps are relatively easy. First, the user defines a calibration data set with both origin and destination locations. Figure 9.11 illustrates this process. As with the primary and secondary files, the routine reads *ArcView* 'shp', *dBase* 'dbf', Ascii 'txt', and *MapInfo* 'dat' files. For both the origin location (e.g., the home residence of the offender) and the destination location (i.e., the crime location), the names of the variables for the X and Y coordinates must be identified as well as the type of coordinate system and data unit (see chapter 3). In the example, the origin locations has variable names of HomeX and HomeY and the destination locations has variable names of IncidentX and IncidentY for the X and Y coordinates of the two locations respectively. However, any name is acceptable as long as the two locations are distinguished.

The program will calculate the distance between the origin location and the destination location for each record. If the units are spherical (i.e., lat/lon), then the calculations use spherical geometry; if the units are projected (either meters or feet), then the calculations are Euclidean (see chapter 3 for details).

Kernel Parameters

Next, the user must define the kernel parameters for calibration. There are five choices that have to be made (Figure 9.12):

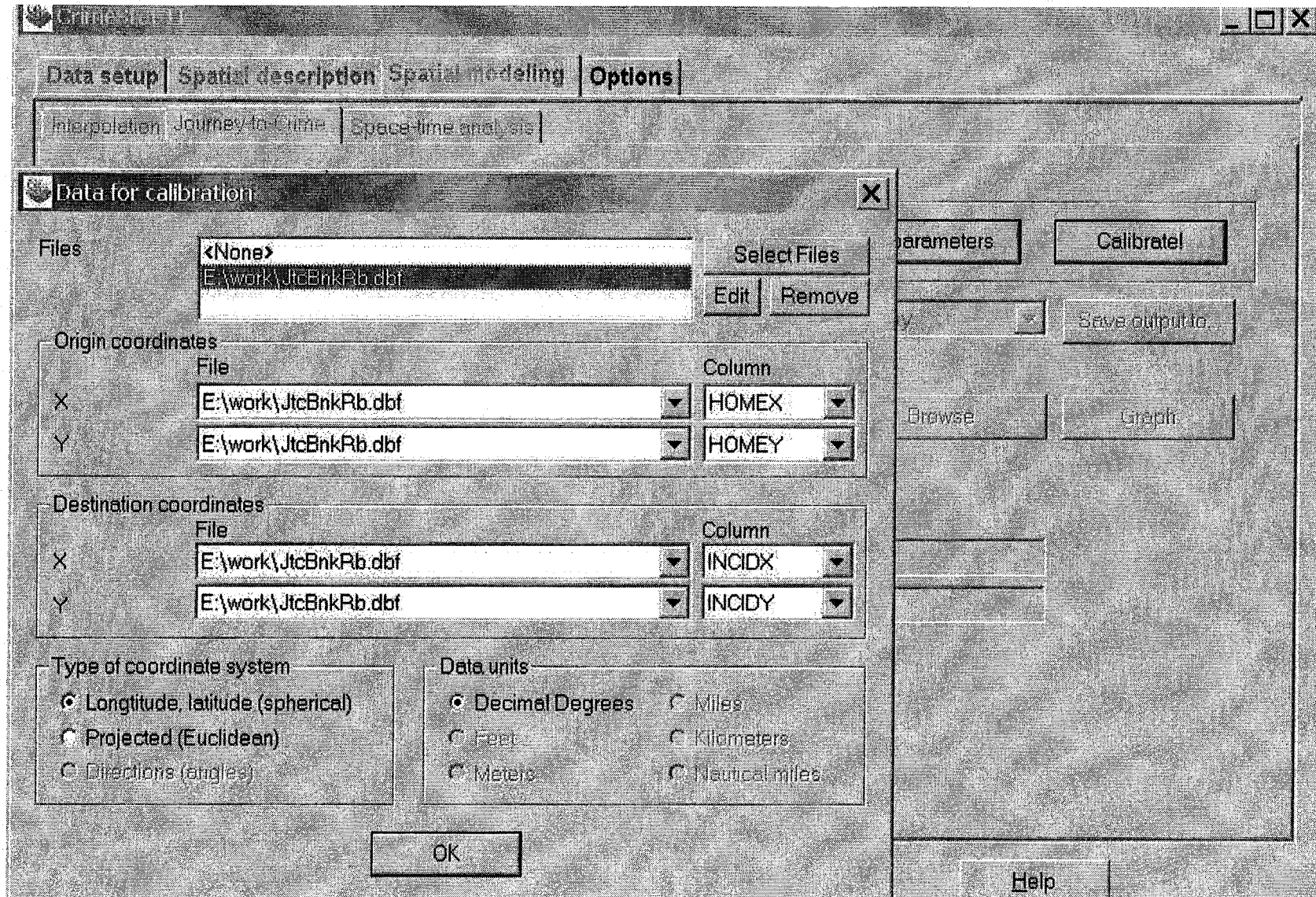385

Figure 9.11: Jtc Calibration Data Input

Figure 9.12: Jtc Calibration Kernel Parameters
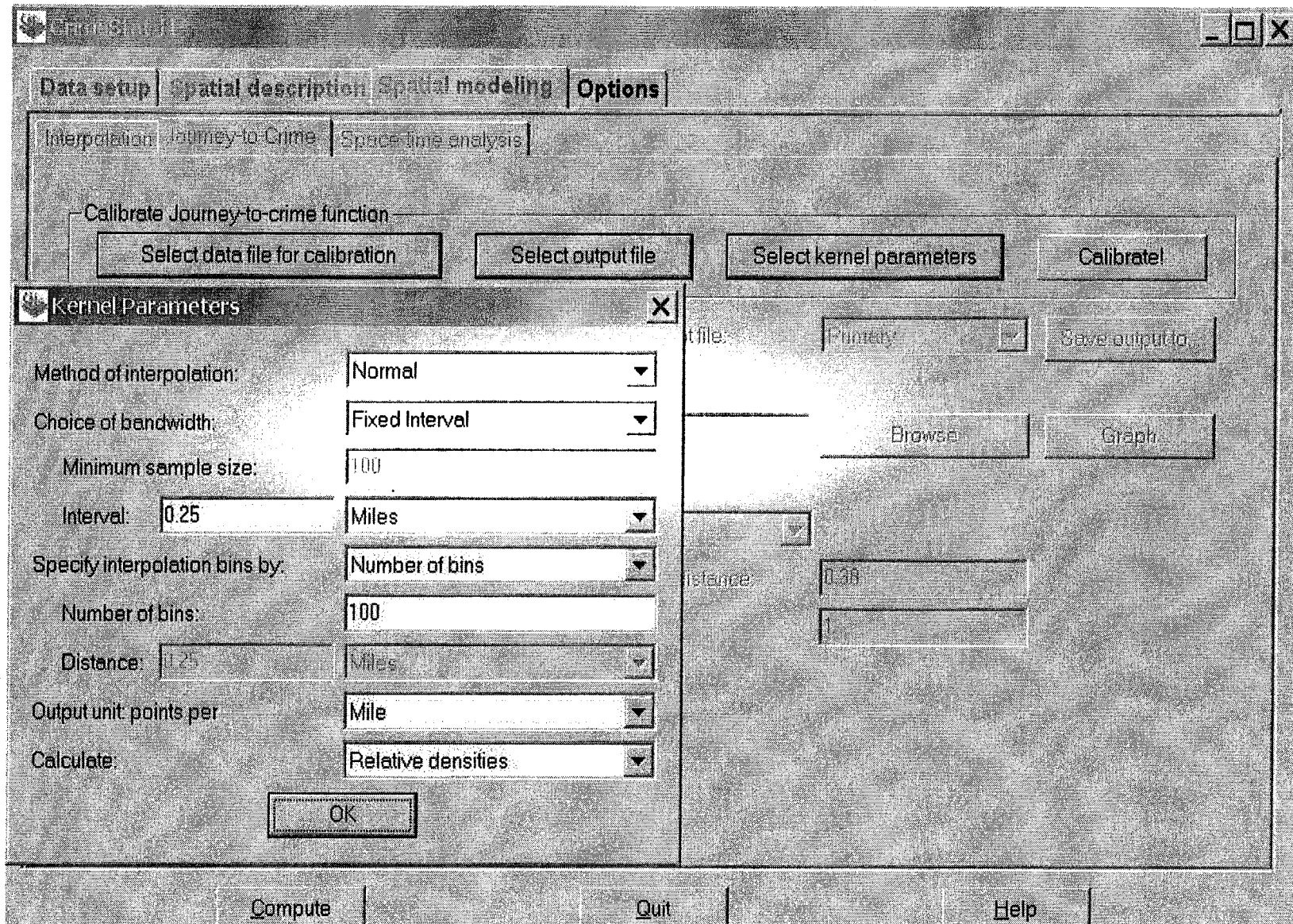
1. The method of interpolation. As with the two-dimensional kernel technique described in chapter 8, there are five possible kernel functions:

    A. Normal (the default);
    B. Quartic;
    C. Triangular (conical);
    D. A negative exponential (peaked); and
    E. A uniform (flat) distribution.

2. Choice of bandwidth. The bandwidth is the width of the kernel function. For a normal kernel, it is the standard deviation of the normal distribution whereas for the other four kernels (quartic, triangular, negative exponential, and uniform), it is the radius of the circle defined by the kernel. As with the two-dimension kernel technique, the bandwidth can be fixed in length or adaptive (variable in length). However, for the one-dimensional kernel, the fixed bandwidth is the default since an even estimate over an equal number of intervals (bins) is desirable. If the fixed bandwidth is selected, the interval size must be specified and the units (in miles, kilometers, feet, meters, and nautical miles). The default is 0.25 mile intervals. If the adaptive bandwidth is selected, the user must identify the minimum sample size that the bandwidth should incorporate; in this case, the bandwidth is widened until the specified sample size is counted.

3. The number of interpolation bins. The bins are the intervals along the distance scale (from 0 up to the maximum distance for a journey to crime trip) and are used to estimate the density function. There are two choices. First, the user can specify the number of intervals (the default choice with 100 intervals). In this case, the routine calculates the maximum distance (or longest trip) between the origin location and the destination location and divides it by the specified number of intervals (e.g., 100 equal-sized intervals). The interval size is dependent on the longest trip distance measured. Second, the user can specify the distance between bins (or the interval size). The default choice is 0.25 miles, but another value can be entered. In this case, the routine counts out intervals of the specified size until it reaches the maximum trip distance.

4. The output units. The user specifies the units for the density estimate (in units per mile, kilometer, feet, meters, and nautical miles).

5. The output calculations. The user specifies whether the output results are in probabilities (the default) or in densities. For probabilities, the sum of all kernel estimates will equal 1.0. For densities, the sum of all kernel estimates will equal the sample size.

Saved Calibration File

Third, the user must define an output file to save the empirically determined function. The function is then used in estimating the likely home residence of a particular function. The choices are to save the file as a 'dbf' or Ascii text file. The saved file then can be used in the Jtc routine. Figure 9.13 illustrates the output file format.

Calibrate

Fourth, the calibrate button runs the routine. A calibration window appears and indicates the progress of the calculations. When it is finished, the user can view a graph illustrating the estimated distance decay function (Figure 9.14). The purpose is to provide quick diagnostics to the user on the function and selection of the kernel parameters. While the graph can be printed, it is not a high quality print. If a high quality graph is needed, the output calibration file should be imported into a graphics program.

Examples from Baltimore County

Let's illustrate this method by showing the results for the same data sets that were calculated above in the mathematical section (figures 9.4-9.8). In all cases, the normal kernel function was used. The bandwidth was 0.25 miles except for the bank robbery data set, which had only 176 cases, and the homicide data set, which only had 137 cases; because of the small sample sizes, a bandwidth of 0.50 miles was used for these two data sets. The interval width selected was a distance of 0.25 miles between bins (0.5 miles for bank robberies and homicides) and probabilities were output.

Figure 9.15 shows the kernel estimate for all crimes (41,426 trips). A frequency distribution was calculated for the same number of intervals and is overlaid on the graph. It was selected to be comparable to the mathematical function (see figure 9.4). Note how closely the kernel estimate fits the data compared to the negative exponential mathematical function. The fit is good for every value but the peak value; that is because the kernel averages several intervals together to produce an estimate.

Figure 9.16 shows the kernel estimate for larceny thefts. Again, the kernel method produces a much closer fit as a comparison with figure 9.5 will show. Figure 9.17 shows the kernel estimate for vehicle thefts. Figure 9.18 shows the kernel estimate for bank robberies and figure 9.19 shows the kernel estimate for homicides. An inspection of these graphs shows how well the kernel function fits the data, compared to the mathematical function, even when the data are irregularly spaced (in vehicle thefts, bank robberies, and homicides). Figure 9.20 compares the distance decay functions for homicides committed against strangers compared to homicides committed against known victims.

In short, the Jtc calibration routine allows a much closer fit to the data than any of the simpler mathematical functions. While it's possible to produce a complex mathematical function that will fit the data more closely (e.g., higher order polynomials), the kernel method is much simpler to use and gives a good approximation to the data.

389

Figure 9.13: Jtc Calibration Output File
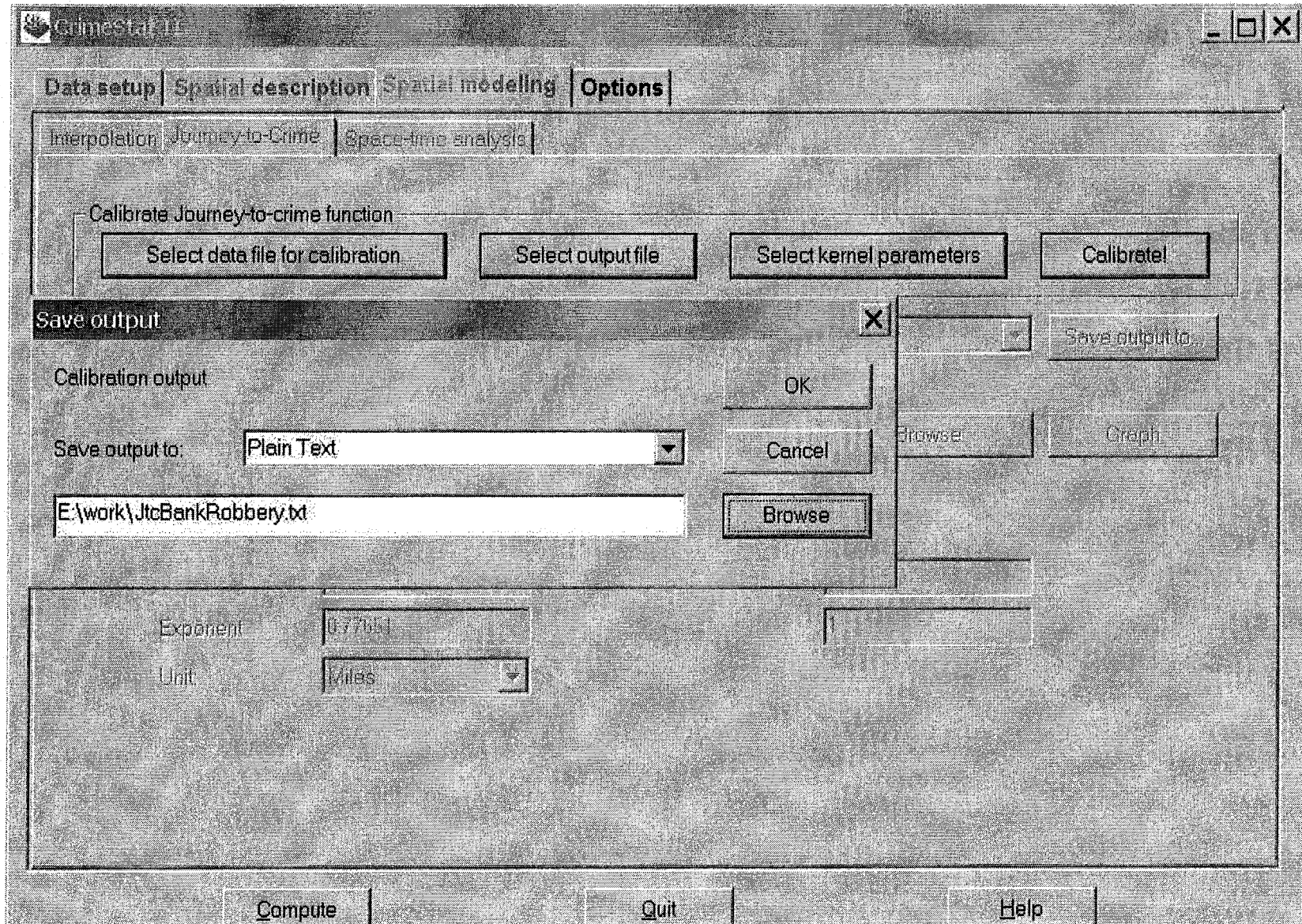
Figure 9.14: Jtc Calibration Graphic Output
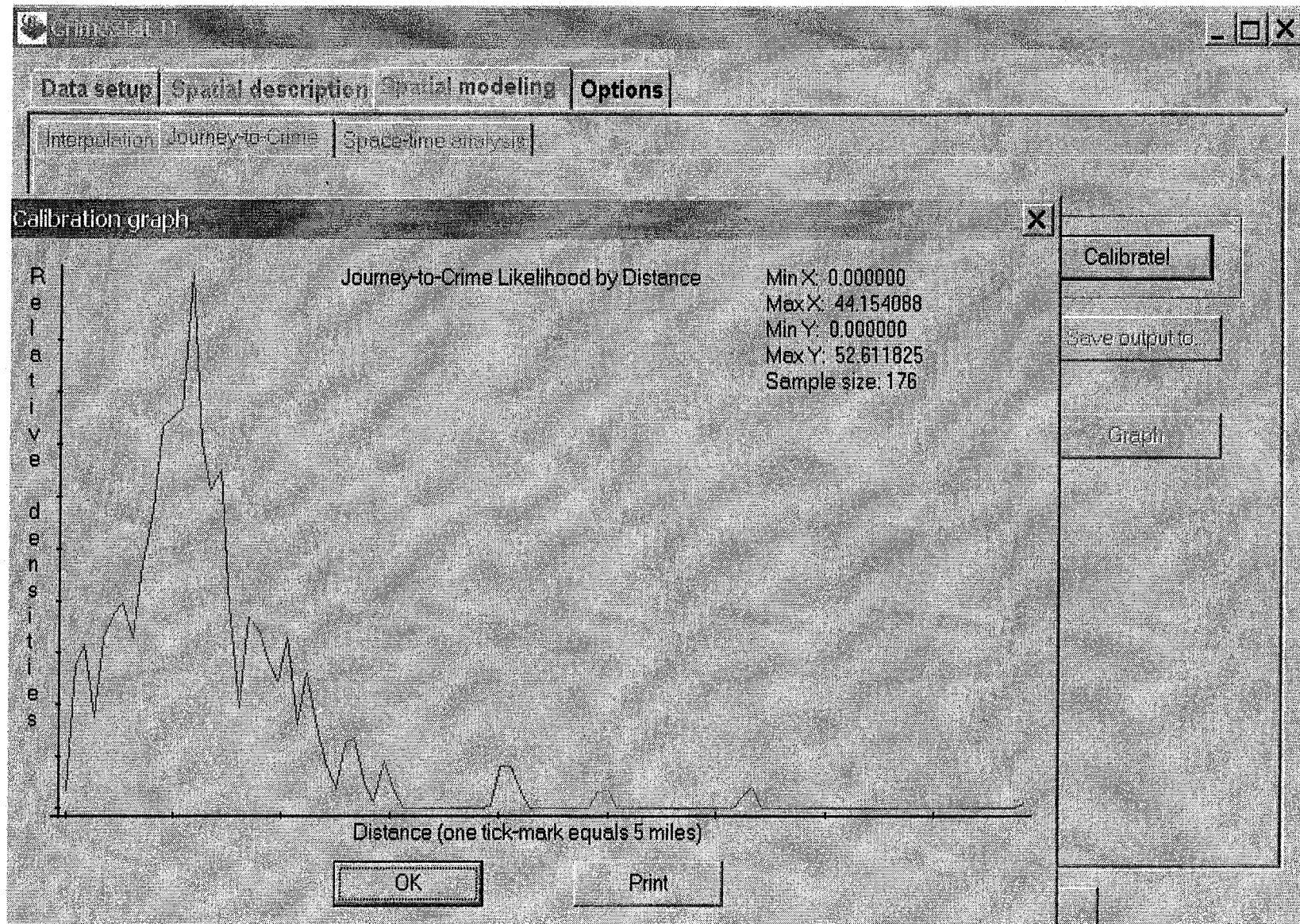
Figure 9.15:

Journey to Crime Distances: All Crimes
Frequencies and Kernel Density Estimate

Legend:
- + Number of trips for distance bin
- Kernel density estimate

Y-axis: Percent of all crimes (10%, 8%, 6%, 4%, 2%, 0%)

X-axis: Distance from offender's home (miles) (0, 5, 10, 15, 20, 25)
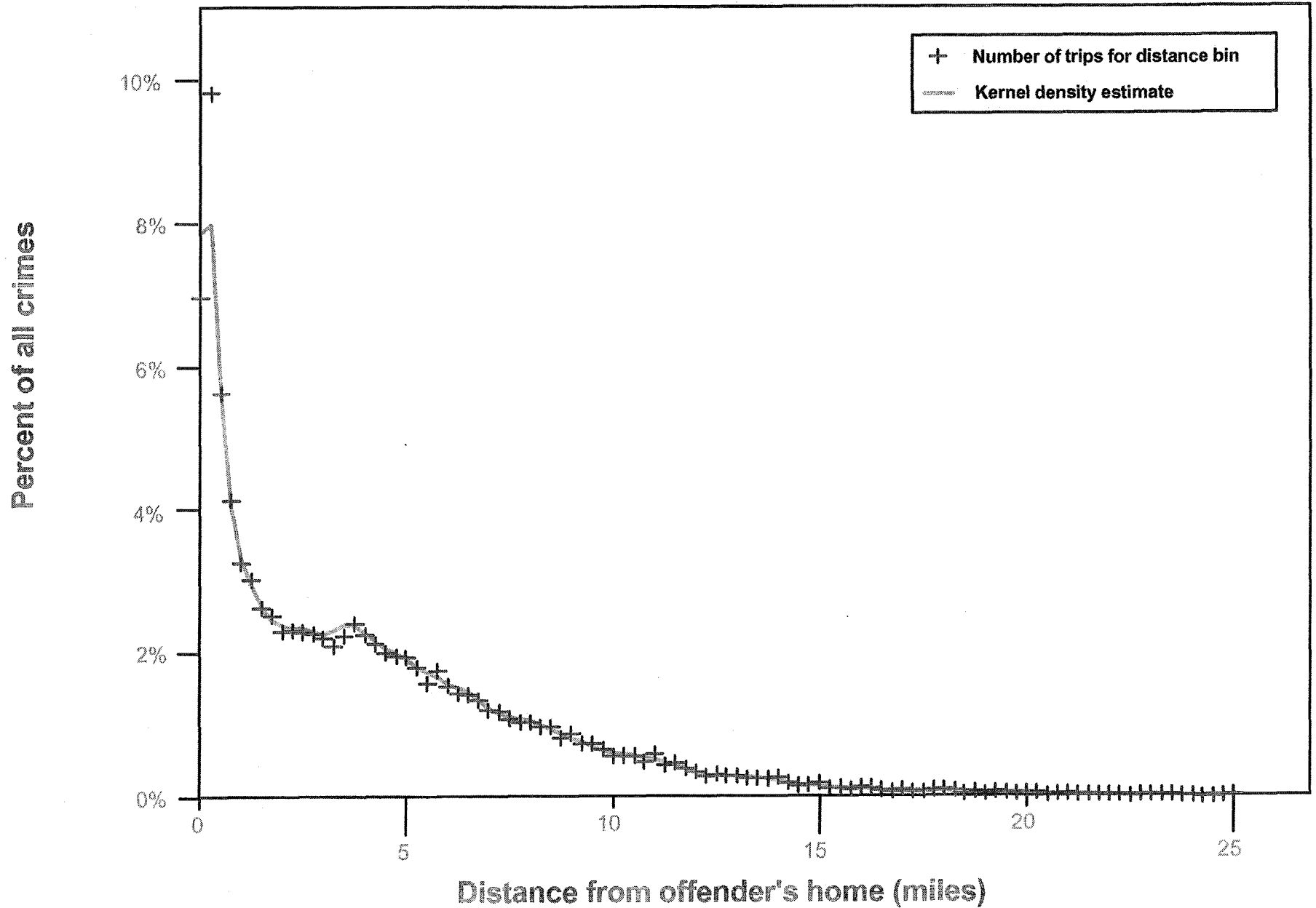
Figure 9.16:
Journey to Crime Distances: Larceny
Frequencies and Kernel Density Estimate

Legend:
- Number of trips for distance bin
- Kernel density estimate

Y-axis: Percent of all larceny thefts
X-axis: Distance from offender's home (miles)
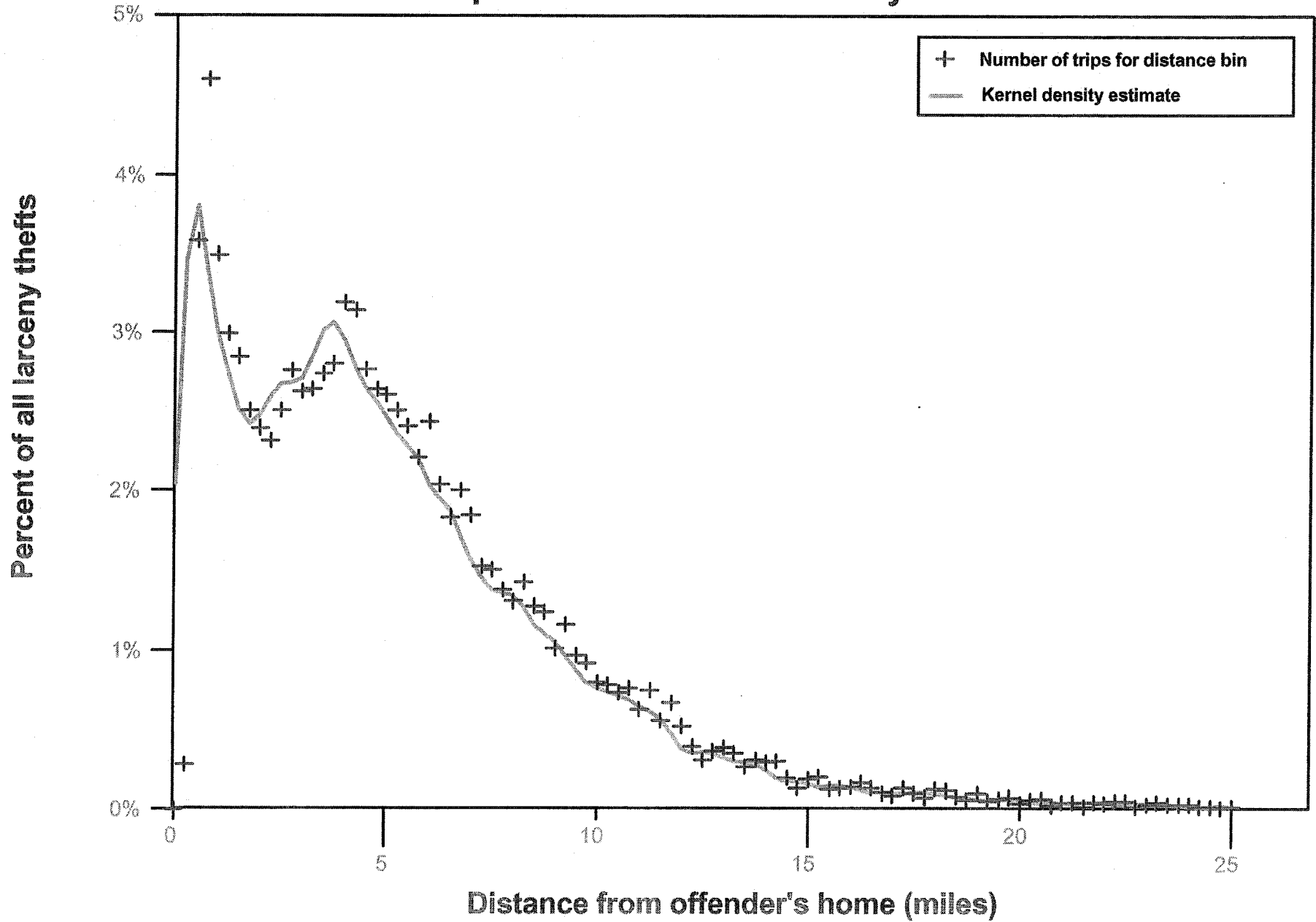
Figure 9.17:

Journey to Crime Distances: Vehicle Theft
Frequencies and Kernel Density Estimate

Legend:
- **+** Number of trips for distance bin
- — Kernel density estimate

Y-axis: Percent of all vehicle thefts (0%, 1%, 2%, 3%, 4%, 5%)

X-axis: Distance from offender's home (miles) (0, 5, 10, 15, 20, 25)
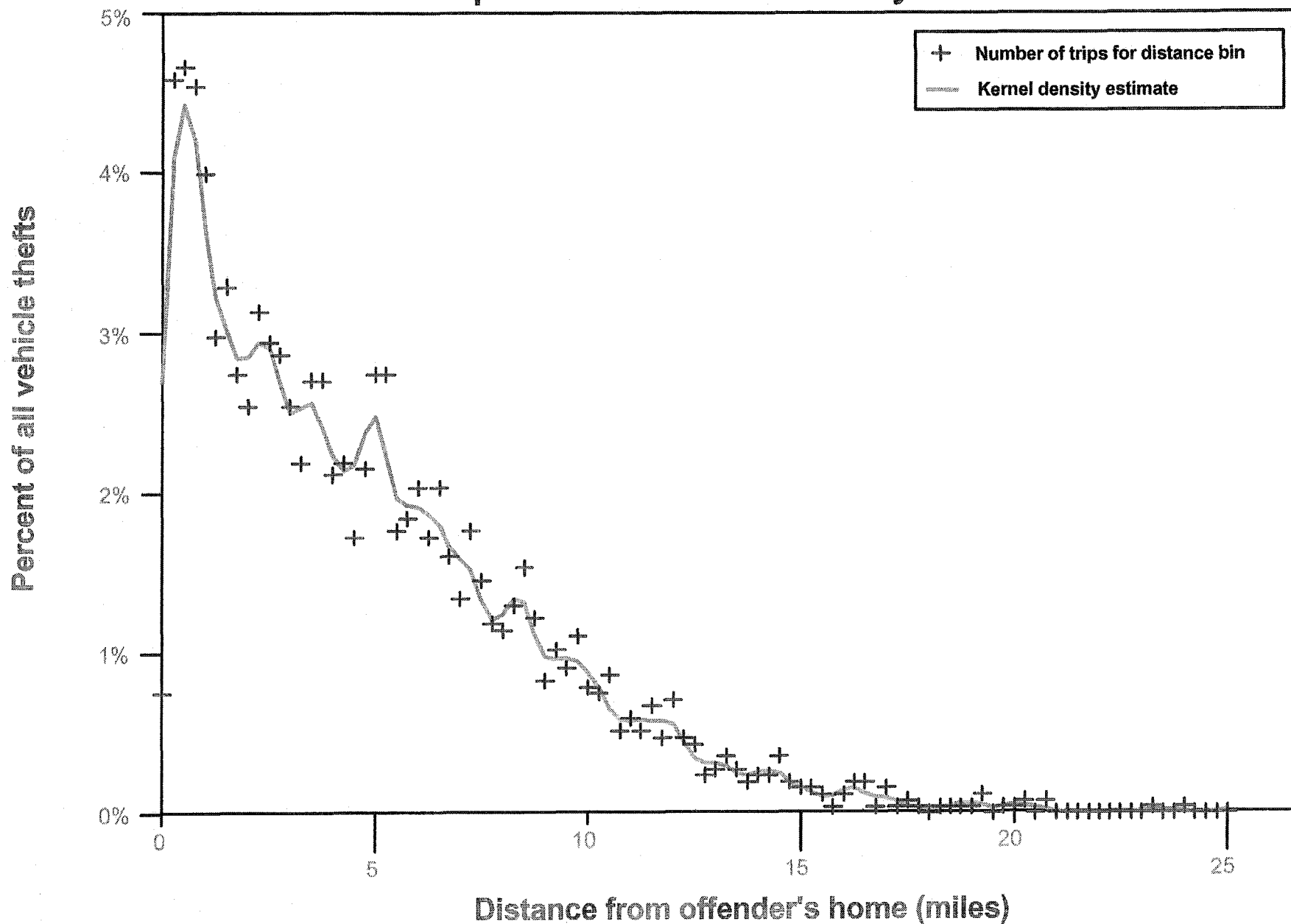
Figure 9.18:

Journey to Crime Distances: Bank Robbery

Frequencies and Kernel Density Estimate
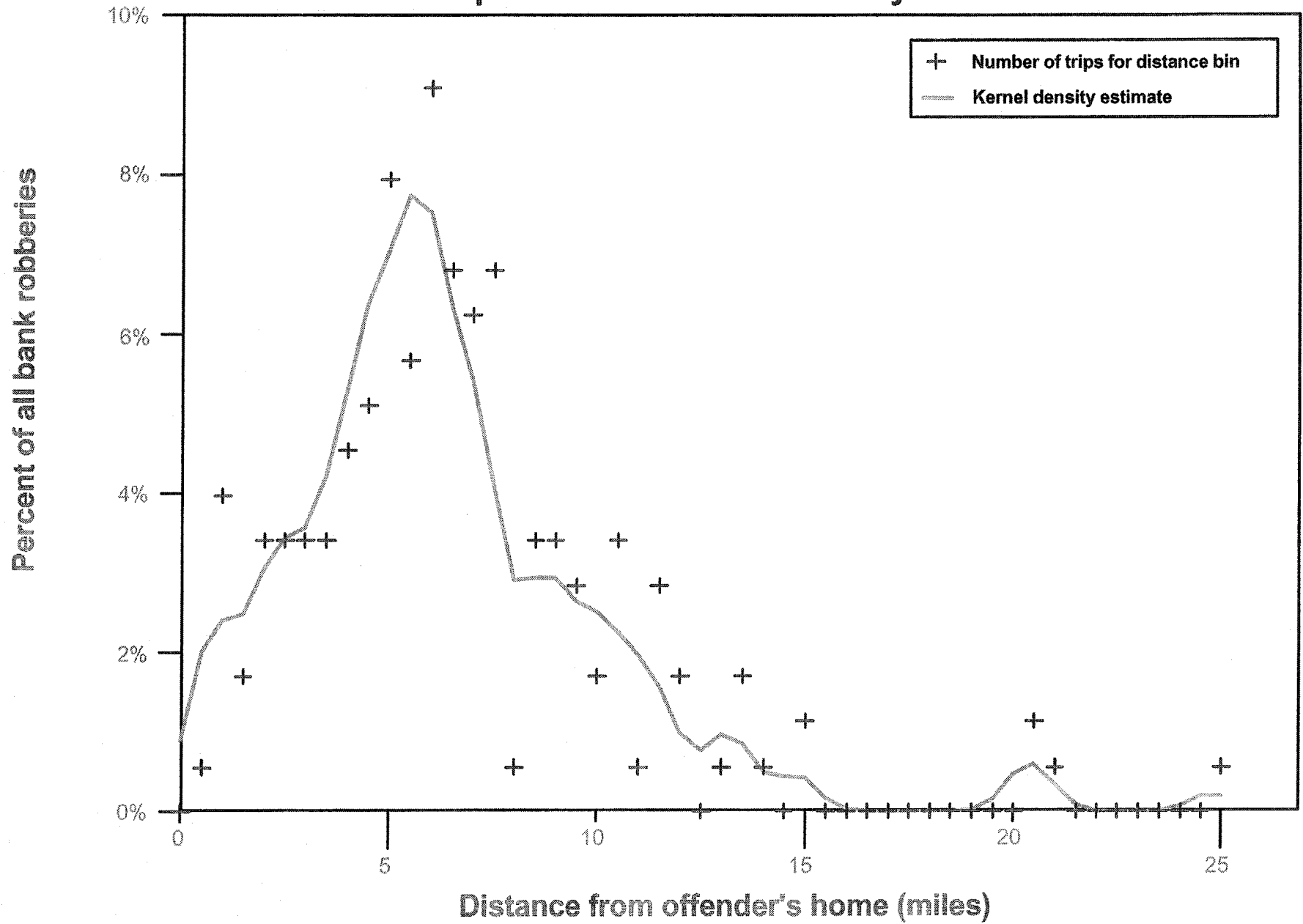
Figure 9.19:
Journey to Crime Distances: Homicide
Frequencies and Kernel Density Estimate
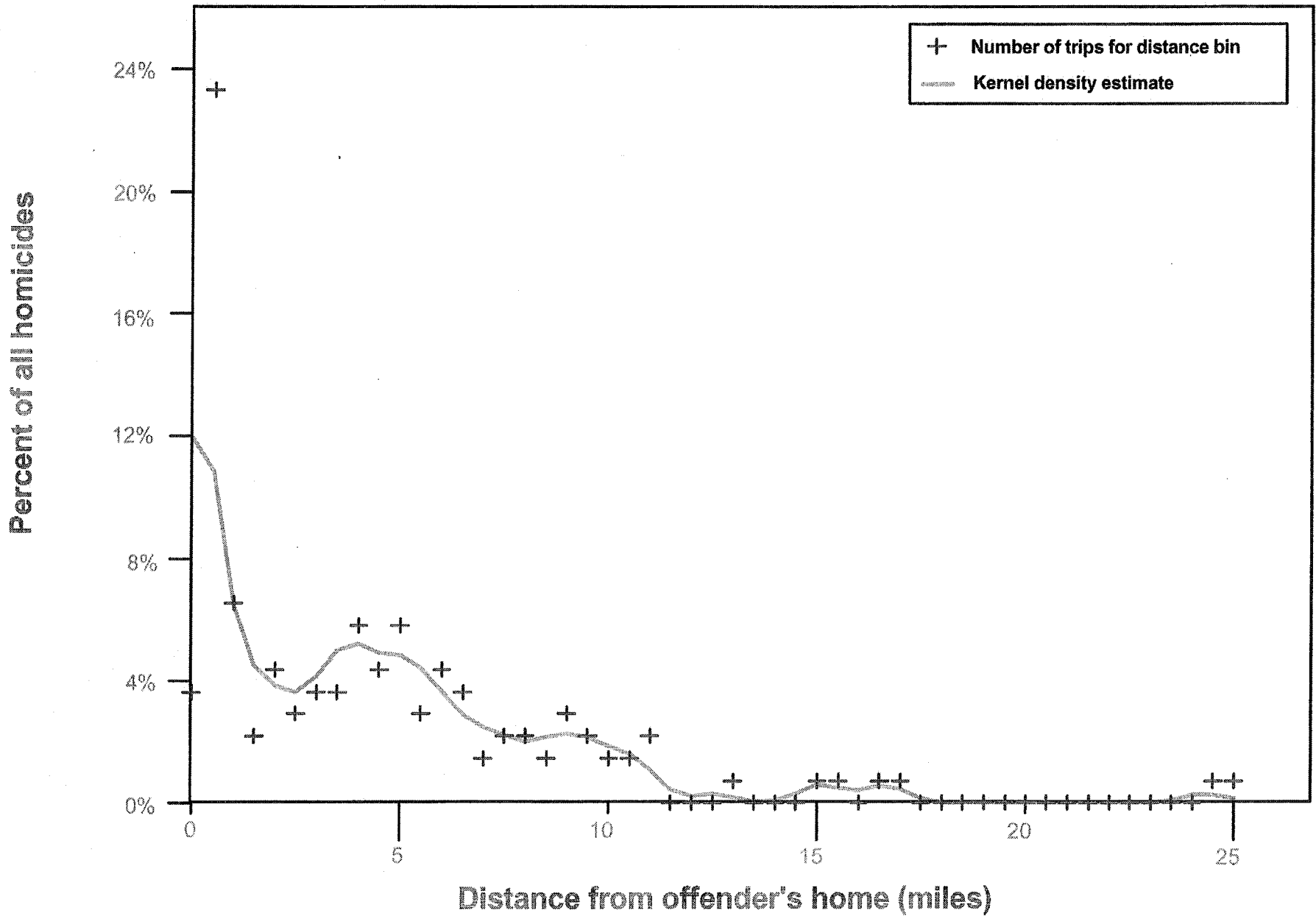
Figure 9.20:
Journey to Crime Distances: Homicide by Victim Relationship
Frequencies and Kernel Density Estimate
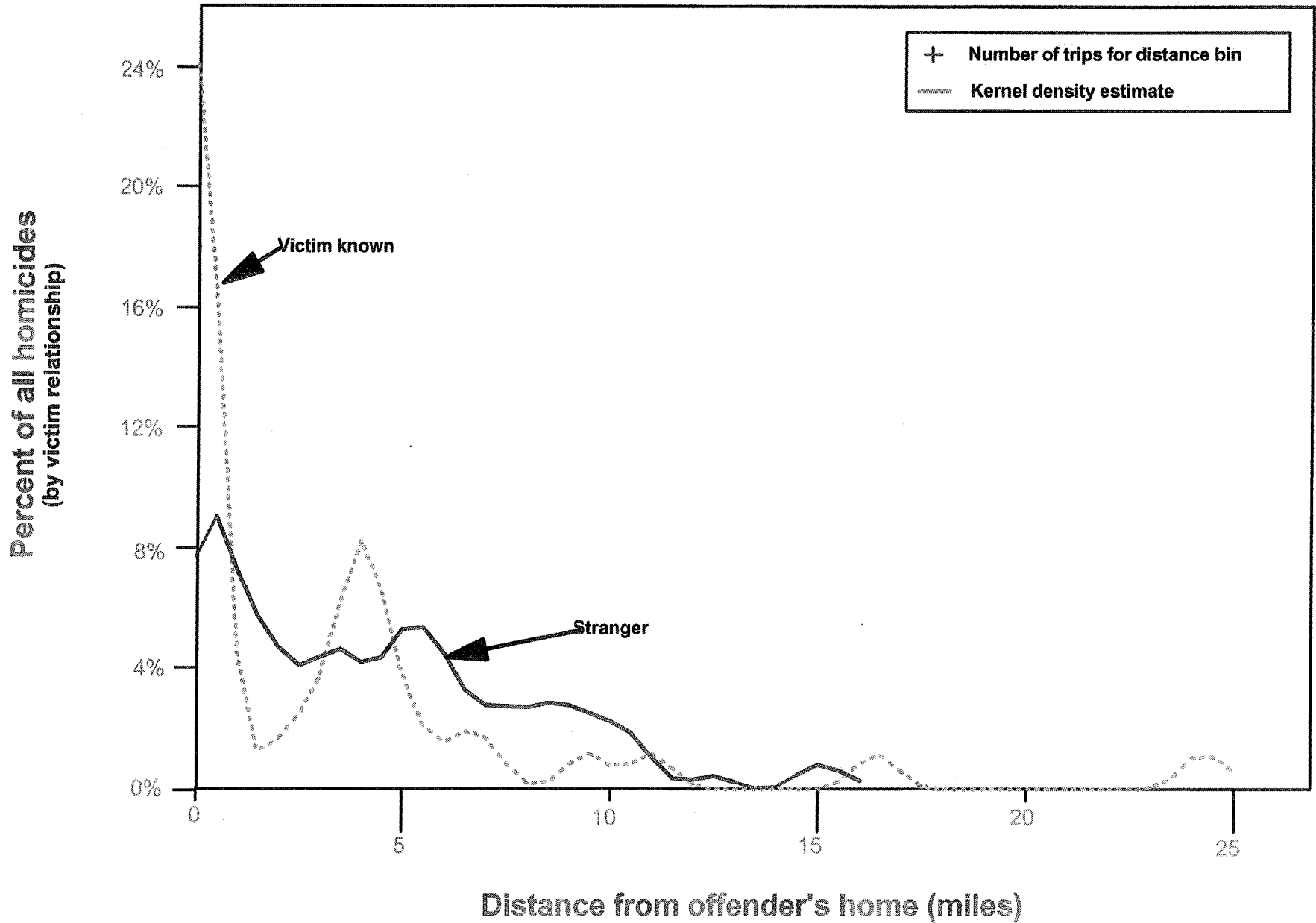
Using Journey-To-Crime Routine for Journey-After-Crime Analysis

Yongmei Lu
Department of Geography
Southwest Texas State University
San Marcos, TX

The study of vehicle theft recovery locations can fill a gap in the knowledge about criminal travel patterns. Although the journey-to-crime routine of *CrimeStat* was designed to analyze the distance between offense location and offender's residential location, it can be used to describe the distance between vehicle theft location and the corresponding recovery location.

There were more than 3000 vehicle thefts in the City of Buffalo in 1998. Matching the offenses with vehicle recoveries in the same year, 1600 location pairs were identified for a journey-after-vehicle-theft analysis. To evaluate the randomness of the distances, 1000 groups of simulations were conducted. Every group contains 1600 simulated trips of journey-after-vehicle-theft. The results indicate that 1) short distances dominate journey-after-vehicle-theft, and 2) the observed trips are significantly shorter than the random trips given the distribution of possible vehicle theft and recovery locations.
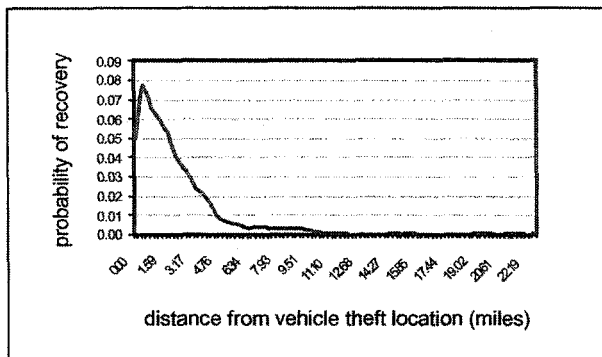
Probability of recovering a stolen vehicle by distance from vehicle theft location
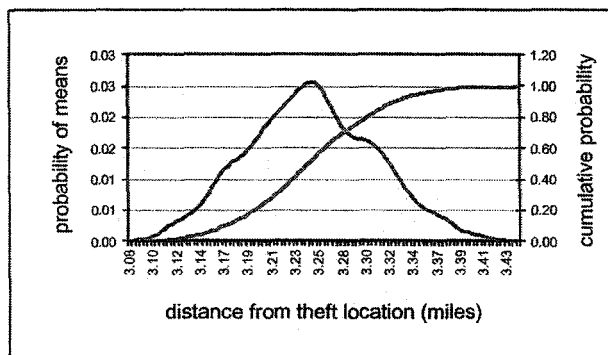
Distribution of mean distances of simulated vehicle theft-recovery location pairs.

Using Journey to Crime for Different Age Groups of Offenders

Renato Assunção, Cláudio Beato, Bráulio Silva
CRISP, Universidade Federal de Minas Gerais , Brazil

CrimeStat offers a method for analysing the distance between the crime scene and the residence of the offender using the journey to crime routine within the spatial modeling module. We analysed homicide incidents in Belo Horizonte, a Brazilian city of 2 million inhabitants, for the period January 1996 – December 2000. We used 496 homicide cases for which the police identified an offender who was living in Belo Horizonte, and for which both the crime location and offender residence could be identified. The cases were divided into three groups according to the offender's age: 1) 14 to 24 (N=201); 2) 25 to 34 (N=176); and 3) 35 or older (N=119). The journey to crime calibration routine was used to produce a probability curve *P(d)* that gives the approximate chance of an offender travelling approximately distance *d* to commit the crime.

We used the normal kernel, a fixed bandwidth of 1000 meters, 100 output bins, and the probability (or proportion of all points) option, rather than densities. This is to allow comparisons between the three age groups since they have different number of homicides. We tested for each age group separately and directed the output to a text file to analyse the three groups simultaneously.

The green, blue, and purple curves are associated with the 14-24, 25-34, 35+ year olds respectively. There are more similarities than differences between the groups. Most homicides are committed near to the residence of the offenders with between 60% t o 70% closer than one mile from their home. However, the curve does not vanish totally even for large distances because there are around 15% of offenders, of any age group, travelling longer than 3 miles to commit the crime. The oldest offenders travel longer distances, on average, followed by the youngest group, with the 25-34 year olds travelling the shortest distances.

Journey to homicide probabilities in Belo Horizonte, Brazil
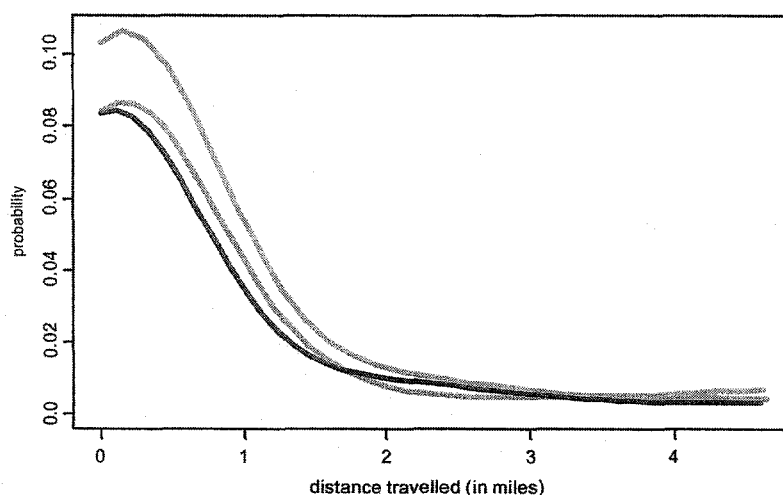
The Journey to Crime Routine Using the Calibrated File

After the distance decay function has been calibrated and saved as a file, the file can be used to calculate the likelihood surface for a serial offender. The user specifies the name of the already-calibrated distance function (as a 'dbf' or an Ascii text file) and the output format. As with the mathematical routine, the output can be to *ArcView*, *MapInfo*, *Atlas*GIS*, *Surfer for Windows*, *Spatial Analyst*, and as an Ascii grid file which can be read by many other GIS packages. All but *Surfer for Windows* require that the reference grid be created by *CrimeStat*.

The result is produced in three steps:

1. The routine calculates the distance between each reference cell of the grid and each incident location;

2. For each distance measured, the routine looks up the calculated value from the saved calibration file; and

3. For each reference grid cell, it sums the values of all the incidents to produce a single likelihood estimate.

Application of the Routine

To illustrate the techniques, the results of the two methods on a single case are compared. The case has been selected because the routines accurately estimate the offender's residence. This was done to demonstrate how the techniques work. In the next section, I'll ask the question about how accurate these methods are in general.

The case involved a man who had committed 24 offenses. These included 13 thefts, 5 burglaries, 5 assaults, and one rape. The spatial distribution was varied; many of the offenses were clustered but some were scattered. Since there were multiple types of crimes committed by this individual, a decision had to be made over which model to use to estimate the individual's residence. In this case, the theft (larceny) model was selected since that was the dominant type of crime for this individual.

For the mathematical function, the truncated negative exponential was chosen from table 9.3 with the parameters being:

Peak likelihood	4.76%
Peak distance	0.38 miles
Exponent	0.193015

For the kernel density model, the calibrated function for larceny was selected (figure 9.16).

Figure 9.21 shows the results of the estimation for the two methods. The output is from *Surfer for Windows* (Golden Software, 1994). The left pane shows the results of the

400

Figure 9.21:

Predicted and Actual Location of Serial Thief
Man Charged with 24 Offenses in Baltimore County
Predicted with Mathematical and Kernel Density Models for Larceny

Residence Location = Square

Crime Locations = Circles
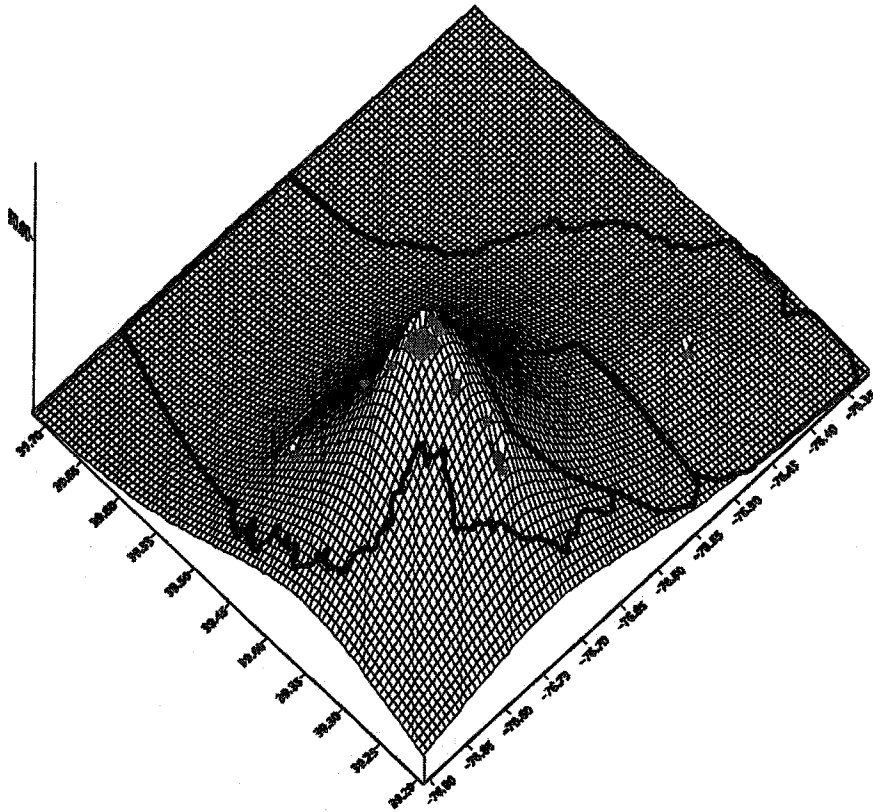
Mathematical Model:
Truncated Negative Exponential
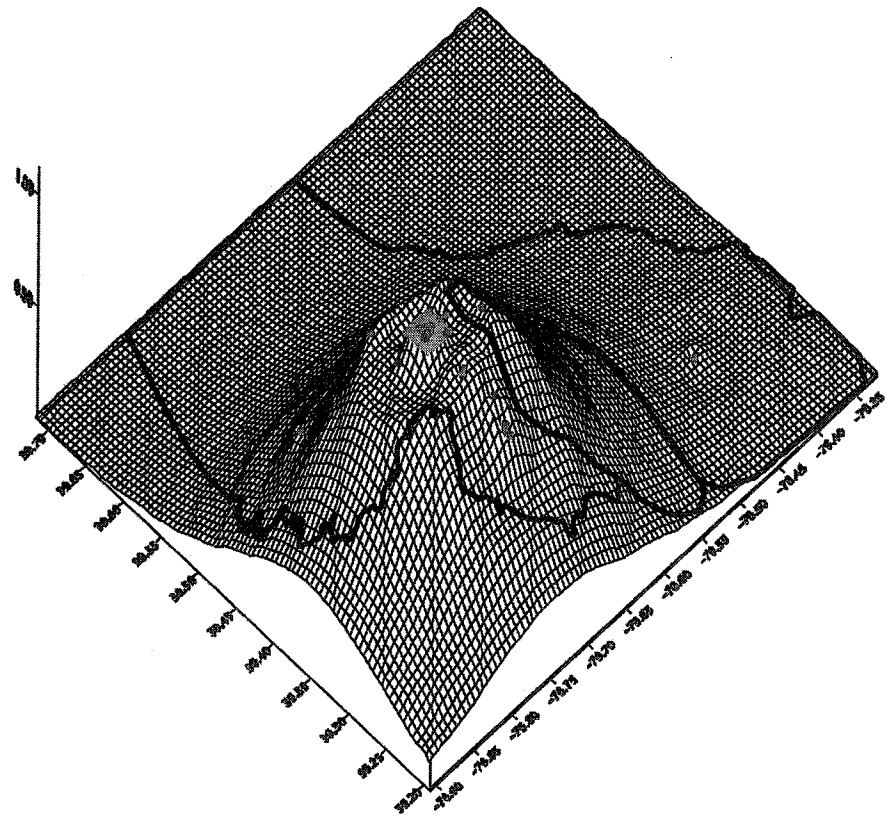
Kernel Density Model

mathematical function while the right pane shows the results for the kernel density function. The incident locations are shown as circles while the actual residence location of the offender is shown as a square. Since this is a surface model, the highest location has the highest predicted likelihood.

In both cases, the models predicted quite accurately. The discrepancy (error) between the predicted peak location and the actual residence location was 0.66 miles for the mathematical function and 0.36 miles for the kernel density function. For the mathematical model, the actual residence location (square) is seen as slightly off from the peak of the surface whereas for the kernel density model the discrepancy from the peak cannot be seen.

Nevertheless, the differences in the two surfaces show distinctions. The mathematical model has a smooth decline from the peak likelihood location, almost like a cone. The kernel density model, on the other hand, shows a more irregular distribution with a peak location followed by a surrounding 'trough' followed a peak 'rim'. This is due to the irregular distance decay function calibrated for larceny (see figure 9.16). But, in both cases, they more or less identify the actual residence location of the offender.

Choice of Calibration Sample

The calibration sample is critical for either method. Each method assumes that the distribution of the serial offender will be similar to a sample of 'like' offenders. Obviously, distinctions can be made to make the calibration sample more or less similar to the particular case. For example, if a distance decay function of all crimes is selected, then a model (of either the mathematical or kernel density form) will have less differentiation than for a distance decay function from a specific type of crime. Similarly, breaking down the type of crime by, say, mode of operation or time of day will produce better differentiation than by grouping all offenders of the same type together. This process can be taken on indefinitely until there is too little data to make a reliable estimate. An analyst should try to find as close a calibration sample to the actual as is possible, given the limitations of the data.

For example, in our calibration data set, there were 4,694 burglary incidents where both the offender's home residence and the incident location were known. The approximate time of the offense for 2,620 of the burglaries was known and, of these, 1,531 occurred at night between 6 pm and 6 am. Thus, if a particular serial burglar for whom the police are interested in catching tends to commit most of his burglaries at night, then choosing a calibration sample of nighttime burglars will generally produce a better estimate than by grouping all burglars together. Similarly, of the 1,531 nighttime burglaries, 409 were committed by individuals who had a prior relationship with the victim. Again, if the analysts suspect that the burglar is robbing homes of people he knows or is acquainted with, then selecting the subset of nighttime burglaries committed against a known victim would produce even better differentiation in the model than taking all nighttime burglars. However, eventually, with further sub-groupings there will be insufficient data.

402

This point has been raised in a recent debate. Van Koppen and De Keijser (1997) argued that a distance decay function that combined multiple incidents committed by the same individuals could distort the estimated relationship compared to selecting incidents committed by different individuals.[6] Rengert, Piquero and Jones (1999) argued that such a distribution is nevertheless meaningful. In our language, these are two different sub-groups - persons committing multiple offenses compared to persons committing only one offense. Combining these two sub-groups into a single calibration data set will only mean that the result will have less differentiation in prediction than if the sub-groups were separated out.

Actually, there is not much difference, at least in Baltimore County. From the 41,426 cases, 18,174 were committed by persons who were only listed once in the database while 23,251 offenses were committed by persons who were listed two or more times (7,802 individuals). Categorizing the 18,174 crimes as committed by 'single incident offenders' and the 23,251 crimes as committed by 'multiple incident offenders', the density distance decays functions were calculated using the kernel density method (Figure 9.22).

The distributions are remarkably similar. There are some subtle differences. The average journey to crime trip distance made by a single incident offender is longer than for multiple incident offenders (4.6 miles compared to 4.0 miles, on average); the difference is highly significant ($p \leq .0001$), partly because of the very large sample sizes. However, a visual inspection of the distance decay functions shows they are similar. The single incident offenders tend to have slightly more trips near their home, slightly fewer for distances between about a mile up to three miles, and slightly more longer trips. But, the differences are not very large.

There are several reasons for the similarity. First, some of the 'single incident offenders' are actually multiple incident offenders who have not been charged with other incidents. Second, some of the single incident offenders are in the process of becoming multiple incident offenders so their behavior is probably similar. Third, there may not be a major difference in travel patterns by the number of offenses an individual commits, certainly compared to the major differences by type of crime (see graphs above). In other words, the distinction between a single offender crime trip and a multiple offender crime trip is just another sub-group comparison and, apparently, not that important. Nevertheless, it is important to choose an appropriate sample from which to estimate a likely home base location for a serial offender. The method depends on a similar sample of offenders for comparison.

Sample Data Sets for Journey to Crime Routines

Three sample data sets from Baltimore County have been provided for the journey to crime routine. The data sets are simulated and do not represent real data. The first file - JtcTest1.dbf, are 2000 simulated robberies while the second file - JtcTest2.dbf, are 2500 simulated burglaries. Both files have coordinates for an origin location (HomeX, HomeY) and a destination location (IncidentX, IncidentY). Users can use the calibration routine to calculate the travel distances between the origins and the destinations. A third data set -

403

Figure 9.22:

Journey to Crime Distances
Kernel Density Estimate of Single and Multiple Incident Offenders

Percent of all crimes (by offender type)

Multiple incident offender

Single incident offender

Distance from offender's home (miles)
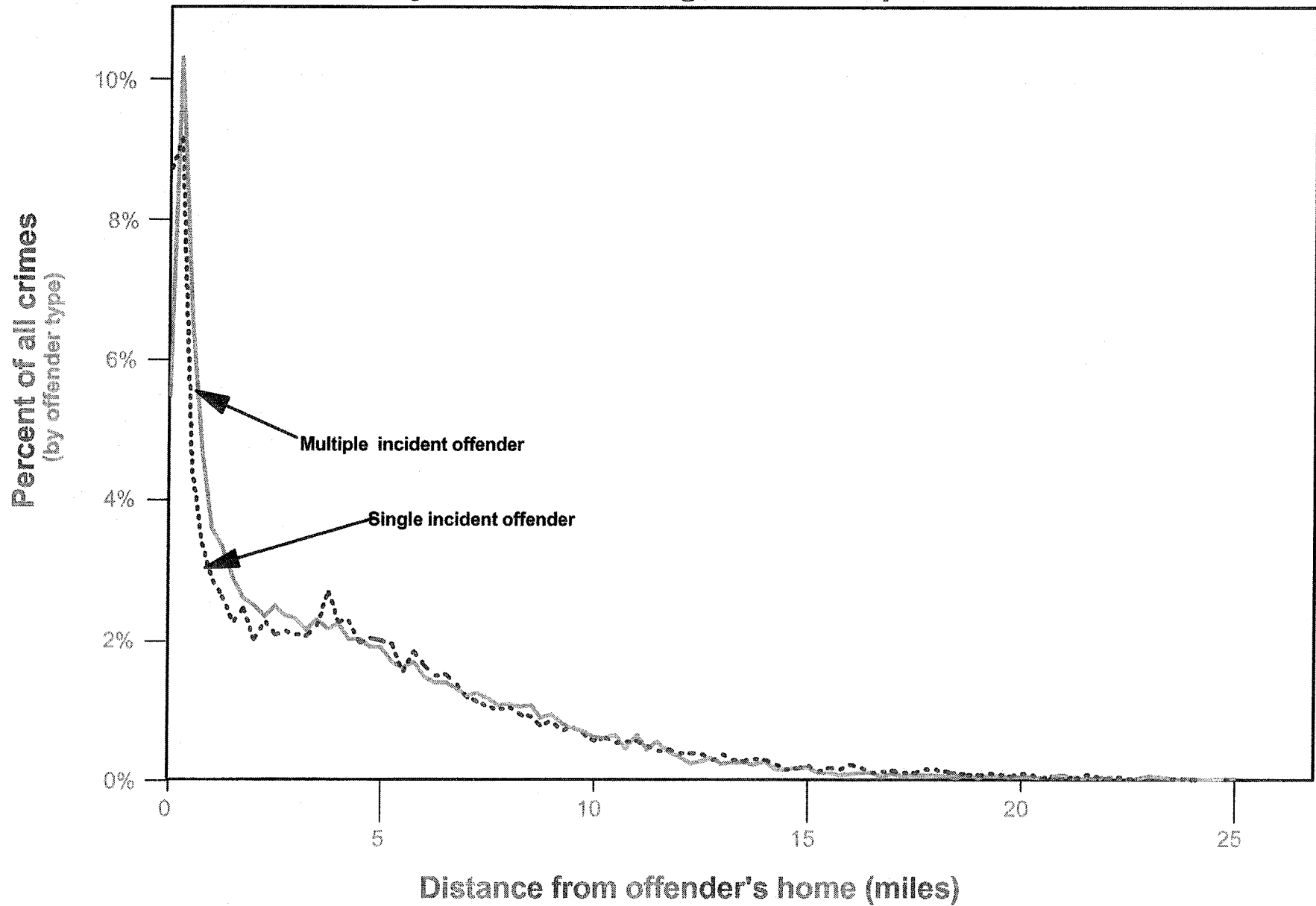
Catching the Bad Guy

Bryan Hill
Glendale Police Department
Glendale, AZ

The City of Glendale, Arizona recently had a string of auto thefts committed by the same individual. The map shows known auto theft suspects and their home address. The red area in the map shows the most probable home address. Prior to the analysis, the Phoenix Police Department's Crime Analysis Unit was able to calibrate the *CrimeStat* Jtc routine with known offender robbery suspect data.

Monthly citation data was used to search for anyone that lived within the area identified by the routine who also drove a red Saturn. A suspect with a felony warrant was identified and proved to be also the suspect in a series of armed robberies and a homicide that occurred in the Phoenix and Glendale jurisdictions. When he was arrested for the felony warrant at his home, evidence of the robberies and homicide were found.
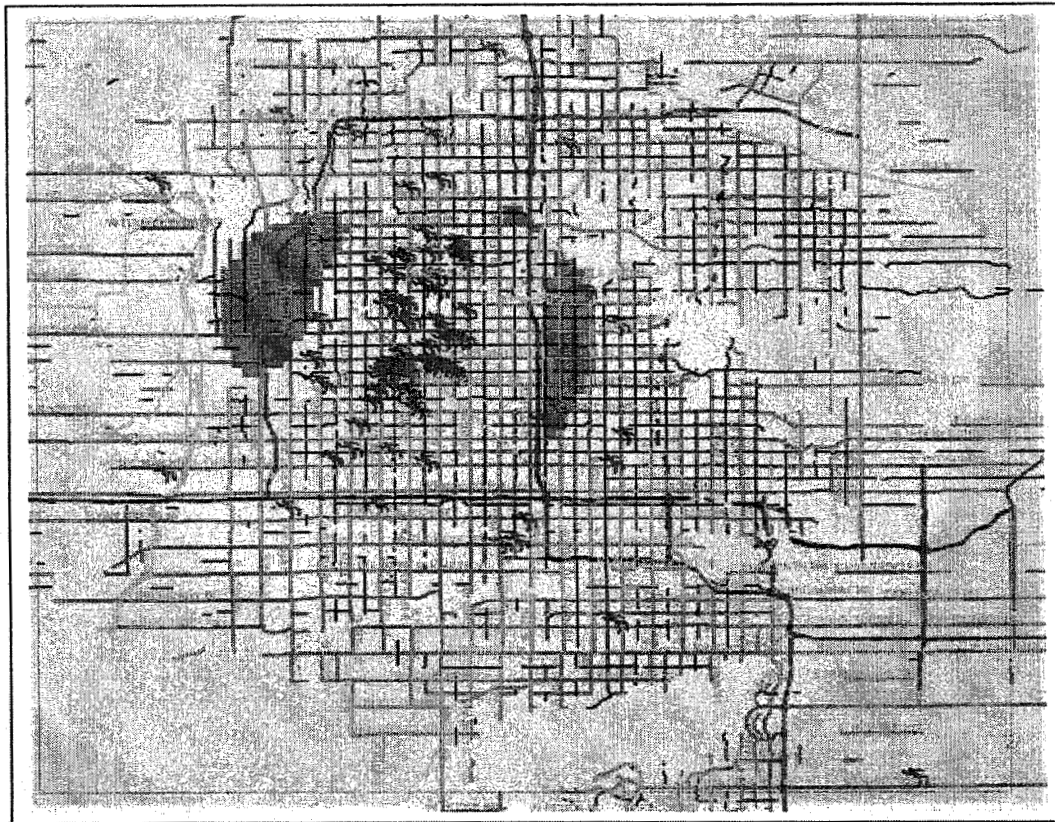
Serial1.dbf, are simulated incident locations for a serial offender. Users can use the Jtc estimation routine to identify the likely residence location for this individual. In running this routine, a reference grid needs to be overlaid (see chapter 3). For Baltimore County, appropriate coordinates for the lower-left corner are -76.91^0 longitude and 39.19^0 latitude and for the upper-right corner are -76.32^0 longitude and 39.72^0 latitude.

How Accurate are the Methods?

A critical question is how accurate are these methods? The journey to crime model is just that, a model. Whether it involves using a mathematical function or an empirically-derived one, the assumption in the Jtc routine is that the distribution of incidents will provide information about the home base location of the offender. In this sense, it's not unlike the way most crime analysts will work when they are trying to find a serial offender. A typical approach will be to plot the distribution of incidents and routinely search a geographic area in and around a serial crime pattern, noting offenders who have an arrest history matching case attributes (MO, type weapon, suspect description, etc.). Because a high proportion of offenses are committed within a short distance of offender residence's, the method can frequently lead to their apprehension. But, in doing this method, the analysts are not using a sophisticated statistical model.

Test Sample of Serial Offenders

To explore the accuracy of the approach, a small sample of 50 serial offenders was isolated from the database and used as a target sample to test the accuracy of the methods. The 50 offenders accounted for 520 individual crime incidents in the database. To test the Jtc method systematically, the following distribution was selected (table 9.4). The sample was not random, but was selected to produce a balance in the number of incidents committed by each individual and to, roughly, approximate the distribution of incidents by serial offenders. Each of the 50 offenders was isolated as a separate file so that each could be analyzed in *CrimeStat*.

Identifying the Crime Type

Each of the 50 offenders was categorized by a crime type. Only two of the offenders committed the same crime for all their offenses; most committed two or more different types of crimes. Arbitrarily, each offender was typed according to the crime type that he/she most frequently committed; in the two cases where there was a tie between two crime types, the most severe was selected (i.e., personal crime over property crime). While I recognize that there is arbitrariness in the approach, it seemed a practical solution. Any error in categorizing an offender would be applicable to all the methods. The crime types for the 50 offenders approximately mirrored the distribution of incidents: larceny (29); vehicle theft (7); burglary (5); robbery (5); assault (2); bank robbery (1); and arson (1).

406

Table 9.4

Serial Offenders Used in Accuracy Evaluation

Number of Offenders	Number of Crimes Committed by Each Person
4	3
4	4
4	5
4	6
4	7
4	8
3	9
3	10
3	11
2	12
2	13
2	14
2	15
1	16
1	17
1	18
1	19
1	20
1	21
1	22
1	24
1	33
50	520

Identifying the Home Base and Incident Locations

In the database, each of the offenders was listed as having a residence location. For the analysis, this was taken as the *origin* location of the journey to crime trip. Similarly, the incident location was taken as the *destination* for the trip. Operationally, the crime trip is taken as the distance from the origin location to the destination location. However, it is very possible that some crime trips actually started from other locations. Further, many of these individuals have moved their residences over time; we only have the last known residence in the database. Unfortunately, there was no other information in the digital database to allow more accurate identification of the home location. In other words, there may be, and probably are, numerous errors in the estimation of the journey to crime trip.

407

However, these errors would be similar across all methods and should not affect their relative accuracy.

Evaluated Methods

Ten methods were compared in estimating the likely residence location of the offenders. Four of the methods used the Jtc routines and six were simple spatial distribution methods (table 9.5).

Table 9.5

Comparison Methods for Estimating the Home Base of a Serial Offender

Journey to Crime Methods

Mathematical model for all crimes

Mathematical model for specific crime type

Kernel density model for all crimes

Kernel density model for specific crime type

Spatial Distribution Methods

Mean center

Center of minimum distance

Directional mean (weighted) calculated with 'lower left corner' as origin

Triangulated mean

Geometric mean

Harmonic mean

The mean center and center of minimum distance are discussed in chapter 4. The center of minimum distance, in particular, is more or less the geographic center of distribution in that it ignores the values of particular locations; thus, locations that are far away from the cluster (extreme values) have no effect on the result. The directional mean and triangulated mean is part of the directional mean routine, discussed in chapter 4 and

408

in the update release notes; the routine has now been modified so that it can be used with ordinary X/Y coordinates. The geometric and harmonic means are discussed in the update release notes; they are both means which discount extreme values.

The Test

Each of these ten methods were run against each of the files created for the serial offenders. For the six 'means' (mean center, geometric mean, harmonic mean, directional mean, triangulated mean, center of minimum distance), the mean was itself the best guess for the likely residence location of the offender. For the four journey to crime functions, the grid cell with the highest likelihood estimate was the best guess for the likely residence location of the offender.

Measurement of Error

For each of the 50 offenders, error was defined as the distance in miles between the 'best guess' and the actual location. For each offender, the distance between the estimated home base (the 'best guess') and the actual residence location was calculated using direct distances. Table 9.6 presents the results. The data show the error by method for each of the 50 offenders. The three right columns show the average error of all methods and the minimum error and maximum errors obtained by a method. The method with the minimum error is boldfaced; for some cases, two or three methods are tied for the minimum. The bottom two rows show the average error and the standard deviation of the errors for each method across all 50 offenders.

There are a number of conclusions from the results. First, the degree of precision for any of these methods varies considerably. The precision of the estimates vary from a low of 0.0466 miles (about 246 feet) to a high of 75.7 miles. The overall precision of the methods is not very high and is highly variable. There are a number of possible reasons for this, some of which have been discussed above. Each of the methods produces a single parameter from what is, essentially, a probability distribution whereas the distribution of many of these incidents are widely dispersed. Few of the offenders had such a concentrated pattern that only a single location was possible. Since these are probability distributions, not everyone follows the 'central tendency'. Also, some of these offenders may have moved during the period indicated by the incidents, thereby shifting the spatial pattern of incidents and making it difficult to identify the last residence.

A second conclusion is that, for any one offender, the methods produce similar results. For many of the offenders the difference between the best estimate (the minimum error) and the worst estimate (the maximum error) is not great. Thus, the simple methods are generally as good (or bad) as the more sophisticated methods.

Third, across all methods, the center of minimum distance had the lowest average error. Thus, the approximate geographic center of the distribution produced as good an estimate as the more sophisticated methods. However, it wasn't particularly close (3.8441 miles, on average). The worst method was the triangulated mean; it had an average error

Table 9.6

Accuracy of Methods for Estimating Serial Offender Residences
(N= 50 Serial Offenders)

Dataset	Number of Crimes	Primary Crime Type	Mean Center Error (miles)	Center of Minimum Distance Error (miles)	Triangulated Mean Error (miles)	Geometric Mean Error (miles)	Harmonic Mean Error (miles)	Jtc Kernel: All Crimes Error (miles)	Jtc Kernel: Crime Type Error (miles)	Jtc Math: All Crimes Error (miles)	Jtc Math: Crime Type Error (miles)	Average Error	All Methods Minimum Error	Maximum Error
3A	3	Larceny	31.5991	32.4477	32.4109	31.5995	31.6000	32.7824	32.7880	32.7824	32.7880	32.3109	31.5991	32.7880
3B	3	Larceny	13.2303	12.1683	24.1531	13.2311	13.2319	10.7526	14.4929	10.7526	11.2501	13.6959	10.7526	24.1531
3C	3	Bank robbery	2.8348	0.9137	2.7767	2.8335	2.8322	0.6775	5.8416	0.6775	6.0946	2.8313	0.6775	6.0946
3D	3	Burglary	2.9733	3.2603	6.1013	2.9728	2.9724	4.6038	3.3883	3.3882	3.7931	3.7170	2.9724	6.1013
4A	4	Vehicle theft	4.2436	4.2670	3.8217	4.2436	4.2436	4.2527	4.2364	4.2527	4.2590	4.2022	3.8217	4.2670
4B	4	Larceny	1.9618	0.3100	2.0563	1.9621	1.9623	0.3125	0.2018	0.3125	0.2784	1.0397	0.2018	2.0563
4C	4	Larceny	4.4733	4.4733	4.6789	4.4733	4.4733	4.9681	4.3563	4.2637	4.3563	4.5018	4.2637	4.9681
4D	4	Assault	0.2925	0.1905	0.0466	0.2925	0.2926	0.0703	0.0703	0.0703	0.4560	0.1979	0.0466	0.4560
5A	5	Larceny	17.3306	16.6459	17.8985	17.3292	17.3276	15.9738	17.8655	15.9739	16.4526	16.9775	15.9738	17.8985
5B	5	Larceny	1.3609	0.2481	1.7733	1.3586	1.3564	0.2068	0.6974	0.5140	0.6974	0.9126	0.2068	1.7733
5C	5	Larceny	2.2458	2.6832	16.4518	2.2450	2.2442	2.7886	2.4205	2.7886	3.0922	4.1067	2.2442	16.4518
5D	5	Larceny	0.9169	0.2250	0.2371	0.9171	0.9174	0.1577	0.4267	0.1577	0.4267	0.4889	0.1577	0.9174
6A	6	Larceny	5.1837	5.2081	7.9621	5.1837	5.1837	5.1271	4.8554	4.9393	5.2256	5.4298	4.8554	7.9621
6B	6	Vehicle theft	1.3720	1.1889	0.9625	1.3710	1.3700	3.1126	2.3600	1.3566	2.0831	1.6983	0.9625	3.1126
6C	6	Larceny	1.3199	0.3157	1.7928	1.3192	1.3184	0.2580	0.5272	0.2580	0.5272	0.8485	0.2580	1.7928
6D	6	Larceny	3.2458	2.3324	6.5209	3.2431	3.2405	1.2506	2.6253	1.9718	1.9718	2.9336	1.2506	6.5209
7A	7	Larceny	3.9023	3.4185	2.3176	3.9022	3.9021	2.7419	3.0532	3.1364	3.0532	3.2697	2.3176	3.9023
7B	7	Larceny	12.4100	9.2973	14.8293	12.4107	12.4115	8.5357	8.6148	8.5357	8.8275	10.6528	8.5357	14.8293
7C	7	Burglary	5.0501	7.1477	10.8567	5.0481	5.0460	7.9975	7.9975	7.9975	7.6274	7.1965	5.0460	10.8567
7D	7	Larceny	2.2686	0.7733	75.7424	2.2684	2.2682	0.0892	0.7191	0.0892	0.7191	9.4375	0.0892	75.7424
8A	8	Larceny	6.0298	6.0165	8.2653	6.0264	6.0229	8.4210	6.2962	6.2022	6.1166	6.3774	6.0165	8.4210
8B	8	Larceny	1.0041	1.1437	2.1776	1.0042	1.0042	1.7475	1.5298	1.3510	1.3510	1.3681	1.0041	2.1776
8C	8	Larceny	1.3058	1.5944	1.3684	1.3043	1.3027	2.1513	1.2020	2.1513	1.8707	1.5946	1.2020	2.1513
8D	8	Vehicle theft	3.5794	2.3760	5.5915	3.5809	3.5825	0.5900	1.3340	1.9133	1.3340	2.6537	0.5900	5.5915
9A	9	Robbery	5.2527	5.7156	4.8574	5.2529	5.2532	7.8257	7.1961	6.2520	5.9265	5.9480	4.8574	7.8257
9B	9	Larceny	8.1923	10.6555	6.9916	8.1896	8.1850	12.4578	10.3957	12.4578	12.0514	9.9529	6.9916	12.4578
9C	9	Robbery	3.7776	3.8454	11.0042	3.7758	3.7738	4.9015	5.1862	4.6206	4.3445	5.0255	3.7738	11.0042
10A	10	Larceny	0.9358	0.5159	1.1003	0.9355	0.9353	0.0606	0.3720	0.2601	0.7172	0.6481	0.0606	1.1003
10B	10	Larceny	2.8581	3.4940	14.2219	2.8536	2.8491	6.4061	6.5709	10.3095	6.4758	6.2264	2.8491	14.2219
10C	10	Larceny	0.8052	0.7251	5.5938	0.8050	0.8049	0.9059	0.8404	0.9060	1.2786	1.4072	0.7251	5.5938
11A	11	Vehicle theft	2.9127	3.2715	3.1192	2.9130	2.9134	3.6936	3.4335	3.4282	3.2087	3.2104	2.9127	3.6936
11B	11	Robbery	0.3250	0.3250	0.2613	0.3250	0.3250	0.4235	0.2263	0.4235	0.7011	0.3695	0.2263	0.7011
11C	11	Vehicle theft	1.2689	1.7157	1.4750	1.2709	1.2729	2.8945	0.6984	2.8945	2.2049	1.7440	0.6984	2.8945
12A	12	Larceny	3.3881	4.2334	10.9241	3.3867	3.3852	6.4050	3.2639	5.5843	5.2132	5.0871	3.2639	10.9241
12B	12	Larceny	0.5562	0.5361	2.8003	0.5562	0.5562	0.7897	0.6709	0.7897	0.9631	0.9132	0.5361	2.8003
13A	13	Larceny	6.3282	7.2857	6.0244	6.3248	6.3213	7.6438	7.4607	7.6438	7.9915	7.0027	6.0244	7.9915
13B	13	Assault	1.4943	1.4943	1.5279	1.4944	1.4944	1.6501	1.5954	1.6501	2.0824	1.6092	1.4943	2.0824
14A	14	Larceny	1.9363	0.8706	1.4498	1.9365	1.9368	0.3434	0.6058	0.2596	0.7631	1.1224	0.2596	1.9368
14B	14	Arson	0.6898	0.3727	0.8086	0.6899	0.6900	0.3359	0.3359	0.3359	0.6213	0.5422	0.3359	0.8086
15A	15	Vehicle theft	0.7282	0.7189	0.3362	0.7277	0.7271	0.8155	0.4855	0.8155	1.5128	0.7630	0.3362	1.5128
15B	15	Robbery	0.4914	0.4914	0.8254	0.4914	0.4914	0.6468	0.5693	0.8468	0.6546	0.5898	0.4914	0.8254
16A	16	Vehicle theft	2.1107	2.0995	8.2311	2.1107	2.1107	1.5957	1.6404	2.5911	2.4033	2.7659	1.5957	8.2311
17A	17	Burglary	1.6484	0.3093	1.0227	1.6461	1.6438	0.2879	0.2879	0.2879	0.5268	0.8512	0.2879	1.6484
18A	18	Larceny	0.6308	0.4198	1.0876	0.6329	0.6349	0.2132	0.3383	0.2132	0.6985	0.5410	0.2132	1.0876
19A	19	Larceny	8.6462	9.4195	8.6772	8.6486	8.6511	10.2869	9.2706	9.7022	9.5548	9.2064	8.6462	10.2869
20A	20	Burglary	6.3520	5.7969	28.3094	6.3486	6.3452	0.5934	0.6673	0.5934	0.7945	6.2223	0.5934	28.3094
21A	21	Burglary	1.2396	0.8661	1.2776	1.2393	1.2390	0.5243	0.5243	1.0253	0.4965	0.9391	0.4965	1.2776
22A	22	Larceny	3.6828	2.6232	2.0949	3.6803	3.6777	2.4937	2.8944	2.4937	2.8944	2.9484	2.0949	3.6828
24A	24	Larceny	1.7959	0.5892	2.3033	1.7975	1.7991	0.2658	0.3574	0.4222	0.6587	1.1099	0.2658	2.3033
33A	33	Robbery	3.9901	5.0481	7.2505	3.9940	3.9979	7.9485	7.6939	8.1907	7.9439	6.2286	3.9901	8.1907
		Mean Error =	4.0434	3.8441	7.6472	4.0429	4.0424	4.0395	4.0305	4.0163	4.1467			
		SD Error =	5.2166	5.3845	12.0642	5.2166	5.2166	5.5678	5.6237	5.5398	5.4177			

of 7.6472 miles. The triangulated mean is produced by vector geometry and will not necessarily capture the center of the distribution. Other than this, there were not great differences. This reinforces the point above that the methods are all, more or less, describing the central tendency of the distribution. For offenders that don't live in the center of their distribution, the error of a method will necessary be high.

Looking at each of the 50 offenders, the methods vary in their efficacy. For example, the Jtc kernel function for all crimes was the best or tied for best for 17 of the offenders, but was also the worst or tied for worst for 9. Similarly, the Jtc kernel function for the specific crimes was best or tied for best for 8 of the offenders, but worse for 4. Even the most consistent was best for 4 offenders, but also worst for one. On the other hand, the triangulated mean, which had the worst overall error, produced the best estimate for 9 of the individuals while it produced the worst estimate for 25 of the individuals. Thus, the triangulated mean tends to be very accurate or very inaccurate; it had the highest variance, by far.

Fourth, the amount of error varies by the number of incidents. Table 9.7 below shows the average error for each method as a function of three size classes: 1-5 incidents; 6-9 incidents; and 10 or more incidents. As can be seen, for each of the ten methods, the error decreases with increasing number of incidents. In this sense, the measured error is responsive to the sample size from which it is based. It is, perhaps, not surprising that with only a handful of incidents no method can be very precise.

Fifth, the relative accuracy of each of these methods varies by sample size. The method or methods with the minimum error are boldfaced. For a limited number of incidents (1-5), the Jtc mathematical function for all crimes (i.e., the negative exponential with the parameters from table 9.5) produced the estimate with the least error, followed by the Jtc kernel function for all crimes; the was the third best. The differences in error between these were not very great. For the middle category (6-9 incidents), the center of minimum distance produced the least error followed by the Jtc mathematical function for the specific crime type. For those offenders who had committed ten or more crimes, the Jtc kernel function for the specific crime type produced the best estimate, followed by the center of minimum distance. The two mathematical functions produced the least accuracy for this sub-group, though again the differences in error are not very big (2.2 miles for the best compared to 2.7 miles for the worst). In other words, only with a sizeable number of incidents does the Jtc kernel density approach for specific crimes produce a good estimate. It is better than the other approaches, but only slightly better than the simple measure of the center of minimum distance.

Cautionary Notes

Of course, this is a limited test. It was a small sample (only 50 cases) from a single jurisdiction (Baltimore County). The sample wasn't even randomly selected, but chosen to examine the accuracy by a range of sample sizes. Thus, the conclusions are only tentative and must be seen as hypotheses for further work. Clearly, more research is needed.

411

Table 9.7

Method Estimation Error and Sample Size:
Average Error of Method by Number of Incidents (miles)

Number of Incidents	Mean Center	Center of Minimum Distance	Triangulated Mean	Geometric Mean	Harmonic Mean	Jtc Kernel: All	Jtc Kernel: Crime types	Jtc Math: All	Jtc Math: Crime types	All Methods Average Error	Minimum Error
3-5	6.9553	6.4861	9.3672	6.9160	6.9545	6.4622	7.2321	6.3278	6.9954	7.0774	6.3278
6-9	4.2596	4.0753	10.6160	4.3331	4.2576	4.4805	4.2489	4.2274	4.2020	4.9667	4.0753
10+	2.3832	2.3149	4.8136	2.4575	2.3827	2.4880	2.2176	2.6725	2.6243	2.7060	2.2176

Nevertheless, there are certain cautions that must be considered in using either of these journey to crime methods (the mathematical or the empirical). First, a simple technique, such as the center of minimum distance, may be as good as a more sophisticated technique. It doesn't always follow that a sophisticated method will produce any more accuracy than a simple one. For the time being, I would advise crime analysts who are trying to detect a pattern in the distribution of the incidents of a serial offender to do exactly what they have been doing, basically looking at the data and making a subjective guess about where the offender may be residing. The kernel density Jtc routine needs an adequate amount of information (i.e, at least 10 incidents) to produce somewhat precise estimates. These techniques should be seen for now as research tools rather than as diagnostics for identifying the whereabouts of an offender. They are just too imprecise and unreliable to depend on, at least until more definitive results are obtained.

Second, there are other limitations to the technique. The model must be calibrated for each individual jurisdiction. Further, it must be periodically re-calibrated to account for changes in crime patterns. For example, in using the mathematical model, one cannot take the parameters estimated for Baltimore County (Table 9.3) and apply them to another city or if using the kernel density method take the results found at one time period and assume that they will remain indefinitely. The model is a probability model, not a guarantee of certainty. It provides guesses based on the similarity to other offenders of the same type of crime. In this sense, a particular serial offender may not be typical and the model could actually orient police wrongly if the offender is different from the calibration sample. It will take insight by the investigating officers to know whether the pattern is typical or not.

Third, as a theoretical model, the journey to crime approach is quite simple. It is based on a distribution of incidents and an assumed travel distance decay function. From the perspective of modeling the travel behavior of offenders, it is limited. As mentioned above, the method does not utilize information on the distribution of target opportunities nor does it utilize information on the travel mode and route that an offender takes. It is purely a statistical model. The research area of geographic profiling attempts to go beyond statistical description and understand the cognitive maps that offenders use as well as how these interact with their motives. This is good and should clearly guide future research. But it has to be understood that the theory of offender travel behavior is not very well developed, certainly compared to other types of travel behavior. Further, some types of crime trips may not even start from an offender's residence, but may be referenced from another location, such as vehicle thefts occurring near disposal locations. Routine activity theory would suggest multiple origins for crimes (Cohen and Felson, 1979).

The existing models of travel demand used by transportation planners (which have themselves been criticized for being too simple) measure a variety of factors that have only been marginally included in the crime travel literature - the availability of opportunities, the concentration of offender types in certain areas, the mode of travel (i.e., auto, bus, walk), the specific routes that are taken, the interaction between travel time and travel route, and other factors. It will be important to incorporate these elements into the understanding of journey to crime trips to build a much more comprehensive theory of how

413

offenders operate. Travel behavior is very complicated and we need more than a statistical distance model to adequately understand it.

Also, it's not clear whether knowing an offender's 'cognitive map' will help in prediction. There have been no evaluations that have compared a strictly statistical approach with an approach that utilizes information about the offender as he or she understands the environment. It cannot be assumed that integrating information about the perception of the environment will aid prediction. In most travel demand forecasts that transportation engineers and planners make, cognitive information about the environment is not utilized except in the definition of trip purpose (i.e., what the purpose of the trip was). The models use the actual trips by origin and destination as the basis for formulating predictions, not the understanding of the trip by the individual. Understanding is important from the viewpoint of developing theory or for ways to communicate with people. But, it is not necessarily useful for prediction. In short, understanding and prediction are not the same thing.

On the other hand, the journey to crime routine, particularly the kernel density approach, can be useful for police departments *if* used carefully. If there are sufficient cases to build an estimate (i.e., 10 or more incidents), it can provide additional information to officers investigating a serial offender by reducing the number of possible suspects that might be linked to a series of crimes. It can also provide some direction in orienting the deployment of officers and detectives investigating what appear to be serial offenses. It provides guesses about where the offender might be living, but based on similarities with previous offenders for the same type of crime. It's not going to give an exact estimate of where an offender is living, but will provide some insights into which areas the individual might be located. The Jtc model should be seen as a supplement to other techniques, not a complete solution. Like all the statistical tools in *CrimeStat*, it must be used carefully and intelligently. The philosophy of crime analysis must always be to use a technique with thought and with a systematic procedure.

414

Endnotes for Chapter 9

1. It should also be pointed out that the use of direct distances will underestimate travel distances particularly if the street network follows a grid.

2. There are, of course, many other types of mathematical functions that can be used to describe a declining likelihood with distance. In fact, there are an infinite number of such functions. However, the five types of functions presented here are commonly used. We avoided the inverse distance function because of its potential to distort the likelihood relationship.

$$f(d) = \frac{1}{d_{ij}^{k}}$$

where k is a power (e.g., 1, 2, 2.5). For large distances, this function can be a useful approximation of the lessening travel interaction with distance. However, for short distances, it doesn't work. As the distance between the reference cell location and an incident location becomes very small, approaching zero, then the likelihood estimate becomes very large, approaching infinity. In fact, for $d_{ij} = 0$, the function is unsolvable. Since many distances between reference cells and incidents will be zero or close to zero, the function becomes unusable.

3. It is actually the inverse of the inverse distance function. If a distance decay function drops off proportional to the inverse of the distance,

$$Y_{ij} = A * 1/d_{ij}$$

where Y_{ij} is the travel likelihood, A is coefficient, and d_{ij} is the distance from the home base, then the opposite - a distance increase is just the inverse of this function

$$Z_{ij} = \frac{1}{A * 1/d_{ij}} = \frac{d_{ij}}{A} = B * d_{ij}$$

4. There are several sources of error associated with the data set. First, these records were arrest records prior to a trial. Undoubtedly, some of the individuals were incorrectly arrested. Second, there are multiple offenses. In fact, more than half the records were for individuals who were listed two or more times in the database. The travel pattern of repeat offenders may be slightly different than for apparent first-time offenders (see figure 9.19). Third, many of these individuals have lived in multiple locations. Considering that many are young and that most are socially not well adjusted, it would be expected that these individuals would have multiple homes. Thus, the distribution of incidents could reflect multiple home bases, rather than one. Unfortunately, the data we have only gives a single residential location, the place at which they were living when arrested.

415

5. If the coordinate system is projected with the distance units in feet, meters or miles, then the distance between two points is the hypotenuse of a right triangle using Euclidean geometry.

$$d_{AB} = \sqrt{(X_A - X_B)^2 + (Y_A - Y_B)^2} \qquad (3.1)$$
$$\text{repeat}$$

where each location is defined by an X and Y coordinate in feet, meters, or miles.

If the coordinate system is spherical with units in latitudes and longitudes, then the distance between two points is the Great Circle distance. All latitudes and longitudes are converted into radians using

$$\text{Radians } (\phi) = \frac{2\pi\,\phi}{360} \qquad (3.2)$$
$$\text{repeat}$$

$$\text{Radians } (\lambda) = \frac{2\pi\,\lambda}{360} \qquad (3.3)$$
$$\text{repeat}$$

Then, the distance between the two points is determined from

$$d_{AB} = 2*\text{Arcsin}\{\text{Sin}^2[(\phi_B - \phi_A)/2] + \text{Cos }\phi_A * \text{Cos}\phi_B * \text{Sin}^2[(\lambda_B - \lambda_A)/2]^{1/2}\} \qquad (3.4)$$
$$\text{repeat}$$

with all angles being defined in radians (Snyder, 1987, p. 30, 5-3a).

6. They also argued that the combination of incidents - which they called 'aggregation', would distort the relationship between distance and incidence likelihood because of the ecological fallacy. To my mind, they are incorrect on this point. Data on a distribution of incidents by distance traveled is an individual characteristic and is not 'ecological' in any way. An ecological inference occurs when data are aggregated with a *grouping* variable (e.g., state, county, city, census tract; see Langbein and Lichtman, 1978). A frequency distribution of individual crime trip distances is an individual probability distribution, similar, for example, to a distribution of individuals by height, weight, income or any other characteristic. Of course, there are sub-sets of the data that have been aggregated (similar to heights of men v. heights of women, for example). Clearly, identifying sub-groups can make better distinctions in a distribution. But, it is still an individual probability distribution. This doesn't produce bias in estimating a parameter, only variability. For example if a particular distance decay function implies that 70% of the offenders live within, say, 5 miles of their committed incidents, then 30% don't live within 5 miles. In other words, because the data are individual level, then a distance decay function, whether estimated by a mathematical or a kernel density model, is an individual probability model (i.e., an attempt to describe the underlying distribution of individual travel distances for journey to crime trips).

416

Chapter 10
Space-Time Analysis

In this chapter, we discuss three techniques that are used to analyze the relationship between space and time. Up to this point, we have analyzed the distribution of incidents irrespective of the order in which they appeared or in which the time frame in which they appeared. The only temporal analysis that was conducted was in Chapter 4 where several spatial description indices, including the standard deviational ellipse, were compared for different time periods.

As police departments usually know, however, the spatial patterning of incidents doesn't occur uniformly throughout the year, but instead are often clustered together during short time periods. At certain times, a rash of incidents will occur in certain neighborhoods and the police often have to respond quickly to these events. In other words, there is both clustering in time as well clustering in space. This area of research has been developed mostly in the field of epidemiology (Knox, 1963, 1988; Mantel, 1967; Mantel and Bailar, 1970; Besag and Newell, 1991; Kulldorf and Nargawalla, 1995; Bailey and Gattrell, 1995). However, most of these techniques are applicable to crime analysis and criminal justice research as well.

CrimeStat includes three space-time techniques: the Knox Index, the Mantel index, and Correlated Walk Analysis. Figure 10.1 shows the Space-Time Analysis screen.

Measurement of Time in *CrimeStat*

Time can be defined as hours, days, weeks, months, or years. The default is days. However, please note that for any of these techniques, in *CrimeStat*, time must be measured as an *integer* or *real* variable, as mentioned in Chapter 3. Time cannot be defined by a formatted date code (e.g., 11/06/01, July 30, 2002). Each of the three space-time routines expect time to be an integer or real variable (e.g., 1, 2, 34527, 2.8). If given formatted dates, they will calculate an answer, but the result will not be correct.

If the time unit is days, a simple transformation is to use the number of days since January 1, 1900. Most spreadsheet and data base programs usually assign an integer number from this reference point. For example, November 12, 2001 has the integer value of 37207 while January 30, 2002 has the integer value of 37286. These are the number of days since January 1, 1900. Any spreadsheet program (e.g., Excel or Lotus 1-2-3) can convert a date format into a real number with the Value function. Also, any arbitrary numbering system will work (e.g., 1, 2, 3).

Space-Time Interaction

There are different types of interaction that could occur between space and time. Four distinctions can be made. First, there could be *spatial clustering* all the time. Certain communities are prone to certain events. For example, robberies often are

417

Figure 10.1: Space-time Analysis

concentrated in particular locations as are vehicle thefts. The hot spot methods that were discussed in chapters 6, 7 and 8 are useful for identifying these concentrations. In this case, there is no space-time interaction since the clustering occurs all the time.

Second, there could be *spatial clustering within a specific time period*. Hot spots could occur during certain time periods. For example, motor vehicle crashes tend to occur with much higher frequencies in the late afternoon and early evening, often as a by-product of congestion on the roads. Crash hot spots will tend to appear at certain times because of the congestion. At most other times, the concentration does not occur because the congestion levels are lower.

Third, there could be *space-time clustering*. A number of events could occur within a short time period within a concentrated area. This type of effect is very common with motor vehicle thefts. A car thief gang may decide to attack a particular neighborhood. After a binge of car thefts, they move on to another neighborhood. In this instance, there are a number of theft incidents that are occurring within a limited period in a limited location. The cluster moves from one location to another. In this case, there is an interaction between space and time in that spatial hot spots appear at particular times, but are temporary. The ability to detect this type of shift is very important to police departments since it affects their ability to respond.

Fourth, there could be *space-time interaction* in which the relationship between space and time in more complex. The interaction could be concentrated, as in the spatial clustering mentioned above, or it could follow a more complex pattern. For example, there could be a diffusion of drug sales from a central location to a more dispersed area. Whereas initially, the drug dealing is concentrated in a few locations, it starts to diffuse to other areas. However, the diffusion may occur at different times of the year (e.g., Christmas and New Years). Alternatively, vehicle thefts may shift towards seaside communities during the summer months when the number of vacationers increases. We saw an example of this in chapter 4 where the ellipse of motor vehicle thefts shifted between June and July to the communities along the Chesapeake River near Baltimore. This type of diffusion is not clustering *per se*, in that it may be spread over a very large coastline. But it is a distinct space-time interaction.

The importance of these distinctions is that many of the space-time tests that exist only measure gross space-time interaction, rather than space-time clustering. For example, the Knox and Mantel tests that follow test for spatial interaction. The interaction could be the result of spatial clustering, but doesn't necessarily have to be. The interaction could occur in a very complex way that would not easily lend itself to more focused intervention by the police. Still, the ability to identify the interaction is an important step in planning an intervention strategy.

Knox Index

The Knox Index is a simple comparison of the relationship between incidents in terms of distance (space) and time (Knox, 1963; 1964). That is, each individual pair is

419

compared in terms of distance and in terms of time interval. Since each pair of points is being compared, there are N*(N-1)/2 pairs. The distance between points is divided into two groups - Close in distance and Not close in distance, and the time interval between points is also divided into two groups - Close in time and Not close in time. The definitions of 'close' and 'Not close' are left to the user.

A simple 2 x 2 table is produced that compares closeness in distance with closeness in time. The number of pairs that fall in each of the four cells are compared (Table 10.1).

Table 10.1

Logical Structure of Knox Index

	Close in time	Not close in time	
Close in Distance	O_1	O_2	S_1
Not close in distance	O_3	O_4	S_2
	S_3	S_4	N

where $N = O_1 + O_2 + O_3 + O_4$
$S_1 = O_1 + O_2$
$S_2 = O_3 + O_4$
$S_3 = O_1 + O_3$
$S_4 = O_2 + O_4$

The actual number of pairs that falls into each of the four cells are then compared to the expected number if there was no relationship between closeness in distance and closeness in time. The expected number of pairs in each cell under strict independence between distance and the time interval is obtained by the cross-products of the columns and row totals (table 10.2).

Table 10.2
Expected Frequencies for Knox Index

	Close in time	Not close in time
Close in Distance	E_1	E_2
Not close in distance	E_3	E_4

420

where $E_1 = S_1 * S_3 / N$
$E_2 = S_1 * S_4 / N$
$E_3 = S_2 * S_3 / N$
$E_4 = S_2 * S_4 / N$

The difference between the actual (observed) number of pairs in each cell and the expected number is measured with a Chi-square statistic (equation 10.1).

$$\chi^2 = \Sigma \; \frac{(O_i - E_i)^2}{E_i} \qquad \text{with 1 degree of freedom} \qquad (10.1)$$

Monte Carlo Simulation of Critical Chi-square

Unfortunately, the usual probability test associated with the Chi-square statistic cannot be applied since the observations are not independent. Interaction between space and time tend to be compounded when calculating the Chi-square statistic. For example, we've noticed that the Chi-square statistic tends to get larger with increasing sample size, a condition that would normally not be true with the independent observations. To handle the issue of interdependency, there is a Monte Carlo simulation of the chi-square value for the Knox Index under spatial randomness (Dwass, 1957; Barnard, 1963). If the user selects a simulation, the routine randomly selects M pairs of a distance and a time interval where M is the number of pairs in the data set (M = N* [N-1]/2) and calculates the Knox Index and the chi-square test. Each pair of a distance and a time interval are selected from the range between the minimum and maximum values for distance and time interval in the data set using a uniform random generator.

The random simulation is repeated K times, where K is specified by the user and Usually, it is wise to run the simulation 1000 or more times. The output includes:

1. The sample size
2. The number of pairs
3. The calculated chi-square value of the Knox Index from the data
4. The minimum chi-square value of the Knox Index from the simulation
5. The maximum chi-square value of the Knox Index from the simulation
6. Ten percentiles from the simulation:

   a. 0.5%
   b. 1%
   c. 2.5%
   d. 5%
   e. 10%
   f. 90%
   g. 95%

421

h.   97.5%
i.   99%
j.   99.5%

Example of the Knox Index

For an example, vehicle thefts in Baltimore County for 1996 were taken. There were 1855 vehicle thefts for which a date was recorded in the data base. The data base was further broken down into twelve separate monthly subsets. Using the median for both distance and time interval, the Knox Index was calculated for the entire set of 1855 incidents. Then, using the median distance for the entire year but a month-specific median time interval, the Knox Index was calculated for each of the twelve months. Table 10.3 presents the Chi-square values and their pseudo-significance levels.

To produce a better test of the significance of the results, 1000 random simulations were calculated for the vehicle theft for the entire year. Table 10.3 below shows the results. Because an extreme value could be obtained by chance with a random distribution, reasonable cut-off points are usually selected from the simulation. In this case, we want a cut-off point that approximates a 5% significance level. Since the Knox Index is a one-tailed test (i.e., only a high chi-square value is indicative of spatial interaction), we adopt an upper threshold of the 95 percentile. In other words, only if the observed chi-square test for the Knox Index is larger than the 95 percentile threshold will the null hypothesis of a random distribution between space and time be rejected.

Methods for Dividing Distance and Time

In the *CrimeStat* implementation of the Knox Index, the user can divide distance and time interval based on the three criteria:

1.   The mean (mean distance and mean time interval). This is the default.

2.   The median (median distance and median time interval)

3.   User defined criteria for distance and time separately.

There are advantage to each of these methods. The mean is the center of the distribution; it denotes a balance point. The median will divide both distance and time interval into approximately equal numbers of pairs. The division is approximate since the data may not easily divide into two equal numbered groups. A user-defined criteria can fit a particular need of an analyst. For example, a police department may only be interested in incidents that occur within two miles of each other within a one week period. Those criteria would be the basis for dividing the sample into 'Close' and 'Not close' distances and time intervals.

422

Table 10.3

Knox Index for Baltimore County Vehicle Thefts
Median Split
N = 1,855 with 1,719,585 comparisons

Month	Actual Chi-square	95 Percentile Simulation Chi-square	Approx. p
January	0.26	6.95	n.s.
February	0.00	6.61	n.s.
March	0.00	6.86	n.s.
April	0.50	6.56	n.s.
May	1.04	7.25	n.s.
June	0.01	6.02	n.s.
July	8.43	8.20	.01
August	5.91	8.29	.025
September	0.27	5.41	n.s.
October	3.33	6.43	n.s.
November	10.79	6.15	.01
December	0.00	6.87	n.s.
All of 1996	**8.69**	**41.89**	**n.s.**

For the entire year, there was not a significant clustering between space and time. Approximately, 26.7% of the incidents were both close in distance (i.e., closer than the median distance between pairs of incidents) and close in time (i.e., closer than the median time interval between pairs of incidents). However, when individual months are examined, three show significant relationships: July, August and November. During these months, there is an interaction between space and time. Typically, incidents that cluster together spatially tend also to cluster together temporally. However, it could be the opposite (i.e., events that cluster together temporally tend to be far apart spatially).

The next step would to identify whether there are particular clusters that occur within a short time period. Using one of the 'hot spot' analysis methods discussed in chapters 6 and 7, an analyst could take the events for the three months and try to identify whether there is spatial clustering during those three months that does not normally occur. We won't do that here, but the point is that the Knox Index is useful to identify *when* there is spatial clustering.

Problems with the Knox Index

The Knox Index is a simple measure of space-time clustering. However, because it is only a 2 x 2 table, different results can be obtained by varying the cut-off points for distance or time. For example, using the mean as the cut-off, the overall Chi-square statistic for all vehicle thefts was 8.67, reasonably close. However, when a cut-off point for

423

distance of 1000 meters and a cut-off point for time of 80 days was used, the Chi-square statistic dropped to 3.16. In other words, the Knox Index will produce different results for different cut-off points.

A second problem has to do with the interpretation. As with any Chi-square test, differences between the observed and expected frequencies could occur in any cell or any combination of cells. Finding a significant relationship does not automatically mean that events that were close in distance were also close in time; it could have been the opposite relationship. However, a simple inspection of the table can indicate whether the relationship is as expected or not. In the above example, all the significant relationships showed a higher proportion of events that were both close in distance and close in time.

Mantel Index

The Mantel Index resolves some of the problems of the Knox Index. Essentially, it is a correlation between distance and time interval for pairs of incidents (Mantel, 1967). More formally, it is a general test for the correlation between two *dissimilarity* matrices that summarizes comparisons between pairs of points (Mantel and Bailar, 1970). It is based on a simple cross-product of two interval variables (e.g., distance and time interval):

$$T = \sum_{i=1}^{N} \sum_{j=1}^{N} (X_{ij} - MeanX)(Y_{ij} - MeanY) \tag{10.2}$$

where X_{ij} is an index of similarity between two observations, i and j, for one variable (e.g., distance) while Y_{ij} is an index of similarity between the same two observations, i and j, for another variable (e.g., time interval).

The cross-product is then normalized by dividing each deviation by its standard deviation:

$$r = \frac{1}{(N-1)} \sum_{i=1}^{N} \sum_{j=1}^{N} (X_{ij} - MeanX)/S_x * (Y_{ij} - MeanY)/S_y \tag{10.3}$$

$$= \sum_{i=1}^{N} \sum_{j=1}^{N} Z_x * Z_y / (N-1)$$

where X_{ij} and Y_{ij} are the original variables for comparing two observations, i and j, and Z_x and Z_y are the normalized variables.

In *CrimeStat*, the Mantel Index routine calculates the correlation between distance and time interval. To illustrate, table 10.4 examines the Mantel correlation for the 1996

424

Table 10.4

Mantel Index for Baltimore County Vehicle Thefts
Median Split
N = 1,855 and 1,719,585 Comparisons

Month	r	Simulation 2.5%	Simulation 97.5%	Approx. p-level
January	-.0047	-0.033	0.033	n.s.
February	-.0023	-0.037	0.042	n.s.
March	-.0245	-0.032	0.039	n.s.
April	0.0077	-0.040	0.041	n.s.
May	0.0018	-0.038	0.043	n.s.
June	0.0043	-0.035	0.041	n.s.
July	0.0348	-0.034	0.033	.025
August	0.0544	-0.034	0.035	.01
September	0.0013	-0.044	0.046	n.s.
October	0.0409	-0.037	0.043	n.s.
November	0.0630	-0.042	0.040	.001
December	0.0086	-0.035	0.038	n.s.
All of 1996	0.0015	-0.009	0.010	n.s.

vehicle thefts in Baltimore County that was illustrated above. As seen, the correlations are all low. However, as with the Knox Index, July, August and November produce relatively higher correlations. If used as an index, rather than an estimate of variance explained, the Mantel Index can identify time periods when spatial interaction is occurring.

Monte Carlo Simulation of Confidence Intervals

Even though the Mantel Index is a Pearson product-moment correlation between distance and time interval, the measures are not independent and, in fact, are highly interdependent. Consequently, the usual significance test for a correlation coefficient is not appropriate. Instead, the Mantel routine offers a simulation of the confidence intervals around the index. If the user selects a simulation, the routine randomly selects M pairs of a distance and a time interval where M is the number of pairs in the data set (M = N* [N-1]/2) and calculates the Mantel Index. Each pair of a distance and a time interval are selected from the range between the minimum and maximum values for distance and time interval in the data set using a uniform random generator.

The random simulation is repeated K times, where K is specified by the user. Usually, it is wise to run the simulation 1000 or more times. The output includes:

1.   The sample size
2.   The number of pairs

425

3.   The calculated Mantel Index from the data
4.   The minimum Mantel value from the simulation
5.   The maximum Mantel value from the simulation
6.   Ten percentiles from the simulation:

    a.   0.5%
    b.   1%
    c.   2.5%
    d.   5%
    e.   10%
    f.   90%
    g.   95%
    h.   97.5%
    i.   99%
    j.   99.5%

To illustrate, 1000 random simulations were calculated for each month using the same sample size as the monthly vehicle theft totals. Table 10.4 above shows the results. Because an extreme value could be obtained by chance with a random distribution, reasonable cut-off points are usually selected from the simulation. In this case, we want cut-off points that approximate a 5% significance level. Since the Mantel Index is a two-tailed test (i.e., one could just as easily get dispersion between space and time as clustering), we adopt a lower threshold of the 2.5 percentile and an upper threshold of 97.5 percentile. Combined, the two cut-off points ensure that approximately 5% of the cases would be either lower than the lower threshold or higher than the upper threshold under random conditions.[1] In other words, only if the observed Mantel Index is smaller than the lower threshold or larger than the upper threshold will the null hypothesis of a random distribution between space and time be rejected.

In Table 10.4, for the entire year, the observed Mantel Index (correlation between space and time) was 0.0015. The 2.5 percentile was -.009 and the 97.5 percentile was 0.01. Since the observed value is between these two cut-off points, we cannot reject the null hypothesis of no relationship between space and time. However, for the individual months, again, July, August and November have correlations above the upper cut-off threshold. Thus, for those three months *only*, the amount of space-time clustering in the vehicle theft data is most likely greater than what would be expected on the basis of a chance distribution. One would, then, have to explore the data further to find out where those vehicle thefts were occurring, using one the hot spot routines in Chapter 6.

Limitations of the Mantel Index

The Mantel Index is a useful measure of the relationship between space and time. But it does have limitations. First, because it is a Pearson-type correlation coefficient, it is prone to the same types of problems that befall correlations. Extreme values of either space or time could distort the relationship, either positively, if there are one or two

426

observations that are extreme in both distance in time interval, or negatively, if there are only one or two observations that are extreme in either distance or in time interval.

Second, because the test is a comparison of all pairs of observations, the correlations tend to be small, as noted above. This makes it less intuitive as a measure than, say, a traditional correlation coefficient which varies between -1 and +1 and in which high values are expected. For most analysts, it is not very intuitive to have an index where 0.05 is a high value. This doesn't fault the statistic as much make it a little non-intuitive for users.

Third, as with any correlation coefficient, the same size needs to be fairly large to produce a stable estimate. In the above, example, one could further break down monthly vehicle thefts by week or, even, day. However, the number of cases will decrease considerably. In the above example, with 1,855 vehicle thefts over a year, the weekly average would be around 36, which is a small sample. Intuitively, a crime analyst wants to know when space-time clustering is occurring and a short time frame is critical for detection; a week would be the largest time interval that would be useful. However, as the sample size gets small, the index becomes unstable. For one thing, the sample size makes the index volatile. While the Monte Carlo simulation will adjust for the sample size, the range of the cut-off thresholds will vary considerably from one week to another with small sample sizes. The analyst will have to run the simulation repeatedly to adjust for the varying sample sizes. For another thing, the shortened time frame allows fewer distinctions in time; if one takes a very narrow time frame (e.g., a day), there can be virtually no time differences observed. One would have to switch to an hourly analysis to produce meaningful differences.

One way to get around this is to have a moving average where the time frame is adjusted to fit a constant number of days (e.g., a 14 day moving average). The advantage is that the sample size tends to remain fairly constant; one could therefore reduce the number of recalculations of the cut-off thresholds since they would not vary much from one day to another. To make this work, however, the data base must be set up to produce the appropriate number of incidents for a moving average analysis.

Nevertheless, the Mantel Index remains a useful tool for analysts. It is still widely used for space-time analysis and it has been generalized to many other types of dissimilarity analysis than just space and time. If used carefully, the index can be a powerful tool for detection of clusters that are also concentrated in time.

Correlated Walk Analysis

Correlated Walk Analysis (CWA) is a tool that is aimed at analyzing the spatial and temporal *sequencing* of incidents committed by a single serial offender. In this sense, it is the 'flip side' of Journey to crime analysis. Whereas journey to crime analysis makes guess about the likely origin location for a serial offender, based on the spatial distribution of the incidents committed by the offender, the CWA routine makes guesses about the time and location of a next event, based on both the spatial distribution of the incidents and the temporal sequencing of them.

427

The statistical origin of CWA is Random Walk Theory. Random Walk Theory has been developed by physicists to explain the distribution of molecules in a rapidly changing environment (e.g., the movements of a particle in a gas which is diffusing - Brownian movement). Sometimes called a 'drunkard's walk', the theory starts with the premise that movement is random in all directions. From an arbitrary starting point, a particle (or person) moves in any direction in a series of steps. The direction of each step is independent of the previous steps. After each step, a random decision is made and the person moves in a random direction. This process is repeated *ad infinitum* until an arbitrary stopping point is selected (i.e., the observer quits looking). It has been shown mathematically that all one and two dimensional random walks must eventually return to their original starting point (Spitzer, 1963; Henderson, Renshaw, and Ford, 1983).[2] This is called a *recurrent random walk*. On the other hand, independent random walks in more than two dimensions are not necessarily recurrent, a state called *transient random walk*.

Figure 10.2 illustrates a random walk of 2000 steps. For a large number of steps in a two-dimensional walk, the likely distance of a person (or particle) from the starting point is

$$E(d) = d_{rms} * \sqrt{N} \tag{10.4}$$

where $d_{rms} = \sqrt{(\Sigma d_i^2 / N)}$. The term, d_{rms} is the *root mean square* of distance.

There are a number of different types of random walks. The simplest is a movement of uniform distance only along a grid cell (i.e., a Manhattan geometry). The person can only move North, South, East or West for a unit distance of 1. A more complex random walk allows angular distances and an even more complex random walk allows varying distances (e.g., normally distributed random distances, uniformly random distances). The walk in figure 10.2 was of this latter type. X and Y values were selected randomly from a range of -1 to +1 using a uniform random number generator. For a conceptual understanding of Random Walk Theory, see Chaitin (1990) and, for a mathematical treatment, see Spitzer (1976). Malkiel (1999) applied the concepts of Random Walk Theory to stock price fluctuations in a book that has now become a classic.

Henderson, Renshaw and Ford (1983; 1984) have introduced the concept of a *correlated random walk*. In a correlated random walk, momentum is maintained. If a person is moving in a certain direction, they are more likely to continue in that direction than to reverse direction or travel orthogonally. In other words, at any one decision point, the probabilities of traveling in any direction are not equal; the same direction has a higher probability than an orthogonal change (i.e., turning 90 degrees) and those, in turn, have a higher probability than completely reversing direction. By implication, the same is true for distance and distance. A longer step than average is likely to be followed by another longer step than average while a shorter step than average is likely to be followed by another short step. Similarly, there is consistency in the time interval between events; a short interval is also likely to be followed by a short interval. In other words, a correlated random walk is a random walk with momentum (Chen and Renshaw, 1992; 1994). These

428

Figure 10.2:

A Random Walk

2000 Random Steps of -1.0 to +1.0 in X and Y Direction

Finish

Start
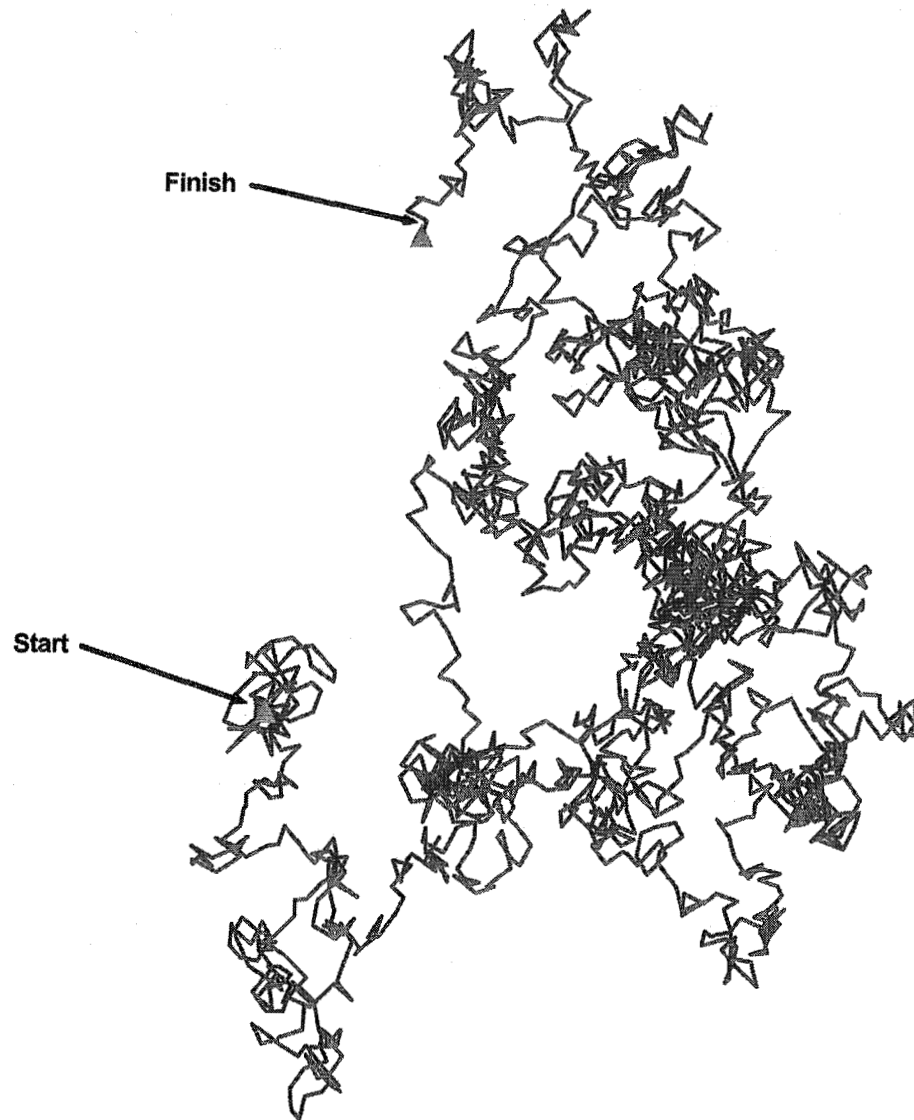
authors have applied the theory to the analysis of the branching of tree roots (Henderson, Ford, Renshaw, and Deans, 1983; Renshaw, 1985).

Correlated Walk Analysis

Correlated Walk Analysis is a set of tools that can help an analyst understand the sequencing of sequential events in terms of time interval, distance and direction. In *CrimeStat*, there are three CWA routines. The first two help the analyst understand whether there are patterns in time, distance or direction while the last routine allows the analyst to make a guess about the next likely event, when it will occur and where it will occur. The three routines are:

1. CWA - Correlogram
2. CWA - Diagnostics
3. CWA - Prediction

CWA - Correlogram

The *Correlogram* routine calculates the correlation in time interval, distance, and bearing (direction) between events. It does this through *lags*. A lag is a separation in the intervals between events. The difference between the first and second event is the first interval. The difference between the second and third events is the second interval. The difference between the third and fourth events is the third interval, and so forth. For each successive interval, there is a time difference; there is a distance; and there is a direction. One could extend this to all the intervals, comparing each interval with the next one; that is, we compare the first interval with the second, the second interval with the third, the third interval with the fourth, and so on until the sample is complete. When comparing successive intervals, this is called a *lag of 1*. It is important to keep in mind the distinction between an event (e.g., an incident) and an interval. It takes two events to create an interval. Thus, for a lag of 1, there are M= N-1 intervals where N is the number of events (e.g., for 3 incidents, there are 2 intervals).

A lag of two compares every other event. Thus, the first interval is compared to the third interval; the second interval is compared to the fourt; the third interval is compared to the fifth; and so on until there are no more intervals left in the sample. Again, the comparison is for time difference, distance, and direction separately. We can extend this logic to a lag of 3 (every third event), a lag of 4 (every fourth event), and so forth.

The CWA - Correlogram routine calculates the Pearson Product-Moment correlation coefficient between successive events. For a lag of 1, it compares successive events and correlates the time interval, distance, and bearing separately for these successive events. For a lag of 2, it compares every other event and correlates the time interval, distance, and bearing separately for these successive events. The routine does this until it reaches a maximum of 7 lags (i.e., every seventh event). However, if the sample size is very small, it may not be able to calculate all lags. It will require 12 incidents (events) to calculate all

430

seven lags since it requires at least four observations per lag (i.e., N - L - 4 where N is the number of events and L is the maximum number of lags calculated).

Adjusted Correlogram

The Correlogram calculates the raw correlation between intervals by lag for time, distance, and bearing. One of the problems that may appear, especially with small samples, is for higher-order lags to be very high, either positive or negative. There are probably two reasons for this. For one thing, with each lag, the sample size decreases by one; with a very small sample size, correlations can become very volatile, jumping from positive to negative, and from low to high. Another reason is that periodicity in the data set is compounded with higher-order lags in the form of 'echos'. For example, if a lag of 2 is high, then a lag of 4 will also be somewhat high since there is a compounding of the lag 2 effect. When combined with a small sample size, it is not uncommon to have higher-order lags with very high correlations, sometimes approaching +/- 1.0. The user must be careful in selecting a higher-order lag because there is an apparent effect which may be due to the above reasons, rather than any real predictability. One of the key signs for spurious higher-order effect is a sudden jump in the strength of the correlation from one lag to the next (though sometimes a high higher-order lag can be real; see examples below).

To minimize these effects, the output also includes an adjusted correlogram that adjusts for the loss of degrees of freedom. The formula is:

$$A = \frac{M - L - 1}{M - 1} \qquad (10.5)$$

where M is the number of intervals (N-1) and L is the number of lags. For example, for a sample size of 13, there will be 12 intervals (M). For a lag of 1, the adjustment will be

$$A = \frac{12 - 1 - 1}{12 - 1} = \frac{10}{11} = 0.909$$

The effect of the adjustment is to reduce the correlation for higher-order lags. It won't completely eliminate the effect, but it should help minimize spurious effects. As will be shown below, however, sometimes high higher-order lags are real.

CWA - Correlogram Output

The routine outputs 10 parameters:

1. The sample size (number of events);
2. Number of intervals;
3. Information on the units of time, distance, and bearing;
4. Final distance to origin in meters (distance between last and first event);

431

5. Expected random walk distance from origin (if sequence was strictly random);
6. Drift (the ratio of actual distance from origin to expected random walk distance);
7. Final bearing from origin (direction between last event and first event);
8. Expected random walk bearing. Defined as 0 because there is no expected direction.
9. Correlations by lag for time, distance, and bearing (up to 7 lags); and
10. Adjusted correlations by lag for time, distance, and bearing (up to 7 lags).

The aim of the CWA - Correlogram is to examine repetitive sequences, whether for time interval, distance or direction. It is possible to have separate repetitions for time, distance and direction. For example, an offender may commit crimes every 7 days or so, say, on the weekend. In this case, the individual is repeating himself/herself about once every week. Similarly, an individual may alternate directions, first going East then going West, then going back to the East, and so forth. In other words, what we're asking with the routine is whether there are any repetitions in the sequence of incidents committed by a serial offender. Does he/she repeat the crimes in time? If so, what is the *periodicity* (the repititious sequence? Does he/she repeat the crimes in distance? Is so, what is the periodicity? Finally, does he/she repeat the crimes in direction? If so, what is the periodicity? The CWA-Correlogram, therefore, analyzes the sequence of incidents committed by an individual and does this separately for time interval, distance, and direction.

Offender repetition

Why is this important? Most crime analysis is predicted on the assumption that offenders (people in general) repeat themselves, consciously or unconsciously. That is, individuals have specific behavior patterns that tend to be repeated. If an individual acts in a certain way (e.g., committing a burglary), then, most likely , the person will repeat himself/herself again. There is no guarantee, of course. But, because human beings do not behave spatially or temporally random but tend to operate in somewhat consistent ways, there is a likelihood that the individual will act in a similar manner again.

This assumption is the basis of profiling which aims at understanding the MO of an offender. If offenders were totally random in their behavior, detection and apprehension would be made much more difficult than it already is. So, between the two extremes of a totally random individual (the 'random walk person') and a totally predictable individual (the 'algorithmic person'), we have the bulk of human behavior, at least in terms of time, distance and direction.

CWA - Diagnostics

The Diagnostics routine is similar to the CWA - Correlogram except that it calculates an Ordinary Least Squares autoregression for a particular lag. That is, it regresses each interval against a previous interval. The user enters the lag number (the

432

default is 1) and the routine produces three regression models for the successive event as the dependent variable against the prior event as the independent variable. There are three equations, for time interval, distance, and bearing separately. The output includes:

1. The sample size (number of events);
2. The number of intervals;
3. Information on the units of time, distance, and bearing;
4. The multiple correlation coefficient;
5. The squared multiple correlation coefficient (i.e., R^2);
6. The overall standard error of estimate;
7. The regression coefficient for the constant and for the prior event;
8. The standard error of the regression coefficients;
9. The t-values for the regression coefficients;
10. The p-value (two-tail) for the regression coefficients;
11. An analysis of variance test for the full model. This includes sum of squares for the regression term and for the residual;
12. The ratio of the regression sum of squares to the residual sum of squares (the F-ratio); and
13. The p-value associated with the F-value.

What the regression diagnostics provides is an indicator of the amount of predictability in the lag. It has the same information as the Correlogram (since the square of the correlation, r^2, is the same as R^2 for a single independent variable regression equation), but it is easier to interpret. Essentially, it is argued below that, unless the R^2 in the regression equation is sufficiently high, that one is better off using the mean or median lag for prediction. Conversely, if the R^2 is very high, then the user should be suspicious about the data.

CWA - Prediction

Finally, after having analyzed the sequential pattern of events, the user can make a prediction about the time and place of the next event. There are three methods for making a prediction, each with a separate lag:

1. Mean difference
2. Median difference
3. Regression equation

The method is applied to the last event in the data set. The *mean difference* applies the mean interval of the data for the specified lag to the last event. For example, for time interval and a lag of 1, the routine calculates the interval between each event and takes the average. It then applies the mean time interval to the last time in the data set as the prediction. The *median difference* applies the median interval of the data for the specified lag to the last event. For example, for bearing and a lag of 1, the routine calculates the direction (bearing) between each event, calculates the median bearing, and applies that median average to the location of the last event in the data set as the predicted value.

433

The *regression equation* calculates a regression coefficient and constant for the specified lag and uses the data value for the last *interval* as input into the regression equation; the result is the predicted value. For example, for distance and a lag of 1, the routine calculates the regression coefficient and constant for a regression equation in which each event is compared to the previous event. The last distance in the data set (i.e., between the last event and the previous event) is used as an input for the regression equation and the predicted distance is marked off from the coordinates of the last event.

In other words, the routine takes the time and location of the last event and adds a time interval, a direction, and a distance as a predicted next event (next time, next location). The method by which this prediction is made can be the mean interval, the median interval, or the regression equation. If the user species a lag other than 1, that lag is applied to the last event. For example, for time with a mean difference and a lag of 2, the routine calculates the time interval between each event and every other event, calculates the average and applies that average to the last event in the data set.

CWA - Prediction Graphical Output

The CWA - Prediction routine outputs five graphical objects in 'shp', 'mif, or 'bna' formats. The user provides a file name and the routine adds five prefixes to the name in 'shp', 'mif' or 'bna' output:

1.   Events - a line indicating the sequence of events. If the user also brings in the points in the data set, it will be possible to number each of these steps;
2.   PredDest - the predicted location for the next event;
3.   Path - a line from the last location in the data set to the predicted location;
4.   POrigL - a point representing the center of minimum distance of the data set. The center of minimum distance is taken as a proxy for the origin location of the offender; and
5.   PW - a line from the expected origin to the predicted destination

For example, if the user provides the file name 'NightRobberies' and specifies a 'shp' output, there will be five objects output:

> EventsNightRobberies.shp
> PredDestNightRobberies.shp
> PathNightRobberies.shp
> POrigLNightRobberies.shp
> PWNightRobberies.shp

Example 1: A Completely Predictable Individual

The simplest way to illustrate the logic of the CWA is to start with a completely predictable individual. This individual commits crimes on a completely systematic basis. Table 10.5 illustrates the behavior of this individual.

434

Starting at an arbitrary origin with an X coordinate of 1 and a Y coordinate of 1 and on day 1, the individual commits 13 incidents in total. In the table, these are numbered events 1 through 13. Let's start with direction and distance. From the origin, the individual always travels in a Northeast direction of 45 degrees (clockwise from due North - 0 degrees). The individual's second incident is at coordinate X=2, Y=2. Thus, the individual traveled at 45 degrees from the previous incident and for a distance of 1.4142 (the hypotenuse of the right angle created by traveling one unit in the X direction and one unit in the Y direction). For the third incident, the individual commits this at X=4, Y=4. Thus, the direction is also at 45 degrees from the previous location but the distance is now 2.8284 (or the square root of 8 which comes from a step of 2 along the X axis and a step of 2 along the Y axis). For the fourth incident, the individual commits the crime at X=7, Y=7. Again, the direction is 45 degrees, but the distance is 4.2426 (or the square root of 18 which comes from a step of 3 along the X axis and a step of 3 along the Y axis).

Table 10.5

Example of a Predictable Serial Offender: 1
(N = 13 incidents)

Event	X	Y	Distance	Days	Time Interval
1	1	1	-	1	-
2	2	2	1.4142	3	2
3	4	4	2.8284	7	4
4	7	7	4.2426	9	2
5	8	8	1.4142	13	4
6	10	10	2.8284	15	2
7	13	13	4.2426	19	4
8	14	14	1.4142	21	2
9	16	16	2.8284	25	4
10	19	19	4.2426	27	2
11	20	20	1.4142	31	4
12	22	22	2.8284	33	2
13	25	25	4.2426	37	4
Logical prediction for next event 14	26	26	1.4142	39	2

For the fifth incident, again the individual travels at 45 degrees to the previous incident, but repeats himself/herself with a step of only 1 unit in both the X and Y directions. The individual then continues the sequence, always traveling in a 45 degree orientation to due North. For distance, a step of 1 in both the X and Y directions is

435

followed by a step of 2 in both directions, and is followed by a step of 3 in both directions. In other words, the individual repeats direction every time and repeats distance every third time. There is a periodicity of 1 for direction and 3 for distance.

For time interval, this individual repeat him/herself every other time. The second event occurs 2 days after the first event. The third event occurs 4 days after the second event; the fourth event occurs 2 days after the third event; the fifth events occurs 4 days after the fourth event; and so forth. In other words, for time interval, the individual repeats him/herself every other interval (i.e., the periodicity is 2). Figure 10.3 illustrates the sequence; the number at each event location is the number of the day that the individual committed the offense (starting at an arbitrary day 1).

Since this fictitious individual is completely predictable, we can easily guess when and where the next event will occur (see table 10.5 above). The direction will, of course, be at 45 degrees from the previous location. Looking at the last known event (event 13), the distance traveled was 4.2426. Thus, we predict that the individual will revert to a move of 1 in the X direction and 1 in the Y direction, or coordinates X=26, Y=26. Finally, for time interval, since the last known time interval was 4 days, then this individual will commit the next event 2 days later, or day number 39.

Example 1: Analysis

The first step is to analyze the sequencing of the events. There are 13 events and 12 intervals. The correlogram produces the following output (table 10.6).

Looking at the unadjusted correlations, it can be seen that time shows an alternating pattern of perfect correlations. The first repeating positive 1.0 correlation is for lag 2, which is the exact periodicity that was specified in the example. This offender repeats the time sequence every other time. Thus, if the individual alternates between committing offenses 2 and 4 days after the last, then knowing the time interval for the last offense, it can be assumed that the next event will repeat the next-to-the-last time interval.

For distance, the highest correlation is for a lag of 3. This offender repeat himself/herself every third time, which is exactly what was programmed into the example. Thus, knowing the location of the last event, it can be assumed that the individual will choose the same distance for the next interval as three earlier. Finally, all lags show a perfect 1.0 correlation for bearing. The lowest one is taken, which is a lag of 1. That is, this individual repeats the direction every single time (i.e., he/she always travels in the same direction). Thus, in summary, the correlogram shows that the individual repeats the time interval every other time, the distance every third time, and the direction every time.

The CWA - Diagnostics routine merely confirms these correlations. The regression equations yield an R^2 of 1.0 (unadjusted) for each of three variables, for the appropriate lag. For example, table 10.7 above shows the regression results for distance for a lag of 3

436

Figure 10.3:
Example of a Predictable Serial Offender: 1
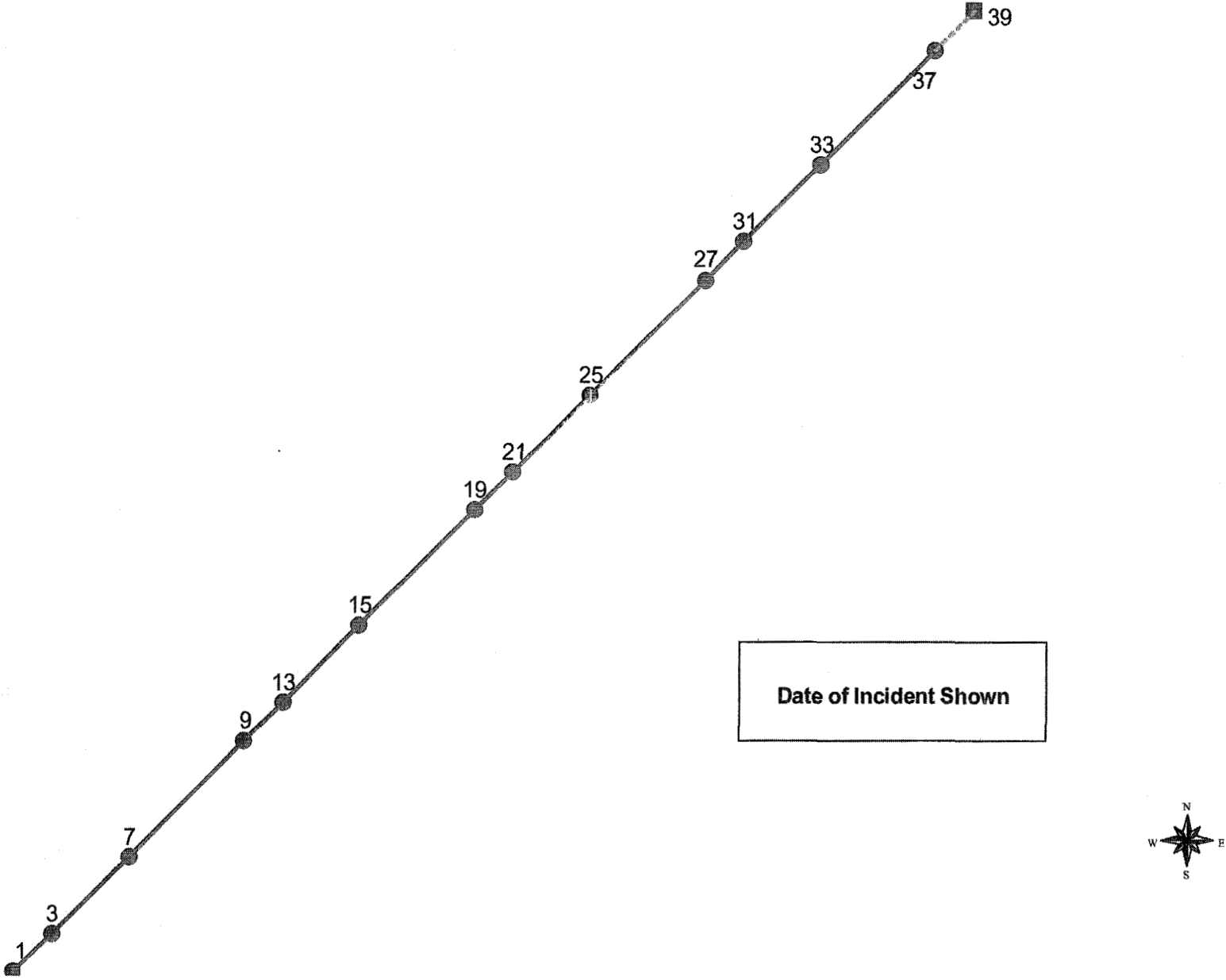## (N=13 Incidents)

Date of Incident Shown

Table 10.6
Correlogram of Predictable Serial Offender: 1

Correlated Walk Analysis -- Correlogram:

Sample size ....: 13
Measurement type ...: Direct
Input units ....: Feet
Time units ....: Days
Distance units ....: Feet
Bearing units ....: Degrees

Adjusted:

Lag	Correlation			Lag	Correlation		
	Time	Distance	Bearing		Time	Distance	Bearing
0	1.00000	1.00000	1.00000	0	1.00000	1.00000	1.00000
1	-1.00000	-0.42105	1.00000	1	-0.90909	-0.38278	0.90909
2	1.00000	-0.56522	1.00000	2	0.81818	-0.46245	0.81818
3	-1.00000	1.00000	1.00000	3	-0.72727	0.72727	0.72727
4	1.00000	-0.38462	1.00000	4	0.63636	-0.24476	0.63636
5	-1.00000	-0.58824	1.00000	5	-0.54545	-0.32086	0.54545
6	1.00000	1.00000	1.00000	6	0.45455	0.45455	0.45455
7	-1.00000	-0.28571	1.00000	7	-0.36364	-0.10390	0.36364

The adjusted correlogram show a similar pattern, though the absolute correlations have been reduced. The best decision would still be for a lag of 2 for time, a lag of 3 for distance, and a lag of 1 for bearing. Figure 10.4 shows a graph of the correlogram. *CrimeStat* has a built-in graph function for the correlogram and adjusted correlogram.

Table 10.7

Regression Results for Serial Offender 1: Distance

| Variable: | distance | | Standard error of estimate: | | 0.00000 |
| Multiple R: | 1.00000 | | Squared multiple R: | | 1.00000 |

	Coefficient	Std Error	t	P(2 Tail)
Constant	0.000000	0.00000	0.00000	0.00000
Coefficient	1.000000	0.00000	0.00000	0.00000

Analysis of Variance

Source	Sum-of-Squares	df	Mean-Square	F-ratio	P
Regression	12.00000	1	12.00000	0.00000	0.00000
Residual	0.00000	8	0.00000		
Total	12.00000	9			

438

Figure 10.4: Correlogram of Serial Offender: 1
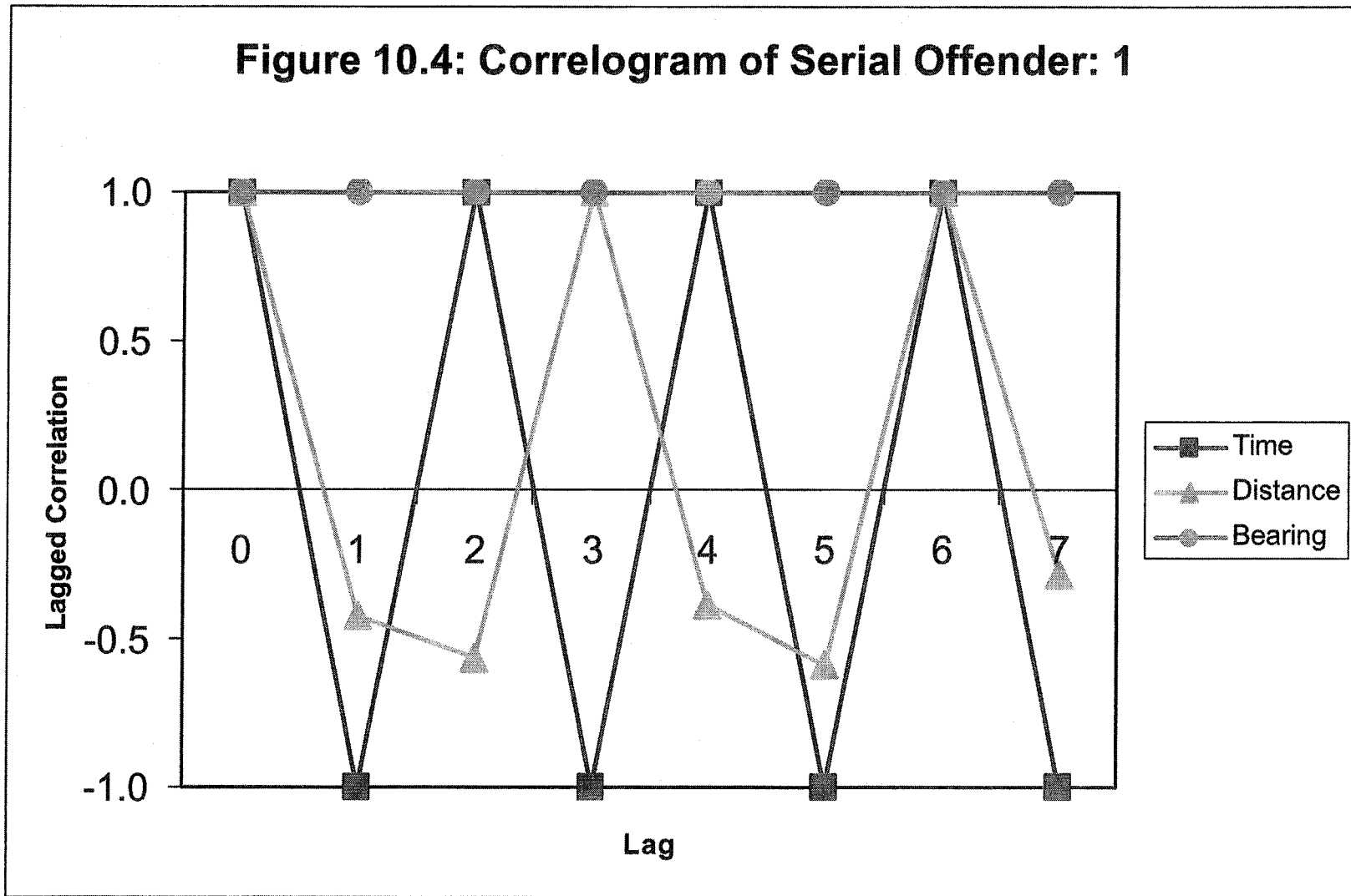
Example 1: Prediction

Finally, for prediction, it is apparent that the best method would be to use a regression equation with lags of 2 for time, 3 for distance, and 1 for bearing. Table 10.8 shows the output. As can be seen, the routine predicts exactly the next time and location. The next event for this completely predictable serial offender will be on day 39 at the location with coordinates X=26, Y=26.

Table 10.8
Predicted Results for Serial Offender 1
Regression Equation with
Lags of 2 for Time, 3 for Distance, 1 for Bearing

Variable	Predicted value	From event	Method	Lag
Time interval	2.00000	13	Regression	2
Distance interval	1.41421	13	Regression	3
Bearing interval	44.99997	13	Regression	1
Predicted time ....:	39.00000			
Predicted X coordinate :	26.00000			
Predicted Y coordinate :	26.00000			

The regression equation is the best model in this case. The other methods produce reasonably close approximations, however. Table 10.9 shows the results of using other methods for prediction. As seen, a model where all three components (time, distance, bearing) were lagged by 1 as well as a model where all three components were lagged by 3 also produces the expected correct answer. The mean interval and median interval methods also produce reasonably close, though not exact, answers. In this particular case, the regression method with the best lags produced the optimal solution.

Example 2: Another Completely Predictable Individual

A second example is also a perfectly predictable individual. This time, the directional component changes. The directional trend is northward, but with changes in angle every third event. The time pattern is completely consistent with subsequent events occurring every two days. Table 10.10 presents the pattern and the logical next event while figure 10.5 displays the pattern

The correlogram reveals that both distance and bearing repeat themselves every third event while the time interval is repeated every time. The regression diagnostics show that there is perfect predictability for time and for distance, and high predictability for bearing (not shown). Finally, a regression model is used for prediction with lags of 1 for

440

Table 10.9
Comparison of Methods for Predictable Serial Offender 1

	EVENT	X	Y	DISTANCE	DAYS	TIME INTERVAL	DIRECTION
Logical Prediction for next event	14	26	26	1.4142	39	2	45
PREDICTION:							
Mean (lag=1)	14	27.0	27.0	2.8	40.0	3.0	45.0
Median (lag=1)	14	27.0	27.0	2.8	41.0	4.0	45.0
Regression:							
Lag=1	14	26.6	26.6	2.3	39.0	2.0	45.0
Lag=2	14	27.0	27.0	2.9	39.0	2.0	45.0
Lag=3	14	26.0	26.0	1.4	39.0	2.0	45.0
Optimal (t=2,d=3,b=1)	14	26.0	26.0	1.4	39.0	2.0	45.0

Table 10.10
Example of a Predictable Serial Offender: 2
(N = 14 incidents)

Event	X	Y	Distance	Days	Time Interval
1	3	1	-	1	-
2	1	3	2.8284	3	2
3	1	5	2.0000	5	2
4	3	7	2.8284	7	2
5	1	9	2.8284	9	2
6	1	11	2.0000	11	2
7	3	13	2.8284	13	2
8	1	15	2.8284	15	2
9	1	17	2.0000	17	2
10	3	19	2.8284	19	2
11	1	21	2.8284	21	2
12	1	23	2.0000	23	2
Logical prediction for next event 13	3	25	2.8284	25	2

441

time, 3 for distance, and 3 for bearing. The model correctly predicts the expected time (days=25) and location (X=3, Y=25). Table 10.11 shows the results.

Methodology for CWA

These two examples illustrate what the CWA routine is doing. There are three steps. First, the sequential pattern is analyzed with the correlogram. This shows which lags have the strongest correlations between lags for time, distance, and bearing separately. Second, the pattern is tested with a regression model. The purpose is to determine how strong a relationship is any particular model. As will be suggested below, if a model is too weak or, conversely, too strong, it most likely will not predict very well. Third, a prediction model is selected. The user can utilize the regression model or use the mean interval or median interval.

Table 10.11

Comparison of Methods for Predictable Serial Offender 2

	EVENT	X	Y	DISTANCE	DAYS	TIME INTERVAL	DIRECTION
Logical Prediction for next event	13	3	25	2.8284	25	2	45
PREDICTION:							
Mean (lag=1)	13	2.2	25.2	2.5	25.0	2.0	28.6
Median (lag=1)	13	3.0	25.0	2.8	25.0	2.0	45.0
Regression:							
Lag=1	13	3.0	25.0	2.8	25.0	2.0	45.0
Lag=2	13	1.9	25.2	2.4	25.2	2.0	22.5
Lag=3	13	3.0	25.0	2.8	25.0	2.0	45.0
Optimal (t=1,d=3,b=3)	13	3.0	25.0	2.8	25.0	2.0	45.0

Example 3: A Real Serial Offender

How well does the CWA routine work with real serial offenders? People are not as predictable as these examples; the examples are algorithmic and people don't work like algorithms. But, to the extent to which there is some predictability in human behavior, the CWA routine can be a useful tool for crime analysis, detection, and apprehension.

442

Figure 10.5:
Example of a Predictable Serial Offender: 2
(N=12 Incidents)
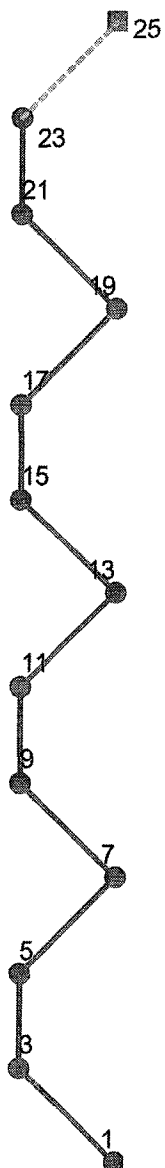
Date of Incident Shown

To illustrate this, a serial offender was identified from a large data set obtained from Baltimore County (see Chapter 9). The individual committed 16 offenses between 1992 and 1997 when he was eventually apprehended. The profile of crimes committed by this individual were quite diverse. There were 11 larceny incidents (shoplifting and bicycle theft), 1 residential burglary, 1 commercial burglary, 2 assaults, and 1 robbery.

To test the model, the first 15 incidents were used to predict the 16th. This allowed the error between the observed and predicted values for time and location to be used for evaluation. Figure 10.6 shows the sequencing of actions of the first 15 incidents committed by this individual, most of which occurred in the eastern part of Baltimore County.

The correlogram revealed a complicated pattern (figure 10.7). The adjusted matrix was used because of the high correlations at higher-order lags. Nevertheless, the optimal lags appeared to be 1 for time, 3 for distance, and 6 for bearing. A regression model was used to test these parameters. Figure 10.6 also shows the predicted location for the next likely location (the red plus sign) and the location where the individual actually committed the 16th event (green triangle). The error in prediction was good. The distance between the actual and predicted locations was 1.8 miles and the error in predicting the time of the next location was 3.9 days. Overall, the model did quite well for this individual.

Event Sequence as an Analogy to a Correlated Walk

Nevertheless, there are problems in the model for this case. First, this is not a true sequence of actions, but a pseudo-sequence. The individual doesn't go from the first event to the second event to the third event, and so forth. A considerable time may elapse between events. Similarly, distance and direction are conceptual only, not real. For example, in figure 10.6, the individual did not actually travel across the inlets of the Chesapeake Bay as the lines indicate. Distance between the events was actually much greater than estimated by the model and direction was more complex. Nevertheless, to the extent to which an individual makes a spatial decision about where to go, implicitly he or she is making a directional and distance decision. In other words, the decision making process may take into account prior locations. In this case, the CWA routines would be useful.

Example 4: A Second Real Serial Offender

A second real example confirms that the method can produce reasonably close predictions. An offender committed 13 crimes, including three incidents of shoplifting, eight incidents of theft from a vehicle, one residential burglary, and one highway robbery. The correlogram showed that a lag of 1 was strongest for time, distance, and bearing (figure 10.8). The R-squares were moderate (0.45 for time; 0.18 for distance; 0.18 for bearing). Using the regression method with a lag of 1 for each component, the likely location of the next event was predicted (Figure 10.9). The error between the predicted event and the actual event was, again, reasonable with a difference in time of 3.3 days and a difference in distance of 2.4 miles.

444

Figure 10.6:
Likely Location for Next Crime:
Serial Offender in Baltimore County
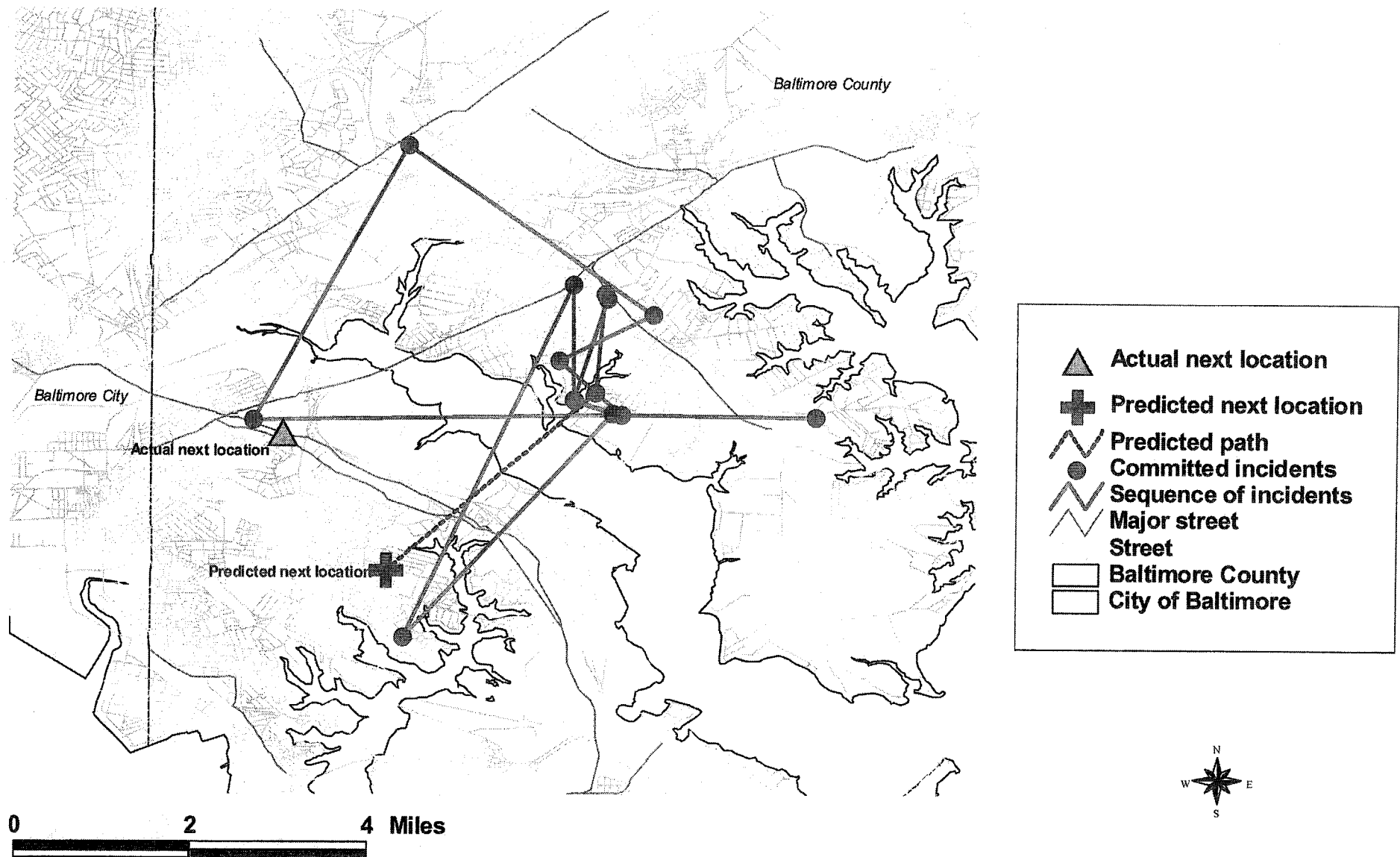## N=16 Incidents

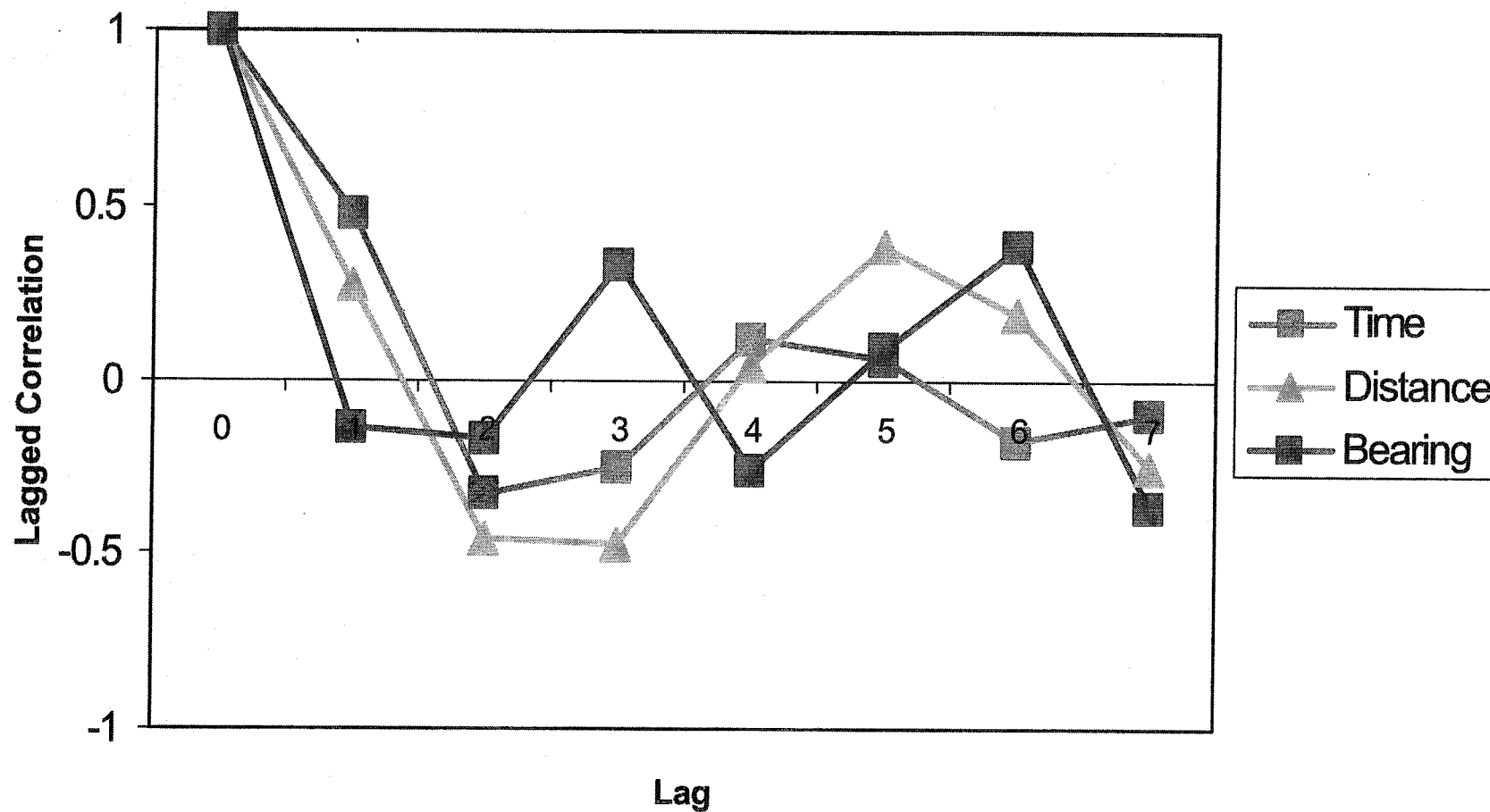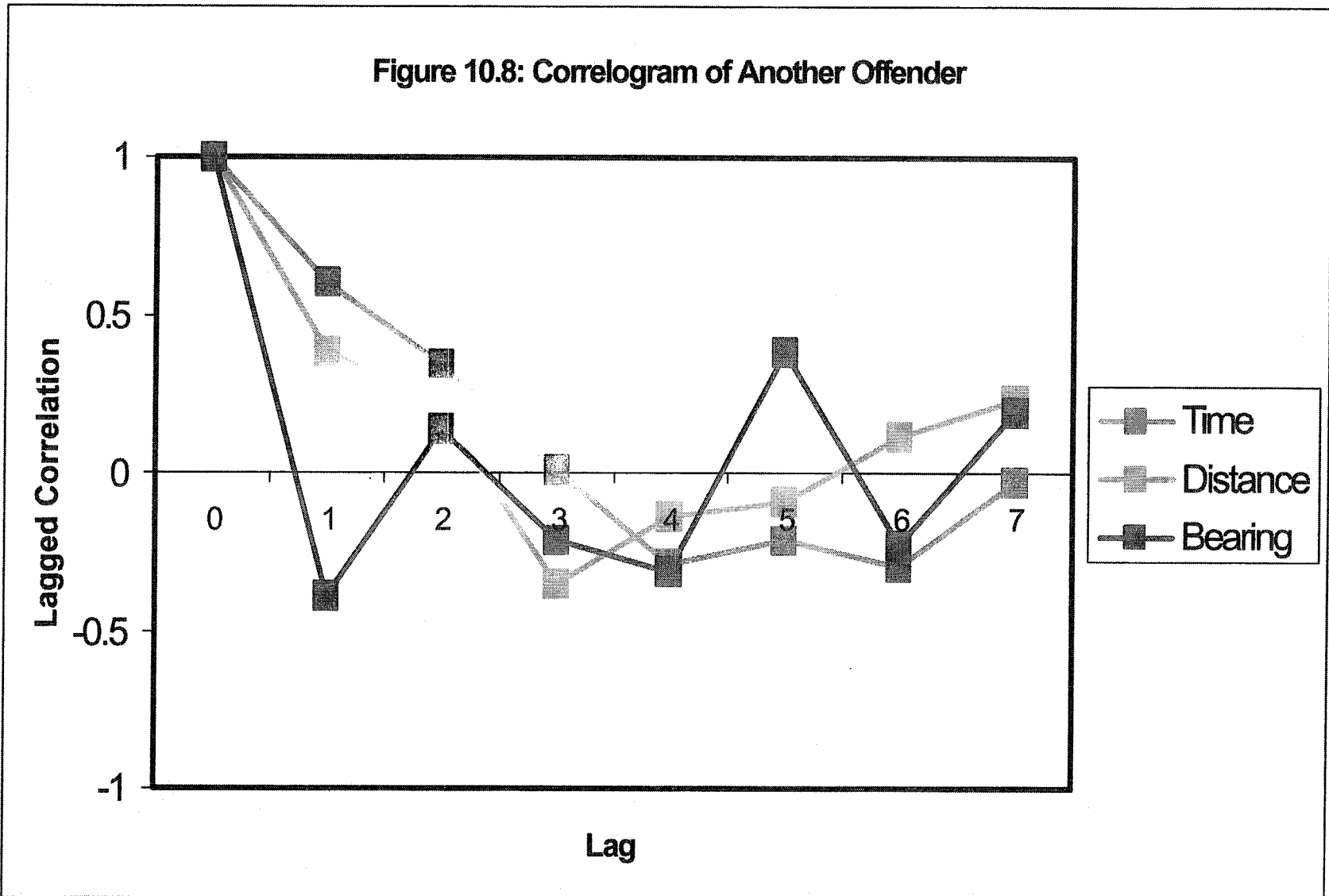Figure 10.7: Correlogram of Actual Offender
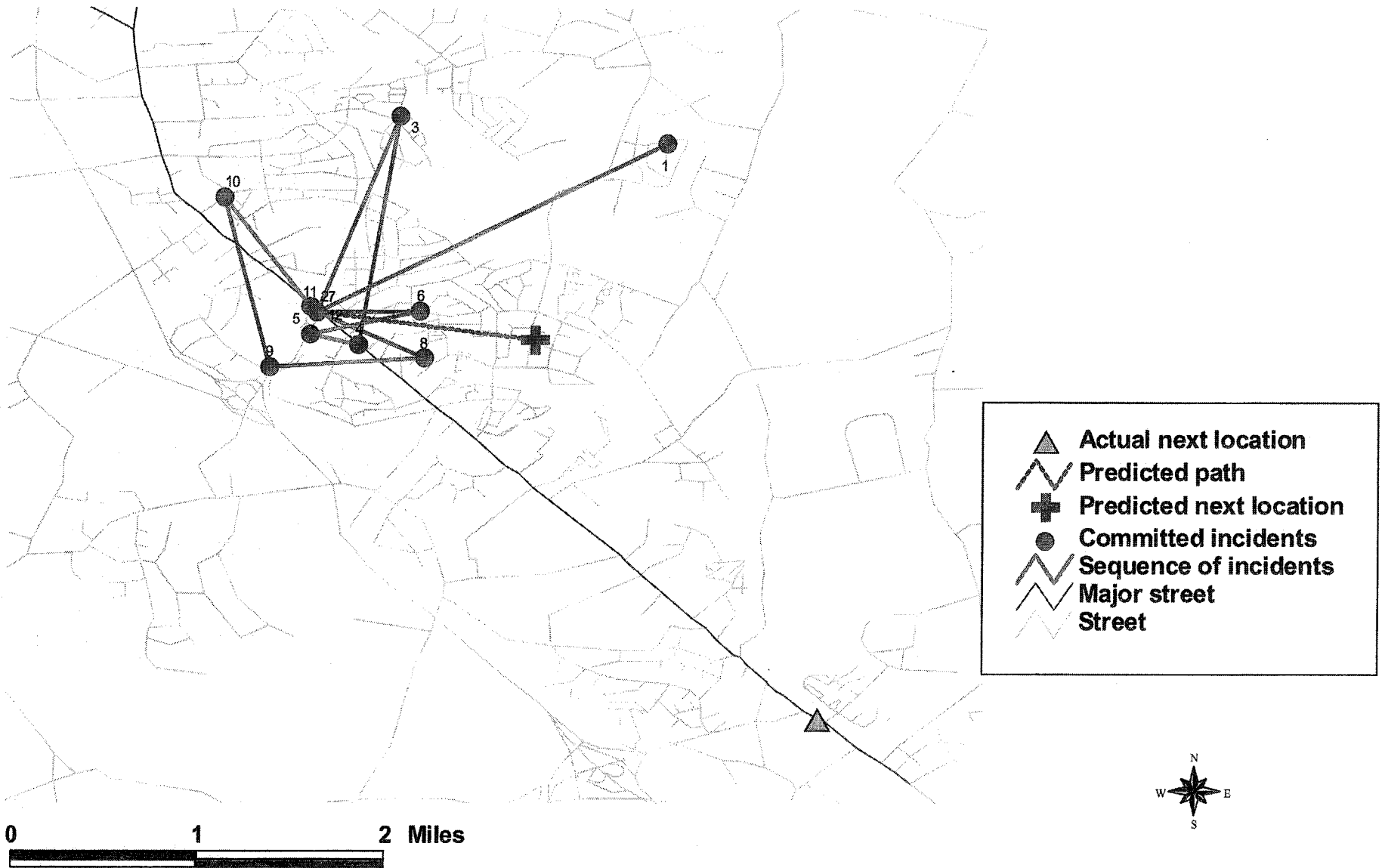
Figure 10.8: Correlogram of Another Offender

Figure 10.9:
Likely Location for Next Crime:
Another Serial Offender in Baltimore County

Legend:
- ▲ Actual next location
- Predicted path
- ✚ Predicted next location
- ● Committed incidents
- Sequence of incidents
- Major street
- Street

0   1   2 Miles

Tracking a Burglary Gang with the Correlated Walk Analysis

Bryan Hill
Glendale Police Department
Glendale, AZ

The space-time analysis tools provided with *CrimeStat II* add an important element to an analyst's review of a tactical prediction effort. Although the method for calculating the Correlated Walk Analysis (CWA) is still more experimental than proven, it allows the analyst to see potential patterns in relation to a suspect's crime travel in terms of time, distance, and direction. In a recent burglary series involving several jurisdictions in our county, the CWA technique was used as part of an aggregate process referred to as the Probability Grid Method. That method combines results from several models to predict the next likely area for a new hit in a crime series. One of the most confusing aspects of these burglaries was the fact that several jurisdictions were involved and the offenders seemed to bounce back and forth from one jurisdiction to the next.

There were also 219 offenses in the series, providing considerable complexity. Because there were so many events, the distances could be anywhere from 0.5 miles to 20 miles, I could never really put my finger on what direction or distance the offender would hit next, but was confident a pattern existed and was likely changing over time. The following map shows the probability grid areas predicted and the CWA points predicted. The triangles shown represent the last four hits. The first hit was near the probability grid prediction in the northern portion of the map; however the subsequent hits were all very close to where the CWA routine predicted they would be. This was also a brand new area for these offenders and was a surprise to the department investigating these incidents. This area was not what was expected based on the SD ellipses and other methods used to predict the next event. The CWA tool requires more testing to determine the accuracy of its predictions, however it may turn out to be a valuable tool in a crime analyst's arsenal.

Burglary Series Prediction of Next Hit
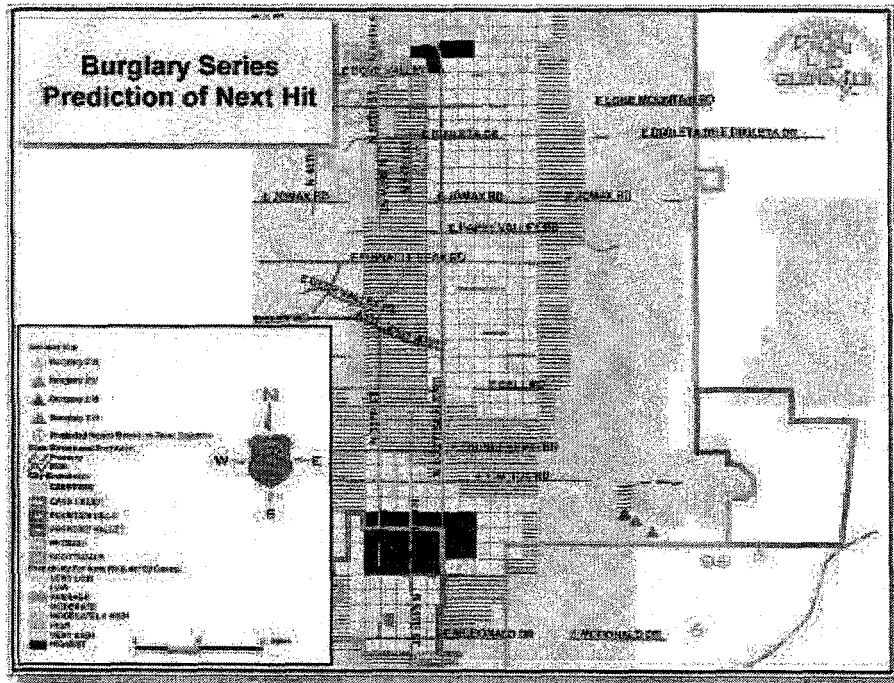
Accuracy of Predictions

However, it's important not to be overly optimistic about the technique. It is always possible to find cases that fit a method very well. The above mentioned cases appear to do that. Unfortunately, the method is not a magic elixir for predicting serial offenders. Like any method, it has error. It is also a fairly new tool in crime analysis so that we don't have a lot of experience with it. The one example of its use was by Helms (1999), who also is cautious about its utility.

Therefore, at this point, I cannot give conclusive results about whether the method is accurate or not and under what conditions it is best used. It will take some experience to know how effective it is for crime analysis.

To explore the accuracy of the method, 50 serial offenders were identified from a large data base of more than 41,000 incidents in Baltimore County between 1993 and 1997 (see Chapter 9). The 50 offenders were identified based on knowing the dates on which they committed crimes, or at least on which they committed crimes for which they were charged and eventually tried. The number of incidents varied from a low of 7 incidents to a high of 38 incidents. An attempt was made to produce balance in the number of incidents, though the actual distribution of cases did reflect the availability of candidates in the data base. For the fifty individuals, the distribution of incidents was 7 (five individuals), 8 (four individuals), 9 (six individuals), 10 (two individuals), 11 (five individuals), 12 (five individuals), 13 (six individuals), 14 (three individuals), 15 (six individuals), 17 (two individuals), and one individual each for 20, 21, 24, 29 and 38 incidents.

To test the CWA model, the last event committed by these individuals was removed so that N-1 events could be used to predict event N. In this way, it is possible to evaluate the accuracy of the method.

Ten methods were compared:

1. The optimal regression method for time with the lag having the strongest relationship being selected;
2. The optimal regression method for location (distance and bearing) where the with the lags for distance and bearing having the strongest relationship being selected;
3. A regression model for time with a lag of 1;
4. A regression model for location with a lag of 1 (for both distance and bearing);
5. The mean interval for time;
6. The mean interval for location (distance and bearing);
7. The median interval for time;
8. The median interval for location (distance and bearing);
9. The mean center of the incidents (for location only); and
10. The center of minimum distance of the incidents (for location only).

The latter two methods were used for reference. In Chapter 9, we saw that center of

450

minimum distance, particularly, was among the best, if not the best, at predicting the origin location of serial offenders. The reason is because this statistic *minimizes the distance* to all incident locations. The mean center was close behind, though not quite as good. As an estimate, the center of minimum distance is a very good index when there is a single origin that is being predicted. On the other hand, where the purpose is to predict the location of a next event, the center of minimum distance and mean center may be less than useful since they will not generally predict the actual next location. They minimize error, but are rarely accurate. For example, in the above mentioned cases (two theoretical and two real), these statistics did not predict accurately the location of the next event. Instead, they identified a point in the middle of the distribution where the sum of the distances to all incident locations was small.

Error Analysis

Each of the models was compared to the actual time and location of the last, removed incident. For time, the error measure was in days (the absolute difference between the actual day and the predicted day). For location, the error measure was in miles (i.e., absolute distance between the actual and predicted location). The results were mixed. Overall, error was moderate. Table 10.12 summarizes the overall error.

Overall, the center of minimum distance and the mean center do produce, as expected, smaller errors for distance than any of the CWA methods; as noted above, locations in the middle of the distribution of incidents will minimize error, but they won't predict accurately the location of a next event nor indicate in which direction it will occur from the last event. On the other hand, the CWA methods are particularly accurate, either. They work very well for a completely predictable offender, as was seen in the examples above, but not necessarily for real offenders.

Among the CWA methods, the mean interval, median interval and the lag 1 regression appears to give better results for time than the optimal regression. Overall, the median interval produces the lowest median error, which is about a month and half. In terms of location, the mean interval and median intervals produce slightly better results than the optimal regression, though the lag 1 regression was just as good.

Comparison of CWA Methods

At this point, it is unclear as when it is best to use this technique. Three variables seem to explain part of the error variation. First, a larger sample size leads to better prediction, as would be expected (Table 10.13).

For time, there is definitely an improvement in predictability with larger sample sizes. Among these methods, the mean interval and lag 1 regression show the smallest error for the largest samples (14 cases). For distance, on the other hand, generally, the error increases with increasing sample size. The one exception is for the optimal regression method where medium-sized samples (10-13 cases) produce the lowest error.

451

Table 10.12

Average and Median Error for CWA Methods
50 Serial Offenders

Method	Average Error	Median Error
Time (days)		
Optimal regression: time	112.2	79.8
Lag 1 regression: time	88.1	70.0
Mean interval: time	89.7	64.9
Median interval: time	91.2	45.5
Distance (miles)		
Optimal regression: location	6.4	5.4
Lag 1 regression: location	5.7	4.2
Mean interval: location	5.8	4.7
Median interval: location	5.3	3.9
Reference Location (miles)		
Mean center	3.3	1.7
Center of minimum distance	3.1	1.2

Variables Affecting Predictability

Long time span

There are a variety of reasons for these strange results, but one reason may be the time span of the events. Some of these offenders committed crimes over a long period, up to five years. Sample size is intrinsically related to the time span ($r=0.55$). The longer the time span that an offender commits crimes, the more incident he/she will perpetrate. With increasing time, the individual's behavior patterns may change.

For those offenders with many incidents, a separate analysis was conducted of the events occurring within the last year. Many of these individuals appeared to have moved their base of operation over time, so the isolation of the most recent events was done in order to produce a clearer behavior pattern. The results, while promising, were not dramatic. Accuracy was improved a little compared to using the full sequence, particularly spatial accuracy. However, even with the last few events, these frequently occurred over a long time period (up to two years). Consequently, the idea of isolating a 'clean' set of events did not materialize, at least with these data. On the other hand, with a data set of only recent events, it may be possible to improve predictability.

452

Strength of predictability

A second variable that appears to have an effect is the strength of predictability, based on the first N-1 cases. For the diagnostics routine, as the overall R-square for the regression equation increases, the regression equation does better. However, with very high R-square coefficients, the error is worse. Table 10.14 shows the relationship.

The lowest error is obtained with moderate R-square coefficients, for both time and distance. This is why one has to be careful with very high lagged correlations in the correlogram and high R-squares in the diagnostics. Unless one is dealing with a perfectly predictable individual (as the two theoretical examples illustrated), high correlations may be a result of a very small sample size, rather than any inherent predictability.

Limitations of the Technique

In short, users should be careful about using the CWA technique. It can be useful for identifying repeating patterns by an offender, but it won't necessarily predict accurately the offender's next actions. There are a variety of reasons for the lack of predictability. First, there may be intermediate events that are unknown. With each of these offenders in the Baltimore County data base, there is always the possibility that the individuals committed more crimes for which they were not charged. The sequential analysis assumes that all the events are known. But this may not be the case.

A simulation on several cases was conducted by removing events and then re-running the correlogram and prediction models. Removing one event did not appreciably alter the relationship, but removing more than one event did. In other words, if there are unknown events, the true sequential behavior pattern of the offender may not be properly identified. Considering that most offenders commit fewer than 10 incidents before they get caught, the statistical effect of missing information may be critical.

A second reason has been alluded to already. In applying the model to crime events, it is not a true sequential model, but a *pseudo-sequential* model since much time may intervene between events. Distance and direction are conceptual in the sense that the individual doesn't directly orient from one event to the other, but returns to his/her living patterns. Thus, what may appear to be a repeating pattern may not be. Here, the issue of sample size is critical. If there are only a few incidents on which to base an analysis, one could see a pattern which actually doesn't exist. One has to be careful about drawing inferences from very small samples.

A third reason is that people are inherently unpredictable. The two algorithmic examples produced excellent results, but few persons are that systematic about their behavior. Therefore, we must be cautious in expecting too much out of the model.

453

Table 10.13

Sample Size and Prediction Error

Sample Size	*	Time				*	Distance				*		
	*	Optimal Regression	Lag 1 Regression	Mean Interval	Median Interval	*	Optimal Regression	Lag 1 Regression	Mean Interval	Median Interval	*	Mean Center	Median Center
6-9	*	143.4	108.5	116.4	120.8	*	7.4	5.2	5.0	4.4	*	2.9	2.4
10-13	*	108.2	86.8	83.4	79.5	*	5.5	6.0	5.7	5.5	*	2.9	3.1
14+	*	79.8	65.1	65.7	71.2	*	6.1	5.9	6.8	6.1	*	4.3	4.1

Table 10.14

Regression Diagnostics and Prediction Error
Comparison of CWA Regression Methods

R-Square	Time (days)		Distance (miles)	
	Optimal Regression	Lag 1 Regression	Optimal Regression	Lag 1 Regression
0 - 0.29	93.7	90.9	6.7	6.3
0.30 - 0.59	89.3	33.8	6.0	5.0
0.60 +	164.3	122.7	6.3	5.2

Conclusion

Nevertheless, the model has utility. First, it can help police identify whether there is a pattern in an offender's behavior. Knowing that there is a pattern can help in planning an arrest strategy. Even if the strategy does not pay off every time, it may improve police effectiveness. In short, the CWA can help a police department analyze the sequential behavior of an offender they are trying to catch. They may be able to anticipate a new event and may be able to warn people who are more likely to be attacked by this individual. If used carefully, the model can be useful for crime analysis and detection.

Second, it can encourage the development of additional predictor tools for individuals. As mentioned above, the center of minimum distance produces a 'best guess' estimate in the sense that it minimizes the distance to the next event. It usually doesn't predict the next event, but it does produce a minimal error. If used in conjunction with the CWA, it may be possible to narrow the search area for the next event.

454

Third, the CWA model can stimulate research into crime prediction. Police are always trying to predict the next event by an offender and will use multiple techniques and a lot of intuition in trying to 'out-guess' an offender. It is hoped that the CWA model will stimulate more research into predicting the sequence of offender behavior as well into how those sequences aggregate into a large spatial pattern. Most of this text has been devoted to analyzing the spatial patterns of a large number of events. The statistics have, perhaps naively, assumed that each of those events were independent. In reality, they aren't since many crimes are committed by the same individuals. In theory, a distribution of crime incidents could be disaggregated into a distribution of *sequences of events* committed by the same offenders, if we had enough information. Understanding how aggregate distributions is a by-product of the behavior of a limited number of individuals is an important research goal that needs to be addressed.

Endnotes for Chapter 10

1. It would be possible to make a one-tailed test with the simulation. For example, if one is only interested in the degree of clustering, one could adopt the 95 percentile as the threshold. An observed Mantel value that was lower than this threshold would be consistent with the null hypothesis.

2. Henderson, Renshaw and Ford (1981) defined the correlated walk as a two-dimensional walk where the sum of the probabilities in four directions along a lattice are:

    $$P = p + q + 2r = 1$$

    where P is the total probability (1), p is the probability of continuing in the same direction, q is the probability of moving in an opposite direction, and r is the probability of moving one unit to the right or to the left. The advantage of this formulation is that the probabilities do not have to be equal (i.e., p could exceed q or r). Nevertheless, the individual steps can be considered a special case of a correlated random walk in the plane (Henderson, 1981).

    The non-lattice two dimensional case can also be considered a recurrent random walk since a step in any direction (not just along a lattice) can be considered the result of two steps, one in the X direction and one in the Y (or, alternatively, a pairing of all steps in the X direction with all steps in the Y direction). Unfortunately, this logic does not apply to more than two dimensions. Such multi-dimensional walks do not have to return to their origin. However, Spitzer (1963) has shown that an independent walk is recurrent if the second moment around the origin is finite.

456

References Used in *CrimeStat* Manual

Aldenderfer, M. and R. Blashfield (1984). *Cluster Analysis.* Sage: Beverly Hills, CA.

Andersson, T. (1897). *Den Inre Omflyttningen.* Norrland: Mälmo.

Anselin, Luc (1995). "Local indicators of spatial association - LISA". *Geographical Analysis.* 27, No. 2 (April), 93-115.

Anselin, Luc. (1992). *SpaceStat: A Program for the Statistical Analysis of Spatial Data.* Santa Barbara, CA: National Center for Geographic Information and Analysis, University of California.

Anselin, Luc and Moss Madden (1990). *New Directions in Regional Analysis.* Belhaven Press: New York.

Aplin, Graeme (1983). *Order-Neighbour Analysis.* Concepts and Techniques in Modern Geography No. 36. Institute of British Geographers, Norwich, England: Geo Books.

Bachi, R. (1957). *Statistical Analysis of Geographical Series.* Central Bureau of Statistics, Kaplan School, Hebrew University: Jerusalem.

Bailey, Trevor C. and Anthony C. Gatrell (1995). *Interactive Spatial Data Analysis.* Longman Scientific & Technical: Burnt Mill, Essex, England.

Ball, G. H. and D. J. Hall (1970). "A clustering technique for summarizing multivariate data". *Behavioral Science*, 12, 153-155.

Barnard, G. A. (1963). "Comment on 'The Spectral Analysis of Point Processes' by M. S. Bartlett", *Journal of the Royal Statistical Society*, Series B, 25, 294.

Beale, E. M. L. (1969). *Cluster Analysis.* Scientific Control Systems: London.

Besag, Julian and James Newell (1991). "The detection of clusters in rare diseases". Journal of the Royal Statistic Society A, 154, Part I, 143-155.

Bezdek, J. C. (1981). *Pattern Recognition with Fuzzy Objective Function Algorithms.* Plenum Press: New York.

Block, Carolyn R. (1994). "STAC hot spot areas: a statistical tool for law enforcement decisions". In *Proceedings of the Workshop on Crime Analysis Through Computer Mapping.* Criminal Justice Information Authority: Chicago, IL.

Block, Richard and Carolyn R. Block (1995). "Space, place and crime: hot spot areas and hot places of liquor-related crime". In John E. Eck and David Weisburd, *Crime Places in Crime Theory*, Rutgers Crime Prevention Studies Series, Criminal Justice Press: Newark.

457

Block, Carolyn R. and Lynn A. Green (1994). *The GeoArchive Handbook: A Guide for Developing a Geographic Database an Information Foundation for Community Policing.* Illinois Criminal Justice Information Authority: Chicago, IL.

Blumin, D. (1973). *Victims: A Study of Crime in a Boston Housing Project.* City of Boston, Mayor's Safe Street Act, Advisory Committee: Boston.

Boggs, S. L. (1965). "Urban crime patterns", *American Sociological Review*, 30, 899-908.

Borland.Com (1998). *dBase IV 2.0.* Inprise Corporation: Scotts Valley, CA.

Bossard, Earl G. (1993). "RETAIL: Retail trade spatial interaction". In Richard E. Klosterman, Richard K. Brail and Earl G. Bossard, *Spreadsheet Models for Urban and Regional Analysis.* Center for Urban Policy Research, Rutgers University: New Brunswick, NJ, 419-448.

Bowman, Adrian W. and Adelchi Azzalini (1997). *Applied Smoothing Techniques for Data Analysis: The Kernel Approach with S-Plus Illustrations.* Oxford Science Publications, Oxford University Press: Oxford, England.

Brantingham, Patricia L. and Paul J. Brantingham (1981). "Notes on the geometry of crime". In Paul J. Brantingham and Patricia L. Brantingham, *Environmental Criminology.* Waveland Press, Inc.: Prospect Heights, IL, 27-54.

Bright, M. L. and D. S. Thomas (1941). "Interstate migration and intervening opportunities", *American Sociological Review*, 6, 773-783.

Burgess, E. W. (1925). "The growth of the city: an introduction to a research project". In R. E. Park, E. W. Burgess and R. D. Mackensie (eds), *The City.* University of Chicago Press: Chicago, 47-62.

Burt, James E. and Gerald M. Barber (1996). *Elementary Statistics for Geographers* (second edition). The Guilford Press: New York.

Çan, Ayşe and Issac Megbolugbe (1996). "The geography of underserved mortgage markets". Paper presented at the American Real Estate and Urban Economics Association meeting. May.

Canter, D., T. Coffey, and M. Huntley (2000). "Predicting serial killers' home base using a decision support system". *Journal of Quantitative Criminology.* In press.

Canter, D. and S. Tagg (1975). "Distance estimation in cities", *Environment and Behaviour*, 7, 59-80.

Canter, D. and P. Larkin (1993). "The environmental range of serial rapists", *Journal of Environmental Psychology*, 13, 63-69.

458

Canter, D. and A. Gregory (1994). "Identifying the residential location of rapists", *Journal of the Forensic Science Society*, 34 (3), 169-175.

Canter, D. (1994). *Criminal Shadows: Inside the Mind of the Serial Killer*. Harper Collins Publishers: London.

Canter, D. (1999). "Modelling the home location of serial offenders". Paper presented at the 3rd Annual International Crime Mapping Research Conference, Orlando, December.

Capone, D. L. and W. W. Nichols Jr. (1975). "Crime and distance: an analysis of offender behaviour in space", *Proceedings, Association of American Geographers*, 45-49.

Chaitin, Gregory (1990). *Information, Randomness and Incompleteness* (second edition). World Scientific: Singapore.

Chen, A. & Renshaw, E. (1994) The general correlated random walk. *Journal of Applied Probability*, 31, 869-884.

Chen, A. & Renshaw, E. (1992). "The Gillis-Domb-Fisher correlated random walk." *Journal of Applied Probability*, 29, 792-813.

Chrisman, Nicholas (1997) *Exploring Geographic Information Systems*. John Wiley and Sons, Inc.: New York.

Clark, P. J. and F. C. Evans (1954). "Distance to nearest neighbor as a measure of spatial relationships in populations". *Ecology*, 35, 445-453.

Cleveland, William S., Eric Grosse, and William M. Shyu (1993). "Local regression models". In John M. Chambers and Trevor J. Hastie, *Statistical Models in S*. Chapman & Hall: London.

Cliff, Andrew D. and Peter Haggett (1988). *Atlas of Disease Distributions*. Blackwell Reference: Oxford.

Cliff, A. and J. Ord (1973). *Spatial Autocorrelation*. Pion: London.

Cohen, L.E. and Felson, M. (1979) "Social change and crime rate trends: a routine activity approach", *American Sociological Review*, 44: 588-608.

Cohen, Lawrence E. 1981 "Modeling crime trends: a criminal opportunity perspective", *Journal of Research in Crime and Delinquency*, 18:138-163.

Cole, A. J. and D. Wishart (1970). "An improved algorithm for the Jardine-Sibson method of generating overlapping clusters". *Comparative Journal*, 13, 156-163.

459

Committee on Map Projections (1986). *Which Map is Best*, American Congress on Surveying and Mapping, Falls Church, VA., 1986.

Cressie, Noel (1991). *Statistics for Spatial Data*. New York: J. Wiley & Sons, Inc.

Cromley, Robert G. (1992). *Digital Cartography*. Prentice Hall: Englewood Cliffs, NJ.

Curtis, L. A. (1974). *Criminal Violence*. Lexington Books: Lexington, MA.

D'andrade,R. 1978, "U-Statistic Hierarchical Clustering" *Psychometrika*, 4,58-67.

Dwass, M (1957). "Modified randomization tests for nonparametric hypotheses". *Annals of Mathematical Statistics*, 28, 181-187.

Ebdon, David (1988). *Statistics in Geography* (second edition with corrections). Blackwell: Oxford.

Eldridge, J. D. and J. P. Jones (1991). "Warped space: a geography of distance decay", *Professional Geographer*, 43 (4), 500-511.

ESRI (1998a). *ArcView GIS 3.1*. Environmental Systems Research Institute: Redland, CA.

ESRI (1998b). *ArcInfo 7.2.1*. Environmental Systems Research Institute: Redland, CA.

ESRI (1998c). *Atlas*GIS 4.0*. Environmental Systems Research Institute: Redland, CA.

ESRI (1997). *ArcView Spatial Analyst*. Environmental Systems Research Institute: Redland, CA.

Everett, Brian (1974). *Cluster Analysis*. Heinemann Educational books, Ltd: London.

Farewell, Daniel (1999). "Specifying the bandwidth function for the kernel density estimator". http://www.iph.cam.ac.uk/bugs/documentation/coda03/node44.html.

Field, Brian and Bryan MacGregor (1987). *Forecasting Techniques for Urban and Regional Planning*. UCL Press, Ltd: London.

Fisher, W. (1958). "On grouping for maximum homogeneity." *Journal of the American Statistical Association*. **53**, 789-798.

Foot, D. (1981). *Operational Urban Models*. Methuen: London.

Fotheringham, A. S. and M. E. O'Kelly (1989). *Spatial Interaction Models: Formulations and Applications*. Kluwer Academic Publishers: Boston.

460

Fotheringham, A. S. and M. E. O'Kelly (1989). *Spatial Interaction Models: Formulations and Applications.* Kluwer Academic Publishers: Boston.

Friedman, H. P. and J. Rubin (1967). "On some invariant criteria for grouping data", *Journal of the American Statistical Association*, **62**, 1159-1178.

Furfey, P. H. (1927). "A note on Lefever's 'Standard deviational ellipse'". *American Journal of Sociology.* XXIII, 94-98.

Gaile, Gary L. and James E. Burt (1980). *Directional Statistics.* Concepts and Techniques in Modern Geography No. 25. Institute of British Geographers, Norwich, England: Geo Books.

Geary, R. (1954). "The contiguity ratio and statistical mapping". *The Incorporated Statistician*, <u>5</u>, 115-145.

Gersho, A. and R. Gray (1992). *Vector Quantization and Signal Compression.* Kluwer Academic Publishers: Dordrecht, Netherlands.

Getis, Arthur (1991). "Spatial interaction and spatial auto-correlation: a cross-product approach". *Environment and Planning A*, 23, 1269-1277.

Getis, Arthur and J. Keith Ord (1996). "Local spatial statistics: an overview". In Paul Longley and Michael Batty (eds), *Spatial Analysis: Modelling in a GIS Environment.* GeoInformation International: Cambridge, England, 261-277.

Getis, Arthur and Barry Boots (1978). *Models of Spatial Processes: An Approach to the Study of Point, Line and Area Patterns.* London: Cambridge University Press.

Golden Software. 1994. *Surfer® for Windows (Ver. 6).* Golden Software, Inc.: Golden, CO.

Gowers, J. C. (1967). "A comparison of some methods of cluster analysis". *Biometrics*, 23, 623-628.

Greenhood, David (1964). *Mapping.* The University of Chicago Press: Chicago.

Griffith, Danel A. (1987). *Spatial Autocorrelation: A Primer.* Resource Publications in Geography, The Association of American Geographers: Washington, DC.

Grubesic, Tony H. and Alan T. Murray (2001). "Detecting hot spots using cluster analysis and GIS". Paper presented at Annual Conference of the Crime Mapping Research Center, Dallas, TX. http://www.ojp.usdoj.gov/cmrc.

Hägerstrand, T. (1957). "Migration and area: survey of a sample of Swedish migration fields and hypothetical considerations on their genesis". *Lund Studies in Geography, Series B, Human Geography*, 4, 3-19.

Haggett, Peter and Edward Arnold (1965). *Locational Analysis in Human Geography* (1st edition). Edward Arnold: London.

Haggett, Peter, Andrew D. Cliff, and Allan Frey (1977). *Locational Analysis in Human Geography* (2nd edition). Edward Arnold: London.

Hammond, Robert, and Patrick McCullagh (1978). *Quantitative Techniques in Geography: An Introduction*. Second Edition. Clarendon Press: Oxford, England.

Härdle, Wolfgang (1991). *Smoothing Techniques with Implementation in S*. Springer-Verlag: New York.

Harries, Keith (1980). *Crime and the Environment*. Charles C. Thomas Press: Springfield.

Harries, Keith and Phil Canter (1998). "The use of GPS in geocoding crime incidents". Personal Communication.

Hartigan, J. A. (1975). *Clustering Algorithms*. John Wiley & Sons, Inc.: New York.

Henderson, Robin (1981). *The Structural Root Systems of Sitka Spruce and Related Stochastic Processes*. PhD Thesis, University of Edinburgh: Edinburgh.

Henderson, Robin, Eric Renshaw, and David Ford (1983). "A note on the recurrence of a correlated random walk". Journal of Applied Probability, 20, 696-699.

Henderson, Robin, Eric Renshaw, and David Ford (1984). "A correlated random walk model for two-dimensional diffusion". Journal of Applied Probability, 21, 233-246.

Henderson, R., E. D. Ford, E. Renshaw, and J. D. Deans (1983). "Morphology of the structural root system of Sitka Spruce 1. Analysis and Quantitative Description". *Forestry*, 56 (2), 121-135.

Hodge, S. and D. Canter (2000). "Spatial patterns of serial murder: an analysis of disposal site location choice". *Journal of Environmental Psychology*. In press.

Huff, D. L. (1963). "A probabilistic analysis of shopping center trade areas". *Land Economics*, 39, 81-90.

Hultquist, J., L. Brown and J. Holmes (1971). "Centro: a program for centrographic measures". Discussion paper no. 21, Department of Geography, Ohio State University: Columbus, OH.

Huxhold, William E. (1991). *An Introduction to Geographic Information Systems*. Oxford University Press: Oxford, New York, 147-184.

Insightful Corporation (2001). *S-PLUS 6.0 Professional for Windows*. Insightful Corporation: Seattle, WA.

Isard, Walter (1979). *Location and Space-Economy: A General Theory Relating to Industrial Location, Market Areas, Land Use, Trade, and Urban Structure* (originally published 1956). Program in Urban and Regional Studies, Cornell University: Ithaca, NY.

Isard, Walter (1960). *Methods in Regional Analysis*. John Wiley and Sons: New York.

Isbel, E. C. (1944). "Internal migration in Sweden and intervening opportunities", *American Sociological Review*, 9, 627-639.

Jardine, N. and R. Sibson (1968). "The construction of hierarchic and non-hierarchic classifications". *Comparative Journal*, 11, 117-184.

Jefferis, Eric (1998). "A multi-method exploration of crime hot spots". Crime Mapping Research Center, National Institute of Justice: Washington, DC.
http://www.ojp.usdoj.gov/cmrc/whatsnew/hotspot/intro.pdf

Johnson,S.C. 1967, "Hierarchical Clustering Schemes" *Psychometrika*, 2,241-254

Jones, K. S. and D. M. Jackson (1967). "Current approaches to classification and clump finding at the Cambridge Language Research Unit". *Comparative Journal*, 10, 29-37.

Kafadar, K. (1996). "Smoothing geographical data, particularly rates of disease". *Statistics in Medicine* 15(23), 2539-2560.

Kaluzny, Stephen P., Silvia C. Vega, Tamre P. Cardoso, and Alice A. Shelly (1998). *S+ Spatial Stats: User Manual for Windows and Unix*. Springer: New York.

Kanji, Gopal K. (1993). *100 Statistical Tests*. Sage Publications: Thousand Oaks, CA.

Kelsall, J. E. and P. J. Diggle (1995a). "Kernel estimation of relative risk", *Bernoulli*, 1, 3-16.

Kelsall, J. E. and P. J. Diggle (1995b). "Non-parametric estimation of spatial variation in relative risk". *Statistical Medicine*, 14, 2335-2342.

Kim, Karl E. and Michael Parke (1996). The use of GPS and GIS in traffic safety. Report to Motor Vehicle Safety Office, State of Hawaii Department of Transportation: Honolulu.

Kind, Stuart S. (1987). "Navigational ideas and the Yorkshire Ripper investigation". *Journal of Navigation*, 40 (3), 385-393.

King, B. F. (1967). "Step wise clustering procedures". *Journal of the American Statistical Association*. 62, 86-101.

Knox, E. G. (1988). "Detection of clusters". In P. Elliott (ed), *Methodology of Enquiries into Disease Clustering*, London School of Hygiene and Tropical Medicine: London.

Knox, E. G. (1964). "The detection of space-time interactions". *Applied Statistics*, 13, 25-29.

Knox, E. G. (1963). "Detection of low intensity epidemicity: application in cleft lip and palate". *British Journal of Preventive and Social Medicine*, 18, 17-24.

Krueckeberg, D. A. and A. L. Silvers (1974). *Urban Planning Analysis: Methods and Models*. John Wiley and Sons: New York.

Kuhn, H.W. and R. E. Kuenne (1962). "An efficient algorithm for the numerical solution of the generalized Weber problem in spatial economics", *Journal of Regional Science* 4, 21-33.

Kulldorff, Martin (1997). "A spatial scan statistic", *Communications in Statistics - Theory and Methods*, 26, 1481-1496.

Kulldorff, M. and G. Williams (1997). *SaTScan v 1.0: Software for the Space and Space-Time Scan Statistics*, Bethesda, MD: National Cancer Institute.

Kulldorff, Martin and Nevelle Nagarwalla (1995). "Spatial disease clusters: Detection and inference", *Statistics in Medicine*, 14, 799-810.

Langbein, Laura Irwin and Allan J. Lichtman (1978). *Ecological Inference*. Sage University Paper series on Quantitative Applications in the Social Sciences, series no. 07-010. Beverly Hills and London: Sage Publications.

Langworthy, Robert H. and Eric Jefferis (1998). "The utility of standard deviational ellipses for project evaluation". Discussion paper, National Institute of Justice: Washington, DC.

LeBeau, James L. (1987a). "The journey to rape: geographic distance and the rapist's method of approaching the victim", *Journal of Police Science and Administration*, 15, 129-136.

LeBeau, James L. (1987b). "The methods and measures of centrography and the spatial dynamics of rape", *Journal of Quantitative Criminology*, 3 (2), 125-141.

Lefever, D. (1926). "Measuring geographic concentration by means of the standard deviational ellipse". *American Journal of Sociology*, 32(1): 88-94.

Levine, Ned (1999). "The effects of local growth management on regional housing production and population redistribution in California", *Urban Studies*, 36(12), 2047-2068.

464

Levine, Ned (1996). "Spatial statistics and GIS: software tools to quantify spatial patterns". *Journal of the American Planning Association*. 62 (3), 381-392.

Levine, Ned and Karl E. Kim (1999). "The spatial location of motor vehicle accidents: A methodology for geocoding intersections". *Computers, Environment, and Urban Systems*. 22 (6), 557-576.

Levine, Levine, Karl E. Kim, and Lawrence H. Nitz (1995a). "Spatial analysis of Honolulu motor vehicle crashes: I. Spatial patterns". *Accident Analysis & Prevention*, 27(5), 663-674.

Levine, Ned, Karl E. Kim, and Lawrence H. Nitz (1995b). "Spatial analysis of Honolulu motor vehicle crashes: II. Generators of crashes". *Accident Analysis & Prevention*, 27(5), 675-685.

Levine, Ned and Martin Wachs (1986a). "Bus Crime in Los Angeles: I - Measuring The Incidence". *Transportation Research*. 20 (4), 273-284.

Levine, Ned and Martin Wachs (1986b). "Bus Crime in Los Angeles: II - Victims and Public Impact". *Transportation Research*. 20 (4), 285-293.

Levine, Ned, Martin Wachs and Elham Shirazi (1986). "Crime at Bus Stops: A Study of Environmental Factors". *Journal of Architectural and Planning Research*. 3 (4), 339-361.

Los Angeles Times (1998). *Eye on the Sky*. Business section, July 20.

Lottier, S. (1938). "Distribution of criminal offences in metropolitan regions", *Journal of Criminal Law, Criminology, and Police Science*, 29, 37-50.

MacQueen, J. (1967). "Some methods for classification and analysis of multivariate observations". *5th Berkeley Symposium on Mathematics, Statistics and Probability*. Vol 1, 281-298.

McBratney, A. B. and J. J. deBruijter (1992). "A continuum approach to soil classification by modified fuzzy k-means with extragrades", *Journal of Soil Science*, 43, 159-175.

McDonnell, Porter W. Jr. (1979). *Introduction to Map Projections*. New York: Marcel Dekker, Inc.

McQuitty, L. L. (1960). "Hierarchical syndrome analysis". *Educational and Psychological Measurement*, 20, 293-304.

Maling, D. H. (1973). *Coordinate Systems and Map Projections* (1973). George Philip and Sons, London.

Malkiel, Burton G. (1999). *A Random Walk Down Wall Street* (revised edition). W. W. Norton & Company: New York.

465

Maltz, Michael D., Andrew C. Gordon, and Warren Friedman (1989). *Mapping Crime in Its Community Setting: A Study of Event Geography Analysis.*

Mantel, Nathan (1967). "The detection of disease clustering and a generalized regression approach". *Cancer Research*, 27, 209-220.

Mantel, N. and J. C. Bailar (1970). "A class of permutational and multinomial test arising in epidemiological research", *Biometrics*, 26, 687-700.

MapInfo (1998). *MapInfo Professional 5.0.1.* MapInfo Corporation: Troy, NY.

Mardia, K.V. (1972). *Statistics of Directional Data.* Academic Press: New York.

Mather, A. S. (1986). *Land Use.* John Wiley & Sons: New York.

Microsoft (1999). *Welcome to the ODBC Section of the Microsoft Universal Data Access Web Site.* Microsoft: Redmond, WA. http://www.microsoft.com/data/obdc.

Microsoft (1998a). *Windows NT Workstation 4.0.* Microsoft: Redmond, WA.

Microsoft (1998b). *Windows NT Server 4.0.* Microsoft: Redmond, WA.

Microsoft (1998c). *Windows 98.* Microsoft: Redmond, WA.

Microsoft (1995). *Windows 95.* Microsoft: Redmond, WA.

Moran, P. A. P. The interpretation of statistical maps. *Journal of the Royal Statistical Society B*, 10, 1948; 243-251.

Moran, P. A. P. Notes on continuous stochastic phenomena. *Biometrika*, 37, 1950; 17-23.

Murray, A.T. and T.H. Grubesic. 2002. "Identifying Non-hierarchical Clusters." *International Journal of Industrial Engineering.* In press.

Needham, R. M. (1967). "Automatic classification in linguistics". *The Statistician*, 17, 45-54.

Neft, David Samuel (1962). *Statistical Analysis for Areal Distributions.* Ph.D. dissertation, Columbia University: New York.

Oppenheim, Norbert (1980). *Applied Models in Urban and Regional Analysis.* Prentice-Hall, Inc.: Englewood Cliffs, NJ.

Openshaw, S. A. W. Craft, M. Charlton, and J. M. Birch (1988). "Invetigation of leukemia clusters by use of a geographical analysis machine", *Lancet*, 1, 272-273.

466

Openshaw, S. A., M. Charlton, C. Wymer, and A. W. Craft (1987). "A mark 1 analysis machine for the automated analysis of point data sets", *International Journal of Geographical Information Systems*, 1, 335-358.

Parzen, E. (1962). "On the estimation of a probability density and mode". *Annuals of Mathematical Statistics*, 33, 1065-1076.

Phillip, P.D. (1980) "Characteristics and typology of the journey to crime." In D.E. Georges-Abeyie and K.D. Harries (eds), *Crime: A Spatial Perspective*, Columbia Univ. Press: New York, 156-166.

Pyle, G. F. (1974). *The Spatial Dynamics of Crime.* Department of Geography Research Paper No. 159, University of Chicago: Chicago.

Ravenstein, E. G. (1885). "The laws of migration". *Journal of the Royal Statistical Society.* 48.

Reilly, W. J. (1929). "Methods for the study of retail relationships". *University of Texas Bulletin*, 2944.

Rengert, G., A. R. Piquero, and P. R. Jones (1999). "Distance decay re-examined", *Criminology*, 37 (2), 427-445.

Rengert, George F. "Comparing cognitive hot spots to crime hot spots". In Carolyn Rebecca Block, Margaret Dabdoub and Suzanne Fregly, *Crime Analysis Through Computer Mapping.* Police Executive Research Forum: Washington, DC. 1995. 33-47.

Rengert, George F (1981). "Burglary in Philadelphia: a critique of the opportunity structure model". In Paul J. Brantingham and Patricia L. Brantingham, *Environmental Criminology.* Waveland Press, Inc.: Prospect Heights, IL, 189-202.

Rengert, George F. (1975). "Some effects of being female on criminal spatial behavior". *The Pennsylvania Geographer*, 13 (2), 10-18.

Renshaw, Eric (1985). "Computer simulation of sitka spruce: spatial branching models for canopy growth and root structure". Journal of Mathematics Applied in Medicine and Biology, 2, 183-200.

Repetto, T. A. (1974). *Residential Crime.* Ballinger: Cambridge, MA.

Rhodes, William M. and Catherine Conly (1981). "Crime and mobility: an empirical study". In Brantingham, Paul J. and Patricia L. Brantingham, *Environmental Criminology.* Waveland Press, Inc.: Prospect Heights, IL, 167-188.

Ripley, Brian D (1981). *Spatial Statistics.* John Wiley & Sons: New York.

Ripley, Brian D. (1976). "The second-order analysis of stationary point processes". *Journal of Applied Probability* 13: 255-66.

Robinson, A. H., R. D. Sale, J. L. Morrison and P. C. Muehrcke (1984). *Elements of Cartography* (5th edition). J. Wiley and Sons: New York.

Rosenblatt, M. (1956). "Remarks on some non-parametric estimates of a density function". *Annuals of Mathematical Statistics*, 27, 832-837.

Rossmo, D. Kim (1993a). "Multivariate spatial profiles as a tool in crime investigation". In Carolyn Rebecca Block and Margaret Dabdoub (eds), *Workshop on Crime Analysis Through Computer Mapping: Proceedings*. Illinois Criminal Justice Information Authority and Loyola University Sociology Department: Chicago. (Library of Congress HV7936.C88 W67 1993).

Rossmo, D. Kim (1993b). "Target patterns of serial murderers: a methodological model". *American Journal of Criminal Justice*, 17, 1-21.

Rossmo, D. Kim (1995). "Overview: multivariate spatial profiles as a tool in crime investigation". In Carolyn Rebecca Block, Margaret Dabdoub and Suzanne Fregly, *Crime Analysis Through Computer Mapping*. Police Executive Research Forum: Washington, DC. 65-97.

Rossmo, D. Kim (1997). "Geographic profiling". In Janet L. Jackson and Debra A. Bekerian, *Offender Profiling: Theory, Research and Practice*. John Wiley and Sons: Chicester, 159-175.

Rushton, Gerard (1979). *Optimal Location of Facilities*. COMPress: Wentworth, NH. SPSS, Inc. (1999). *SPSS 9.0 for Windows*. SPSS, Inc.: Chicago.

SAS Institute Inc. (1998). *Statistical Analysis System, Version 7*. Cary, NC.

Scott, D. W. (1992). *Multivariate Density Estimation: Theory, Practice, and Visualization*. John Wiley and Sons: New York.

Shaw, C. R. (1929). *Delinquency Areas*. University of Chicago Press: Chicago.

Sherman, Lawrence W. and David Weisburd (1995). "General deterrent effects of police patrol in crime "hot spots": a randomized controlled trial". *Justice Quarterly*. 12, 625-648.

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

Smith, T. S. (1976). "Inverse distance variations for the flow of crime in urban areas". *Social Forces*, 25(4), 804-815.

468

Sneath, P. H. A. (1957). "The application of computers to taxonomy". *Journal of General Microbiology*, 17, 201-226.

Snyder, John P. (1987). *Map Projections - A Working Manual.* U.S. Geological Survey Professional Paper 1395. U. S. Government Printing Office: Washington, DC.

Snyder, John P. and Philip M. Voxland (1989). *An Album of Map Projections.* U.S. Geological Survey Professional Paper 1453. U. S. Government Printing Office: Washington, DC.

Sokal, R. R. and P. H. A. Sneath (1963). *Principles of Numerical Taxonomy.* W. H. Freeman and Co.: San Francisco.

Sokal, R. R. and C. D. Michener (1958). "A statistical method for evaluating systematic relationships". *University of Kansas Science Bulletin*, 38, 1409-1438.

Spitzer, Frank (1976). *Principles of Random Walk* (second edition). Springer: New York.

Stephenson, L. (1980). "Centrographic analysis of crime". In D. George-Abeyie and K. Harries (eds), *Crime, A Spatial Perspective*, Columbia University Press: New York.

Stewart, J. Q. (1950). "The development of social physics". *American Journal of Physics*, 18, 239-53.

Stopher, Peter R. and Arnim H. Meyburg (1975). *Urban Transportation Modeling and Planning.* Lexington, MA: Lexington Books.

Stouffer, S. A. (1940). "Intervening opportunities: a theory relating mobility and distance". *American Sociological Review*, 5, 845-67.

Systat, Inc. (2000). *Systat 10: Statistics I.* SPSS, Inc.: Chicago.

Talbot T. O., Kulldorff M., Forand S. P., Haley V. B. (2000). "Evaluation of spatial filters to create smoothed maps of health data". *Statistics in Medicine*, 19, 2399-2408.

Taylor, Peter James (1970). *Interaction and Distance: An Investigation into Distance Decay Functions and a Study of Migration at a Microscale.* PhD thesis, University of Liverpool: Liverpool.

Thompson, H. R. (1956). "Distribution of distance to nth neighbour in a population of randomly distributed individuals". *Ecology*, 37, 391-394.

Thorndike, R. L. (1953). "Who belongs in a family?". *Psychometrika*, 18, 267-276.

Thrasher, F. M. (1927). *The Gang,* University of Chicago Press: Chicago.

Thünen, J. H. Von (1826). *Der Isolierte Staat in Beziehung auf Landwirtshaft und Nationalökonomie* (The Isolated State in Relation to Agriculture). Hamburg.

Turnbull, B. W., E. J. Iwano, W. S. Burnett, H. L. Howe, and L. C. Clark (1990). "Monitoring for clusters of disease: application to leukemia incidence in upstate New York", *American Journal of Epidemiology*, 132, S136-S143.

Turner, S. (1969). "Delinquency and distance". In M. E. Wolfgang and T. Sellin (eds), *Delinquency: Selected Studies*. John Wiley and Sons: New York.

U.S. Census Bureau (1998). *TIGER/Line 1997*. Bureau of the Census, U. S. Department of Commerce: Washington, DC.

USDOJ (2000). *Regional Crime Analysis Geographic Information System (RCAGIS)*. Criminal Division, U.S. Department of Justice: Washington, DC. http://www.usdoj.gov/criminal/gis/rcagishome.htm.

Van Koppen, P. J. and J. W. De Keijser (1997). "Desisting distance decay: on the aggregation of individual crime trips", *Criminology*, 35, 2: 505-513.

Venables, W.N. and B.D. Ripley (1997). *Modern Applied Statistics with S-Plus (second edition)*. Springer-Verlag: New York.

Ward, J. H. (1963). "Hierarchical grouping to optimize an objective function". *Journal of the American Statistical Association*. 58, 236-244.

Weber, A. (1909). *Über den Standort der Industrien* (Theory of Location of Industries).

Weisburd, David and Tom McEwen (1998). *Crime Mapping Crime Prevention*. Criminal Justice Press: Monsey, NY.

Weisburd, David and Lorraine Green (1995). "Policing drug hot spots: the Jersey City drug market analysis experiment". *Justice Quarterly*. 12 (4), 711-735.

Whittle, P. (1958). "On the smoothing of probability density functions". *Journal of the Royal Statistical Society, Series B*, 55, 549-557.

Wilson, A. G. (1970). *Entropy in Urban and Regional Planning*. Leonard Hill Books: Buckinghamshire.

White, R. C. (1932). "The relation of felonies to environmental factors in Indianapolis". *Social Forces*, 10(4), 498-509.

Xie, X. L. and G. Beni (1991). "A validity measure for fuzzy clustering". IEEE Trans. Pattern Analysis Machine Intell., 13, 841-847.

Zipf, G. K. (1949). *Human Behaviour and the Principle of Least Effort*. Cambridge.

Appendix A

Dynamic Data Exchange (DDE) Support

CrimeStat supports Dynamic Data Exchange (DDE). This allows the program to be linked to another program, which can call up *CrimeStat* as a routine. The following are the programming codes used to support DDE commands.

CrimeStat's DDE Topics That Support the DDE "poke" Command

Topic	Item	Data format
Primary File	File	RemoveAll
		AddTextFile| <name>| <columns>| <separator>| <header rows>
		AddDbfFile| <name>
		AddShpFile| <name>
	X	<file name>| <column name>
	Y	<file name>| <column name>
	Weight	<file name>| <column name>
	Intensity	<file name>| <column name>
	Direction	<file name>| <column name>

	Coordinate	<coordinate>| <unit>
		Valid coordinates:
		  Longtitude, latitude
		  Projected
		  Direction
		Valid units:
		  Decimal degrees
		  Feet
		  Meters

Secondary File	File	See Primary File
	X	See Primary File
	Y	See Primary File
	Weight	See Primary File
	Intensity	See Primary File
	Direction	See Primary File

Reference File	Source	<source>
		Valid sources:
		  From File
		  Generated

	File	See Primary File

A-1

	X	See Primary File
	Y	See Primary File
	True Grid	\<number of columns\> a value of zero (0) will uncheck the check box.
	Bound	\<lower-left x\>| \<lower-left y\>| \<upper-right x\>| \<upper-right y\>

	Cell specification	\<source\>| \<value\>
		Valid sources:
		  By cell-spacing
		  By number of columns

Measurement Parameters	Measurement Type	\<type\>
		Valid types:
		  Direct
		  Indirect

	Area	\<area\>| \<area unit\> Valid units: See Dialog
	Length	\<length\>| \<length unit\> Valid units: See Dialog

CrimeStat's DDE Topics That Support the DDE "request" Command

Topic	Item	Return value
System	SysItems	Name of the supported items of the "system" topic.
	ReturnMessage	Detailed information (if any) of the last message.
	Status	Server status, either "Ready" or "Busy".
	Formats	Supported data formats.
	Help	Detailed help on CrimeStat's DDE support.
	TopicItemList	Name of the supported items of the current topic.

CrimeStat's DDE Topics That Support the DDE "execute" Command

Topic	Command	Description
System	Quit	Close CrimeStat.
Primary File	Select	Select the Primary file tab.
Secondary File	Select	Select the Secondary file tab.
Reference File	Select	Select the Reference file tab.

A-2

Measurement Parameters	Select	Select the Measurement parameters tab.
Spatial Distribution	Select	Select the Spatial distribution tab.
Distance Analysis	Select	Select the Distance analysis tab.
'Hot spot' Analysis I	Select	Select the 'Hot spot' analysis I tab.
'Hot spot' Analysis II	Select	Select the 'Hot spot' analysis II tab.
Interpolation	Select	Select the Interpolation tab.

Example: Controlling *CrimeStat* from within Visual basic

```
Public Function OpenCrimeStat(topic As String) As Variant
   On Error Resume Next
   Dim channel, I
   Dim file As String
   file = "CrimeStat.exe"
   channel = DDEInitiate("CrimeStat", topic)
   If Err Then
      Err = 0
      I = Shell(file, 1)
      If Err Then
         Return
      End If
      channel = DDEInitiate("CrimeStat", topic)
   End If
   OpenCrimeStat = channel
End Function


Public Sub TestCrimeStatDde(foo As String)
   On Error Resume Next
   Dim file As String
   Dim channel

   file = "SampleData.dbf"
   channel = OpenCrimeStat("Primary File")
   DDEPoke channel, "Coordinate", "Projected| Feet"
   DDEPoke channel, "File", "RemoveAll"
   DDEPoke channel, "File", "AddDbfFile| " & file
   DDEPoke channel, "X", file & "| LON"
   DDEPoke channel, "Y", file & "| LAT"
   DDEPoke channel, "Coordinate", "Longtitude, latitude| Decimal degrees"
   DDETerminate channel

   file = "Grid.dbf"
   channel = OpenCrimeStat("Reference File")
```

A-3

```
    DDEPoke channel, "Source", "From File"
    DDEPoke channel, "True Grid", "0"
    DDEPoke channel, "File", "RemoveAll"
    DDEPoke channel, "File", "AddDbfFile| " & file
    DDEPoke channel, "X", file & "| LON"
    DDEPoke channel, "Y", file & "| LAT"
    DDEPoke channel, "True Grid", "108"
    DDEPoke channel, "Source", "Generated"
    DDEPoke channel, "Bound", "-78.5| 22.4| -75.3| 24.2"
    DDEPoke channel, "Cell Specification", "By cell-spacing| 0.5"
    DDEPoke channel, "Cell Specification", "By number of columns| 20"
    DDETerminate channel

    channel = OpenCrimeStat("Measurement Parameters")
    DDEPoke channel, "Measurement Type", "Direct"
    DDEPoke channel, "Measurement Type", "Indirect"
    DDEPoke channel, "Area", "734.12| Square meters"
    DDEPoke channel, "Length", "1734.12| meters"
    DDETerminate channel

    channel = OpenCrimeStat("Interpolation")
    DDEExecute channel, "select"
    DDETerminate channel
End Sub

Private Sub CrimeStatQuit_Click()
    On Error Resume Next
    Dim channel
    channel = OpenCrimeStat("System")
    DDEExecute channel, "quit"
    DDETerminate channel
End Sub
Private Sub TestCrimeStat_Click()
    TestCrimeStatDde "bar"
End Sub
```

A-4

Some Notes on the Statistical Comparison of Two Samples

The following presents methods for testing the spatial differences between two distributions. At this point, *CrimeStat* does not include routines for testing the differences between two or more samples. The following is provided for the reader's information. Chapter 4 discussed the calculation of these statistics as a single distribution.

Differences in the Mean Center of Two Samples

For differences between two samples in the mean center, it is necessary to test both differences in the X coordinate and differences in the Y coordinates. Since *CrimeStat* outputs both the mean X, mean Y, standard deviation of X, and standard deviation of Y, a simple t-test can be set up. The null hypothesis is that the mean centers are equal

$$H_0: \quad \mu_{XA} = \mu_{XB}$$
$$\mu_{YA} = \mu_{YB}$$

and the alternative hypothesis is that the mean centers are not equal

$$H_1: \quad \mu_{XA} \neq \mu_{XB}$$
$$\mu_{YA} \neq \mu_{YB}$$

Because the true standard deviations of sample A, σ_{XA} and σ_{YA}, and sample B, σ_{XB} and σ_{YB}, are not known, the sample standard deviations are taken, S_{XA}, S_{YA}, S_{XB} and S_{YB}. However, since there are two different variables being tested (mean of X and mean of Y for groups 1 and 2), the alternative hypothesis has two fundamentally different interpretations:

Comparison I: That EITHER $\mu_{XA} \neq \mu_{XB}$ OR $\mu_{YA} \neq \mu_{YB}$ is true

Comparison II: That BOTH $\mu_{XA} \neq \mu_{XB}$ AND $\mu_{YA} \neq \mu_{YB}$ are true

In the first case, the mean centers will be considered not being equal if either the mean of X *or* the mean of Y are significantly different. In the second case, both the mean of *and* the mean of Y must be significantly different for the mean centers to be considered not equal. The first case is clearly easier to fulfill than the second.

Significance levels

By tradition, significance tests for comparisons between two means are made at the $\alpha \leq .05$ or $\alpha \leq .01$ levels, though there is nothing absolute about those levels. The significance levels are selected to minimize *Type 1 Errors*, inadvertently declaring a difference in the means when, in reality, there is not a difference. Thus, a test establishes that the

B-1

likelihood of falsely rejecting the null hypothesis be less than one-in-twenty (less strict) or one-in-one hundred (more strict).

However, with multiple comparisons, the chances increase for finding 'significance' due to the multiple tests. For example, with two tests - a difference in the means of the X coordinate and a difference in the means of the Y coordinate, the likelihood of rejecting the first null hypothesis ($\mu_{XA} \neq \mu_{XB}$) is one-in-twenty and the likelihood of rejecting the second null hypothesis ($\mu_{YA} \neq \mu_{YB}$) is also one-in-twenty, then the likelihood of rejecting either one null hypothesis or the other is actually one-in-ten.

To handle this situation, comparison I - the 'either/or' condition, a Bonferoni test is appropriate (Anselin, 1995; Systat, 1996). Because the likelihood of achieving a given significance level increases with multiple tests, a 'penalty' must be assigned in finding either the differences in means for the X coordinate or differences in means for the Y coordinates significant. The Bonferoni criteria divides the critical probability level by the number of tests. Thus, if the $\alpha \leq .05$ level is taken for rejecting the null hypothesis, the critical probability for each mean must be .025 (.05/2); that is, differences in either the mean of X or mean of Y between two groups must yield a significance level less than .025.

For comparison II - the 'both/and' condition, on the other hand, the test is more stringent since the differences between the means of X and the means of Y must both be significant. Following the logic of the Bonferoni criteria, the critical probability level is multiplied by the number of tests. Thus, if the $\alpha = .05$ level is taken for rejecting the null hypothesis, then *both* tests must be significant at the $\alpha \leq .10$ level (i.e., .05*2).[1]

Tests

The statistics used are for the t-test of the difference between means (Kanji, 1993).

a. First, test for equality of variances by taking the ratio of the variances (squared sample standard deviations) of both the X and Y coordinates:

$$F_X = \frac{S_{XA}^2}{S_{XB}^2} \qquad\qquad (B.1)$$

$$F_Y = \frac{S_{YA}^2}{S_{YB}^2} \qquad\qquad (B.2)$$

with ($N_A - 1$) and ($N_B - 1$) degrees of freedom for groups *A* and *B* respectively. This test is usually done with the larger of the variances in the numerator. Since there are two variances being compared (for X and Y, respectively), the logic should follow either *I* or *II* above (i.e., if either are to be true, then the critical α will be actually $\alpha/2$ for each; if both must be true, then the critical α

B-2

will be actually $2*\alpha$ for each).

b.   Second, if the variances are considered equal, then a t-test for two group means with unknown, but equal, variances can be used (Kanji, 1993; 28). Let

$$S_{XAB} = SQRT \left[\frac{\sum\limits_{i=1}^{N(A)} (X_{Ai} - \bar{X}_A)^2 + \sum\limits_{i=1}^{N(B)} (X_{Bi} - \bar{X}_B)^2}{(N_A + N_B - 2)} \right] \qquad (B.3)$$

$$S_{YAB} = SQRT \left[\frac{\sum\limits_{i=1}^{N(A)} (Y_{Ai} - \bar{Y}_A)^2 + \sum\limits_{i=1}^{N(B)} (Y_{Bi} - \bar{Y}_B)^2}{(N_A + N_B - 2)} \right] \qquad (B.4)$$

where the summations are for i=1 to N within each group separately. Then the test becomes

$$t_X = \frac{(\bar{X}_A - \bar{X}_B) - (\mu_{XA} - \mu_{XB})}{S_{XAB} * SQRT\left[\frac{1}{N_A} + \frac{1}{N_B} \right]} \qquad (B.5)$$

$$t_Y = \frac{(\bar{Y}_A - \bar{Y}_B) - (\mu_{YA} - \mu_{YB})}{S_{YAB} * SQRT\left[\frac{1}{N_A} + \frac{1}{N_B} \right]} \qquad (B.6)$$

with $(N_A + N_B - 2)$ degrees of freedom for each test.

c.   Third, if the variances are not equal, then a t-test for two group means with unknown and unequal variances should be used (Kanji, 1993; 29).

$$t_X = \frac{(\bar{X}_A - \bar{X}_B) - (\mu_{XA} - \mu_{XB})}{\text{SQRT} \left\{ \left[\dfrac{S_{XA}^2}{N_A} + \dfrac{S_{XB}^2}{N_B} \right] \right\}} \tag{B.7}$$

$$t_Y = \frac{(\bar{Y}_A - \bar{Y}_B) - (\mu_{YA} - \mu_{YB})}{\text{SQRT} \left\{ \left[\dfrac{S_{YA}^2}{N_A} + \dfrac{S_{YB}^2}{N_B} \right] \right\}} \tag{B.8}$$

with degrees of freedom

$$v = \left\{ \frac{\left[\dfrac{S_A^2}{N_A} + \dfrac{S_B^2}{N_B} \right]}{\left[\dfrac{S_A^4}{N_A^2(N_A + 1)} + \dfrac{S_B^4}{N_B^2(N_B + 1)} \right]} \right\} - 2 \tag{B.9}$$

for both the X and Y test. Even though this latter formula is cumbersome, in practice, if the sample size of each group is greater than 100, then the t-values for infinity can be taken as a reasonable approximation and the above degrees of freedom need not be tested (t=1.645 for α=.05; t=1.960 for α=.01).

d. The significance levels are those selected above. For comparison I - that either differences in the means of X or differences in the means of Y are significant, the critical probability level is $\alpha/2$ (e.g., .05/2 = .025; .01/2 = .005). For comparison II - that both differences in the means of X and differences in the means of Y are significant, the critical probability level is $\alpha*2$ (e.g., .05*2 = .10; .01*2 - .02).

e. Reject the null hypothesis if:

Comparison I: Either tested t-value (t_x or t_y) is greater than the Critical t for $\alpha/2$

Comparison II: Both tested t-values (t_x and t_y) are greater than the critical t for $\alpha*2$

Example 1: Burglaries and Robberies in Baltimore County

To illustrate, compare the distribution of burglaries in Baltimore County with those of robberies, both for 1996. Figure B.1 shows the mean center of all robberies (blue square) and all residential burglaries (red triangle). As can be seen, the mean centers are located within Baltimore City, a property of the unusual shape of the county (which surrounds the city on three sides). Thus, these mean centers cannot be considered an unbiased estimate of the metropolitan area, but unbiased estimates for the County only. When the relative positions of the two mean centers are compared (figure 4.12 in chapter 4), the center of robberies is south and west of the center for burglaries. Is this difference significant or not?

To test this, the standard deviations of the two distributions are first compared and the F-test of the larger to the smaller variance is used (equations B.1 and B.2). *CrimeStat* provides the standard deviation of both the X and Y coordinates; the variance is the square of the standard deviation. In this case, the variance for burglaries is slightly larger than for robberies for both the X and Y coordinates.

$$F_X = \frac{S_{XA}^2}{S_{XB}^2} = \frac{0.0154}{0.0145} = 1.058$$

$$F_Y = \frac{S_{YA}^2}{S_{YB}^2} = \frac{0.0058}{0.0029} = 2.007$$

Because both samples are fairly large (1180 robberies and 6051 burglaries), the degrees of freedom are also very large. The F-tables are a little indeterminate with large samples, but the variance ratio approaches 1.00 as the sample reaches infinity. An approximate critical F-ratio can be obtained by the next largest pair of values in the table (1.22 for $p \leq .05$ and 1.32 for $p \leq .01$). Using this criteria, differences in the variances for the X coordinate are probably not significant while that for the Y coordinates definitely are significant. Consequently, the test for a difference in means with unequal variances is used (equations B.7, B.8 and B.9).
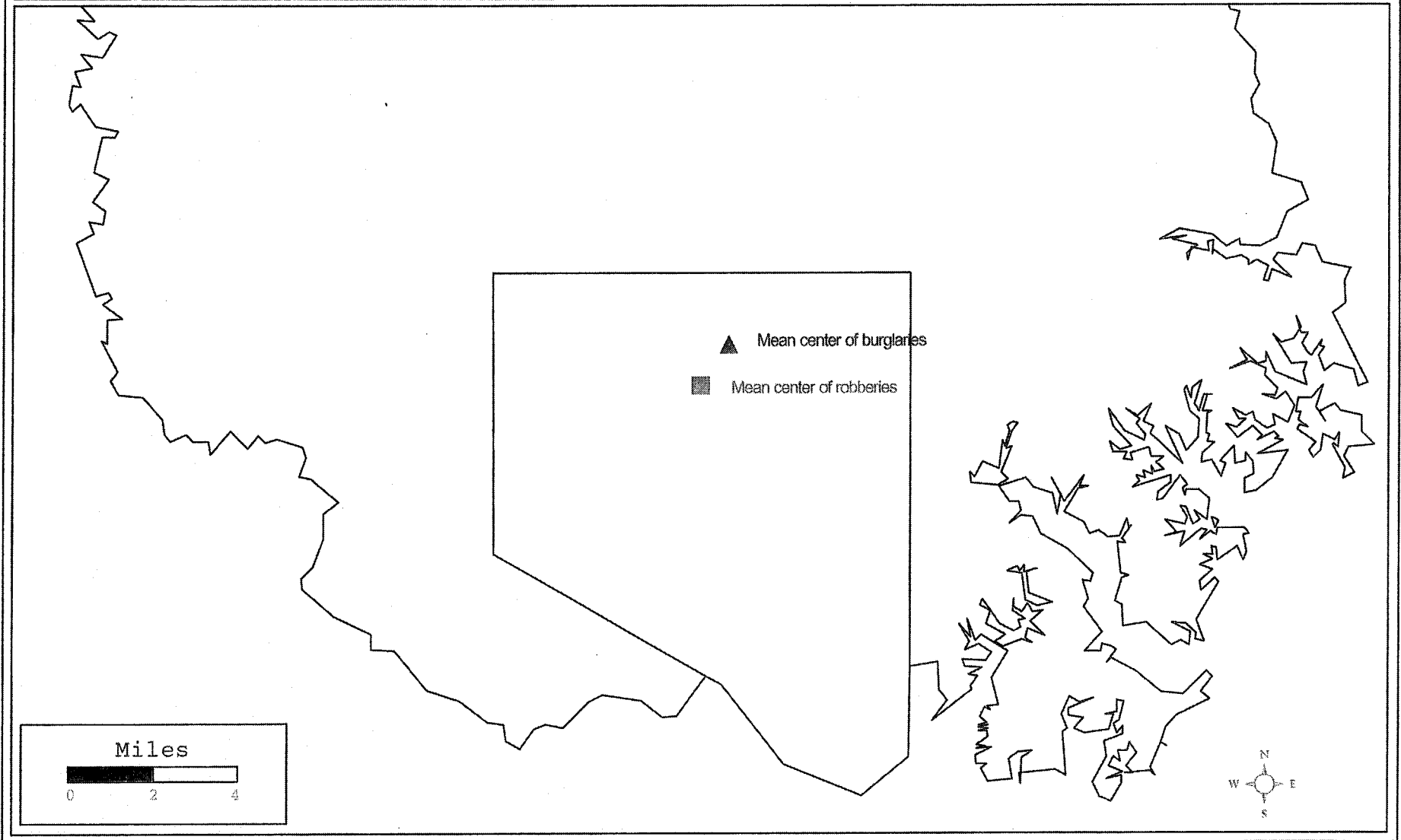
$$t_X = \frac{(\overline{X}_A - \overline{X}_B) - (\mu_{XA} - \mu_{XB})}{SQRT\left\{\left[\dfrac{S_{XA}^2}{N_A} + \dfrac{S_{XB}^2}{N_B}\right]\right\}} = \frac{-76.608482 - (-76.620838)}{SQRT\left\{\left[\dfrac{0.0154}{6051} + \dfrac{0.0145}{1180}\right]\right\}}$$

$$= \frac{0.0124}{0.0039} = 3.21 \; (p \leq .005)$$

B-5

Figure B.1:

1996 Burglaries and Robberies in Baltimore County

Comparison of Mean Centers

▲ Mean center of burglaries

■ Mean center of robberies

Miles

0   2   4

$$t_Y = \frac{(\overline{Y}_A - \overline{Y}_B) - (\mu_{YA} - \mu_{YB})}{SQRT\{[\frac{S_{YA}^2}{N_A} + \frac{S_{YB}^2}{N_B}]\}} = \frac{39.348368 - 39.334816}{SQRT\{[\frac{0.0058}{6051} + \frac{0.0029}{1180}]\}}$$

$$= \frac{0.0136}{0.0018} = 7.36 \ (p \leq .005)$$

Therefore, whether we use the 'either/or' test (critical $\alpha \leq .025$) or the 'both/and' test (critical $\alpha \leq .1$), we find that the difference in the mean centers is highly significant. Burglaries have a different center of gravity than robberies in Baltimore County.

Differences in the Standard Distance Deviation of Two Samples

Since the standard distance deviation, S_{XY} (equation 4.6 in chapter 4) is a standard deviation, differences in the standard distances of two groups can be compared with an equality of variance test (Kanji, 1993, 37).

$$F = \frac{S_{XYA}^2}{S_{XYB}^2} \tag{B.10}$$

with $(N_A - 1)$ and $(N_B - 1)$ degrees of freedom for groups A and B, respectively. This test is usually done with the larger of the variances in the numerator. Since there is only one variance being compared, the critical α are as listed in the tables.

From *CrimeStat*, we find that the standard distance deviation of burglaries is 8.44 miles while that for robberies is 7.42 miles. In chapter 4, figure 4.12 displayed these two standard distance deviations. As can be seen, the dispersion of incidents, as defined by the standard distance deviation, is greater for burglaries than for robberies. The F-test of the difference is calculated by

$$F = \frac{S_{XYA}^2}{S_{XYB}^2} = \frac{8.44^2}{7.42^2} = 1.29$$

with 6050 and 1180 degrees of freedom respectively. Again, the F-tables are slightly indeterminate with respect to large samples, but the next largest F beyond infinity is 1.25 for $p \leq .05$ and 1.38 for $p \leq .01$. Thus, it appears that burglaries have a significantly greater dispersion than robberies, at least at the $p \leq .05$ level.

Differences in the Standard Deviational Ellipse of Two Samples

In a standard deviational ellipse, there are actually six variables being compared:

> Mean of X
> Mean of Y
> Angle of rotation
> Standard deviation along the transformed X axis
> Standard deviation along the transformed Y axis
> Area of the ellipse

Differences in the mean centers

Comparisons between the two mean centers can be tested with the above statistics.

Differences in the angle of rotation

Unfortunately, to our knowledge, there is not a formal test for the difference in the angle of rotation. Until this test is developed, we have to rely on subjective judgements.

Differences in the standard deviations along the transformed axes

The differences in the standard deviations along the transformed axes (X and Y) can be tested with an equality of variance test (Kanji, 1993, 37).

$$F_{Sx} = \frac{S_{X1}^2}{S_{X2}^2} \tag{B.11}$$

$$F_{SY} = \frac{S_{Y1}^2}{S_{Y2}^2} \tag{B.12}$$

with $(N_A - 1)$ and $(N_B - 1)$ degrees of freedom for groups A and B respectively. This test is usually done with the larger of the variances in the numerator. The example above for comparing the mean centers of Baltimore County burglaries and robberies illustrated the use of this test.

Differences in the areas of the two ellipses

Since an area is a variance, the differences in the areas of the two ellipses can be compared with an equality of variance test (Kanji, 1993, 37).

B-8

$$F = \frac{\text{Area}_A}{\text{Area}_B} \qquad\qquad (B.13)$$

with $(N_A - 1)$ and $(N_B - 1)$ degrees of freedom for groups 1 and 2 respectively. This test is done with the larger of the variances in the numerator.

Significance levels

The testing of each of these parameters for the difference between two ellipses is even more complicated than the difference between two mean centers since there are up to six parameters which must be tested (differences in mean X, mean Y, angle of rotation, standard deviation along transformed X axis, standard deviation along transformed Y axis, and area of ellipse). However, as with differences in mean center of two groups, there are two different interpretations of differences.

Comparison I: That the two ellipses differ on ANY of the parameters

Comparison II: That the two ellipses differ on ALL parameters.

In the first case, the critical probability level, α, must be divided by the number of parameters being tested, α/p. In theory, this could involve up to six tests, though in practice some of these may not be tested (e.g., the angle of rotation). For example, if five of the parameters are being estimated, then the critical probability level at $\alpha \leq .05$ is actually $\alpha \leq .01$ (.05/5).

In the second case, the critical probability level, α, is multiplied by the number of parameters being tested, $\alpha*p$, since *all* tests must be significant for the two ellipses to be considered as different. For example, if five of the parameters are being estimated, then the critical probability level, say, at $\alpha \leq .05$ is actually $\alpha \leq .25$ (.05*5).

Differences in Mean Direction Between Two Groups

Statistical tests of different angular distributions can be made with the directional mean and variance statistics. To test the difference in the angle of rotation between two groups, a Watson-Williams test can be used (Kanji, 1993; 153-54). The steps in the test are as follows:

1.   All angles, θ_i, are converted into radians

       $\text{Radian}_i \qquad = \qquad \text{Angle}_i * \pi/180 \qquad\qquad (B.14)$

2.   For each sample separately, A and B, the following measures are calculated

B-9

$$C_j = \sum_{A=1}^{N_1} \cos \theta_j \qquad S_j = \sum_{A=1}^{N_1} \sin \theta_j \qquad \text{(B.15)}$$

$$C_k = \sum_{B=1}^{N_2} \cos \theta_k \qquad S_k = \sum_{B=1}^{N_2} \sin \theta_k \qquad \text{(B.16)}$$

where θ_j and θ_k are the individual angles for the respective groups, A and B.

3.   Calculate the resultant lengths of each group

$$R_A = SQRT[\, C_A^2 + S_A^2\,] \qquad \text{(B.17)}$$

$$R_B = SQRT[\, C_B^2 + S_B^2\,] \qquad \text{(B.18)}$$

4.   Resultant lengths for the combined sample are calculated as well as the length of the resultant vector.

$$C = C_A + C_B \qquad \text{(B.19)}$$

$$S = S_A + S_B \qquad \text{(B.20)}$$

$$R = SQRT[\, C^2 + S^2\,] \qquad \text{(B.21)}$$

$$N = N_A + N_B \qquad \text{(B.22)}$$

$$R^* = \frac{(R_A + R_B)}{N} \qquad \text{(B.23)}$$

5.   An F-test of the two angular means is calculated with

$$F = g\,(N-2)\,\frac{R_A + R_B - R}{N - (R_A + R_B)} \qquad \text{(B.24)}$$

where

$$g = 1 - \frac{3}{8k} \qquad \text{(B.25)}$$

with k being identified from a maximum likelihood Von Mises distribution by referencing R^* with 1 and N-2 degrees of freedom (Mardia, 1972; Gaile and Burt, 1980). Some of the reference k's are given in table B.1 (from Mardia, 1972; Kanji, 1993, table 38).

B-10

Table B.1

Maximum Likelihood Estimates for Given R* in the Von Mises Case
(from Mardia, 1972; Kanji, 1993, table 38)

R*	k
0.00	0.00000
0.05	0.10013
0.10	0.20101
0.15	0.30344
0.20	0.40828
0.25	0.51649
0.30	0.62922
0.35	0.74783
0.40	0.87408
0.45	1.01022
0.50	1.15932
0.55	1.32570
0.60	1.51574
0.65	1.73945
0.70	2.01363
0.75	2.36930
0.80	2.87129
0.85	3.68041
0.90	5.3047
0.95	10.2716
1.00	infinity

Table B.2

Comparison of Two Groups for Angular Measurements
Angle of Deviation From Due North

Group A		Group B	
Incident	Measured Angle	Incident	Measured Angle
1	160	1	196
2	184	2	212
3	240	3	297
4	100	4	280
5	95	5	235
6	120	6	353
		7	190
		8	340

B-11

6. Reject the null hypothesis of no angular difference if the calculated F is greater than the critical value $F_{1, N-2}$.

Example 2: Angular comparisons between two groups

A fourth example is that of sets of angular measurements from two different groups, A and B. Table B.2 provides the data for the two sets. The angular mean for Group A is 144.83^0 with a directional variance of 0.35 while the angular mean for Group B is 258.95^0 with a directional variance of 0.47. The higher directional variance for Group B suggests that there is more angular variability than for Group A.

Using the Watson-Wheeler test, we compare these two distributions.

1. All angles are converted into radians (equation B.14).

2. The cosines and sines of each angle are taken and are summed within groups (equations B.15 and B.16).

   $C_A = -3.1981 \qquad S_A = 2.2533$
   $C_B = -.8078 \qquad S_B = -4.1381$

3. The resultants are calculated (equations B.17 and B.18).

   $R_A = 3.9121$
   $R_B = 4.2162$

4. Combined sample characteristics are defined (equations B.19 through B.23).

   $C = -4.0059$
   $S = -1.8848$
   $R = 4.4271$
   $N = 14$
   $R^* = 0.5806$

5. Once the parameter, k, is obtained (approximated from table 4.1 or obtained from Mardia, 1972 or Kanji, 1993), g is calculated, and an F-test is constructed (equations B.24 and B.25).

   $k = 1.44$
   $g = 0.7396$
   $F = 5.59$

6. The critical F for 1 and 12 degrees of freedom is 4.75 ($p \le .05$) and 9.33 ($p \le .01$). The test is significant at the $p \le .05$ level and we reject the null hypothesis of no angular differences between the two groups. Group A has a different angular distribution than Group B.

Endnotes

1. There are limits to the Bonferoni logic. For example, if there were 10 tests, having a threshold significance level of .005 (.05 / 10) for the 'either/or' conditions and a threshold significance level of .50 (.05 * 10) for the 'both/and' would lead to an excessively difficult test in the first case and a much too easy test in the second. Thus, the Bonferoni logic should be applied to only a few tests (e.g., 5 or fewer).

B-13