Geostatistics

NOEL CRESSIE*

Most data have a space and time label associated with them; data that are close together are usually more correlated than those that are far apart. Prediction (or forecasting) of a process at a particular label where there is no datum, from observed nearby data, is the subject of this article. One approach, known as geostatistics, is featured, from which linear methods of spatial prediction (kriging) will be considered. Brief reference is made to other linear/nonlinear, stochastic/deterministic predictors. The (linear) geostatistical method is applied to piezometric-head data around a potential nuclear-waste repository site.

KEY WORDS: Anisotropy; Covariance function; Isotropy; Kriging; Spatial statistics; Stationarity; Trend; Variogram.

1. INTRODUCTION

Statistics, the science of uncertainty, attempts to find structure in chaos. The simplest structure imposed assumes that observations on a phenomenon are taken under identical conditions, and independently from one observation to another; that is, the data are independent and identically distributed (iid). Traditionally, this is generalized by allowing nonidentical means but retaining iid errors. Standard statistical techniques (e.g., see Hogg and Craig 1978) can then be used to build a statistical model and to estimate its parameters. But the iid assumption should not be taken for granted, particularly when there are good physical reasons to abandon it. It is the relaxation of the independence part of the assumption that I shall address in this article.

We have not yet been able to escape the three-dimensional world in which we live, nor the unidirectional flow of time through which we live. The notion that data close together, in time or space, are likely to be correlated (i.e., cannot be modeled as independent) is natural and has been used successfully by statisticians to model the processes generating the data. Pure temporal models, or time series models as they have come to be known, are usually based on identically distributed errors that are dependent and occur at equally spaced time points [e.g., the autoregressive moving average models of Box and Jenkins (1970)]. The equal spacing and unidirectional flow of time underlie the construction of these models.

Spatial models are a more recent addition to the statistics literature. Geology, soil science, crop science, forestry, astronomy, or simply any discipline that works with data collected from different spatial locations needs to develop (not necessarily statistical) models that indicate when there is dependence between measurements at different locations. The models need to be more flexible than their temporal counterparts; for example, it is not reasonable to assume that spatial locations of data occur regularly, and "past," "present," and "future" make way for (possibly anisotropic) dependence in a multitude of directions. Cliff and Ord (1981), Ripley (1981), and Upton and Fingleton (1985) have written books concerned with the statistical analysis of spatial data.

A scientific problem that is *both* temporal and spatial in nature is the study of the effects of atmospheric pollution, and in particular acid rain (e.g., see Bilonick 1985; Cressie, Gotway, and Grondona, in press). Daily data collection over a number of years at various locations throughout the northeast United States yields a massive data set. But most of it is temporal so that, spatially speaking, the data are still rather sparse. Nevertheless, spatial prediction is just as important as temporal prediction, since people living in those cities and rural districts without monitoring stations have the same right to know whether their water or their air is polluted. Purely spatial problems are considered here, and to provide a clear comparison between spatial methods and temporal methods a data set that does not have a temporal component will be chosen in Section 4.

Geostatistics is usually concerned with spatial prediction, but there are other important areas (such as stationary-distribution estimation, effect of aggregation, and spatial design) that offer fruitful open problems. These could all be lengthy articles in themselves; however, in this article I present the basics of a spatial-prediction method known as (ordinary) *kriging*. For a comparative survey of various forms of kriging with other stochastic and nonstochastic methods of spatial prediction, see Cressie (1989a), where 15 methods are considered. These include disjunctive kriging, Markov-random-field prediction, inverse-distance-squared weighted averaging, and spline smoothing.

2. THE VARIOGRAM

First, a measure of the (second-order) spatial dependence exhibited by the data is needed. I define a model-based parameter (which is a function) known as the variogram; then its estimate will provide such a measure. Statisticians are used to dealing with autocovariance functions. It will be demonstrated that the class of processes with variogram functions includes the class of processes with autocovariance functions; hence kriging can be carried out on a wider class of processes than the one traditionally used.

Let $\{Z(\mathbf{s}): \mathbf{s} \in D \subset \mathbf{R}^d\}$ be a real-valued stochastic process defined on a domain D of the d-dimensional space $\mathbf{R}^d$, and suppose that differences of variables lagged $\mathbf{h}$ apart vary in a way that depends only on $\mathbf{h}$. That is,

$$\text{var}(Z(\mathbf{s}+\mathbf{h}) - Z(\mathbf{s})) = 2\gamma(\mathbf{h}) \quad \text{for all } \mathbf{s}, \mathbf{s}+\mathbf{h} \in D; \tag{1}$$

*Noel Cressie is Professor, Department of Statistics, Iowa State University, Ames, IA 50011. Helpful discussions with I. Clark are gratefully acknowledged, as are the comments of the referees and the research assistance of C. Gotway. This work was partially supported by National Science Foundation Grant DMS-8703083.

typically the spatial index $\mathbf{s}$ is two- or three-dimensional (i.e., $d = 2$ or 3). The quantity $2\gamma(\cdot)$, which is a function only of the *difference* between the spatial locations $\mathbf{s}$ and $\mathbf{s} + \mathbf{h}$, has been called the *variogram* by Matheron (1963), although earlier appearances in the scientific literature can be found. It has been called a *structure function* by Yaglom (1957) in probability and Gandin (1963) in meteorology, and a *mean-squared difference* by Jowett (1952) in time series. Kolmogorov (1941a) introduced it in physics to study the local structure of turbulence in a fluid. Nevertheless, it has been Matheron's mining terminology that has persisted. The variogram must satisfy the conditional negative semi-definiteness condition, $\Sigma_{i=1}^{k} \Sigma_{j=1}^{k} a_i a_j 2\gamma(\mathbf{s}_i - \mathbf{s}_j) \leqq 0$, for any finite number of spatial locations $\{\mathbf{s}_i: i = 1, \ldots, k\}$, and real numbers $\{a_i: i = 1, \ldots, k\}$ satisfying $\Sigma_{i=1}^{k} a_i = 0$. When $2\gamma(\mathbf{h}) = 2\gamma^o(\|\mathbf{h}\|)$, the variogram is said to be *isotropic*. Various parametric models were presented by Journel and Huijbregts (1978, sec. III.B).

Replacing (1) with the stronger assumption

$$\operatorname{cov}(Z(\mathbf{s}+\mathbf{h}), Z(\mathbf{s})) = C(\mathbf{h}) \quad \text{for all } \mathbf{s}, \mathbf{s}+\mathbf{h} \in D, \tag{2}$$

and specifying the mean function to be constant,

$$E(Z(\mathbf{s})) = \mu \quad \text{for all } \mathbf{s} \in D, \tag{3}$$

defines the class of *second-order* (or wide-sense) *stationary* processes in D, with covariance function $C(\cdot)$. Time series analysts usually assume (2) and work with the quantity $\rho(\cdot) \equiv C(\cdot)/C(\mathbf{0})$. Conditions (1) and (3) define the class of *intrinsically stationary* processes, which is now shown to contain the class of second-order stationary processes.

Assuming only (2),

$$\gamma(\mathbf{h}) = C(\mathbf{0}) - C(\mathbf{h}); \tag{4}$$

that is, the *semivariogram* (half of the variogram) is related very simply to the covariance function. An example of a process for which $2\gamma(\cdot)$ exists but $C(\cdot)$ does not is a one-dimensional standard Wiener process $\{W(t): t \geqq 0\}$. Here $2\gamma(h) = |h|$ $(-\infty < h < \infty)$, but $\operatorname{cov}(W(t), W(u)) = \min(t, u)$, which is not a function of $|t-u|$. Thus the class of intrinsically stationary processes *strictly* contains the class of second-order stationary processes.

Now I turn to finding an estimator of the variogram, which also serves as a measure of spatial dependence among the data $\{Z(\mathbf{s}_i): i = 1, \ldots, n\}$. Suppose these are observations on an intrinsically stationary process [i.e., a process that satisfies (1) and (3)], taken at the n spatial locations $\{\mathbf{s}_i: i = 1, \ldots, n\}$. Because of (3), $\operatorname{var}(Z(\mathbf{s}+\mathbf{h}) - Z(\mathbf{s})) = E(Z(\mathbf{s}+\mathbf{h}) - Z(\mathbf{s}))^2$; hence the method-of-moments estimator of the variogram $2\gamma(\mathbf{h})$ is

$$2\hat{\gamma}(\mathbf{h}) = \sum_{N(\mathbf{h})} (Z(\mathbf{s}_i) - Z(\mathbf{s}_j))^2 / |N(\mathbf{h})|, \qquad \mathbf{h} \in \mathbf{R}^d, \tag{5}$$

where the average in (5) is taken over $N(\mathbf{h}) = \{(\mathbf{s}_i, \mathbf{s}_j): \mathbf{s}_i - \mathbf{s}_j = \mathbf{h}\}$ and $|N(\mathbf{h})|$ is the number of distinct elements in $N(\mathbf{h})$. For irregularly spaced data, $N(\mathbf{h})$ is usually modified to $\{(\mathbf{s}_i, \mathbf{s}_j): \mathbf{s}_i - \mathbf{s}_j \in T(\mathbf{h})\}$, where $T(\mathbf{h})$ is a tolerance region of $\mathbf{R}^d$ surrounding $\mathbf{h}$ (see Sec. 4). Other estimators, more robust than (5), are given in Cressie and Hawkins (1980). Parametric models can be fit to the estimator (5) by various means; as a compromise between efficiency and simplicity, Cressie (1985) proposes minimizing a weighted sum of squares,

$$\sum_{k=1}^{K} \left\{ \frac{2\hat{\gamma}(\mathbf{h}(k))}{2\gamma(\mathbf{h}(k); \boldsymbol{\theta})} - 1 \right\}^2 |N(\mathbf{h}(k))|,$$

with respect to variogram model parameters $\boldsymbol{\theta}$. The sequence $\mathbf{h}(1), \ldots, \mathbf{h}(K)$ denote the "lags" at which an estimator (5) was obtained, and that satisfy conditions such as those given by Journel and Huijbregts (1978, p. 194, sec. III.42). Zimmerman and Zimmerman (1989) summarize and compare several methods of variogram-parameter estimation. They find that no method dominates, but their simulation results indicate that the weighted least squares approach usually performs well and never does poorly.

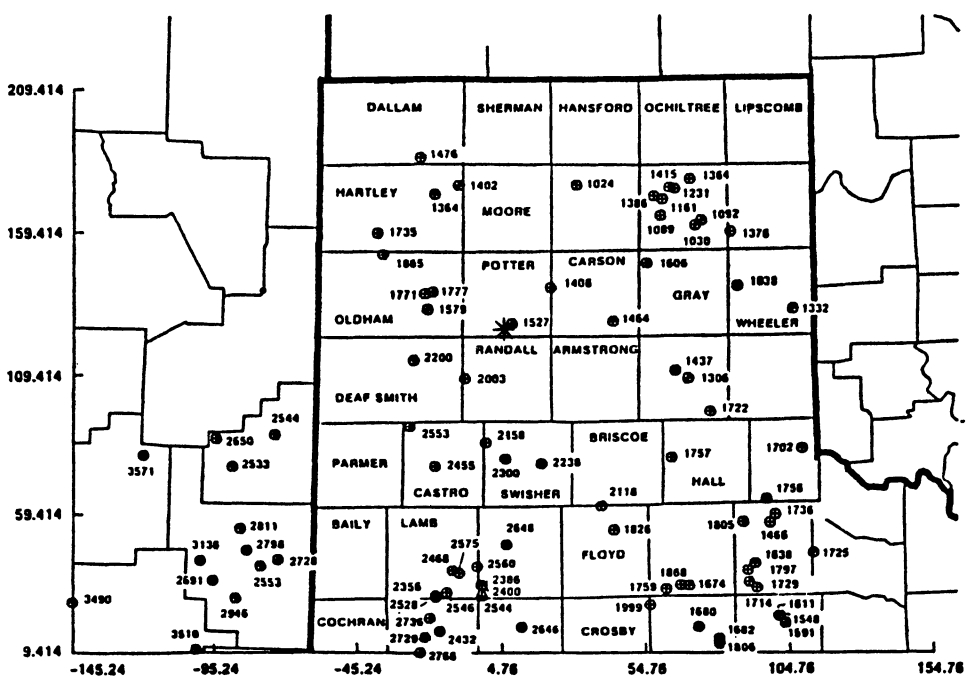

*Figure 1. Location and Levels of Piezometric-Head Data in the Wolfcamp Aquifer: *, Location of Amarillo, Texas (Potter County).*

Table 1. Wolfcamp-Aquifer Data

x	y	Z(x, y)	x	y	Z(x, y)
42.78275	127.62282	1,464	103.26625	20.34239	1,591
−27.39691	90.78732	2,553	−14.31073	31.26545	2,540
−1.16289	84.89600	2,158	−18.13447	30.18118	2,352
−18.61823	76.45199	2,455	−18.12151	29.53241	2,528
96.46549	64.58058	1,756	−9.88796	38.14483	2,575
108.56243	82.92325	1,702	−12.16336	39.11081	2,468
88.36356	56.45348	1,805	11.65754	18.73347	2,646
90.04213	39.25820	1,797	61.69122	32.49406	1,739
93.17269	33.05852	1,714	69.57896	33.80841	1,674
97.61099	56.27887	1,466	66.72205	33.93264	1,868
90.62946	35.08169	1,729	−36.65446	150.91456	1,865
92.55262	41.75238	1,638	−19.55102	137.78404	1,777
99.48996	59.15785	1,736	−21.29791	131.82542	1,579
−24.06744	184.76636	1,476	−22.36166	137.13680	1,771
−26.06285	114.07479	2,200	21.14719	139.26199	1,408
56.27842	26.84826	1,999	7.68461	126.83751	1,527
73.03881	18.88140	1,680	−8.33227	107.77691	2,003
80.26679	12.61593	1,806	56.70724	171.26443	1,386
80.23009	14.61795	1,682	59.00052	164.54863	1,089
68.83845	107.77423	1,306	68.96893	177.24820	1,384
76.39921	95.99380	1,722	70.90225	161.38136	1,030
64.46148	110.39641	1,437	73.00243	162.98960	1,092
43.39657	53.61499	1,828	59.66237	170.10544	1,161
39.07769	61.99805	2,118	61.87429	174.30178	1,415
112.80450	45.54766	1,725	63.70810	173.91453	1,231
54.25899	147.81987	1,606	5.62706	79.08730	2,300
6.13202	48.32772	2,648	18.24739	77.39191	2,238
−3.80469	40.40450	2,560	85.68824	139.81701	1,038
−2.23054	29.91113	2,544	105.07646	132.03181	1,332
−2.36177	33.82002	2,386	−101.64278	10.65106	3,510
−2.18890	33.68207	2,400	−145.23654	28.02333	3,490
63.22428	79.49924	1,757	−73.99313	87.97270	2,594
−10.77860	175.11346	1,402	−94.48182	86.62606	2,650
−18.98889	171.91694	1,364	−88.84983	76.70991	2,533
−38.57884	158.52742	1,735	−120.25898	80.76485	3,571
83.14496	159.11558	1,376	−86.02454	54.36334	2,811
−21.80248	15.02551	2,729	−72.79097	43.09215	2,728
−23.56457	9.41441	2,766	−100.17372	42.89881	3,136
−20.11299	22.09269	2,736	−78.83539	40.82141	2,553
−16.62654	17.25621	2,432	−83.69063	46.50482	2,798
29.90748	175.12875	1,024	−95.61661	35.82183	2,691
100.91568	22.97808	1,611	−87.55480	29.39267	2,946
101.29544	22.96385	1,548			

NOTE: The first two columns are data locations $\mathbf{s}_i = (x_i, y_i)$, in miles from an arbitrary origin; the third column is piezometric head $Z(\mathbf{s}_i)$, in feet above sea level. Here $i = 1, \ldots, 85$.

To derive the kriging equations in Section 3, I assume that the variogram is known. In practice the estimation and fitting stage of the spatial analysis has to be carried out first, and this I do for the example in Section 4. Discussion of the effect of this initial variogram fitting is given in Section 5.

3. KRIGING

In Matheron's geostatistics and Gandin's objective analysis, the variogram is used to define coefficients in an optimal linear predictor. Suppose it is desired to predict $Z(\mathbf{s}_0)$ at some unsampled spatial location $\mathbf{s}_0$ using a linear function of the data

$$\hat{Z}(\mathbf{s}_0) = \sum_{i=1}^{n} \lambda_i Z(\mathbf{s}_i), \tag{6}$$

which is unbiased and *minimizes* the mean squared prediction error $E(Z(\mathbf{s}_0) - \hat{Z}(\mathbf{s}_0))^2$. Then a straightforward minimization using Lagrange multipliers to ensure unbiasedness yields optimal λ's that satisfy an $(n+1)$-dimensional equation depending on $2\gamma(\mathbf{s}_i - \mathbf{s}_j)$, $i, j = 0, 1, \ldots, n$:

$$\Gamma \boldsymbol{\lambda} = \boldsymbol{\gamma}, \tag{7}$$

where $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_n, m)'$, $\boldsymbol{\gamma} = (\gamma(\mathbf{s}_1 - \mathbf{s}_0), \ldots, \gamma(\mathbf{s}_n - \mathbf{s}_0), 1)'$, and

$$\begin{aligned} \Gamma &= \gamma(\mathbf{s}_i - \mathbf{s}_j), && i = 1, \ldots, n, \quad j = 1, \ldots, n \\ &= 1, && i = n+1, \quad j = 1, \ldots, n \\ &= 0, && i = n+1, \quad j = n+1; \end{aligned}$$

Γ is a symmetric $(n+1) \times (n+1)$ matrix. The minimized root mean squared prediction error (sometimes called kriging standard error) is given by

$$\sigma_k(\mathbf{s}_0) = \left\{ \sum_{i=1}^{n} \lambda_i \gamma(\mathbf{s}_i - \mathbf{s}_0) + m \right\}^{1/2}. \tag{8}$$

Matheron (1963) called this spatial prediction method *kriging*, after D. G. Krige (a South African mining engineer who, in the 1950s, developed empirical methods for determining true ore grade distributions from distributions based on sampled ore grades), whereas Gandin (1963) called it *optimal interpolation*. In fact, derivation of an optimal linear predictor (with known mean μ) can be found earlier in the literature in works of Wold (1938), Kolmogorov (1941b), and Wiener (1949); see Cressie (1989b) for further historical

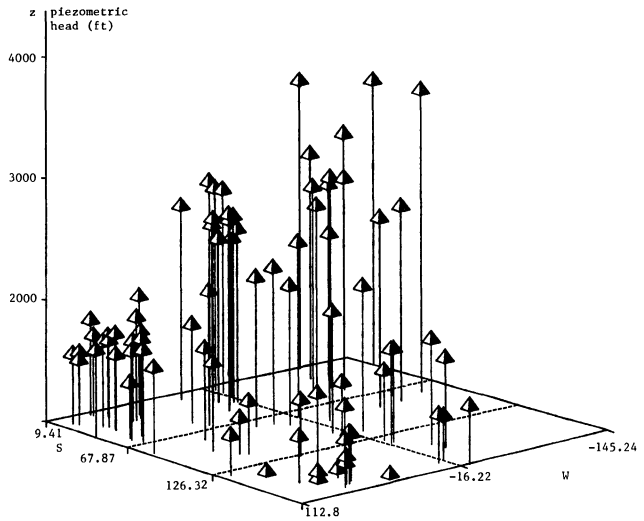

Figure 2. Scatterplot of Piezometric-Head Data. Verticle lines are sited at the spatial locations, and their heights are equal to the corresponding piezometric heads (in feet above sea level).

observations. The kriging predictor given by (6) and (7) assumes that the process Z contains no measurement error; Cressie (1988) has details on how to predict noiseless versions of Z.

4. FLOW OF GROUNDWATER FROM A NUCLEAR-WASTE SITE

Several years ago there were three potential high-level nuclear-waste sites proposed in Nevada, Texas, and the state of Washington. This section contains a geostatistical analysis of groundwater flow in the surrounding regions of one of those sites.

The chosen site will eventually contain more than 68,000 high-level waste canisters placed about 30 feet apart in holes or trenches surrounded by salt, covering an area of about two square miles, and buried deep in the ground. Leaks could occur, however, or the radioactive heat could cause the tiny quantities of water in the salt to migrate toward the heat until eventually each canister is surrounded by water (about six gallons). The chemical reaction of salt and water would create hydrochloric acid that could slowly corrode the canisters.

Using geostatistics, I address the question Where would radionuclide contamination flow for the site in Deaf Smith County, Texas? (As a matter of interest, a decision was made in December 1987 by the U.S. Congress to locate the site in Nevada, probably at Yucca Mountain.) From piezometric-head data for the Wolfcamp Aquifer, a variogram will be fit and kriging will be used to predict the head surface throughout a region in West Texas/New Mexico.

Figure 1 shows the 85 data and their locations (in principle, obtained by drilling a narrow pipe into the aquifer and letting the water find its own level in the pipe); the measurements are in units of feet above sea level. This figure is essentially that found in Harper and Furr (1986), which was my source for the piezometric-head data along with their spatial coordinates. The data are repeated here in Table 1, for completeness.

The proposed nuclear-waste site was located in Deaf Smith County, seen in Figure 1 to border with New Mexico, in the Texas panhandle. For country-music listeners who would like to get their bearings, Amarillo, Texas, is also shown on the map.

The goal in analyzing this (purely) spatial data set is to draw a map of a predicted surface, based on the (irregularly located) 85 data available. An advantage of using a stochastic method of prediction (kriging) is that a map of root mean squared prediction error can also be drawn, quantifying the uncertainty in the predicted piezometric-head surface.

In the notation of Section 3, it is desired to use the predictor $\hat{Z}(\mathbf{s}_0)$ defined by (6), along with $\sigma_k(\mathbf{s}_0)$ defined by (8), as $\mathbf{s}_0$ varies over the extent of the Wolfcamp Aquifer. The kriging equations (7) are straightforward to solve once the variogram (1) is known.

Assume that the data $\{Z(\mathbf{s}_1), \ldots, Z(\mathbf{s}_{85})\}$ given in Table 1 are observed without measurement error, and that they are a sampling from an intrinsically stationary stochastic process. [In fact, Harper and Furr (1986) did not make Assumption (3), assuming instead linear deterministic trend (see Sec. 5 for further discussion).] Estimation of the variogram defined by (1) allows (7) to be solved and hence maps based on (6) and (8) to be drawn.

It is clear from the data shown in Figure 2 that the behavior of the process in the NE–SW direction is different from that in the NW–SE direction. For this reason variogram estimators based on (5) were calculated in these two directions.

Figure 3a shows $2\hat{\gamma}(\mathbf{h})$ for $\mathbf{h} \in \{h(1)\mathbf{e}, \ldots, h(24)\mathbf{e}\}$; $\mathbf{e}$ is a vector of length $\|\mathbf{e}\| = 5$ miles, and direction $\pi/4$; $h(k) = k - \frac{1}{2}$ for $k = 1, \ldots, 24$. Figure 3b shows a similar plot, but $\mathbf{e}$ has direction $3\pi/4$. Since the data are

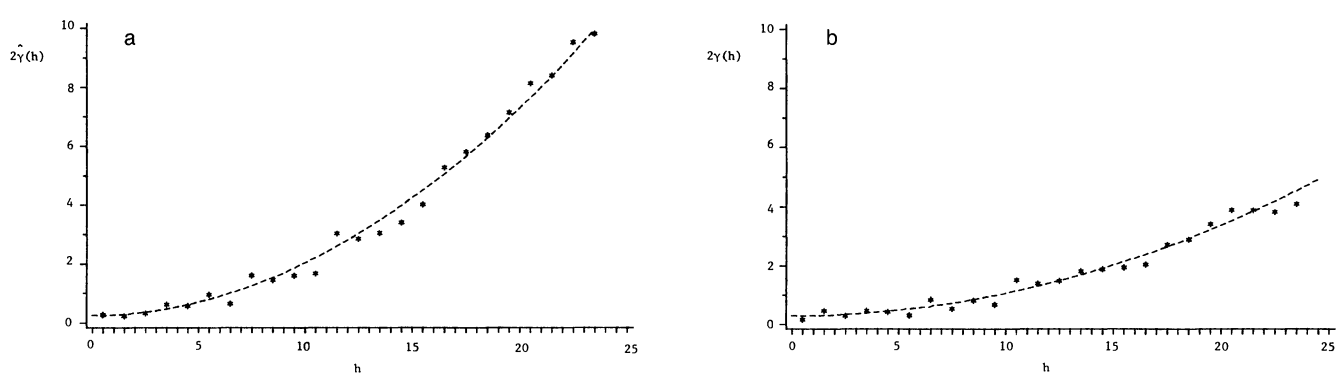

Figure 3. Experimental Variogram in the (a) NE–SW and (b) NW–SE Directions. The dashed line is the weighted-least-squares fit of a power model: $2\gamma(h; c_0, b, p) = 2\{c_0 + b|h|^p\}$ $(h > 0)$. On the horizontal axis, 1 unit = 5 miles; on the vertical axis, 1 unit = 10^5ft.^{2}.

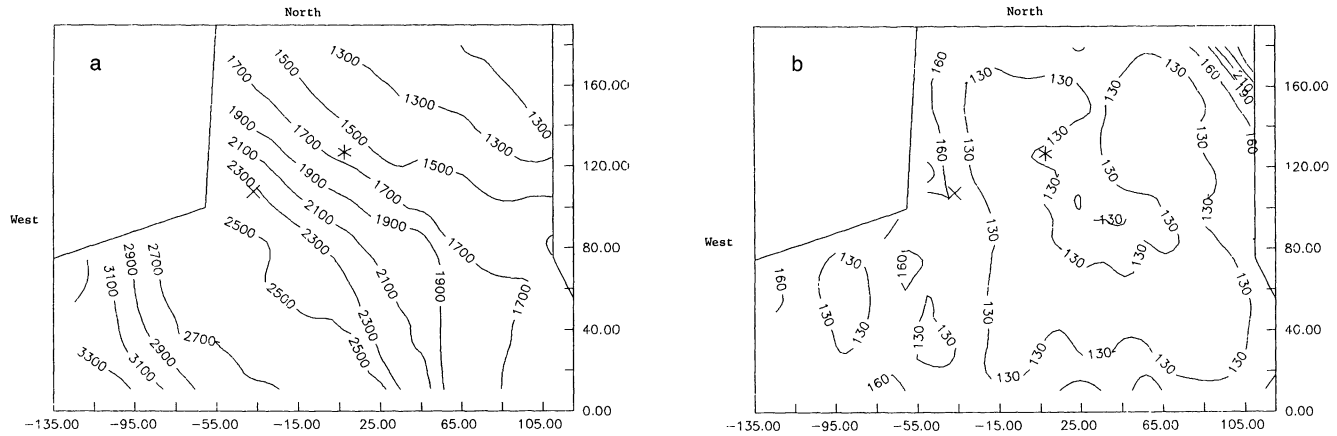

*Figure 4. (a) Predicted Surface (6) and (b) Standard Error Surface (8) (in ft.) Obtained by Kriging: ×, Center of Deaf Smith County; *, Location of Amarillo, Texas.*

not on a regular grid, tolerance regions for $N(\mathbf{h})$ in (5) have to be declared. The pair of locations $(\mathbf{s}, \mathbf{u}) \in N(\mathbf{h})$ if $\|\mathbf{s}-\mathbf{u}\| - \|\mathbf{h}\|$ is bounded between ± 2.5 miles, and the direction of $(\mathbf{s}-\mathbf{u})$ minus the direction of $\mathbf{h}$ is bounded between $\pm \pi/4$ radians.

Superimposed on Figure 3 are weighted-least-squares fits of the variogram model

$$\begin{aligned} 2\gamma(h;\boldsymbol{\theta}) &= 0, && h = 0 \\ &= 2\{c_0 + b|h|^p\}, && h > 0,\ c_0 > 0, \\ & && b > 0,\ 0 \leqq p < 2. \end{aligned} \quad (9)$$

In Figure 3a (NE–SW direction),

$$\hat{\boldsymbol{\theta}} = (\hat{c}_0, \hat{b}_1, \hat{p}) = (14 \times 10^3 \text{ft.}^2,\ 38,\ 1.99);$$

in Figure 3b (NW–SE direction),

$$\hat{\boldsymbol{\theta}} = (\hat{c}_0, \hat{b}_2, \hat{p}) = (14 \times 10^3 \text{ft.}^2,\ 15,\ 1.99).$$

Thus the process is exhibiting *anisotropy*; the fitted two-dimensional anisotropic semivariogram model used for kriging is

$$\gamma(\mathbf{h};\hat{\boldsymbol{\theta}}) = \hat{c}_0 + \left\{ \hat{b}_1^{2/\hat{p}}\, r^2 \cos^2\left(\frac{\pi}{4} - \phi\right) + \hat{b}_2^{2/\hat{p}}\, r^2 \cos^2\left(\frac{\pi}{4} + \phi\right) \right\}^{\hat{p}/2}, \quad (10)$$

where $\mathbf{h} = (h_1, h_2) \equiv (r\cos\phi,\ r\sin\phi)$. Geometrical anisotropy (Journel and Huijbregts 1978, p. 179) was assumed in deriving (10).

From the (fitted) variogram (10), contour maps of the kriging predictor (6) and the kriging standard error (8) can be drawn. The Toolkit package by Geostokos Ltd., London, was used; it is interactive and well-documented, has the anisotropy option built in, and outputs prediction and standard error values for easy contouring. The results are displayed in Figure 4. Figure 5 shows a three-dimensional plot of the ordinary-kriging surface of Figure 4a, viewed from the northeast corner of the region under study. It can be concluded from these maps that contaminated groundwater from Deaf Smith County, Texas, would flow directly "downhill" toward Amarillo, Texas.

5. CONCLUDING REMARKS

In attempting to paint a picture of the geostatistical method in broad brushstrokes, many subtleties have been glossed over. David (1977) and Journel and Huijbregts (1978) are good sources for these, or patient readers might wait to consult chapters 2–5 of Cressie (in press).

One of the least subtle observations one might make from Section 4 is that the constant-mean assumption (3) may not hold. Write

$$Z(\mathbf{s}) = \mu(\mathbf{s}) + \delta(\mathbf{s}),$$

where $E(Z(\mathbf{s})) = \mu(\mathbf{s})$ and $\delta(\cdot)$ is a zero mean, intrinsically stationary stochastic process with $\text{var}(\delta(\mathbf{s}+\mathbf{h}) - \delta(\mathbf{s})) = \text{var}(Z(\mathbf{s}+\mathbf{h}) - Z(\mathbf{s})) = 2\gamma(\mathbf{h})$. The "large-scale variation" $\mu(\cdot)$ and the "small-scale variation" $\delta(\cdot)$ are modeled as deterministic and stochastic processes, respectively, but without any way of making each of them identifiable. What is one person's mean structure could be another person's correlation structure.

Often this problem is resolved in a substantive application

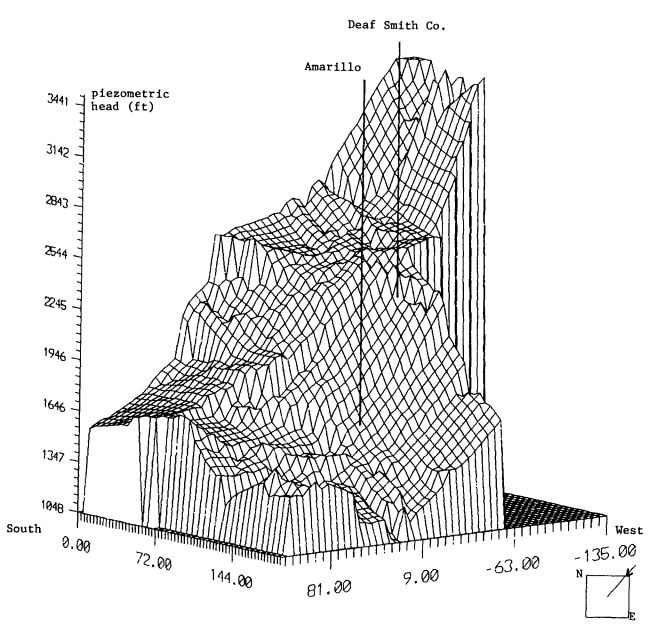

Figure 5. Three-dimensional View of Ordinary-Kriging Surface $\{\hat{Z}(\mathbf{s}_0)$: $\mathbf{s}_0 \in D\}$, From the Northeast Corner of D.

by relying on scientific or habitual reasons for determining the mean structure. Cressie (1986) discussed various ways of kriging in the presence of nonconstant mean. The contour map of predicted values $\{\hat{Z}(\mathbf{s}_0)\colon \mathbf{s}_0 \in D\}$ typically does not change much, but the map of kriging standard errors does; Cressie (in press, sec. 4.1) illustrates this point on the Wolfcamp aquifer data.

Another problem that geostatistics has yet to address properly is the effect of estimating variogram parameters on mean squared prediction errors. Intuitively, this estimation should lead to larger actual prediction standard errors than $\sigma_k(\mathbf{s}_0)$ given by (8). Zimmerman and Cressie (1988) have been able to quantify this intuition in several special cases.

In this article I have summarized the geostatistical method, featuring a particular method of spatial prediction known as kriging. Many other prediction methods, both stochastic and nonstochastic, are available, and are summarized in Cressie (1989a). Confronted with a set of data, which is the best method to apply? Depending on which model is true, the answer may change drastically. For this reason, several authors have conducted investigations where the values to be predicted were actually known. They then compared the performance of various methods by determining the closeness of predicted to actual. For example, Laslett, McBratney, Pahl, and Hutchinson (1987) found interpolating methods in general to be poor, and Laplacian smoothing splines to be as good as kriging (for predicting soil pH); the relationship between kriging and splines was reviewed by Watson (1984). In all comparisons, on real and simulated data, universal kriging generally did as well or better than the other methods.

Intuitively, the prediction method must be flexible, adapting to the underlying spatial variation in the data. Kriging has this flexibility, since the spatial dependence structure is first gauged from an initial data analysis, before the kriging equations are solved. Unlike splines, kriging adapts to the quality and quantity of spatial dependence demonstrated by the data, produces prediction standard errors, and easily handles prediction of an unobservable whose support is different from those of the data (Matheron 1963).

[*Received October 1987. Revised May 1989.*]

REFERENCES

Bilonick, R. A. (1985), "The Space-Time Distribution of Sulfate-Deposition in the Northeastern United States," *Atmospheric Environment*, 19, 1829–1845.

Box, G. E. P., and Jenkins, G. M. (1970), *Time Series Analysis, Forecasting and Control*, San Francisco: Holden-Day.

Cliff, A., and Ord, J. K. (1981), *Spatial Processes: Models and Applications*, London: Pion.

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

——— (1986), "Kriging Nonstationary Data," *Journal of the American Statistical Association*, 81, 625–634.

——— (1988), "Spatial Prediction and Ordinary Kriging," *Mathematical Geology*, 20, 405–421.

——— (1989a), "The Many Faces of Spatial Prediction," in *Geostatistics* (Vol. 1), ed. M. Armstrong, Dordrecht, Holland: Kluwer, pp. 163–176.

——— (1989b), "The Origins of Kriging," unpublished manuscript submitted to *Mathematical Geology*.

——— (in press), *Statistics for Spatial Data*, New York: John Wiley.

Cressie, N., Gotway, C. A., and Grondona, M. O. (in press), "Spatial Prediction From Networks," *Chemometrics and Intelligent Laboratory Systems*, 6.

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

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

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

Harper, W. V., and Furr, J. M. (1986), "Geostatistical Analysis of Potentiometric Data in the Wolfcamp Aquifer of the Palo Duro Basin, Texas," Technical Report ONWI-587, Battelle Memorial Institute, Columbus, OH.

Hogg, R. V., and Craig, A. T. (1978), *Introduction to Mathematical Statistics* (4th ed.), New York: Macmillan.

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

Jowett, G. H. (1952), "The Accuracy of Systematic Sampling From Conveyor Belts," *Applied Statistics*, 1, 50–59.

Kolmogorov, A. N. (1941a), "The Local Structure of Turbulence in an Incompressible Fluid at Very Large Reynolds Numbers," *Doklady Akademii Nauk SSSR*, 30, 229–303.

——— (1941b), "Interpolation and Extrapolation of Stationary Random Sequences," *Izvestiia Akademii Nauk SSSR, Seriia Matematicheskaia*, 5, 3–14.

Laslett, G. M., McBratney, A. B., Pahl, P. J., and Hutchinson, M. F. (1987), "Comparison of Several Spatial Prediction Methods for Soil pH," *Journal of Soil Science*, 38, 325–341.

Matheron, G. (1963), "Principles of Geostatistics," *Economic Geology*, 58, 1246–1266.

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

Upton, G. J. G., and Fingleton, B. (1985), *Spatial Data Analysis by Example*, Chichester, U.K.: John Wiley.

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

Wiener, N. (1949), *Extrapolation, Interpolation and Smoothing of Stationary Time Series*, Cambridge, MA: MIT Press.

Wold, H. (1938), *A Study in the Analysis of Stationary Time Series*, Uppsala, Sweden: Almqvist & Wiksells.

Yaglom, A. M. (1957), "Some Classes of Random Fields in *n*-Dimensional Space, Related to Stationary Random Processes," *Theory of Probability and Its Applications*, 2, 273–320.

Zimmerman, D. L., and Cressie, N. (1988), "Improved Estimation of the Kriging Variance," Preprint 88-37, Iowa State University, Statistical Laboratory.

Zimmerman, D. L., and Zimmerman, M. B. (1989), "A Monte Carlo Comparison of Spatial Variogram Esstimators, Kriging Predictors, and Kriging Variances," unpublished manuscript submitted to *Technometrics*.