CHAPTER 8

Spatial Exposure Data

I speculate. Mapmakers are entitled to do so, since they readily acknowledge that they are rarely in possession of all the facts. They are always dealing with secondary accounts, the tag ends of impressions. Theirs is an uncertain science.

James Cowan, *A Mapmaker's Dream, the Meditations of Fra Mauro, Cartographer to the Court of Venice*, 1996, Warner Books, p. 11

We now turn to a different aspect of the spatial analysis of public health data: spatial exposure information. When analyzing spatial patterns of disease, we may also want to study spatial patterns of potential exposures. In fact, we often need to consider spatial exposure information to understand the observed spatial variation in disease. As we noted in Chapter 1, Palm (1890) was among the first to link the spatial distribution of a disease (rickets in Great Britain) to a spatial exposure when he noted that sunlight deficiency was an important component in the etiology of this disease. Spatial analysis also led Blum (1948) and Lancaster (1956) to identify sunlight as a causal factor in skin cancer. While latitude is a surrogate for sunlight that is easily recorded, other studies may require exposure assessment and maps that are based on field measurements. This is particularly true in environmental health applications in which environmental measures (e.g., air pollution, groundwater contamination) are thought to exacerbate disease. Spatial exposure assessment is also important from a health policy standpoint since environmental monitoring data often formulate regional air and water quality standards.

Maps of both disease and potential exposures form the basis for *geographical correlation studies* that attempt to draw inferences about disease risk in relation to spatially varying risk factors. We discuss statistical methods for such studies in Chapter 9. In this chapter we focus on the exposure component alone and statistical methods for mapping spatial exposure data. We will borrow much of the methodology from the field of *geostatistics*, a field of statistics concerned with the study of spatial data that have a continuous spatial index (i.e., data can be observed at any location within a domain of interest, at least conceptually). This is in contrast to aggregated spatial data (discussed in Chapter 7), which are associated with areal

Applied Spatial Statistics for Public Health Data, by Lance A. Waller and Carol A. Gotway
ISBN 0-471-38771-1

regions for which there is no opportunity for measurement between locations. In geostatistics, the locations of the data are assumed to be fixed and known, not random as is the case with spatial point patterns discussed in Chapters 5 and 6. In using geostatistical methodology, we seek a general statistical model to infer the characteristics of the spatial process that gives rise to the data we observe. Much of the material in this chapter draws heavily from that in Cressie (1993) and from Carol Gotway's coursework and interaction with Noel Cressie as one of his first Ph.D. students in spatial statistics, and his continued long-term mentoring as an incredible resource on spatial statistics.

8.1 RANDOM FIELDS AND STATIONARITY

Suppose that we have spatial exposure data $Z(\mathbf{s}_1), Z(\mathbf{s}_2), \ldots, Z(\mathbf{s}_N)$ that represent observations of a variable Z at spatial locations $\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_N$. The spatial locations may be aligned on a regular grid or distributed irregularly throughout some domain of interest, D. We restrict our attention to two dimensions, so each location is a two-dimensional vector, $\mathbf{s} = (x, y)$, referencing a point in the plane.

In geostatistical applications, the data are assumed to be a partial realization of a random process (called a *stochastic process* or *random field*)

$$\{Z(\mathbf{s}) : \mathbf{s} \in D\}$$

where D is a fixed subset of $\Re^2$, and the spatial index, $\mathbf{s}$, varies continuously throughout D. Thus, for a fixed location $\mathbf{s}$, $Z(\mathbf{s})$ is a random variable to which the laws of probability apply; for a fixed realization of this process, we observe a function of space: namely, the data at locations $\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_N$. The data are only a partial realization of a spatial function since we cannot, for practical reasons, observe the process at every point in D.

This model for spatial data makes traditional statistical inference difficult since we do not have independent replication. As mentioned in Chapter 5, a facsimile of replication is provided by the concept of *stationarity*. With a random field, if

$$E[Z(\mathbf{s})] = \mu \qquad \text{for all } \mathbf{s} \in D \tag{8.1}$$

(i.e., the mean of the process does not depend on location) and

$$\operatorname{Cov}\big(Z(\mathbf{s}_i), Z(\mathbf{s}_j)\big) = C(\mathbf{s}_i - \mathbf{s}_j) \qquad \text{for all } \mathbf{s}_i, \mathbf{s}_j \in D \tag{8.2}$$

(i.e., the covariance depends only on the difference between locations $\mathbf{s}_i$ and $\mathbf{s}_j$, *not on the locations themselves*), then $Z(\cdot)$ is said to be *second-order stationary*. The function $C(\cdot)$ defined by equation (8.2) is called the *covariance function* and is one measure of spatial autocorrelation. Thus, through the assumption of stationarity, the process essentially repeats itself in space, providing the replication necessary for estimation and inference. If, in addition, $C(\mathbf{s}_i - \mathbf{s}_j)$ is a function only of the distance between $\boldsymbol{s}_i$ and $\boldsymbol{s}_j$ and not direction, the process is called *isotropic*. If $C(\boldsymbol{s}_i - \boldsymbol{s}_j)$ depends on both distance and direction, the spatial process is called *anisotropic*. The notions of stationarity and isotropy correspond conceptually to those defined for spatial point processes in Chapter 5, although equations (8.1) and (8.2) phrase the concepts directly in terms of the Z process.

We do not need stationarity to work with random fields and geostatistics, but it provides a convenient place to begin our development of geostatistical methods for public health data. Methods for handling nonstationary exposure data are described in Section 9.2.

8.2 SEMIVARIOGRAMS

We have seen in Chapter 7 that one of the fundamental attributes of spatial data is spatial autocorrelation: observations closer together tend to be more alike than observations farther apart. In geostatistics, this idea of autocorrelation is quantified through a function called a *semivariogram*.

Suppose that in addition to the constant mean assumption given in equation (8.1), $\{Z(\boldsymbol{s}) : \boldsymbol{s} \in D\}$ also satisfies

$$\mathrm{Var}(Z(\boldsymbol{s}_i) - Z(\boldsymbol{s}_j)) = 2\gamma(\boldsymbol{s}_i - \boldsymbol{s}_j), \quad \boldsymbol{s}_i, \boldsymbol{s}_j \in D. \tag{8.3}$$

Then $Z(\cdot)$ is said to be *intrinsically stationary* and the function $2\gamma(\cdot)$ defined by equation (8.3) is called a *variogram*. If a process is intrinsically stationary [i.e., it satisfies equations (8.1) and (8.3)], $2\gamma(\boldsymbol{h})$ is a function of the *spatial lag,* $\boldsymbol{h} = \boldsymbol{s} - \boldsymbol{u}$, *but not of the locations* $\boldsymbol{s}$ *and* $\boldsymbol{u}$. Note that the definition of intrinsic stationarity is very similar to that of second-order stationarity, where the former is defined in terms of the variogram and the latter in terms of the covariance function. In fact, the variogram is a generalization of the covariance function, and under the assumption of second-order stationarity, the two functions are related, as we shall see below.

The function $\gamma(\cdot)$ is called a *semivariogram*, as it is one-half the variogram. The semivariogram is central to the field of geostatistics. Although many authors use the terms *variogram* and *semivariogram* interchangeably, they clearly differ: One is twice the other. This may not matter in some calculations, but in others the distinction can be crucial. Since we have seen too many studies (and even theoretical results) off by a factor of 2, we will distinguish the variogram from the semivariogram and use just the latter term.

The semivariogram is a function of the spatial process and as such satisfies certain properties:

(i) $\gamma(-\boldsymbol{h}) = \gamma(\boldsymbol{h})$ [i.e., the autocorrelation between $Z(\boldsymbol{s})$ and $Z(\boldsymbol{u})$ is the same as that between $Z(\boldsymbol{u})$ and $Z(\boldsymbol{s})$].

(ii) $\gamma(\boldsymbol{0}) = 0$, since by definition, $\mathrm{Var}(Z(\boldsymbol{s}) - Z(\boldsymbol{s})) = 0$.

(iii) $\gamma(\boldsymbol{h})/\|\boldsymbol{h}\|^2 \longrightarrow 0$ as $\|\boldsymbol{h}\| \longrightarrow \infty$, where $\|\boldsymbol{h}\|$ denotes the length of the vector $\boldsymbol{h}$.

(iv) $\gamma(\cdot)$ must be *conditionally negative definite*; that is,

$$\sum_{i=1}^{m}\sum_{j=1}^{m} a_i a_j \gamma(\boldsymbol{s}_i - \boldsymbol{s}_j) \le 0 \tag{8.4}$$

for any finite number of locations $\{\boldsymbol{s}_i : i = 1, \dots, m\}$ and real numbers $\{a_1, \dots, a_m\}$ satisfying $\sum_{i=1}^{m} a_i = 0$. This condition is the analog of the positive-definite condition for covariance functions and variance–covariance matrices, ensuring that all variances are nonnegative.

(v) If the spatial process is isotropic, $\gamma(\boldsymbol{h}) \equiv \gamma(h)$, where $h = \|\boldsymbol{h}\|$ (i.e., the semivariogram is a function of distance alone).

A graph of a semivariogram plotted against separation distance, $\|\boldsymbol{h}\|$, conveys information about the continuity and spatial variability of the process. This graph starts at zero, and if observations close together are more alike than those farther apart, increases as the separation distance increases. In this way, increasing variation in pairwise differences with increasing distance reflects decreasing spatial autocorrelation, since $Z(\boldsymbol{s})$ and $Z(\boldsymbol{u})$ can vary more with respect to each other as locations $\boldsymbol{s}$ and $\boldsymbol{u}$ move farther apart. Often, the semivariogram will level off to nearly a constant value (called the *sill*) at a large separation distance (called the *range*). Beyond this distance, observations are spatially uncorrelated, reflected by a (near) constant variance in pairwise differences. These properties are depicted in Figure 8.1. Note that if the spatial process is not isotropic, the semivariogram and the information it conveys will differ with direction and we can envision several graphs like Figure 8.1, one for each direction. If there is no autocorrelation between $Z(\boldsymbol{s})$ and $Z(\boldsymbol{u})$, the semivariogram will be a horizontal line.

The shape of the semivariogram near the origin is of particular interest since it indicates the degree of smoothness or spatial continuity of the spatial variable under study. A parabolic shape near the origin arises with a very smooth spatial variable

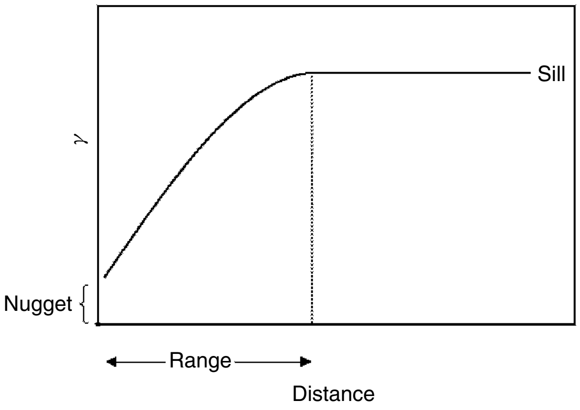

FIG. 8.1 Typical semivariogram.

that is both continuous and differentiable. A linear shape near the origin reflects a variable that is continuous but not differentiable, and hence less regular. A discontinuity, or vertical jump, at the origin [i.e., $\lim_{\boldsymbol{h}\to\mathbf{0}} \gamma(\boldsymbol{h}) = c_0 > 0$] indicates that the spatial variable is not continuous and has highly irregular spatial variability. This discontinuity is called the *nugget effect* in the geostatistical literature. If the process has a large nugget effect, two observations fairly close together have very different values. This is often due to measurement error, but can also simply indicate a spatially discontinuous process. (The term *nugget effect* comes from mining. In mining gold ore, we may not find ore at one location, but then at a nearby location find a gold nugget.) The nugget effect is a discontinuity; by definition [see equation (8.3)] the semivariogram at the origin (i.e., at zero separation distance) is always zero.

8.2.1 Relationship to Covariance Function and Correlogram

The covariance function defined in equation (8.2) is related to the semivariogram in the following way. If $Z(\cdot)$ is second-order stationary [i.e., satisfied equations (8.1) and (8.2)], then

$$\gamma(\boldsymbol{h}) = C(\mathbf{0}) - C(\boldsymbol{h}).$$

If $C(\boldsymbol{h}) \longrightarrow 0$ as $\|\boldsymbol{h}\| \longrightarrow \infty$, then $\gamma(\boldsymbol{h}) \longrightarrow C(\mathbf{0})$. So $C(\mathbf{0})$ is the variance of $Z(\boldsymbol{s})$ and the sill of the semivariogram. When there is a nugget effect, the *partial sill* is defined as the difference between the process variance (sill) and the nugget effect, or $C(\mathbf{0}) - c_0$. The term *relative nugget effect* refers to the percentage of the total sill comprised of the nugget effect. Formally, the *range* of the semivariogram in the direction $\boldsymbol{r}_0/\|\boldsymbol{r}_0\|$ is the smallest length $\|\boldsymbol{r}_0\|$ such that $\gamma(\boldsymbol{r}_0(1+\varepsilon)) = C(\mathbf{0})$ for any $\varepsilon > 0$ (i.e., for any distance larger than $\|\boldsymbol{r}_0\|$, the semivariogram equals the sill). Practically, for any fixed direction, it is the minimum distance, r, for which $\gamma(r) = C(0)$.

Sometimes, particularly if we are comparing two spatial processes, it is useful to use a measure of correlation instead of covariance. Thus, we can define the spatial *correlogram* as

$$\rho(\boldsymbol{h}) = C(\boldsymbol{h})/C(\mathbf{0}).$$

The definition of $\rho(\boldsymbol{h})$ is analogous to that of a typical correlation [i.e., scaled so that $|\rho(\boldsymbol{h})| \le 1$].

Since these functions are all related, why do we use the semivariogram instead of a covariance function? Theoretically, the class of intrinsically stationary processes (those with a valid semivariogram) is more general than the class of second-order stationary processes (those with a valid covariance function). But only barely, and processes that have a semivariogram but not a covariance function rarely arise in public health applications. The curious can refer to Cressie (1993, p. 68) for an example of a process (Brownian motion) for which $\gamma(\cdot)$ is defined but $C(\cdot)$ is not. A practical reason for preferring the semivariogram to the covariance function is that estimation of the semivariogram from the data observed is more reliable than estimation of the covariance function, since estimation of the semivariogram does not require estimation of the mean (see Cressie 1993).

8.2.2 Parametric Isotropic Semivariogram Models

There are many parametric functions that satisfy the properties of the semivariogram (see, e.g., Journel and Huijbregts 1978; Chilès and Delfiner 1999). We say that a semivariogram model is valid in d dimensions (i.e., in $\Re^d$) if it satisfies the conditional negative-definite property defined in Section 8.2 [equation (8.4)]. A closer look at some of these will give us a better understanding of the semivariogram and its relationship to the spatial process of interest. We also use these as models for the empirical semivariogram in Section 8.2.4. Since these are isotopic models, we write them in terms of a generic lag distance, denoted h. Graphs of these theoretical semivariograms illustrate their differences (Figure 8.2).

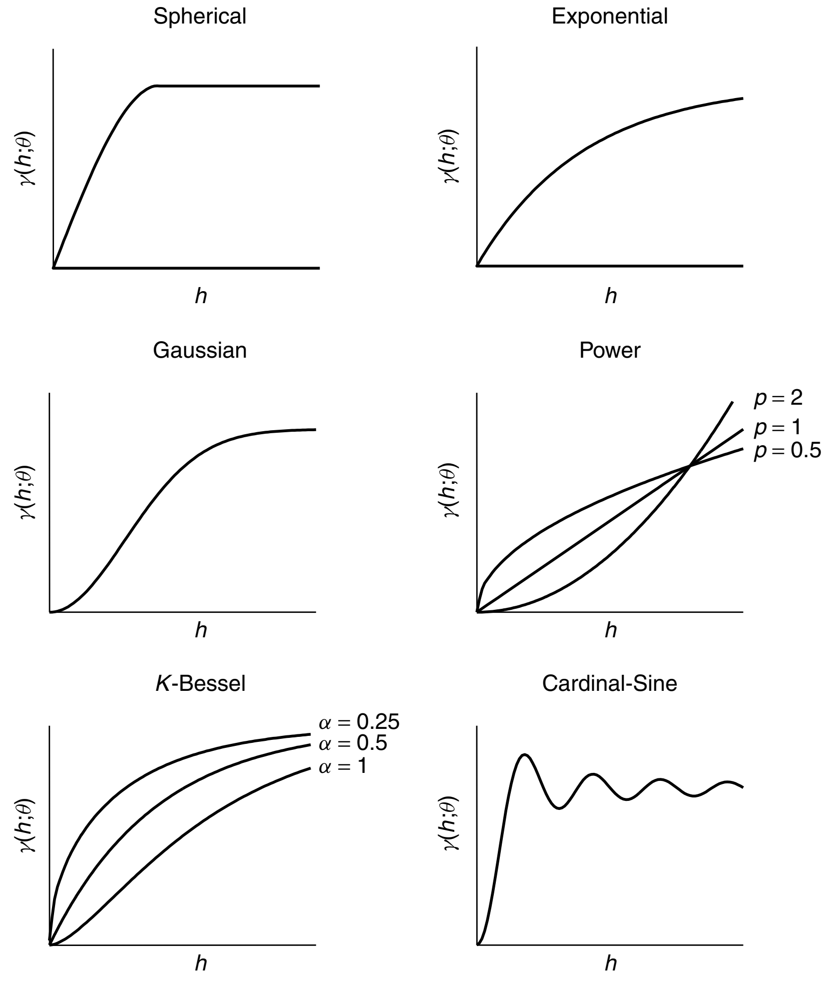

FIG. 8.2 Selected theoretical semivariogram models.

- *Spherical:*

$$\gamma(h;\boldsymbol{\theta}) = \begin{cases} 0, & h = 0 \\ c_0 + c_s\left[(3/2)(h/a_s) - (1/2)(h/a_s)^3\right], & 0 < h \le a_s \\ c_0 + c_s, & h > a_s, \end{cases} \tag{8.5}$$

where $\boldsymbol{\theta} = (c_0, c_s, a_s)'$; $c_0 \ge 0, c_s \ge 0, a_s > 0$. It is valid in $\Re^d, d = 1, 2, 3$. The spherical semivariogram is nearly linear near the origin. The parameter c_0 measures the nugget effect, c_s is the partial sill (so $c_0 + c_s$ is the sill), and a_s is the range.

- *Exponential:*

$$\gamma(h;\boldsymbol{\theta}) = \begin{cases} 0, & h = 0 \\ c_0 + c_e\left[1 - \exp(-h/a_e)\right], & h > 0, \end{cases} \tag{8.6}$$

where $\boldsymbol{\theta} = (c_0, c_e, a_e)'$; $c_0 \ge 0, c_e \ge 0, a_e > 0$. It is valid in $\Re^d, d \ge 1$. The exponential semivariogram rises more slowly from the origin than does the spherical semivariogram. As with the spherical model, c_0 is the nugget effect and c_e is the partial sill (so $c_0 + c_e$ is the sill). However, this model approaches the sill asymptotically, so the range is not a_e, since $\gamma(a_e) \ne c_0 + c_e$. The "effective range" (traditionally defined as the distance at which the autocorrelation is 0.05) is $3a_e$.

- *Gaussian:*

$$\gamma(h;\boldsymbol{\theta}) = \begin{cases} 0, & h = 0 \\ c_0 + c_g\left\{1 - \exp[-(h/a_g)^2]\right\}, & h > 0, \end{cases}$$

where $\boldsymbol{\theta} = (c_0, c_g, a_g)'$; $c_0 \ge 0, c_g \ge 0, a_g > 0$. It is valid in $\Re^d, d \ge 1$. The Gaussian semivariogram model is parabolic near the origin, indicative of a very smooth spatial process. Many argue that such processes rarely arise in practice, although the Gaussian model is often deemed best by automatic model-fitting criteria. This is a valid semivariogram model; however, its use can often lead to singularities in spatial prediction equations (Davis and Morris 1997). Wackernagel (1995) calls it "pathological" since it corresponds to a deterministic process and thus contradicts the underlying randomness assumption in geostatistics. Although we do not take such an extreme view, we do recommend that the Gaussian semivariogram model only be used with caution and never without a lot of closely spaced data to assess behavior near the origin. Similar to the previous models, c_0 measures the nugget effect and c_g is the partial sill (so $c_0 + c_g$ is the sill). The effective range is $\sqrt{3}\, a_g$.

- *Power:*

$$\gamma(h;\boldsymbol{\theta}) = \begin{cases} 0, & h = 0 \\ c_0 + bh^p, & h > 0, \end{cases}$$

where $\boldsymbol{\theta} = (c_0, b, p)'$; $c_0 \geq 0, b \geq 0, 0 \leq p < 2$. It is valid in $\Re^d, d \geq 1$. Models in this family do not have a sill or a range, so spatial correlation does not level off for large lag distances. They play an important role in fractal processes and the estimation of the fractal dimension [see Gallant et al. (1994) for a comparative overview of fractal dimension estimation]. The linear model, obtained by taking $p = 1$, is the most common member of this class.

- *Stable:*

$$\gamma(h; \boldsymbol{\theta}) = \begin{cases} 0, & h = 0 \\ c_0 + c_t \left\{1 - \exp[(-h/a_t)^{\alpha}]\right\}, & h > 0, \end{cases}$$

where $\boldsymbol{\theta} = (c_0, c_t, a_t, \alpha)'$; $c_0 \geq 0, c_t \geq 0, a_t > 0, 0 \leq \alpha \leq 2$. It is valid in $\Re^d, d \geq 1$. Near the origin, models in the stable semivariogram family have the same behavior as models in the power family. The behavior near the origin is determined by α. However, they do reach a sill (c_0 measures the nugget effect and c_t is the partial sill, so $c_0 + c_t$ is the sill). As with the exponential model that is a member of this family, models in this family approach the sill asymptotically, so the range is not a_t. The effective range depends on both a_t and α. Given a_t, models with smaller values of α will approach the sill more slowly.

- *K-Bessel (Matérn):*

$$\gamma(h; \boldsymbol{\theta}) = \begin{cases} 0, & h = 0 \\ c_0 + c_k \left[1 - \dfrac{1}{2^{\alpha-1}\Gamma(\alpha)}\left(\dfrac{h}{a_k}\right)^{\alpha} K_{\alpha}\dfrac{h}{a_k}\right], & h > 0, \end{cases}$$

where $\boldsymbol{\theta} = (c_0, c_k, a_k, \alpha)'$; $c_0 \geq 0, c_k \geq 0, a_k > 0, \alpha \geq 0$, $K_{\alpha}(\cdot)$ is the modified Bessel function of the second kind of order α, and $\Gamma(\cdot)$ is the gamma function. It is valid in $\Re^d, d \geq 1$. This family of models has long been referred to as the *K-Bessel model* in geostatistics, due to its dependence on $K_{\alpha}(\cdot)$. Recently, statisticians rediscovered the utility of this family and renamed it the *Matérn class*, based on its initial presentation by Matérn (1960). Here, c_0 measures the nugget effect and c_k is the partial sill (so $c_0 + c_k$ is the sill). Models in this family approach the sill asymptotically. As in the stable family, the behavior near the origin is determined by α, and the parameter a_k controls the range. The Gaussian semivariogram model is obtained in the limit by letting $\alpha \to \infty$, and the exponential semivariogram corresponds to the case where $\alpha = \frac{1}{2}$. An advantage of this family of models (and the stable family) is that the behavior of the semivariogram near the origin can be estimated from the data rather than assumed to be of a certain form. However, the computation of $K_{\alpha}(\cdot)$ needed for this estimation is cumbersome and, as for the Gaussian model, requires some closely spaced data.

- *Cardinal-Sine:*

$$\gamma(h;\boldsymbol{\theta}) = \begin{cases} 0, & h = 0 \\ c_0 + c_w\left(1 - \dfrac{a_w}{h}\sin\dfrac{h}{a_w}\right), & h > 0, \end{cases}$$

  where $\boldsymbol{\theta} = (c_0, c_w, a_w)'$; $c_0 \geq 0, c_w \geq 0, a_w > 0$. It is valid in $\Re^d$, $d = 1, 2, 3$. The cardinal-sine model is one member of a family of models called *hole-effect* models that are parameterized by a more general form called the *J-Bessel model* [see Chilès and Delfiner (1999) for the equation of the J-Bessel model]. These models, and the cardinal-sine model in particular, are useful for processes with negative spatial autocorrelation or processes with cyclical or periodic variability. It reaches a maximum and then continues to oscillate around the sill with a period of a_w.

There are many more parametric semivariogram models not described here [see, e.g., Armstrong (1999), Chilès and Delfiner (1999), and Olea (1999) for several others]. In addition, the sum of two semivariogram models that are both valid in $\Re^d$ is also a valid semivariogram model in $\Re^d$, so more complex processes can be modeled by adding two or more of these basic semivariogram models (Christakos 1984). Semivariogram models created this way are referred to as models of *nested structures.* Note that a model valid in $\Re^{d_2}$ is also valid in $\Re^{d_1}$, where $d_2 > d_1$, but the converse is not true. An important example is the piecewise linear or "tent" function:

$$\gamma(h;\boldsymbol{\theta}) = \begin{cases} 0, & h = 0 \\ c_0 + c_s h, & 0 \leq h \leq a_s \\ c_0 + c_s, & h > a_s, \end{cases} \tag{8.7}$$

where $\boldsymbol{\theta} = (c_0, c_s, a_s)'$; $c_0 \geq 0, c_s \geq 0, a_s \geq 0$. This model is a valid semivariogram in $\Re^1$ but not in $\Re^2$. Thus, this model should not be used with spatial processes.

8.2.3 Estimating the Semivariogram

The semivariogram can be estimated easily from data $\{Z(\boldsymbol{s}_i) : i = 1, \ldots, N\}$ under the assumption of intrinsic stationary so that equations (8.1) and (8.3) hold. Using rules of expectation, we can write the variogram as

$$\begin{aligned} 2\gamma(\boldsymbol{h}) &= \mathrm{Var}(Z(\boldsymbol{s}+\boldsymbol{h}) - Z(\boldsymbol{s})) \\ &= E[(Z(\boldsymbol{s}+\boldsymbol{h}) - Z(\boldsymbol{s}))^2] - [E(Z(\boldsymbol{s}+\boldsymbol{h}) - Z(\boldsymbol{s}))]^2. \end{aligned}$$

From equation (8.1), $E[Z(\boldsymbol{s}_i)] = \mu$ for all i, so the second term is zero. Thus, to estimate the variogram we need only estimate $E[(Z(\boldsymbol{s}+\boldsymbol{h}) - Z(\boldsymbol{s}))^2]$. Since expectations are just statistical averages, one way to estimate this term is to average all observed squared differences $[Z(\boldsymbol{s}_i) - Z(\boldsymbol{s}_j)]^2$ for pairs of observations taken the same distance apart in the same direction (i.e., for all $\boldsymbol{s}_i, \boldsymbol{s}_j$ such that $\boldsymbol{s}_i - \boldsymbol{s}_j = \boldsymbol{h}$). This is the rationale behind the *method of moments estimator* of the semivariogram, given by

$$\widehat{\gamma}(\boldsymbol{h}) = \frac{1}{2|N(\boldsymbol{h})|} \sum_{N(\boldsymbol{h})} [Z(\boldsymbol{s}_i) - Z(\boldsymbol{s}_j)]^2, \qquad \boldsymbol{h} \in \Re^2, \tag{8.8}$$

where $N(\boldsymbol{h})$ is the set of distinct pairs separated by $\boldsymbol{h}$ [i.e., $N(\boldsymbol{h}) = \{(\boldsymbol{s}_i, \boldsymbol{s}_j) : \boldsymbol{s}_i - \boldsymbol{s}_j = \boldsymbol{h},\ i, j = 1, \ldots, N\}$ and $|N(\boldsymbol{h})| =$ the number of distinct pairs in $N(\boldsymbol{h})$].

Equation (8.8) gives what is often referred to as the *classical semivariogram estimator*. It gives point estimates of $\gamma(\cdot)$ at observed values of $\boldsymbol{h}$. If the process is isotropic, we need only consider pairs lagged $||\boldsymbol{h}||$ apart. If the process is anisotropic, the semivariogram can be estimated in different directions by selecting a particular direction and averaging pairs of data lagged $||\boldsymbol{h}||$ apart in that particular direction.

If we have data on a regular grid, we can easily define the lag distances, $||\boldsymbol{h}||$, and the directions using the grid. With irregularly spaced data, there may be only one pair of locations that is $\boldsymbol{h}$ apart (two for $\|\boldsymbol{h}\|$). Averages based on only one or two points are poor estimates with large uncertainties. We can reduce this variation and increase the accuracy of our point estimates by allowing a *tolerance* on the lags. Thus, we will define *tolerance regions* and group the sample pairs into these regions prior to averaging. This is analogous to the procedure used in making a histogram, adapted to two dimensions (see Figure 8.3). We average the pairwise squared differences for pairs of points in the tolerance regions to produce point estimates of the semivariogram at the average lag distances in each region. Each region should be small enough so that we retain enough spatial resolution to define the structure of the semivariogram, but also large enough so that we base each point estimate on a relative large number of paired differences. Typically, one specifies tolerance regions through the choice of five parameters: the direction of interest; the *angle tolerance*, which defines a sector centered on the direction of interest; the *lag spacing*, which defines the distances at which the semivariogram is estimated; the *lag tolerance*, which defines a distance interval centered at each lag; and the total number of lags at which we wish to estimate the semivariogram. Tolerance regions should include at least 30 pairs of points each to ensure that the empirical semivariogram at each point is well estimated (Journel and Huijbregts 1978). Usually, a set of directions and associated angle tolerances are chosen together so that they completely cover two-dimensional space. For example, we might choose the four main directions east-west, north-south, northeast-southwest, and northwest-southeast, and an angle tolerance of 22.5°. This partitions the space into four sectors that cover the plane completely. A more complete analysis of potential anisotropies might specify directions every 10° with a 5° angle tolerance. An isotropic semivariogram results when the angle tolerance is 90°, so we use all pairs to estimate the semivariogram. In a similar fashion, one typically defines distance classes by specifying the lag spacing and lag tolerance (directly analogous to the class spacing and number of

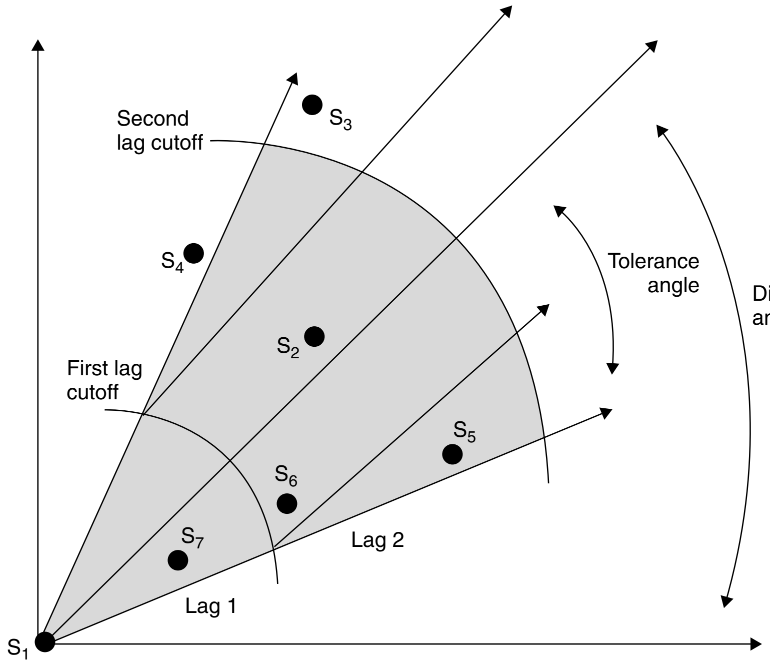

FIG. 8.3 Tolerance regions.

classes in a histogram). As a rule of thumb, one should construct these intervals so that the total number of lags is between 10 and 20 in order to see the structure of the semivariogram. Finally, note that estimates of the semivariogram at large lags rely only on points at the opposite ends of the domain. This usually results in very few pairs of data locations and wide variation in the estimates. Thus, in practice, we usually take the maximum lag distance to be about half the maximum separation distance (Journel and Huijbregts 1978). One should be careful of the use of very short maximum lag distances. The semivariogram is a picture of your data *spatially*: the sill and the range, if they exist, provide estimates of the process variance and the zone of influence of the observations, and information at larger lags can indicate large-scale trends that may be important to interpret.

DATA BREAK: Smoky Mountain pH Data The pH of a stream can affect organisms living in the water, and a change in the stream pH can be an indicator of pollution. Values of pH can range from 0 to 14, with 7 considered neutral; values less than 7 are termed *acidic* while pH values greater than 7 are called *alkaline*. Acidic stream water (with pH < 5) is particularly harmful to fish. In a study of the chemical properties of streams in the mid-Atlantic and southeastern United States, Kaufman et al. (1988) measured water pH at many locations within the Great Smoky Mountains. The locations of these measurements are shown in Figure 8.4. The relative variation in the pH values is also indicated in Figure 8.4,

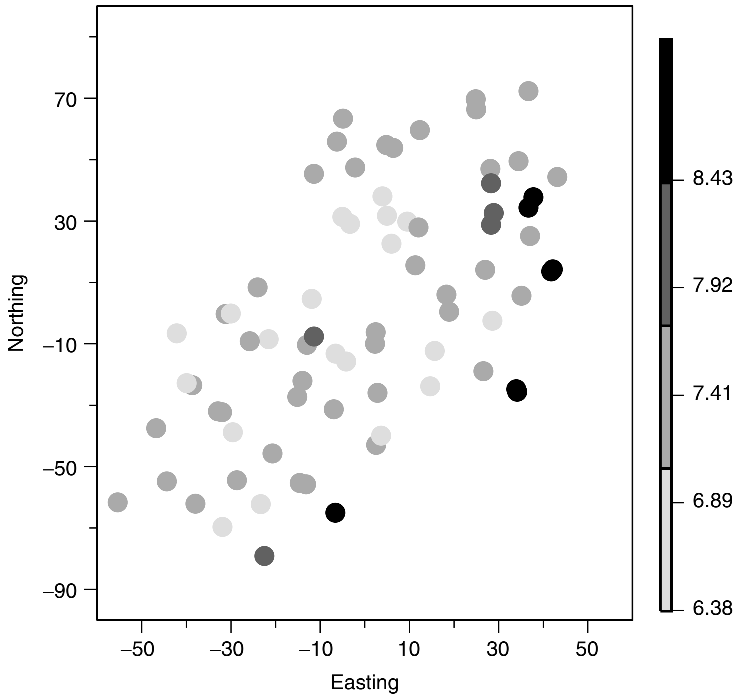

FIG. 8.4 Smoky Mountain pH sampling locations. Distances are measured in miles.

with the darker circles indicating higher values of pH. Initially, we assume that the water pH process is isotropic and estimate what is called an *omnidirectional* semivariogram. Here, we are simply concerned with the distances between points and not in directional variation. Since the data do not lie on a regular grid, we will have to define tolerance regions (which for an omnidirectional semivariogram are tolerance intervals on the lag distances) for semivariogram estimation.

We begin with the following specifications as a starting point: Let the maximum lag distance be one-half the maximum separation distance, which in our case is $\sqrt{(43.2 + 55.4)^2 + (72.3 + 79.2)^2}/2 = 90.4$ miles; set the number of lags equal to 20 (arbitrarily chosen); take the lag spacing = maximum lag distance/number of lags, or 4.5 miles; and set the lag tolerance = lag spacing/2, in this case 2.3 miles. Using these defaults with the pH values, we obtain the empirical semivariogram in shown Figure 8.5. The distance column gives the average distance between locations in each tolerance interval. It is also common simply to use the midpoint of each tolerance interval to indicate the spacing of the intervals. Notice that the estimates of the semivariogram at the first few lags are based on only a few pairs of observations and may not be very accurate. Also, the empirical semivariogram is a bit noisy and we may be able to obtain a clearer structure by taking a larger tolerance interval of, say, 10 miles. We may also want to extend the maximum

Lag	Distance	$\widehat{\gamma}(h)$	$N(h)$
1	1.20	0.02	7
2	4.11	0.04	15
3	9.58	0.12	59
4	13.38	0.13	93
5	18.05	0.10	99
6	22.79	0.17	122
7	26.75	0.14	132
8	31.70	0.17	167
9	35.80	0.22	174
10	40.80	0.17	152
11	45.03	0.13	163
12	49.39	0.19	139
13	54.34	0.27	142
14	58.39	0.22	144
15	63.13	0.14	125
16	67.60	0.32	146
17	72.04	0.23	105
18	76.44	0.17	107
19	80.50	0.21	74
20	85.78	0.24	85

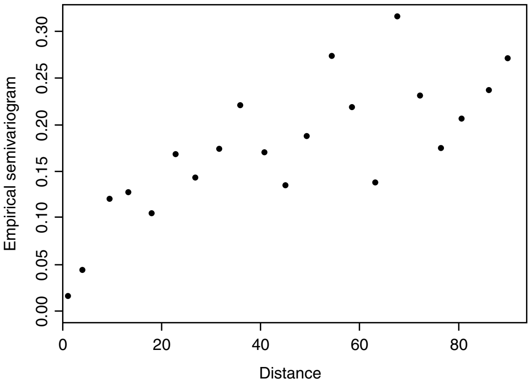

FIG. 8.5 Empirical semivariogram of pH values. The lag spacing is 4.5 miles.

lag distance to be sure that there is a well-defined sill. The resulting empirical semivariogram is shown in Figure 8.6.

The latter empirical semivariogram shows a much clearer relationship. It is always a good idea to experiment with several choices for the maximum lag distance and lag spacing. Semivariogram estimation is a mixture of both science and art. The goal is accurate estimation, a clear structure, and at least a total of 10 to 20 lags for modeling and inference. The guidelines presented here should work well in many applications. However, in others, the empirical semivariogram may appear erratic for many reasonable choices of the lag spacing and lag tolerance. Common problems with semivariogram estimation and more complex solutions are discussed in Section 8.4.1.

8.2.4 Fitting Semivariogram Models

The empirical semivariogram, $\widehat{\gamma}(\cdot)$, is not guaranteed to be conditionally nonnegative definite. This is not a problem if we limit ourselves to inferences about the spatial continuity of the process, but it can lead to problems when used for spatial prediction and mapping where we need reliable estimates of prediction uncertainty. Thus, we will need to find a valid theoretical semivariogram function that closely reflects the features of our empirical semivariogram. We limit our choices to a parametric family of theoretical variograms (like those described in Section 8.2.2) and seek to find the parameter estimates that best fit the data.

Nonlinear Least Squares Regression The idea here is to select a theoretical semivariogram family and find a vector of parameters $\widehat{\boldsymbol{\theta}}$ that makes this theoretical

Lag	Distance	$\widehat{\gamma}(h)$	$N(h)$
1	2.77	0.04	19
2	11.33	0.13	138
3	20.49	0.14	249
4	30.42	0.17	341
5	39.98	0.18	352
6	49.75	0.18	314
7	59.64	0.21	310
8	69.67	0.27	281
9	79.23	0.21	184
10	89.55	0.23	192
11	99.86	0.23	138
12	110.30	0.22	101
13	119.87	0.22	68

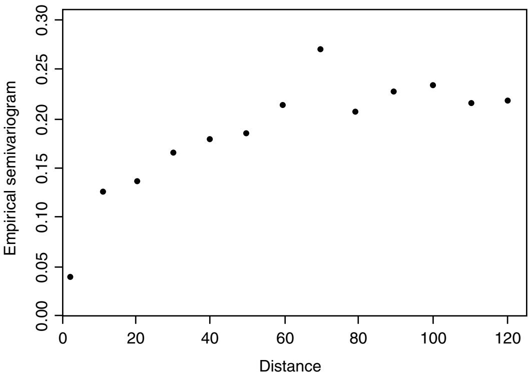

FIG. 8.6 Empirical semivariogram of pH values. The lag spacing is 10 miles.

model "close" to the empirical semivariogram. Let $\widehat{\gamma}(\cdot)$ be the *empirical semivariogram* estimated at K lags, $h(1), \ldots, h(K)$ and let $\gamma(h;\boldsymbol{\theta})$ be the *theoretical semivariogram* model whose form is known up to $\boldsymbol{\theta}$. Since the relationship between $\widehat{\gamma}(h)$ and h is usually nonlinear, nonlinear least squares regression can be used to estimate $\boldsymbol{\theta}$.

Nonlinear ordinary least squares (OLS) finds $\widehat{\boldsymbol{\theta}}$ minimizing the squared vertical distance between the empirical and theoretical semivariograms, that is, minimizing

$$\sum_{j=1}^{K}\left[\widehat{\gamma}(h(j)) - \gamma(h(j);\boldsymbol{\theta})\right]^2.$$

However, the estimates $[\widehat{\gamma}(h(j))]$ are correlated and have different variances, violating the general assumptions underling OLS theory. The usual statistical adjustment to OLS when observations are correlated and heteroskedastic is generalized least squares (GLS). Cressie (1985) applied nonlinear GLS to semivariogram estimation, finding $\widehat{\boldsymbol{\theta}}$ minimizing the objective function

$$[\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}(\boldsymbol{\theta})]'\boldsymbol{V}(\boldsymbol{\theta})^{-1}[\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}(\boldsymbol{\theta})],$$

where $\hat{\boldsymbol{\gamma}} = [\widehat{\gamma}(h(1)), \ldots, \widehat{\gamma}(h(K))]'$ with variance–covariance matrix $\boldsymbol{V}(\boldsymbol{\theta})$ and $\boldsymbol{\gamma}(\boldsymbol{\theta}) = \left[\gamma(h(1);\boldsymbol{\theta}), \ldots, \gamma(h(K), \boldsymbol{\theta})\right]'$. Since $\boldsymbol{V}(\boldsymbol{\theta})$ depends on $\boldsymbol{\theta}$ and $\boldsymbol{\theta}$ is unknown, this estimator is computed iteratively from starting values (one for each parameter in $\boldsymbol{\theta}$) that are improved (using, e.g., the Gauss–Newton algorithm) until the objective function is minimized. Taking $V(\boldsymbol{\theta}) \equiv I$ gives the OLS estimator, and taking $V(\boldsymbol{\theta}) \equiv \text{diag}\{\text{Var}[\hat{\gamma}(h(1))], \ldots, \text{Var}[\hat{\gamma}(h(K))]\}$ gives a nonlinear weighted least squares (WLS) estimator.

Determining the elements of $\boldsymbol{V}(\boldsymbol{\theta})$ requires knowledge of the fourth-order moments of $\mathbf{Z}$. Although these have been derived (see Cressie 1985), they are tedious to compute. Cressie (1985) showed that a nonlinear WLS estimator based on

$$\mathrm{var}[\widehat{\gamma}(h(j))] \approx 2[\gamma(h(j);\boldsymbol{\theta})]^2/N(h(j)) \tag{8.9}$$

yields an estimation procedure that often works well in practice. Thus, weighting the OLS objective function inversely proportional to the (approximate) variance of the empirical semivariogram estimator gives an estimator of $\boldsymbol{\theta}$ that minimizes the weighted regression sum of squares:

$$\mathrm{WRSS}(\boldsymbol{\theta}) = \frac{1}{2}\sum_{j=1}^{K} \frac{N(h(j))}{[\gamma(h(j);\boldsymbol{\theta})]^2}\left[\widehat{\gamma}(h(j)) - \gamma(h(j);\boldsymbol{\theta})\right]^2. \tag{8.10}$$

This approach is an approximation to WLS and offers a pragmatic compromise between OLS and GLS. It gives more weight where there is more "data" [large $N(h(j))$] and near the origin [small $\gamma(\boldsymbol{h};\boldsymbol{\theta})$], thus improving on OLS. Although it will not be as efficient as GLS, the ease of computation is a definite advantage. It can be used even when the data are not Gaussian, and empirical studies have shown (e.g., Zimmerman and Zimmerman 1991) this approach to be fairly accurate in a variety of practical situations.

An approximation to the covariance matrix of $\widehat{\boldsymbol{\theta}}$ can also be obtained from the regression, based on the matrix of partial derivatives of $\gamma(\boldsymbol{h};\boldsymbol{\theta})$ with respect to each parameter in $\boldsymbol{\theta}$. Approximate confidence limits are then $\widehat{\boldsymbol{\theta}} \pm t_{K-g,1-\alpha/2}\mathrm{s.e.}(\widehat{\boldsymbol{\theta}})$, where $\mathrm{s.e.}(\widehat{\boldsymbol{\theta}})$ is the standard error of $\widehat{\boldsymbol{\theta}}$, $t_{K-g,1-\alpha/2}$ is the $1-\alpha/2$ percentage point of a t-distribution with $K-g$ degrees of freedom, and $g = \dim(\boldsymbol{\theta})$. Textbooks on nonlinear regression (e.g., Seber and Wild 1989) provide the theoretical and computational details of nonlinear regression.

Maximum Likelihood Estimation If the data, $\mathbf{Z}$, follow a multivariate Gaussian distribution with mean $\mathbf{1}\mu$ (here $\mathbf{1}$ is a vector of 1's) and variance–covariance matrix $\Sigma(\boldsymbol{\theta})$, likelihood-based techniques can be used to estimate $\boldsymbol{\theta}$. Maximizing the multivariate Gaussian likelihood with respect to $\boldsymbol{\theta}$ yields the maximum likelihood estimator (MLE) of $\boldsymbol{\theta}$. With restricted maximum likelihood (REML), the likelihood derived from error contrasts (or differences) of the data is maximized. Since maximizing the likelihood is equivalent to minimizing twice the negative log likelihood, in practice, the following objective functions are minimized:

$$\begin{aligned}\text{ML:}\ \ l(\boldsymbol{\theta}) &= \log(|\Sigma(\boldsymbol{\theta})|) + (\mathbf{Z}-\mathbf{1}\mu)'\Sigma(\boldsymbol{\theta})^{-1}(\mathbf{Z}-\mathbf{1}\mu) + N\log(2\pi)\\ \text{REML:}\ \ l_R(\boldsymbol{\theta}) &= (N-1)\log(2\pi) + \log(|\Sigma(\boldsymbol{\theta})|) + \log(|\mathbf{1}'\Sigma(\boldsymbol{\theta})^{-1}\mathbf{1}|)\\ &\quad + \mathbf{Z}'\{\Sigma(\boldsymbol{\theta})^{-1} - \Sigma(\boldsymbol{\theta})^{-1}\mathbf{1}(\mathbf{1}'\Sigma(\boldsymbol{\theta})^{-1}\mathbf{1})^{-1}\mathbf{1}'\Sigma(\boldsymbol{\theta})^{-1}\}\mathbf{Z}.\end{aligned}$$

An approximate covariance matrix of $\widehat{\boldsymbol{\theta}}$ can be obtained as the inverse of the Fisher information matrix. Approximate confidence limits are then $\widehat{\boldsymbol{\theta}} \pm z_{1-\alpha/2}\mathrm{s.e.}(\widehat{\boldsymbol{\theta}})$, where $\mathrm{s.e.}(\widehat{\boldsymbol{\theta}})$ is the standard error of $\widehat{\boldsymbol{\theta}}$, and $z_{1-\alpha/2}$ is the $1-\alpha/2$ percentage point of a standard normal distribution. Searle et al. (1992) is a good reference text for the computational details and distributional properties of likelihood-based estimators.

Empirical Model Comparison A major advantage of statistical approaches to fitting semivariogram models (as opposed to fitting them "by eye" as is often done in many disciplines) is the availability of objective criteria for comparing the fits of two or more competing models. Although the details depend on whether nonlinear least squares (NLS) or likelihood-based (LB) approaches are used to estimate the model parameters, such criteria take three broad forms.

1. *Single-parameter tests.* These are *Wald tests* for each parameter computed as $\widehat{\theta_i}/\text{s.e.}(\widehat{\theta_i})$. For NLS, this test statistic is compared to a t-distribution on $K - g$ degrees of freedom, and for LB tests, it is compared to a standard normal distribution. Such tests are similar to the familiar significance tests for individual parameters in a linear regression.
2. *Full and reduced model tests.* The best-fitting model should have the smallest value of the objective function [either $\text{WRSS}(\boldsymbol{\theta})$, $l(\boldsymbol{\theta})$, or $l_R(\boldsymbol{\theta})$]. The question is whether or not the difference in these criteria between two models is "significant." When the two candidate models are nested (i.e., one is a special case of the other obtained by putting one or more of the parameters equal to zero), formal tests can be made.

   Consider comparing two models of the same form, one based on parameters $\boldsymbol{\theta}_1$ and a larger model based on $\boldsymbol{\theta}_2$, with $\dim(\boldsymbol{\theta}_2) > \dim(\boldsymbol{\theta}_1)$ (i.e., $\boldsymbol{\theta}_1$ is obtained by setting some parameters in $\boldsymbol{\theta}_2$ equal to zero). With NLS, the test statistic for testing $H_0 : \gamma(h; \boldsymbol{\theta}) = \gamma(h; \boldsymbol{\theta}_1)$ vs. $H_1 : \gamma(h; \boldsymbol{\theta}) = \gamma(h; \boldsymbol{\theta}_2)$ is (Webster and McBratney 1989)

$$\frac{\text{WRSS}(\boldsymbol{\theta}_1) - \text{WRSS}(\boldsymbol{\theta}_2)}{\dim(\boldsymbol{\theta}_2) - \dim(\boldsymbol{\theta}_1)} \Bigg/ \frac{\text{WRSS}(\boldsymbol{\theta}_2)}{K - \dim(\boldsymbol{\theta}_2)}, \tag{8.11}$$

   and the test is made by comparing this statistic to an F distribution with $(\dim(\boldsymbol{\theta}_2) - \dim(\boldsymbol{\theta}_1), K - \dim(\boldsymbol{\theta}_2))$ degrees of freedom. With REML, the comparable test is based on comparing

$$l_R(\boldsymbol{\theta}_2) - l_R(\boldsymbol{\theta}) \tag{8.12}$$

   to a χ^2 with $\dim(\boldsymbol{\theta}_2) - \dim(\boldsymbol{\theta}_1)$ degrees of freedom. [An analogous test can be done for ML using $l(\boldsymbol{\theta})$.] An important exception occurs when the parameters to be tested lie on the boundary of the parameter space. In such cases, the test statistic in equation (8.12) is a mixture of χ^2 distributions. Such boundary exceptions arise in practice when testing whether variance components (e.g., nugget and sill) are equal to zero. If we are testing only one of these variance components against 0, the test statistic has a distribution that is a 50:50 mixture of a degenerate χ^2 that places all mass at {0} and a

χ^2 with $\dim(\boldsymbol{\theta}_2) - \dim(\boldsymbol{\theta}_1)$ degrees of freedom (Self and Liang 1987; Littell et al. 1996). Thus, to make the test, simply divide by 2 the p-value obtained from a χ^2 with $\dim(\boldsymbol{\theta}_2) - \dim(\boldsymbol{\theta}_1)$ degrees of freedom.

3. *Penalized objective functions.* The tests described above should be used with caution. Wald tests can be unreliable with small samples and for variance components that have a skewed distribution. The full and reduced F-test is based on assumptions of independence and normality, assumptions not met by the empirical semivariogram. Moreover, when the models are not nested (e.g., we want to compare the fit of a spherical to that of an exponential), full and reduced tests are not applicable, even for likelihood-based methods. Thus, rather than relying on a statistical test, we should simply choose the model that has the smallest value for the objective function. Since the value of this function is reduced by increasing the number of parameters in the model, other criteria that penalize the objective functions for additional parameters have been developed. Akaike's information criterion (AIC) (Akaike 1974) is perhaps the most commonly used of these. This criterion was originally developed for likelihood methods and then adapted to regression fitting of semivariogram models by Webster and McBratney (1989):

$$\text{AIC}(\boldsymbol{\theta}) = K \log\left(\frac{\text{WRSS}(\boldsymbol{\theta})}{K}\right) + 2p \qquad \text{(NLS)} \qquad (8.13)$$

$$\text{AIC}(\boldsymbol{\theta}) = l_R(\boldsymbol{\theta}) + 2p, \qquad \text{(REML)} \qquad (8.14)$$

where p is the number of parameters in the model. We should prefer the model with the smallest AIC value.

Practical Notes on Model Fitting There are many different ways to parameterize each semivariogram model (particularly the exponential, Gaussian, and K-Bessel models), and different computer software programs may use slightly different forms. Also, some programs simply ask for or return "the range" and "the sill," which may or may not be one of the parameters of the model. It is a good idea to check the software manual or online help for the equation of each model to be sure of the exact form of the models you are using.

Good starting values are very important in iterative fitting methods. We recommend using a parameter search to determine good starting values, even if you feel your initial values are very close. If convergence is still a problem, using a different optimization approach (e.g., use of the Marquardt method in NLS or Fisher scoring in LB methods) may be helpful (Seber and Wild 1989). NLS can be affected adversely by a noisy semivariogram, so you may need to try a different lag spacing to define the structure of the semivariogram more clearly. LB methods are based on a covariance function, not on a semivariogram, and they do not calculate an empirical covariance function analogous to equation (8.8). They also use all the data to estimate the model parameters, whereas the empirical semivariogram is often estimated only for lags less than half the maximum separation distance. Thus, judging the model fits from LB approaches by superimposing the fitted model on the empirical semivariogram can be misleading.

Finally, remember that although the semivariogram is estimated from the data available, it is describing the variability *of a spatial process*. So even though a particular model is deemed best for a particular data set by a statistical comparison, it may not be the best choice. For example, the Gaussian model is often selected as best with automatic fitting criterion, but it also corresponds to a process that is often unrealistically smooth. Ultimately, the final choice of model should reflect both the results of the statistical model fitting procedure and an interpretation consistent with the scientific understanding of the process being studied.

DATA BREAK: Smoky Mountain pH Data (*cont.*) To continue our Smoky Mountain pH data break, consider fitting a theoretical semivariogram model to the empirical semivariogram shown in Figure 8.6. The semivariogram appears to reach a definite sill, but we do not have a lot of information about the shape of the semivariogram near the origin or the value of the nugget effect. In the absence of such information, a model that is approximately linear near the origin is a good choice since it does not assume that the process is too smooth. Thus either an exponential model or a spherical model might be a good choice. The

Table 8.1 Weighted Nonlinear Least Squares Fit of the Exponential Model to the Empirical Semivariogram in Figure 8.6[a]

Parameter	Estimate	Approximate Standard Error	Approximate 95% Confidence Limits
a_e	34.432	12.352	(6.909, 61.955)
c_e	0.191	0.025	(0.137, 0.246)
c_0	0.057	0.025	(0.002, 0.111)

$\text{WRSS}(\widehat{\boldsymbol{\theta}}) = 26.81$ with 10 degrees of freedom

[a] WRSS is the weighted residual sum of squares as discussed in the text.

Table 8.2 Weighted Nonlinear Least Squares Fit of the Spherical Model to the Empirical Semivariogram in Figure 8.6[a]

Parameter	Estimate	Approximate Standard Error	Approximate 95% Confidence Limits
a_s	110.312	31.939	(39.138, 181.521)
c_s	0.185	0.026	(0.126, 0.243)
c_0	0.084	0.016	(0.048, 0.121)

$\text{WRSS}(\widehat{\boldsymbol{\theta}}) = 29.28$ with 10 degrees of freedom

[a] WRSS is the weighted residual sum of squares as discussed in the text.

values of pH may have some measurement error associated with them, so we initially include a nugget effect in both models. The NLS fit statistics for the exponential model are shown in Table 8.1, and those for the spherical model are given in Table 8.2. (These results depend on the options used in the NLS fitting. We compared the results from three different software packages using their default options and obtained different results from each. Hence your results may differ from those presented here, but they should be close.) The fitted models are superimposed on the empirical semivariogram in Figures 8.7 and 8.8.

Since the models contain the same number of parameters, comparing AIC criteria computed from equations (8.13) is the same as comparing the weighted residual sums of squares [shown as $\text{WRSS}(\widehat{\boldsymbol{\theta}})$ in each table]. Since the fit to the exponential model has the smallest WRSS, this model fits the empirical semivariogram better than the spherical model.

We have several criteria to help us decide whether or not a nugget effect is needed. First, the Wald confidence limits do not contain zero, indicating that the nugget effect is significantly different from 0. Refitting the model without a nugget effect gives WRSS = 36.89, much higher than WRSS = 26.81 for the full model. The full and reduced F statistic is then $(36.89 - 26.81)/1/(26.81/10) = 3.76$ on (1,10) degrees of freedom, giving a corresponding p-value for the test of 0.08. Although this is not technically significant at 0.05, it is significant at 0.10 and provides some evidence against our null hypothesis of a zero nugget effect. Finally, a comparison of the AIC criteria gives AIC = 15.41 for the full model and AIC = 17.56 for the reduced model. Taken together, these statistics lead us to conclude that including a nugget effect probably results in a better-fitting model.

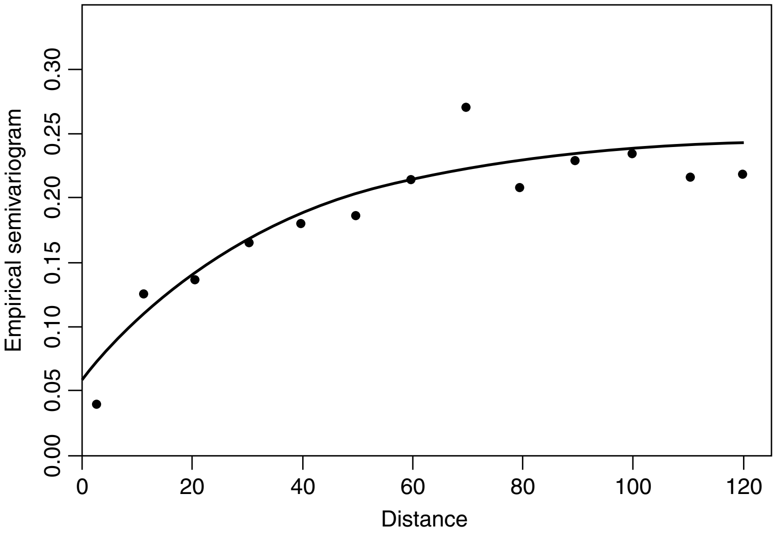

FIG. 8.7 Exponential model fit to the empirical semivariogram of pH values.

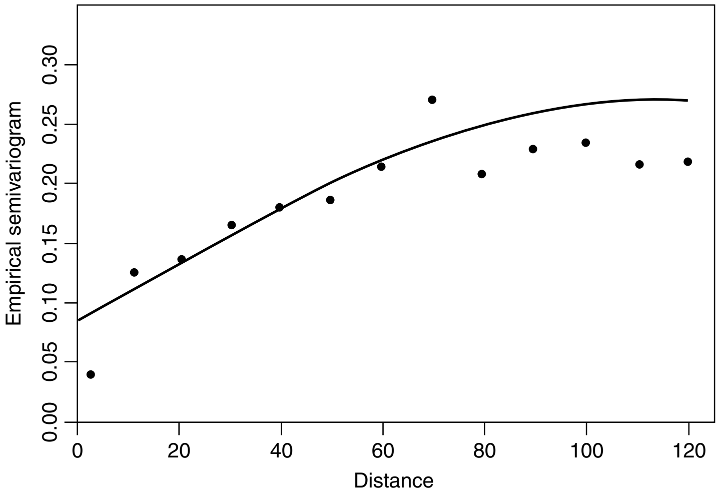

FIG. 8.8 Spherical model fit to empirical semivariogram of pH values.

8.2.5 Anisotropic Semivariogram Modeling

The nature of the semivariogram may change with direction. Then $\gamma(\boldsymbol{h})$ will be a function of both the magnitude and direction of the lag vector $\boldsymbol{h}$. Such spatial processes are referred to as *anisotropic*. Anisotropies result from the differential behavior of a physical process. For example, the pH of a stream may vary with the direction of stream flow. Although anisotropic processes are probably more common than isotropic ones, they receive less attention in the literature because they are more difficult mathematically and require more data for inference.

Since we must refer to direction as well as distance, we need a way to describe directions succinctly. The term *northwest* is too vague, and notation such as "N30°E" can be too cumbersome. In this book, we report directions as an angle, ϕ, measured in degrees counterclockwise from east. This allows the use of standard geometrical definitions.

Geometric Anisotropy Geometric anisotropy is a particular type of anisotropy characterized by two properties: (1) the directional semivariograms have the same shape and sill but different ranges; and (2) the semivariogram in direction ϕ has the maximum range of any direction, $a_{\max}$, and perpendicular to this direction, the semivariogram in direction $\phi \pm 90^\circ$ has the minimum range of any direction, $a_{\min}$, and all the ranges delineate an ellipse with major and minor radii equal to $a_{\max}$ and $a_{\min}$, respectively. This is depicted graphically in Figure 8.9. To work with anisotropy we need to construct anisotropic semivariogram models that are conditionally negative definite. To do this, we adapt the isotropic semivariogram models described in Section 8.2.2 using elliptical geometry. Eriksson and Siska (2000) provide an excellent discussion of anisotropy, and our presentation draws greatly from their work.

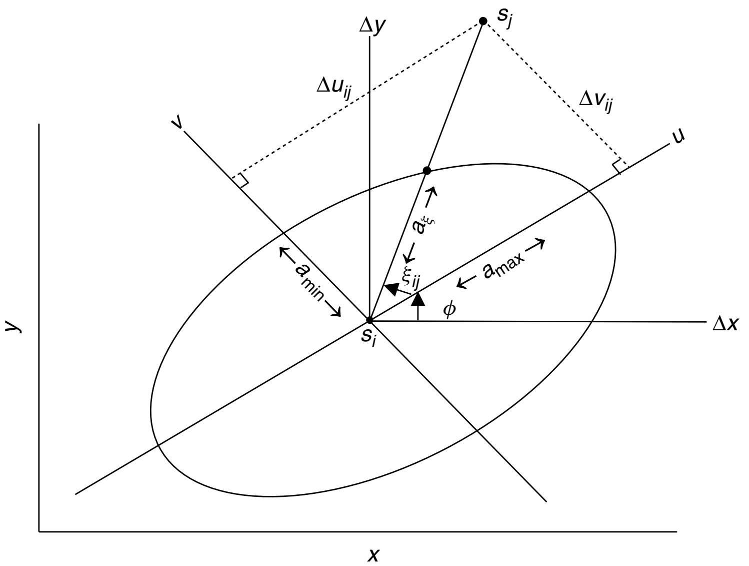

FIG. 8.9 Geometric anisotropy. The maximum range is in the direction ϕ. The points s_i and s_j are separated by a distance h_{ij} in the direction $\eta_{ij} = \xi_{ij} + \phi$. [Adapted from Eriksson and Siska (2000).]

Let u and v be the axes defined by the major and minor axes of the ellipse and convert these to rectangular coordinates:

$$u = a_\xi \cos\xi, \qquad v = a_\xi \sin\xi, \qquad \text{with } a_\xi = \sqrt{u^2 + v^2}.$$

Here ξ is the angle between the vector $\overrightarrow{s_i s_j}$ and the u-axis relative to the (u, v) coordinate system and a_ξ is the range (length) in the direction ξ. Then, substituting into the equation of an ellipse, $u^2/a_{\text{max}}^2 + v^2/a_{\text{min}}^2 = 1$, gives

$$\frac{a_\xi^2 \cos^2\xi}{a_{\text{max}}^2} + \frac{a_\xi^2 \sin^2\xi}{a_{\text{min}}^2} = 1.$$

So, relative to the u, v coordinate system, the range in direction ξ is

$$a_\xi = a_{\text{max}} a_{\text{min}} \Bigg/ \sqrt{a_{\text{min}}^2 \cos^2\xi + a_{\text{max}}^2 \sin^2\xi}\ . \tag{8.15}$$

With respect to the original x, y coordinate system, the vector $\overrightarrow{s_i s_j}$ makes an angle $\eta_{ij} = \xi_{ij} + \phi$ which, with respect to the u, v system, is the same as the angle $\xi_{ij} = \eta_{ij} - \phi$. Thus, the range in direction η_{ij} is

$$a_{\eta_{ij}} = a_{\text{max}} a_{\text{min}} \Bigg/ \sqrt{a_{\text{min}}^2 \cos^2(\eta_{ij} - \phi) + a_{\text{max}}^2 \sin^2(\eta_{ij} - \phi)}\ .$$

This now plays the role of the usual range parameter in an isotropic semivariogram model. For example, using a spherical model with unit sill and zero nugget effect, a geometric anisotropic model can be written as

$$\gamma(h, \eta; \boldsymbol{\theta}) = \begin{cases} 0 \quad \text{for} \;\; h = 0 \\ \dfrac{3}{2}\left(\dfrac{h\sqrt{a_{\min}^2 \cos^2(\eta - \phi) + a_{\max}^2 \sin^2(\eta - \phi)}}{a_{\max}a_{\min}}\right) \\ \quad -\dfrac{1}{2}\left(\dfrac{h\sqrt{a_{\min}^2 \cos^2(\eta - \phi) + a_{\max}^2 \sin^2(\eta - \phi)}}{a_{\max}a_{\min}}\right)^3 \\ \quad \text{for} \;\; 0 \le h \le a_{\max}a_{\min}\Big/\sqrt{a_{\min}^2 \cos^2(\eta - \phi) + a_{\max}^2 \sin^2(\eta - \phi)} \\ 1 \quad \text{otherwise} \end{cases} \tag{8.16}$$

where $\boldsymbol{\theta} = (a_{\min}, a_{\max}, \phi)'$.

Unfortunately, few books or papers ever write anisotropic models in this form. Instead they usually describe a method of rotation and shrinkage and a “reduced distance” notation. The idea here is first to rotate the coordinate axes so they are aligned with the major and minor axes of the ellipse. Then shrink the axes so that the ellipse is now a circle with radius 1. This can be done using

$$\begin{aligned} \boldsymbol{h}' &= \begin{bmatrix} 1/a_{\max} & 0 \\ 0 & 1/a_{\min} \end{bmatrix} \begin{bmatrix} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{bmatrix} \begin{bmatrix} \Delta x_{ij} \\ \Delta y_{ij} \end{bmatrix} \\ &\equiv \begin{bmatrix} 1/a_{\max} & 0 \\ 0 & 1/a_{\min} \end{bmatrix} \begin{bmatrix} \Delta u_{ij} \\ \Delta v_{ij} \end{bmatrix}, \end{aligned} \tag{8.17}$$

where Δu_{ij} and Δv_{ij} are distance components of $\boldsymbol{s}_i - \boldsymbol{s}_j$ in the u and v directions relative to the u and v axes, and Δx_{ij} and Δy_{ij} are the same distance components in the x and y directions relative to the x and y axes. After transformation, the ellipse is now a circle with radius 1, so isotropic semivariogram models can be used with the transformed distances. Continuing with our example of a spherical model without a nugget effect, the value of the anisotropic model in various directions is equal to the value of the isotropic semivariogram with range 1 using

$$||\boldsymbol{h}'|| = \sqrt{\frac{\Delta u_{ij}^2}{a_{\max}^2} + \frac{\Delta v_{ij}^2}{a_{\min}^2}}. \tag{8.18}$$

The quantity $||\boldsymbol{h}'||$ given in equation (8.18) is called the *reduced distance*. Then, the values of the anisotropic semivariogram model given in equation (8.16) are equal to

$$\mathrm{Sph}_1(||\boldsymbol{h}'||) = \begin{cases} 0, & ||\boldsymbol{h}'|| = 0 \\ (3/2)||\boldsymbol{h}'|| - (1/2)||\boldsymbol{h}'||^3, & 0 \le ||\boldsymbol{h}'|| \le 1 \\ 1, & \text{otherwise,} \end{cases} \tag{8.19}$$

where $\mathrm{Sph}_a(\cdot)$ denotes an isotropic spherical semivariogram model with zero nugget effect, unit sill, and range a (Isaaks and Srivastava 1989). In this formulation the dependence on direction is not obvious but is implicit in the reduced distance (each pair of points may have different distance components). A geometric argument for the use of the reduced distance can be found in Eriksson and Siska (2000).

Zonal Anisotropy The term *zonal anisotropy* refers to the case where the sill changes with direction but the range remains constant. This type of anisotropy is common with three-dimensional spatial processes, where the vertical direction (depth, height, or time) behaves differently from the two horizontal directions. To model zonal anisotropy, assume that in the direction ϕ, the sill is $c_{\max}$, the sill in the direction perpendicular to ϕ is $c_{\min}$, and denote the constant range by a. This type of anisotropy is usually modeled using the sum of two isotropic semivariogram models, referred to as *structures*. (Recall from Section 8.2 that the sum of two valid semivariograms is also a valid semivariogram.) The first structure is an isotropic model with sill $c_{\min}$ and range a. The second structure is a contrived geometric model with sill $c_{\max} - c_{\min}$. The range in the direction of maximum sill is taken to be the common range a and the range in the direction of minimum sill is set to a very large value so that this structure does not contribute to the overall model in the direction of minimum sill. For example, suppose that the maximum sill is 9 in the direction ϕ and that the smallest sill of 5 is observed in the perpendicular direction. Assume that a spherical model can be used to fit both directions and that the constant range is 100. Then the values for the zonal model can be computed from

$$5 \cdot \mathrm{Sph}_1(h/100) + 4 \cdot \mathrm{Sph}_1(||\boldsymbol{h}'||) \quad \text{with} \quad ||\boldsymbol{h}'|| = \sqrt{\frac{\Delta u_{ij}^2}{100} + \frac{\Delta v_{ij}^2}{100{,}000}}.$$

In the direction of minimum sill, the first component of the reduced distance used in the second structure is zero and the contribution of the second term is negligible because of the large range in this direction. Hence we obtain a spherical model with range 100 and a sill of 5. In the direction of maximum sill, both structures contribute to the computations, but the second component of the reduced distance used in the second structure is negligible. Thus, the values can be obtained using $5 \cdot \mathrm{Sph}_1(h/100) + 4 \cdot \mathrm{Sph}(h/100) = 9 \cdot \mathrm{Sph}(h/100)$.

Detecting Anisotropy There are two primary tools used to explore anisotropy. The first is a contour or image map of the empirical semivariogram surface. This map is constructed by partitioning the domain into cells of length Δx in the x direction and Δy in the y direction (so the "pixels" are rectangular tolerance regions). To calculate the empirical semivariogram for locations separated by $\mathbf{h} = (h_x, h_y)$, we average the pairs separated by $h_x \pm \Delta x$ in the x direction and by $h_y \pm \Delta y$ in the y direction. Then we draw a contour or image map depicting the empirical semivariogram as a function of the x and y distances. The center of the map corresponds to (0,0), with distances increasing in each direction from this central point. If the process is isotropic, no strong directional trends will be evident in the map and we should see circular contours. Elliptical contours indicate anisotropy and the direction of maximum range or sill will be indicated by a trough of low values connected in a particular direction. This type of map is a good visual tool for detecting anisotropy and suggesting plausible anisotropic models.

Contour plots can be difficult to use for modeling where we need more precise values for the nugget effect, ranges, and the sills. Directional empirical semivariograms are used to determine these. Directions of interest may be indicated by the contour map of the semivariogram surface, determined from scientific information, or investigated systematically going from 0° in increments of $d°$. Once a suite of directional semivariograms is estimated from the data, the first place to start is to see if the process exhibits geometric anisotropy. Fix a common nugget effect and sill and then record the range in each direction. It may be useful to plot the length of each range as a line in the direction of interest; this plot is called *a rose diagram* (Isaaks and Srivastava 1989). If geometric anisotropy is present, the diagram will resemble an ellipse and the quantities for constructing a model of geometric anisotropy can be read easily from this diagram. The same may be done to investigate zonal anisotropy by fixing the range and the nugget effect and plotting the sills.

When investigating anisotropy, patience, creativity, and flexibility are key. The ellipses are based on semivariograms estimated from the data, and perfectly defined ellipses are rare. Investigations of anisotropy split the data set, so directional semivariograms based on a much reduced sample size are consequently noisier. In many applications, there will not be enough data to investigate anisotropy adequately, and some compromises will have to made (e.g., consider only a few directions, use large angle tolerances, assume isotropy). The semivariogram surface could show several directions of anisotropy, and the directional semivariograms may indicate both zonal and geometric anisotropy. An irregular domain can make an isotropic process appear anisotropic. One of the advantages of interactive semivariogram modeling, such as that done by ESRI's Geostatistical Analyst (Johnston et al. 2001), is that it can show hypothetical fits of an anisotropic model to several directions simultaneously. When the anisotropy is complicated, it may be easier to consider a trend process model (see the data break following Section 9.1.2) since systematic large-scale trends in the data can manifest themselves as directional spatial autocorrelation. The difference between trend and anisotropy is subtle, and it can simply be one of interpretation and preference (Zimmerman 1993; Gotway and

Hergert 1997). Remember that all models are abstract approximations of reality; the goal is not *the* model, but *a good* model that is defensible, parsimonious, and reflects understanding of the spatial processes under study.

DATA BREAK: Smoky Mountain pH Data (*cont.*) In this portion of our data break, we investigate anisotropy in the pH data. First, consider the semivariogram surface displayed as an image map in Figure 8.10. This map was based on a lag spacing of 12 miles (with a tolerance of ±6 miles) in both the x and y directions. We initially chose a lag spacing of 10 miles based on the omnidirectional semivariogram (Figure 8.6), but this resulted in too few pairs for estimation in certain directions. After experimenting with several different choices for the lag spacing, including choices that allowed different spacings in the x and y directions, we decided that this lag spacing gave the best balance between reliable estimation and map detail. From Figure 8.10 we can see a trough of low values oriented at about 70°, indicating strong spatial continuity in this direction (i.e., low variance of pairwise differences for observations in the "trough").

To explore this pattern further, we estimated four directional semivariograms, starting with direction 70° and rotating counterclockwise 45°, using an angle tolerance of ±22.5°. We considered only four directions since, with only 75 data points, restricting semivariogram estimation to data within small sectors may result in poor estimates. Figure 8.11 indicates anisotropy, with the 70° direction showing the largest range and smallest sill and the 160° direction showing the smallest range and highest sill. Thus, there is evidence of anisotropy, but the type of anisotropy is not immediately apparent. Since the sill and range depend on each other, fixing one of these parameters allows us to vary the other. Since geometric anisotropy is the simplest type of anisotropy, we start by fixing a common nugget effect and sill

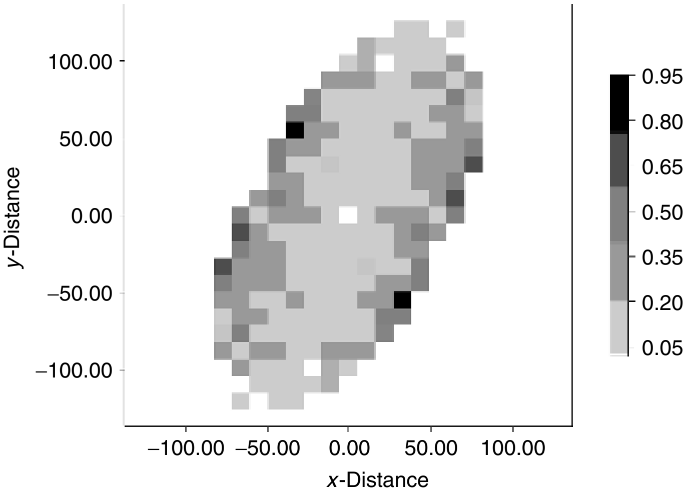

FIG. 8.10 Empirical semivariogram contour map for Smoky Mountain pH data.

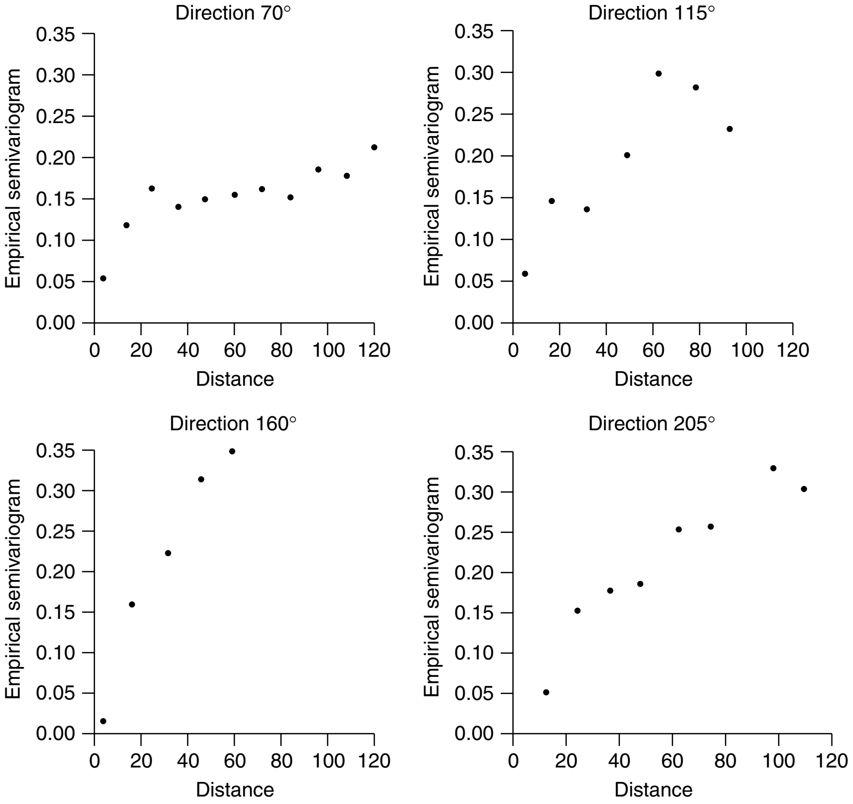

FIG. 8.11 Empirical directional semivariograms for Smoky Mountain pH data.

and then vary the ranges to see if a geometrically anisotropic model can adequately fit these empirical semivariograms. We initially chose a common nugget effect of 0 since the 160° direction does not show a nugget effect and WLS estimation will be easier without this extra parameter. Then, fixing $\phi = 70°$ and $c_0 = 0$, we used the estimates from all four directions to fit an exponential model with range given in equation (8.15) using WLS. This gave $\hat{c}_e = 0.2725$, $\hat{a}_{\max} = 36.25$, and $\hat{a}_{\min} = 16.93$, and the resulting fits are shown in Figure 8.12. Considering that we have only 75 data points, the overall fit seems adequate, although the fit to the 70° direction could be improved by including a nugget effect, but only at the expense of the fit in the 160° direction. Although we did have evidence of anisotropy, the fit of the isotropic model from Table 8.1 to these directions seems almost equally adequate. The isotropic model fit seems better in the 70° and 115° directions, but worse in the 160° and 205° directions. However, the isotropic model is a much simpler model. To help decide if a geometrically anisotropic model is better, we can compare the AIC criteria from fitting this model to that obtained by refitting an isotropic model to the directional semivariograms. This gives AIC = 67.95 for the

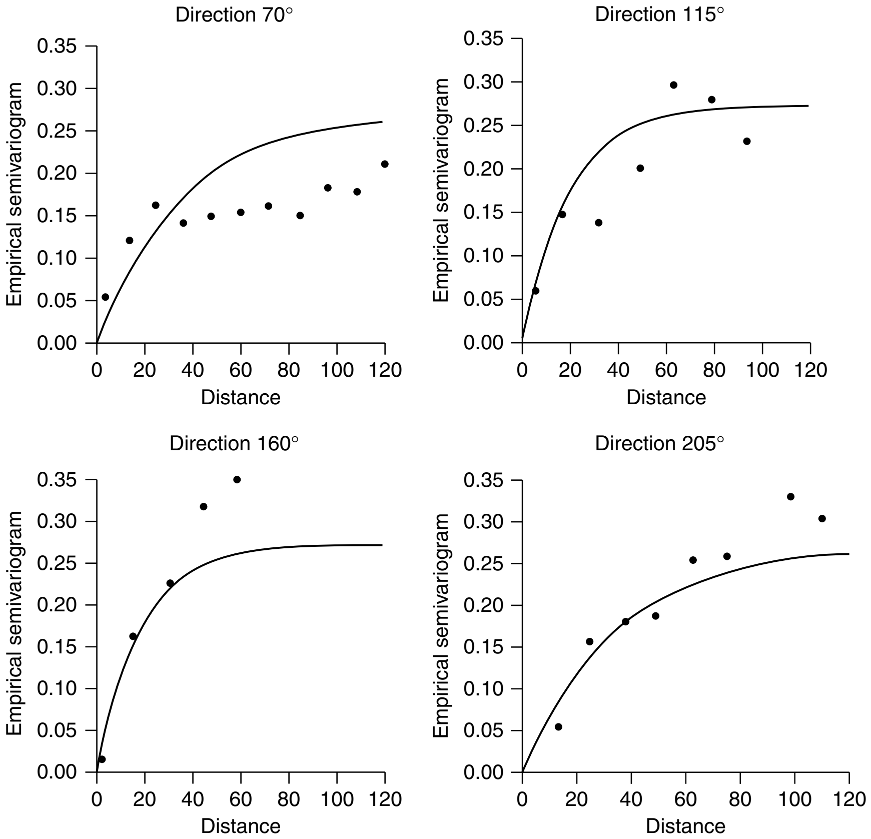

FIG. 8.12 Empirical directional semivariograms and fitted models for Smoky Mountain pH data. The solid curve is the fit of the geometric anisotropic model, and the dashed curve shows the fit of the isotropic model.

geometrically anisotropic model and AIC = 72.99 for the isotropic model. Even if we further penalize the geometrically anisotropic model for estimation of ϕ, it appears that this model gives a slightly better fit than the isotropic model. Knowledge of stream flow or other stream parameters affecting pH would also be helpful in deciding between the two models.

In this analysis we assumed that the pH process was anisotropic and estimated the direction of maximum continuity (range for geometric anisotropy and sill for zonal anisotropy) "by eye" using the empirical semivariogram surface in Figure 8.10. Instead of considering just four directions defined by nonoverlapping sectors, we could have estimated empirical semivariograms every 20° (or so) using an angle tolerance of 22.5° or even 45°, allowing the binning sectors to overlap. This would have allowed us to construct a rose diagram of the ranges in each direction and permitted a more refined estimate of the direction of maximum range, although the notion of "different directions" in the case of overlapping sectors and large angle tolerances is rather vague. More precise and objective (e.g., WLS) estimation of this direction requires many empirical directional semivariograms (i.e., measured every 10° to 20°) computed with a small angle tolerance. We usually do not have enough data to do this, and some compromise between many blurred directions and a few more precise ones must be made. This compromise must also balance precision and accuracy against the time and effort involved; obtaining good estimates of many directional semivariograms can be a tedious chore. One such compromise, based on smoothing of the semivariogram surface, is implemented with ESRI's Geostatistical Analyst (Johnston et al. 2001). A kernel smoother, similar to those described in Chapter 5, is used to smooth the semivariogram values in each cell of the semivariogram surface, borrowing strength from values in neighboring cells. Locations are weighted based on their distance from the center of each cell. A modified weighted least squares algorithm, implemented in stages, can then be used to fit an anisotropic model to the smoothed surface. Using this approach with the pH data, we obtained WLS estimates $\hat{\phi} = 69.6^\circ$, $\hat{c}_0 = 0.0325$, $\hat{c}_e = 0.2015$, $\hat{a}_{\max} = 44.9$, and $\hat{a}_{\min} = 16.65$. These values, computed from the empirical semivariogram surface, are subject to errors induced by the rectangular binning and the smoothing, but are obtained quickly and objectively (and the fit is not all that different from that obtained previously with much greater effort).

8.3 INTERPOLATION AND SPATIAL PREDICTION

In exposure assessment, we may want to predict exposure at a location where we have not recorded an observation, say at location s_0. *Interpolation* is the process of obtaining a value for a variable of interest [denoted here as $Z(s_0)$] at an unsampled location based on surrounding measurements. An example of the interpolation problem is given in Figure 8.13. Here five data values are recorded at locations s_1, s_2, s_3, s_4, and s_5, and we would like to predict the value at s_0 from these observations.

It is often useful to have a map of the spatial variation in exposure. *Gridding* refers to the systematic interpolation of many values identified by the nodes of a regular grid. These interpolated values are then displayed graphically, usually by means of a contour or surface map.

There are many methods for interpolating spatial data. These fall into two broad classes: deterministic and probabilistic. Deterministic methods have a mathematical development based on assumptions about the functional form of the interpolator. Probabilistic methods have a foundation in statistical theory and assume a statistical model for the data. When probabilistic methods are used for interpolation, they are referred to as methods for *spatial prediction*. These predictors have standard errors that quantify the uncertainty associated with the interpolated values. Deterministic interpolators do not have a measure of uncertainty associated with them. Sometimes, interpolation methods can be developed from both points of view (as with least squares regression, for example). We discuss two of the most commonly used methods: inverse distance interpolation (deterministic) and kriging (probabilistic).

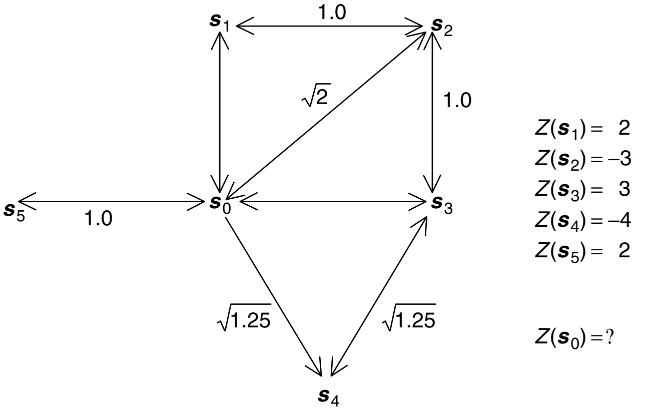

FIG. 8.13 The interpolation problem. The numbers next to each line are distances.

Cressie (1989) gives a comparative overview of many other interpolation methods not described here.

8.3.1 Inverse-Distance Interpolation

An inverse-distance interpolator is simply a weighted average of neighboring values. The weight given to each observation is a function of the distance between that observation's location and the grid point s_0 at which interpolation is desired. Mathematically, the general inverse-distance interpolator is written as

$$\hat{Z}_{\mathrm{ID}p} = \sum_{i=1}^{N} Z(s_i)\, d_{0,i}^{-p} \Bigg/ \sum_{i=1}^{N} d_{0,i}^{-p} . \tag{8.20}$$

Here $d_{0,i}$ is the distance from the grid point location s_0 to the ith data location s_i. The weighting power, p, is selected to control how fast the weights tend to zero as the distance from the grid node increases, based on assumed increasing similarity between observations taken closer together. As the power increases, the contribution (to the interpolated value) from data points far from the grid node decreases. Distance powers between 1 and 3 are typically chosen, and taking $p = 2$ gives the popular inverse-distance-squared interpolator. [See Burrough (1986) for a general discussion and illustration of these interpolators.]

As an example, consider the spatial layout in Figure 8.13. Intuitively, a good guess for the value at s_0 is 0, the average of these values. But in inverse-distance interpolation, the values at s_2 and s_4 will receive lower weight since they are farther away. Applying equation (8.20) with $p = 2$, we see that $\hat{Z}(s_0)_{\mathrm{ID}2} = 0.53$.

Most interpolation methods use only some of the "neighboring" data. The neighborhood is usually taken to be a circle centered on the grid node unless the study domain is elongated (e.g., elliptical) or the spatial process is anisotropic. If there are many points in the neighborhood, we may further restrict the calculations and use just the closest m points. (In many computer programs, defaults for m vary from 6 to 24.) These parameters (search neighborhood size, shape, and number of points used for interpolation), together with the choice of the weighting power, can affect the nature of the interpolated surface. Higher weighting powers and small search neighborhoods with few data points retained for interpolation produce very localized, choppy surfaces; lower powers of distance and large neighborhoods with many data points used for interpolation result in smoother surfaces. This idea is similar to the role of bandwidth in kernel density estimators of spatial intensity functions discussed in Section 5.2.5. Care must be taken to ensure that the interpolated values are based on enough data so that the averaging gives a fairly accurate value, particularly for grid nodes near the edges of the domain.

Inverse-distance interpolators are popular in many disciplines since they are relatively simple conceptually, require no modeling or parameter estimation, and are computationally fast. They can be exact interpolators (i.e., the interpolated surface passes through the original observations), or smoothers (i.e., predictions at observed locations are adjusted toward their neighborhood mean). However, mapped surfaces tend to have flat-topped peaks and flat-bottomed valleys giving a characteristic bull's-eye pattern produced by concentric contours around data points that can detract from the visual interpretation of the map. Also, since there is no underlying statistical model, there is no easy measure of the uncertainty associated with the value interpolated.

8.3.2 Kriging

Kriging is a geostatistical technique for optimal spatial *prediction*. We emphasize the distinction between *prediction,* which is inference on *random* quantities, and *estimation,* which is inference on *fixed but unknown* parameters. Georges Matheron, considered by many to be the founding father of geostatistics, introduced this term in one of his early works developing geostatistical theory (Matheron 1963). A Soviet scientist, L. S. Gandin, simultaneously developed the same theory in meteorology, where it is known by the name *objective analysis* (Gandin 1963). Since these beginnings, the original development of kriging has been extended in many ways. There are now many different types of kriging, differing by underlying assumptions and analytical goals. The following represents a partial list of the different types of kriging appearing in the literature, along with a brief definition:

- *Simple kriging:* linear prediction (i.e., predictor is a linear combination of observed data values) assuming a known mean
- *Ordinary kriging:* linear prediction with an constant unknown mean
- *Universal kriging:* linear prediction with a nonstationary mean structure
- *Filtered kriging:* smoothing and prediction for noisy data; also known as kriging with measurement error
- *Lognormal kriging:* optimal spatial prediction based on the lognormal distribution

- *Trans-Gaussian kriging:* spatial prediction based on transformations of the data
- *Cokriging:* multivariate linear prediction (i.e., linear prediction based on one or more interrelated spatial processes)
- *Indicator kriging:* probability mapping using indicator functions (binary data)
- *Probability kriging:* probability mapping based on both indicator functions of the data and the original data
- *Disjunctive kriging:* nonlinear prediction based on univariate functions of the data
- *Bayesian kriging:* incorporates prior information about the mean and/or covariance functions into spatial prediction
- *Block kriging:* optimal linear prediction of areal data from point data

This list is not exhaustive, and many combinations of these are possible (e.g., universal cokriging is multivariate spatial prediction with a nonstationary mean structure).

A comprehensive discussion of all of these methods is beyond the scope of this book. Instead, our discussion focuses on a few of these methods, written to balance completeness, theoretical development, and practical implementation. Our choice was indeed a struggle. Our readers are encouraged to consult other books (e.g., Journel and Huijbregts 1978; Isaaks and Srivastava 1989; Cressie 1993; Rivoirard 1994; Wackernagel 1995; Chilès and Delfiner 1999; Olea 1999; Stein 1999) to learn more about the methods and theoretical considerations we cannot discuss here.

Ordinary Kriging Assume that $Z(\cdot)$ is intrinsically stationary [i.e., has a constant unknown mean, μ, and known semivariogram, $\gamma(\boldsymbol{h})$]. Assume that we have data $\mathbf{Z} = [Z(\boldsymbol{s}_1), \ldots, Z(\boldsymbol{s}_N)]'$ and want to predict the value of the $Z(\cdot)$ process at an unobserved location, $Z(\boldsymbol{s}_0)$, $\boldsymbol{s}_0 \in D$. As with the inverse distance methods described in Section 8.3.1, the ordinary kriging (OK) predictor is a weighted average of the data:

$$\hat{Z}_{\text{OK}}(\boldsymbol{s}_0) = \sum_{i=1}^{N} \lambda_i Z(\boldsymbol{s}_i). \tag{8.21}$$

However, instead of specifying an arbitrary function of distance, we determine the weights based on the data using the semivariogram and two statistical optimality criteria: unbiasedness and minimum mean-squared prediction error. For unbiasedness, the predicted value should, on average, coincide with the value of the unknown random variable, $Z(\boldsymbol{s}_0)$. In statistical terms, unbiasedness requires $E[\hat{Z}_{\text{OK}}(\boldsymbol{s}_0)] = \mu = E[Z(\boldsymbol{s}_0)]$, which means that $\sum_{i=1}^{N} \lambda_i = 1$. To ensure the second optimality criterion, we need to minimize *mean-squared prediction error* (MSPE), defined as $E[\hat{Z}_{\text{OK}}(\boldsymbol{s}_0) - Z(\boldsymbol{s}_0)]^2$, subject to the unbiasedness constraint. One method for solving constrained optimization problems is the method of *Lagrange multipliers*.

With this method, we need to find $\lambda_1, \ldots, \lambda_N$ and a Lagrange multiplier, m, that minimize the objective function

$$E\left[\left(\sum_{i=1}^{N} \lambda_i Z(\boldsymbol{s}_i) - Z(\boldsymbol{s}_0)\right)^2\right] - 2m\left(\sum_{i=1}^{N} \lambda_i - 1\right). \tag{8.22}$$

The second term is essentially a penalty, minimized when $\sum_{i=1}^{N} \lambda_i = 1$, thus ensuring that our overall minimization incorporates our unbiasedness constraint. Now $\sum_{i=1}^{N} \lambda_i = 1$ implies that

$$\left[\sum_{i=1}^{N} \lambda_i Z(\boldsymbol{s}_i) - Z(\boldsymbol{s}_0)\right]^2 = -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \lambda_i \lambda_j \left[Z(\boldsymbol{s}_i) - Z(\boldsymbol{s}_j)\right]^2 + \sum_{i=1}^{N} \lambda_i \left[Z(\boldsymbol{s}_0) - Z(\boldsymbol{s}_i)\right]^2.$$

Taking expectations of both sides of this equation gives

$$E\left[\left(\sum_{i=1}^{N} \lambda_i Z(\boldsymbol{s}_i) - Z(\boldsymbol{s}_0)\right)^2\right] = -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \lambda_i \lambda_j E\left[\left(Z(\boldsymbol{s}_i) - Z(\boldsymbol{s}_j)\right)^2\right] + \sum_{i=1}^{N} \lambda_i E\left[\left(Z(\boldsymbol{s}_0) - Z(\boldsymbol{s}_i)\right)^2\right],$$

so that the objective function given in equation (8.22) becomes

$$-\sum_{i=1}^{N}\sum_{j=1}^{N} \lambda_i \lambda_j \gamma(\boldsymbol{s}_i - \boldsymbol{s}_j) + 2\sum_{i=1}^{N} \lambda_i \gamma(\boldsymbol{s}_0 - \boldsymbol{s}_i) - 2m\left(\sum_{i=1}^{N} \lambda_i - 1\right). \tag{8.23}$$

To minimize (8.23), we differentiate with respect to $\lambda_1, \ldots, \lambda_N$, and m in turn and set the partial derivatives equal to zero. This gives a system of equations, referred to as the *ordinary kriging equations*,

$$\sum_{j=1}^{N} \lambda_j \gamma(\boldsymbol{s}_i - \boldsymbol{s}_j) + m = \gamma(\boldsymbol{s}_0 - \boldsymbol{s}_i), \qquad i = 1, \ldots, N$$

$$\sum_{i=1}^{N} \lambda_i = 1. \tag{8.24}$$

We solve these equations for $\lambda_1, \ldots, \lambda_N$ (and m), and use the resulting optimal weights in equation (8.21) to give the ordinary kriging predictor. Note that $\hat{Z}(\boldsymbol{s}_0)$

has weights that depend on both the spatial correlations between $Z(\boldsymbol{s}_0)$ and each data point $Z(\boldsymbol{s}_i),\ i = 1, \dots, N$, and the spatial correlations between all pairs of data points $Z(\boldsymbol{s}_i)$ and $Z(\boldsymbol{s}_j),\ i = 1, \dots, N,\ j = 1, \dots, N$.

It is often more convenient to write the system of equations in (8.24) in matrix form as

$$\boldsymbol{\lambda}_O = \boldsymbol{\Gamma}_O^{-1}\boldsymbol{\gamma}_O \tag{8.25}$$

where

$$\boldsymbol{\lambda}_O = (\lambda_1, \dots, \lambda_N, m)'$$
$$\boldsymbol{\gamma}_O = [\gamma(\boldsymbol{s}_0 - \boldsymbol{s}_1), \dots, \gamma(\boldsymbol{s}_0 - \boldsymbol{s}_N), 1]'$$

and the elements of Γ_O are

$$\Gamma_{O_{ij}} = \begin{cases} \gamma(\boldsymbol{s}_i - \boldsymbol{s}_j), & i = 1, \dots, N \\ & j = 1, \dots, N \\ 1, & i = N+1;\ j = 1, \dots, N \\ & j = N+1;\ i = 1, \dots, N \\ 0, & i = j = N+1. \end{cases}$$

So (8.25) becomes

$$\begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_N \\ m \end{bmatrix} = \begin{bmatrix} \gamma(\boldsymbol{s}_1 - \boldsymbol{s}_1) & \cdots & \gamma(\boldsymbol{s}_1 - \boldsymbol{s}_N) & 1 \\ \gamma(\boldsymbol{s}_2 - \boldsymbol{s}_1) & \cdots & \gamma(\boldsymbol{s}_2 - \boldsymbol{s}_N) & 1 \\ \vdots & \ddots & \vdots & \vdots \\ \gamma(\boldsymbol{s}_N - \boldsymbol{s}_1) & \cdots & \gamma(\boldsymbol{s}_N - \boldsymbol{s}_N) & 1 \\ 1 & \cdots & 1 & 0 \end{bmatrix}^{-1} \begin{bmatrix} \gamma(\boldsymbol{s}_0 - \boldsymbol{s}_1) \\ \gamma(\boldsymbol{s}_0 - \boldsymbol{s}_2) \\ \vdots \\ \gamma(\boldsymbol{s}_0 - \boldsymbol{s}_N) \\ 1 \end{bmatrix}.$$

Note that we must calculate $\boldsymbol{\lambda}_O$ for each prediction location, $\boldsymbol{s}_0$. However, only the right-hand side of equation (8.25) changes with the prediction locations (through $\boldsymbol{\gamma}_O$). Since Γ_O depends only on the data locations and not on the prediction locations, we need only invert Γ_O once and then multiply by the associated $\boldsymbol{\gamma}_O$ vector to obtain a prediction for any $\boldsymbol{s}_0$ in D.

The minimized MSPE, also known as the *kriging variance*, is

$$\begin{aligned} \sigma_k^2(\boldsymbol{s}_0) &= \boldsymbol{\lambda}_O'\boldsymbol{\gamma}_O \\ &= \sum_{i=1}^{N} \lambda_i \gamma(\boldsymbol{s}_0 - \boldsymbol{s}_i) + m \\ &= 2\sum_{i=1}^{N} \lambda_i \gamma(\boldsymbol{s}_0 - \boldsymbol{s}_i) - \sum_{i=1}^{N}\sum_{j=1}^{N} \lambda_i \lambda_j \gamma(\boldsymbol{s}_i - \boldsymbol{s}_j), \end{aligned} \tag{8.26}$$

and the kriging standard error, $\sigma_k(\boldsymbol{s}_0)$, is a measure of the uncertainty in the prediction of $Z(\boldsymbol{s}_0)$. If we assume the prediction errors, $\hat{Z}(\boldsymbol{s}_0) - Z(\boldsymbol{s}_0)$, follow a Gaussian distribution, then a 95% prediction interval for $Z(\boldsymbol{s}_0)$ is

$$(\widehat{Z}(\boldsymbol{s}_0) \pm 1.96\sigma_k(\boldsymbol{s}_0)).$$

As an example, consider the spatial configuration shown in Figure 8.13. For illustration, assume a spherical semivariogram with $c_0 = 0, c_s = 1$, and $a = 1.5$. Then the ordinary kriging equations are

$$\begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_5 \\ m \end{bmatrix} = \begin{bmatrix} 0 & 0.852 & 0.995 & 1.00 & 0.995 & 1 \\ 0.852 & 0 & 0.852 & 1.00 & 1.00 & 1 \\ 0.995 & 0.852 & 0 & 0.911 & 1.00 & 1 \\ 1.00 & 1.00 & 0.911 & 0 & 1.00 & 1 \\ 0.995 & 1.00 & 1.00 & 1.00 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 & 0 \end{bmatrix}^{-1} \begin{bmatrix} 0.852 \\ 0.995 \\ 0.852 \\ 0.911 \\ 0.852 \\ 1 \end{bmatrix},$$

and the ordinary kriging predictor of $Z(\boldsymbol{s}_0)$ is $\hat{Z}_{\text{OK}}(\boldsymbol{s}_0) = 0.88$. Note that this is slightly larger than $\widehat{Z}(\boldsymbol{s}_0)_{\text{ID2}} = 0.53$, because $Z(\boldsymbol{s}_2) = -3$ is given smaller with kriging than it is with inverse-distance-squared interpolation. The 95% prediction interval for $Z(\boldsymbol{s}_0)$ is $(-1.07, 2.83)$.

The ordinary kriging predictor has the smallest mean-squared prediction error in the class of all linear unbiased predictors. Consequently, it is often referred to as the BLUP (best linear unbiased predictor). In practice, OK is also called the EBLUP ("E" for empirical) since the unknown semivariogram is estimated and modeled parametrically, as described in Sections 8.2.3 and 8.2.4. The resulting empirical semivariogram model then provides the values of $\gamma(\boldsymbol{s}_i - \boldsymbol{s}_j)$ and $\gamma(\boldsymbol{s}_0 - \boldsymbol{s}_j)$ needed to solve the ordinary kriging equations [equations (8.24)]. The ordinary kriging predictor is always the BLUP, regardless of the underlying statistical distribution of the data (i.e., the data need not be Gaussian to use OK). Prediction intervals are also valid for non-Gaussian data, although we do have to assume that the prediction *errors* are Gaussian to construct such intervals. However, OK may not always be the best predictor. In statistical prediction theory, the best predictor of $Z(\boldsymbol{s}_0)$ given the data is always $E[Z(\boldsymbol{s}_0)|Z(\boldsymbol{s}_1), \ldots, Z(\boldsymbol{s}_N)]$. For Gaussian data, this conditional expectation is a linear function of the data and is equivalent to simple kriging (kriging with a known mean). When the mean is unknown and the data are Gaussian, ordinary kriging serves as a very good approximation to this best linear predictor. However, when the data are not Gaussian, $E[Z(\boldsymbol{s}_0)|Z(\boldsymbol{s}_1), \ldots, Z(\boldsymbol{s}_N)]$ may not be linear, and ordinary kriging, being a linear predictor, may not provide the best approximation to this conditional mean.

As with the inverse distance methods described in Section 8.3.1, kriging is usually done locally using search neighborhoods. In fact, when kriging any particular point, there is usually no need to use data outside the range of the semivariogram because these data will have negligible weights. The use of local search neighborhoods can result in great computational savings when working with large data sets. However, the search neighborhood should be selected with care, as it will affect the characteristics of the kriged surface. Global kriging using all the data produces a relatively smooth surface. Small search neighborhoods with few data points for prediction will produce surfaces that show more detail, but this detail may be misleading if predictions are unstable. In general, try to use at least 7–12 points for each value predicted; kriging with more than 25 points is usually unnecessary. If the data are sparse or the relative nugget effect is large, distant points may have important information and the search neighborhood should be increased to include them. When the data are evenly distributed within the domain, a simple search that defines the neighborhood by the closest points is usually adequate. However, when observations are clustered, very irregularly spaced, or located on widely spaced transects, *quadrant* or *octant searches* can be useful. Here, one divides the neighborhood around each target node into quadrants or octants and uses the nearest two or three points from each quadrant or octant in the interpolation. This ensures that neighbors from several different directions, not just the closest points, will be used for prediction. It is also possible to change the search strategy node by node. It is always a good idea to experiment with several different neighborhoods and to check the results of each prediction to be sure of the calculations.

Filtered Kriging Ordinary kriging is an *exact interpolator* that *honors the data* [i.e., the kriging surface *must* pass through all data points so that $\hat{Z}(\boldsymbol{s}_i) = Z(\boldsymbol{s}_i)$ whenever a data value is observed at location $\boldsymbol{s}_i$]. However, when the data are measured with error, it would be better to predict a less noisy or filtered version of the data instead of forcing the kriged surface to honor the errors. Suppose that we can write

$$Z(\boldsymbol{s}) = S(\boldsymbol{s}) + \epsilon(\boldsymbol{s}), \quad \boldsymbol{s} \in D,$$

where $S(\cdot)$ is the true, noiseless version of the process we are studying and $\epsilon(\boldsymbol{s})$ is a measurement error process. $S(\cdot)$ need not be a smooth process; it may also exhibit small-scale or nugget variation. Cressie (1993) gives a more general model that distinguishes between several sources of variation (measurement error variation and small-scale or nugget variation being just two of these). We have adapted his development for our discussion here. We assume that $S(\cdot)$ is intrinsically stationary and that $\epsilon(\boldsymbol{s})$ is a zero-mean white noise (i.e., without spatial correlation) process, independent of $S(\cdot)$, with $\mathrm{Var}(\epsilon(\cdot)) = \sigma^2_{\mathrm{ME}}$. Repeated measurements at the same location allow an estimate of σ^2_{ME}. When the measurements are recorded by a laboratory instrument (e.g., chemical concentrations), the precision of the instrument may be known from validation studies and can be used to estimate σ^2_{ME} when replicates are not available. Note that $\gamma_Z(\boldsymbol{h}) = \gamma_S(\boldsymbol{h}) + \gamma_\epsilon(\boldsymbol{h})$ with $\gamma_\epsilon(\boldsymbol{h}) = \sigma^2_{\mathrm{ME}}$ for $||\boldsymbol{h}|| > 0$ and $\gamma_\epsilon(\boldsymbol{h}) = 0$ otherwise.

The ordinary kriging predictor of $S(\boldsymbol{s}_0)$ is derived in a manner analogous to that given above. This predictor is

$$\widehat{S}(\boldsymbol{s}_0) = \sum_{i=1}^{N} \nu_i Z(\boldsymbol{s}_i),$$

with optimal weights satisfying

$$\boldsymbol{\Gamma}_O \nu_O = \boldsymbol{\gamma}^*_O. \tag{8.27}$$

The matrix $\boldsymbol{\Gamma}_O$ is the same as in the ordinary kriging equations (8.25), with elements $\gamma_Z(\boldsymbol{s}_i - \boldsymbol{s}_j)$. The matrix $\boldsymbol{\gamma}^*_O = [\gamma^*(\boldsymbol{s}_0 - \boldsymbol{s}_1), \ldots, \gamma^*(\boldsymbol{s}_0 - \boldsymbol{s}_N), 1]'$ is slightly different since, in the minimization, the elements of this matrix are derived from $E[(Z(\boldsymbol{s}_0) - Z(\boldsymbol{s}_i))^2] = \gamma_S(\boldsymbol{s}_0 - \boldsymbol{s}_i) + \sigma^2_{\text{ME}} \equiv \gamma^*(\boldsymbol{s}_0 - \boldsymbol{s}_i)$. At prediction locations, $\boldsymbol{s}_0 \neq \boldsymbol{s}_i$, and $\gamma^*(\boldsymbol{s}_0 - \boldsymbol{s}_i) = \gamma_Z(\boldsymbol{s}_0 - \boldsymbol{s}_i), i = 1, \ldots, N$. At data locations, $\boldsymbol{s}_0 = \boldsymbol{s}_i$, and $\gamma^*(\boldsymbol{s}_0 - \boldsymbol{s}_i) = \sigma^2_{\text{ME}}(\neq 0)$. Thus, this predictor "smooths" the data, and larger values of σ^2_{ME} result in more smoothing.

The minimized MSPE is given by

$$\tau_k^2(\boldsymbol{s}_0) = \sum_{i=1}^{N} \nu_i \gamma^*(\boldsymbol{s}_0 - \boldsymbol{s}_i) + m - \sigma^2_{\text{ME}}.$$

Note that this is not equal to the ordinary kriging variance, $\sigma_k^2(\boldsymbol{s}_0)$, defined in equation (8.26), unless $\sigma^2_{\text{ME}} = 0$. Prediction standard errors associated with filtered kriging are smaller than those associated with ordinary kriging (except at data locations) since $S(\cdot)$ is less variable than $Z(\cdot)$.

Software programs sometimes implement filtered kriging by putting σ^2_{ME} on the diagonal of $\boldsymbol{\Gamma}_O$ [i.e., by replacing $\gamma_Z(\mathbf{0}) = 0$ with σ^2_{ME}]. This gives an incorrect set of kriging equations since, by definition, the semivariogram is always 0 at the origin. The quantity $\boldsymbol{\gamma}^*(\boldsymbol{h})$ results from the minimization and is really a cross-semivariogram between $Z(\boldsymbol{s}_i)$ and $S(\boldsymbol{s}_0)$ and not the semivariogram of either $Z(\cdot)$ or $S(\cdot)$.

DATA BREAK: Smoky Mountain pH Data (*cont.*) We now use the semivariogram models developed for the Smoky Mountain pH data (beginning at the end of section 8.2) in ordinary kriging to predict the pH value anywhere in the study area. We make a map of predicted stream pH by predicting several values on a regular grid of points within the domain of interest and then drawing a contour plot. To specify the grid, we must choose the starting and ending points of the grid in both directions and the grid spacing. A general guideline is to choose about 50 grid points in the *x* direction and use the resulting grid spacing in this direction to specify the grid in the *y* direction. These specifications can then be adjusted based on the domain dimensions and the spacing of the data points in both directions. For the stream pH data, we used this guideline, taking the minimum and maximum values of easting and northing to define the beginning and ending points of the grid, and using a grid spacing of 3 miles in each direction. Since it is easier to specify the grid as a rectangle, we will make predictions on this rectangular grid and then trim them to a polygon containing the data points.

We use ordinary kriging, filtered kriging, and inverse-distance-squared interpolation to predict pH values at the grid locations. For ordinary kriging, we use the geometrically anisotropic semivariogram model developed earlier: exponential with $\phi = 70^\circ$, $\hat{a}_{\max} = 36.25$, $\hat{a}_{\min} = 16.93$, $\hat{c}_e = 0.2725$, and $c_0 = 0$. For filtered kriging, we assume a nugget effect of $\hat{c_0} = 0.0325$ as indicated by Geostatistical Analyst and assume that this nugget effect is due entirely to measurement error in the pH values. We use the same semivariogram as with ordinary kriging but adapt it to include this nugget effect: exponential with $\phi = 70^\circ$, $\hat{a}_{\max} = 36.25$, $\hat{a}_{\min} = 16.93$, $\hat{c}_e = 0.2400$, and $\hat{c_0} = 0.0325$. The search neighborhood for all approaches was taken to be an ellipse oriented in the 70° direction, with the lengths of the major and minor axes equal to $\hat{a}_{\max}$ and $\hat{a}_{\min}$, respectively, and we based each prediction on the closest eight values in this neighborhood. The results appear in Figure 8.14.

Comparing the three pH maps, we can see that the filtered kriging map is slightly smoother than the ordinary kriging map, reflecting the extra smoothing done to filter the measurement error. The bull's-eye pattern characteristic of inverse-distance approaches is evident in the inverse-distance-squared pH map. We also notice that the root mean-squared prediction errors for the filtered kriging approach are higher at and near the data locations, but lower overall, also reflecting our measurement error modeling. The root-mean-squared prediction errors closely reflect the data locations, being smaller where pH samples were taken and higher in areas where there are relatively few pH measurements.

All three pH maps depict the general spatial pattern in the Smoky Mountain pH measurements, and any one of these is a decent picture of the spatial distribution of pH values. However, they do show subtle differences, reflecting the different models of spatial dependence that we used to construct them.

The accuracy of kriged maps compared to those constructed using inverse-distance-squared interpolation depends on the statistical and spatial characteristics of the data. Kriging is often more accurate (Isaaks and Srivastava 1989; Weber and Englund 1992; Gotway et al. 1996; Zimmerman et al. 1999) and preferred in many applications since the data determine the nature of the spatial autocorrelation, and a prediction standard error is associated with each value predicted. More objective approaches that can be used to select the "best" map include *validation* and *cross validation*, where subsets of the data are withheld from the analysis and then predicted with the remaining values (see, e.g., Isaaks and Srivastava 1989).

Lognormal Kriging When the data are very skewed, a linear predictor may not be the best choice, since the best predictor, the conditional expectation mentioned at the end of the discussion of ordinary kriging, may be highly nonlinear. In addition, the empirical semivariogram may be a poor estimator of the true semivariogram. Statisticians often deal with such problems by transforming the data so that the transformed data follow a Gaussian distribution and then performing analyses with the transformed data. If we want predictions on the original scale, we can transform back, but the resulting predictions will be biased. However, in certain cases, we can adjust the back transformation so that the resulting predictions are unbiased.

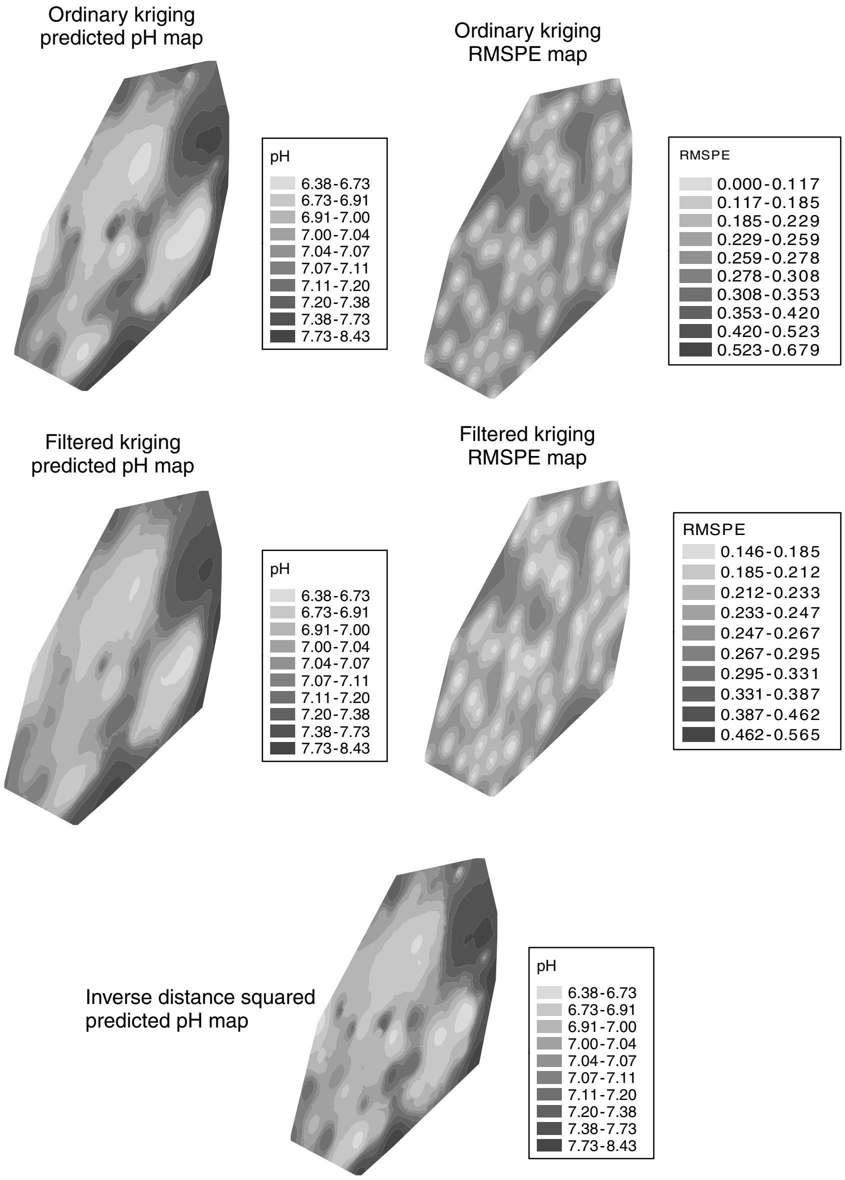

FIG. 8.14 Prediction and prediction standard error maps of Smoky Mountain pH.

With lognormal kriging, we transform the data $Z(\boldsymbol{s}_i), i = 1, \ldots, N$ to a Gaussian distribution using $Y(\boldsymbol{s}) = \log(Z(\boldsymbol{s}))$, and assume that $Y(\cdot)$ is intrinsically stationary with mean μ_Y and semivariogram $\gamma_Y(\boldsymbol{h})$. Ordinary kriging of $Y(\boldsymbol{s}_0)$ using data $Y(\boldsymbol{s}_1), \ldots, Y(\boldsymbol{s}_N)$ gives $\hat{Y}_{\mathrm{OK}}(\boldsymbol{s}_0)$ and $\sigma^2_{Y,k}(\boldsymbol{s}_0)$, obtained from equations (8.21), (8.24) and (8.26) using $\gamma_Y(\boldsymbol{h})$. Now, if we transform predictions back to the $Z(\cdot)$

scale using the exponential function, the resulting predictor $\hat{Z}(\boldsymbol{s}_0) = \exp[\hat{Y}_{\text{OK}}(\boldsymbol{s}_0)]$ is biased. However, we can use the properties of the lognormal distribution to construct an unbiased predictor. Aitchison and Brown (1957) showed that if

$$\begin{aligned} \mathbf{Y} &= (Y_1, Y_2)' \sim MVN(\boldsymbol{\mu}, \Sigma) \\ \boldsymbol{\mu} &= (\mu_1, \mu_2)', \qquad \Sigma = \sigma_{ij}, i, j = 1, 2, \end{aligned}$$

then $[\exp(Y_1), \exp(Y_2)]'$ has mean ν and covariance matrix T, where

$$\nu = (\nu_1, \nu_2)' = [\exp(\mu_1 + \sigma_{11}/2), \exp(\mu_2 + \sigma_{22}/2)]'$$

and

$$T = \begin{bmatrix} \nu_1^2[\exp(\sigma_{11}) - 1] & \nu_1\nu_2[\exp(\sigma_{12}) - 1] \\ \nu_1\nu_2[\exp(\sigma_{21}) - 1] & \nu_2^2[\exp(\sigma_{22}) - 1] \end{bmatrix}.$$

Applying this result to twice, first to $\hat{Z}(\boldsymbol{s}_0)$ and then inversely to μ_Y, gives

$$E[\hat{Z}(\boldsymbol{s}_0)] = E[\exp(\hat{Y}_{\text{OK}}(\boldsymbol{s}_0))] = \mu_Z \exp\left\{-\sigma_Y^2/2 + \text{Var}[\hat{Y}_{\text{OK}}(\boldsymbol{s}_0)]/2\right\},$$

where $\sigma_Y^2 = \text{Var}(Y(\boldsymbol{s}_i))$. Then the bias-corrected predictor of $Z(\boldsymbol{s}_0)$ (see Cressie 1993), denoted here as $\hat{Z}_{\text{OLK}}$ (for ordinary lognormal kriging) is

$$\begin{aligned} \hat{Z}_{\text{OLK}} &= \exp\left\{\hat{Y}_{\text{OK}}(\boldsymbol{s}_0) + \sigma_Y^2/2 - \text{Var}(\hat{Y}_{\text{OK}}(\boldsymbol{s}_0))/2\right\} \\ &= \exp\left\{\hat{Y}_{\text{OK}}(\boldsymbol{s}_0) + \sigma_{Y,k}^2(\boldsymbol{s}_0)/2 - m_Y\right\} \qquad (8.28) \end{aligned}$$

where m_Y is the Lagrange multiplier on the Y scale. The bias-corrected MSPE (see, e.g., David 1988) is

$$\begin{aligned} E\left[\left(\hat{Z}_{\text{OLK}} - Z(\boldsymbol{s}_0)\right)^2\right] &= \exp(2\mu_Y + \sigma_Y^2)\exp(\sigma_Y^2) \\ &\quad \cdot \left\{1 + \left[\exp(-\sigma_{Y,k}^2(\boldsymbol{s}_0) + m_Y)\right]\left[\exp(m_Y) - 2\right]\right\}. \end{aligned}$$

Thus, unlike ordinary kriging, we will need to estimate μ_Y and $\sigma_Y^2(\cdot)$ as well as $\gamma_Y(\cdot)$ in order to use lognormal kriging.

The bias correction makes the lognormal kriging predictor sensitive to departures from the lognormality assumption and to fluctuations in the semivariogram. Thus, some authors (e.g., Journel and Huijbregts 1978) have recommended calibration of $\hat{Z}$, forcing the mean of kriged predictions to equal the mean of the original Z data. This may be a useful technique, but it is difficult to determine whether or not it is needed and to determine the properties of the resulting predictor. Others (e.g., Chilès and Delfiner 1999) seem to regard *mean* unbiasedness as unnecessary, noting that $\exp(\hat{Y}_{\text{OK}}(\boldsymbol{s}_0))$ is *median* unbiased (i.e., $\Pr[\exp(\hat{Y}_{\text{OK}}(\boldsymbol{s}_0)) > Z_0] =$

$\Pr[\exp(\hat{Y}_{\mathrm{OK}}(\boldsymbol{s}_0)) < Z_0] = 0.5)$. Since the back-transformed predictor will not be optimal (have minimum MSPE), correcting for bias can be important, so we prefer to use the mean-unbiased lognormal kriging predictor given in equation (8.28).

Indicator Kriging Indicator kriging provides a simple way to make a probability map of an event of interest. Suppose that we are interested in mapping an *exceedance* probability [i.e., $\Pr(Z(s_0) > z|Z_1, \ldots, Z_N)$]. This probability can be estimated by kriging the indicator $I(Z(s_0) > z)$ from indicator data $I(Z(s_1) > z), \ldots, I(Z(s_N) > z)$ (Journel 1983), where

$$I(Z(\boldsymbol{s}) > z) = \begin{cases} 1 & \text{if } Z(s) > z \\ 0 & \text{otherwise.} \end{cases} \tag{8.29}$$

This gives an estimate of the optimal predictor, $E(I(Z(s_0) > z)|I(Z(s_i) > z)$, which for indicator data is an estimate of $\Pr(Z(s_0) > z|Z_1, \ldots, Z_N)$. The indicator kriging predictor is simply

$$\hat{I}_{\mathrm{OK}}(\boldsymbol{s}_0) = \sum_{i=1}^{N} \lambda_i I(Z(\boldsymbol{s}_i) > z),$$

where the kriging is performed as described earlier using the semivariogram estimated and modeled from the data indicator.

As some information is lost by using indicator functions instead of the original data values, *indicator cokriging* (Journel 1983), which use k sets of indicators corresponding to various threshold levels, z_k, and *probability kriging* (Sullivan 1984), which uses both the indicator data and the original data to estimate conditional probabilities, have been suggested as better alternatives. There is no guarantee, even with these methods, that the predicted probabilities will lie in [0,1] as probabilities should. In practice, various corrections are used (see Goovaerts 1997), some of which force the kriging weights to be positive. Negative kriging weights usually occur when the influence of one data value is reduced or *screened* by a closer value. Thus, one common solution to the problem of negative probabilities with indicator kriging is to decrease the search neighborhood. This can often also correct the problem of "excessive" probabilities (those > 1). If this does not work, another common correction is to reset any unrealistic values to their nearest bound, either 0 for negative probabilities or 1 for excessive probabilities.

Kriging Areal Regions Thus far, our discussion has focused on predicting values associated with spatial *locations*. In some applications we may want to predict an average value associated with a region of interest (e.g., county, census tract) from either data at individual locations or data associated with the regions themselves. Suppose that instead of observing a realization of the process $\{Z(\boldsymbol{s}) : \boldsymbol{s} \in D\}$, data $Z(B_1), Z(B_2), \ldots, Z(B_N)$ are collected where

$$Z(B_i) = \frac{1}{|B_i|} \int_{B_i} Z(\boldsymbol{s})\, d\boldsymbol{s}.$$

Here $Z(B_i)$ is the average value of the process within region $B_i \subset D$, and $|B_i|$ is the area of $B_i, i = 1, 2, \ldots, N$. In geostatistics, B_i is called the spatial *support* of $Z(B_i)$. The support of the data reflects the size, shape, and spatial orientation of the specific regions (and not just their areas) being considered. The *change of support problem* is concerned with drawing inference on $Z(B)$ from data $Z(B_1), Z(B_2), \ldots, Z(B_N)$.

A common special case of the change of support problem is the prediction of the average value of a region, $Z(B)$, from "point" samples $Z(\mathbf{s}_1), \ldots, Z(\mathbf{s}_N)$. These samples may or may not lie within region B. We consider linear prediction using

$$\hat{Z}(B) = \sum_{i=1}^{N} \lambda_i Z(s_i), \tag{8.30}$$

and derive the optimal weights, $\{\lambda_i\}$, by minimizing the mean-squared prediction error subject to the unbiasedness constraint $E(\hat{Z}(B)) = E(Z(B))$. Since $E(Z(B)) = E(Z(\mathbf{s}_i)) = \mu$, the unbiasedness constraint implies that $\sum \lambda_i = 1$. To minimize the mean-squared prediction error, we follow a development similar to that described earlier and minimize

$$-\sum_{i=1}^{N}\sum_{j=1}^{N} \lambda_i\lambda_j\gamma(\mathbf{s}_i - \mathbf{s}_j) + 2\sum_{i=1}^{N}\lambda_i\gamma(\mathbf{s}_i, B) - 2m\left(\sum_{i=1}^{N}\lambda_i - 1\right), \tag{8.31}$$

which is analogous to equation (8.23) with $\gamma(\mathbf{s}_0 - \mathbf{s}_i)$ replaced with $\gamma(\mathbf{s}_i, B)$, the *point-to-block semivariogram*. This semivariogram can be derived from the semivariogram of the $\{Z(\mathbf{s})\}$ process as (Journel and Huijbregts 1978; Cressie 1993)

$$\gamma(\mathbf{s}, B) \equiv \int_B \gamma(\mathbf{u} - \mathbf{s}) d\mathbf{u}/|B|. \tag{8.32}$$

Differentiating equation (8.31) with respect to $\lambda_1, \ldots, \lambda_N$, and m in turn and setting each partial derivative equal to zero gives the system of equations

$$\sum_{j=1}^{N} \lambda_k \gamma(\mathbf{s}_i - \mathbf{s}_j) + m = \gamma(\mathbf{s}_i, B), \qquad i = 1, \ldots, N$$

$$\sum_{j=1}^{N} \lambda_j = 1. \tag{8.33}$$

These equations are referred to as the *block kriging equations* in geostatistics. The term *block* comes from mining applications for which this approach was developed, where the goal was the prediction of the grade of a block of ore prior to mining recovery.

The minimized mean-squared prediction error, called the *block kriging variance*, is

$$\sigma_K^2(B) = \sum_{i=1}^{N} \lambda_i \gamma(\mathbf{s}_i, B) + m.$$

The point-to-point semivariogram, $\gamma(\boldsymbol{u} - \boldsymbol{s})$, is assumed known for theoretical derivations, but is then estimated from the point data and modeled with a valid conditional nonnegative definite function (as described in Section 8.2.3). In practice, the integral in equation (8.32) is computed by discretizing B into N_u points, $\{\boldsymbol{u}'_j\}$, and using the approximation $\gamma(\boldsymbol{s}_i, B) \approx 1/N_u \sum_{j=1}^{N_u} \gamma(\boldsymbol{u}'_j, \boldsymbol{s}_i)$.

The block kriging predictor, $\hat{Z}(B)$ given in equation (8.30) with weights satisfying equations (8.33), is identical to that obtained by averaging N_u ordinary point kriging predictions at the discretized nodes $\{\boldsymbol{u}'_j\}$. However, block kriging will reduce the computational effort involved in solving many ordinary point kriging systems and will also assure the correct prediction standard error.

CASE STUDY: Hazardous Waste Site Remediation This case study is based on an investigation of dioxin-contaminated soils described by Zirschky and Harris (1986). In 1971, a truck transporting dioxin-contaminated residues dumped an unknown quantity of waste in a rural area of Missouri to avoid citations for being overweight. Although the highest concentration of wastes occurred where the waste was dumped, contamination had spread to other areas. In November 1983, the U.S. Environmental Protection Agency (EPA) collected soil samples in several areas and measured the TCDD (tetrachlorodibenzo-*p*-dioxin) concentration (in µg/kg) in each sample. Figure 8.15 shows the locations of the TCDD samples, where for illustration we have transformed the study domain by dividing the *x*-coordinate by 50 to produce a region that is almost square. The objective of the study was to determine where EPA should concentrate soil remediation efforts.

One way to address the study objective is to make a map of the TCDD concentration predicted. Since concentration values must be positive and often have skewed distributions, it is common to assume a lognormal distribution, so we will

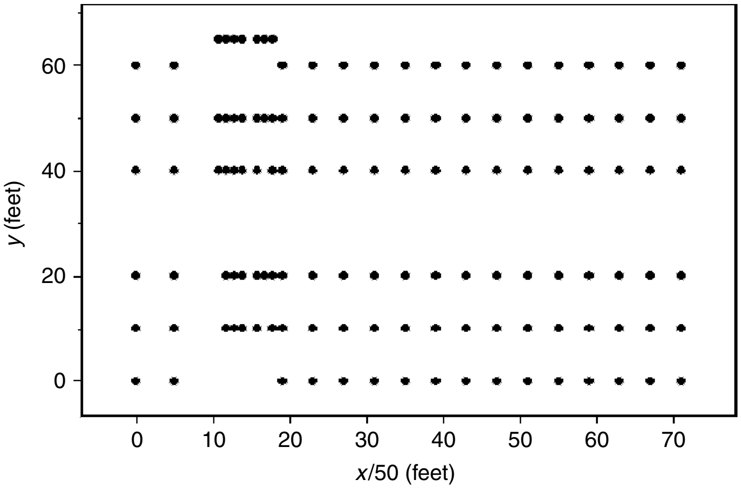

FIG. 8.15 Locations of EPA TCDD samples.

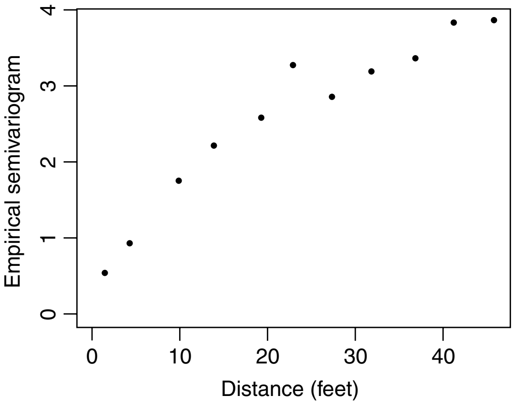

FIG. 8.16 Empirical omnidirectional semivariogram of log-TCDD data.

use lognormal kriging to make this map. The first step in lognormal kriging is to take the logarithm of the concentration data and then compute the omnidirectional semivariogram using the methods discussed in Section 8.2.3. This is shown in Figure 8.16. The initial increase in the semivariogram indicates strong spatial autocorrelation with a fairly large range and a negligible nugget effect. Given that the spill occurred along a roadway, we might expect anisotropy and can explore this using a contour map of the semivariogram surface (Figure 8.17) (see Section 8.2.5). From this figure it appears that the range of spatial autocorrelation is greatest in the east-west (0° direction), corresponding to the primary orientation of the road.

If we assume geometric anisotropy with the maximum range in the east-west direction we can fit an anisotropic model. With no other information about the laboratory process used to obtain the TCDD concentrations, we assume that the nugget effect is zero. Assuming an exponential model, we used the modified WLS approach implemented in ESRI's Geostatistical Analyst (Johnston et al. 2001) to fit a geometrically anisotropic model and obtain the estimates $\hat{a}_{\max} = 18.30$, $\hat{a}_{\min} = 8.18$, and $\hat{c}_e = 3.45$. Model fits to the east-west and north-south directional semivariograms are shown in Figure 8.18.

We predict values on a closely spaced regular grid superimposed on the domain depicted in Figure 8.15 using lognormal kriging. We use an elliptical search neighborhood oriented in the east-west direction with the length of the major and minor axes equal to $\hat{a}_{\max} = 18.30$ and $\hat{a}_{\min} = 8.18$, respectively, and retain the nearest five points from each of four quadrants for each prediction. The resulting map is shown in Figure 8.19. Based on this map, EPA should concentrate remediation efforts in the area with the highest predicted TCDD concentration, indicated by the large dark area in the center of the study domain.

An alternative approach is to recommend remediation only for those areas for which it is likely that the TCDD concentration exceeds the EPA standard. Zirschky and Harris (1986) considered 1 µg/kg as the cleanup criterion, but today, the EPA standard for disposal of sewage sludge used in agriculture is 0.3 µg/kg. We will use

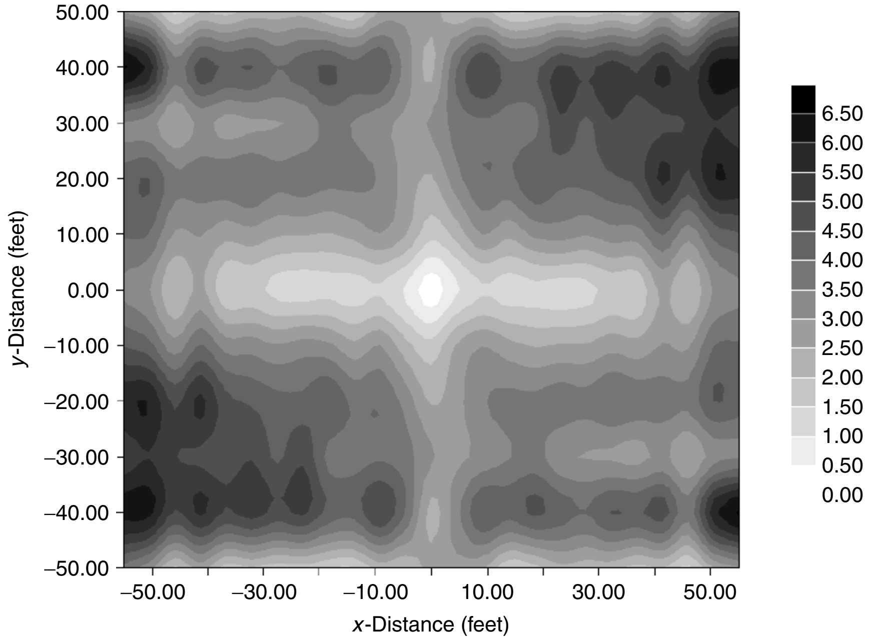

FIG. 8.17 Semivariogram surface of log-TCDD data.

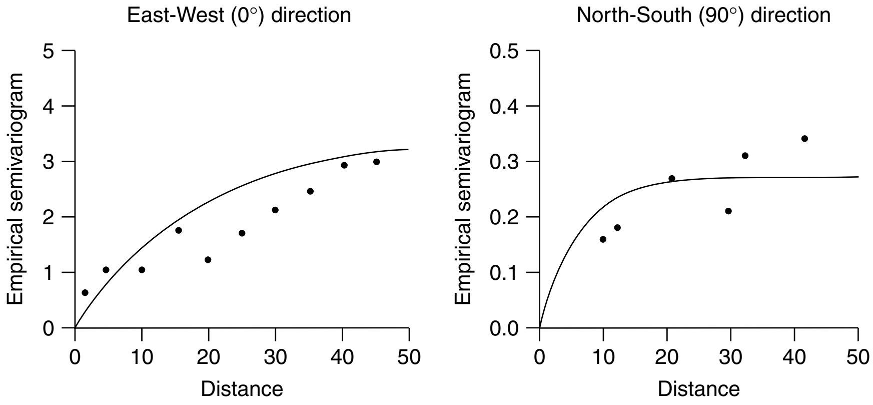

FIG. 8.18 Empirical directional semivariograms of log-TCDD data and model fit.

indicator kriging to determine areas that exceed this standard. First, we transform the data to indicators defined as

$$I(\text{TCDD}(\boldsymbol{s}) > 0.3) = \begin{cases} 1 & \text{if } \ \text{TCDD}(\boldsymbol{s}) \ > 0.3 \\ 0 & \text{otherwise.} \end{cases}$$

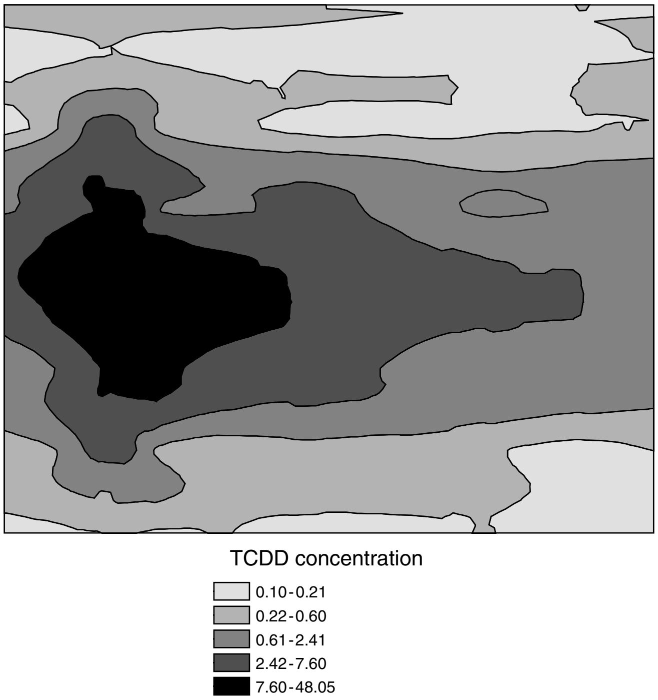

FIG. 8.19 Lognormal kriging prediction map of TCDD contamination.

The indicator process reflects the same type of anisotropy as the original concentration process, so we will consider the same type of model as used for the TCDD concentration process. Refitting this model to the indicator data, we obtain $\hat{a}_{\max} = 14.02$, $\hat{a}_{\min} = 5.57$, and $\hat{c}_e = 0.2716$. Model fits to the east-west and north-south directional indicator semivariograms are shown in Figure 8.20.

We used this fitted model to predict the indicator data on the same grid and with the same type of search neighborhood as with the original TCDD data. The resulting map, depicting the probability of exceeding the EPA standard for TCDD concentration of 0.3 μg/kg appears in Figure 8.21. This map leads to a very different remediation strategy than that inferred from the lognormal kriging map. Since most of the values that exceed our cleanup criterion of 0.3 μg/kg occur on either side of the roadway, the area with the greatest probability of exceeding this criterion occurs along the roadway as well. Thus, based on this map, we would recommend concentrating remediation efforts along a rather large swath of land centered along the roadway.

An alternative approach to constructing such a probability map is to use the lognormal kriging predictions. If we assume that the logarithms of the TCDD concentration follow a Gaussian distribution, then at each location we can calculate the probability of exceeding a threshold value. The mean and standard deviation necessary for these probabilities are the predicted value obtained from kriging and

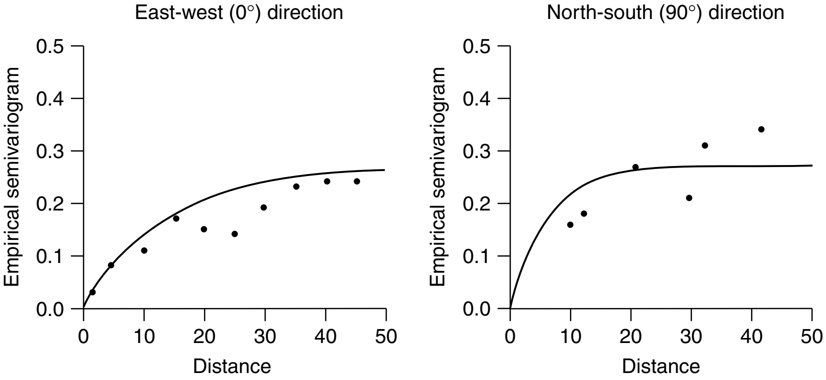

FIG. 8.20 Empirical directional indicator semivariograms and model fit.

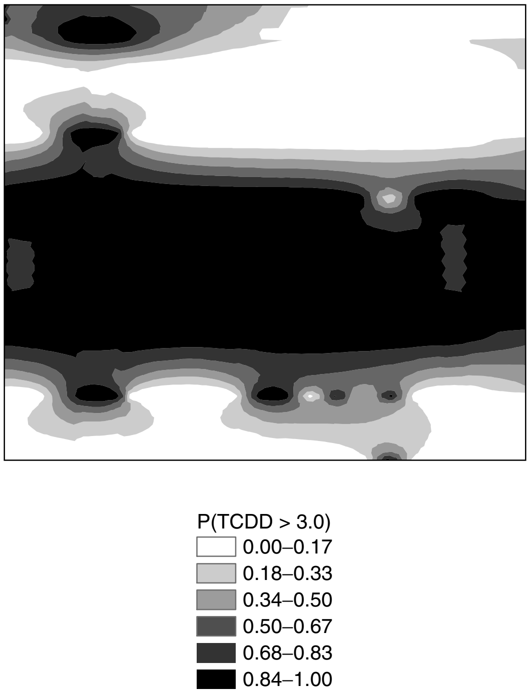

FIG. 8.21 Indicator kriging prediction map of Pr(TCDD > 0.3).

the standard error, respectively. We applied this approach to the TCDD concentration data using the same threshold of 0.3 μg/kg, and the resulting map appears in Figure 8.22. This map is similar to that made using indicator kriging, but since the actual TCDD concentration values are used in the calculations, it shows more detail.

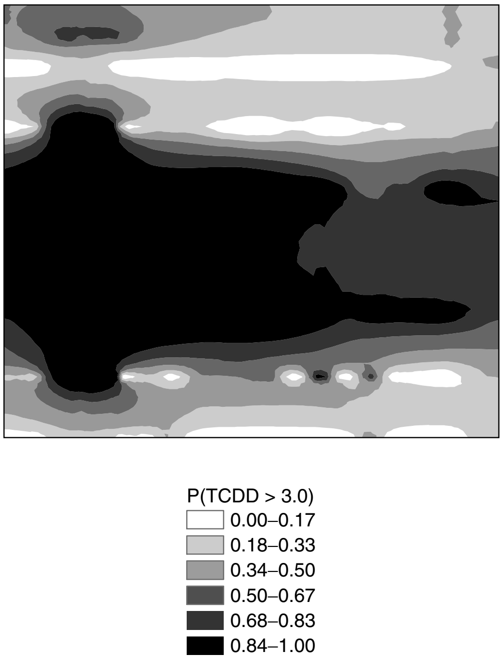

FIG. 8.22 Probability map of Pr(TCDD > 0.3) using the Gaussian distribution.

All the maps at least roughly indicate the same area of highest contamination. However, they differ greatly in the uncertainty associated with each predicted value or probability. Obtaining standard error maps for the three methods is left as an exercise.

This case study illustrates the type of analyses that geostatistical methods can provide. We would be remiss if we led you to believe that these analyses are definitive. The dioxin data represent an example of a "real" data set with many "problems" that often accompany real data. The data do not meet all of the assumptions and expectations discussed in this chapter. For example, the data are left-censored and of differing supports. In kriging the TCDD data, we essentially assumed that the spread of contamination was not affected by the placement of roadway. Nevertheless, our geostatistical methods provide valuable insight and a preliminary analysis that can form the basis for more complex models, and the case study highlights both the potential and the limitations of geostatistical analysis of environmental monitoring data.

8.4 ADDITIONAL TOPICS AND FURTHER READING

8.4.1 Erratic Experimental Semivariograms

Semivariogram estimation and modeling can be difficult, particularly if the data are noisy (either statistically or spatially), the data distribution is skewed, or there are influential observations. Good discussions of many of the common problems encountered in semivariogram estimation and model fitting, together with some suggested solutions, are given in Armstrong (1984) and Webster and Oliver (2001). Basu et al. (1997) considered the effect of influential observations on semivariogram estimation and illustrated graphical techniques for identifying such observations. Cressie and Hawkins (1980) developed several robust semivariogram estimators, using power transformations, medians, and trimmed means. One of these, based on fourth roots of squared differences, is available in many software packages as either the *robust* or the *Cressie–Hawkins semivariogram estimator*. Haining (1990) provides a good summary of robust and resistant semivariogram estimators and the results of several theoretical and empirical evaluations and comparisons.

8.4.2 Sampling Distribution of the Classical Semivariogram Estimator

Davis and Borgman (1979) tabulated the exact sampling distribution of $2\hat{\gamma}(\boldsymbol{h})$ assuming equally spaced, one-dimensional Gaussian data with a linear theoretical semivariogram. Analogous results for more general cases may be possible, but given the difficulty of this approach, attention has focused on asymptotic sampling distributions. Davis and Borgman (1982) showed that $[\hat{\gamma}(h) - \gamma(h)]/\sqrt{\mathrm{Var}(\hat{\gamma}(h))}$ converges to a standard normal distribution as the number of data values becomes large. The variance of $\hat{\gamma}(h)$ needed to compute this standardized quantity (and the covariances for the multivariate generalization needed for simultaneous confidence bands) were derived by Cressie (1985). As discussed in Section 8.2.4, these variances and covariances are not easy to obtain from the data, so approximations like the one given in equation (8.9) are used. An alternative based on iteratively reweighted generalized least squares (cf. Section 9.2) is described by Genton (1998). A comprehensive discussion and illustration of these methods and several others for assessing the uncertainty in semivariogram estimates is given in Pardo-Iqúsquiza and Dowd (2001).

8.4.3 Nonparametric Semivariogram Models

Using spectral properties of the covariance function, Shapiro and Botha (1991) introduced a nonparametric representation of the semivariogram

$$\gamma(h) = \sum_{i=1}^{k} [1 - \Omega_d(ht_i)]p_i,$$

where the values t_i are the *nodes* of the semivariogram, $\{p_i\}$ are nonzero weights and

$$\Omega_d(x) = \left(\frac{2}{x}\right)^{(d-2)/2} \Gamma\left(\frac{d}{2}\right) J_{(d-2)/2}(x),$$

where d is the spatial dimension ($d = 2$ for two-dimensional space), $\Gamma(\cdot)$ is the gamma function, and $J_m(x)$ is the Bessel function of the first kind of order m. The idea here is to build more flexible semivariogram models by adding several valid semivariograms together. This is the same idea behind the nested structure models in geostatistics that we described briefly in Section 8.2.2.

Given the nodes, constrained nonlinear least squares (constrained so that $p_i \geq 0$) can be used to estimate the weights. Various methods have been suggested for determining the number of nodes and their placement. Cherry et al. (1996) suggest using many nodes (e.g., 200) and placing half of them at equally spaced increments in [0,4] and the other half in equally spaced increments in [4.16, 20]. Ecker and Gelfand (1997) recommend using a small number of nodes (e.g., 5), spaced equally in an interval determined by the number of desired sign changes in the Bessel function and use a Bayesian approach to infer either the placement of the nodes when the weights are equal to $1/k$, or to infer the weights when the placement of the nodes is specified.

Although nonparametric semivariogram families provide a flexible alternative to parametric models, they have two main drawbacks (other than the subjectivity required in determining the nodes): their inability to fit a nugget effect and their difficulty in modeling anisotropy. Ecker and Gelfand (1999) overcome the latter difficulty by extending their Bayesian approach to geometrically anisotropic data. Ver Hoef and Barry (1996) overcome both problems using a family of semivariograms based on integration of a moving-average function over white noise random processes.

8.4.4 Kriging Non-Gaussian Data

As we discussed in Section 8.3.2, linear predictors such as ordinary and universal kriging may not be the best choice for spatial prediction with non-Gaussian data. Based on the ideas behind lognormal kriging, Cressie (1993) derives a trans-Gaussian predictor using a general transformation to a Gaussian distribution [e.g., $Y(\boldsymbol{s}) = \phi(Z(\boldsymbol{s}))$]. He uses a second-order Taylor series expansion to adjust for bias in the back transformation. Gotway and Stroup (1997) extend universal kriging to the class of generalized linear models that account for variance-to-mean relationships in the data (e.g., models for Poisson data for which the mean equals the variance). Diggle et al. (1998) had a similar goal but used a conditional specification for the generalized linear model and inference based on Bayesian hierarchical modeling. Yasui and Lele (1997) compare and illustrate both marginal and conditional generalized estimating equations for estimation of spatial disease rates using Bayesian methods, and Gotway and Wolfinger (2003) make a similar comparison of spatial prediction methods using penalized quasi-likelihood approaches.

8.4.5 Geostatistical Simulation

For many methods in spatial statistics, Monte Carlo simulation is used for inference (cf. inference methods in Chapters 5, 6, and 7). Simulation is also a powerful tool for the analysis of geostatistical data. Typically, simulation is used here for *uncertainty analysis* in order to generate a distribution of a spatial response from uncertain spatial inputs. The result is an entire probability distribution of values at each spatial location, and thus an ensemble of maps or surfaces, all possible given the data. One of the main differences between geostatistical simulation and other simulation approaches is that geostatistical simulation methods can constrain each simulated surface to pass through the given data points. This is called *conditional simulation*.

There are many different conditional simulation algorithms [see, e.g., Deutsch and Journel (1992), Dowd (1992), and Gotway and Rutherford (1994) for descriptions and comparisons]. One of the most familiar to statisticians is called *LU decomposition*, based on a Cholesky-type decomposition of the covariance matrix between data observed at study locations and data at grid or prediction locations. This covariance matrix can be written as

$$C = \begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{pmatrix},$$

where C_{11} is the covariance among data at study locations, C_{12} is the covariance between data at study locations and data to be predicted at grid locations, and C_{22} is the covariance among data to be predicted at grid locations. It can be decomposed into a product of a lower triangular matrix and an upper triangular matrix (hence the name *LU decomposition*) as

$$C = \begin{pmatrix} L_{11} & 0 \\ L_{21} & L_{22} \end{pmatrix} \begin{pmatrix} U_{11} & U_{12} \\ 0 & U_{22} \end{pmatrix}$$

(a well-known result from matrix algebra). A conditional Gaussian simulation is obtained by simulating a vector, $\boldsymbol{\epsilon}$, of independent Gaussian random variables with mean 0 and variance 1 and using the data vector **Z** in the transformation

$$\begin{pmatrix} L_{ZZ} & 0 \\ L_{21} & L_{22} \end{pmatrix} \begin{pmatrix} L_{11}^{-1} \\ \boldsymbol{\epsilon} \end{pmatrix} = \begin{pmatrix} \mathbf{Z} \\ L_{21}L_{11}^{-1}\mathbf{Z} + L_{22}\boldsymbol{\epsilon} \end{pmatrix}. \tag{8.34}$$

This transformation induces spatial correlation (specified by C) in the simulated values and also forces them to be equal to the Z values at observed data locations. It is also possible to generate realizations that do not honor the data. More details about this approach as well as other geostatistical simulation algorithms can be found in Deutsch and Journel (1992), Dowd (1992), and Gotway and Rutherford (1994).

8.4.6 Use of Non-Euclidean Distances in Geostatistics

Throughout this chapter we have used Euclidean distance as a basis for measuring spatial autocorrelation. In many applications, particularly those in the environmental sciences, this measure of distance may not be realistic. For example, mountains, irregularly shaped domains, and partitions in a building can present barriers to movement. Although two points on either side of a barrier may be physically close, it may be unrealistic to assume that they are related. Geostatistical analyses can also be done using non-Euclidean distances (e.g., city-block distance) provided that two conditions hold: (1) the distance measure is a valid metric in $\Re^2$ (i.e., it must be nonnegative, symmetric, and satisfy the triangle inequality); and (2) the semivariogram used with this metric must satisfy the properties of a semivariogram (i.e., it must be conditionally negative definite). Using these criteria, Curriero (1996) showed that the exponential semivariogram is valid with city-block distance, but the Gaussian semivariogram is not. This general issue and a proposed alternative using multidimensional scaling is given in Curriero (1996). Rathbun (1998) used these results to evaluate the use of *water distance* (the shortest path between two sites that may be traversed entirely over water) for kriging estuaries.

8.4.7 Spatial Sampling and Network Design

When we have the luxury of choosing the data locations, some thought should go into the *design* of our sampling plan. Olea (1984), Cressie (1993), and Gilbert (1997) review many different spatial sampling plans in the context of different objectives (e.g., estimation of a mean or total, estimation of spatial patterns, detection of hot spots). A systematic sampling plan using a regular grid with a random start was recommended for estimating trends and patterns of spatial variability in fixed (not mobile) populations. A triangular grid is the most efficient for semivariogram estimation and kriging and EPA's Environmental Monitoring and Assessment Program (EMAP) is based on this design [see, e.g., Stevens (1994) for an overview of this program]. Instead of a systematic triangular grid design, Stehman and Overton (1994) suggest a *tessellation-stratified design*, where the strata are defined by the squares or triangles of a regular or triangular grid and the sampling locations are chosen randomly within each stratum. This type of design allows greater variability in distances, including some that are small, and so may allow for better estimation of the semivariogram.

Since all of these sampling plans are probability based, probability sampling theory offers many tools for estimation and inference (e.g., the Horvitz–Thompson estimator and a variety of methods for variance estimation from systematic sampling). Model-based analysis (e.g., regression and kriging) can also be done. Another approach to the construction of spatial sampling plans useful in network design is based on the kriging variance. Since the kriging variance depends only on the spatial locations and not on the data values themselves, prediction errors from kriging can be determined for any particular sampling plan before the sampling is actually performed. We may then choose the sampling plan that minimizes the average (or the maximum) prediction error. Cressie et al. (1990) illustrate this approach. More recently, Bayesian methods have been used to provide a flexible framework for network design (Zidek et al. 2000).

8.5 EXERCISES

8.1 The data in Table 8.3 were collected to assess the suitability of a waste isolation pilot plant (WIPP) in southeastern New Mexico for disposal of transuranic wastes (see, e.g., Gotway 1994). Transmissivity values measuring the rate of water flow through the Culebra aquifer that lies just above the WIPP site were collected from 41 wells.

Table 8.3 WIPP Transmissivity Data[a]

East (km)	North (km)	log(*T*)	East (km)	North (km)	log(*T*)
14.2850	31.1240	−4.6839	7.0330	17.6060	−2.9136
7.4450	29.5230	−3.3692	16.7480	17.3390	−5.6089
16.7400	26.1450	−6.6023	15.3600	16.7980	−6.4842
24.1450	25.8250	−6.5535	21.3860	16.7940	−10.1234
16.7020	21.7380	−4.0191	18.2220	16.7770	−4.9271
13.6130	21.4520	−4.4500	18.3650	15.5740	−4.5057
19.8910	21.2450	−7.0115	11.7210	15.3210	−5.6897
15.6630	20.6910	−4.1296	13.6430	15.1910	−7.0354
9.4040	20.4720	−3.5412	0.0000	15.1380	−2.9685
16.7290	19.9680	−6.9685	15.3990	14.9270	−5.9960
16.7540	19.6230	−6.4913	16.2100	14.4930	−6.5213
15.2830	19.6100	−5.7775	18.7370	13.9570	−6.6361
16.7580	19.2260	−6.1903	16.9450	13.9100	−5.9685
16.7580	19.0970	−6.4003	20.0420	11.8960	−6.7132
16.7620	18.7630	−6.5705	11.1430	11.0920	−2.8125
16.3880	18.6560	−6.1149	25.9940	8.9170	−7.1234
12.1030	18.4200	−3.5571	9.4810	5.9030	−3.2584
16.7150	18.4020	−6.2964	17.0080	4.7050	−3.9019
18.3340	18.3030	−6.8804	17.9720	3.8980	−4.3350
16.4420	18.1280	−6.0290	11.7020	0.0000	−5.0547
15.6700	18.0950	−6.2005			

[a] Data are the UTM coordinates in kilometers measured from a fixed location and the log transmissivity (T) in $\log_{10}(\text{m}^2/\text{s})$.

(a) Estimate the omnidirectional semivariogram and the empirical semivariograms in the east-west and north-south directions. For each, provide the average lag distance, number of pairs, and semivariogram estimate as well as a graph of the empirical semivariograms. Do you see evidence of trend or anisotropy? Discuss.

(b) Repeat your analysis deleting the large negative log(*T*) value of −10.12. What effect does this have on your results?

8.2 Suppose that we have data $Z(\boldsymbol{s}_1), \ldots Z(\boldsymbol{s}_N)$ from an intrinsically stationary process with covariance function $C(\boldsymbol{s}_i - \boldsymbol{s}_j)$. Find the variance of the sample mean, $\overline{Z} = \sum_{i=1}^{N} Z(\boldsymbol{s}_i)/n$. Express this result in terms of the autocorrelation function $\rho(\boldsymbol{s}_i - \boldsymbol{s}_j)$ and compare it to the result you obtain assuming that the data are independent. What implication does this have for survey design based on spatial data?

8.3 Referring to the dioxin case study, Zirschky and Harris (1986) provide the TCDD data. Use ordinary kriging to map the logarithm of the TCDD concentration. How do your conclusions about waste remediation differ from those obtained using lognormal kriging and indicator kriging?

8.4 Using the logarithm of the TCDD concentration values, use different nugget effects, sills, ranges, and search neighborhoods to investigate systematically the effects of your choices on the maps of log(TCDD) concentration and prediction standard errors obtained from ordinary kriging. What conclusions can you draw about the effect of these parameters on the kriged values and standard errors? Do the maps you draw reflect the differences?

8.5 Using the TCDD data, obtain standard error maps for lognormal kriging, indicator kriging, and a probability map from the lognormal kriging using properties of the Gaussian distribution. Discuss the differences in the maps and relate them to the assumptions underlying each method.

8.6 Derive the filtered kriging equations given in equations (8.27).

8.7 Refer to the graveyard data set described in Chapter 5 and given in Table 5.2. Suppose that the affected grave sites are coded as 1's and the nonaffected grave sites are coded as 0's, so that the data set consists of *x* location, *y* location, and grave site type. Discuss the implications of using indicator kriging to map the probability of an affected grave site. What assumptions need to be made? How realistic are these?