CHAPTER

9

Models for spatial data

9.1 Introduction

Location information is essential for linking crashes with site-specific data, analyzing crash patterns, and identifying crash hot spots. Traditional crash prediction models assume that crash events are independent of each other in space (see Chapter 4—*Crash-Frequency Modeling*). This aspatial assumption may hold true for crashes occurring within a small area where geographical conditions are relatively homogeneous; however, the same cannot be said for a county- or a state-wide analysis. Weather, animal population, enforcement discretion, driver behavior, and other spatial phenomena have a great impact on crash occurrence. Ignoring spatial dependence can lead to serious errors in statistical analysis for crash data because features lying in space influenced by geographical factors are bound to exhibit some sort of spatial dependencies (Getis and Ord, 1992; Ord and Getis, 1995; Cressie, 1993).

Spatial dependence suggests values at one location is influenced by values in their geographical proximity. As described in the Tobler's First Law of geography: "*everything is related to everything else, but nearby things are more related than distant things*" (Tobler, 1970). The relative proximity among many pairs of spatial data features can be expressed as a spatial weights matrix, often denoted by **W**. This matrix defines and measures the spatial influence among geographic objects as a function of adjacency or (inverse) distance. Typical distance measurements include the straight-line measure (or Euclidean distance) on the plane with constant elevation, the great-circle distance (or orthorhombic distance) on the surface of a sphere, and the travel distance measured from a street network. Measuring and modeling of spatial dependencies in conjunction with geographic information systems (GIS) can help analyze spatial patterns in crash data as well as help improve modeling procedures and coefficient estimates with an explicit spatial structure in statistical models.

https://doi.org/10.1016/B978-0-12-816818-9.00009-3

This chapter is dedicated to analyzing and modeling crash data within a spatial context. The chapter begins with an overview of the characteristics of spatial data and commonly used data models. Then, spatial indicators such as Getis G and Moran's I are introduced to help determine the distribution of crash locations as clustering, dispersed, or random. Next, specific techniques for analyzing crash point data are presented to facilitate the discovery of the underlying process that generates these points. Finally, spatial regression methods are introduced to explicitly consider the spatial dependency of crash observations and spatial heterogeneity in the relationship between crashes and their contributing factors.

9.2 Spatial data and data models

Spatial data identify the geographic location of features, boundaries, and other geographic phenomena on the surface of the Earth. Spatial data are usually recorded by coordinates, pixels, and typology. The main spatial data types are vectors and rasters.

Points, lines, and polygons, collectively, are called vector data. In the vector model, the physical representation of these features includes two components: one that describes the location of the feature and one that describes the characteristics (i.e., attributes) of the feature. A variant of line data is network data. Network data include a set of connected lines in which the first and last vertices in the line represent junctions in the network.

Other data types such as elevation, temperature, and rainfall precipitation have no distinct shape. Instead, they can be measured for any location on the Earth's surface. Such geographic phenomena are better represented as surfaces than as shapes. The most typical surface is a raster, which is a matrix of identically sized square cells. Each cell represents a unit of surface area and contains a measured or estimated value for that location. The raster model stores only the data values and does not include the location information pertaining to the position of individual grid cells. The data values are ordered either by column within row or by row within column. Fig. 9.1 illustrates the vector and the raster data model, respectively.

9.3 Measurement of spatial association

Spatial series data possess certain patterns that may be a result of the concentration of weighted points or the areas represented by weighted

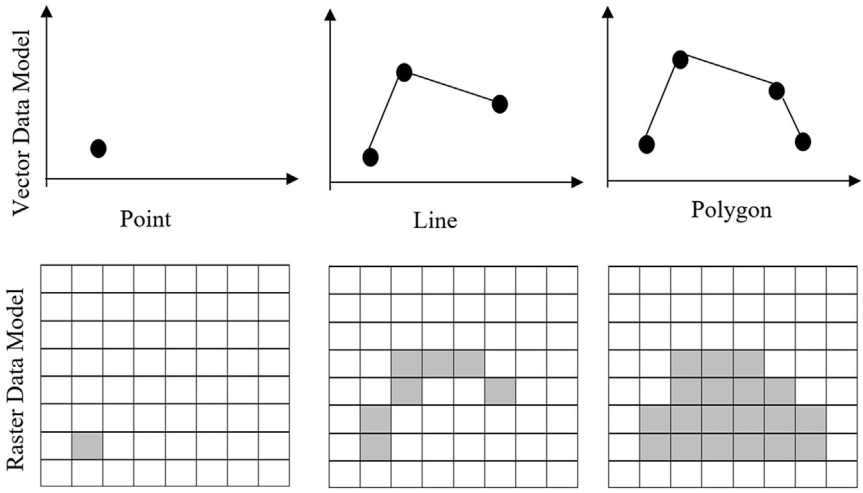

FIGURE 9.1 Vector data model and raster data model.

points. Many methods have been developed to characterize the structure embedded in spatially referenced data and the strength of the correlation using the statistical concepts and methods. Measurement can be taken globally or locally. A global measure provides the overall trend for the entire region under study, but sometimes it is beneficial to examine patterns at a local level, particularly if the pattern generating process is varying over space. This section introduces the most representative statistics for measuring spatial association.

9.3.1 Global statistics for spatial association

This section describes the two most popular global statistics.

9.3.1.1 Getis–Ord general $G^*(d)$

Spatial association or specifically, spatial autocorrelation, can be measured by several best-known spatial statistics, including Getis–Ord General G^* or simply $G^*(d)$ and Moran's I or simply $I(d)$. Getis and Ord (1992) proposed $G^*(d)$ as a global statistic to measure the concentration of the high or low values for an entire study area as a function of distance d. Assume a study area is divided into n subareas, each subarea i ($i = 1, 2, \ldots, n$) is identified with its centroid associated with a value x (a weight or

attribute) taken from a variable X. $G_i^*(d)$ for subarea i and its related statistics $G_i(d)$ are defined as follows:

$$G_i^*(d) = \frac{\sum_{j=1}^{n} w_{ij}(d)x_j}{\sum_{j=1}^{n} x_j}, \forall j \tag{9.1a}$$

$$G_i(d) = \frac{\sum_{j=1}^{n} w_{ij}(d)x_j}{\sum_{j=1}^{n} x_j}, \; j \text{ not equal to } i \tag{9.1b}$$

where w_{ij} is a symmetric 0−1 spatial weights matrix with the value of one for all subareas defined as being within distance d of a given subarea i; all others are 0, including the i itself. The concept of w_{ij} is to construct a matrix that reflects the assumption about the spatial phenomenon of interest. The matrix can have different forms and the selection of spatial weights matrix should be guided by theory about the spatial phenomenon in study.

$G_i^*(d)$ and $G_i(d)$ measure similar spatial phenomenon and the difference is denoted by "*," inclusion of i itself, implying that any concentration of the x values includes the x at i (Getis and Ord, 1992). For brevity, only $G_i^*(d)$ is discussed here. As $G_i^*(d)$ is a proportion of the sum of all x_j values that are within distance d of i, $G_i^*(d)$ is high if high value x_js are within d of i, and $G_i^*(d)$ is low if low value x_js are within d of i. A more general statistic can be defined based on all pairs of values (x_i, x_j) if i and j are within the distance of d of each other. When no particular location i is fixed, $G^*(d)$ is specified as:

$$G^*(d) = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(d)x_i x_j}{\sum_{i=1}^{n}\sum_{j=1}^{n} x_i x_j}, \; \forall j \tag{9.2}$$

The $G^*(d)$ statistic measures the concentration or lack of concentration of all pairs of values (x_i, x_j) within d of each other. The statistical significance of $G^*(d)$ depends on its distribution that can be approximated as a normal distribution. If the absolute value of the Z score of $G^*(d)$ is greater than a predetermined value, strong spatial association or clustering is present. A positive Z score means that high values cluster together, and a negative Z score means that low values cluster together.

9.3.1.2 Moran's I

Given a set of geographic features and their associated attributes in the study area, Moran's I evaluates whether the pattern is spatially clustered,

dispersed, or random on a global scale in a study area. Moran (1950) developed Moran's I in Eq. (9.3) to measure the correlation of each x_i with all neighboring x_js, including itself.

$$I(d) = \frac{\sum_i \sum_j w_{ij}(x_i - \overline{x})\left(x_j - \overline{x}\right)}{W\sum_i (x_i - \overline{x})^2/n} \tag{9.3}$$

where $\overline{x}$ is the mean of x; w_{ij} is a matrix of spatial weights with zeros on the diagonal (i.e., $w_{ii} = 0$), and W is the sum of all w_{ij}, $W = \sum_i \sum_j w_{ij}$. Distance d is used to determine the neighbors j. As a correlation statistic,[1] values of $I(d)$ range from -1 to $+1$. A Moran's Index value near $+1$ indicates a clustering pattern while an index value near -1 indicates a dispersed pattern.

9.3.2 Local indicators of spatial association

This section covers the two most popular local indicators.

9.3.2.1 Local $G_i^*(d)$

Global statistics that measure overall association may dismiss significant spatial clustering at the local level. Alternative statistics, or local indicators of spatial association (LISA), have been introduced to help detect local clusters. LISA assesses the significance of local statistics at each location, to identify the locations of spatial clusters and spatial outliers irrespective of the presence of a global spatial association.

Ord and Getis (1995) developed the local version of $G^*(d)$. If G_i is redefined as a standard variate by taking the statistic minus its expectation and then divide by the square root of its variance, the resulting $G_i^*(d)$ statistic can be specified as follows:

$$G_i^*(d) = \frac{\sum_{j=1}^{n} w_{ij}(d)x_j - \overline{x}\sum_{j=1}^{n} w_{ij}(d)}{S\sqrt{\dfrac{N\sum_{j=1}^{n} w_{ij}^2(d) - \left(\sum_{j=1}^{n} w_{ij}(d)\right)^2}{N-1}}}, \ \forall j \tag{9.4}$$

[1] Moran's I equation bears a high resemblance to the Pearson correlation coefficient, $\rho = \dfrac{\sum_i (x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_i (x_i - \overline{x})^2}\sqrt{\sum_i (y_i - \overline{y})^2}}$ because it is a correlation statistic. It is the slope $\left(\widehat{\beta}\right)$ of a simple linear regression $y = \alpha + \beta x$ where y is the weighted neighbors, that is, x_js, of x.

with $\overline{x} = \frac{\sum_{j=1}^{n} x_j}{n}$ and $s = \sqrt{\frac{\sum_{j=1}^{n} x_j^2}{n} - (\overline{x})^2}$ where all the terms are the same as previously defined for $G^*(d)$. This standardized G calculates a single Z score value for each location in the study area. A positive Z value indicates the clustering of high attribute value locations, and a negative Z value indicates the clustering of low attribute value locations. Thus, $G_i^*(d)$ statistics are often used for hot spot/cold spot analysis. The underlying theory is that a feature with a high value is interesting, but it must be surrounded by other features with high values to be qualified as statistically significant. A Z score is statistically significant when the local sum (i.e., $\sum_{j=1}^{n} w_{ij}(d)x_j$) is very different than the expected local sum (i.e., $\overline{x}\sum_{j=1}^{n} w_{ij}(d)$) and the difference is too large to be a random chance.

9.3.2.2 *Local Moran's* $\mathbf{I_i}$

Anselin (1995) modified Moran's I for the local Moran's I index as follows:

$$I_i = \frac{(x_i - \overline{x})}{\sum_i (x_i - \overline{x})^2 / n} \sum_j w_{ij}\left(x_j - \overline{x}\right) \tag{9.5}$$

Note that $\sum_i I_i = nI$. Therefore, global Moran is the average of local Moran statistics.

The Z-score and p-value represent the statistical significance of the I value. A positive value for I indicates a clustering pattern, meaning the feature is surrounded by features with similar values. A negative value for I indicates an outlier, meaning the feature is surrounded by features with dissimilar values. Hence, the local Moran's I can help identify the cluster of high values (HH), cluster of low values (LL), an outlier in which a high value is surrounded primarily by low values (HL), and an outlier in which a low value is surrounded primarily by high values (LH).

Both $G_i^*(d)$ and $I_i(d)$ can measure the association among the set of weighted points or areas represented by points, but they are different in the formulation and may yield different results. $G_i^*(d)$ measures the concentration or lack of concentration of all pairs of (x_i, x_j) such that i and j are within d of each other. $I_i(d)$ is used to measure the correlation of each x_i with all x_js within d. This difference means that G statistics are useful when only positive spatial autocorrelation is of interest (i.e., hot spots (clustering of high values) or cold spots (clustering of low values)), whereas Moran's I identifies both spatial clusters and outliers. Table 9.1 shows the $G_i^*(d)$ and $I_i(d)$ under various situations.

TABLE 9.1 Standard normal variates for G(d) and $I(d)$.

Situation	Z(G)	Z(I)
HH	++	++
HM	+	+
MM	0	0
Random	0	0
HL	−	−−
ML	−#	−
LL	−−	++

#, tends to be more negative than HL; +, moderate positive association; ++, strong positive association (high positive Z scores); *0*, no association; *HH*, pattern of high values of *x*s within *d* of other high *x* values; *L*, low values; *M*, moderate values; *H*, high values; *Random*, no discernible pattern of *x*s.
Table is adapted from Getis, A., Ord, J.K. 1992. The analysis of spatial association by use of distance statistics. Geogr. Anal. 24 (3), 189–206.

Example 9.1

Analyze Spatial Patterns of Weather-related Crashes.

In Khan et al. (2008), 3 years of Wisconsin crash data (2000–02) were collected, and snow-related crashes (i.e., crashes reported in snowy weather conditions and/or roads covered with snow) are aggregated by 72 Wisconsin counties. $G_i^*(d)$ is used to analyze patterns of snow-related crashes at a county level, and d was a fixed distance band of 75,000 m with row-standardized weights (sum of row quantities in spatial weight matrix equals one). The fixed distance band was selected because the minimum distance at which all counties had at least one neighbor was calculated to be 71,910 m.

The results of the $G_i^*(d)$ analysis of snow-related crashes present a spatial clustering of high and low attribute value (relative crash rates for snow-related type crash) locations in different regions of the state. Each county is represented by a standardized Z-score value in four categories as calculated by the G_i^* statistic. The Z-score values above +2 represent counties that lie in a cluster of the high attribute value (relative crash rate) area, statistically significant at the 5% significance level. The reason for spatial clusters of snow-related crashes in the northern region is likely due to the fact that northern counties in Wisconsin experience more snowfall and snowstorm events. On the other hand, counties with a Z-score between +2 and −2 represent locations that may have a high or low relative crash rate value but are not part of a statistically significant spatial pattern or cluster. See (Khan et al., 2008) for details.

9.4 Spatial weights and distance decay models

The spatial weights matrix (or spatial weighting matrix, weighting factor) is the proximity measure that determines the influence of site j on site i where $i \neq j$. The measurement can be either adjacency-based or distance-based. In some textbooks, adjacency-based is also called contiguity-based. The rule of thumb is that the adjacency-based measure is more common for area or zonal variables, and the distance-based measure is more common for point data. When the rule is relaxed, the concept of adjacency can be extended for point data based on the distance d_{ij} against a predetermined value; or the concept of distance can be applied to zonal variables in which the distance is measured from zonal centroid to centroid.

The weights express the neighbor structure between the observations as an $n \times n$ matrix $\mathbf{W}$ denoted as, $\boldsymbol{W} = \begin{bmatrix} w_{11} & \cdots & w_{1n} \\ \vdots & \ddots & \vdots \\ w_{n1} & \cdots & w_{nn} \end{bmatrix}$, in which the elements w_{ij} of the matrix are the spatial weights. The spatial weights w_{ij} are nonzero when i and j are neighbors and are zero otherwise. The self-neighbor relation is usually excluded so that the diagonal elements of W are zero, $w_{ii} = 0$.

In its simplest form, the spatial weights matrix expresses the presence of a neighbor as a binary relationship with weights 1 and 0. In a row-standardized form, it takes the given weights w_{ij} and divides them by the row sum: $w_{ij(s)} = \frac{w_{ij}}{\sum_j w_{ij}}$. Hence, each row sum of the row-standardized weights equals to one and the sum of all weights $\sum_i \sum_j w_{ij}$ equal to the number of observations.

Contiguity weights: The most common neighboring relation is contiguity, which means the two spatial features share a common border of nonzero length. Contiguity can be further distinguished by the queen and the rook criterion as shown in Fig. 9.2. The queen criterion defines neighbors as features sharing a common edge or a common vertex. The rook criterion defines neighbors by the existence of a common edge between two features.

Distance-band weights: adjacency relation can be constructed from distance as follows:

$$w_{ij}(d_{ij}) = \begin{cases} 1, & d_{ij} < d \\ 0, & otherwise \end{cases} \tag{9.6}$$

where d is a predetermined cutoff value. When d is infinite, all data points are used for calibrating point i, and the weight is unity.

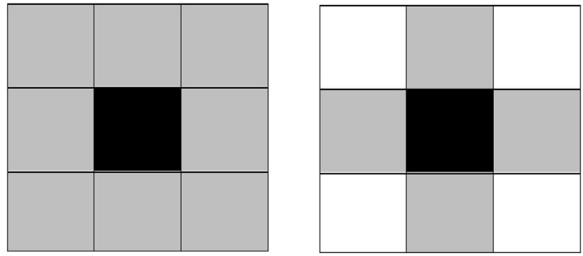

FIGURE 9.2 Contiguity case: (A) Queen's case contiguity; (B) Rook's case contiguity. (■ central cell; □ cells excluded; and other cells included in calculation).

The adjacency-based measure might cause the issue of discontinuity and abrupt change along the border. Rather than expressing spatial influence as a binary value based on adjacency, the spatial weight is often expressed as a continuous value using a distance decay function. The distance decay model measures how fast the strength of spatial dependence decays as distance increases. The common functional forms include the following:

- Inverse distance weighting:

$$w(d_{ij}) = \frac{1}{d_{ij}^{k}} \tag{9.7}$$

- A generalized powered exponential family:

$$w(d_{ij}) = \exp\left[-(\varphi d_{ij})^{k}\right], k \in [0,\ 2]; \varphi > 0 \tag{9.8}$$

where φ is the principal decay parameter; k is a smoothing factor. When $k = 2$, this is a Gaussian distance decay function. Readers will see similar distance decay functions in the kernel density estimation (KDE) and geographically weighted regression (GWR) later in the chapter where h, also known as bandwidth, is introduced. A small bandwidth results in a rapid distance decay, whereas a larger value results in a smooth change. A variant of Eq. (9.8) is $w(d_{ij}) = \exp\left[-\frac{1}{2}\left(\frac{d_{ij}}{h}\right)^{2}\right]$, where h is the bandwidth.

The spatial weighting function can either be universal (i.e., applied equally at each point) or adaptive, depending on the location of a point as shown in Fig. 9.3.

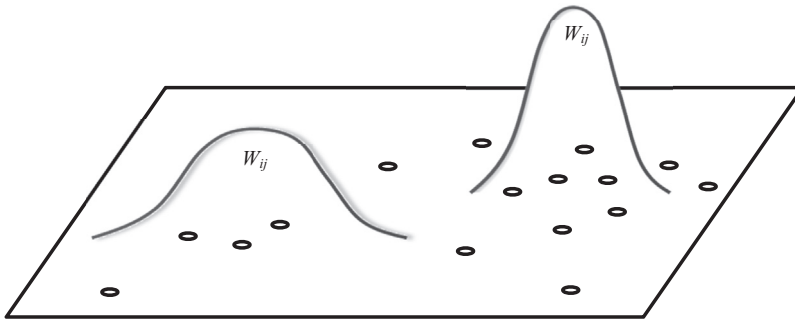

FIGURE 9.3 Adaptive spatial weighting function (⬭ data point).

The location-specific spatial weighting function is presumed to improve accuracy in prediction, but it can also be computational challenging. The following is used for a spatial adaptive weighting function:

- Bi-square:

$$w_{ij}(d_{ij}) = \begin{cases} \left[1 - \dfrac{d_{ij}^2}{k}\right]^2, & d_{ij} < d \\ 0, & \textit{otherwise} \end{cases} \tag{9.9}$$

Where k is determined by the kth nearest-neighbor distance from i. In this way, the exact number of points (i.e., k nearest neighbors) can be used to calibrate the relationship at point i.

- Ranked distance:

$$w_{ij}(d_{ij}) = \exp(-\delta R_{ij}) \tag{9.10}$$

where R_{ij} is the rank of the distance for data point j from calibration point i.

- Constraints to one:

$$\sum_j w_{ij} = 1, \quad \text{for all } i \tag{9.11}$$

All three location-specific functions will ensure that the weighting kernels will expand in areas where the data are scarce and concentrate in areas where data are rich.

9.5 Point data analysis

Point data analysis is used to study the distribution of the location of point data in the hope that the spatial patterns observed will provide information about the underlying process that generates the points. During a safety analysis, analysts sometimes aggregate crashes by location based on preestablished boundaries such as traffic analysis zones (TAZ), roadway segments, or intersections. However, the patterns observed can vary by the choice of zone boundaries. Other techniques such as kernel density estimation allow researchers to analyze crash point patterns directly.

Analysts may also be interested in studying the location association between two sets of data, such as the association between a run-off road crash location and a horizontal curve location or an ice-related crash location with a bridge location. In these cases, the goal is to establish the relationships between several sets of point data rather than within a single set of points and to perform a multivariate analysis instead of a univariate analysis.

An exploratory data analysis and a model-based analysis are two methods for analyzing point patterns. Exploratory approaches consist mostly of graphic techniques such as scatterplots, box plots, or bubble plots in which the degree of clustering or dispersion of points can be measured by the "mean nearest neighbor distance" (i.e., the mean value of the distance from one point to the closest point in the dataset) (see Chapter 5—*Exploratory Analyses of Safety Data*). Larger mean nearest neighbor distances usually suggest more regularity, while smaller values suggest clustering.

9.5.1 First- and second-order process

Point patterns can be defined by running the statistical test of a point pattern against the point pattern generated through a random process, also known as complete spatial randomness (CSR). We use Fotheringham et al. (2000)'s mathematical models to define point patterns. Let us first introduce the first-order and second-order property of point processes. In the first-order property of point process, a point dataset $\mathbf{X}$ consists of a series of two-dimensional points $\{x_1, \cdots, x_n\}$. All points are contained in a study area R. In a random process, the mean intensity M is defined as: $M = N(A)/|A|$, where $|A|$ is the area of zone A. Now, assume x is a point in R and that $C(x, r)$ is a circular subarea of R centered on x with radius r. Let r infinitely approach zero, this expected value $E(N(C(x, r), \mathbf{X}))$ will approach a limit defined as intensity at the point x, denoted by $\lambda(x)$. In the

first-order properties of point processes, expected values associate with individual points defined over R as follows:

$$\lambda(x) = \lim_{r \to 0} \frac{E(N(C(x,r), \mathbf{X}))}{\pi r^2} \tag{9.12}$$

Let us normalize $\lambda(x)$ by dividing by the expected number of events in the entire area of R. This gives the probability density function of the process, denoted as follows:

$$f(x) = \frac{\lambda(x)}{E(N(R,\ \boldsymbol{X}))} \tag{9.13}$$

It is important to distinguish between the intensity function and probability density function (PDF), which is a normalized intensity. The PDF is used when comparing the spatial distribution of two datasets.

The first-order process is associated with individual points, while the second-order property of point process examines the correlations (i.e., interactions) between two distinct points in R.

When considering the interaction between a pair of points in the study area R, the second-order intensity function can be formulated as follows:

$$\gamma(x_1, x_2) = \lim_{r_1 \to 0, r_2 \to 0} \frac{E(N(C(x_1, r_1), \boldsymbol{X}) N(C(x_2, r_2), \boldsymbol{X}))}{\pi r_1^2 \pi r_2^2} \tag{9.14}$$

The distance $d = (x_1 - x_2)$ between a pair of points can also be called displacement or pairwise distance. When only the distance between the two points is of interest, the expression of $\gamma(d)$ called a stationary process is more appropriate than $\gamma(x_1, x_2)$. Ripley's K function is probably the most known second-order measure for summarizing a point pattern.

9.5.2 Kernel density estimation

The most straightforward way to analyze data patterns at first-order intensity is to divide the area into grid squares, count the number of events in each square, divide the counts by the area of the squares, and then formulate a density function over space. The process is similar to generating a histogram, where the range of values would be binned (i.e., divide the entire range into equal length intervals) and counted to see how many fell into each interval. The values would then be displayed using bars of different heights (frequency). It is evident that the method is sensitive to the size of the grid squares being used. In addition, the value within each grid cell is independent of its neighbor cells, which contradicts the data with strong spatial autocorrelation.

A better alternative is the KDE. In statistics, KDE is a nonparametric method used to estimate the probability density function of a random

variable. Let $P = \{x_1, \cdots, x_n\}$ be a univariate independent and identically distributed sample drawn from some distribution with an unknown density function. The shape of this unknown function can be estimated through its kernel density estimates as follows:

$$\widehat{f}_k(x) = \frac{1}{nh^d} \sum_{i=1}^{n} \mathrm{K}\left(\frac{x - x_i}{h}\right) \tag{9.15}$$

where x is the variable of interest; h is the bandwidth that controls the amount of smoothing; d is the dimension (e.g., $d = 1$ is one-dimensional kernel like a roadway link; $d = 2$ is two-dimensional kernel like an area); and, K(.) is the kernel function.

Basically, the KDE smooths each data point x_i into a small density bumps and then adds all these small bumps together to obtain the final density estimate. In a kernel density, a small hump is a center at every crash point and the shape of the hump represents the extent of the influence of each crash. The influence of a crash is at its peak right at the crash location and decreases as it moves away from the location. The crash intensity of any location is the accumulated density of individual kernels of all crash points.

Kernel functions can take on different forms: uniform, normal, exponential, and spherical.

- Uniform:

$$\mathrm{K} = \begin{cases} 1, & d_{ij} \le h \\ 0, & d_{ij} > h \end{cases} \tag{9.16a}$$

- Normal (or Gaussian):

$$\mathrm{K} = \exp\left(-\frac{\left(\frac{d_{ij}}{h}\right)^2}{2} \right) \tag{9.16b}$$

- Exponential:

$$\mathrm{K} = \exp\left(-\frac{d_{ij}}{h}\right) \tag{9.16c}$$

- Spherical:

$$K = \begin{cases} \left[1 - \left(\frac{d_{ij}}{h}\right)^2\right]^2, & d_{ij} \le h \\ 0, & d_{ij} > h \end{cases} \tag{9.16d}$$

The kernel function determines the shape of the hump, and the parameter h (bandwidth) controls the spread of the hump. A small bandwidth results in a very rapid distance decay, whereas a larger value will result in a smoother decay. In addition, a large h value tends to smooth out local effects, while a smaller h value tends to produce rugged surfaces with many spikes. Cross-validation is a common approach for choosing an optimal h value (Bowman, 1984), a balance between "spreadness" and "spikeness." When kernel density is assumed to be a Gaussian distribution with variance σ^2, the one-dimensional and two-dimensional bandwidth are, respectively, defined in Eqs. (9.17a and b):

$$h_{opt} = 1.06n^{-1/5}\sigma \tag{9.17a}$$

$$h_{opt} = \left(\frac{2}{3n}\right)^{1/4}\sigma \tag{9.17b}$$

where variance σ^2 can be estimated from the sample standard deviation and n is the sample size.

The kernel function can be either universal (i.e., applied equally at each point) or adaptive to the location of a point. The location-specific kernel is presumed to improve accuracy in prediction, but it can be computationally challenging. Readers can refer to Section 9.4 "Spatial Weights and Distance Decay Models" for more information.

Example 9.2

Use local spatial autocorrelation and the kernel method for identifying crash hot spots.

This example illustrates that Flahaut et al. (2003) used the kernel density estimation method to define the location and the length of crash hot spots. The basic spatial unit (BSU) of a roadway segment is hectometers (hm) (100 m).

The study is based on 510 crashes that occurred on Belgium's N29, a 527 hm roadway, between 1992 and 1996. Local Moran's I (I_i^*) are calculated for the n BSUs based on the number of crashes per hectometer and a spatial contiguity matrix (0−1 matrix) is considered. The intensity of the crash risk depends on the value of I_i and the length of the hot spot

Example 9.2 (*cont'd*)

depends on the number of neighbors and w_{ii} matrix. The authors adopted adaptive length for a hot spot by trying 10 different levels of contiguity for a given *i*th BSU. The number of neighbors varying from 2 to 20 is symmetrically located on either side and thus, 10 values for I_i^* are calculated. The length of the hot spot corresponding to the maximum I_i^* index is used in the study. Note that weights (w_{ii}) are inversely proportional to the distance in various forms $\left(d_{ij}^0,\ d_{ij}^{-1},\ d_{ij}^{-1.5},\ d_{ij}^{-2}\right)$. Fig. 9.4 displays the hot spots ranked from the first three quantiles based on the Moran's I_i^*.

The kernel function for KDE is assumed to be a normal distribution. The reference bandwidth ($h_{ref} = 59.6$ hm) and the optimal bandwidth ($h_{opt} = 26.5$ hm) are compared using the minimization of the integrated mean square error. As expected, a small bandwidth displays more peaks than a large bandwidth.

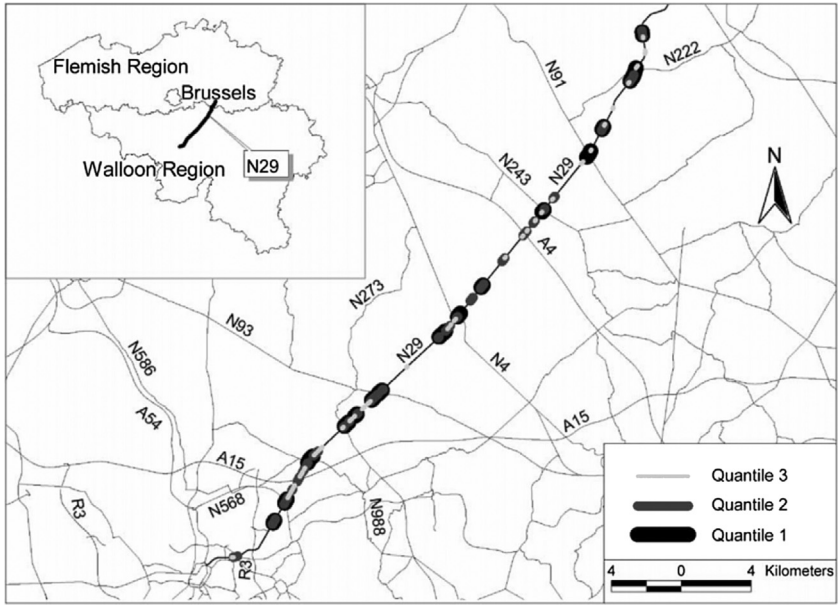

FIGURE 9.4 The N29: its location in Belgium and the hot spots defined by the local autocorrelation indices *I** i (weights function of d_{ij}^{-2}). *From Flahaut, B., Mouchart, M., Martin, E.S., Thomas, I., 2003. The local spatial autocorrelation and the kernel method for identifying black zones: a comparative approach. Accid. Anal. Prev. 35 (6), 991–1004.*

Fig. 9.5 displays resemblances between the local Moran's I and KDE when h is set as 2.5 hm for the KDE and when d_{ij}^{-2} is used for the spatial autocorrelation method.

continued

Example 9.2 (*cont'd*)

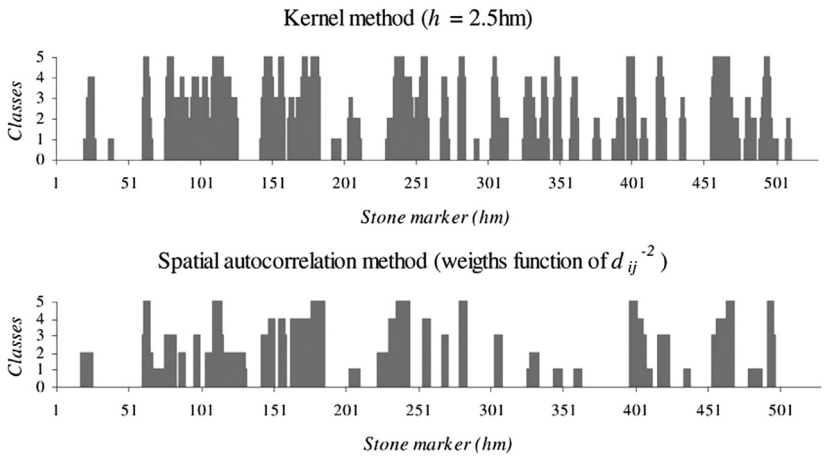

FIGURE 9.5 Comparison of the dangerousness of the N29 determined by kernel and the spatial correlation methods. *From Flahaut, B., Mouchart, M., Martin, E.S., Thomas, I., 2003. The local spatial autocorrelation and the kernel method for identifying black zones: a comparative approach. Accid. Anal. Prev. 35 (6), 991–1004.*

9.5.3 Ripley's K-function

When the second-order property of data is characterized as an isotropic and stationary process, the second-order intensity function depends only on the distance between the two points as follows:

$$K(d) = \frac{E\,(N(C(\mathrm{x}, d)))}{\overline{\lambda}} \tag{9.18}$$

where C is the circular region operator, N is the count of pairs within the distance d, and $\overline{\lambda}$ is the unit count, and x is the data.

An estimate of $K(d)$ is

$$\widehat{K}(d) = \frac{|A| \sum_i N(C(\mathrm{x}_i, d))}{n^2} \tag{9.19}$$

where n is the number of points in the dataset and $|\mathrm{A}|$ is the size of the study area. So, the K function for a distance d is the average number of points found in a circle of radius d centered on an event, divided by the intensity of points, which is the number of events divided by the study area.

Ripley's K is the best-known second-order statistic in spatial statistics. Ripley (1976) introduced the planar K-function analysis to depict the spatial distribution of point events in a given dataset **P**. Ripley's K tests point patterns on various spatial scales, handles all event to event distances, and does not aggregate points into areas.

The analysis examines a function, $K^{pl}(d)$, defined as follows:

$$K^{pl}(d) = \frac{1}{\lambda} E(\text{number of points in } \mathbf{P} \text{ within Euclidean distance d of a point in } \mathbf{P}) \tag{9.20}$$

where $E(\,)$ is the expected value. When the point pattern is CSR following the Poisson point process and where events are independently and uniformly distributed over the study area, the theoretical value of $K^{pl}(d)$ is given by πd^2. For a set of n points, $P = \{x_1, \cdots, x_n\}$, distributed over a study area A, the estimator of the planar K-function is given by

$$\widehat{K}^{pl}(d) = \frac{1}{\widehat{\lambda} n} \sum_{i \neq j} \sum I_h(d_{ij}) = \frac{|A|}{n^2} \sum_{i \neq j} \sum I_d(d_{ij}) \tag{9.21}$$

where $\widehat{\lambda}$ is estimated as $n/|\boldsymbol{A}|$, d_{ij} is the Euclidean distance between points p_i and p_j, and $I_d(d_{ii})$ is an indicator function defined as follows: $I_d(d_{ij}) = \begin{cases} 1 & \text{if } d_{ij} \leq d0 \text{ otherwise.} \end{cases}$ Note that $\sum_{i \neq j} \sum I_d(d_{ij})$ can be expressed as $\sum_i N(C(\mathrm{x}_i, d))$.

In practice, the Monte Carlo simulation is often used to calculate pseudo significance levels by repeated randomization. This technique determines the expected values of $\widehat{K}^{pl}(d)$, the upper and lower significance envelopes under the null hypothesis of CSR. Cressie (1993) had a more detailed discussion of the planar K-function analysis.

Okabe and Yamada (2001) and Yamada and Thill (2004) extended from a Euclidean space to a network space. The modified network K-function is

$$K^{net}(d) = \frac{1}{\rho} E(\text{number of points in } \mathbf{P} \text{ within network distance d of a point } p \text{ on } L_T) \tag{9.22}$$

where ρ is the density of points with respect to a unit length of the network links.

Let $L_p(d)$ be the subset of a finite network L_T such that the network distance between p and any point in $L_p(d)$ is less than or equal to d. The network K-function is formulated as below:

$$K^{net}(d) = \frac{1}{|L_T|} \int_{p \in L_T} |L_p(d)| dp \tag{9.23}$$

where dp means the intergarion of p along L_T, and $|L_T|$ is the length of L_T.

For an observed point pattern, the estimator of the network K-function $\widehat{K}^{net}(d)$ is given by

$$\widehat{K}^{net}(d) = \frac{|L_T|}{n(n-1)} \sum_{i \neq j} \sum I_d(d_{ij}) \tag{9.24}$$

where d_{ij} is the network distance between points x_i and x_j, and $I(\ )$ is the indicator function defined previously. Note that an unbiased estimator of the density of points, $\frac{n-1}{L_T}$, is used here rather than, $\frac{n}{L_T}$. See Okabe and Yamada (2001) for more details on the method derivation.

Because the mathematical formulation of the expected value of $K^{net}(h)$ and its upper and lower envelopes for a given network is complicated, Monte Carlo simulation is a conventional approach to generate these values. Okabe et al. (2006) developed the user-friendly toolbox, called SANET (Spatial Analysis on a NETwork), for network spatial analysis. SANET can handle two types of network spatial phenomena: (a) phenomenon that occurs on a network (e.g., traffic crashes); and (b) phenomenon that occurs along a network (e.g., retail stores in an urbanized area). The toolbox contains 13 tools, including the K-function and cross K-function methods.

Yamada and Thill (2004) used the K-function to explore whether highway crashes in a region in New York are spatially clustered. To better compare the differences between the planar and network K-functions, the authors added the third type of K-function that satisfies the constraint that points are on the network, but distances between points are measured as a straight-line, so-called the planar K-function with network constraint. The authors observed values greatly exceed the corresponding upper envelope of all three K functions (i.e., planar, network, planar with network constraint) over the entire range of distances, and they concluded that crashes occurring in these regions are spatially clustered. Further, they noticed that the degree of exceedance is most significant in the case of the planar K-function, and suspected that an observed distribution can possibly be misidentified as a clustered pattern according to the planar K-function while it is actually not. To show this, the authors simulated random patterns of the same number of crashes in the same highway network. The observed K-function values for this simulated data prove that the planar K-function detects significant spatial clustering over the

entire range of distances, whereas the other two methods suggest randomness. Hence, the planar K-function analysis mistakenly detects a clustered pattern although the observed pattern is indeed random.

9.5.4 Cross-K function

Cross-K function is used for bivariate point patterns instead of univariate point patterns. The network cross-K function describes the relationship between the patterns of two sets of points (e.g., $\mathbf{A} = \{a_1, a_2, \ldots, a_{na}\}$ and $\mathbf{B} = \{b_1, b_2, \ldots, b_{nb}\}$) placed on a finite planar network, and shows whether the set of points **B** affects the location of the set of points **A**. The effect can be examined with the following null hypothesis: the set of points **A** is randomly distributed, regardless of the location of the set of points **B**. A rejected null hypothesis confirms that the location of the set of points **B** affects the distribution of the set of points **A**. The cross-K function is defined as follows:

$$K^{ba}(d) = \frac{1}{\rho_a} E(\textit{number of points of } \boldsymbol{A} \textit{ within network distance } d \textit{ of a point} b_i \textit{ in } \boldsymbol{B}) \tag{9.25}$$

where $K^{ba}(d)$ is the network cross-K function of **A** relative to **B** for the binomial point process. **A** is a set of point events, and **B** is a different set of point events. $E(\,)$ is the expected value of A following the binomial point process, with respect to $b_1, \ldots, b_{nb}$, $\rho_a = na/L_T$, density of points of **A**. na is the total number of point events of **A**, and L_T is the length of finite planar network. The network cross-K function analysis can be conducted using the SANET tool developed by Okabe et al. (2006).

Both the Ripley K function and cross-K function are used to measure the second-order property of point events as a function of pairwise distance on a global scale. Ripley K is interested in testing the correlation or intersection between any pair of points within the same dataset, whereas cross-K is interested in testing the colocation or coexistence of two different datasets. The estimated K function is a sample function, and it is therefore subject to random error. The standard procedure is to check whether the estimated K function is located between the upper- and lower-bound of the 95th confidence interval is useful in these instances. These tests can be generated by sampling methods such as bootstrap or Monte Carlo simulation.

Example 9.3

Apply Lattice Data Analysis, Network K-Functions to Study Ice-Related Crashes.

Khan et al. (2009) combined advanced spatial statistical methods with GIS functionality to analyze spatial patterns of safety data. The aim is to

continued

EXAMPLE 9.3 (*cont'd*)

show the usefulness of these techniques by analyzing county-level data and identifying spatial patterns of ice-related crashes that have occurred in the selected counties. The study uses 4 years of Wisconsin crash data (2003 through 2006) between November 1 and April 30, which is the winter season in Wisconsin. Ice-related crashes were identified as those that occurred with ice on the pavement, sleet falling, or both. The ice-related crash rate is calculated as crashes per 100 million VMT for study counties. Data on bridge locations and on roads comprising the Wisconsin State Trunk Network system were also obtained.

The results of the local Moran's I analysis were used to select counties for which the network cross K-function analysis was conducted. Counties with a Z score of greater than +1.96 represent locations that are part of statistically significant clusters of similar ice-related crash rates at a 95% confidence level, and vice versa. The results of the K-function analysis show for the northwest counties, suggesting that bridge locations in Barron County are more prone to ice-related crashes than are the locations in Rusk. Refer to Khan et al. (2009) for details.

9.6 Spatial regression analysis

The purpose of applying spatial regression analysis is to explicitly consider the spatial dependence of observations in the regression model. In spatial regression, the residuals are not independent of each other but are spatially structured and correlated. Ignoring the spatial dependence of data will result in inefficient, biased, and inconsistent parameter estimates and statistical inferences. In the highway safety analysis, spatial regression analysis has been used to identify and rank areas with potential improvements for safety, estimate the varying effects of crash contributing factors over space, and recognize spatial patterns and scopes to inform decision-making.

In general, the two techniques in the spatial regression analysis are as follows: (1) spatial econometrics models for continuous spatial data and (2) Bayesian hierarchical models for nonnegative random count spatial data.

9.6.1 Spatial econometrics methods

Spatial econometrics is a subdiscipline of econometrics that handles spatial autocorrelation and spatial heterogeneity in regression models. These methods are particularly appealing in modeling crashes because crash events are regarded as the consequence of complicated interactions between many factors: driver, vehicle, roadway, and environment. Crashes have locational and geographic implications to consider when deciding safety treatments. Models used to estimate such geographic phenomena require the specification of spatial effects in econometric models. This section focuses on two well-known spatial econometrics methods: the spatial autoregressive model (SAR) and the spatial error model (SEM).

9.6.1.1 Spatial autoregressive model

Autocorrelation means a correlation exists between the values of the same variable. Crash data such as frequency or severity within proximity in space or time can be correlated due to shared weather, land use characteristics, driver behavior, policies, and enforcement practices (Levine et al., 1995). The SAR (also known as the lagged response model or spatial lag model) deals with spatial autocorrelation by adding an explanatory variable in the form of a spatially lagged dependent variable to a multivariable regression model, as formulated below:

$$\mathbf{y} = \mathbf{x}'\boldsymbol{\beta} + \boldsymbol{\rho}\mathbf{W}\mathbf{y} + \boldsymbol{\varepsilon} \tag{9.26}$$

where $\mathbf{y}$ is an n $\times$ 1 vector of observations for all locations, i; $\mathbf{W}$ is the spatial matrix that specifies the spatial dependence between observations; $\mathbf{Wy}$ is the spatially lagged dependent variable; ρ is the coefficient of the spatial lag; $\mathbf{x}$ is an $n \times k$ matrix of observations on the explanatory variables, β is a $k \times 1$ vector of regression coefficients, and ε is an $n \times 1$ vector of normally distributed random error terms, with zero mean and constant variance σ^2.

The SAR model expresses the notion that the value of a variable at a given location is related to the values of the same variable measured at nearby locations through the spatial weights matrix, $\mathbf{W}$, which is almost always row standardized. The SAR model rearranges and multiplies both sides by $(\mathbf{I} - \rho\mathbf{W})$, and is reformulated as:

$$\mathbf{y} = (\mathbf{I} - \rho\mathbf{W})^{-1}\mathbf{x}'\boldsymbol{\beta} + (\mathbf{I} - \rho\mathbf{W})^{-1}\boldsymbol{\varepsilon} \tag{9.27}$$

Thus, the new variance–covariance matrix is $C = \sigma^2\left[(\mathbf{I} - \rho\mathbf{W})^{-1}\right]'(\mathbf{I} - \rho\mathbf{W})^{-1}$, where a spatial structure is reflected in error dependence. Once ρ is known, $\boldsymbol{C}$ can be computed and $\boldsymbol{\beta}$ can be estimated. If ρ is statistically significant, the spatial dependence exists and should be included in the model.

Model performance in terms of goodness of fit can be evaluated with the Akaike information criterion (AIC) described in Chapter 2—*Fundamentals and Data Collection*.

9.6.1.2 Spatial error model

The SEM is applied when data are suspected to be autocorrelated. In the SEM, the error term—not the dependent variable—is considered as autoregressive. The SEM is formulated as

$$\mathbf{y} = \mathbf{x}'\boldsymbol{\beta} + \varepsilon \text{ where } \boldsymbol{\varepsilon} = \lambda \mathbf{W}\boldsymbol{\varepsilon} + \boldsymbol{\mu} \tag{9.28}$$

The basic SEM is similar to the standard linear model, but the error term is made up of a spatially weighted error vector $\lambda\mathbf{W}\boldsymbol{\varepsilon}$ where λ is the spatial autoregressive coefficient and a vector of *i.i.d* random errors $\mathbf{u}$ with zero mean and variance σ^2. We can rearrange the expression for $\boldsymbol{\varepsilon}$ to obtain $\varepsilon = \left[(\mathbf{I} - \lambda\mathbf{W})^{-1}\right]\mu$.

Hence, the error variance–covariance matrix, $\boldsymbol{C}$, is given by $E(\varepsilon\varepsilon^T) = \sigma^2[(\mathbf{I} - \lambda\mathbf{W})^T(\mathbf{I} - \lambda\mathbf{W})]^{-1}$ and the spatial error model can be arranged as

$$\mathbf{y} = \mathbf{x}'\boldsymbol{\beta} + \lambda\mathbf{W}\mathbf{y} - \lambda\mathbf{W}\mathbf{x}'\boldsymbol{\beta} + \boldsymbol{\mu} \tag{9.29}$$

In the SEM model, the dependent variable $\mathbf{y}$ is a combination of a general (global) linear trend component $\mathbf{x}'\boldsymbol{\beta}$, plus a pure spatial autocorrelation component $\lambda\mathbf{W}\mathbf{y}$, plus a (negative) weighted average of predicted neighboring value $\lambda\mathbf{W}\mathbf{x}'\boldsymbol{\beta}$, plus an *i.i.d* random error term $\mathbf{u}$. So, the spatial error model can be viewed as a form of the mixed spatial lag model with the additional spatial component of the neighboring trend $\lambda\mathbf{W}\mathbf{x}'\boldsymbol{\beta}$. When models for spatial lag effects do not provide a significant improvement, readers can consider using SEM.

Example 9.4

Model area-wide crash count with spatial correlation and heterogeneity.

Quddus (2008) collected ward-level safety data from the 633 census wards from the Greater London metropolitan area. The spatial distribution of crash frequency at the ward-level is displayed in Fig. 9.6.

Zonal-level crash counts (2000–02) are modeled by traffic characteristics, roadway characteristics, and sociodemographic factors. The SAR and the SEM are formulated as:

$$\text{SAR: } \ln\left(\frac{y_i}{EV_i}\right) = \rho\mathbf{W}\ln\left(\frac{y_i}{EV_i}\right) + \mathbf{x}_i'\boldsymbol{\beta} + \varepsilon_i \text{ where } \varepsilon_i \tilde{} N(0,\ \sigma^2\boldsymbol{I})$$

Example 9.4 (*cont'd*)

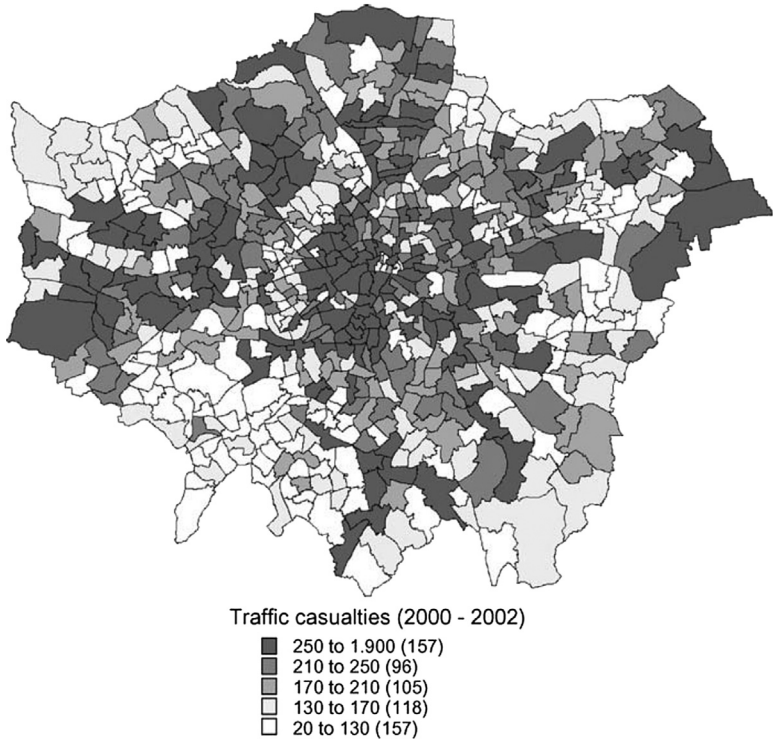

FIGURE 9.6 Spatial distribution of crash frequency. *From Quddus, M.A., 2008. Modelling area-wide count outcomes with spatial correlation and heterogeneity: an analysis of London crash data. Accid. Anal. Prev. 40 (4), 1486–1497.*

$$\text{SEM}: \ln\left(\frac{y_i}{EV_i}\right) = \mathbf{x}_i'\boldsymbol{\beta} + \varepsilon_i$$
$$\varepsilon i = \lambda \mathbf{W}\varepsilon_i + \mu_i \text{ where } \mu_i \tilde{} N\left(0,\ \sigma^2 \boldsymbol{I}\right)$$

where y_i are a number of crashes at ward i; EV_i is the exposure variable; $\mathbf{x}$ are covariates; and β is the estimable coefficients. $\mathbf{W}$ is the spatial weights matrix.

The model results are summarized in Table 9.2.

continued

TABLE 9.2 Model results.

Spatial error models (SEMs)	Total serious		Total slight		Motorized slight	
	Coef	*t*-stat	Coef	*t*-stat	Coef	*t*-stat
Traffic characteristics						
ln(Traffic flow (PCU km/h))	0.5713	3.08	0.5882	3.57	0.3499	1.56
Average speed (km/h)	−0.0215	−2.77	−0.0163	−1.34	−0.0101	−1.12
Road characteristics						
Length of motorway (km)	0.1025	2.89	0.1228	3.89	0.0730	1.70
Length of A-road (km)	0.0507	2.07	0.0358	1.64	0.0873	2.93
Length of B-road (km)	0.0091	0.40	−0.0071	−0.35	0.0174	0.64
Length of minor road (km)	0.0191	2.57	0.0166	2.51	0.0207	2.29
ln(Road curvature)	–	–	–	–	−1.9405	−0.29
Socio-demographic factors						
ln(Resident population aged less than 60)	0.1209	0.83	0.0562	0.43	0.1656	0.92
ln(Resident population aged 60 or over)	−0.2588	−2.67	−0.2295	−2.64	−0.2493	−2.14
ln(# of employees)	0.1204	3.25	0.1269	3.85	0.0382	0.85
ln(# of households with no cars)	0.3160	3.26	0.3344	3.86	0.1843	1.56
Constant	−16.95	−6.10	−15.14	−6.12	−11.21	−3.33
Spatial autoregressive coef (lamda)	0.8246	16.02	0.8478	17.71	0.7778	13.23
Observations	633		633		633	

Example 9.4 (*cont'd*)

TABLE 9.2 Model results.—cont'd

Spatial error models (SEMs)	Total serious		Total slight		Motorized slight	
	Coef	*t*-stat	Coef	*t*-stat	Coef	*t*-stat
Tests for lamda = 0						
Wald test	Reject H0: of no spatial dependence		Reject H0: of no spatial dependence		Reject H0: of no spatial dependence	
Likelihood ratio test	Reject H0: of no spatial dependence		Reject H0: of no spatial dependence		Reject H0: of no spatial dependence	
Lagrange multiplier test	Reject H0: of no spatial dependence		Reject H0: of no spatial dependence		Reject H0: of no spatial dependence	

Estimation results for SEM models.

From Quddus, M.A., 2008. Modelling area-wide count outcomes with spatial correlation and heterogeneity: an analysis of London crash data. Accid. Anal. Prev. 40 (4), 1486–1497.

Modeling crash rate, that is, $\frac{Y_i}{EV_i}$, rather than directly crash count is a compromising strategy because in spatial econometric models, the error term is normally distributed. Another methodological issue emerges due to the natural logarithm of the dependent variable, a crash rate which is a nonnegative real. Hence, the ward-level traffic casualty data with zero counts cannot be modeled; only total serious injuries, total slight injuries, and motorized slight injuries were modeled. The autoregressive coefficient ρ in the SAR model was statistically insignificant, suggesting SAR may not be an appropriate model specification. An SEM model with an autoregressive coefficient λ was statistically significant.

9.6.2 Generalized linear model with spatial correlation

This section describes the characteristics of the generalized linear mixed model (GLMM) and Hierarchical Bayesian model (HBM).

9.6.2.1 Generalized linear mixed model

As can be seen in Example 9.4, the spatial econometric models are designed for a continuous variable, count data needs to be converted to meet the model assumption. The GLMM can accommodate spatial

random effects and provide a flexible means of modeling spatially correlated counts. Let's begin with specifying a nonspatial GLMM for the number of crashes as:

$$\lambda_i = \exp\left(\beta_0 + \sum_{k=1}^{K} \beta_k x_{ik} + \mu_i\right) \tag{9.30}$$

where x_{ik} are kth explanatory covariate, βs are the regression coefficients to be estimated from the data, and μ_i is unstructured random error assumed to be distributed as $N(0, \sigma^2)$. The parameter σ^2 represents the variance in the population, and therefore the degree of heterogeneity of the subjects.

Although μ_i can be treated as the random effects due to spatial influence, a more appropriate formulation is to explicitly add the neighboring effects through the spatial weights matrix. Thus, the spatial Poisson GLMM model can be specified as follows:

$$\lambda_i = \exp\left(\beta_0 + \sum_{k=1}^{K} \beta_k x_{ik} + \varphi\right) \tag{9.31}$$

where φ_i is assumed to have conditional autoregressive (CAR) (Besag, 1974, 1975). A subclass[2] of CAR called the intrinsic conditional autoregressive (ICAR) is used here. The conditional distribution of the random effects $\varphi_i|\varphi_{-i}$ is

$$\varphi_i \Big| \varphi_{-i} \tilde{} N\left(\mu_{\varphi_i}, \sigma^2_{\varphi_i}\right) \tag{9.32}$$

where $\mu_{\varphi_i} = \sum_i \frac{w_{ij}\varphi_j}{w_{i+}}$; $w_{i+} = \sum_{j \in i} w_{ij}$ and $\sigma^2_{\varphi_i} = \frac{\sigma_\varphi}{w_{i+}}$.

9.6.2.2 Hierarchical Bayesian model

A more popular approach to modeling crash count with spatial correlation in safety literature is the HBM with spatial random effects. This modeling approach is more flexible than GLMM in fitting hierarchical models with nonnormal random effects. The HBM, covered in Chapter 2—*Fundamentals and Data Collection*, is capable of accounting for high data variance due to unobserved heterogeneity, discovering geographic patterns and trends, and increasing accuracy of model estimates through "borrowing strengths" from neighboring sites.

In Chapter 2, it is discussed that HBMs are flexible in configuring complicated models by specifying a variety of components and prior

[2] A more general CAR model is called the Besag et al. (BYM) (1991) model. It is a lognormal Poisson model, which includes both an ICAR component for spatial smoothing (i.e., spatial autocorrelation) and an ordinary random-effect2s component for nonspatial heterogeneity.

distributions in a hierarchical structure. At the first level of the model hierarchy, the number of crashes (Y_{it}) at site i and the time period t is assumed to be a mutually independent and Poisson distributed random variable and is defined as follows:

$$y_{it}|\mu_{it} \sim Poisson(\mu_{it}) \text{ for } i = 1,2,\ldots,I \text{ and } t = 1,2,\ldots,T \tag{9.33}$$

The mean (μ_{it}) of Poisson is modeled as follows:

$$\mu_{it} = \nu_{it}\lambda_{it} \tag{9.34}$$

where v_{it} is the traffic exposure variable (e.g., million vehicle miles traveled (MVMT) AADT) that is usually treated as an offset.[3] λ_{it} is the crash rate.

The rate λ_{it}, as a nonnegative real, can be specified as

$$\lambda_{it} = \exp\left(\beta_0 + \sum_{k=1}^{K}\beta_k x_{ik} + \delta_t + \varphi_i + \varepsilon_{it}\right) \tag{9.35}$$

where log is the natural logarithm, x_{itk} are kth explanatory covariate, βs are the regression coefficients to be estimated from the data, δ_t represents the time effect, φ_i represents the random spatial effect; and ε_{it} is an exchangeable, unstructured random error. All φ_i, δ_t, and ε_{it} are mutually independent.

At the second level of hierarchy, the prior distribution of unknown coefficients βs and the temporal, δ_t, and spatial effect, φ_i, components represent the knowledge of the analyst's. Their prior distributions may contain new unknown parameters called hyperparameters whose distributions, also called hyperpriors, constitute the third level of the hierarchy.

Below is the configuration of the commonly used model specifications and assumptions in the highway safety analysis:

- Covariates x_{itk} can be treated either as fixed effects or random effects. The model that includes both fixed and random effects of covariates is called a mixed effect model. It is also important to note that values of covariates are usually centered (i.e., $\frac{x-\bar{x}}{SE(x)}$) for computational efficiency). For fixed effects, a noninformative[4] prior is assumed (e.g., $\beta \sim N(0, 1000)$ (Aguero-Valverde and Jovanis, 2006, 2008 and 2010), $\beta \sim N(0, 10000)$ (Quddus, 2008)). The prior for random effects follows a certain probabilistic structure with unknown parameters or hyperpriors whose probabilistic distribution needs to be specified. For instance, in Miaou et al. (2003) and Miaou and Song (2005), the random effects β_k ($k = 1, 2, \ldots K$) are assumed to be

[3] For an offset variable, the power coefficient is forced to be 1.

[4] The word "noninformative" indicates our attitude toward this part of the model and is not intended to imply that this particular distribution has any special properties.

independent and normally distributed with mean μ_{β_k}, and variance $\sigma^2_{\beta_k}$, $N\left(\mu_{\beta_k},\ \sigma^2_{\beta_k}\right)$. The hyperprior densities of μ_{β_k} and $\sigma^2_{\beta_k}$ are assumed to follow a noninformative normal and inverse gamma distribution, respectively.

$$\mu_{\beta_k} \sim N(,) \ \text{ and } \sigma^2_{\beta_k} \sim IG(,) \tag{9.36}$$

According to Miaou et al. (2003), fixed effects βs are favored over random effects βs. Most of the safety literature using HBMs with random spatial effects treat βs as fixed effects.

- δ_t represents the time effect, which is assumed to be fixed with noninformative normal priors. Miaou et al. (2003) suggested two types of fixed effects for δ_t: (a) fixed effects that vary by time t (e.g., if the data are described in yearly crash count, then use a year-wise fixed effect model with $\delta_1 = 0$ for the first year as the baseline); and (b) order-one autoregressive (AR(1)) with the same coefficient for all ts.
- φ_i represents the random spatial effect. Besag's Gaussian CAR model (Besag, 1974, 1975) and its variants are probably the most popular spatial models. The CAR model is formulated as follows:

$$P(\varphi_i|\varphi_{-i}) \propto \frac{1}{\sigma_\varphi} \exp\left\{ -\frac{1}{2\sigma^2_\varphi} \sum\nolimits_{j \in C_i} w_{ij} \left(\varphi_i - \rho\varphi_j\right)^2 \right\} \tag{9.37}$$

where $P(\varphi_i|\varphi_{-i})$is the conditional probability of φ_i given φ_{-i} and φ_{-i} represents all φ except φ_i. $\propto$ means "proportional to," C_i is a set of sites that are neighbors of site i or have an influence on site i. σ^2_φ controls the variability of φ and is a fixed effect parameter across all sites. w_{ij} is a spatial weighting factor associated with the pair (i, j), and ρ is a parameter that determines the sign and magnitude of the overall spatial dependence.

The basic idea is that the probability of values estimated at any given location is conditional on the neighboring values. For the spatial autocorrelation parameter ρ, when $\rho = 0$, there is no spatial dependence and the model is reduced to an independent model. A larger value of ρ corresponds to a stronger spatial dependence, and vice versa, although the interpretation of the actual value of ρ demands caution (Banergjee et al., 2004). Gelfand and Vounatsou (2003) discussed extensively on the value of ρ. In practice, ρ value is expected to be between 0 and one and the prior of ρ is usually assumed to be a uniform distribution as (Cressie, 1993; Qin et al., 2009):

$$\rho \sim Uniform\left(e^{-1}_{\min}, e^{-1}_{\max}\right) \tag{9.38}$$

where $e_{\min} < 0$ and $e_{\max} > 0$ are the minimum and the maximum eigenvalues of $\boldsymbol{D}^{1/2}\boldsymbol{W}\boldsymbol{D}^{1/2}$ where $\boldsymbol{W}$ denotes the weight matrix with entries w_{ij} and $\boldsymbol{D}$ denotes a diagonal matrix with entries w_{i+} to ensure positive definiteness. However, in the highway safety applications, ρ is almost always assumed to be 1.

When ρ is assumed to be 1, Eq. (9.37) can be simplified as Eq. (9.32). σ_{φ}^2 controls the variability of φ, and its inverse is precision (precision $= 1/$ variance). Most safety literature (Aguero-Valverde and Jovanis, 2006, 2008, 2010; Quddus, 2008) adopts the suggestion in (Wakefield et al., 2000) to diffuse the hyperprior density of the precision $\left(\text{or}\frac{1}{\sigma_{\varphi}^2}\right)$: $Ga(0.5,\ 0.0005)$.

- ε_{it} can take different forms. As discussed previously, when the exchangeable random error ε_{it} is assumed to be normal, $\varepsilon_{it} \tilde{} N(0, \sigma_{\varepsilon}^2)$, the hierarchical model is the Poisson-lognormal model. When $\exp(\varepsilon_{it})$is assumed to be gamma, $\exp(\varepsilon_{it}) \tilde{} Gamma(\psi, \psi)$, then the hierarchical model is the Poisson-gamma model with the dispersion parameter ψ. According to Miaou et al. (2003), the gamma distribution is preferred over the lognormal distribution. When ψincreases, the amount of overdispersion (due to spatial effects) decreases. Noninformative inverse gamma and gamma distributions are assumed for hyperprior σ_{ε}^2 and ψ, respectively, such as $Ga(0.5, 0.0005)$. Gelman has extensively discussed the choice of prior distributions for variance parameters in hierarchical models (Gelman, 2006).

For model assessment, the deviance information criterion (DIC) (Spiegelhalter et al., 2002) can be used for this purpose and is described in Chapter 2—*Fundamentals and Data Collection*.

Example 9.5

Spatial analysis of fatal and injury crashes in Pennsylvania.

Aguero-Valverde and Jovanis (2006) developed spatial models of road crash frequency for the State of Pennsylvania at the county level. The models include socioeconomic, transportation-related, and environmental factors. The results from full Bayesian (FB) hierarchical spatial models are compared with the traditional negative binomial (NB) model. Model results are presented in Table 9.3.

continued

Example 9.5 (*cont'd*)

TABLE 9.3 Model results.

	Estimate		Credible set 95%	
Variable	**Mean**	**SD**	**2.50%**	**97.50%**
Intercept	3.110	0.764	1.593	4.582
DVMT	−0.092	0.013	−0.118	−0.067
Infrastructure mileage	0.323	0.051	0.227	0.425
Mileage density	0.131	0.022	0.086	0.174
Mileage of federal aid roads (%)	0.026	0.007	0.013	0.040
Persons 0−14 (%)	0.047	0.020	0.008	0.088
Persons 15−24 (%)	0.015	0.012	−0.008	0.039
Persons 65 and over (%)	0.013	0.016	−0.018	0.044
Total precipitation	5.86E−05	2.27E−04	−3.85E−04	5.01E−04
σ_u^2	0.176	0.038	0.115	0.263
ϕ	−0.015	0.004	−0.022	−0.008
σ_δ^2	6.66E−04	2.15E−04	3.31E−04	0.001164

Full Bayes model of injury crashes with spatial correlation (u_i), time trend (ϕ), and space × time interactions (δ_i)$\overline{D}$, 3261.0; $D(\overline{\theta})$, 3159.0; DIC, 3363.0; and p_D, 102.1.
From Aguero-Valverde, J., Jovanis, P.P., 2006. Spatial analysis of fatal and injury crashes in Pennsylvania. Accid. Anal. Prev. 38 (3), 618−625.

The authors found no evidence of spatial correlation in fatal crashes but statistically significant spatial correlation in injury crashes. The authors also found that the effects of the covariates on fatal and injury crash risk are mostly consistent between the negative binomial and full Bayes models. Variables just marginally significant in the NB models are generally not significant in the FB models.

9.6.3 Modeling local relationships in crash data

The GWR model is a more sophisticated modeling technique to help understand local variations of the data (Fotheringham et al., 2002). Site-specific parameters are useful in revealing the influence of unobserved data heterogeneity and are more accurate in prediction compared with a global model in which the parameters do not vary across space. GWR can be formulated as

$$y_i = \beta_0(\mu_i, \upsilon_i) + \sum\nolimits_k \beta_k(\mu_i, \upsilon_i) x_{ik} + \varepsilon_i \tag{9.39}$$

where $\beta_k(u_i, v_i)$ ($k = 1$, m) are a set of specific coefficients to site (u_i, v_i) of point i; x_i are explanatory variables; y_i are dependent variables; and, ε_i is the error term.

In the GWR framework, the geographically weighted Poisson regression (GWPR) and the geographically weighted negative binomial regression (GWNBR) model are appropriate and consistent with the state of practice in crash count modeling. Nakaya et al. (2005) applied GWPR as a new statistical technique to study the spatially varying relationships between disease rate and socioeconomic characteristics. Hadayeghi et al. (2010), Li et al. (2013) utilized GWPR in the software GWRx3.0 (Charlton et al., 2003) to develop zonal-level and county-level crash count models. Although it is well accepted that crash data are overdispersed, the GWRx3.0 does not support the calibration of the GWR model estimates with a negative binomial structure. Till 2014, da Silva and Rodrigues (2014) proposed GWNBR to model nonstationary count data with overdispersion. In the general form of the GWNBR model, the overdispersion parameter $\alpha(u_i, v_i)$ varies by location (u_i, v_i). Silva and Rodrigues (2016) implemented both GWPR and GWNBR with a set of SAS/IML© macros. Gomes et al. (2017) used spatial local GWPR and GWNBR to develop zonal level safety performance functions for Fortaleza/Brazil.

Here, only GWPR is discussed for the rest of the section. Readers are referred to da Silva and Rodrigues (2014) for more details about GWNBR. Let the number of crashes y_i at ith site follow a Poisson distribution as $\mu_i \tilde{} Poisson(\mu_i)$; the Poisson mean μ_i is formulated as $\mu_i = V_i^{\beta_1(\mu_i, v_i)} \exp\left(\sum_k \beta_k(\mu_i, v_i) x_{ik}\right)$ where V_i is the traffic exposure variable for site i. The logarithm transformation of μ_i is specified as

$$\ln\mu_i = \beta_0(\mu_i, v_i) + \beta_1(\mu_i, v_i) lnV_i + \beta_2(\mu_i, v_i)x_{i2} + \ldots + \beta_k(\mu_i, v_i)x_{ik} \quad (9.40)$$

Coefficients βs can be formulated as

$$\boldsymbol{\beta}(\mu_i, v_i) = (\mathbf{x}'\mathbf{W}(\mu_i, v_i)\mathbf{x})^{-1}\mathbf{x}'\mathbf{W}(\mu_i, v_i)\mathbf{y} \quad (9.41)$$

where $\mathbf{W}(u_i, v_i) = \begin{pmatrix} w_{i1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & w_{in} \end{pmatrix}$ is a $n \times n$ matrix whose off-diagonal elements are zero and whose diagonal elements denote the geographical weighting of observed data for point i. w_{ij} represents the weight of data at point j on the calibration of the model around point i.

In GWPR, coefficients are calibrated for each crash site based on the observations in nearby sites. The estimation process is repeated for all regression points. The resultant models may not be spatially transferable because the set of local parameters are for a specific geographic region. Probably, the nonstationary coefficients themselves are more interesting than model equations as they may reveal hidden patterns and shed new light on the potential causes for crashes.

When calibrating GWR, it is assumed that the observations that are closer to point i are more influential in the estimation of βs than the observations that are farther away. This weight determines how many points will be used to calibrate the coefficients for point i and their contributions in the model calibration. The degree of influence is commonly specified through a distance-decayed weight function, **W**. The Gaussian function and bi-square function (adaptive kernel) are commonly used. The Gaussian function is unbounded, while the bi-square function can be bounded by the bandwidth. The selection of bandwidth is important, even more important than the choice of kernel density. In most of the safety literature on GWRP, either an adaptive kernel or a trial-and-error procedure is adopted to determine the optimal bandwidth (Hadayeghi et al., 2010; Li et al., 2013; Xu and Huang, 2015). Readers can refer to "Spatial Weights and Distance Decay Models" for the choice of weighting function.

The AIC, described in Chapter 2—*Fundamentals and Data Collection*, can be used as an indicator not only for choosing between fixed and adaptive kernels, but also for bandwidth selection in adaptive kernels. The best GWRP model is the one with the lowest AIC; however, Nakaya et al. (2005) argued that because the degrees of freedom for GWPR models are likely to be small, a small sample bias adjustment should be included in the AIC definition. Hence, the corrected Akaike information criterion (AICc) is defined as follows:

$$AIC_c = -2 \times LL + 2k + 2\frac{k(k+1)}{N-k-1} \tag{9.42}$$

where LL is the log-likelihood and k is the effective number of parameters in the model, respectively. If the effective number of parameters, k, is small relative to the number of observations, N, then the difference between AIC and AICc can be neglected.

Example 9.6

Develop planning level transportation safety tools using geographically weighted Poisson regression

Hadayeghi et al. (2010) investigated the spatial variations in the relationship between the number of zonal crashes and the transportation planning predictors using the GWPR modeling technique. This study was based on 481 traffic analysis zones in 2001, Toronto, CA. Other data sources include one-day travel survey data, traffic volume, and street network and land use data. The dependent variable of each developed model is the number of zonal crashes per year. The model parameters were estimated using the maximum likelihood method for GWPR in the "GWRx3.0" package. Model results are presented in Table 9.4.

TABLE 9.4 Model results.

Parameters	GWPR model number							
	#1	#2	#3	#4	#5	#6	#7	#8
ln(*A*)	−6.4, 8.10 (1.09, 2.4,8.1)	1.06, 4.60 (3.31,3.84,4.14)	−3.23, 6.92 (1.2,2.3,3.3)	−6.42, 7.41 (1.32,2.53,3.74)	−4.58, 7.31 (1.55,2.67,3.74)	−3.81, 5.88 (1.50,2.56,3.47)	−3.97, 6.23 (1.41,2.42,3.47)	−2.839, 5.526 (1.49,2.50,3.40)
ln(VKT)	−0.36, 1.37 (0.12,0.33,0.47)	−0.009, 0.38 (0.03,0.07,0.13)	−0.33, 0.89 (0.14,0.27,0.38)	−0.32, 1.298 (0.06,0.23,0.38)	−0.26,1.17 (0.07,0.22,0.35)	−0.224, 0.979 (0.08,0.21,0.34)	−0.22, 0.99 (0.09,0.21,0.34)	−0.165, 0.857 (0.10,0.21,0.34)
Total arterial road kilometers		0.07, 0.33 (0.17,0.22,0.26)						
Total expressway kilometers		−0.15, 0.23 (0.05,0.08,0.12)						
Total collector kilometers		−0.02, 0.17 (0.05,0.09,0.12)						
Total laneway kilometers		−0.023, 0.48 (0.04,0.06,0.10)						
Total local road kilometers		−0.048, 0.016 (−0.02,−0.009,0.002)						
Total ramp kilometers		−0.29, 0.26 (0.01,0.07,0.13)						
Total road kilometers			−0.04, 0.16 (0.03,0.04,0.06)			−0.060, 0.175 (0.01,0.03,0.05)	−0.043, 0.156 (0.01,0.03,0.05)	−0.038, 0.118 (0.01,0.03,0.05)
Number of 4-legged signalized intersections				−0.165, 0.487 (0.08,0.14,0.21)		−0.083, 0.372 (0.05,0.10,0.17)	−0.098, 0.350 (0.05,0.10,0.16)	−0.069, 0.347 (0.05,0.09,0.15)
Number of 3-legged signalized intersections				−0.514, 0.564 (0.04,0.15,0.23)		−0.195, 0.543 (0.06,0.13,0.19)	−0.23, 0.474 (0.06,0.12,0.18)	−0.269, 0.419 (0.06,0.13,0.19)
Total number of signalized intersections					−0.067, 0.348 (0.09,0.13,0.18)			
Total rail kilometers								−0.56, 1.12 (−0.11,0.01,0.15)
Number of schools							−0.26, 0.24 (−0.06,0.00,0.05)	
GWPR AICc	16,304	13,404	14,602	8,137	9,910	8,220	7,629	8,032
Global AICc	27,410	20,431	24,373	20,041	20,161	18,805	18,634	18,779

Macrolevel CWPR collision prediction models based on traffic and road network variables, total collisions. Minimum, Maximum (Lower Quartile, Median, Upper Quartile). *From Hadayeghi, A., Shalaby, A.S., Persaud, B.N., 2010. Development of planning level transportation safety tools using Geographically Weighted Poisson Regression. Accid. Anal. Prev. 42 (2), 676–688.*

Example 9.6 (*cont'd*)

It is clear from the table that the signs of traffic exposure (e.g., VKT) coefficients for each TAZ are not always the same. This is counterintuitive because traffic exposure is expected to have a positive effect on the number of crashes; therefore, its coefficient should be positive. Fig. 9.7 depicts the local parameters for both total and severe crash models for each TAZ, respectively. The parameters clearly demonstrate spatial variations across the city.

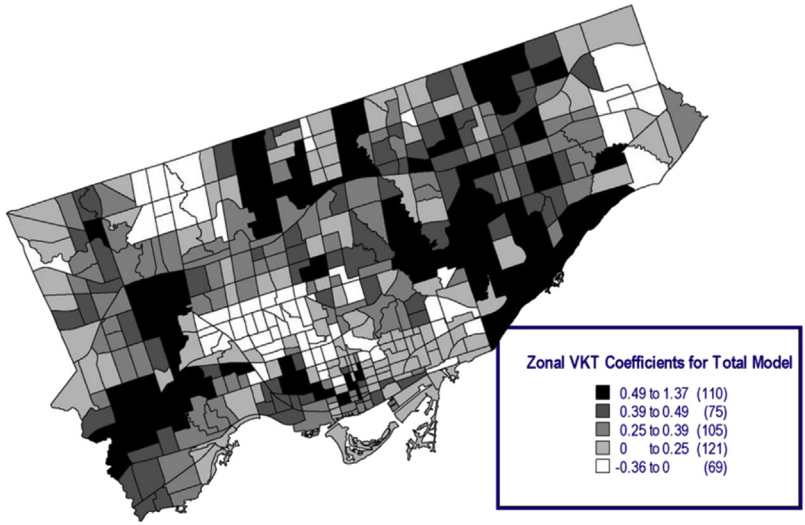

FIGURE 9.7 Spatial variations in observed relationships. *From Hadayeghi, A., Shalaby, A.S., Persaud, B.N., 2010. Development of planning level transportation safety tools using Geographically Weighted Poisson Regression. Accid. Anal. Prev. 42(2), 676–688.*

References

Aguero-Valverde, J., Jovanis, P.P., 2006. Spatial analysis of fatal and injury crashes in Pennsylvania. Accid. Anal. Prev. 38 (3), 618–625.

Aguero-Valverde, J., Jovanis, P.P., 2008. Analysis of road crash frequency with spatial models. Transport. Res. Rec. 2061 (1), 55–63.

Aguero-Valverde, J., Jovanis, P.P., 2010. Spatial correlation in multilevel crash frequency models: effects of different neighboring structures. Transport. Res. Rec. 2165 (1), 21–32.

Anselin, L., 1995. Local indicators of spatial association – LISA. Geogr. Anal. 27, 93–115.

Banergjee, S., Carlin, B.P., Gelfand, A.E., 2004. Hierarchical Modeling and Analysis for Spatial Data. CRC Press.

Besag, J., 1974. Spatial interaction and the statistical analysis of lattice systems. J. Roy. Stat. Soc. B 36 (2), 192–225.
Besag, J., 1975. Statistical analysis of non-lattice data. J. Roy. Stat. Soc. Series D (The Stat.) 24 (3), 179–195.
Besag, J., York, J., Mollíe, A., 1991. Bayesian image restoration, with two applications in spatial statistics. Ann. Inst. Stat. Math. 43 (1), 1–20.
Bowman, A.W., 1984. An alternative method of cross-validation for the smoothing of density estimates. Biometrika 71, 353–360.
Charlton, M., Fotheringham, A.S., Brunsdon, C., 2003. Software for Geographically Weighted Regression. University of Newcastle, Newcastle, United Kingdom.
Cressie, N.A.C., 1993. Statistics for Spatial Data. Wiley, New York.
da Silva, A.R., Rodrigues, T.C.V., 2014. Geographically weighted negative binomial regression—incorporating overdispersion. Stat. Comput. 24 (5), 769–783.
Flahaut, B., Mouchart, M., Martin, E.S., Thomas, I., 2003. The local spatial autocorrelation and the kernel method for identifying black zones: a comparative approach. Accid. Anal. Prev. 35 (6), 991–1004.
Fotheringham, A.S., Brunsdon, C., Charlton, M.E., 2000. Quantitative Geography, Perspectives on Spatial Data Analysis. SAGE.
Fotheringham, A.S., Brunsdon, C., Charlton, M.E., 2002. Geographically Weighted Regression: the Analysis of Spatially Varying Relationship. Wiley, Chichester.
Gelfand, A.E., Vounatsou, P., 2003. Proper multivariate conditional autoregressive models for spatial data analysis. Biostatistics 4 (1), 11–15.
Gelman, A., 2006. Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). Bayesian Anal. 1 (3), 515–534.
Getis, A., Ord, J.K., 1992. The analysis of spatial association by use of distance statistics. Geogr. Anal. 24 (No. 3), 189–206.
Gomes, M.J.T.L., Cunto, F., da Silva, A.R., 2017. Geographically weighted negative binomial regression applied to zonal level safety performance models. Accid. Anal. Prev. 106, 254–261.
Hadayeghi, A., Shalaby, A.S., Persaud, B.N., 2010. Development of planning level transportation safety tools using Geographically Weighted Poisson Regression. Accid. Anal. Prev. 42 (2), 676–688.
Khan, G., Qin, X., Noyce, D.A., 2008. Spatial analysis of weather crash patterns. J. Transport. Eng. 134 (5), 191–202.
Khan, G., Santiago-Chaparro, K.R., Qin, X., Noyce, D.A., 2009. Application and integration of lattice data analysis, network K-functions, and geographic information system software to study ice-related crashes. Transport. Res. Rec. 2136 (1), 67–76.
Levine, N., Kim, K.E., Nitz, L.H., 1995. Spatial analysis of Honolulu motor vehicle crashes: II. Zonal generators. Accid. Anal. Prev. 27 (5), 675–685.
Li, Z., Wang, W., Pan, L., Bigham, J.M., Ragland, D.R., 2013. Using geographically weighted Poisson regression for county-level crash modeling in California. Saf. Sci. 58, 89–97.
Miaou, S.P., Song, J.J., 2005. Bayesian ranking of sites for engineering safety improvements: decision parameter, treatability concept, statistical criterion, and spatial dependence. Accid. Anal. Prev. 37, 699–720.
Miaou, S.P., Song, J.J., Mallick, B.K., 2003. Roadway traffic crash mapping: a space-time modeling approach. J. Transport. Stat. 6 (1), 33–57, 1094-8848.
Moran, P.A.P., 1950. Notes on continuous Stochastic phenomena. Biometrika 37 (1), 17–23, 10.2307/2332142. JSTOR 2332142.
Nakaya, T., Fotheringham, A.S., Brunsdon, C., Charlton, M., 2005. Geographically weighted Poisson regression for disease association mapping. Stat. Med. 24 (17), 2695–2717.
Okabe, A., Yamada, I., 2001. The K-function method on a network and its computational implementation. Geogr. Anal. 33 (3), 271–290.

Okabe, A., Okunuki, K.I., Shiode, S., 2006. SANET: a toolbox for spatial analysis on a network. Geogr. Anal. 38 (1), 57–66.
Ord, J.K., Getis, A., 1995. Local spatial autocorrelation statistics: distributional issues and an application. Geogr. Anal. 27, 286–306.
Qin, X., Han, J., Zhu, J., 2009. Spatial analysis of road weather safety data using a Bayesian hierarchical modeling approach. Adv. Transport. Stud. 18, 69–78.
Quddus, M.A., 2008. Modelling area-wide count outcomes with spatial correlation and heterogeneity: an analysis of London crash data. Accid. Anal. Prev. 40 (4), 1486–1497.
Ripley, B.D., 1976. The second-order analysis of stationary point process. J. Appl. Probab. 13, 255–266.
Silva, A.R., Rodrigues, T.C.V., 2016. A SAS. Available at: http://support.sas.com/resources/papers/proceedings16/8000-2016.pdf.
Spiegelhalter, D., Best, N., Carlin, B.P., Linde, A., 2002. Bayesian measures of model complexity and fit. J. Roy. Stat. Soc. B 64 (4), 583–639.
Tobler, W., 1970. A computer movie simulating urban growth in the Detroit region. Econ. Geogr. 46, 234–240. Supplement.
Wakefield, J.C., Best, N.G., Waller, L., 2000. Bayesian Approaches to Disease Mapping, on Spatial Epidemiology: Methods and Applications. Oxford University Press.
Xu, P., Huang, H., 2015. Modeling crash spatial heterogeneity: random parameter versus geographically weighting Accid. Anal. Prev. 16–25.
Yamada, I., Thill, J.C., 2004. Comparison of planar and network K-functions in traffic accident analysis. J. Transport Geogr. 12 (2), 149–158.