

CrimeStat IV

Part IV: Spatial Modeling I

Chapter 10:
Kernel Density Interpolation

Ned Levine
Ned Levine & Associates
Houston, TX

Table of Contents

Introduction	10.1
Kernel Density Interpolation	10.1
Kernel Estimates as an Alternative to Histograms	10.3
Kernel Functions	10.7
Kernel Parameters	10.10
Shape and size of the bandwidth	10.11
Three-dimensional kernels	10.11
<i>CrimeStat</i> Kernel Density Methods	10.13
Single Kernel Density Interpolation	10.13
File to be Interpolated	10.13
Method of Interpolation	10.16
Choice of Bandwidth	10.16
Fixed interval	10.16
Adaptive interval	10.17
Output Unit	10.17
Intensity or Weighting Variable	10.17
Density Calculation	10.18
Output File	10.18
Example 1: Kernel Density Estimate of Baltimore County Street Robberies	10.19
Dual Kernel Density Interpolation	10.22
File to be Interpolated	10.24
Method of Interpolation	10.24
Choice of Bandwidth	10.24
Fixed interval	10.25
Adaptive interval	10.25
Variable interval	10.25
Use kernel bandwidths that produce stable estimates	10.26
Output Unit	10.26
Intensity or Weighting Variable	10.26
Density Calculation	10.26
Output File	10.28
Example 2: Kernel Density Estimates of Vehicle Thefts	
Relative to Population	10.28
Example 3: Kernel Density Estimates and Risk-adjusted Clustering of	
Robberies Relative to Population	10.31
Visually Presenting Kernel Estimates	10.34

Table of Contents (continued)

Advantages and Limitations of Kernel Density Interpolation	10.34
Advantages of Kernel Density Interpolation	10.34
Limitations of Kernel Density Interpolation	10.34
Conclusion	10.36
References	10.37
Endnotes	10.40
Attachments	10.42
A. Kernel Density Interpolation to Estimate Sampling Bias in the Climatic Response of <i>Sphagnum</i> Spores in North America By Mike Sawada	10.43
B. Describing Crime Spatial Patterns By Time of Day in Belo Horizonte By Renato Assunção, Cláudio Beato, & Bráulio Silva	10.44
C. Using Kernel Density Smoothing and Linking to <i>ArcGIS</i> : Examples from London, England By Spencer Chainey	10.45
D. Infant Death Rate and Low Birth Weight in the I-5 Corridor of Seattle and King County By Richard Hoskins	10.46
E. The Risk of Violent Incidents Relative to Population Density in Cologne Using the Dual Kernel Density Routine By Dietrich Oberwittler & Marc Wiesenhütter	10.47
F. Kernel Density Interpolation of Police Confrontations in Buenos Aires Province, Argentina: 1999 By Gastón Pezzuchi	10.48
G. Evolution of the Urbanization Process in the Brazilian Amazonia By Silvana Amaral, Antônio Miguel V. Monteiro, Gilberto Câmara, & José A. Quintanilha	10.49
H. Using Small Area Estimation to Target Health Services in Harris County, TX By Thomas F. Reynolds	10.50
I. Identifying Voucher Holders and Crime Concentrations Using Dual Kernel Density Estimation By Ron Wilson	10.51

Chapter 10:

Kernel Density Interpolation

Introduction

In this chapter, we discuss tools aimed at interpolating incidents, using the kernel density approach. *Kernel Density Interpolation* (sometimes called *Kernel Density Estimation*) 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 & Shyu, 1993; Venables & 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 & Gatrell, 1995; Burt & Barber, 1996; Bowman & Azalini, 1997).

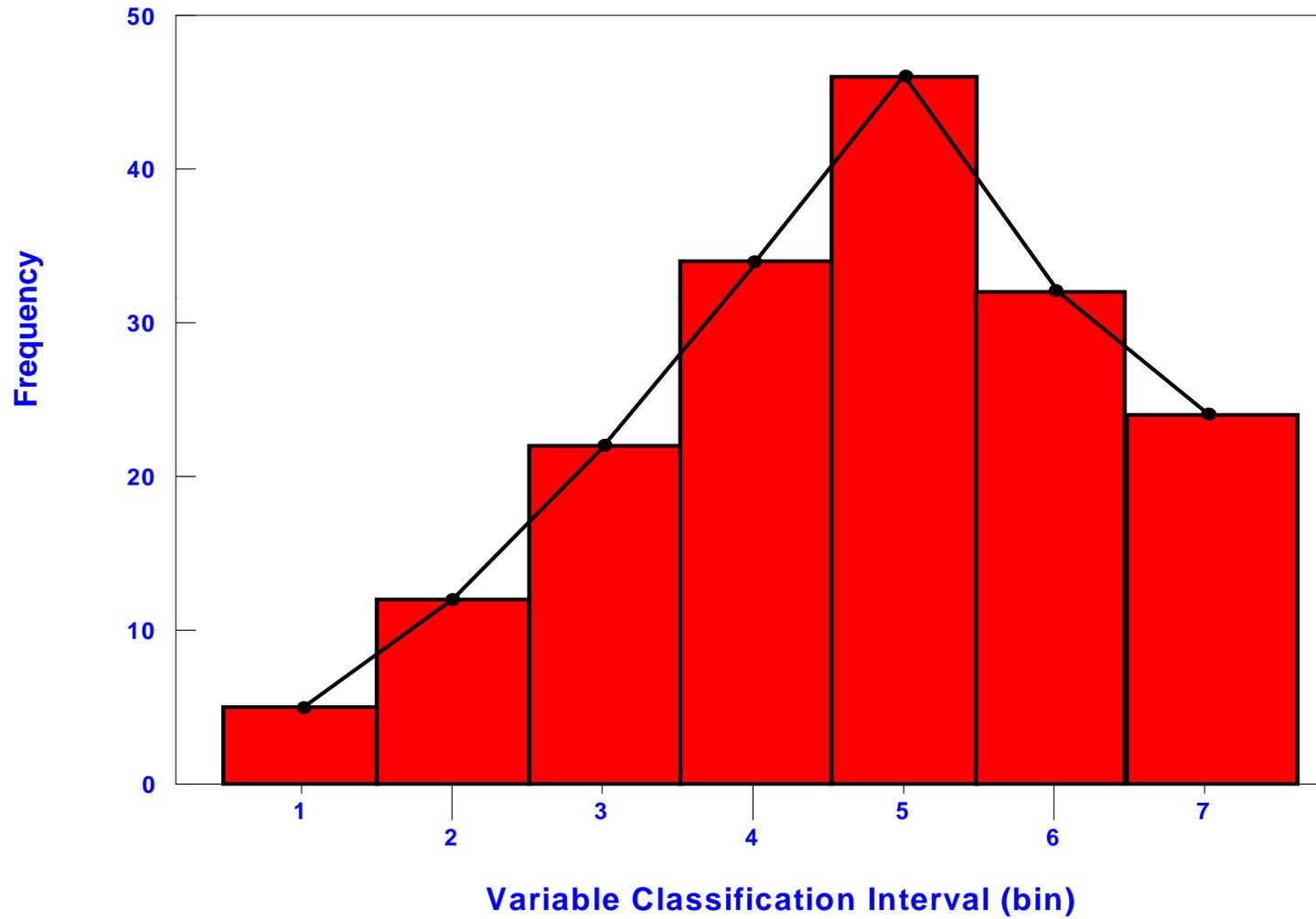
Kernel Density Estimation

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 10.1).

Figure 10.1:

CONSTRUCTING A DENSITY ESTIMATE FROM HISTOGRAM

Method of Connecting Midpoints



Kernel Estimates as an Alternative to Histograms

Unfortunately, doing this causes three statistical problems (Bowman & 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 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* over each point and sum the functions for each location on the scale. Figure 10.2 illustrates the process with five point locations. As seen, over each location, a symmetrical kernel function is placed; by symmetrical is meant that it 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 10.3 shows a kernel with a narrow bandwidth placed over the same five points while Figure 10.4 shows a kernel with a wider bandwidth placed over the points. Clearly, the smoothness of the resulting density function is a result of the bandwidth size.

Figure 10.2:

Kernel Density Estimate

Summing of Normal Kernel Function for 5 Points

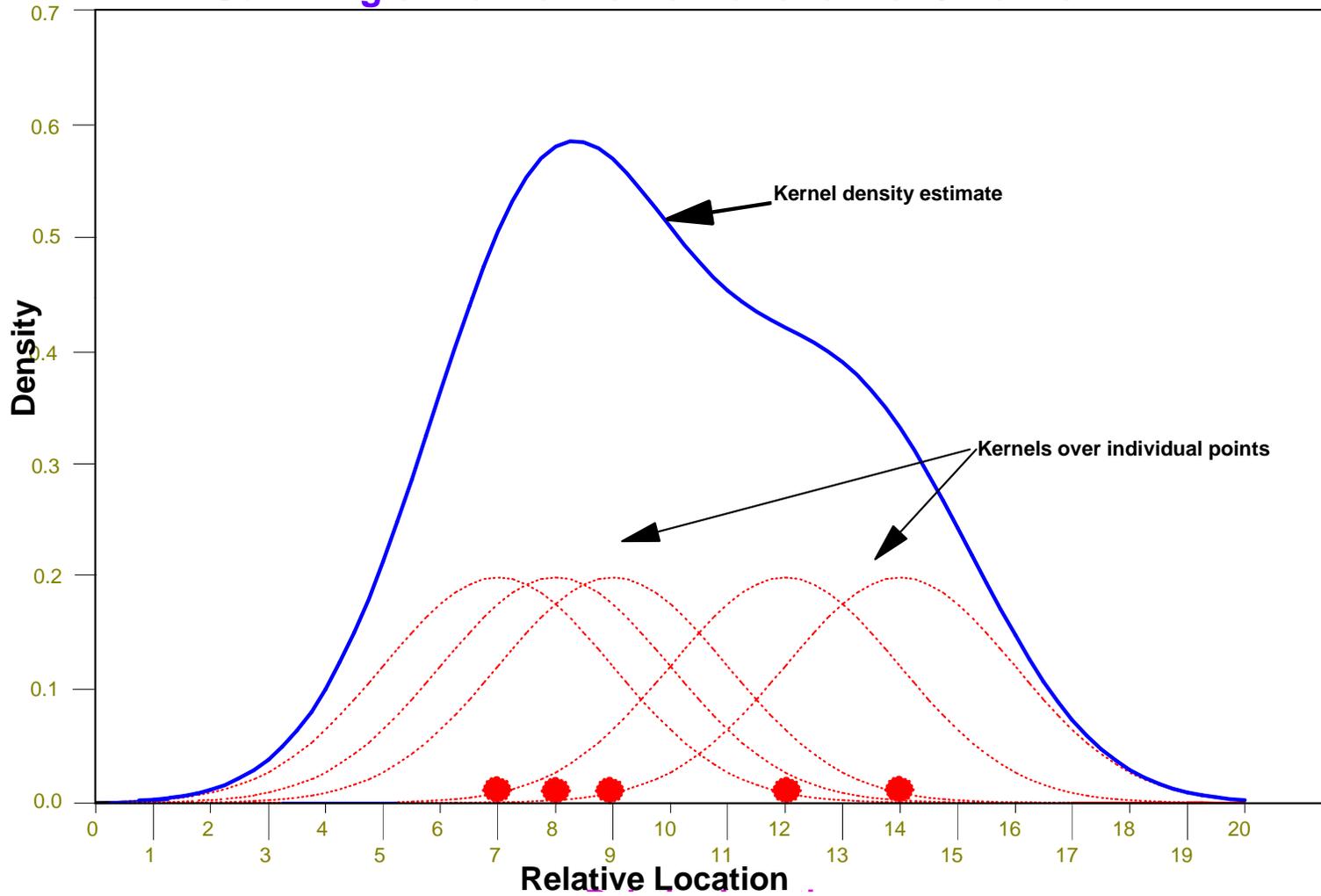


Figure 10.3:
Kernel Density Estimate
Narrower Bandwidth

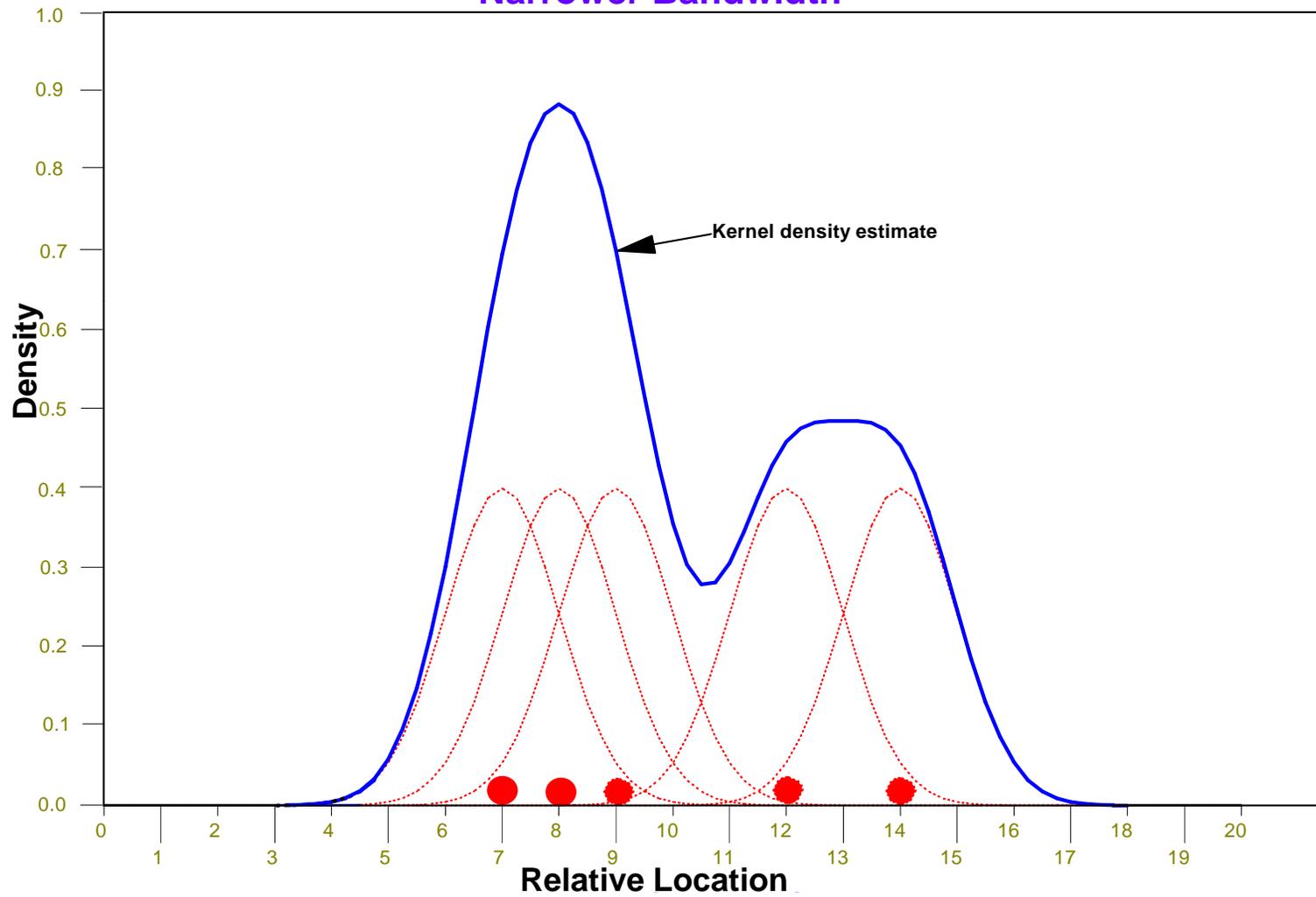
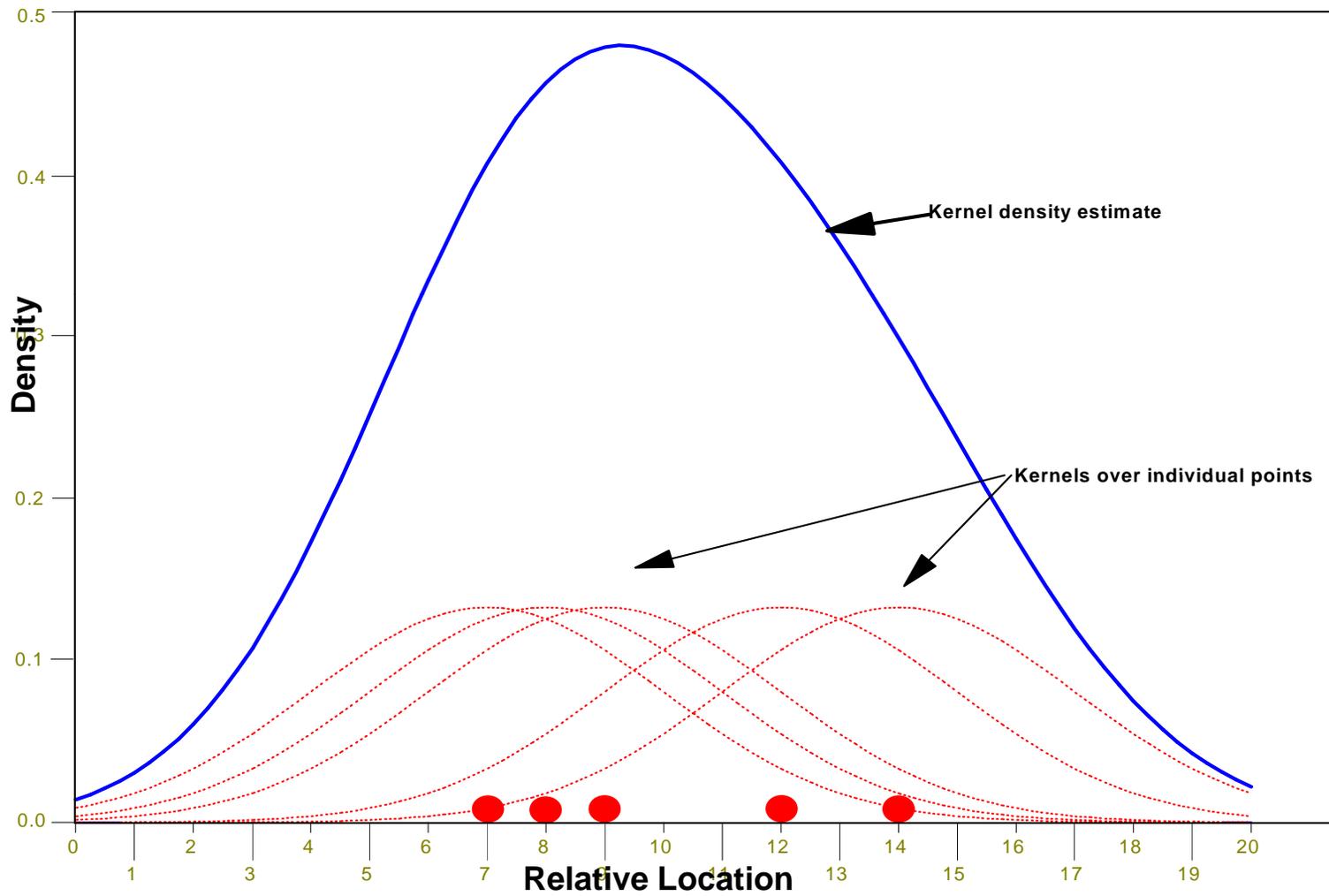


Figure 10.4:
Kernel Density Estimate
Wider Bandwidth



Kernel Functions

There are a number of different kernel functions that have been used in applications. Figure 10.5 illustrates five different kernel functions that are available in *CrimeStat*.

The first is the **normal distribution** and is the most commonly used (Kelsall & Diggle, 1995a). It has the following functional form:

$$g(j) = \sum_{i=1}^N \left[KW_i I_i \frac{1}{h^2 2\pi} e^{-\frac{d_{ij}^2}{2h^2}} \right] \quad (10.1)$$

where $g(x_j)$ is the density of cell j , d_{ij} is the distance between cell j and an incident location, i , h is the standard deviation of the normal distribution (the bandwidth), K is a constant, 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. In *CrimeStat*, the constant K 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).

In other words, the density of cell j is the sum over all incidents of a distance function where the function is the normal distribution. Each cell, in turn, is evaluated with this function and the result is a density estimate for every cell in the reference grid.

In *CrimeStat*, there are four alternative kernel functions that can be used, all of which have a circumscribed bandwidth (search area) 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 bandwidth radius is reached. Its functional form is:

1. Outside the specified bandwidth, h :

$$g(j) = 0 \quad (10.2)$$

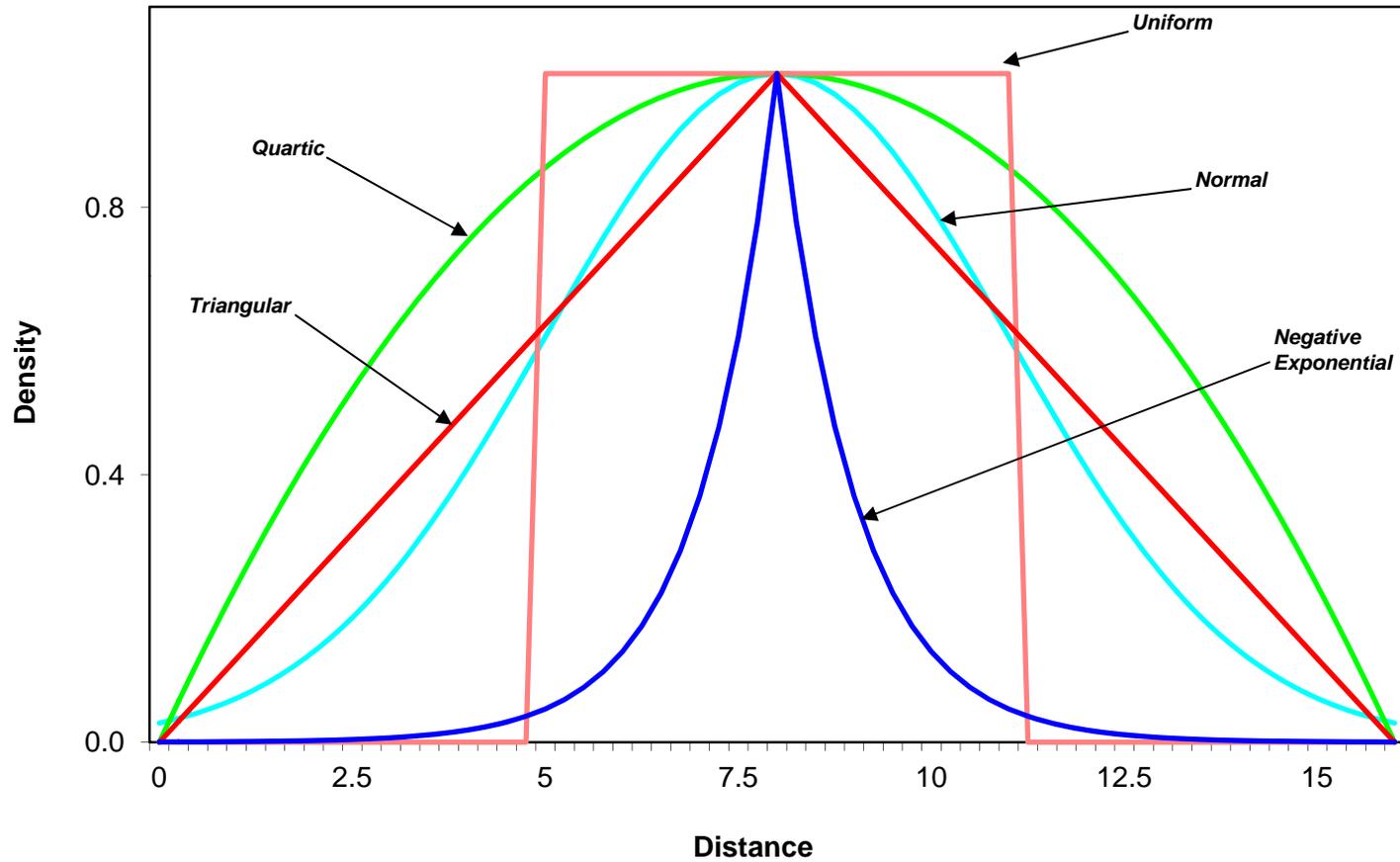
2. Within the specified bandwidth, h :

$$g(j) = \sum_{i=1}^{M_j} \left[KW_i I_i \frac{3}{h^2 2\pi} \left(1 - \frac{d_{ij}^2}{h^2}\right)^2 \right] \quad (10.3)$$

where $g(j)$ is the density of cell j , d_{ij} is the distance between cell j and an incident location, i , h is the radius of the search area (the bandwidth), K is a constant, W_i is a weight at the point location, and I_i is an intensity at the point location. The summation is for the incidents that are within the

Figure 10.5:

Five Types of Kernel Functions



bandwidth. Thus, each cell, j , has a different number of incidents that fall within the bandwidth search area, M_j . In *CrimeStat*, the constant K 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).

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

1. Outside the specified bandwidth, h :

$$g(j) = 0 \tag{10.4}$$

2. Within the specified bandwidth, h :

$$g(j) = \sum_{i=1}^{M_j} [W_i I_i (K - \frac{K}{h}) d_{ij}] \tag{10.5}$$

where $g(x_j)$ is the density of cell j , d_{ij} is the distance between cell j and an incident location, i , h is the radius of the search area (the bandwidth), K is a constant, W_i is a weight at the point location, and I_i is an intensity at the point location. The summation is for the incidents that are within the bandwidth. Thus, each cell, j , has a different number of incidents that fall within the bandwidth search area, M_j . 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:

1. Outside the specified bandwidth, h :

$$g(j) = 0 \tag{10.6}$$

2. Within the specified bandwidth, h :

$$g(j) = \sum_{i=1}^{M_j} W_i I_i K e^{-A d_{ij}} \tag{10.7}$$

where $g(x_j)$ is the density of cell j , d_{ij} is the distance between cell j and an incident location, i , h is the radius of the search area (the bandwidth), K is a constant, A is an exponent, W_i is a weight at the point location, and I_i is an intensity at the point location. The summation is for the

incidents that are within the bandwidth. Thus, each cell, j , has a different number of incidents that fall within the bandwidth search area, M_j . In *CrimeStat*, A is set to 3 while K 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:

1. Outside the specified bandwidth, h :

$$g(j) = 0 \tag{10.8}$$

2. Within the specified bandwidth, h :

$$g(j) = \sum_{i=1}^{M_j} W_i I_i K \tag{10.9}$$

where $g(x_j)$ is the density of cell j , K is a constant, W_i is a weight at the point location, and I_i is an intensity at the point location. The summation is for the incidents that are within the bandwidth. Thus, each cell, j , has a different number of incidents that fall within the bandwidth search area, M_j . 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).

Kernel Parameters

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 points much more highly than far points within the circle and the decay is very rapid.

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 weight more evenly (e.g., the normal distribution, the quartic, or the uniform); these latter ones tend to *smooth* the distribution more.

Shape and size of the bandwidth

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 & Ripley, 1997).

There have also been variations on the size of the bandwidth with various formulas and criteria proposed (Silverman, 1986; Härdle, 1991; Venables & Ripley, 1997). Generally, bandwidth choices fall into either fixed or adaptive (variable) kernels (Kelsall & Diggle, 1995a; Bailey & Gatrell, 1995). *CrimeStat* follows this distinction, which will be explained below.

Another suggestion is to use the Moran correlogram, which was discussed in Chapter 5, to estimate the shape of the weighting function (Cliff & Haggett, 1988; Bailey & Gatrell, 1995). This would be appropriate for variables that have *weights*, such as population or employment. The Moran correlogram displays the degree of spatial autocorrelation as a function of distance. Whether the autocorrelation falls off quickly or more slowly can be used to select an approximate kernel function (e.g., a negative exponential function decays quickly whereas a quartic function decays very slowly). The bandwidth could also be selected by the distance at which the Moran correlogram levels off (i.e., approaches the global I value). This would lead to an estimate that minimizes spatial autocorrelation in the data set. It would be good for capturing major trends in the data, but would not be good for identifying local clusters (hot spots) since the bandwidth distance would incorporate most of a metropolitan area.

Three-dimensional kernels

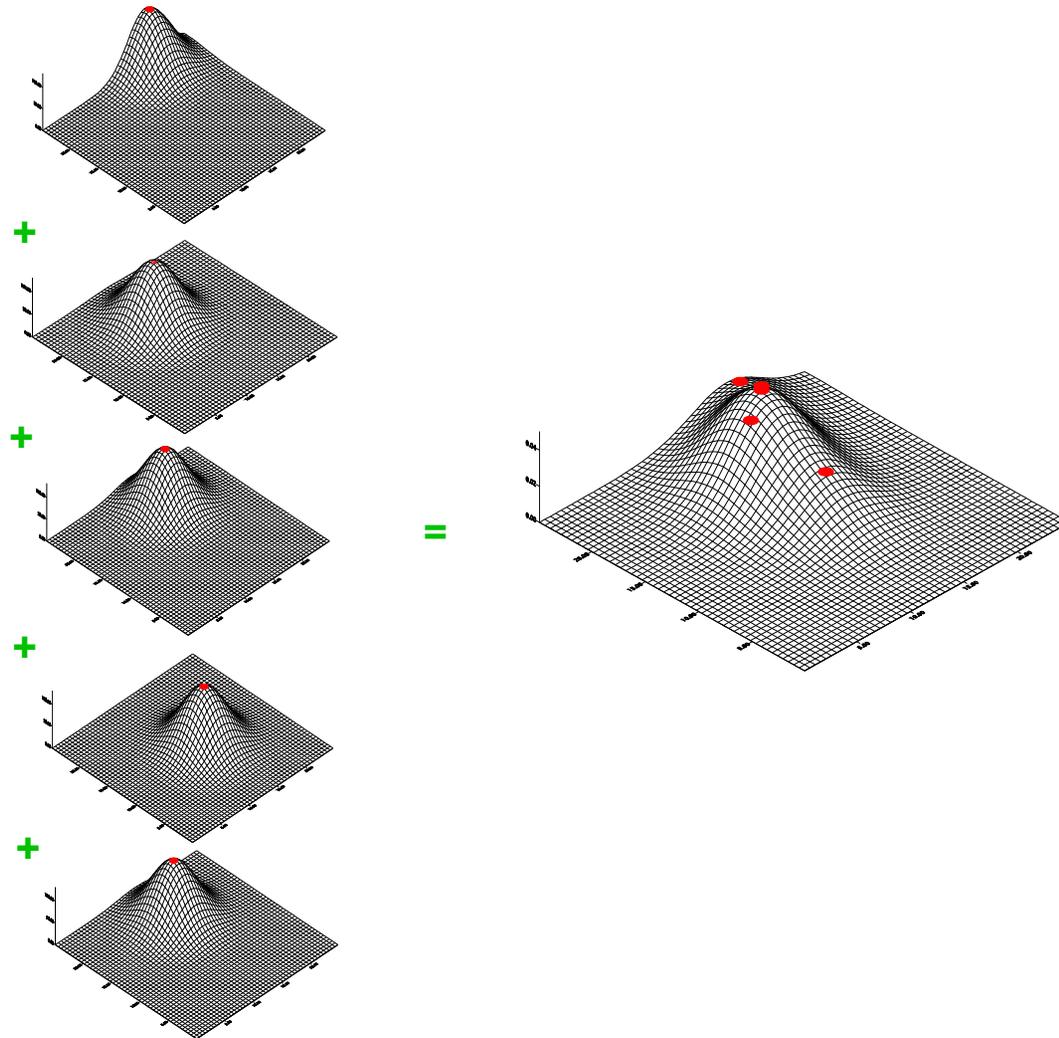
The kernel function can be expanded to more than two dimensions (Härdle, 1991; Bailey & Gatrell, 1995; Burt & Barber, 1996; Bowman & Azalini, 1997). Figure 10.6 shows a three-dimensional normal distribution placed over each of five points with the resulting density surface being a sum of all five individual 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 & Diggle, 1995a; Bowman & Azalini, 1997).

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

Figure 10.6:

Kernel Density Surface

Summing of Normal Kernel Surface for 5 Points



***CrimeStat* Kernel Density Methods**

CrimeStat has two kernel density interpolation routines. The first applies to a single variable while the second to the relationship between two variables. Both routines have a number of options. In addition, kernel density interpolation is used in several other *CrimeStat* routines including journey-to-crime modeling, Bayesian journey-to-crime modeling, and Head-Bang interpolation. Those latter techniques will be discussed in Chapters 11, 13, and 14.

Figure 10.7 shows the Interpolation I screen in *CrimeStat* and the two routines that are available. 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 10.8 illustrates a reference grid over the Baltimore region with 100 columns and 90 rows.

Single Kernel Density Interpolation

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 weighting and intensity variables as the routine will use both variables in calculating densities, which could lead to double weighting.

It is necessary to define the appropriate file on the Primary or Secondary file pages. 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.

File to be Interpolated

First, the user must indicate whether the Primary file or the Secondary file (if used) is to be interpolated.

Figure 10.7:
Interpolation I Screen

CrimeStat IV

Spatial Modeling II | Crime Travel Demand | Options

Data Setup | Spatial Description | Hot Spot Analysis | Spatial Modeling I

Interpolation I | Interpolation II | Space-time analysis | Journey-to-Crime | Bayesian Journey-to-Crime Estimation

Kernel density estimate: Single Dual First file: Second file:

File to be interpolated: Primary Primary Secondary

Method of interpolation: Normal Normal

Choice of bandwidth: Adaptive Adaptive

Minimum sample size: 100 100

Interval: 1 1 1

Interval unit: Miles Miles Miles

Area units: points per Square Miles Square Miles

Use intensity variable:

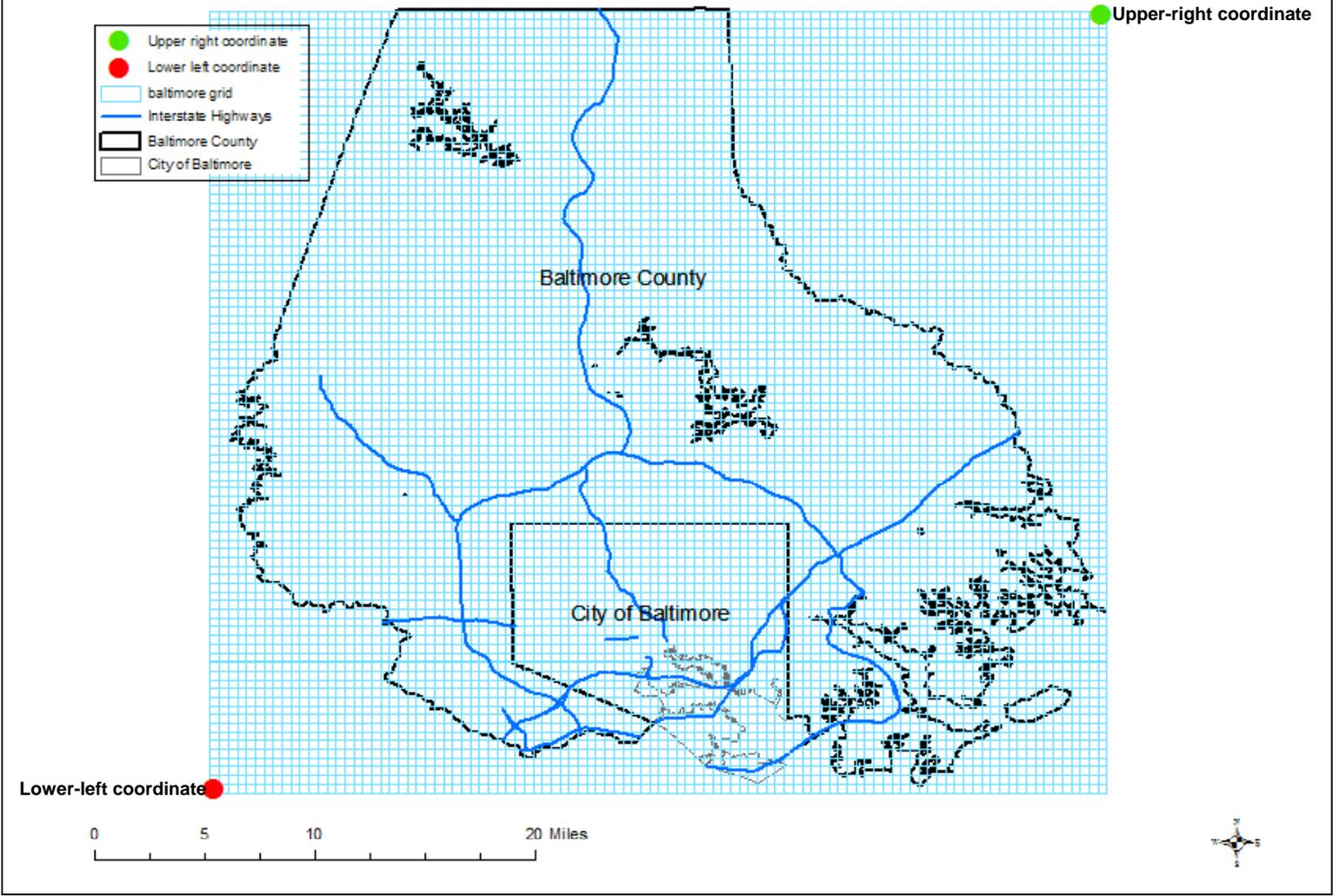
Use weighting variable:

Output units: Absolute Densities Ratio of densities

Output: Save result to... Save result to...

Compute Quit Help

Figure 10.8:
Grid Cell Structure for Baltimore Region
100 Columns by 90 Rows



Method of Interpolation

Second, 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

Third, 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 (square miles, square nautical miles, square feet, square kilometers, or square 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. On the other hand, 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 & Ripley, 1997). Some have argued that the bandwidth be no larger than

the finest resolution that is desired and others have argued for a variation on random nearest neighbor distances (see Spencer Chainey's article in the attachments section of this chapter). Others have argued for particular sizes (Silverman, 1986; Härdle, 1991; Kadafar, 1996; Farewell, 1999; Talbot, Kulldorff, Forand, & Haley, 2000; see endnote *i*). 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 lots of '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, and the 'peaks and valleys' may show 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 sparser, 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 Unit

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

Intensity or Weighting Variable

If an intensity or weighting variable is to be used (and has been defined on the Primary or Secondary file page), the appropriate box must be checked. Be careful about using both intensity and weighting variables to avoid 'double weighting'.

Density Calculation

Finally, 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 equations 10.1-10.9, 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. That is, the estimate is the number of incidents/points that occurred in each grid cell. This is the default choice.
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 square mile or points per square kilometer.
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 outputting the results as probabilities would make more sense. For display purposes, however, it makes no difference as both look the same.

Output File

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, 2008) or *ArcGIS Spatial Analyst* 'asc' format (ESRI, 2012); or 2) Polygon grids in *ArcGIS* '.shp', *MapInfo* '.mif' or various Ascii formats. However, all but *Surfer for Windows* require that the reference grid be created by *CrimeStat*.¹

1 *CrimeStat* will output the geographical boundaries of the reference grid and will assign a third-variable

Example 1: Kernel Density Estimate of Baltimore County Street Robberies

An example can illustrate the use of the single kernel density routine. Figure 10.9 shows a *Surfer for Windows* output of 1180 street robberies for 1996 in Baltimore County. The reference grid was generated by *CrimeStat* and had 100 columns and 90 rows. Thus, the routine calculated the distance between each of the 10,800 reference cells and each of the 1180 robbery incident locations, evaluated the kernel function for each measured 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 10.10 shows an *ArcGIS* map of 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 on the west side, one on the northern border between Baltimore City and Baltimore County, and one on the east side which blends with a smaller peak in the southeast corner of the County.

From one perspective, the kernel estimate is a better 'hot spot' identifier than the cluster analysis routines discussed in Chapters 7 and 8. 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

(called *Z*) as the density estimate. *ArcGIS* '.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 files must be imported to a *MapInfo* '.tab' file. For both output formats, the values of *Z* can be shown as a thematic map but the ranges must be adjusted to illustrate the high density locations (i.e., the default values in most GIS programs will not display the densities very well). On the other hand, the default interval values for *Surfer for Windows* and *ArcGIS Spatial Analyst* provide a reasonably good visualization.

Figure 10.9:
Baltimore County Robberies: 1996-97
Kernel Density Interpolation

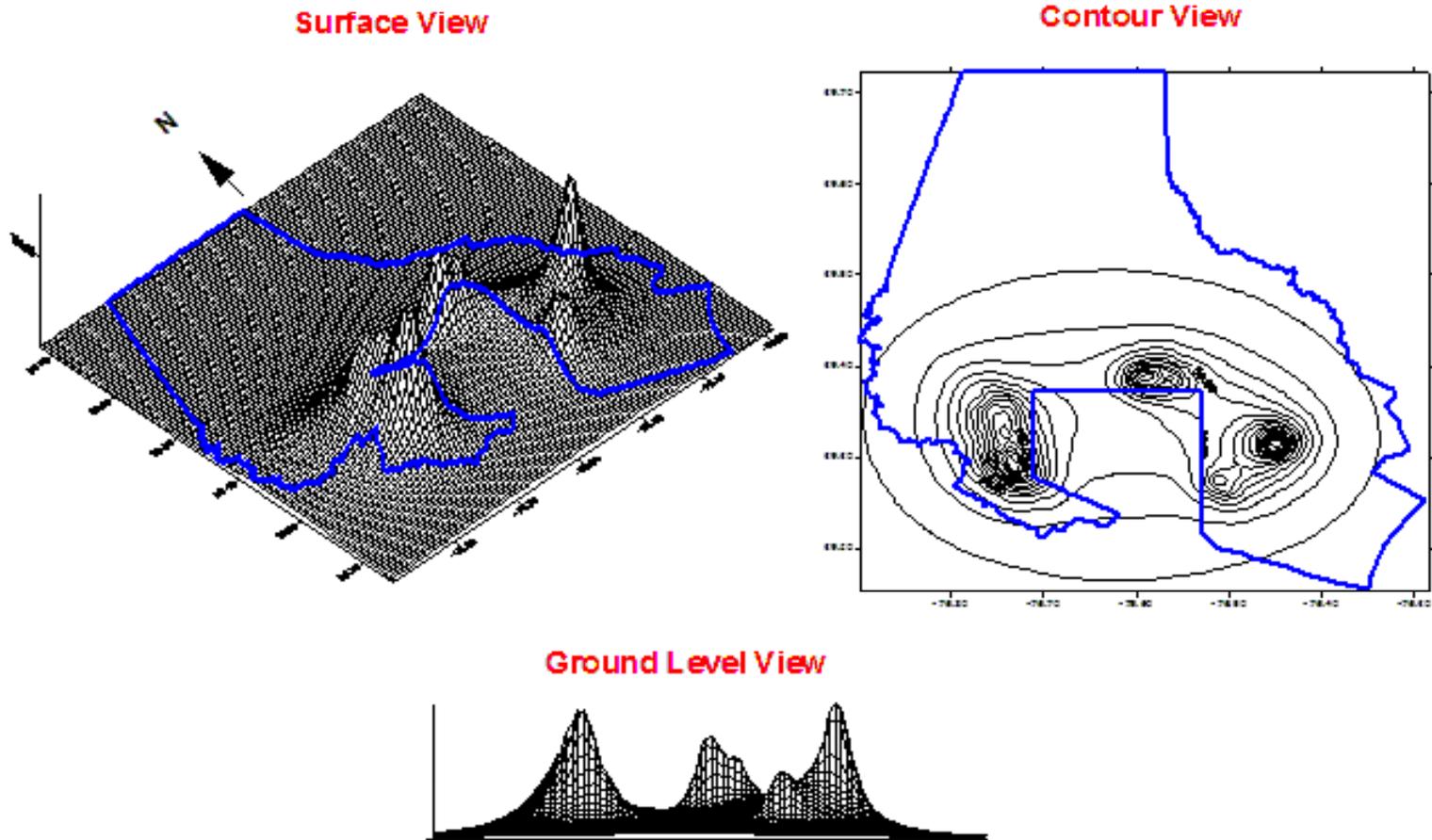
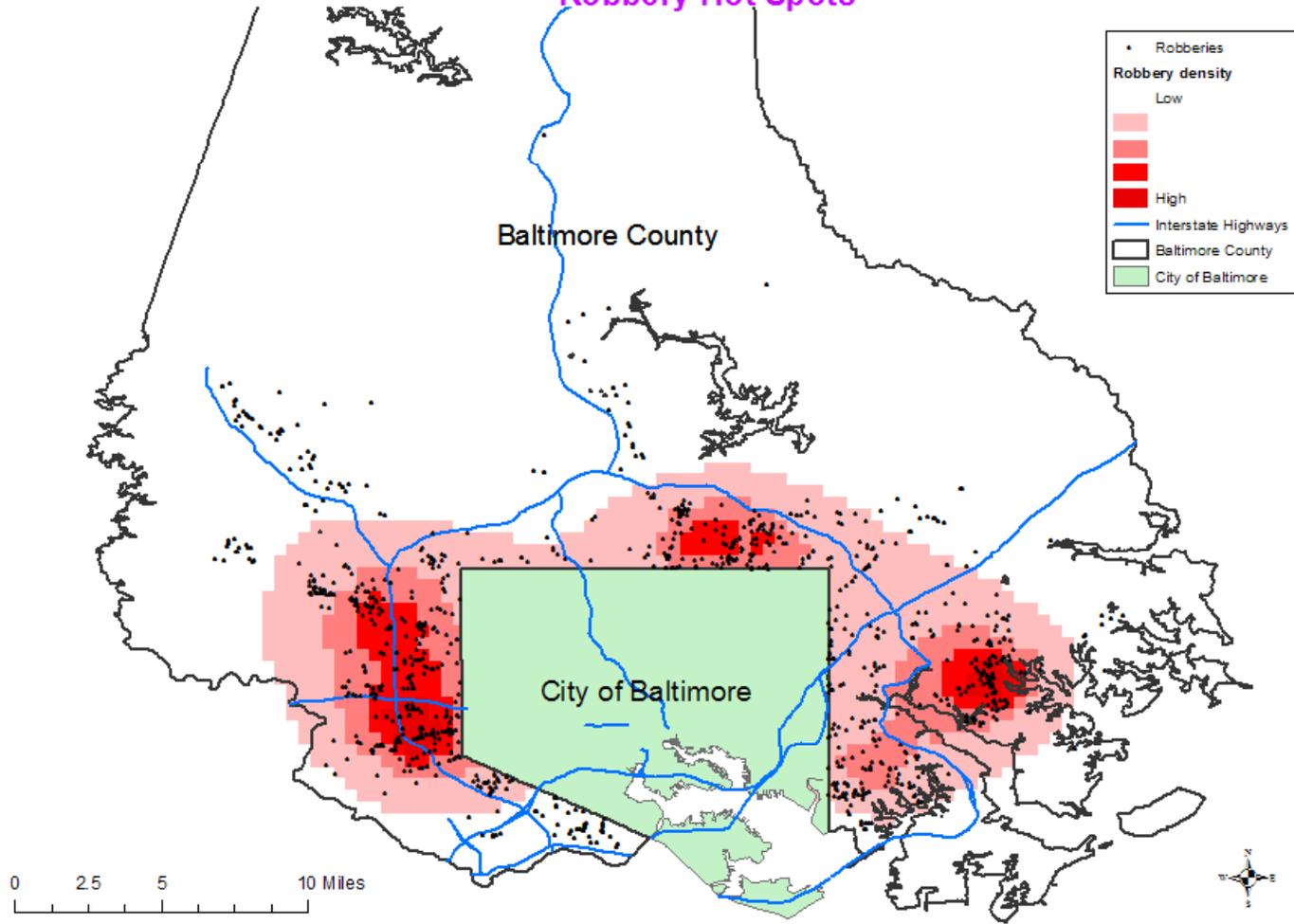


Figure 10.10:
Baltimore County Robberies: 1996
Robbery Hot Spots



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 10.11 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 produced a finer mesh interpolation than an interval of 2 miles, which tended 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 gave very specific peak locations whereas the adaptive interval with a minimum sample size of 100 gave 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 by-products 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. Again, experimentation and discussions with beat officers will be necessary to establish which bandwidth choice should be used in future interpolations.

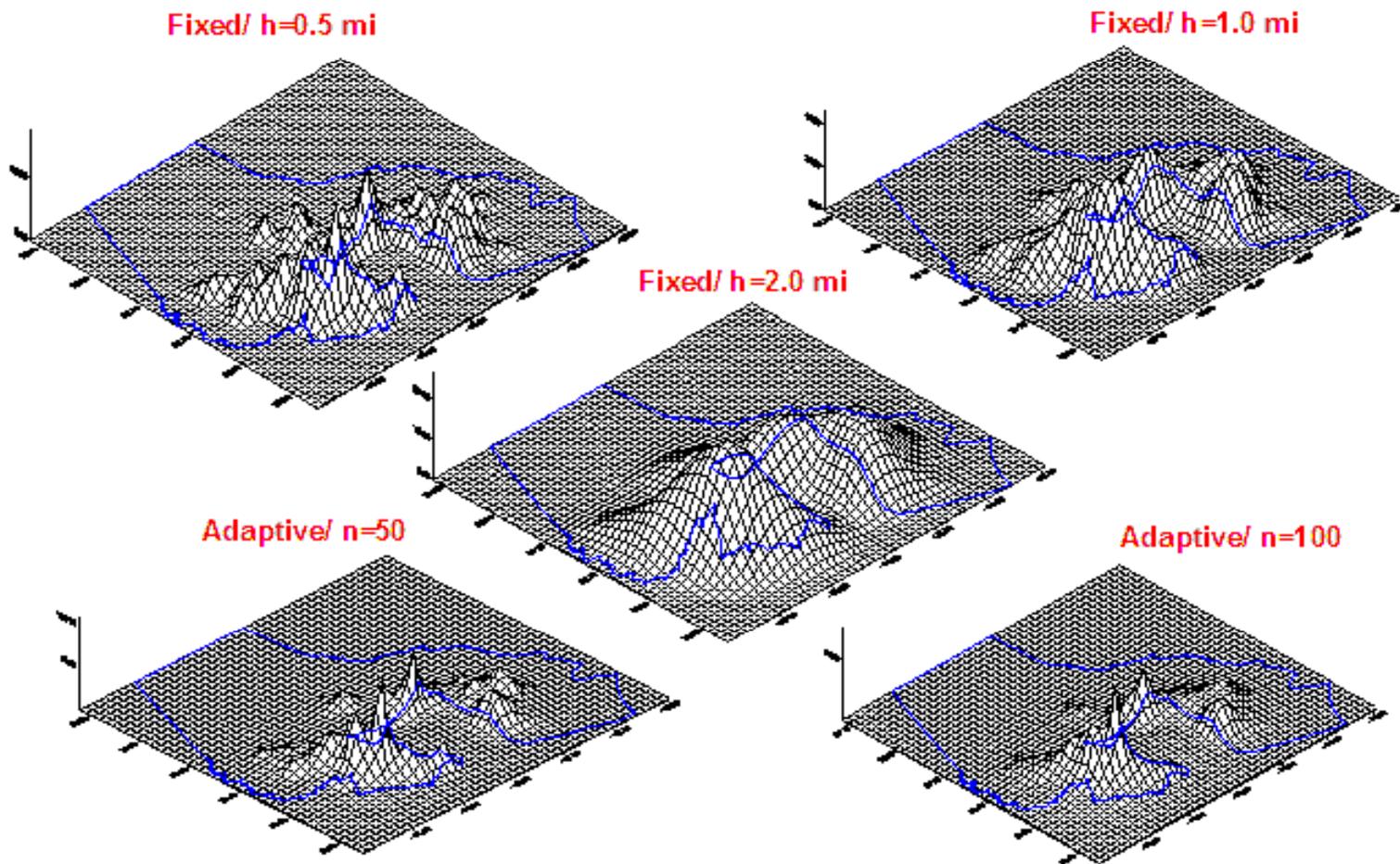
Note in all five of the interpolations, there is some bias at the edges with the City of Baltimore (the three-sided area in the central southern part of the map). Since the primary file only included incidents for the County, the interpolation nevertheless has estimated some likelihood at the edges; these are *edge biases* and need to be ignored or removed with an ASCII editor.² Further, the wider the interval chosen, the more bias was produced at the edge.

Dual Kernel Density Interpolation

The **dual kernel density** routine in *CrimeStat* is applied to *two* distributions. 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

² All the *CrimeStat* outputs except for *ArcGIS* '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 *ArcGIS* 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.

Figure 10.11:
Interpolation of Baltimore County Vehicle Thefts: 1996
Different Smoothing Parameters



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

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

Fixed interval

With a fixed bandwidth, the user must specify the interval to be used and the units of measurement (square miles, square nautical miles, square feet, square kilometers, or square 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.

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 for the kernel estimate over the entire region. Thus, in areas that have a high concentration of points, the bandwidth is narrow while in areas where the concentration of points is sparser, the bandwidth will be wider. This is the default bandwidth choice in *CrimeStat* since 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 (e.g., 200).

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 (square miles, square nautical miles, square feet, square kilometers, or square meters).

There is a good reason why a user might want to use 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 for the denominator variable 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 to reduce these spikes. For example, the interval for the first file (the numerator) could be 0.5 miles whereas the interval for the second file (the denominator) could be 1 mile. Experimentation will be necessary to see whether this is warranted. But, in our experience, excessively large densities happen when either there are too few cases or there is an irregular boundary to the region with a number of incidents grouped at one of the edges.

Use kernel bandwidths that produce stable estimates

Note that with a dual kernel calculation, particularly the ratio of one variable to another, it is important not to choose too small a 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 Unit

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

Intensity or Weighting Variable

If an intensity or weighting variable is to be used (and has been defined on the Primary or Secondary file page), the appropriate box must be checked. Be careful about using both intensity and weighting variables to avoid 'double weighting'.

Density Calculation

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*. This 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

$$\text{Log ratio of densities} = \text{Ln} [g(x_j) / g(y_j)] \quad (10.10)$$

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 & Diggle, 1995b).

3. There is the *absolute difference in densities*. This 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 variable, the difference in population between two different census years can be estimated. Since the spatial arrangements of the block groups changes slightly from one census to the next (the U. S. Census Bureau suggests that census units be drawn so that there are approximately equal populations in each unit), estimating the difference in kernel densities between two census can show where the changes have occurred irrespective of the particular census units.
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 in familiar units (square miles or square kilometers) 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*. This is the density estimate for the first file plus the density estimate for the second file. 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*. The relative sum of densities first standardizes the densities of each file by dividing by the grid cell area in familiar units (square miles or square kilometers) and then adds the secondary file relative density to 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 File

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, 2008) and *ArcGIS Spatial Analyst* format as an 'asc' file (ESRI, 2012). Third, the results can be output as polygon grids into *ArcGIS* '.shp', *MapInfo* '.mif' and various Ascii formats (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 dual 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*.

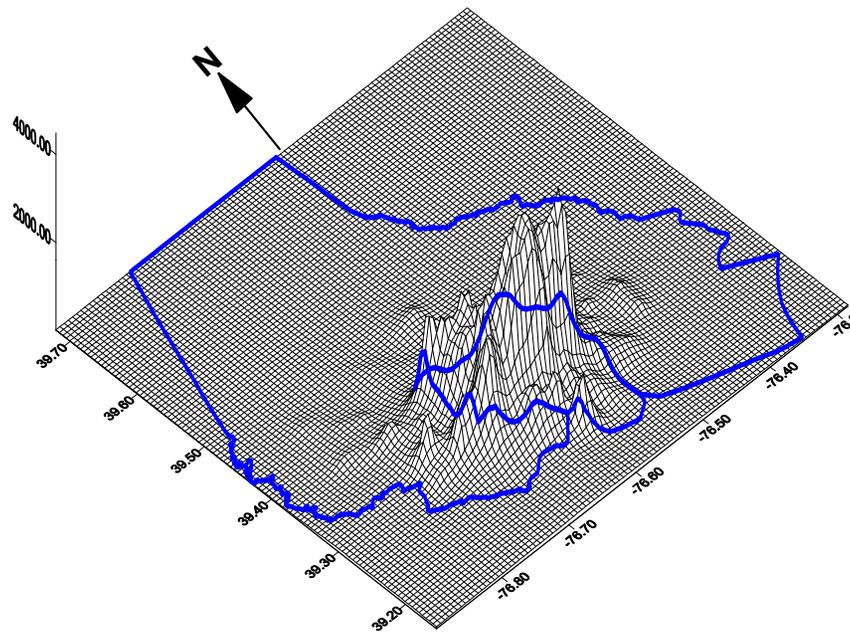
Figure 10.12 shows the resulting density estimate as a *Surfer for Windows* output; again, there is a map view, a surface view, and a contour view. The normal kernel function was used and an adaptive bandwidth of 100 points was selected. As seen, there is a very high concentration of auto theft incidents within the central part of the metropolitan area. The contour view suggests 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 10.13, 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.

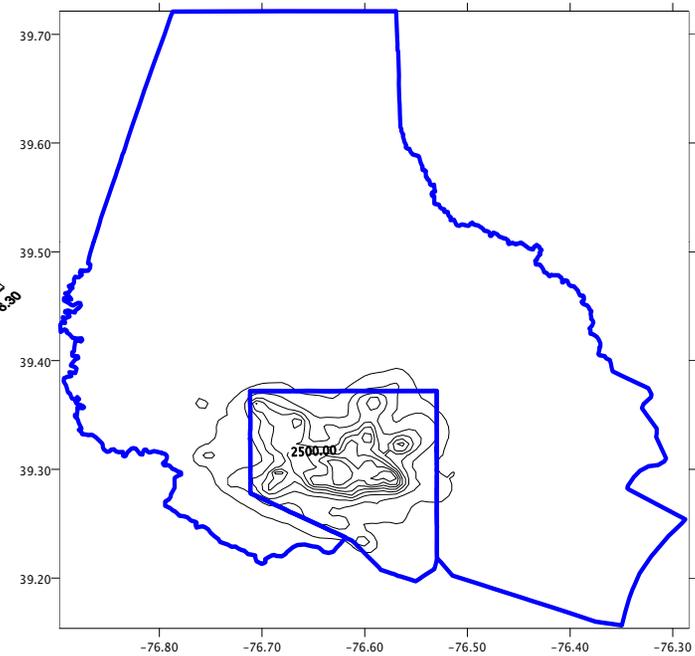
When these two kernel estimates are compared using the dual kernel density routine, a more complicated picture emerges (Figure 10.14). This routine has conducted three operations: 1) it calculated the distance between each of the 10,800 reference cells and the 14,853 auto theft locations, evaluated the kernel function for each measured distance, and summed the results for each reference cell; 2) it calculated the distance between each of the 10,800 reference cells and

Figure 10.12:
Baltimore Metropolitan Vehicle Thefts: 1996
Three Views of Kernel Density Interpolation

Surface View



Contour View



Ground Level View

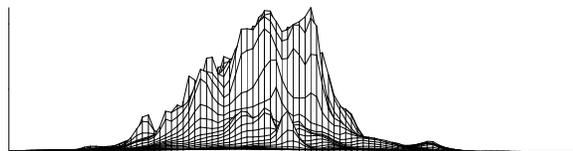
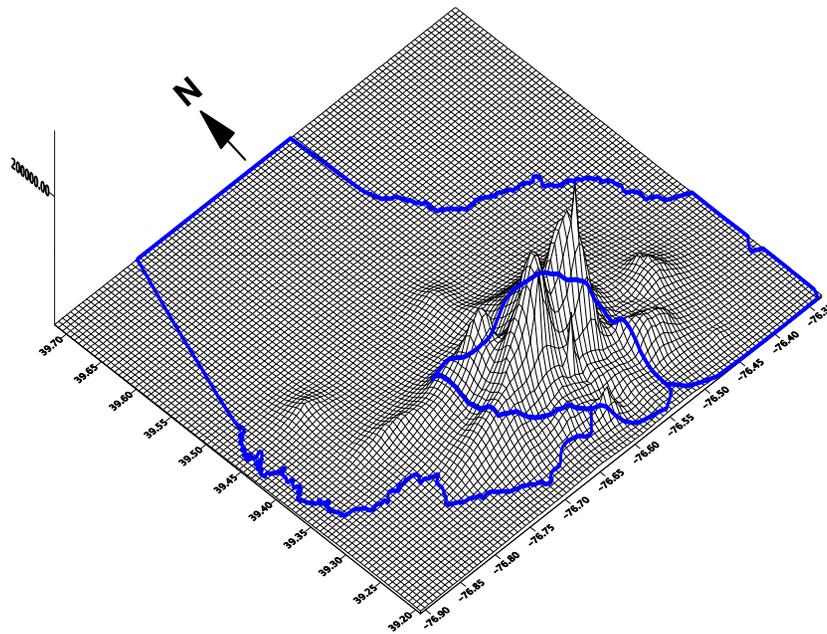
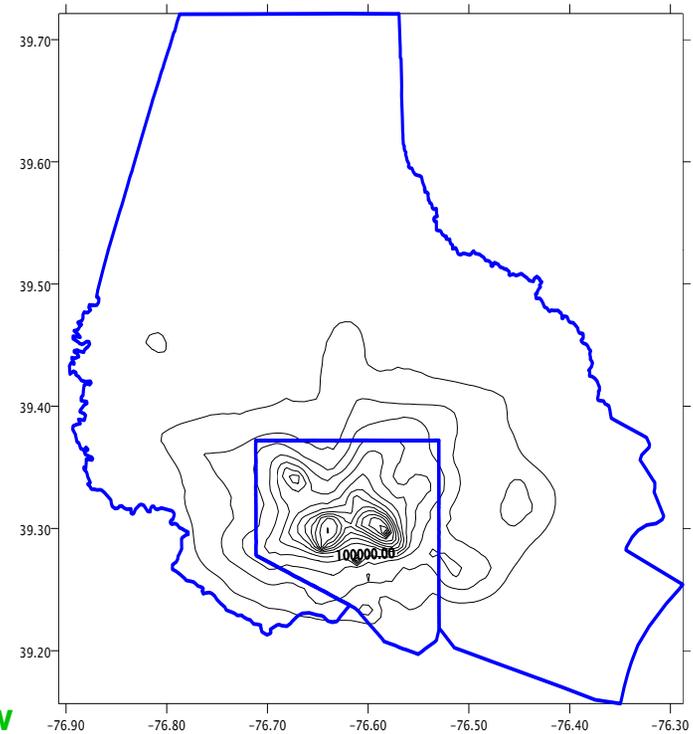


Figure 10.13:
Baltimore Metropolitan Population: 1990
Three Views of Kernel Density Interpolation

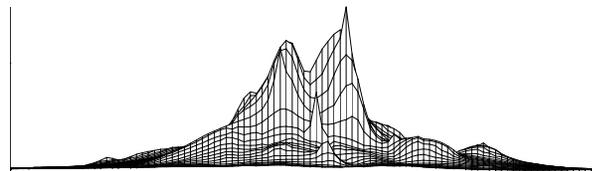
Surface View



Contour View



Ground Level View



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 that 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, & Nitz, 1995a; 1995b).

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

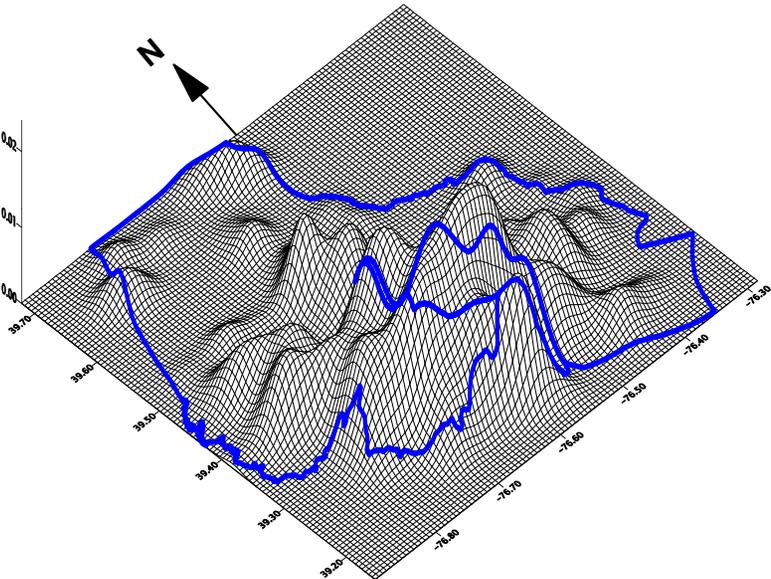
The final example shows how the dual kernel interpolation compares with the risk-adjusted nearest neighbor clustering, discussed in chapter 6. Figure 10.15 shows 15 first-order and two second-order risk-adjusted clusters overlaid on the dual kernel estimate of 1996 robberies relative to 1990 population.³ As seen, there is a correspondence between the identified risk-adjusted clusters and the dual 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.

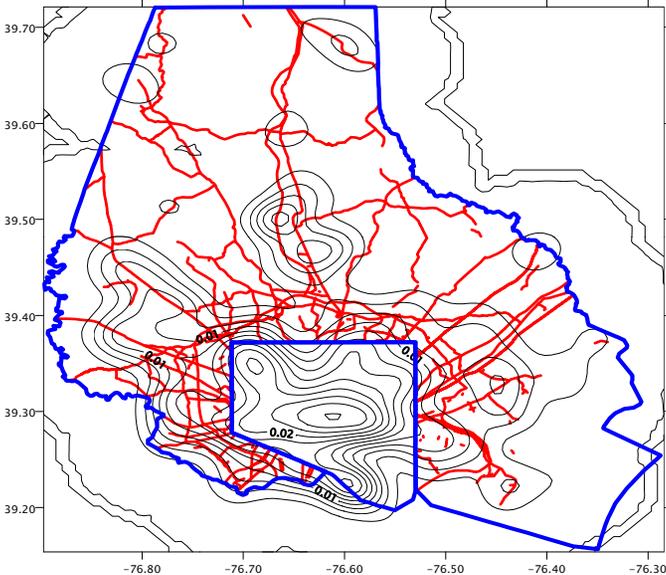
³ The risk-adjusted hierarchical clustering (Rnnh) method defined the largest search radius but a minimum of 25 points being required to be clustered. The kernel estimate for both the Rnnh and the dual-kernel routines used the normal distribution function with an adaptive bandwidth of 25 points.

Figure 10.14:
Baltimore Metropolitan Vehicle Theft Risk: 1990-1996
Ratio of Interpolation of 1996 Auto Thefts to 1990 Population

Surface View



Contour View



Ground Level View

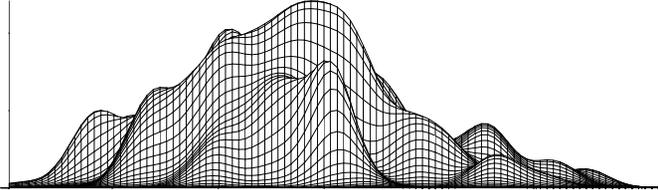
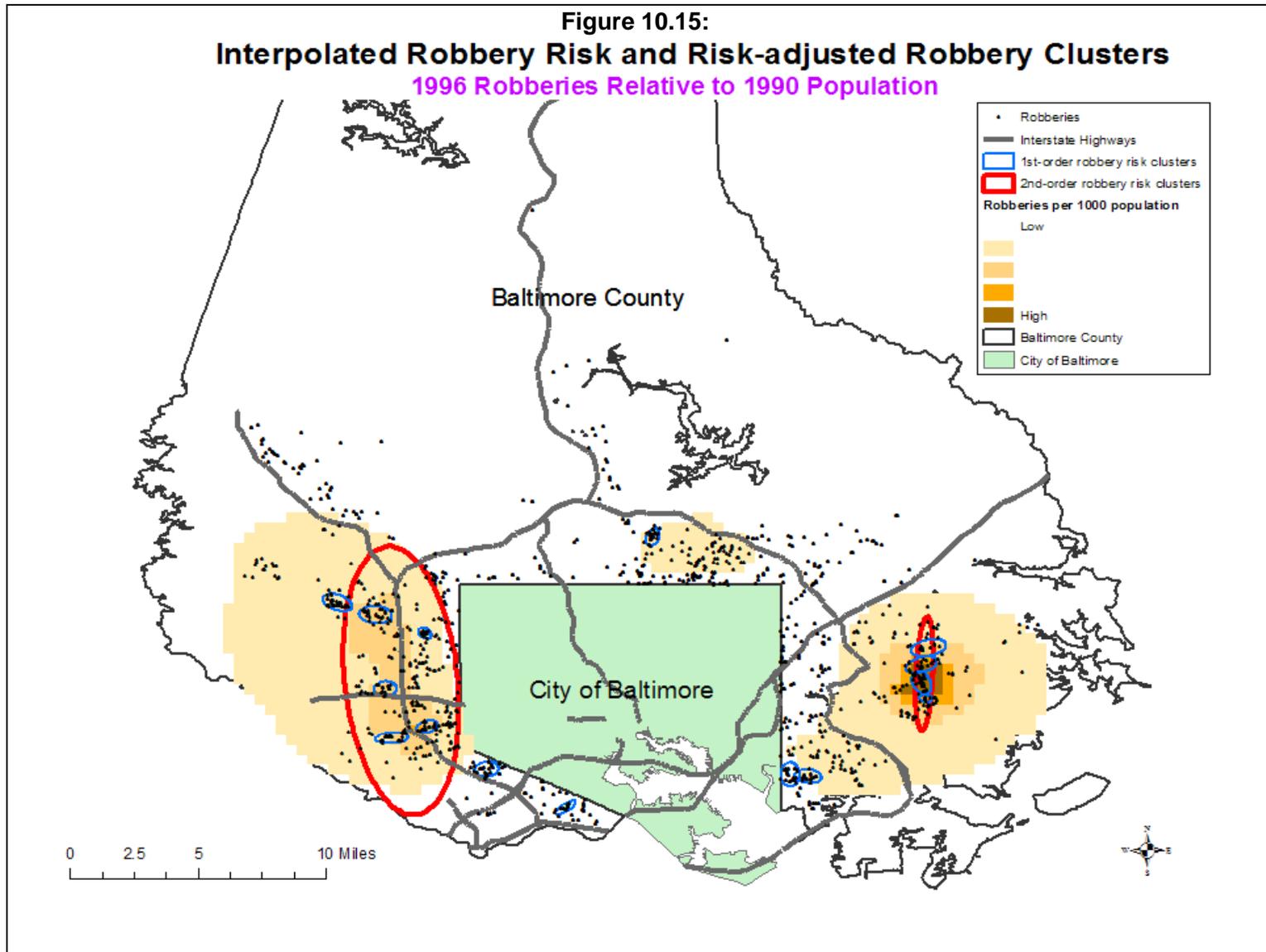


Figure 10.15:
Interpolated Robbery Risk and Risk-adjusted Robbery Clusters
 1996 Robberies Relative to 1990 Population



Visually Presenting Kernel Estimates

Whether the single- or dual-kernel estimate is used, the result is a grid interpretation of the data. By scaling these values 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.

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.

Advantages and Limitations of Kernel Density Interpolation

There are advantages and limitations to the kernel density interpolation method for hot spot analysis.

Advantages of Kernel Density Interpolation

The main advantage of kernel density interpolation is its ability to visualize a broad, regional view of events. Whereas each of the hot spot analysis techniques discussed in Chapters 7 and 8 (and 9 for the Zonal Nnh) drew boundaries around the hot spots, kernel density interpolation provides density estimates through the study area. One can see all the high density and low density areas simultaneously. For example, this can provide a police department with an overview of the high crime areas and can form the basis of patrol deployment. Essentially, for a city-wide or region-wide view, there is no better technique (Chainey, Thompson & Uhlig, 2008).

Limitations of Kernel Density Interpolation

At the same time, there are limitations to the approach for hot spot analysis. There are three statistical problems. First, the method does depend on overgeneralizing data. By interpolating N data points to M grid cells where M is almost always much greater than N , means that the data are being shared across many grid cells. This can lead to overgeneralization of results. For example, 10,000 cases seems like a large data set (which it normally is), but when it is generalized to 10,000 grids (100 columns x 100 rows), this leaves an average of 1 data point

per grid cell. It is well known that sampling error is very high with small samples and almost infinitely high with a sample size of 1. Yet the method pools the data so that every grid cell is represented by all (for the normal distribution kernel) or most (for the other kernels) data points. This leads to additional spatial autocorrelation among the estimates since each grid cell shares the same data points with adjacent grid cells. The practical effect of this is that a hot spot can appear to be larger than it truly is. Too many users are taking kernel density interpolations as evidence of hot spots even with very small samples. These hot spots will turn out to be nothing more than random variation. Again, if used carefully, the method can provide an overview of crime density in a study area. But, one has to be very, very careful in using the method to define specific hot spots.

Second, like other hot spot analysis methods, kernel density interpolation is effected by the choice of kernel used and the selected bandwidth. The normal distribution kernel, for example, will smooth the data and eliminate small nodules ('peaks and valleys') whereas the quartic and exponential kernels will emphasize the small nodules. Whether the more granular variation in density estimates is valid or not depends on the sample size. With a small sample size, small hot spots may be nothing more than random variation and may not be real. Unless the sample size is very large (meaning 10,000 or more cases), we recommend using the normal distribution kernel to avoid finding *false hot spots*. In addition, the selected bandwidth determines the smoothness of the visualization. Again, if the sample size is large, a smaller bandwidth is appropriate whereas a larger bandwidth is more desirable for smaller samples. One has to consider the *precision* of the estimates, which is a function of the sample size (i.e., larger is better).

Third, because the technique smoothes data, it is often inappropriate for small area analysis. It will lead to generalization of data points into adjacent areas from where the events occur and can lead to false conclusions (Levine, 2008). For example, motor vehicle crashes typically occur on freeways, highways, major arterial roads, and minor arterial roads. Few occur on residential (neighborhood) streets, typically less than 15%. Levine (2009) found that only 11% of motor vehicle crashes in the Houston metropolitan area occurred on local roads even though these roads accounted for 61% of the total road mileage in the region. The likelihood of a crash occurring on any particular local road is extremely small. However, since nearly one half of the crashes occurred at intersections, the method would generalize crashes at two intersecting arterial roads into the adjacent neighborhood streets when, in reality, very few crashes will occur on those streets.

Similarly, Levine (2008) showed how vehicle thefts that were concentrated in parking lots in a commercial area of Houston were generalized by kernel density interpolation into the local residential neighborhoods. In other words, the method produces spatial distortion

especially for small area (large-scale), neighborhood-level analysis. Very often, hot spots are very limited spatially, sometimes into an area less than half a block wide.

Taking another example, in a 1986 study of dangerous bus stops in Los Angeles, the most serious one was identified on one corner of a Hollywood (CA) intersection (Levine, Wachs & Shirazi, 1986). The hot spot involved a drug trade that occurred at a food stand on the corner and was supplied by drug dealers who used a bar near the intersection for cover. The crimes occurred only on that corner of the intersection; the other three corners had no crime events.

In other words, the amount of smoothing involved in kernel density interpolation will distort spatial relationships for very small hot spots and will make it appear as if there is a risk in nearby blocks when there might not be such a risk. The use of one of the cluster routines discussed in chapters 7 and 8 would be more appropriate for small area analysis.

Conclusion

Kernel density estimation is one of the most utilized 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 regional 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. Caution has to be used in adapting the method for small area (large scale) neighborhood type of analysis. Other techniques are more appropriate for that level.

References

- Anselin, L. (1992). *SpaceStat: A Program for the Statistical Analysis of Spatial Data*. Santa Barbara, CA: National Center for Geographic Information and Analysis, University of California.
- Bailey, T. C. & Gatrell, A. C. (1995). *Interactive Spatial Data Analysis*. Longman Scientific & Technical: Burnt Mill, Essex, England.
- Bowman, A. W. & Azzalini, A. (1997). *Applied Smoothing Techniques for Data Analysis: The Kernel Approach with S-Plus Illustrations*. Oxford Science Publications, Oxford University Press: Oxford, England.
- Burt, J. E. & Barber, G. M. (1996). *Elementary Statistics for Geographers* (second edition). The Guilford Press: New York.
- Chainey, S., Thompson, L., & Uhlig, S. (2008). The utility of hotspot mapping for predicting spatial patterns of crime. *Security Journal*, **21**, 4-28.
- Cleveland, W. S., Grosse, E., & Shyu, W. M. (1993). Local regression models. In John M. Chambers & Trevor J. Hastie, *Statistical Models in S*. Chapman & Hall: London.
- Cliff, A. D. & Haggett, P. (1988). *Atlas of Disease Distributions*. Blackwell Reference: Oxford.
- ESRI (2012). *ArcGIS 10.0*. Environmental Systems Research Institute: Redlands, CA. <http://www.esri.com/software/arcgis/index.html>.
- Farewell, D.I (1999). Specifying the bandwidth function for the kernel density estimator. <http://www.iph.cam.ac.uk/bugs/documentation/coda03/node44.html>.
- Golden Software. 2008. *Surfer® for Windows (Ver. 10)*. Golden Software, Inc.: Golden, CO.
- Härdle, W. (1991). *Smoothing Techniques with Implementation in S*. Springer-Verlag: New York.
- Kafadar, K. (1996). Smoothing geographical data, particularly rates of disease. *Statistics in Medicine* 15(23), 2539-2560.

References (continued)

- Kelsall, J. E. & Diggle, P.J. (1995a). Kernel estimation of relative risk, *Bernoulli*, 1, 3-16.
- Kelsall, J. E. & Diggle, P.J. (1995b). Non-parametric estimation of spatial variation in relative risk. *Statistical Medicine*, 14, 2335-2342.
- Levine, N. (2009). "A motor vehicle safety planning support system: The Houston experience". In S. Geertman and J. Stillwell, *Planning Support Systems: Best Practice and New Methods*. Springer. 93-111.
- Levine, N. (2008). "The 'hottest' part of a crime hotspot: Comments on "The utility of hotspot mapping for predicting spatial patterns of crime" by Spencer Chainey, Lisa Tompson, and Sebastian Uhlig". *Security Journal*, 21, 295-302.
- Levine, N., Kim, K. E. & Nitz, L. H. (1995a). Spatial analysis of Honolulu motor vehicle crashes: I. Spatial patterns. *Accident Analysis & Prevention*, 27(5), 663-674.
- Levine, N., Kim, K. E. & Nitz, L. H. (1995b). Spatial analysis of Honolulu motor vehicle crashes: II. Generators of crashes. *Accident Analysis & Prevention*, 27(5), 675-685.
- Levine, N., Wachs, M. & Shirazi, E. (1986). Crime at Bus Stops: A Study of Environmental Factors. *Journal of Architectural and Planning Research*. 3 (4), 339-361.
- Parzen, E. (1962). On the estimation of a probability density and mode. *Annals of Mathematical Statistics*, 33, 1065-1076.
- Rosenblatt, M. (1956). Remarks on some non-parametric estimates of a density function. *Annals of Mathematical Statistics*, 27, 832-837.
- Scott, D. W. (1992). *Multivariate Density Estimation: Theory, Practice, and Visualization*. John Wiley & Sons: New York.
- Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. Chapman & Hall: London.
- 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.

References (continued)

Venables, W.N. & Ripley, B. D. (1997). *Modern Applied Statistics with S-Plus (second edition)*. Springer-Verlag: New York.

Whittle, P. (1958). On the smoothing of probability density functions. *Journal of the Royal Statistical Society, Series B*, 55, 549-557.

Endnotes

- i. 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, $\text{sd}(D)$ is the standard deviation of distance, and $\text{iqr}(D)$ is the inter-quartile range of distances between points. These would be substituted for MeanZ , $\text{sd}(Z)$, and $\text{iqr}(Z)$ respectively

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

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

where \min is the minimum of the next two terms, $\text{sd}(Z)$ is the standard deviation of the variable, Z , being interpolated, $\text{iqr}(Z)$ 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:

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

To avoid being influenced by outlier, they suggested using the median absolute deviation estimator for $\text{sd}(Z)$:

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

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

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

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

$$h = 0.68N^{-\frac{1}{5}}$$

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 theory (Bailey & 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:

$$95\% \text{ confidence interval} = \bar{Z} \pm 1.96 \frac{0.5}{N(h)^{\frac{1}{2}}} \text{sd}(Z)$$

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 criterion requires that the bandwidth be increased until it captures the specified number of points.

Endnotes (continued)

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:

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

and the average bandwidth, $Mean(h)$, is:

$$Mean(h) = \sqrt{\frac{A(p)}{\pi}} = \sqrt{\frac{AN(p)}{N\pi}}$$

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).

Attachments

Kernel Density Interpolation to Estimate Sampling Bias in the Climatic Response of *Sphagnum* Spores in North America

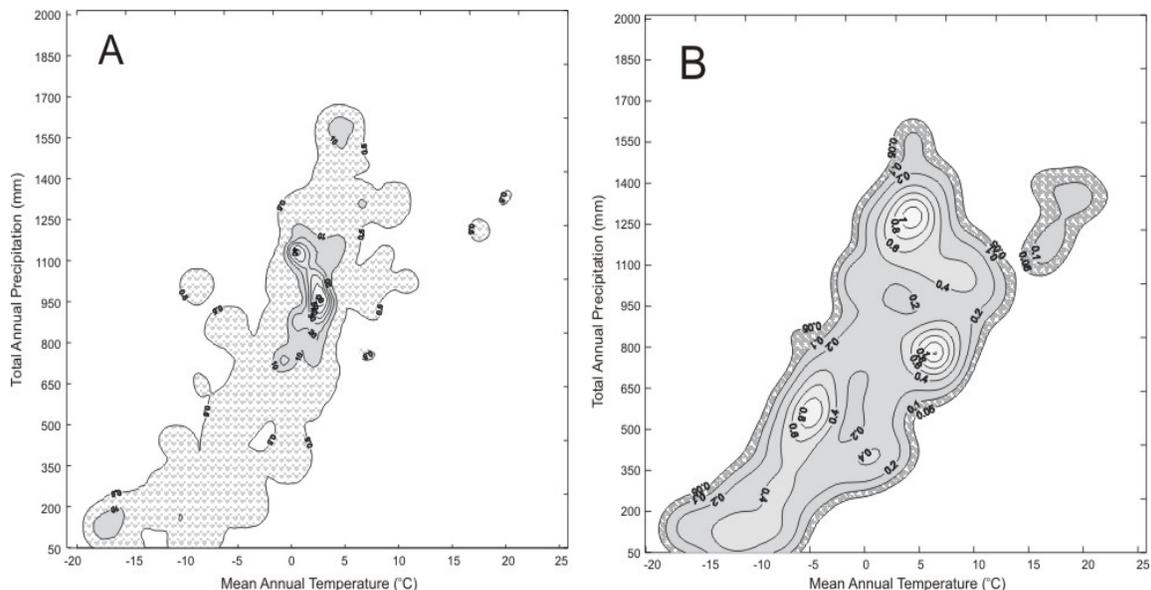
Mike Sawada

Laboratory for Applied Geomatics and GIS Science
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 in Belo Horizonte

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

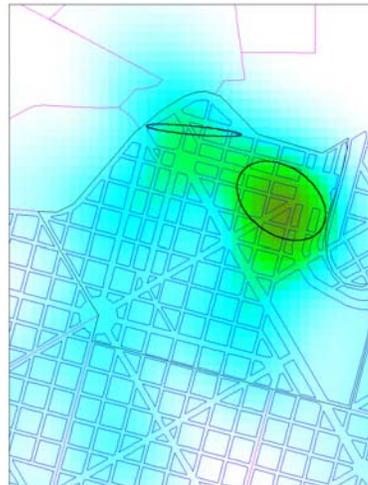
We used the kernel density estimate to visualize time trends for crime occurrences on a typical weekday. We found markedly different spatial distributions depending on the time, with the amount of crime varying and the hot spots, identified by the ellipses, appearing in different places.

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.

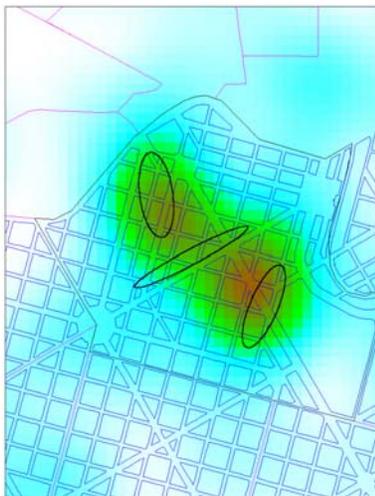
9:00 AM



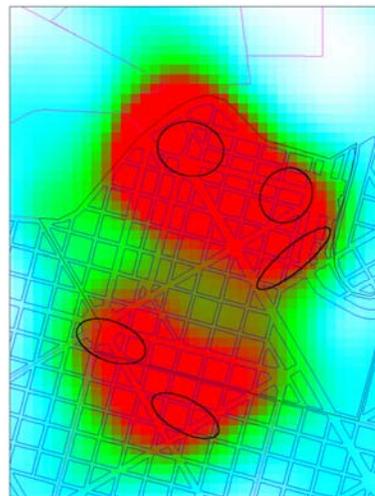
1:00 PM



7:00 PM



11:00 PM



Legend

- High density
- Medium density
- Low density
- Very low density

Using Kernel Density Smoothing and Linking to *ArcGIS*: Examples from London, England

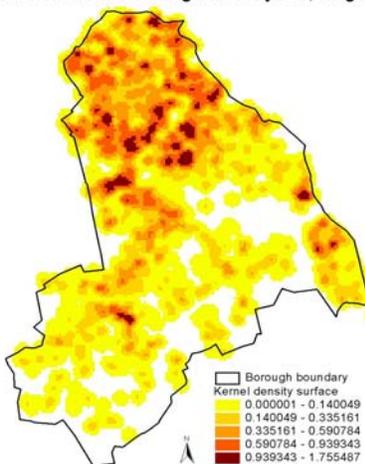
Spencer Chainey
Jill Dando Institute of Crime Science
University College
London, England

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 arbitrary 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 *ArcGIS*.

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.

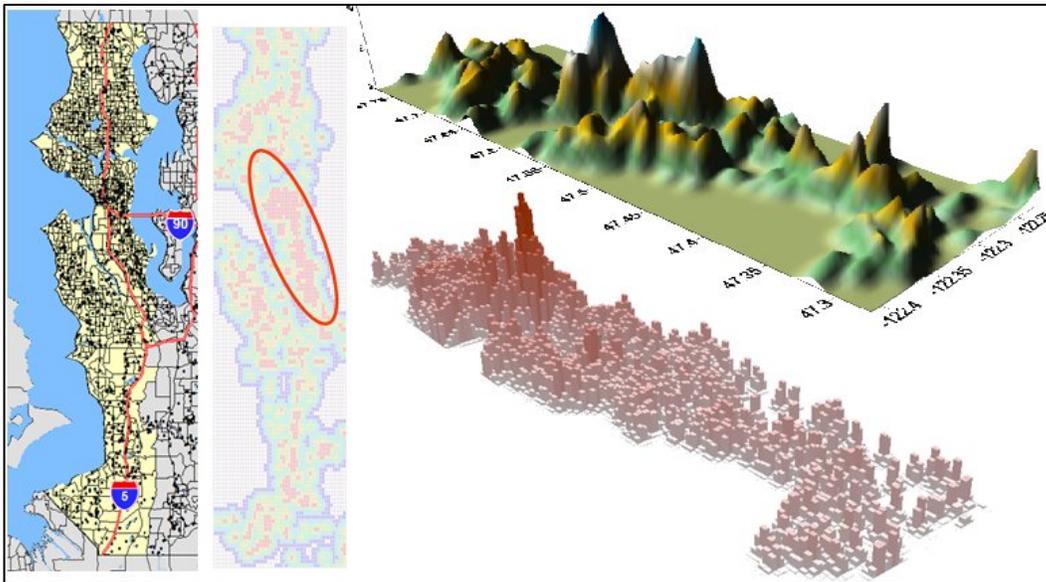


Infant Death Rate and Low Birth Weight in the I-5 Corridor of Seattle and King County

Richard Hoskins
Washington State Department of Health
Olympia, Washington

Although the infant death rate (< 1 year old) has been steadily declining in Washington, the incidence of low birth weight (< 2500 gms) is increasing. This is a significant public health problem, resulting in suffering and high medical cost. If we know where the rates are high at a neighborhood level we can develop more efficient and effective programs. The goal is to determine regions where rates are clustered and to characterize those regions with respect to SES variables from the US Census.

Birth and infant death data were geocoded to the street level. In order to detect clusters of high infant death *and* low birth weight, several *CrimeStat* tools were used. We find that using several tools at once helps detect regions where something untoward is going on and also helps develops guesses about where other problems might be expected develop.



I-5 corridor in
King County

Kernel density
interpolation

Top: 3-D map: empirical Bayes rate
Bottom: Prism map: SMR

The result of a kernel density interpolation using a normal estimator is shown above along with an empirical Bayes rate and standardized mortality ratio (SMR) calculated in SAS and mapped in Maptitude (www.caliper.com). Starting with over 2,500 infant deaths, about 25,000 low weight births (out of over 500,000 live births) occurred in the Seattle I-5 corridor region in King County from 1989-2002. The kernel density method was used to detect high rate regions. A clearly articulated region and ridge appears on the grid of the kernel density map and the 3D and prism maps.

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

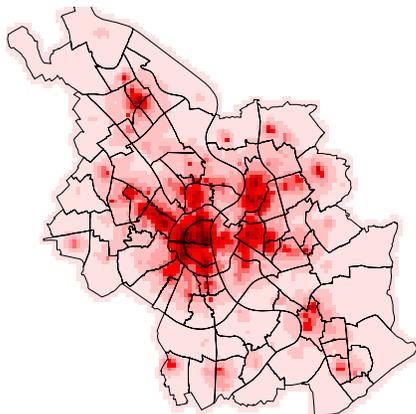
When estimating the density of street crimes within a metropolitan area by interpolating crime incidents, the result is usually a very high concentration in the city center. However, there is also a very high concentration of people either living or pursuing their daily routine activities in these areas. 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* dual 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'.

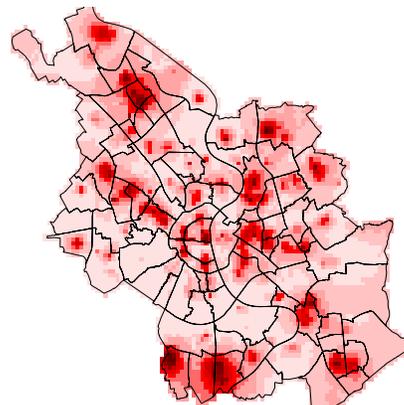
In the dual kernel routine, the density estimate of crime incidents is compared to the density estimate of the population at risk, defined by the centroids of census tracts with the number of persons as an intensity variable. We chose the normal method of interpolation and adaptive intervals with a minimum 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)



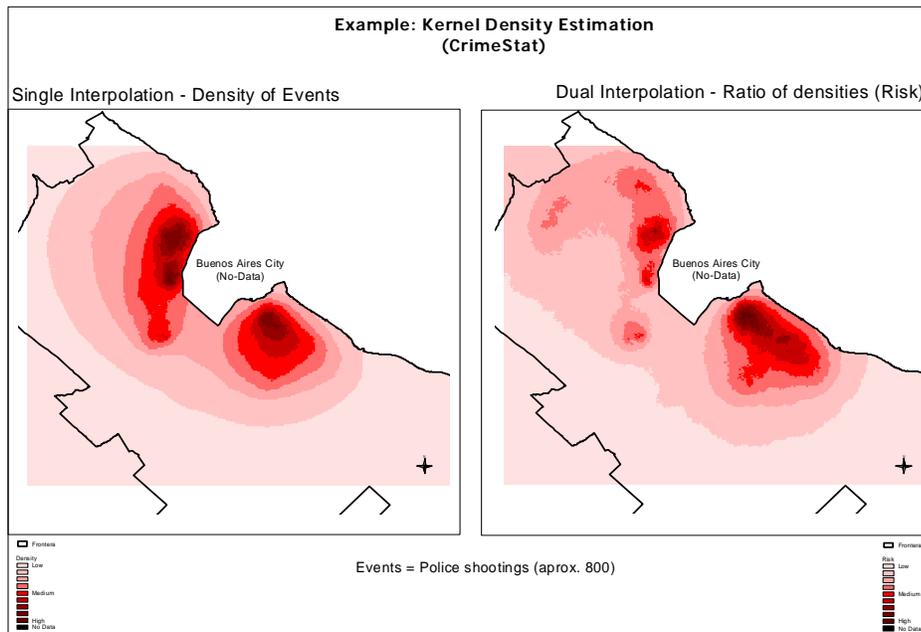
Dual kernel density of crime incidences
relative to population at risk



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 dual 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.\



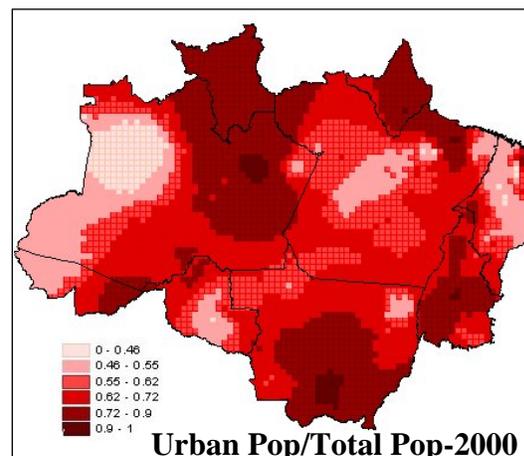
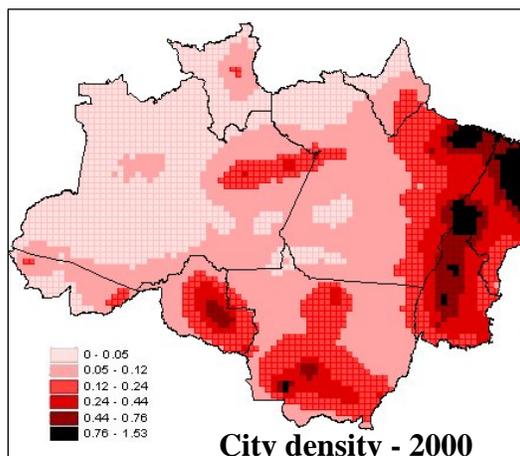
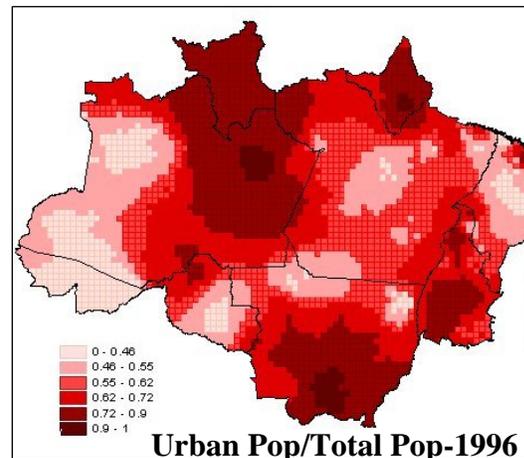
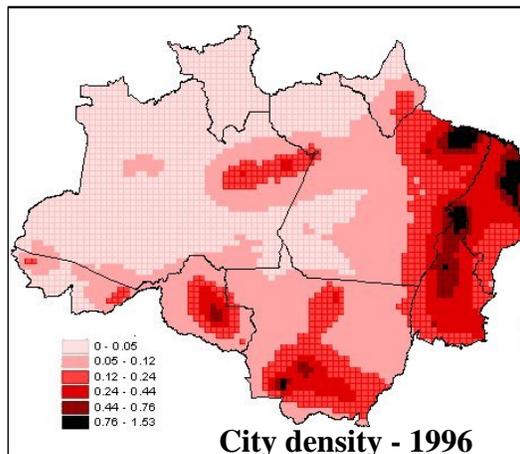
Both images are quite different, suggesting varying policing strategies. For example, though there are two well-defined hot spot areas in the Province (one in the north, the other in the south), the high levels of risk detected in the southern areas came as a complete surprise. The northern area has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

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 dual 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 with federal 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.



Using Small Area Estimation to Target Health Services in Harris County, TX

Thomas F. Reynolds, MS
University of Texas-Houston School of Public Health

In Texas, the City of Houston and Harris County organized a Public Health Task force to make recommendations concerning the provision of health services for those without health insurance. Task force members wanted to know approximately how many area citizens did not have health insurance.

Data from the two most recent Current Population Survey Annual Social and Economic Supplements (CPS-ASEC, 2003-04) were used to derive a synthetic estimate using a stratified model. Estimates were calculated at census tract and block group levels. Selected political divisions were clipped from base maps for political officials and legislators.

Percentages are indicative of risk. On the other hand, numbers are essential for targeting physical resources. There is seldom a perfect correspondence between high percentages and large numbers. For example, an area with a concentration of multi-family housing may have a relatively small percentage, but a large number, of uninsured. Percentage maps of the uninsured (figure 1) are generally clustered and informative; however, due to large variations in population numbers at both levels of census geography, maps of the population densities of uninsured proved most valuable to officials (figure 2).

CrimeStat was used to develop the density maps. The single kernel density routine was used to estimate the density of block group values using the centroid to represent the values and the number of uninsured as an intensity value. The Moran Correlogram was used to select the type of kernel for the single-kernel interpolation (a uniform distribution) and an optimal bandwidth.

Fig. 1: Percent Uninsured

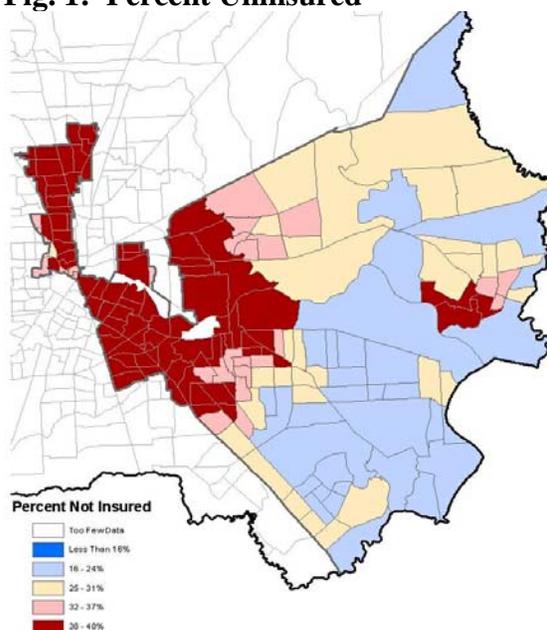
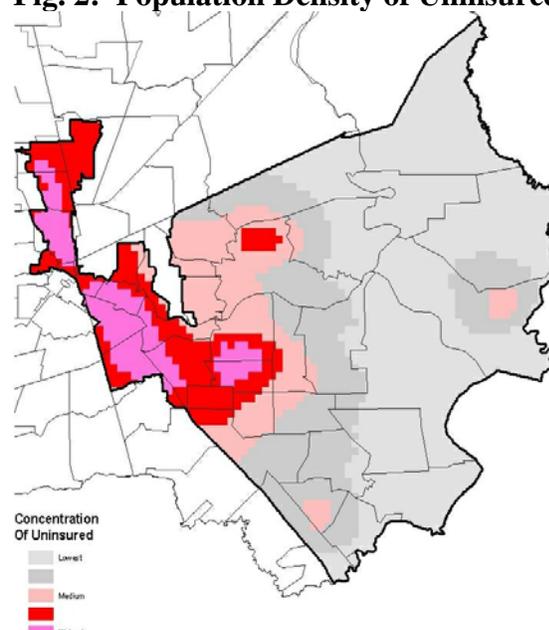


Fig. 2: Population Density of Uninsured



Identifying Voucher Holder and Crime Concentrations using Dual Kernel Density Estimation

Ron Wilson, U.S. Department of Housing and Urban Development

Public housing authorities and law enforcement are cooperating to reduce neighborhood crime where Housing Choice Voucher Program (HCVP) participants concentrate. When police departments and housing authorities can identify geographically combined voucher holder and crime concentrations, more specific strategies can be employed to reduce those crimes and prevent victimization. Aggravated assaults are common in and around neighborhoods where HCVP households concentrate. I used the relative sum of densities option of the Dual Kernel Density Estimation routine in *CrimeStat IV* to identify where HCVP households and aggravated assaults were concentrated in 2010.

Several areas of voucher holder and aggravated assault concentrations are revealed with gradations in density, some in census tracts with high poverty. The Dallas Police Department might deploy varying community policing approaches in these areas based on concentration grade to reduce assault opportunities while building relationships with neighborhood residents. The Dallas Public Housing Authority may help voucher holders find safer neighborhoods to relocate outside the concentrated areas, in particular to areas with low poverty. These findings may also help Dallas city officials craft separate place-based polices that work to eliminate the root causes of aggravated assaults in these areas.

