Kriging Nonstationary Data

NOEL CRESSIE*

Spatial data modeled to have come from a random function with a nonstationary mean are considered. The spatial prediction method known as kriging exploits second-order spatial correlation structure to obtain minimum variance predictions of certain average values of the random function. But to do so, it must be assumed that either the mean function (the drift) is known up to a constant or the second-order structure (the variogram) is known exactly. Knowledge of the drift allows the (stationary) variogram to be estimated and leads to ordinary kriging. Knowledge of the variogram allows the drift to be estimated and leads to universal kriging. More usually, neither is known. This article shows how median polish of gridded spatial data provides a resistant and relatively bias-free way of kriging in the presence of drift, yet yields results as good as the mathematically optimal (but operationally difficult) universal kriging. Comparisons are performed on two data sets.

KEY WORDS: Covariogram; Drift; Resistance; Spatial statistics; Stationarity; Variogram.

1. GEOSTATISTICS

Geostatistics was developed to allow geological data spatially distributed throughout an ore body to be analyzed, the goal being ore-reserve prediction. Its chief proponent has been Matheron (1963, 1965, 1971, 1973), although Whittle (1963, chap. 4) presented some of the same elements of the underlying model and the little-known work of Gandin (1963) in meteorology parallels much of the geostatistics methodology. The basic idea is that evaluations of ore grade at nearby points are (usually positively) correlated, so inferences about mean ore grade over a fixed block B, made from neighboring observations, should reflect these spatial relationships. Kriging (here used as a noun; it is used as a verb in the title) is the technique developed by Matheron to do this [the name comes from Matheron's ascription of the idea to Krige (1951)]. Part of the reason why geostatistics has taken some time to become known is that much of the earlier work was written in French. Since the publication of David (1977) and Journel and Huijbregts (1978), texts written in English, its use is becoming considerably more widespread. Ripley (1981, chap. 4) has a discussion of the various techniques of smoothing and interpolation, including kriging.

The presence of positive correlations, for good physical reasons, is the source of the success geostatistics has enjoyed in the highly competitive and profit-minded mining industry. Potential applications abound, however; for example, in rainfall data (Ord and Rees 1979), in soil mapping (Burgess and Webster 1980), and in regional groundwater geochemical data (Myers, Begovitch, Butz, and Kane 1982). Although the data analyzed in this article are the results of ore-exploration drilling, the techniques have been successfully applied to tillage trials (Cressie and Horton 1986; Hamlett, Horton, and Cressie, in press), crop yields (Cressie 1985a), and public health data (Cressie and Read 1986).

The basic model, called the *intrinsic hypothesis* by Matheron, essentially assumes stationarity of differences. Let $\{Z_t; t \in D\}$ be a real-valued stochastic process defined over a domain D of $\mathbf{R}^2$ or $\mathbf{R}^3$; Matheron calls $\{Z_t\}$ a *regionalized variable*. Note that the index t is a vector and is used throughout as an index over *space*. Then the intrinsic hypothesis is

$$E(Z_{t+h} - Z_t) = 0, \tag{1.1}$$

$$\text{var}(Z_{t+h} - Z_t) = 2\gamma(h), \qquad t, t + h \in D. \tag{1.2}$$

The quantity $2\gamma(h)$ is known as the *variogram* and depends only on the relative position of the two variables Z_{t+h}, Z_t. Its effective estimation and use in kriging (prediction) make up much of the practice of geostatistics (see Sec. 2).

The implication of (1.1) is that $E(Z_t) = m$, a constant, although basing prediction on (1.2) finesses the necessity for first estimating m before estimating the spatial relationship expressed by $2\gamma(h)$. A special case is when $D \subset \mathbf{R}$ and Z_t is observed at equally spaced time points. Here many of the modeling and inference problems have already been solved by time series techniques (see, e.g., Fuller 1976). Typically, second-order stationarity is assumed; that is, the covariance and correlation are functions of h only. Generalizations of smoothing, differencing, and autoregressive moving average time series models to the spatial context have not been all that successful (Martin 1974; Ord 1975). In space, the index t often varies continuously, and data may be at irregularly located $\{t_i\}$.

In general, define *second-order stationarity* to be (1.1) and

$$C(h) = \text{cov}(Z_{t+h}, Z_t), \qquad t, t + h \in D, \tag{1.3}$$

where C depends only on the relative position of the two variables Z_{t+h}, Z_t. But it is possible for (1.2) to hold when (1.3) does not (take, e.g., one-dimensional Brownian motion, which does not have a stationary covariance function but for which a variogram *is* defined). When (1.3) holds,

* Noel Cressie is Professor, Department of Statistics, Iowa State University, Ames, IA 50011. This article owes a lot to the ideas of J. W. Tukey, and the author would like to wish him well for his "retirement." Comments from referees and the editor have led to an improved presentation. Thanks go to Gary Glonek, who gave computational assistance. This material is based on work partially supported by National Science Foundation Grant DMS-8503693 and by The Flinders University of South Australia.

Journal of the American Statistical Association
September 1986, Vol. 81, No. 395, Applications

$2\gamma(h) = 2(C(0) - C(h))$. It is worth noting here that covariance functions and variograms must satisfy certain positive-definiteness conditions to avoid possibly negative prediction variances (Christakos 1984).

In practice, I have found that the most often violated assumption is (1.1); in other words, it is highly unrealistic to assume constant mean. Data virtually never exhibit stationarity. In this article I give robust/resistant techniques that deal with this nonstationarity. Other techniques are available depending on the goals of the data analyst. Apart from the various geostatistical techniques described in Section 2, *spline interpolation* and the *NN method* are worth mentioning.

The objective of the spline method is to estimate globally a curve or surface, the smoothness of which is prespecified, from data that might or might not be "noisy" (DeBoor, 1978; Wahba 1975). Here the smooth surface X, say, typically belongs to a Sobolev space, has zero mean and covariance $E(X_s \cdot X_t) = bR(s, t)$, where $R(\cdot, \cdot)$ is a known real-valued function of two variables and the noise is iid. Kriging, on the other hand, tries to predict *locally* the value of Z at a point t or averaged over a small region B. It assumes stationarity of Z for practical reasons, namely to produce a workable solution to the prediction problem. Setting the differing goals of the two techniques aside, it has been shown by Kimeldorf and Wahba (1971) that smoothing by splines is equivalent to kriging with a certain (generalized) covariance function. See Matheron (1980) and Watson (1984) for further comparisons.

The NN (nearest neighbor) method (Wilkinson, Eckert, Hancock, and Mayo 1983) has been proposed for analyzing field experiments where treatment effects are superimposed onto trend (assumed locally quadratic) plus *independent* error. In the mining context there are no treatment effects, so for the purposes of this article, the NN method would be applied to data that are "homogeneous." The locally-quadratic-trend-plus-independent-error assumption, however, is far too severe for mining data, and the NN method will not be considered further.

After presenting the universal kriging solution to prediction in the presence of nonstationarity and showing how difficult it is to apply to data (Sec. 2), I develop an easy-to-use method called *median polish kriging* (Sec. 3). Section 4 compares the two methods on two data sets and shows little difference, endorsing the easy-to-use method. Conclusions are presented in Section 5.

2. KRIGING

2.1 Ordinary Kriging

Suppose that the intrinsic hypothesis of (1.1) and (1.2) holds. Kriging, in its original form, is simply a minimum variance method of predicting certain average values of the random function $\{Z_t; t \in D\}$ from a linear combination of observed values $\{Z_{t_i}; i = 1, \ldots, n\}$. With the stationary mean unknown in general, Matheron (1971) coined the term *ordinary kriging* to represent this "optimal" linear prediction procedure.

Let B be a volume in D; then

$$Z_B \equiv \int_B Z_t \, dt/|B| \tag{2.1}$$

is the average value of Z over B, where $|B|$ denotes the volume of B. The data $\{Z_{t_i}\}$ are used linearly to provide an ordinary kriging predictor (or estimator) of Z_B,

$$\hat{Z}_B = \sum_{i=1}^{n} \lambda_i Z_{t_i}, \tag{2.2}$$

where the weights $\{\lambda_i; i = 1, \ldots, n\}$ are chosen so that $\hat{Z}_B$ is unbiased ($\sum\lambda_i = 1$), and $\hat{Z}_B$ is minimum variance among all linear unbiased predictors. In other words, choose $\{\lambda_i\}$ to minimize the estimation error

$$\sigma_E^2 \equiv E\{(Z_B - \sum \lambda_i Z_{t_i})^2\}, \tag{2.3}$$

subject to

$$\sum \lambda_i = 1. \tag{2.4}$$

Expanding (2.3) gives

$$\sigma_E^2 = 2\sum \lambda_i \bar{\gamma}(t_i, B) - \bar{\gamma}(B, B) - \sum\sum \lambda_i\lambda_j\gamma(t_i - t_j),$$

where $\gamma(t_i - t_j)$ is given by (1.2), $\bar{\gamma}(B, B) = \int_B\int_B\gamma(u - v)du\,dv/|B|^2$, and $\bar{\gamma}(t_i, B) = \int_B\gamma(t_i - v)dv/|B|$. Differentiating with respect to $\{\lambda_i\}$ [with a Lagrange multiplier μ for condition (2.4)] results in the following linear system of equations:

$$\Gamma(\mathbf{t} \times \mathbf{t})\boldsymbol{\lambda} = \boldsymbol{\gamma}(\mathbf{t}, B), \tag{2.5}$$

where

$$\boldsymbol{\lambda} = (\lambda_1\lambda_2 \cdots \lambda_n\mu)',$$

$$\boldsymbol{\gamma}(\mathbf{t}, B) = (\bar{\gamma}(t_1, B) \cdots \bar{\gamma}(t_n, B)1)',$$

$$\Gamma(\mathbf{t} \times \mathbf{t})_{ij} = \gamma(t_i - t_j), \quad i = 1, \ldots, n, \; j = 1, \ldots, n$$
$$= 1, \quad i = n + 1, \; j = 1, \ldots, n$$
$$= 0, \quad i = n + 1, \; j = n + 1,$$

and $\Gamma(\mathbf{t} \times \mathbf{t})$ is a *symmetric* $(n + 1) \times (n + 1)$ matrix. Thus

$$\boldsymbol{\lambda} = \Gamma^{-1}\boldsymbol{\gamma} \tag{2.6}$$

provides the optimal coefficients in (2.2), and the minimum value of the estimation error σ_E^2, called the *kriging variance*, is

$$\sigma_K^2 = \sum \lambda_i\bar{\gamma}(t_i, B) + \mu - \bar{\gamma}(B, B). \tag{2.7}$$

Equations (2.2), (2.6), and (2.7) define the kriging predictor and variance for predicting Z_B. If $\{Z_t; t \in D\}$ can be written as $Z_t = X_t + U_t$, where $\{U_t; t \in D\}$ is an independent process of iid, zero-mean random variables, then provided $|B| > 0$, it can easily be seen that $Z_B = X_B$. Thus prediction of Z_B is equivalent to prediction of X_B and the equations are unchanged for $|B| > 0$. The case of point kriging, where $B = \{t_0\}$ and $|B| = 0$, requires separate discussion. Think of the component $\{U_t; t \in D\}$ as being made up of both measurement error and microscale struc-

ture. Strictly speaking, it is the latter that should be called the *nugget effect*; common usage, however, embodies both into the nugget effect. Therein lies confusion for those doing and defining kriging (Yakowitz and Szidarovsky 1985). Throughout this article I use kriging to refer to prediction of (integrals of) the Z process; for some purposes, however, it may be required to predict a less noisy version of Z_{t_0} (e.g., without the measurement error component). Then it is easy to see that when t_0 is not one of the sample locations $\{t_i;\ i = 1, \ldots, n\}$, the optimal predictor is $\sum \lambda_i Z_{t_i}$, where $\boldsymbol{\lambda}$ is given by (2.6), and hence is unchanged. [The kriging variance (2.7) is reduced by the measurement error variance, however.]

When $B = \{t_0\}$ is one of the sample locations, say t_1, then it is often touted that kriging is an exact interpolator. The way kriging is done in this article, it is, but it does not have to be. If a less noisy version of Z_{t_0} is predicted, (2.6) does change slightly; the entry $\bar{\gamma}(t_1, \{t_0\})$ in $\boldsymbol{\gamma}$ reflects the measurement error variance and is not zero, as it is when predicting Z_{t_0}. Exact interpolation is seen as both a good thing (by mining companies in joint ventures, who insist in their agreements that map contours of grade "honor" the data) and a bad thing (by, e.g., soil scientists who are looking for a smooth representation, over a large area, of a particular soil property that is sampled throughout the area).

There are a number of further observations to make. Solution (2.6) does not depend on any observed data value and so can also be used to answer design questions, such as where to take one more observation to maximize the reduction in σ_K^2. Although a large-order matrix inversion is needed in (2.6), if the sampling pattern is regular and the Z_B's of interest are regularly positioned throughout D, then the number of times different inversions are performed is small. Implicit in all of the above is that the variogram $2\gamma(\cdot)$ is known. In practice it is estimated, and there has been some recent research into how to do this efficiently but robustly (Armstrong and Delfiner 1980; Cressie 1979; Cressie and Hawkins 1980; Hawkins and Cressie 1984; Switzer 1984). What effect this estimation has on kriged values and kriging variances is at present unknown. Diamond and Armstrong's (1984) results, and the many successful applications of kriging in the mining industry, offer encouragement that there is only a "small-order" effect to worry about here.

The best predictor (without assuming linearity) is $E(Z_B|Z_{t_1}, \ldots, Z_{t_n})$. To compute this it is necessary to know joint $(n + 1)$-dimensional distributions, which in practice is not feasible. When $\{Z_t\}$ is a Gaussian process, the best predictor *is* a linear predictor. In what is to follow, I will assume that an appropriate transformation has been made that converts the problem into Gaussian data (with possible additive outliers, here modeled as heavy-tailed contamination in the stationary error distribution); see (2.11) and Mancey and Howarth (1980). Of course, prediction is needed on the original scale, and bias and variance corrections must be made when untransforming (Dowd 1982). When such a transformation is not obvious, a number of approaches have been proposed: disjunctive kriging (Matheron 1976), the multigaussian approach (Verly 1983), indicator kriging (Journel 1983), probability kriging (Sullivan 1984). The goal of this article is to study the more fundamental question of how to krige in the presence of trend-like *nonstationarity*.

2.2 Current Ways of Handling Nonstationarity

Suppose now that (1.1) of the intrinsic hypothesis is modified so that there is nonstationarity in the mean. Let

$$E(Z_t) = d(t); \tag{2.8}$$

$d(\cdot)$ is called the *drift*. The harder problem of how to work with nonstationarity in the variogram is outside the scope of this article; Cressie (1985b) considered scaling the variogram when nonstationarity is caused by a heterogeneous regionalized variable. There it is assumed that subregions of D can be found where at least (2.8) and (1.2) hold.

Two ways of kriging in the presence of nonconstant drift have been proposed. The first assumes that $d(\cdot)$ can be represented as a polynomial of finite order, whereas the second postulates that a certain finite-order difference of Z's is (weakly) stationary. An extensive discussion and comparison is given in Cressie and Laslett (1986); the main issues, however, are set out here.

The first method is called *universal kriging* (Matheron 1969) and is a simple extension of ordinary kriging. The order k of the polynomial and the variogram γ of the error must be known. Let $B = \{t_0\}$ for simplicity, so an unbiased predictor $\hat{Z}_{t_0} = \sum_{i=1}^n \lambda_i Z_{t_i}$, of Z_{t_0}, is desired. Suppose that f_l is a polynomial of order l $(l = 0, 1, \ldots, k)$. Then the unbiasedness condition becomes

$$E\left(\sum_{i=1}^n \lambda_i Z_{t_i}\right) = \sum_{l=0}^k a_l f_l(t_0),$$

and a set of sufficient conditions for this is

$$\sum_{i=1}^n \lambda_i f_l(t_i) = f_l(t_0), \qquad l = 0, 1, \ldots, k. \tag{2.9}$$

There are $(k + 1)$ Lagrange multipliers $\mu_0, \mu_1, \ldots, \mu_k$ for each of these conditions, yielding essentially the same equation as (2.6), namely

$$\boldsymbol{\lambda}_U = \boldsymbol{\Gamma}_U^{-1} \boldsymbol{\gamma}_U, \tag{2.10}$$

where now

$$\boldsymbol{\lambda}_U = (\lambda_1 \cdots \lambda_n \mu_0 \cdots \mu_k)',$$

$$\boldsymbol{\gamma}_U = (\gamma(t_1 - t_0) \cdots \gamma(t_n - t_0) f_0(t_0) \cdots f_k(t_0))',$$

$$\Gamma_U = \lambda(t_1 - t_j), \qquad i = 1, \ldots, n, \quad j = 1, \ldots, n,$$

$$= f_{i-n-1}(t_j), \qquad i = n + 1, \ldots, n + k + 1, \quad j = 1, \ldots, n,$$

$$= 0, \qquad i = n + 1, \ldots, n + k + 1, \quad j = n + 1, \ldots, n + k + 1,$$

and Γ_U is a *symmetric* $(n + k + 1) \times (n + k + 1)$ matrix.

There are two problems associated with the solution; k

is never known and has to be guessed, and γ is never known and has to be estimated from the *residuals* (= data − guessed drift). Matheron (1971) pointed out that such estimators are biased, and unless the variogram of errors is known, say from a calculation in a direction where there is no perceived drift, universal kriging has serious operational difficulties. It will be seen in Section 3 that a cognizance of the spatial nature of the data and a median-based analysis ameliorates the bias problem.

The second method of kriging, that of *intrinsic random functions of order k* (Matheron 1973), has a more general model assumption than the first, but in practice it reduces to guessing an order k and estimating a *generalized covariance function* from kth-order differences (Delfiner 1976). Although this method is more pleasing mathematically, its usefulness has been limited because (a) it has not generally been possible to estimate the generalized covariance nonparametrically (as for the variogram), (b) there is no easy interpretability of the parameters of the generalized covariance, (c) the order k of differences has to be guessed (usually $k = 2$ is used, in subregions of D), and (d) edge effects lead to a drastic reduction of those points that can be kriged this way. Nevertheless there is at least one geostatistics package, BLUEPACK (Delfiner, Renard, and Chiles 1978), which fits parametric generalized covariances to data. It has a tendency to declare the spatial process to be white noise when a graphical analysis indicates otherwise, and its parameter estimation appears to be biased (Cressie and Laslett 1986).

Section 3 introduces a resistant and relatively bias-free way of kriging in the presence of drift.

2.3 Gaussianity, Contamination, and Robust Kriging

Although it has already been stated that the most important problem to deal with is nonstationarity, I want to make a brief aside here to mention how one might handle outliers in a stationary data set. Outliers are typically not removed by a single transformation that transforms "the bulk" of the data to have a marginal (at least) Gaussian distribution.

These aberrant values may come from various sources. Hawkins and Cressie (1984) postulated a smooth underlying stationary Gaussian distribution $\{X_t; t \in D\}$, but assumed that a noisy version of this is observed, as follows:

$$Z_t = X_t + U_t, \qquad t \in D, \tag{2.11}$$

where $\{U_t; t \in D\}$ embody any measurement error plus any microscale variation. The contribution of $\{U_t\}$ is sometimes called the nugget effect (Journel and Huijbregts 1978, p. 39); the $\{U_t\}$ are assumed to have zero mean, to be independent of each other, and to have a *contaminated* Gaussian distribution. Hawkins and Cressie (1984) proposed a robust (to the contamination) kriging estimator that is *not* linear in a global sense. But locally, if all values in the kriging neighborhood are not "too different," then any predictor using *those* values would be linear in them. If, however, this cross-validation approach produces a value that is "too different" from its neighbors, then it is "brought in" to where it is not discordant with values in the kriging neighborhood. In the mining context, one would certainly *not* want to delete such points, since they may indicate that the company has "struck it rich." Robust kriging essentially formalizes what geostatisticians have in the past developed from common sense. The reason why I have briefly explained the approach is because it was necessary, on the coal ash data in Section 4, to krige robustly because of the presence of outliers.

Another way there may be departures from the Gaussian assumption is through small pockets of nonstationarity. Cressie (1984) presented data analytic tools to detect such pockets, but one would still like to construct an approach that is robust to them.

3. MEDIAN POLISH KRIGING

Anyone dealing with time series or spatial data knows that the decomposition of

observable surface

= large-scale variation + small-scale variation

cannot be specified uniquely, so in practice it is difficult actually to assign contributions to the different sources. What is one person's nonstationarity (in mean) may be another person's random (correlated) variation. Watson (1972) included a nice discussion of the usual decompositions and pointed out that most geological problems have a small-scale variation that should be modeled as a *stationary* zero mean stochastic process, with nearby points showing strong positive correlation.

The aim of kriging is to predict a value Z_B from data $\{Z_{t_i}; i = 1, \ldots, n\}$, exploiting the correlation between neighboring observations. We have seen in Section 2 two operationally unsatisfactory methods for dealing with the problem when $\{Z_t\}$ has nonconstant drift. The drift $d(t)$ is thought of as the large-scale variation, and the stationary error is thought of as the small-scale variation. My working approach is to let the spatial positions of the data determine the scale of variation. It is very much in the spirit of exploratory data analysis in that modeling decisions are based more on what is seen than on what might have been seen.

A surface could be fit exactly through the observed points, but there would be little predictive power in this approach. I want to extract the "apparent" drift and try to model the rest through weak stationarity; then the repeatability of the stationary part throughout space gives the method its predictive power. The spline approach is similar in that it allows the grid size to determine the large-scale variability, together with added conditions about differentiability and smoothness. With the type of applications I have in mind, however, it is the small-scale variability that I want to model, to interpret, and subsequently to exploit in kriging.

A data-analytic approach is taken; that is, I write

$$\text{data} = \text{fit} + \text{residual}. \tag{3.1}$$

Then "residual" itself is analyzed as a fresh data set to give

$$\text{residual} = \text{new fit} + \text{new residual}, \tag{3.2}$$

and so on. Questions about whether residuals are properly

representing the stationary error will be addressed subsequently.

Suppose that the spatial locations of the data are a subset of the points of a finite rectangular grid L. Not every point of L has to be the location of a datum, nor do the rows (or columns) have to contain equal numbers of observations. Spatial data are configured like this often enough for the problem to be an important one to analyze. Cressie and Read (1986) solved the problem of irregularly located data by assigning each datum to the nearest node of an overlaid grid.

Cressie (1984) suggested that the fit f_{ij} at location $t = (i, j)$ be obtained by median polish (Emerson and Hoaglin 1983; Tukey 1977) of the observed Z values. More specifically,

$$f_{ij} = a + r_i + c_j,$$
$$R_{ij} = Z_{ij} - f_{ij}, \tag{3.3}$$

where R_{ij} is the residual from median polish, having the property that $\text{med}_i\{R_{ij}\} = 0 = \text{med}_j\{R_{ij}\}$. And the row effects $\{r_i\}$ and column effects $\{c_j\}$ fitted by median polish are such that $\text{med}_i\{r_i\} = 0 = \text{med}_j\{c_j\}$.

This spatial analysis by rows and columns exploits "replication" down columns and across rows to estimate the large-scale variation, something that is not possible for time series. With just $r + c - 1$ df, a flexible estimator of the mean surface is defined. The "median" part of the median polish algorithm gives resistance to outliers (in particular there is almost no leakage of outliers to other parts of the table) and relatively bias-free residuals.

Median polish removal of trend is very grid-orientation dependent. Although experimenters often choose their grids in meaningful directions (e.g., where the geology dictates the exploration drilling), one should still check that the residuals do not contain "cross-product drift," or interaction, in the terminology of contingency tables. The simplest departure from (3.3) is through one extra parameter g, which accounts for a quadratic term in the fit (Scheffé 1959, p. 130); that is, fit $= a + r_i + c_j + g.(i - \text{ave}(i))(j - \text{ave}(j))$, where g is estimated using (3.2). In both examples considered in Section 4, I found g to be zero.

So, I propose a median polish of the gridded data, followed by a check for an interaction term, followed by the usual geostatistical analyses on the residuals (including robust analyses; see Hawkins and Cressie 1984). When kriging at the point t, say, the predictor is simply obtained by adding the estimated large-scale variation (the fit) to a predictor of the small-scale variation obtained by ordinary (robust) kriging (see Sec. 2.1).

I shall now indicate why in practice (see Sec. 4) good results are achieved with median polish followed by ordinary (robust) kriging. The large-scale variation estimation is median based, leading to resistant estimators and residuals where there is no "leakage" of outliers to other parts of the grid. Because there are no *linear* constraints such as $\text{mean}_i\{R_{ij}\} = 0 = \text{mean}_j\{R_{ij}\}$ on the residuals from median polish, it should follow that the variogram (or co-

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
24																
23										8.59	9.00	11.86	8.91	9.99		
22							11.62	10.91	8.76	8.89	9.10	7.62	9.65			
21					10.39	10.65	10.36	9.58	10.66	8.92	7.80	7.84	9.03	8.60		
20				9.79	9.06	10.70	11.21	8.98	9.27	8.19	7.88	7.61	8.20	8.77		
19				10.74	12.80	10.03	9.36	8.57	9.01	9.04	7.28	9.58	9.69	9.96	9.91	
18				11.21	9.89	10.34	8.20	9.82	10.06	8.58	8.89	8.64	7.04	8.81	7.95	
17			9.97	9.70	9.84	10.29	9.84	10.01	9.01	7.68	9.25	7.83	9.14		7.63	9.07
16	11.17	10.14	9.93	10.27	10.21	11.09	10.63	8.82	10.18	9.34	8.61					
15	9.92	10.82	11.65	8.96	9.88	8.90	10.18	9.34	10.56	9.06						
14	10.21	10.73	9.46	9.35	9.78	10.38	9.79	8.91	9.22	11.43						
13			12.50	9.63	10.82	10.12	9.40	9.48	10.99	9.92	7.85	8.21				
12		9.92	11.05	10.11	11.46	10.41	8.45	8.90	8.07	7.96	7.00	7.90				
11		11.31	9.41	9.37	11.21	9.93	10.70	9.27	9.28	10.13	8.61	8.78				
10		11.15	9.91	10.17	10.55	11.61	9.16	10.04	11.19	8.10	11.30					
9			10.82	11.75	9.78	11.00	9.79	10.19	9.15	8.15	9.20					
8		10.01	8.23	11.04	10.28	13.07	10.47	11.58	9.46	8.54	10.87					
7			10.39	11.11	10.96	10.83	10.09	8.69	11.17	9.39	9.56					
6			10.41	10.82	17.61	10.87		13.06	11.41	9.96	9.15					
5			9.76	11.10	10.80	8.86	9.48	9.22	9.61	8.20						
4			10.93	10.94	9.53	10.61	10.27	9.59	9.82	7.81						
3				9.64	9.52	10.06	12.65	9.63								
2				9.29	8.75	8.96	8.27	8.14								
1				10.59	10.43	9.32										

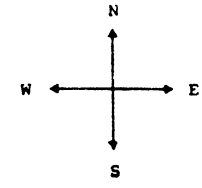

Figure 1. Original Coal Ash Data With Locations Indexed by Row and Column Numbers.

															ROW
									-0 20	0 00	3 29	0.00	0 51		0. 00
						1 35	1 15	-1 32	-0 00	0 00	-1 05	0.64			0. 10
				0. 10	0. 01	0 06	-0 20	0 56	0 00	-1 32	-0. 85	-0 00	-1 00		0. 12
			0. 01	-0. 40	0. 89	1. 74	0. 03	0 00	0 11	-0 41	-0 25	-0. 00	0. 00		-0. 71
			-0. 00	2 38	-0. 74	-1. 06	-1. 34	-1 22	0 00	-1 97	0 76	0 53	0 23	1 53	0. 25
			0. 90	-0. 10	0. 00	-1. 80	0. 34	0 26	-0 03	0 07	0 25	-1. 69	-0. 49	0. 00	-0. 18
		0. 27	-0. 46	0. 00	0. 10	-0. 00	0. 68	-0 64	-0. 78	0 58	-0. 41	0. 56		-0 17	0 00 -0. 33
0. 53	-0. 67	-0. 14	-0. 26	0. 00	0. 53	0. 42	-0. 88	0 16	0 51	-0 43					0. 04
-0. 54	0. 19	1. 77	-1. 39	-0. 15	-1. 48	0 15	-0. 18	0 72	0 41						-0. 14
0. 00	0. 35	-0. 18	-0. 75	-0. 00	0. 25	0. 00	-0. 36	-0 37	3 03						-0. 39
		2. 74	-0. 59	0. 92	-0. 13	-0. 50	0. 09	1. 28	1 40	-0 88	-0 09				-0. 27
	-0. 09	1. 78	0. 38	2 05	0. 65	-0. 96	0. 00	-1 15	-0 07	-1 24	0 09				-0. 76
	0. 93	-0. 23	-0. 73	1. 43	-0. 20	0 92	0. 00	-0 31	1 73	0 00	0 60				-0. 39
	0. 00	-0. 50	-0. 70	0 00	0. 71	-1. 39	0. 00	0 83	-1 07	1 92					0. 38
		0. 59	1. 06	-0 59	0. 28	-0. 58	0. 33	-1 03	-0 84	0 00					0. 20
	-0. 96	-2. 00	0. 35	-0. 09	2. 35	0 09	1. 72	-0 72	-0 45	1 67					0. 20
		-0 20	0. 06	0. 23	-0. 25	-0 64	-1 53	0 63	0 04	0 00					0. 56
		-0. 39	-0. 45	6. 66	-0. 42		2 62	0 66	0 39	-0 63					0. 78
		0. 14	1. 01	1. 03	-1. 25	-0 29	-0. 04	0 04	-0 19						-0. 40
		0 89	0. 44	-0. 65	0. 08	0. 08	-0 08	-0 17	-0 99						0. 01
			-0 39	-0 19	0 00	2 93	0 43								-0. 46
			0. 32	0. 10	-0. 04	-0 38	0 00								-1. 52
			0. 00	0. 16	-1 30										0. 10
0. 78	0. 95	0. 21	0. 67	0. 35	0. 70	0. 36	-0. 16	0. 16	-1. 03	-0. 82	-1. 25	-0. 91	-0. 34	-1. 69	-0. 42 9. 82
COLUMN															ALL

Figure 2. Results of Median Polish on Coal Ash Data. Residuals are shown in the main body of the figure.

variogram) analysis of residuals is less biased. Cressie and Glonek (1984) showed that a median-based analysis of time series data reduces the bias. This, together with recognition of the spatial configuration of the data, ameliorates considerably the first-order bias problems that hinder universal kriging (see Cressie 1984). It is important that the residuals behave like (2.11) so that a robust, efficient estimator of the variogram can be calculated. The regularity of the grid means that most of the time the data configurations used for kriging unknown Z_B or Z_{t_0} are balanced and remain the same throughout various parts of the domain. The final result is an accurate robust/resistant kriging estimator in the presence of nonstationarity.

In summary, *median polish kriging* (MPK) proceeds as follows. An observation at spatial location t is given by

$$Z_t = f_t + R_t, \tag{3.4}$$

where f_t is the drift estimated by interpolating planes between the median polish fit given by (3.3) and R_t is a regionalized variable (estimating the error). Suppose that Z_{t_0} is to be predicted (i.e., point kriging). Then MPK says

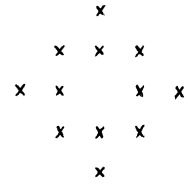

Figure 3. Kriging Configuration Used for Coal Ash Data.

to krige robustly [see (2.3)] to obtain a predictor $\tilde{R}_{t_0}$ of R_{t_0} and to add this back to the estimated drift f_{t_0}, giving the predictor

$$\tilde{Z}_{t_0} = f_{t_0} + \tilde{R}_{t_0}. \tag{3.5}$$

Because $\int_B Z_t\,dt = \int_B f_t\,dt + \int_B R_t\,dt$, the identical approach of kriging to obtain a predictor $\tilde{R}_B$ for R_B, and adding back, yields $\tilde{Z}_B$ the MPK predictor of Z_B, the average value of Z in a block B.

4. TWO SPATIAL DATA SETS

4.1 Coal Ash

Data on coal ash obtained from Gomez and Hazen (1970, tables 19 and 20) for the Robena Mine Property in Greene County, Pennsylvania, are analyzed from an exploratory point of view in Cressie (1984). The data (in %) and their spatial locations are repeated here in Figure 1; they are appropriately rotated so as to appear on a N–S, E–W grid (Cressie shows the original orientation); grid spacings are 2,500 feet by 2,500 feet. A median polish was performed, and the table of residuals is given in Figure 2. A "plus-one-fit" was tried, and no interaction structure was found in these residuals; a stem-and-leaf plot (Tukey 1977) confirmed an approximate Gaussian distribution apart from obvious and possible outliers. When the (robust) variogram estimator was computed it showed no correlation structure in either the N–S or the E–W directions; see Cressie [1985c, fig. 1(c)] and Cressie (1984, fig. 11). Using

```
 7
 7   8
 8   112
 8   5566667799
 9   00000222222223444444
 9   555555667777888888888999                                  (a)
10   000011111111111111222233444
10   55668888888
11   00
11


 7
 7   8
 8   0023
 8   5666788
 9   0011112222223344
 9   555666667777888999999                                     (b)
10   000000000000111122222222333344444
10   5555556666777789
11   002
11


-1
-1   001
-0   555555667778889
-0   00000001111112222222233333333444444444                    (c)
 0   0000000011111111222222222333344444444
 0   5555578
 1   1
 1
```

Figure 4. (a) Stem-and-Leaf Plot of Coal Ash Values Predicted From UK (mean = 9.75, variance = .48). (b) Stem-and-Leaf Plot of Coal Ash Values Predicted From MPK (mean = 9.86, variance = .54). (c) Stem-and-Leaf Plot of Difference Between UK and MPK Predictions (mean = −.10, variance = .20).

the fitting methods of Cressie (1985c), I obtained the semivariogram model

$$\gamma(h) = 0, \qquad h = 0$$
$$= .75, \qquad h \neq 0, \tag{4.1}$$

for the error. The self-scaling of median polish leads to the conclusion that here all of the "interesting" variation is in the large-scale model, leaving nothing but white noise in the small-scale model. This does not always happen, as is shown in the second example, on iron ore.

To demonstrate the predictive power of this robust/resistant approach to kriging in the presence of nonstationarity, I deleted one point at a time and then predicted it via (a) median polish kriging (MPK) and (b) universal kriging (UK). The thorny problems of UK (see Sec. 2.2) can be avoided for this data set by noting that in the N–S direction there is essentially no drift in the originals and in the E–W direction there seems locally to be at most quadratic drift (Cressie 1984, fig. 4). Thus the semivariogram fitted in the N–S direction in Cressie (1985c), namely

$$\gamma(h) = c_0 + c_s\{(3/2)(h/a_s) - (1/2)(h/a_s)^3\}, \quad 0 < h \leqq a_s$$
$$= c_0 + c_s, \qquad h \geqq a_s, \tag{4.2}$$

with $c_0 = .89$, $c_s = .14$, and $a_s = 4.31$ grid spacings, is appropriate for the UK equations (2.10). Figure 3 shows the configuration of data points that were used for UK (and for MPK) of the central deleted point.

I only predicted points for which Figure 3 could be realized, and the results given in Figure 4 were startling. UK can be thought of as the yardstick that, because there is no drift in the N–S direction, can be applied in this case. Figure 4(a) shows the stem-and-leaf plot of $\{\tilde{Z}_{t,\mathrm{UK}}\}$, and Figure 4(b) the stem-and-leaf plot of $\{\tilde{Z}_{t,\mathrm{MPK}}\}$. In terms of overall shape and sample mean and variance, there is very little difference between the two. Figure 4(c) shows a stem-and-leaf plot of $\{\tilde{Z}_{t,\mathrm{UK}} - \tilde{Z}_{t,\mathrm{MPK}}\}$, with tight cluster around 0. Finally, it turns out that the kriging variances are virtually the same; $\sigma_K^2 = 7.09$ for UK, and $\sigma_K^2 = 6.63$ for MPK. The small difference is easily accounted for by the estimation variance of f_t in (3.4).

4.2 Iron Ore

The original data have been modified slightly (for confidentiality reasons) by multiplying by and adding unspecified constants. Figure 5 shows the modified data (% Fe_2O_3)

```
                  60.8  54.8  58.9  59.1  57.4  59.1  57.9  54.4

                  57.8  58.1  57.8  56.1  58.7  58.7  57.3  58.1  51.5  54.2

          N       55.1  59.6  53.2  58.4  55.9  58.6  56.1  58.5  56.9  56.3  56.6  54.3  52.5  54.3  51.3
     W    E       54.3  51.9  52.0  56.7  57.8  50.3  58.2  58.3  59.0  58.9  59.7  52.9  52.8  57.1
          S       56.2  55.4  54.3  50.0  58.9  56.5  51.6  57.1  58.4  55.7  59.6  56.6  56.9  59.0

 45.6  46.9  53.5  52.1  53.5  55.7  52.2  47.2  54.1  57.6  58.7  58.6  59.1  56.1  56.7  55.0  56.4

 47.1  50.5  51.4  55.8  47.5  51.9  55.0  48.0  45.4  56.9  53.7  57.5  55.1  53.5  47.5  57.6  60.3

                  50.4  54.0  51.3  47.8  52.6  50.5  51.2  52.1  53.5  56.7  51.5  56.2  56.8  59.7

                                                    55.9  55.1  53.1
```

Figure 5. Iron Ore Data.

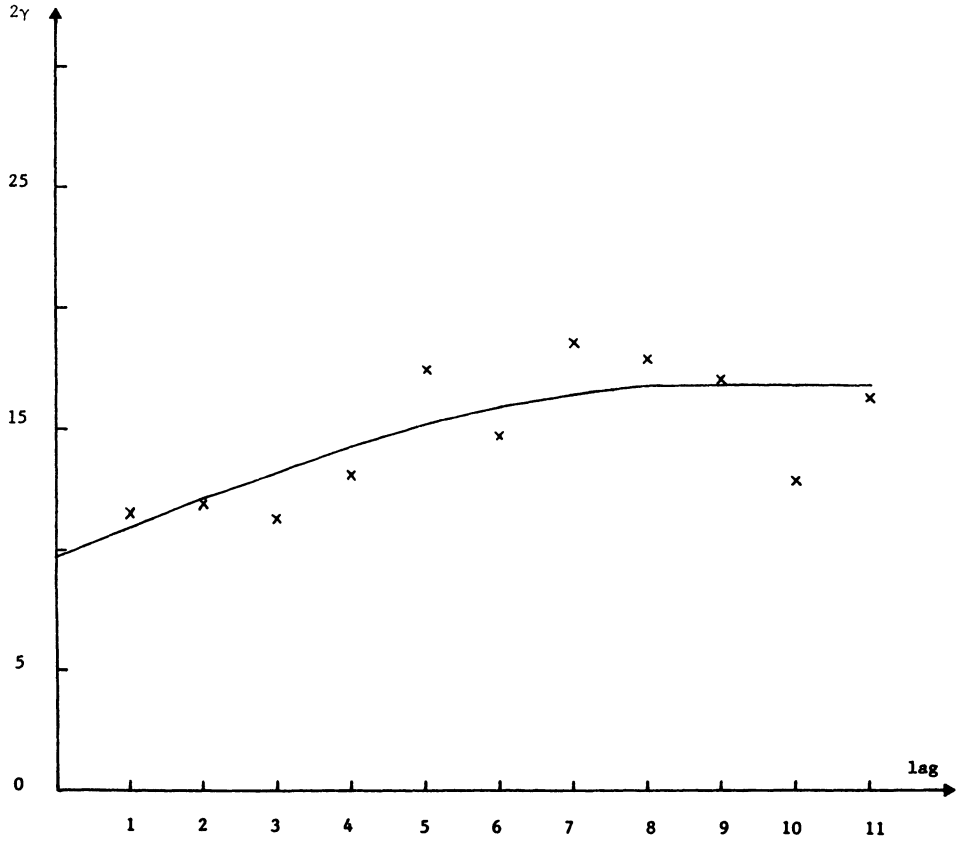

Figure 6. Robust Estimated Variogram of Iron Ore Residuals From Median Polish, E–W Direction. Weighted least squares fit of the spherical variogram is given by the solid line.

and their spatial locations; grid spacings are 50 meters by 50 meters. A strong linear trend in grades emerged in the N–S direction, with nothing perceptible in the E–W direction. A median polish gave *residuals* that appeared stationary and Gaussian and whose variogram in the E–W direction was modeled with a spherical form given by (4.2) with $c_0 = 4.83$, $c_s = 3.59$, and $a_s = 8.73$ grid spacings (Cressie 1985c); see Figure 6. Unlike the first example, this is *not* white noise; there is a large nugget effect superimposed onto positive correlation. After taking into account the N–S variogram, which showed roughly the same nugget effect c_0, roughly the same sill (the constant asymptote, which is equal to the stationary variance; see Journel and Huijbregts 1978, p. 37) $c_0 + c_s$, but *half* the range a_s, it was decided to use an *anisotropic* model. An equivalent way to do this is to double the scale in the N–S direction when computing kriging estimates (see Fig. 7) and use an isotropic model; here I used (4.4) with $c_0 = 5.4$, $c_s = 3.1$, and $a_s = 9$, obtained from both E–W and N–S fits. This gives a representation for the iron ore data in terms of (3.4); most important, there is still small-scale correlation structure left in the residuals.

The same UK versus MPK comparison as for coal ash was possible, since there is no drift in the E–W direction; the E–W variogram of the original data yielded the fit (Cressie 1985c) $\gamma(h) = 5.17 + 1.11h$. Figure 7 shows the configuration of data points that were used in the UK equations (2.10) (and in MPK).

I only predicted points for which Figure 7 could be realized; once again the results, given in Figure 8, were encouraging. Figures 8(a) and 8(b) show the stem-and-leaf plots of $\{\tilde{Z}_{t,\mathrm{UK}}\}$ and $\{\tilde{Z}_{t,\mathrm{UK}}\}$, respectively. Sample means, sample variances, and distribution shapes show very little difference. Figure 8(c) shows a stem-and-leaf plot of $\{\tilde{Z}_{t,\mathrm{UK}} - \tilde{Z}_{t,\mathrm{MPK}}\}$, with clustering around 0. Finally, it turns out that the kriging variances are almost the same; $\sigma_K^2 = 1.016$ for UK, and $\sigma_K^2 = .813$ for MPK. The small difference can be accounted for by the estimation variance of f_t in (3.4).

5. CONCLUSIONS

The mathematically sound approach of universal kriging runs into serious estimation problems when it is applied to data. In Section 4, I showed that median polish kriging (detailed in Sec. 3) is an alternative that can give nearly identical results, is operationally easy, and carries with it at every stage robust/resistant methods of estimation and prediction.

I have purposely not confused the discussion with anal-

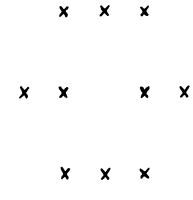

Figure 7. Kriging Configuration Used for Iron Ore Data. The scale in the N–S direction has been doubled to account for anisotropy.

```
50   55
51   49
52   22266
53   125577
54   1133
55   147                                  (a)
56   111266677799999
57   1366688889
58   113344447
59

50   669
51   067
52   568
53   12349
54   034666                               (b)
55   00599
56   0114566778
57   111112355679
58   245789
59   358

-3   23
-2   2389
-1   00149
-0   11122245556678999
 0   001156677799                         (c)
 1   0112233456689
 2   2
 3   2
 4   5
```

Figure 8. (a) Stem-and-Leaf Plot of Iron Ore Values Predicted From UK (mean = 55.76, variance = 5.42). (b) Stem-and-Leaf Plot of Iron Ore Values Predicted From MPK (mean = 55.67, variance = 5.79). (c) Stem-and-Leaf Plot of Differences Between UK and MPK Predictions (mean = .065, variance = 2.39).

ysis of data for which the bulk was not approximately Gaussian (or could be transformed to Gaussianity). Nonlinear geostatistical methods and the controversy that surrounds them would divert attention from the main ideas in this article.

I had good reason for presenting *two* examples. The first, a set of coal ash data, shows small-scale variation that can be modeled simply by white noise. The second example, a set of iron ore data, demonstrates that the small-scale variation in general does still possess spatial correlation. Data are given in the article, for those who wish to try their own spatial analyses. I attribute part of the lack of understanding of the geostatistical method to a cloak of confidentiality over source data for most published case studies.

Kriging is a "sturdy" procedure, in that it really only needs a *local* stationarity for it to be an optimal predictor. The problem is that the local variogram can never be estimated well from only a few points. I have tried to combine this sturdiness with a sturdy way of arriving at the local variability and the spatial relationship, namely median polish kriging on gridded data. This method is accurate, conceptually simple, operationally easy, and resists the effects of unusual observations.

[*Received December 1983. Revised November 1985.*]

REFERENCES

Armstrong, M., and Delfiner, P. (1980), "Towards a More Robust Variogram: A Case Study on Coal," Internal Note N-671, Fontainebleau, France: Centre de Geostatistique.

Burgess, T. M., and Webster, R. (1980), "Optimal Interpolation and Isarithmetic Mapping of Soil Properties, I: The Semi-variogram and Punctual Kriging," *Journal of Soil Science*, 31, 315–331.

Christakos, G. (1984), "On the Problem of Permissible Covariance and Variogram Models," *Water Resources Research*, 20, 251–265.

Cressie, N. A. C. (1979), "Straight Line Fitting and Variogram Estimation" (with discussion), *Bulletin of the International Statistical Institute*, 48, Book 3, 573–582.

——— (1984), "Towards Resistant Geostatistics," in *Geostatistics for Resources Characterization*, eds. G. Verly, M. David, A. G. Journel, and A. Marechal, Dordrecht: D. Reidel, pp. 21–44.

——— (1985a), "A Geostatistical Analysis of the Mercer and Hall Wheat Data," *Bulletin of the International Statistical Institute*, 51 (contributed papers vol.), 277–278.

——— (1985b), "When Are Relative Variograms Useful in Geostatistics?," *Journal of the International Association for Mathematical Geology*, 17, 693–702.

——— (1985c), "Fitting Variogram Models Using Weighted Least Squares," *Journal of the International Association for Mathematical Geology*, 17, 563–586.

Cressie, N. A. C., and Glonek, G. F. (1984), "Median Based Covariogram Estimators Reduce Bias," *Statistics and Probability Letters*, 2, 299–304.

Cressie, N. A. C., and Hawkins, D. M. (1980), "Robust Estimation of the Variogram: I," *Journal of the International Association for Mathematical Geology*, 12, 115–125.

Cressie, N. A. C., and Horton, R. (1986), "A Robust/Resistant Spatial Analysis of Soil-Water Infiltration," *Water Resources Research*, submitted for publication.

Cressie, N. A. C., and Laslett, G. M. (1986), "In Search of Generalized Covariances for Kriging," unpublished manuscript.

Cressie, N. A. C., and Read, T. R. C. (1986), "A Spatial Analysis of Sudden Infant Deaths," *Biometrics*, submitted for publication.

David, M. (1977), *Geostatistical Ore Reserve Estimation*, Amsterdam: Elsevier.

DeBoor, C. (1978), *A Practical Guide to Splines*, New York: Springer-Verlag.

Delfiner, P. (1976), "Linear Estimation of Nonstationary Phenomena," in *Advanced Geostatistics in the Mining Industry*, eds. M. Guarascio, M. David, and C. Huijbregts, Dordrecht: D. Reidel, pp. 49–68.

Delfiner, P., Renard, D., and Chiles, J. P. (1978), *BLUEPACK-3D Manual*, Fontainebleau, France: Centre de Geostatistique.

Diamond, P., and Armstrong, M. (1984), "Robustness of Variograms and Conditioning of Kriging Matrices," *Journal of the International Association for Mathematical Geology*, 16, 809–822.

Dowd, P. A. (1982), "Lognormal Kriging—The General Case," *Journal of the International Association for Mathematical Geology*, 14, 475–492.

Emerson, J. D., and Hoaglin, D. C. (1983), "Analysis of Two-Way Tables by Medians," in *Understanding Robust and Exploratory Data Analysis*, eds. D. C. Hoaglin, F. Mosteller, and I. W. Tukey, New York: John Wiley, pp. 166–210.

Fuller, W. A. (1976), *Introduction to Statistical Time Series*, New York: John Wiley.

Gandin, L. S. (1963), *Objective Analysis of Meteorological Fields*, Leningrad: Gidrometeorologicheskoe Izdatel'stvo (GIMIZ) (reprinted in 1965 by the Israel Program for Scientific Translations, Jerusalem).

Gomez, M., and Hazen, K. (1970), "Evaluating Sulfur and Ash Distribution in Coal Seams by Statistical Response Surface Regression Analysis," U.S. Bureau of Mines Report RI 7377.

Hamlett, J. M., Horton, R., and Cressie, N. A. C. (in press), "Resistant and Exploratory Techniques for Use in Semivariogram Analyses," *Soil Science Society of America Journal.*

Hawkins, D. M., and Cressie, N. A. C. (1984), "Robust Kriging—A Proposal," *Journal of the International Association for Mathematical Geology*, 16, 3–18.

Journel, A. G. (1983), "Nonparametric Estimation of Spatial Distributions," *Journal of the International Association for Mathematical Geology*, 15, 445–468.

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

Kimeldorf, G., and Wahba, G. (1971), "Some Results on Tchebycheffian

Spline Functions," *Journal of Mathematical Analysis and Its Applications*, 33, 82–95.
Krige, D. G. (1951), "A Statistical Approach to Some Mine Valuations and Allied Problems on the Witwatersrand," unpublished master's degree thesis, University of the Witwatersrand.
Mancey, S. J., and Howarth, R. J. (1980), "Power-Transform Removal of Skewness From Large Data Sets," *Transactions of the Institution of Mining and Metallurgy*, 89, Sec. B, B92–B97.
Martin, R. L. (1974), "On Spatial Dependence, Bias and the Use of First Spatial Differences in Regression Analysis," *Area*, 6, 185–194.
Matheron, G. (1963), "Principles of Geostatistics," *Economic Geology*, 58, 1246–1266.
——— (1965), *Les Variables Regionalisees et Leur Estimation*, Paris: Masson.
——— (1969), *Le Krigeage Universel* (Cahiers du Centre de Morphologie Mathematique, No. 1), Fontainebleau, France: Centre de Morphologie Mathematique.
——— (1971), *The Theory of Regionalized Variables and Its Applications* (Cahiers du Centre de Morphologie Mathematique, No. 5), Fontainebleau, France: Centre de Morphologie Mathematique.
——— (1973), "The Intrinsic Random Functions and Their Applications," *Advances in Applied Probability*, 5, 437–468.
——— (1976), "A Simple Substitute for Conditional Expectation: The Disjunctive Kriging," in *Advanced Geostatistics in the Mining Industry*, eds. M. Guarascio, M. David, and C. Huijbregts, Dordrecht: D. Reidel, pp. 221–236.
——— (1980), "Splines and Kriging: Their Formal Equivalence," Internal Note N-667, Fontainebleau, France: Centre de Geostatistique.
Myers, D. E., Begovitch, C. L., Butz, T. R., and Kane, V. E. (1982), "Variogram Models for Regional Groundwater Geochemical Data," *Journal of the International Association for Mathematical Geology*, 14, 629–644.
Ord, J. K. (1975), "Estimation Methods for Models of Spatial Interaction," *Journal of the American Statistical Association*, 70, 120–126.
Ord, J. K., and Rees, M. (1979), "Spatial Processes: Recent Developments With Applications to Hydrology," in *The Mathematics of Hydrology and Water Resources*, eds. E. H. Lloyd, T. O'Donnell, and J. C. Wilkinson, London: Academic Press, pp. 95–118.
Ripley, B. D. (1981), *Spatial Statistics*, New York: John Wiley.
Scheffé, H. (1959), *The Analysis of Variance*, New York: John Wiley.
Sullivan, J. (1984), "Conditional Recovery Estimation Through Probability Kriging—Theory and Practice," in *Geostatistics for Natural Resources Characterization*, eds. G. Verly, M. David, A. G. Journel, and A. Marechal, Dordrecht: D. Reidel, pp. 365–384.
Switzer, P. (1984), "Inference for Spatial Autocorrelation Functions," in *Geostatistics for Natural Resources Characterization*, eds. G. Verly, M. David, A. G. Journel, and A. Marechal, Dordrecht: D. Reidel, pp. 127–140.
Tukey, J. W. (1977), *Exploratory Data Analysis*, Reading, MA: Addison-Wesley.
Verly, G. (1983), "The Multigaussian Approach and Its Applications to the Estimation of Local Reserves," *Journal of the International Association for Mathematical Geology*, 15, 259–286.
Wahba, G. (1975), "Smoothing Noisy Data With Spline Functions," *Numerische Mathematik*, 24, 383–393.
Watson, G. S. (1972), "Trend Surface Analysis and Spatial Correlation," *Geological Society of America, Special Paper*, 146, 39–46.
——— (1984), "Smoothing and Interpolation by Kriging and With Splines," *Journal of the International Association for Mathematical Geology*, 16, 601–615.
Whittle, P. (1963), *Prediction and Regulation*, London: English Universities Press.
Wilkinson, G. N., Eckert, S. R., Hancock, T. W., and Mayo, O. (1983), "Nearest Neighbor (NN) Analysis of Field Experiments," *Journal of the Royal Statistical Society*, Ser. B, 45, 151–211.
Yakowitz, S. J., and Szidarovsky, F. (1985), "A Comparison of Kriging With Nonparametric Regression Methods," *Journal of Multivariate Analysis*, 16, 21–53.