Journal of the American Statistical Association

ISSN: 0162-1459 (Print) 1537-274X (Online) Journal homepage: https://www.tandfonline.com/loi/uasa20

Nonparametric Estimation of Nonstationary Spatial Covariance Structure

Paul D. Sampson & Peter Guttorp

To cite this article: Paul D. Sampson & Peter Guttorp (1992) Nonparametric Estimation of Nonstationary Spatial Covariance Structure, Journal of the American Statistical Association, 87:417, 108-119

To link to this article: https://doi.org/10.1080/01621459.1992.10475181

Nonparametric Estimation of Nonstationary Spatial Covariance Structure

PAUL D. SAMPSON and PETER GUTTORP*

Estimation of the covariance structure of spatial processes is a fundamental prerequisite for problems of spatial interpolation and the design of monitoring networks. We introduce a nonparametric approach to global estimation of the spatial covariance structure of a random function $Z(x, t)$ observed repeatedly at times t_i ($i = 1, \cdots, T$) at a finite number of sampling stations x_i ($i = 1, 2, \ldots, N$) in the plane. Our analyses assume temporal stationarity but do not assume spatial stationarity (or isotropy). We analyze the *spatial dispersions* $\text{var}(Z(x_i, t) - Z(x_j, t))$ as a natural metric for the spatial covariance structure and model these as a general smooth function of the geographic coordinates of station pairs (x_i, x_j). The model is constructed in two steps. First, using nonmetric multidimensional scaling (MDS) we compute a two-dimensional representation of the sampling stations for which a monotone function of interpoint distances δ_{ij} approximates the spatial dispersions. MDS transforms the problem into one for which the covariance structure, expressed in terms of spatial dispersions, is stationary and isotropic. Second, we compute thin-plate splines to provide smooth mappings of the geographic representation of the sampling stations into their MDS representation. The composition of this mapping f and a monotone function g derived from MDS yields a nonparametric estimator of $\text{var}(Z(x_a, t) - Z(x_b, t))$ for any two geographic locations x_a and x_b (monitored or not) of the form $g(|f(x_a) - f(x_b)|)$. By restricting the monotone function g to a class of conditionally nonpositive definite variogram functions, we ensure that the resulting nonparametric model corresponds to a nonnegative definite covariance model. We use *biorthogonal grids*, introduced by Bookstein in the field of morphometrics, to depict the thin-plate spline mappings that embody the nature of the anisotropy and nonstationarity in the sample covariance matrix. An analysis of mesoscale variability in solar radiation monitored in southwestern British Columbia demonstrates this methodology.

KEY WORDS: Biorthogonal grids; Dispersion; Kriging; Multidimensional scaling; Thin-plate spline; Variogram.

The development of nonparametric procedures for extrapolating (to unobserved sites) the spatial covariances of a random function sampled at a finite number of locations has lagged well behind the development of interpolation methods for the expected (or observed) value of the underlying function. Among the latter are finite element methods, stiff lamina methods, natural neighbor interpolation, kriging and/or objective analysis, splining, and local regularization [see the discussions in Ripley (1981), Sibson (1981), Creutin and Obled (1982), Watson (1984), Myers (1988), and Weerahandi and Zidek (1988)]. Kriging and objective analysis are distinguished by easily computed estimates of the variance of interpolation error at arbitrary locations. These methods, as well as the method of regularization, depend explicitly on the spatial covariance or variogram functions.

In the analysis of most spatial–temporal processes underlying environmental studies, there is little reason to expect spatial covariance structures to be stationary over the spatial scales of interest. Certainly the heuristic arguments underlying assumptions of stationarity in time series do not readily apply to spatial processes. For example, in studies of the dispersion of atmospheric pollutants, landscape or topography affects weather patterns and, thereby, the transportation of pollutants. Empirical spatial covariances may be computed using time series of residuals from some form of estimated location (and possibly time) dependent process means. These covariances may reflect paths of pollutant transportation (for data averaged over appropriate temporal scales), and should, therefore, be nonstationary insofar as the landscape varies from place to place. Recent literature includes a number of examples of data analyses addressing such nonstationary spatial covariance structure. Analyses of Escoufier, Camps, and Gonzalez (1984) and Chami and Gonzalez (1984) suggest that stationarity should not be assumed in studies of atmospheric pollution data. Berndtsson (1988) showed the spatial correlation structure of daily rainfall fields to be nonstationary and Obled and Creutin (1986) accommodated nonstationarity in rainfall events. Loader and Switzer (1992) studied nonstationary structure in monthly monitoring data on sulfate concentrations in rain, and Braud (1990) has addressed nonstationarity in monthly summaries of mean sea surface temperature.

Chami and Gonzalez (1984), Switzer (1989), and Loader and Switzer (1992) provided estimation procedures for correlations between sampling sites and unmonitored locations. These articles do not, however, provide general models for estimation at arbitrary pairs of unmonitored locations. The empirical orthogonal function approach to modeling of space–time processes (Obled and Creutin 1986; Braud 1990) provides a more general approach to estimation of nonstationary correlation structure.

We present a new approach to the estimation and graphical depiction of spatial covariance. The method is nonparametric and assumes neither isotropy nor stationarity. We suppose that a random function is sampled repeatedly at a fixed number of sampling stations so that point estimates of spatial covariance among the sampling stations can

* Paul D. Sampson is Research Associate Professor and Peter Guttorp is Associate Professor, Department of Statistics, University of Washington, Seattle WA 98195. This work was carried out in part under a contract with the Electric Power Research Institute and under support to SIMS (Societal Institute of the Mathematical Sciences) and the Department of Statistics, University of Washington, from the United States Environmental Protection Agency. The authors thank their SIMS colleagues at the University of British Columbia and Stanford University, especially Jim Zidek and Paul Switzer, and also Don Myers of the University of Arizona for helpful comments and discussion. They thank John Hay for the solar radiation monitoring problem and data.

be computed. The fundamental idea is to relate the geographic coordinates of the sampling stations to a new set of coordinates representing spatial covariance as expressed through $\text{var}(Z_{it} - Z_{jt})$, where Z_{it} and Z_{jt} represent (mean-centered) observations at stations x_i and x_j. We will refer to $D^2(x_i, x_j) = \text{var}(Z_{it} - Z_{jt})$ as the *spatial dispersion* function rather than the variogram, since the latter term conventionally suggests an assumption of stationarity. The word dispersion was chosen both for its statistical and physical connotations. We model the spatial dispersion $D^2(x_i, x_j)$ nonparametrically as a smooth function of the geographic coordinates (x_i, x_j); the only constraint to be satisfied is that the resulting model for the dispersions be conditionally nonpositive definite (Matheron 1971).

Our approach requires two tools. The first tool is multidimensional scaling (MDS), which generates a two-dimensional coordinate representation of the sampling stations with interpoint distances representing (smoothed versions of) the sample spatial dispersions. The second tool is a thin-plate spline interpolation providing a mapping between the representations of the stations in the two coordinate systems. Together, these tools provide estimates of the covariance between observations at *any* two locations in the geographic sampling region. Estimates of covariance between observations at two observed stations are smoothed versions of the original sample estimates, and, by continuity, estimates at unsampled locations are consistent with the estimates at the sampling stations. We apply the method of *biorthogonal grids* (borrowed from the field of morphometrics; Bookstein 1978a, b) to provide a graphical depiction of the spatial covariance structure of the spatial–temporal process.

This methodology was motivated by problems of environmental monitoring: mapping the spatial distribution of the deposition of chemical pollutants and helping design "optimal" networks of sampling stations for observing these distributions. Our estimates of spatial covariance can be utilized in kriging-type estimation equations to provide better spatial estimates and standard errors than those based on a misspecified stationary covariance structure. The next section of this article explains our method, with descriptions of the application of the two principal tools, multidimensional scaling and thin-plate spline interpolation (illustrated with biorthogonal grids). Section 2 explains how we carry out nonparametric estimation so as to ensure a conditionally nonpositive definite model for the spatial dispersion. As a special case, this approach also provides a basis for a general class of nonparametric models for (stationary) isotropic variograms. Section 3 presents an application of the method to solar radiation data exhibiting strong spatial covariance structure (Hay 1984). We conclude with a list of theoretical and applied questions to be addressed in further research.

1. OUTLINE OF THE METHOD

1.1 Overview

We suppose that data are available from each of N sampling stations at the same T points in time. Let $Z_{it} = Z(x_i, t)$ denote the observation taken at location x_i at time t ($i = 1, 2, \ldots, N$, $t = 1, 2, \ldots, T$). We consider models of the form

$$Z(x, t) = \mu(x) + E_\tau(x, t) + E_\epsilon(x, t),$$

where $\mu(x)$ represents a mean field, $E_\tau(x, t)$ is a mean zero, spatial–temporal process, and $E_\epsilon(x, t)$ represents measurement error and small-scale spatial variability. The process E_τ is assumed to be temporally stationary with a spatial covariance function that depends smoothly on geographic coordinates. The distribution of measurement errors is assumed to be independent of location x and time t, as well as of the process E_τ. Under this model, the spatial dispersion can be expressed as

$$\begin{aligned} D^2(x, y) &= \text{var}[Z(x, t) - Z(y, t)] \\ &= \text{var}[E_\tau(x, t) - E_\tau(y, t)] \\ &\quad + \text{var}[E_\epsilon(x, t) - E_\epsilon(y, t)]. \end{aligned}$$

In most applications of interest, the variation in the temporal process will be substantially greater than the measurement error. We assume the process to be spatially continuous at each time point, and so

$$\lim_{x \to y} D^2(x, y) = 2\,\text{var}(E_\epsilon(y, t)).$$

This "nugget effect," as it is called in geostatistics, is thus determined only by the measurement error process (see Section 2.2).

Translating monitoring data into the framework of this model is rarely trivial. For instance, as suggested previously, there is often a question about the relevant time scale for averaging. Some monitoring networks for acid rain, such as MAP3S (MAP3S/RAINE Research Community 1982), provide data on a rain-event basis, while others, such as NADP (National Atmospheric Deposition Program 1980), provide only weekly totals. The spatial covariation of a process depends on the temporal scale on which it is defined and on modeling of temporal trends that might be removed from raw data (see, for example, Haslett and Raftery 1989). Data averaged over periods of time commensurate with the time required for weather systems to travel and influence multiple sampling stations result in larger observed covariances among stations. There may also be a direct dependence of spatial variability on time or season, as Suckling and Hay (1978) observed in a study of solar radiation. These factors and others may need to be considered in determining an appropriate or relevant temporal scale for analysis.

Supposing that issues of temporal scale have been resolved, we then have an $N \times T$ data matrix Z. After subtracting the column vector of station means, this is naturally viewed as N points in $\mathbf{R}^{T-1}$ since our analysis concerns relationships among, and perhaps changes in, the set of N stations. The sample covariance between values at stations x_i and x_j is

$$s_{ij} = \frac{1}{T} \sum_{t=1}^{T} (Z_{it} - \bar{Z}_{i\cdot})(Z_{jt} - \bar{Z}_{j\cdot}).$$

A natural metric for the spatial covariance structure is

provided by the variances of differences between observations at stations x_i and x_j, estimated without assuming stationarity:

$$d_{ij}^2 = \hat{\text{var}}(Z_{it} - Z_{jt}) = s_{ii} + s_{jj} - 2s_{ij}.$$

This is occasionally referred to as the sample variogram (Matheron 1971), but most often the term variogram is associated with the assumption of second-order stationarity. We, therefore, refer to the d_{ij}^2 as spatial dispersions. Our aim is to provide for the d_{ij}^2 a general nonparametric model, computed as a smooth function of geographic coordinate pairs (x_i, x_j), that permits extrapolation to observations at unmonitored locations. The model should not impose an assumption of stationarity, but it must correspond to a nonnegative definite covariance structure. This means that the spatial dispersion model estimates $\hat{d}_{ij}^2$ (and their extrapolation to unmonitored locations) must be conditionally nonpositive definite (cf. Matheron 1971, Armstrong and Jabin 1981, Armstrong and Diamond 1984):

$$-\sum_{i=1}^{N}\sum_{j=1}^{N} \lambda_i\lambda_j\hat{d}_{ij}^2 \geq 0, \qquad \text{for all } \lambda_1, \cdots, \lambda_N, \text{ such that } \sum_{i=1}^{N} \lambda_i = 0. \quad (1.1)$$

Perhaps the simplest way to assure conditional nonpositive definiteness is through a model that specifies the d_{ij} as euclidean distances. Note that these spatial root-dispersions provide the standard transformation of the covariance matrix $S = [s_{ij}]$, considered as a similarity matrix, to a distance matrix D (Mardia, Kent, and Bibby 1979, Greenacre and Underhill 1982). It is easy to show that the matrix of d_{ij}^2 is conditionally nonpositive definite if and only if the d_{ij} are interpoint distances for a configuration in some euclidean space (see Mardia, Kent, and Bibby 1979, sec. 14.2). But the centered $N \times T$ data matrix underlying computation of the d_{ij} precisely specifies the N sampling stations as points in $\mathbf{R}^{T-1}$, with interpoint distances d_{ij}. Therefore, any continuous function f that maps the sampling stations x_i (in $\mathbf{R}^2$) into their corresponding representations in $\mathbf{R}^{T-1}$ will provide a suitable, conditionally nonpositive definite extrapolation of the spatial dispersion function observed at a finite number of sampling stations to the whole plane; that is, for any two locations x_a, x_b, $D^2(x_a, x_b) = |f(x_a) - f(x_b)|^2$. Solutions can be computed using multidimensional splines, although constraining the resulting mapping f to be an injective function (i.e., assuring that it does not map two locations into the same point in $\mathbf{R}^{T-1}$) may be difficult.

Rather than use the $(T - 1)$-dimensional representation of the sampling stations, we will consider another two-dimensional representation. To motivate this approach, we first note that stationary variogram models with elliptical anisotropy can be written $\gamma(h) = g((h'Bh)^{1/2})$, h representing a vector $x_a - x_b$ and g an appropriately chosen monotone function. (We do not consider "hole effect" and periodic variograms that are not monotone). In this case, a linear transformation of the station coordinates, $x \to y = f(x) = Ax$, where $A'A = B$, results in a problem with isotropic covariance structure, $\gamma(h_y) = g((h_y'h_y)^{1/2})$, where $h_y = Ax_a - Ax_b$. Thus nearly all the variogram models common in geostatistical practice assuming stationarity can be expressed as

$$\gamma(h_y) = D^2(x_a, x_b) = g(|f(x_a) - f(x_b)|), \quad (1.2)$$

where f is a linear transformation in the case of elliptical anisotropy and the identity function in the case of an isotropic covariance structure.

We introduce a generalization accommodating nonstationary structure by permitting f to be a 1–1 *nonlinear* mapping $\mathbf{R}^2 \to \mathbf{R}^2$ (over the geographic domain of interest). Our procedure for estimating the functions g and f that define the model (1.2) in this general case involves two steps. First, we determine a two-dimensional configuration of points $\{y_i, i = 1, 2, \ldots, N\}$ and an appropriately chosen monotone function $\hat{g}$ to approximate the dispersions from interpoint distances, $d_{ij}^2 \approx \hat{g}(|y_i - y_j|)$ for $i, j = 1, 2, \ldots, N$. Second, we determine a nonlinear 1–1 mapping $\hat{f}$ such that $\hat{f}(x_i) = y_i$, or, if necessary for smoothness, an approximate mapping $\hat{f}(x_i) \approx y_i$. The estimated model thus has the form

$$\hat{D}^2(x_a, x_b) = \hat{g}(|\hat{f}(x_a) - \hat{f}(x_b)|). \quad (1.3)$$

These steps are explained in further detail in Sections 1.3 and 1.4. A more general class of models includes mappings $f: \mathbf{R}^2 \to \mathbf{R}^p$ $(2 \leq p \leq T - 1)$.

1.2 The G Plane and the D Plane

We refer to the plane of geographical coordinates x_i of the sampling stations as the *G plane*. (This will typically be expressed in terms of rectangular coordinates derived from latitude and longitude by a map projection.) The second planar representation $\{y_i\}$ of interest is determined from the spatial dispersions so that distances $h_{ij} = |y_i - y_j|$ between stations approximate the spatial dispersions d_{ij}^2 through a monotone function. Call this representation the *D plane*, "D" for dispersion. Because the spatial dispersions are functions only of h_{ij} in the D plane, we may view the D plane as a transformation of the G plane determined so that the spatial dispersion structure is both stationary and isotropic.

1.3 Computing the D Plane Representation: Multidimensional Scaling

We compute the D plane representation of the sampling stations using the Shepard–Kruskal nonmetric multidimensional scaling algorithm as described, for example, in Mardia, Kent, and Bibby (1979, chap. 14). This algorithm determines a monotone function δ such that the matrix of $\delta_{ij} = \delta(d_{ij})$ can be accurately represented by interpoint distances in a two-dimensional metric rescaling. That is, it determines a configuration of points $\{y_i\}$ so that

$$\delta(d_{ij}) = \delta_{ij} \approx |y_i - y_j|. \quad (1.4)$$

Solving this relationship for d_{ij}^2 we obtain a function, g, that serves a role analogous to the usual variogram but applied to distances in the D plane:

$$d_{ij}^2 \equiv (\delta^{-1}(\delta_{ij}))^2 \approx g(|y_i - y_j|). \quad (1.5)$$

The multidimensional scaling algorithm proceeds by searching for a configuration $\{y_i\}$ so that the distances $h_{ij} = |y_i - y_j|$ minimize a stress criterion defined as

$$\min_{\delta} \sum_{i<j} \frac{(\delta(d_{ij}) - h_{ij})^2}{\sum_{i<j} h_{ij}^2},$$

where the minimum is taken over all monotone functions, so that $\delta(d_{ij})$ represents a least squares monotone regression of the h_{ij} on the d_{ij}.

In this application of MDS it may be appropriate to use a weighted stress criterion with weights w_{ij} introduced in the summations in the numerator and denominator. Because methods of spatial analysis focusing on interpolation depend primarily on local structure of the spatial covariance field, we may take weights proportional to inverse squared (or some other function of) distance in the G plane, $w_{ij} = 1/|x_i - x_j|^2$. The graphical depiction of spatial covariance structure explained in Section 1.4 also concerns only local behavior. We do not propose such weighting schemes explicitly on the basis of probabilistic models (although inverse variance weighting schemes, which would be difficult to estimate accurately, should be similar). They are, however, useful to robustify results against the influence of especially high or low sample dispersions between distant sampling locations.

We require the function g to provide a fitted model for the d_{ij}^2 that is conditionally nonpositive definite. Section 2.1 shows that we can meet this requirement by constraining g to a broad class of functions. One could conceivably build such a constraint into a scaling algorithm. (The usual Shepard–Kruskal algorithm constrains the function δ only to be monotone; it does not produce legitimate variograms.) In practice we have generally found the standard nonmetric scaling algorithm to provide a configuration yielding a relationship between the d_{ij} and the h_{ij} that is smooth enough to be well approximated by a function in the class defined in Section 2.1. Thus the analysis in Section 3 uses the standard nonmetric scaling algorithm, followed by the fitting of a function g to the scatterplot of the d_{ij}^2 versus the h_{ij}.

An important issue in multidimensional scaling is the choice of an initial configuration. Different initial configurations will often result in convergence to different final configurations. Fortunately we have good prior information about what preferred solutions should look like. Because the mapping f deforming the G plane into the D plane is required to be smooth, a logical choice for the initial configuration is the geographic configuration itself. With this starting point it would be especially interesting to view kinematically the search for the optimal configuration (Buja 1982). We would then see the spatial covariance pattern dynamically deforming the geographic configuration.

1.4 Mapping between the G and D Planes: Thin-Plate Splines and Biorthogonal Grids

Once the stations are located in the D plane, we must compute a smooth 1–1 function f: $\mathbf{R}^2 \to \mathbf{R}^2$ that maps the coordinate representation of the sampling stations in the G plane, x_i, into their scaled spatial dispersion coordinates in the D plane, y_i. Here we describe the idea of plane mappings, as it applies to the extrapolation of spatial dispersions. Most of the mathematical details are available in Bookstein (1978a, b, 1987).

1.4.1 Interpolation of a Differentiable $\mathbf{R}^2 \to \mathbf{R}^2$ Mapping. Let Cartesian coordinates in the G plane be denoted by (x_1, x_2) and those in the D plane by (y_1, y_2). Given N pairs of corresponding locations, (x_{1i}, x_{2i}) and (y_{1i}, y_{2i}), for $i = 1, 2, \ldots, N$, we must determine a bivariate function f: $\mathbf{R}^2 \to \mathbf{R}^2$ such that

$$\begin{pmatrix} y_{1i} \\ y_{2i} \end{pmatrix} = f\begin{pmatrix} x_{1i} \\ x_{2i} \end{pmatrix}, \qquad i = 1, 2, \ldots, N.$$

Interpolation algorithms generally select a function from a specified vector space of functions (e.g., polynomials) or choose a function to minimize a certain roughness criterion. For example, in $\mathbf{R}^1$ the familiar cubic spline minimizes $\int (f''x))^2\, dx$ subject to constraints at the knots and at the boundary. The generalization of this measure for a two-dimensional problem, $\mathbf{R}^2 \to \mathbf{R}^1$, is

$$J_2(f) = \int_{\mathbf{R}^2} \left[\left(\frac{\partial^2 f}{\partial x_1^2} \right)^2 + 2\left(\frac{\partial^2 f}{\partial x_1 \partial x_2} \right)^2 + \left(\frac{\partial^2 f}{\partial x_2^2} \right)^2 \right] dx_1\, dx_2.$$

This measure of roughness (or smoothness) is proportional to the bending energy or quadratic variation of an idealized thin plate of infinite extent (Wahba and Wendelberger 1980; Bookstein 1987). The interpolators minimizing this bending energy are simple nonlinear functions defined by linear combinations of basis functions (or kernels) centered at the N observations in the G plane. They use a total of $N + 3$ parameters. Let $x_i = (x_{i1}, x_{i2})'$, and let $u_i(x) = |x - x_i|^2 \log|x - x_i|$. Then the value of a spline at an arbitrary point $x = (x_1, x_2)'$ can be expressed as

$$f(x) = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \sum_{i=1}^{N} \beta_i u_i(x). \qquad (1.6)$$

See Meinguet (1979, 1984) or Bookstein (1987) for further details.

For the bivariate problem with locations (x_{i1}, x_{i2}) in the G plane and corresponding (y_{i1}, y_{i2}) in the D plane, we compute a bivariate function $f: \mathbf{R}^2 \to \mathbf{R}^2$ as two thin-plate splines, one for y_1 and one for y_2. For given geographic coordinates, $J_2(f)$ can be expressed as a quadratic form in the D plane coordinates $\{y_i\}$. Minimization of the sum of $J_2(f)$ over both y coordinates is invariant under rotation, translation, and scaling of the coordinate axes.

1.4.2 Smoothed Mappings and Stationary Models. The thin-plate spline just introduced may be inappropriate for one of two reasons. First, as noted in Section 1.1, the mapping f should be injective—a 1–1 mapping (over the geographic domain of interest)—so that it does not map two distinct geographic locations x_a and x_b into the same image point y. This would result in two distinct locations having a spatial dispersion estimate given only by the nugget effect for measurement error and short-scale spatial variation (see Section 2). But this is appropriate only for monitoring sites

that are arbitrarily close in the G plane (as in the siting of colocated samplers for assessment of measurement error). We may visualize a noninjective mapping as resulting from a folding of the G plane to obtain the D plane. Thin-plate splines can fold in this manner.

Second, there is a concern that the MDS calculation (Section 1.3) may result in a model (1.3) that fits the sample dispersions d_{ij}^2 too closely in view of the sampling error in the estimation of d_{ij}^2 from finite series of observations on the monitoring network. (Recall from the discussion in Section 1.1 that we could compute a model that fits the sample d_{ij}^2 exactly by representing the sampling sites in $\mathbf{R}^{T-1}$ rather than $\mathbf{R}^2$.)

In our experience, a mapping that folds usually results in a model that overfits the sample d_{ij}^2—and hence results in poor extrapolations to unmonitored locations. We can address both problems by enforcing greater smoothness on f: computing f as a smoothing spline rather than an interpolating spline. That is, we compute $f = (f_1, f_2)$ to minimize

$$\sum_{j=1}^{2} \sum_{i=1}^{N} (y_{ij} - f_j(x_{ij}))^2 + \lambda(J_2(f_1) + J_2(f_2)),$$

for specified smoothing parameter λ. The solution to this problem has the same parametric form as that given in Equation (2.6). Setting $\lambda = 0$ results in an interpolating spline, while letting $\lambda \to \infty$ drives the β coefficients to zero so that the mapping f is just the least squares linear mapping (which has zero bending energy).

Recall, as explained in Section 1.1, that a linear mapping f corresponds to a stationary variogram model with elliptical anisotropy. We can thus compute a nonparametric, stationary model for the spatial dispersion function, or variogram, by computing f as the least squares linear mapping of the G plane into the D plane.

In general, we would suggest a conventional cross-validatory approach to the choice of the smoothing parameter. However, rather than choose λ to minimize the cross-validated sum of squared errors in prediction of the G plane coordinates y_{ij}, we should choose λ to minimize the cross-validated error in prediction of the d_{ij}^2. That is, leaving out the ith site, we compute the functions $f^{(i)}$ and $g^{(i)}$ and evaluate

$$\sum_{k} (d_{ik}^2 - g^{(i)}(|f^{(i)}(x_i) - f^{(i)}(x_k)|))^2.$$

One must be careful to take into account the spatial sampling design in interpreting the results of such a cross-validation. For some problems a single site may determine local nonstationary spatial covariance structure. But, for sufficiently dense spatial sampling (with respect to the scale of the process being monitored), cross-validation should prove useful. We have not applied cross-validation to our example in Section 3 because of the limited number of sampling sites.

1.4.3 Depiction of Smooth Mappings with Biorthogonal Grids. Linear or affine transformations of the G plane, $f = Ax$, appropriate for a stationary spatial dispersion structure, are characterized by a single pair of *principal axes* given by the eigenvectors of $B = A'A$. The geographic direction or axis corresponding to the largest eigenvalue is the direction in which the geographic plane is (relatively) most stretched. Given two locations x_a and x_b such that $|x_a - x_b| = c$, the spatial dispersion is greatest when x_a and x_b fall on a line parallel to this principal axis since $|Ax_a - Ax_b|$ is thus maximized. Conversely, for a given geographic separation, spatial dispersion is least (spatial covariance greatest) when x_a and x_b are aligned parallel to the principal axis of smallest eigenvalue.

The affine derivative matrix A, together with the variogram function g, completely characterize stationary spatial dispersion models (with elliptical anisotropy). We characterize smooth nonstationary spatial dispersion structures as those for which the principal axes of the affine derivative matrix of the mapping f vary smoothly from place to place. For any differentiable transformation f we can evaluate the affine derivative matrix

$$\begin{pmatrix} \partial f_1/\partial x_1 & \partial f_2/\partial x_1 \\ \partial f_1/\partial x_2 & \partial f_2/\partial x_2 \end{pmatrix}$$

at any point (x_1, x_2). The (local) principal axes computed from this matrix are differentials at 90°. It is a theorem that through almost every point of a differentiable transformation pass just two differentials, which are at 90° both before and after the transformation. (See Bookstein 1978a). The integral curves of these differentials, sampled finitely, form a curvilinear grid whose intersections are at 90° in both images. These are called the *biorthogonal grids* for the transformation because there are two such grids, one in each image (or plane), corresponding curve for curve and intersection for intersection. They provide a useful depiction of the character of the nonlinear transformation.

The simplest example of such a biorthogonal grid is given in Figure 1 for the mapping of a square into a parallelogram. This is a linear mapping and hence the affine derivative matrix is constant. As the principal axes are constant, the biorthogonal grid is just a rectangular grid indicating relative stretching of the plane along a northwest–southeast axis and contraction in the orthogonal direction. If this mapping represented a spatial process, it would indicate that spatial covariance was weakest along the northwest–southeast direction and strongest along the southwest–northeast direction. The figure refers to the factors of relative extension or contraction (ratios of distance in domain to distance in image) as gradients (grad).

A nonlinear mapping is illustrated in Figure 2 (replicating figure VI-6 in Bookstein 1978a). In this example one of the integral curves of the differentials, corresponding all along its length to the local principal axis of greater extension, follows a direction that nearly connects sites 2 and 4. The estimated spatial dispersion is greatest along these curves, while curves of the orthogonal family lie along directions of weakest spatial dispersion (strongest spatial covariance). The nonstationarity and the degree of local anisotropy is reflected in the value of the gradients, which have been coded into four intervals by different line types.

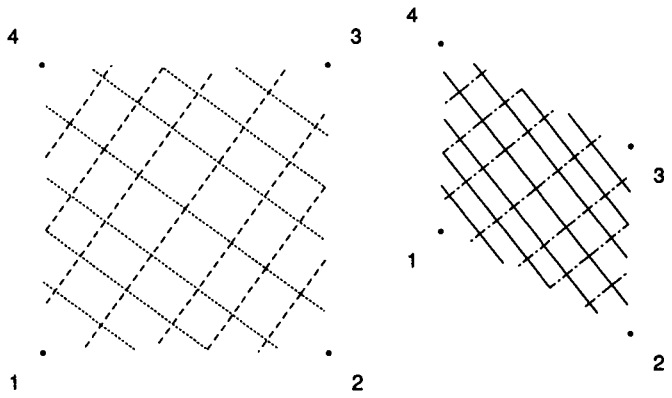

Figure 1. Principal Axes Illustrating a Linear Mapping of a Square (G Plane) into a Parallelogram (D Plane). The ratio of distance in the image to distance in the domain, labeled "grad," is .86 in the direction of the first principal axis (dotted line) and .51 in the direction of the second principal axis (dashed line). The grid on the right side encodes these values for the inverse mapping. On the left side grid ——— indicates grad < 1/4, —·—·— indicates 1/4 < grad < 1/2, — — — indicates 1/2 < grad < 3/4, and ····· indicates 3/4 < grad. On the right side grid ——— indicates grad < 4/3, —·—·— indicates 4/3 < grad < 2, — — — indicates 2 < grad < 4, and ····· indicates 4 < grad.

2. A CLASS OF CONDITIONALLY NONPOSITIVE DEFINITE VARIOGRAM MODELS

Identifying useful parametric variogram models and testing variogram models for conditional nonpositive definiteness have received much attention in the geostatistics literature. See, for example, Armstrong and Jabin (1981), Armstrong and Diamond (1984), Dunn (1983), and Myers (1984). Journel and Huijbregts (1978) provided a catalog of the most commonly used models. Any one of these parametric models could be used as the function g in the general procedure outlined in Section 1. Estimation of spatial covariance from the composition of the function g together with the thin-plate spline mapping f as in (2.3), however, leaves us with essentially no guidance for specification of a parametric model. It would clearly be preferable to estimate a monotone function g nonparametrically.

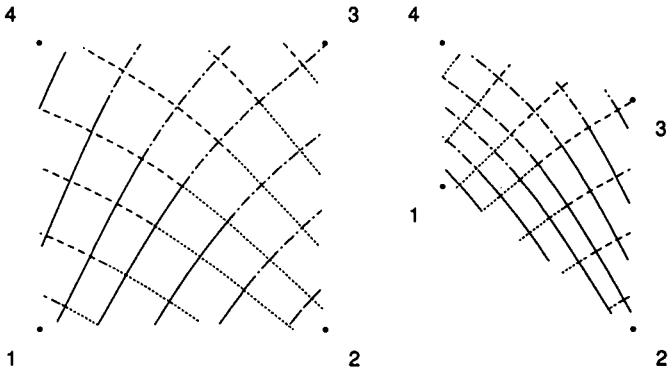

Figure 2. Biorthogonal Grid Illustrating a Nonlinear Deformation From G Plane (Left) to D Plane (Right) for Four Sampling Stations. Line types code the gradient of the deformation (local ratio of distance in the image to distance in the domain) along the curves of the grid. On the left side grid ——— indicates grad < 1/4, —·—·— indicates 1/4 < grad < 1/2, — — — indicates 1/2 < grad < 3/4, and ····· indicates 3/4 < grad. On the right side grid ——— indicates grad < 4/3, —·—·— indicates 4/3 < grad < 2, — — — indicates 2 < grad < 4, and ····· indicates 4 < grad.

Results of Matérn (1986, sec. 2.3), dating back at least to 1960, provide a basis for general nonparametric estimates suitable both for our modeling of nonstationary spatial dispersion and for traditional variogram modeling under isotropy. Specifically, Matérn showed that *all* continuous isotropic correlation functions in two dimensions can be expressed as probability mixtures of Bessel functions of the first kind of order zero. Mixtures of functions of higher order Bessel functions are necessary in higher dimensions. See also Whittle (1963) and Vecchia (1985).

Here we use a result of Schoenberg (1938) (cf. Matérn 1986, p. 17) that states that the class of continuous isotropic correlation functions valid in *any* dimension is the class of probability mixtures of Gaussian type correlations

$$r(|h|) = \int_{-\infty}^{\infty} \exp(-t^2|h|^2)\, dF(t),$$

where F is an arbitrary one-dimensional distribution function. This is useful as, in general, we can consider representations of the monitoring sites in $\mathbf{R}^p$ ($p \geq 2$). Matérn (1986, pp. 17ff) described how certain large subclasses of correlation functions can be obtained by suitable choices of F.

We obtain a family of conditionally nonpositive definite spatial dispersion models by taking

$$g(|h|; \mu) = c(1 - r(|h|)) = \int_{-\infty}^{\infty} (1 - \exp(-t^2 \mid h \mid))\, d\mu(t), \quad (2.1)$$

where μ is an arbitrary finite measure. This class of models is based on assumed finite covariances. (In geostatistics variograms are commonly introduced because of the "intrinsic hypothesis" of stationarity of differences, which allows processes with infinite variance. We see no reason to consider infinite variances for the environmental applications we have in mind.) Within this relatively flexible class of models, we compute a nonparametric estimate of spatial dispersion to approximate the observed spatial dispersions d_{ij}^2. Given n observed dispersions and corresponding displacements h_{ij} in the D plane (or, in the case of a stationary isotropic model, the G plane), we determine a least squares fit to the observed spatial dispersions in terms of a discrete distribution μ, specified by its support $t_1, \ldots, t_k$ and corresponding nonnegative numbers $\alpha_1, \ldots, \alpha_k$, where $k \leq n/2$. Details are provided in the Appendix. A weighted least squares procedure, as in Cressie (1985), could also be used.

If the observations include substantial measurement errors (as is the case in environmental monitoring), it is necessary to incorporate a nugget effect in the fitting. The general model in Section 1 implies that $D^2(x, y) \to 2\sigma_\epsilon^2$ as $x \to y$. For the estimate $\hat{D}^2(x, y) = \hat{g}(|\hat{f}(x) - \hat{f}(y)|)$ outlined in the previous section, we get $\lim_{x \to y} \hat{D}^2(x, y) = \hat{g}(0) = 0$.

A simple adjustment is to increase slightly the class of correlation functions used. Let

$$g^*(|h|; \mu) = |\mu| \left(g_0 1_{(|h|=0)} + (1 - g_0) \times \int_{-\infty}^{\infty} (1 - \exp(-t^2|h|^2)) \frac{d\mu(t)}{|\mu|} \right), \quad (2.2)$$

where $0 \leq g_0 \leq 1$, $1_{(\ )}$ denotes the indicator function, and $|\mu| = \int_{-\infty}^{\infty} d\mu(t)$. Then (cf. Matérn 1986, eq. 2.2.2) g^* is a conditionally nonpositive definite function, and hence a valid dispersion. We can easily incorporate fitting the nugget effect in the algorithm, as outlined in the Appendix.

3. EXAMPLE: SPATIAL VARIATION IN SOLAR RADIATION

We present here a preliminary analysis of data collected from a solar radiation monitoring network in southwestern British Columbia, Canada (Hay 1984). Accurate assessments of mesoscale variations in solar radiation—and hence monitoring network design and spatial interpolation—are important for solar energy system design. This example manifests a somewhat extreme, but easily explained, form of nonstationarity in the spatial covariance structure of the solar radiation field. Figure 3, taken from Hay (1983), displays the locations of the 12 monitoring stations.

The data consist of daily solar radiation totals ($MJ\ m^{-2}\ day^{-1}$) for the years 1980–1983. Figure 4a plots the data for the monitoring station at Vancouver International Airport. Note the relatively sharp upper bound on the maximum solar radiation as a function of season. Following Duffie and Beckman (1974), we computed the expected solar radiation outside of the atmosphere for southwestern British Columbia. This curve of extraterrestrial solar radiation, $H_0(t)$, very nearly matches a cosine function with maximum at the vernal equinox. We based our analysis on the idea that a reasonable stochastic model for daily observations Z_t^* at one location is

$$Z_t^* = \theta_t(\alpha + \beta H_0(t))(1 + \epsilon_t),$$

where observations are taken daily ($t = 1, 2, \cdots, 365$), ϵ_t represents a mean zero measurement error effect, and θ_t is a random variable expressing atmospheric effects and taking values on the interval (0, 1]. That is, we chose to estimate coefficients α and β to linearly scale the radiation outside the atmosphere to estimate the maximum expected solar radiation at the earth's surface (across all sites) on

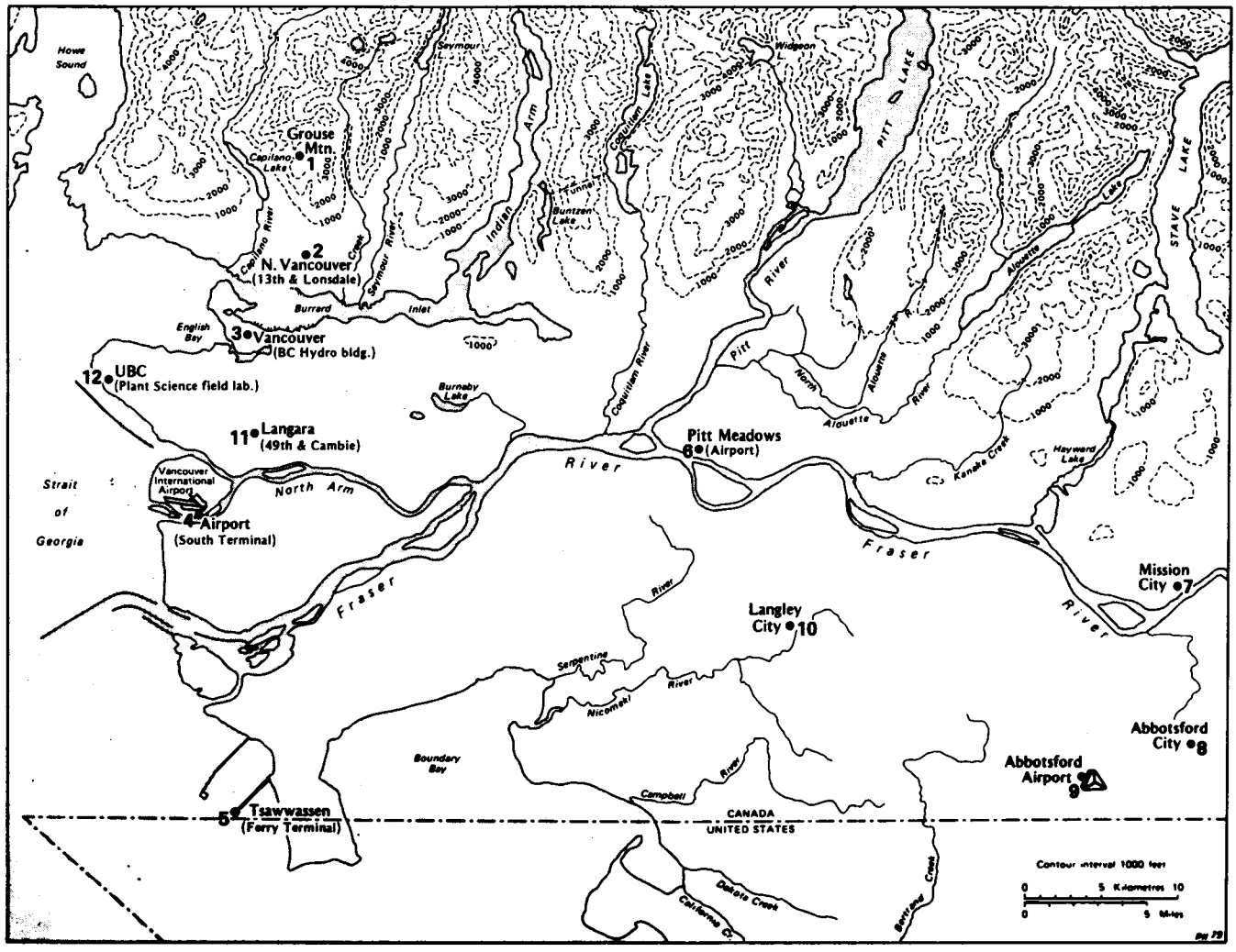

Figure 3. The 12-station Solar Radiation Monitoring Network in Lower Mainland British Columbia (from Hay 1983).

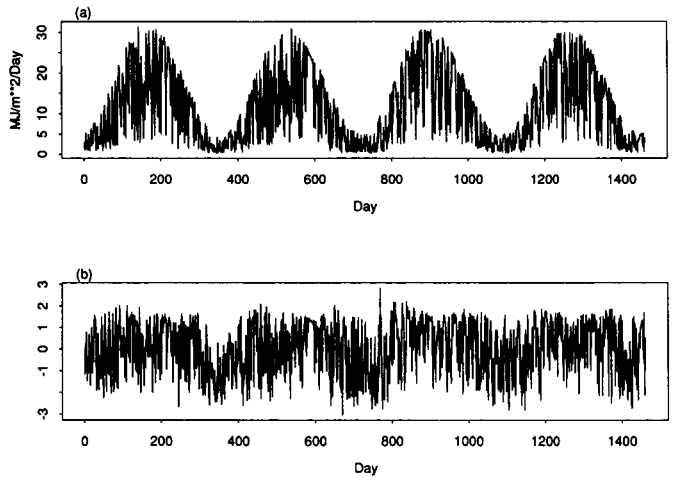

Figure 4. Daily Solar Radiation Totals for Vancouver International Airport (Site 4): (a) Raw, (b) Transformed.

cloudless days. Cloudiness is the principal factor determining θ_t, which reduces observed solar radiation below the maximum expected on cloudless days.

We estimated α and β using a nonlinear optimization routine to determine the least upper bound in the family of curves $\alpha + \beta H_0(t)$ for the data across all sites and years. We then scaled all the data as a percentage of this estimated seasonally varying maximum possible solar radiation. This transformation removed the major component of seasonality in the range of the data. Substantial seasonal differences remained, however, in both the marginal distributions and in the spatial covariances due to systematic variation in the predominant meteorological patterns. We, therefore, analyzed the data separately for autumn–winter (September 21 to March 21) and spring–summer seasons (March 22 to September 20).

The solar radiation distributions on this ratio scale have very short tails, with concentrations of observations near 100%. (Yes, there are sunny days here in the Pacific Northwest.) We, therefore, further applied a logit transformation. The observations subject to spatial analysis were then

$$Z_t = \log\left(\frac{Z_t^* - (\hat{\alpha} + \hat{\beta}H_0(t))}{\hat{\alpha} + \hat{\beta}H_0(t)}\right).$$

This transformation does not produce symmetric distributions, but it does improve the tail behavior of the marginal distributions. The transformed data for Vancouver International Airport is plotted in Figure 4b.

The daily observations are temporally correlated at lags of up to five days. However, our interest is not primarily in modeling this temporal structure. We thus proceeded to compute sample spatial correlations r_{ij} and spatial dispersions d_{ij}^2 among the 12 stations using the transformed solar radiation series Z_t. Although the temporal correlation structure yields dispersion estimates that have larger variability than for uncorrelated spatial series, this is not a major concern for the purposes of this article. Table 1 reports the spatial dispersions for summer.

Interstation correlations range from .80 to .99, and a casual inspection of Table 1, together with Figure 3, shows that the dispersions are indeed closely related to geographic distances among the stations. Figure 5 shows the conventional variogram plot of spatial dispersion versus geographic distance. Clearly no isotropic variogram model will accurately reflect these data. The distinct spatial dispersions involving station 1 (marked by ×'s in Figure 5) are reminiscent of a feature of the spatial covariances of wind power data analyzed by Haslett and Raftery (1989); see Guttorp and Sampson (1989).

The right side of Figure 6 shows the distribution of monitoring stations in the D plane, as determined by multidimensional scaling applied to the matrix of root-dispersions (d_{ij}). In Figure 7 we see that a scatterplot of distances in the D plane (h_{ij}) versus the observed spatial dispersions (d_{ij}^2) for summer displays nearly a straight line. The tight scatter about this line (in contrast to Figure 5) shows that this two-dimensional MDS solution accurately reflects the data in the spatial dispersion matrix. Because the mapping depicted in Figure 6 is smooth (with no folding), no further smoothing was judged necessary in this case. For this scatter we fitted isotropic variogram functions g from the family of Gaussian mixtures described in Section 2. The solid curve in Figure 7 shows the solution for a discrete mixing distribution with three support points, and the dotted line is the solution based on two support points. They are nearly identical, with residual standard deviations of .081 and .083,

Table 1. Observed Spatial Summer Dispersions for the 12 Solar Radiation Monitoring Sites

Site	1	2	3	12	11	4	5	6	10	7	8	9
1	—	34	36	50	41	50	63	38	42	47	49	49
2	34	—	5	14	9	15	26	16	18	31	31	29
3	36	5	—	12	6	12	25	14	17	31	31	29
12	50	14	12	—	7	4	12	22	21	41	36	32
11	41	9	6	7	—	6	17	14	15	33	30	27
4	50	15	12	4	6	—	10	19	17	37	31	29
5	63	26	25	12	17	10	—	26	24	40	34	31
6	38	16	14	22	14	19	26	—	6	14	14	14
10	42	18	17	21	15	17	24	6	—	15	11	10
7	47	31	31	41	33	37	40	14	15	—	6	9
8	49	31	31	36	30	31	34	14	11	6	—	3
9	49	29	29	32	27	29	31	14	10	9	3	—

NOTE: Values in the table are 100 × (d_{ij}^2). Rows and columns of this table have been sorted to reflect the geographic clustering of sites.

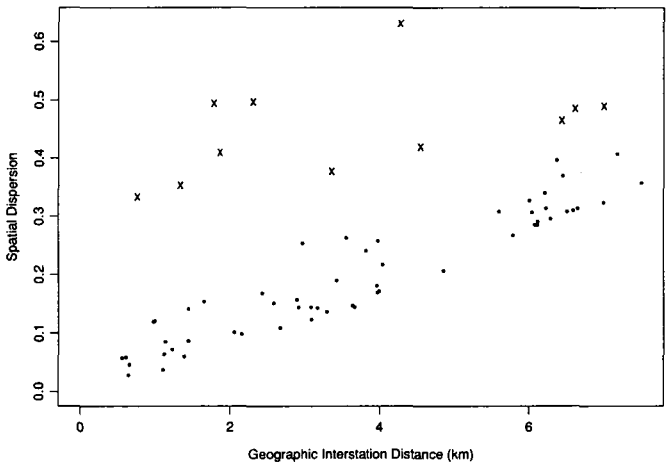

Figure 5. Spatial Dispersion Versus Geographic (G plane) Interstation Distance for the Solar Radiation Data. Dispersions involving station 1 (Grouse Mountain) are marked with ×'s.

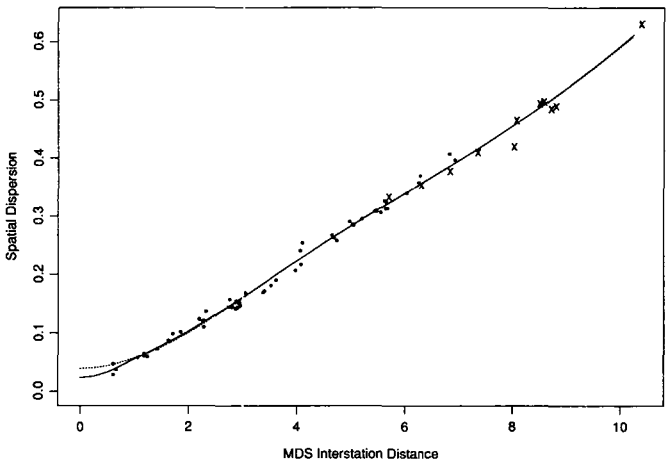

Figure 7. Spatial Dispersion Versus MDS (D plane) Interstation Distance for the Solar Radiation Data. Dispersions involving station 1 (Grouse Mountain) are marked with ×'s. The solid curve is the fitted variogram model g using a three-point mixture of Gaussian terms. The dashed line is the fitted g using a two-point mixture.

respectively, for fitting the d_{ij}^2. They differ primarily in their extrapolations to the intercept or nugget effect, .023 versus .039, respectively. Both these nugget effects are reasonably consistent with the known measurement error of about 3% on the raw scale.

The most obvious deviation between the G plane and D plane representations is in the relative location of station 1, Grouse Mountain. The Grouse Mountain station is at an elevation of 1,128 meters, while all other stations lie below 130 meters of altitude. This orographic feature explains the relatively high dispersions (low covariances) between station 1 and all the others, as reflected in the scaling of Figure 6.

Figure 6 also depicts the thin-plate spline mapping between the G plane and D plane representations of the monitoring stations using the image of a rectangular grid of points located on the G plane. In Figure 8 we display a biorthogonal grid for the deformation. As in Figures 1 and 2, the curves in the grid are coded to represent the principal directional gradients of the deformation and hence the relative directional magnitude of the spatial dispersion. The clearest feature in Figure 8 is the depiction of relatively weaker spatial covariance in a north–south direction, especially along the coast, as indicated by the dotted north–south curves and the relatively denser spacing of curves running generally east–west. Reanalysis without station 1 showed that the estimated nonstationarity is not due to just the influence of station 1.

We can also display the covariance structure through a series of plots of contours of constant spatial dispersion relative to fixed stations x_i, $\{y : D^2(x_i, y) = c\}$. Plots for spatial dispersions relative to stations 2, 4, 7, and 10 are pre-

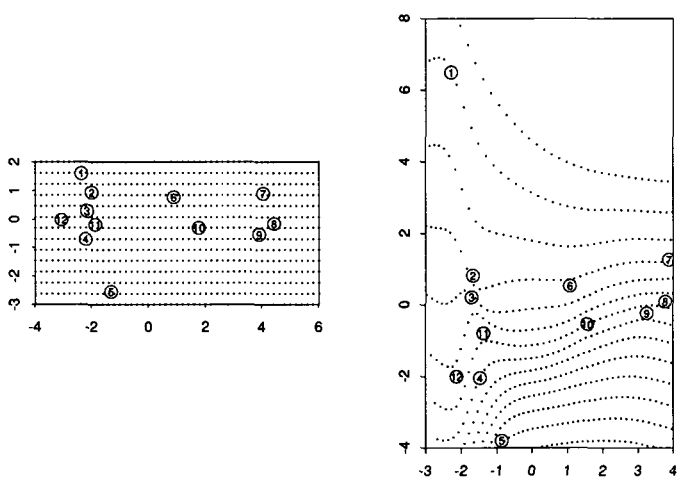

Figure 6. Transformation of the G-plane Configuration of Solar Radiation Monitoring Stations (Left) into the D-plane Configuration (Right). The D-plane configuration was determined using multidimensional scaling. The interpolation of the mapping of monitoring stations by thin-plate splines is illustrated by the image of a rectangular grid of points located in the D plane.

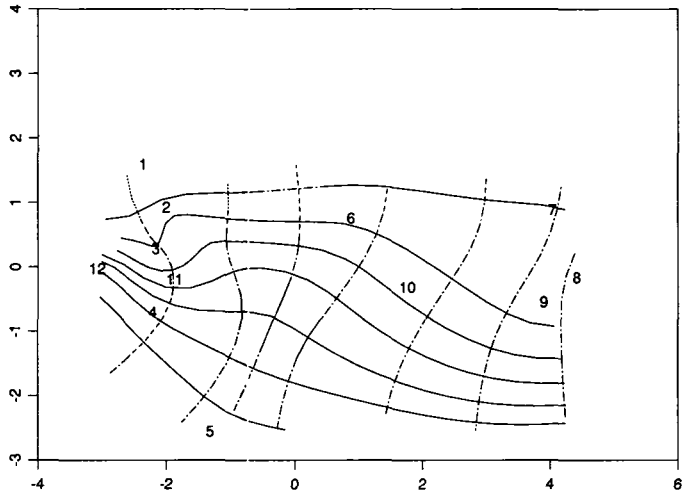

Figure 8. Biorthogonal Grid for the Interpolated G-plane to D-plane Mapping Representing the Spatial Dispersion Structure of the Solar Radiation Data. Line types code the gradient of the interpolated mapping along the curves of the grid: ——— indicates grad <1, —·—·— indicates 1 < grad < 2, — — — indicates 2 < grad < 4, ····· indicates 4 < grad.

sented in Figure 9. Note the correspondence between the orientation of these contours and the directions of the principal axes of the biorthogonal grids at these locations.

4. DISCUSSION

4.1 Nonparametric Inferences

Standard errors of dispersion estimates may be computed, conditional on the given configuration of monitoring stations, by bootstrapping over (independent) replications in time. We have demonstrated this approach elsewhere in an analysis of acid deposition data (Guttorp, Sampson, and Newman 1992).

In presenting a general approach to nonparametric estimation of nonstationary spatial covariance structure, we noted that stationary (but anisotropic) models can be fitted by constraining the mapping from the G plane to the D plane to be affine. In this case the biorthogonal grid is rectangular and uniform across the region of study, indicating a single direction of greatest spatial covariance. By examining the two-dimensional residual vectors from a (preferably robust) linear fit of the D plane coordinates $\{y_i\}$ to the G plane coordinates $\{x_i\}$ we can investigate, formally or informally, the hypothesis of stationarity. Significant spatial structure in these residual vectors is evidence against the assumption of stationarity.

A formal test of nonstationarity could be based on a bootstrap procedure. Compute bootstrap estimates (bootstrapping, again, over time replicates) of the standard error of the spatial dispersion estimates for each pair of stations. Then fit a stationary model and compare the size of the errors in the fitted spatial dispersion estimates at the station pairs to the bootstrap standard errors for those pairs.

We may also assess local features of nonstationarity in the neighborhood of the monitoring stations by leaving out stations one at a time. We would predict the spatial dispersions involving, say, station i, using the data without station i, and then compare these predictions with the observed estimates d_{ij}^2 and their standard errors. If the predictions based on the data without the ith observation agree well with the estimates d_{ij}^2, then there are no local features of nonstationarity not captured by the model based on the other N-1 stations.

4.2 Alternative Approaches

We have proposed estimation of the functions g and f in the model (1.2) in two fixed steps. For the current application, this approach appears satisfactory. There should, however, be more efficient (and more elegant) algorithmic approaches to joint estimation of the functions f and g that do not involve an intervening multidimensional scaling step. It would be natural to consider an alternating estimation scheme: estimating the variogram function g for fixed mapping f, then estimating the mapping f for fixed g, and so on. Even such an alternating scheme, however, does not address the estimation of a smoothing parameter for the thin-plate splines. Our application to solar radiation monitoring data is an exceptionally clean problem in that little or no smoothing is called for because of the high spatial correlations. An analysis of acid deposition in Guttorp et al. (1992) concerns a process for which the spatial covariance structure is much weaker and limited time series of data result in noisy estimates of these spatial covariances. In that application it was necessary to compute f as a smoothing spline. We are currently investigating other methods for estimating f and g.

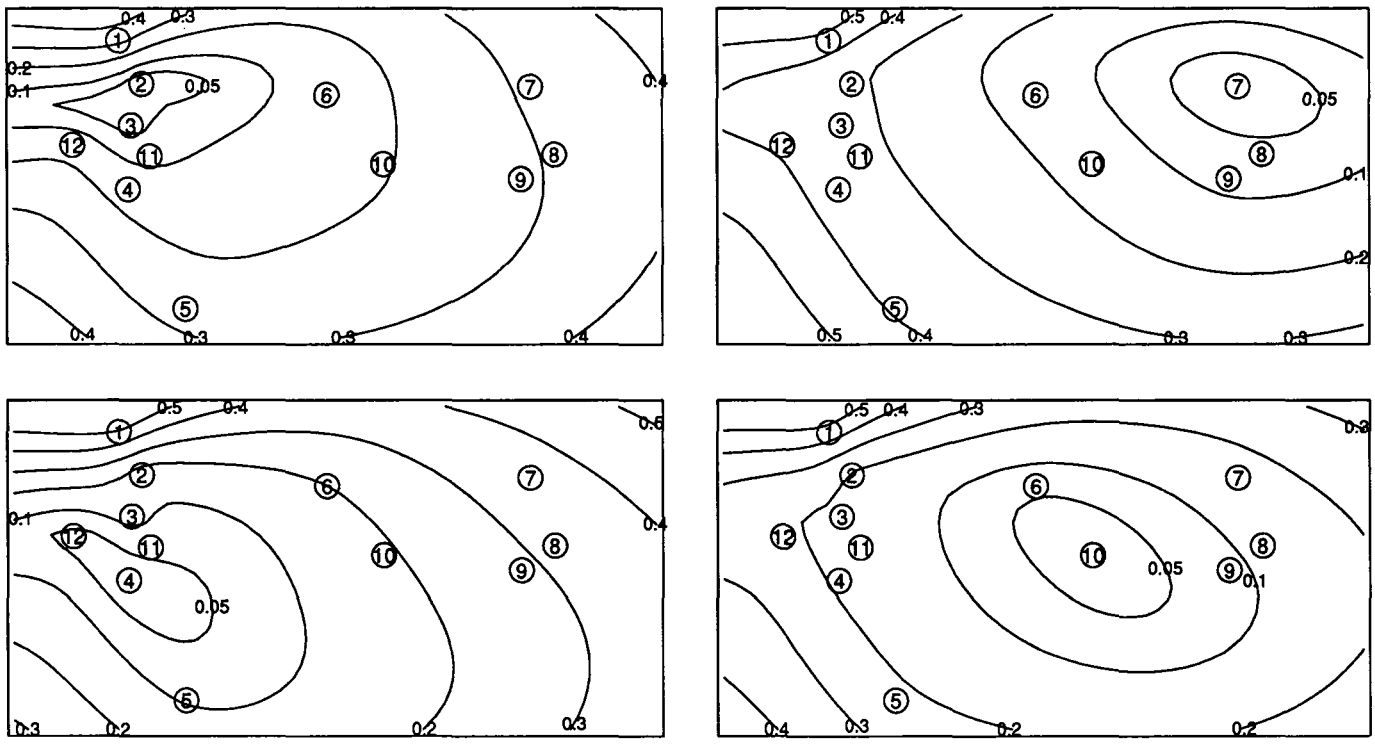

Figure 9. Estimated Contours of Constant Spatial Dispersion Relative to Sites 2, 7, 4, and 10.

We began by suggesting that sample dispersions d_{ij}^2 could be interpolated exactly by representing the sampling stations as points in $\mathbf{R}^{T-1}$ and computing an interpolating mapping from the G plane ($\mathbf{R}^2$) to this higher dimensional space. We have not found two-dimensional representations of the sampling stations to be adequate for all applications. If, for example, spatial dispersion happens to depend on latitude, longitude, and elevation (as is the case, to some extent, for the solar radiation example in Section 3), then a three-dimensional multidimensional scaling solution would probably be appropriate (presuming sufficient variation in elevation over the region of interest). Our approach carries over readily to higher dimensional representations, even including the graphical display by biorthogonal grids. If we scale into $\mathbf{R}^3$, the D plane is replaced by a two-dimensional manifold, and the biorthogonal grids (still) depict the local extrema of ratios of image distance (in the D manifold) to domain distance (in the G plane).

The empirical orthogonal function approach to spatial analysis (Obled and Creutin 1986; Braud 1990) effectively represents the monitoring sites in a (typically) high-dimensional euclidean space and expresses the dispersion function as linear in distance in this high-dimensional space. An analysis of rain events by Obled and Creutin (1986) resulted in a 39-dimensional representation.

Recently Switzer (1989) and Loader and Switzer (1992) have also proposed an approach to estimation of nonstationary spatial covariance structure from monitoring data. They use a Bayesian paradigm to shrink observed spatial covariance estimates among the sampling stations toward estimates provided by a parametric model. On the basis of the parametric (prior) model, they then extend the procedure to estimate covariances between monitored and unmonitored locations. This approach, however, does not appear to extend to estimation of covariances for pairs of unmonitored locations. Such covariance estimates are necessary for some monitoring network design problems (Caselton and Zidek 1984; Guttorp et al. 1992).

4.3 Extensions and Applications

Although our title refers to spatial covariances, we have, in fact, restricted ourselves to modeling spatial dispersions. To estimate spatial covariances we must combine our estimated dispersion model with site variances. This is a simple matter for processes with constant variance. (Note that nonstationary or heterogeneous dispersion structure may very well occur with spatially stationary or homogeneous variances.) If the variance process itself is also heterogeneous, then the natural approach would be to standardize the data to unit variance and hence analyze the dispersion structure as determined from the spatial correlation matrix rather than the covariance matrix. The estimated dispersion model can then be combined with a smooth spatial model for the variance to estimate spatial covariances.

This article is concerned only with modeling and estimation of spatial covariance or spatial dispersion. These estimates may be applied directly to the analysis of network design questions using an information-theoretic approach, like that of Caselton and Zidek (1984) and Guttorp et al. (1992), or the more common approach aiming to minimize mean squared prediction errors [cf. Hughes and Lettenmaier (1981) or Olea (1984)]. On the other hand, spatial estimation or interpolation will also require a model for the mean field $\mu(x)$, which has not been considered here. This may be addressed in a variety of ways. For example, one may use an interpolation or spatial smoothing of the observed station means (Loader and Switzer 1992) or consider the mean as a random spatial process as in Eynon and Switzer (1983) or Egbert and Lettenmaier (1986). A promising approach for the analysis of acid deposition is to specify the expected value of the process using an empirical long-range transport model, as demonstrated by Venkatram (1988).

APPENDIX

To fit a general isotropic dispersion function of the form given in (2.1),

$$g(|h|; \mu) = \int_{-\infty}^{\infty} (1 - \exp(-t^2|h|^2))\, d\mu(t),$$

where $\mu(t)$ is a finite measure, we apply an algorithm discussed by Böhning (1982). We want to minimize the functional $T(\mu)$ given by

$$T(\mu) = \sum_{i<j} (d_{ij}^2 - g(|h_{ij}|; \mu))^2, \tag{A.1}$$

that is, compute a least squares fit of the observed dispersions d_{ij}^2 to the dispersion function evaluated at D plane distances $|h_{ij}|$. A weighted least squares analysis could be carried out similarly, if that is thought more appropriate.

To minimize (A.1), we choose a discrete measure μ, supported on points $t_1, \ldots, t_k$, with corresponding masses $\alpha_1, \ldots, \alpha_k$, where $k \le n/2$ and $n = N(N-1)/2$, the number of station pairs. Starting with a distribution μ_1 that puts mass 1 at t_1 and has $T(\mu_1) < \infty$, we select the next point in the support of μ by minimizing the directional derivative of the functional (A.1) in the direction of point mass at t, denoted μ_t. This is defined as

$$\lim_{\epsilon \to 0} \frac{1}{\epsilon} (T((1-\epsilon)\mu + \epsilon\mu_t) - T(\mu)),$$

which simplifies to

$$-2 \sum_{i<j} (g(|h_{ij}|; \mu_t) - g(|h_i|; \mu))(d_{ij} - g(|h_{ij}|; \mu)).$$

This derivative is minimized at the solution to

$$\sum_{i<j} |h_{ij}|^2 \exp(-|h_i|^2 t^2)(d_{ij} - g(|h_{ij}|; \mu)) = 0.$$

Having added this value of t to the support of μ, we next minimize $T(\mu)$ as a function of the masses α_i constrained to be nonnegative. We use a Newton-type nonlinear optimization program to solve this problem, enforcing the nonnegativity constraints by reexpressing the parameters α_i as $\exp(a_i)$, with the a_i unconstrained. We now iterate the whole procedure, choosing first a new t, and then a new set of α_i, until either we have $[n/2]$ points in the support, or the directional derivative for adding a new point to the support is less than a prespecified amount. The results of Böhning (1982) can be used to show that this procedure converges to the minimum value, and the argument in Simar (1976) shows that the minimum is achieved at a distribution with discrete support. Jewell (1982) discussed fitting of the subclass of mixtures of exponential functions, which yields the class of com-

pletely monotone isotropic dispersion functions, from much the same point of view.

We can include a nugget effect by replacing $T(\mu)$ in (A.1) by

$$T^*(\mu) = \sum \left(d_{ij}^2 - \alpha_0 - \sum_k \alpha_k g(|h_{ij}|; \mu_{t_k}) \right).$$

[*Received April 1990. Revised June 1991.*]

REFERENCES

Armstrong, M., and Jabin, R. (1981), "Variogram Models Must Be Positive-Definite," *Mathematical Geology*, 13, 455–459.

Armstrong, M., and Diamond, P. (1984), "Testing Variograms for Positive-Definiteness," *Mathematical Geology*, 16, 407–421.

Berndtsson, R. (1988), "Temporal Variability in Spatial Correlation of Daily Rainfall," *Water Resources Research*, 24, 1511–1517.

Böhning, D. (1982), "Convergence of Simar's Algorithm for Finding the Maximum Likelihood Estimate of a Compound Poisson Process," *The Annals of Statistics*, 10, 1006–1008.

Bookstein, F. L. (1978a), *The Measurement of Biological Shape and Shape Change* (Lecture Notes in Biomathematics Vol. 24), New York: Springer-Verlag.

——— (1978b), "Linear Machinery for Morphological Distortion," *Computers in Biomedical Research*, 11, 435–458.

——— (1987), "Principal Warps: Thin-Plate Splines and the Decomposition of Deformations." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11, 567–585.

Braud, I. (1990), "Etude Methodologique de l'Analyze en Composantes Principales de Processus Bidimensionels. Effets des Approximations Numeriques et de l'Echantillonnage et Utilisation Pour la Simulation de Champs Aleatoires. Application au Traitement des Temperatures Mensuelles de Surface de la Mer sur l'Atlantique Intertropical," unpublished doctoral thesis, l'Institut National Polytechnique de Grenoble, France.

Buja, A. (1982), "Multidimensional Scaling in Real-Time Implementation: A Movie" (approx. 30 min), Stanford University, Dept. of Statistics.

Caselton, W. F., and Zidek, J. V. (1984), "Optimal Monitoring Network Designs," *Statistics and Probability Letters*, 2, 223–227.

Chami, H., and Gonzalez, P. L. (1984), "Amélioration d'un Réseau de Surveillance de la Pollution Atmosphérique," Montpellier Cedex, Unité de Biométrie.

Cressie, N. (1985), "Fitting Variogram Models by Weighted Least Squares," *Mathematical Geology*, 17, 565–586.

Creutin, J. D., and Obled, C. (1982), "Objective Analyses and Mapping Techniques for Rainfall Fields: An Objective Comparison," *Water Resources Research*, 18, 413–431.

Duffie, J. A., and Beckman, W. A. (1974), *Solar Energy Thermal Processes*, New York: John Wiley.

Dunn, M. R. (1983), "A Simple Sufficient Condition for a Variogram Model to Yield Positive Variances Under Restrictions," *Mathematical Geology*, 15, 553–564.

Egbert, G. D., and Lettenmaier, D. P. (1986), "Stochastic Modeling of the Space–Time Structure of Atmospheric Chemical Deposition," *Water Resources Research*, 22, 165–179.

Escoufier, Y,. Camps, R., and Gonzalez, P. L. (1984), "Bilan et Perspectives dans l'Approche Statistique de la Constitution d'un Réseau d'Alerte à la Pollution Atmosphérique," Matapli 3, Montpellier Cedex, Unité de Biométrie.

Eynon, B. P., and Switzer, P. (1983), "The Variability of Rainfall Acidity," *Canadian Journal of Statistics*, 11, 11–24.

Greenacre, M. J., and Underhill, L. G. (1982), "Scaling a Data Matrix in a Low-Dimensional Euclidean Space," in *Topics in Applied Multivariate Analysis*, ed. D. M. Hawkins, New York: Cambridge University Press, pp. 183–268.

Guttorp, P., and Sampson, P. D. (1989), "Discussion of Haslett and Raftery," *Journal of the Royal Statistical Society*, Ser. C, 38, 32–34.

Guttorp, P., Sampson, P. D., and Newman, K. (1992), "Nonparametric Estimation of Spatial Covariance With Application to Monitoring Network Evaluation," in *Statistics in Environmental and Earth Sciences*, eds. P. Guttorp and A. Walden, London: Charles W. Griffin, pp. 39–57.

Haslett, J., and Raftery, A. E. (1989), "Space–Time Modelling With Long-Memory Dependence: Assessing Ireland's Wind Resource" (with discussion), *Journal of the Royal Statistical Society*, Ser. C, 38, 1–50.

Hay, J. E. (1983), "Solar Energy System Design: The Impact of Mesoscale Variations in Solar Radiation," *Atmosphere-Ocean*, 21, 138–157.

——— (1984), "An Assessment of the Mesoscale Variability of Solar Radiation at the Earth's Surface," *Solar Energy*, 32, 425–434.

Hughes, J. P., and Lettenmaier, D. P. (1981), "Data Requirements for Kriging: Estimation and Network Design," *Water Resources Research*, 17, 1641–1650.

Jewell, N. P. (1982), "Mixtures of exponential distributions," *The Annals of Statistics*, 10, 479–484.

Journel, A. G., and Huijbregts, C. J. (1978), *Mining Geostatistics*, New York: Academic Press.

Loader, C., and Switzer, P. (1992), "Spatial Covariance Estimation for Monitoring Data," in *Statistics in Environmental and Earth Sciences*, eds. P. Guttorp and A. Walden, London: Charles W. Griffin, pp. 52–69.

MAP3S/RAINE Research Community (1982), "The MAP3S/RAINE Precipitation Chemistry Network: Statistical Overview for the Period 1976–1979," *Atmospheric Environment*, 16, 1603–1631.

Mardia, K. V., Kent, J. T., and Bibby, J. M. (1979), *Multivariate Analysis*, New York: Academic Press.

Matérn, B. (1986), *Spatial Variation* (2nd ed; Lecture Notes in Statistics Vol. 36), New York: Springer-Verlag.

Matheron, G. (1971), "The Theory of Regionalized Variables and its Applications," Fascicule 5, Les Cahiers du Centre de Morphologie Mathematique de Fontainebleau.

Meinguet, J. (1979), "Multivariate Interpolation at Arbitrary Points Made Simple," *Journal of Applied Mathematics and Physics* (ZAMP), 30, 292–304.

——— (1984), "Surface Spline Interpolation: Basic Theory and Computational Aspects," in *Approximation Theory and Spline Functions*, eds. S. P. Singh, J. W. H. Burry, and B. Watson, Dordrecht: D. Reidel, pp. 97–126.

Myers, D. E. (1984), Letter to the Editor, *Mathematical Geology*, 16, 431–434.

——— (1988), "Interpolation With Positive Definite Functions," *Etudes Géostatistiques* V (Sciences de la Terre, Séries Informatique, Nancy), 28, 251–265.

National Atmospheric Deposition Program (1980), "National Atmospheric Deposition Program Data Report, Vol. II," NADP Coordinator's Office, Natural Resource Ecology Laboratory, Colorado State University.

Obled, C., and Creutin, J. D. (1986), "Some Developments in the Use of Empirical Orthogonal Functions for Mapping Meteorological Fields," *Journal of Climate and Applied Meteorology*, 25, 1189–1204.

Olea, R. A. (1984), "Sampling Design Optimization for Spatial Functions," *Mathematical Geology*, 16, 369–392.

Ripley, B. D. (1981), *Spatial Statistics*, New York: John Wiley.

Schoenberg, I. J. (1938), "Metric Spaces and Completely Monotone Functions," *The Annals of Mathematics*, 79, 811–841.

Sibson, R. (1981), "A Brief Description of Natural Neighbour Interpolation," in *Interpreting Multivariate Data*, ed. V. Barnett, New York: John Wiley, pp. 21–36.

Simar, L. (1976), "Maximum Likelihood Estimation of a Compound Poisson Process," *The Annals of Statistics*, 4, 1200–1209.

Suckling, P. W., and Hay, J. E. (1978), "On the Use of Synoptic Weather Map Typing to Define Solar Radiation Regimes," *Monthly Weather Review*, 106, 1521–1531.

Switzer, P. (1989), "Non-Stationary Spatial Covariances Estimated from Monitoring Data," in *Geostatistics*, ed. M. Armstrong, Amsterdam: Kluwer Academic Publishers, pp. 127–138.

Vecchia, A. V. (1985), "A General Class of Models for Stationary Two-Dimensional Random Processes," *Biometrika*, 72, 281–291.

Venkatram, A. (1988), "On the Use of Kriging in the Spatial Analysis of Acid Precipitation Data," *Atmospheric Environment*, 22, 1963–1975.

Wahba, G., and Wendelberger, J. (1980), "Some New Mathematical Methods for Variational Objective Analysis Using Splines and Cross Validation," *Monthly Weather Review*, 108, 36–57.

Watson, G. S. (1984), "Smoothing and Interpolation by Kriging and With Splines," *Journal of the International Association for Mathematical Geology*, 16, 601–615.

Weerahandi, S., and Zidek, J. V. (1988), "Bayesian Nonparametric Smoothers," *Canadian Journal of Statistics*, 16, 61–74.

Whittle, P. (1963), "Stochastic Processes in Several Dimensions," *Bulletin of the International Statistical Institute*, 40, 974–985.