Mathematical Geology, Vol. 16, No. 5, 1984

Analysis of Nonintrinsic Spatial Variability by Residual Kriging with Application to Regional Groundwater Levels[1]

Shlomo P. Neuman[2] and Elizabeth A. Jacobson[2,3]

A method for obtaining pointwise or spatially averaged estimates of a nonintrinsic function is introduced based on residual kriging. The method relies on a stepwise iterative regression process for simultaneously estimating the global drift and residual semivariogram. Estimates of the function are then obtained by solving a modified set of simple kriging equations written for the residuals. The modification consists of replacing the true variogram in the kriging equations by the variogram of the residual estimates as obtained from the iterative regression process. The method is illustrated by considering groundwater levels in an Arizona aquifer. The results are compared with those obtained for the aquifer by the generalized covariance package BLUEPACK-3D.

KEY WORDS: Spatial variability, nonintrinsic, nonstationary, drift, kriging, residuals, groundwater levels.

INTRODUCTION

A variety of geostatistical problems involve nonintrinsic phenomena exhibiting a spatial drift. Examples include the topography of a hill slope, the elevation of a dipping structural feature, and the thickness of a wedge-shaped stratum. Of particular concern in this paper is the spatial variation of vertically averaged hydraulic heads in an aquifer. For groundwater to flow laterally within the aquifer, the hydraulic head surface must have a distinct inclination in the main direction of fluid movement. Suppose that the hydraulic head is represented by water levels measured in a number of wellbores completed within the aquifer. To estimate the hydraulic head at selected points between these wellbores by kriging, one

[1]Manuscript received 7 March 1983; revised 1 August 1983.

[2]Department of Hydrology and Water Resources, University of Arizona, Tucson, Arizona 85721 U.S.A.

[3]Now with Geosciences Research and Engineering Department, Battelle, Pacific Northwest Laboratories, P.O. Box 999, Richland, Washington, 99352 U.S.A.

0020-5958/84/0700-0499$03.50/0

needs to know the drift and semivariogram of the differences between actual and mean head values. Our problem is to estimate the drift and semivariogram from the available wellbore data and incorporate these estimates in the kriging process.

Let $Z(\mathbf{x})$ be the nonintrinsic function of interest where $\mathbf{x}$ is a vector of coordinates. We assume that $Z(\mathbf{x})$ can be expressed as the sum of a deterministic drift component, $\mu(\mathbf{x})$, and a zero-mean intrinsic residual component, $R(\mathbf{x})$, such that

$$Z(\mathbf{x}) = \mu(\mathbf{x}) + R(\mathbf{x}) \tag{1}$$

Since $E[R(\mathbf{x})] = 0$, it follows that $E[Z(\mathbf{x})] = \mu(\mathbf{x})$, E being the expectation symbol. We further assume that $R(\mathbf{x})$ can be characterized by an isotropic semivariogram, $\gamma_R(s)$, defined as

$$\gamma_R(s) = \tfrac{1}{2} E[R(\mathbf{x} + \mathbf{s}) - R(\mathbf{x})]^2 \tag{2}$$

where $\mathbf{s}$ is a displacement vector and $s = |\mathbf{s}|$.

Suppose that $Z(\mathbf{x})$ was measured at a discrete number of points and we wish to use these data to determine the semivariogram of Z by means of a standard procedure (e.g., Journel and Huijbregts, 1978). Theoretically, such a semivariogram is an estimate of a function, γ_Z, defined as

$$\gamma_Z(\mathbf{x}, \mathbf{s}) = \tfrac{1}{2} E[Z(\mathbf{x} + \mathbf{s}) - Z(\mathbf{x})]^2 \tag{3}$$

Substituting (1) into (3), expanding and using the above definition of $\gamma_R(s)$, we find that

$$\gamma_Z(\mathbf{x}, \mathbf{s}) = \gamma_R(s) + \tfrac{1}{2} [\mu(\mathbf{x} + \mathbf{s}) - \mu(\mathbf{x})]^2 \tag{4}$$

This shows that for arbitrary drifts, $\gamma_Z(\mathbf{x}, \mathbf{s})$ remains a function of both $\mathbf{x}$ and $\mathbf{s}$, indicating that $Z(\mathbf{x})$ is neither intrinsic nor isotropic. In the particular case where $\mu(\mathbf{x})$ varies linearly with $\mathbf{x}$, $\gamma_Z(\mathbf{x}, \mathbf{s})$ increases quadratically with s without limit. This feature, together with a strong anisotropy, can be used to diagnose the presence of a drift by examining the shape of a sample semivariogram, determined from the data. If the diagnosis is positive, the appropriate variogram is $\gamma_R(s)$, not $\gamma_Z(\mathbf{x}, \mathbf{s})$. The difficulty is that $\gamma_R(s)$ cannot be determined from measured values of $Z(\mathbf{x})$ without knowing $\mu(\mathbf{x})$, and $\mu(\mathbf{x})$ cannot be determined without knowing $\gamma_R(s)$.

Past attempts to resolve this difficulty have led to three key methods: Universal kriging, generalized covariance, and simple kriging modified to account for a prior estimate of the drift. Universal kriging was introduced by Matheron (1969) and applied to various problems by Delhomme (1976), Haas and Jouseline (1976), Chiles (1977), and others. Good descriptions of the method have been published by Delfiner (1976) and Beucher, Delhomme, and de Marsily (1981). The method differs from simple kriging in that $\mu(\mathbf{x})$ is introduced formally into the kriging model as a low-order polynomial with undetermined coefficients. Since such a low-order approximation may not be valid over the entire

domain of interest, universal kriging is usually applied locally over small neighborhoods. Theoretically, the model is written in terms of $\gamma_R(s)$. In practice, however, $\gamma_R(s)$ is unknown. One approach is to replace $\gamma_R(s)$ by γ_Z. This can be justified as long as γ_Z is applied to small neighborhoods within which s is small enough to disregard the square term in (4), so that $\gamma_Z \approx \gamma_R(s)$. Another approach is to estimate the drift on the basis of an assumed $\gamma_R(s)$ and then obtain a new semivariogram from the computed residuals. The resulting bias depends on the structure of the nonintrinsic function and the method used to estimate the drift (Matheron, 1969; Huijbregts and Matheron, 1971; Sabourin, 1976; David, 1977, p. 247). The error in both the drift and the residual semivariogram can be minimized by adopting an iterative generalized least-squares procedure similar to that employed in this paper and alluded to by others (e.g., Ripley, 1981, p. 58). This, however, eliminates the need for universal kriging, as we will point out later. The (unnecessary) practice of applying universal kriging to local neighborhoods restricts the method to sufficiently dense data sets, otherwise there may not be enough points for kriging in each neighborhood. The local approach also suffers from a lack of objective criteria to choose the order of the polynomial drift and the size of the neighborhood within which universal kriging is valid. Such choices can only be justified a posteriori by cross validation, a technique in which the analyst eliminates one datum point at a time and compares its measured value with that estimated by kriging.

Closely related to universal kriging is the generalized covariance method of Matheron (1971, 1973) and Delfiner (1976). Here again $\mu(\mathbf{x})$ is taken to be a low-order polynomial over a small neighborhood, leading to a system of equations similar to those arising in universal kriging. However, instead of working with explicit forms of the drift function, the method seeks to determine $\gamma_R(s)$ by filtering out a polynomial drift. Such filtering is accomplished by considering not only zero-order spatial increments of $Z(\mathbf{x})$ as one does in simple or universal kriging but also higher-order increments. In general, an increment of order m is able to filter out a polynomial drift of similar order. Thus, if the residuals are intrinsic, so are the mth order increments of $Z(\mathbf{x})$, and $Z(\mathbf{x})$ is therefore said to be an "intrinsic random function of order m." Clearly, the semivariogram of such mth order increments, $\gamma_m(s)$, is identical to $\gamma_R(s)$, the semivariogram of zero-order increments of the residual, $R(\mathbf{x})$. The function $\gamma_m(s)$ is called "generalized semivariogram of order m."

The number of data points required to estimate $\gamma_m(s)$ increases rapidly with m. For this reason, m in practice is seldom taken to be greater than 2. Since a second-order polynomial approximation of $\mu(\mathbf{x})$ may not be valid on the global scale, the method is usually applied to small neighborhoods. The approach, based on theoretical considerations, is to approximate $\gamma_m(s)$ by a low-order ($\leqslant 5$) polynomial whose coefficients must satisfy certain predetermined criteria. These coefficients are evaluated automatically by fitting theoretical mth order increments to their experimental values with the aid of ordinary least

squares (in some contrast to γ_Z and $\gamma_R(s)$, which are often fitted to the sample semivariogram by eye). Just as in the case of universal kriging, the choice of neighborhood size and m can only be justified a posteriori by cross validation.

Conventional universal kriging and generalized covariance techniques make no attempt to define the global nature of the drift. One attempt to determine the global drift explicitly has been reported by Gambolati and Volpi (1979) in connection with land subsidence near Venice, Italy. Based on hydrogeological considerations they argued that the drift, $\mu(\mathbf{x})$, of the groundwater head surface, $Z(\mathbf{x})$, must vary logarithmically with distance from the subsidence-causing center of pumping outside of Venice. They estimated the coefficients of this logarithmic function by fitting it to measured groundwater level data with the aid of ordinary nonlinear least squares. This enabled them to compute the residuals at all well locations and to construct an approximate residual semivariogram. The authors then developed a kriging model incorporating the estimates of $\mu(\mathbf{x})$ and $\gamma_R(s)$ that includes fewer constraints, and is thus simpler, than equivalent models based on universal kriging or generalized covariance (this is true even if $\mu(\mathbf{x})$ in the latter is treated locally as a polynomial of just first order). Unfortunately, their approach is internally inconsistent because ordinary least squares treat the residuals as being uncorrelated, whereas the existence of a semivariogram depending on s implies that $R(\mathbf{x})$ has an autocorrelation structure.

In the present paper we describe a consistent method for obtaining pointwise or spatially averaged estimates of a nonintrinsic function based on simple kriging of the residuals. The method is based on a stepwise iterative regression process that yields simultaneous estimates of the global drift and the residual semivariogram. A case study is included to illustrate the method, dealing with groundwater level data from the Avra Valley aquifer in southern Arizona. Other case studies can be found in the dissertation of Binsariti (1980) and in the thesis of Fennessy (1982). The results of the Avra Valley case are compared with those obtained for the same aquifer by the generalized covariance package, BLUEPACK-3D, developed at the Ecole des Mines in Fontainebleau, France.

STEPWISE ITERATIVE REGRESSION

Let $\mathbf{Z}$ be a vector of $Z(\mathbf{x})$ values measured at I discrete locations, $\mathbf{x}_i$, $i = 1, 2, \ldots, I$, and let $\boldsymbol{\mu}$ be the corresponding drift vector. Then, by virtue of (1), $\mathbf{Z}$ and $\boldsymbol{\mu}$ are related through

$$\mathbf{Z} = \boldsymbol{\mu} + \mathbf{R} \tag{5}$$

where $\mathbf{R}$ is the vector of residuals at $\mathbf{x}_i$. The latter vector is characterized by a covariance matrix, $\boldsymbol{V}$, defined as

$$E[\mathbf{R}\mathbf{R}^T] = \boldsymbol{V} \tag{6}$$

If $R(\mathbf{x})$ is weakly stationary and $\gamma_R(s)$ is known, the (i,j)th term of $\boldsymbol{V}$ can be determined from

$$V_{ij} = \rho_R(s_{ij}) = \rho_R(o) - \gamma_R(s_{ij}) \tag{7}$$

where $\rho_R(s)$ is the covariance function of $R(\mathbf{x})$, defined as

$$E[R(\mathbf{x})\, R(\mathbf{x}+\mathbf{s})] = \rho_R(s) \tag{8}$$

and $s_{ij} = |\mathbf{x}_i - \mathbf{x}_j|$. Clearly, $\rho(o)$ is the "sill" of the semivariogram, defined as the constant maximum value of $\gamma_R(s)$ attained beyond a certain "range" of s values, $s \geqslant s_o$.

Suppose that the drift at the pth step of our procedure is expressed as

$$\mu_p(\mathbf{x}) = \sum_{j=1}^{J_p} a_j f_j(\mathbf{x}) \tag{9}$$

where J_p is an integer satisfying $J_{p+1} > J_p$, a_j are coefficients to be determined, and $f_j(\mathbf{x})$ are prescribed linearly independent basis functions. We can then rewrite (5) in the form

$$\mathbf{Z} = \boldsymbol{F}\mathbf{a} + \mathbf{R} \tag{10}$$

where $\boldsymbol{F}$ is an $I \times J_p$ matrix of the basis function values, $f_j(\mathbf{x}_i)$, and $\mathbf{a}$ is a vector of the J_p coefficients, a_j. If $\boldsymbol{V}$ is known, one can obtain an unbiased minimum variance estimate of $\mathbf{a}$, $\hat{\mathbf{a}}$, by minimizing the generalized least-squares criterion (Schweppe, 1973, p. 101)

$$\Omega(\hat{\mathbf{a}}) = (\mathbf{Z} - \boldsymbol{F}\hat{\mathbf{a}})^T\, \boldsymbol{V}^{-1} (\mathbf{Z} - \boldsymbol{F}\hat{\mathbf{a}}) \tag{11}$$

The resulting estimate of $\mathbf{a}$ is given by

$$\hat{\mathbf{a}} = \boldsymbol{V}_a \boldsymbol{F}^T \boldsymbol{V}^{-1} \mathbf{Z} \tag{12}$$

where $\boldsymbol{V}_a$ is the covariance matrix of the estimation error, $\hat{\mathbf{a}} - \mathbf{a}$, defined as $\boldsymbol{V}_a = E[(\hat{\mathbf{a}} - \mathbf{a})(\hat{\mathbf{a}} - \mathbf{a})^T]$. This matrix is computed independently of $\mathbf{Z}$ from

$$\boldsymbol{V}_a = (\boldsymbol{F}^T \boldsymbol{V}^{-1} \boldsymbol{F})^{-1} \tag{13}$$

Estimates of the drift, $\boldsymbol{\mu}$, and residuals, $\mathbf{R}$, are given, respectively, by

$$\hat{\boldsymbol{\mu}} = \boldsymbol{F}\hat{\mathbf{a}} \quad \text{and} \tag{14}$$

$$\hat{\mathbf{R}} = \mathbf{Z} - \hat{\boldsymbol{\mu}} \tag{15}$$

These can be rewritten with the aid of (12) as

$$\hat{\boldsymbol{\mu}} = \boldsymbol{P}\mathbf{Z} \quad \text{and} \tag{16}$$

$$\hat{\mathbf{R}} = (\boldsymbol{I} - \boldsymbol{P})\, \mathbf{Z} \tag{17}$$

where $\boldsymbol{P} = \boldsymbol{F}\boldsymbol{V}_a\boldsymbol{F}^T\boldsymbol{V}^{-1}$. According to (A1) in Appendix A, $(\boldsymbol{I} - \boldsymbol{P})\,\boldsymbol{F} = \boldsymbol{O}$ and thus, $\hat{\mathbf{R}}$ can also be expressed as

$$\hat{\mathbf{R}} = (\boldsymbol{I} - \boldsymbol{P})\,\mathbf{R} \tag{18}$$

Equation (A2) shows that rank $(\boldsymbol{I} - \boldsymbol{P}) \geqslant I - J_p$, implying that the computed residuals are generally linearly dependent. The rank becomes strictly equal to $(I - J_p)$ when $\boldsymbol{V} = \boldsymbol{I}$ (Seber, 1977, p. 46).

Since $\mathbf{R}$ has mean zero, the same clearly holds true for $\hat{\mathbf{R}}$. On the other hand, the covariance of $\hat{\mathbf{R}}$, $\hat{\boldsymbol{V}} = E[\hat{\mathbf{R}}\hat{\mathbf{R}}^T]$, is given by $(\boldsymbol{I} - \boldsymbol{P})\,\boldsymbol{V}(\boldsymbol{I} - \boldsymbol{P})^T$ which, by virtue of (A3) and (A4), takes the simpler form

$$\hat{\boldsymbol{V}} = (\boldsymbol{I} - \boldsymbol{P})\,\boldsymbol{V} \tag{19}$$

It follows that

$$\boldsymbol{V} - \hat{\boldsymbol{V}} = \boldsymbol{P}\boldsymbol{V} = \boldsymbol{F}\boldsymbol{V}_a\boldsymbol{F}^T \tag{20}$$

is positive semidefinite and the covariance structure of $\hat{\mathbf{R}}$ is not exactly the same as that of $\mathbf{R}$. This means that the semivariogram, $\gamma_{\hat{R}}(s)$, determined from the computed residuals, $\hat{\mathbf{R}}$, will differ from $\gamma_R(s)$: in particular, the sill of $\gamma_{\hat{R}}(s)$ will generally be smaller than that of $\gamma_R(s)$. For this reason, we develop our kriging equations in terms of $\gamma_{\hat{R}}$, instead of γ_R, as is usually the custom. Implicit in this approach, of course, is the assumption that a positive definite $\gamma_{\hat{R}}$ exists. This is true at best as an approximation when $I \gg J_p$.

Let

$$\boldsymbol{\nu} = \hat{\mathbf{R}} - \mathbf{R} = \boldsymbol{\mu} - \hat{\boldsymbol{\mu}} \tag{21}$$

be the estimation error of $\hat{\mathbf{R}}$ and $\hat{\boldsymbol{\mu}}$. By virtue of (18), $E[\boldsymbol{\nu}] = \mathbf{0}$, meaning that both $\hat{\mathbf{R}}$ and $\hat{\boldsymbol{\mu}}$ are unbiased. The covariance of $\boldsymbol{\nu}$, $\boldsymbol{V}_\nu = E[\boldsymbol{\nu}\boldsymbol{\nu}^T]$, is obtained directly from (13) and (14) as

$$\boldsymbol{V}_\nu = \boldsymbol{F}\boldsymbol{V}_a\boldsymbol{F}^T \tag{22}$$

If the drift at an arbitrary point, $\mathbf{x}$, is computed according to

$$\hat{\mu}(\mathbf{x}) = \sum_{j=1}^{J_p} \hat{a}_j f_j(\mathbf{x}) \tag{23}$$

then the covariance, $C_\mu(\mathbf{x}, \mathbf{x} + \mathbf{s}) = E\{[\hat{\mu}(\mathbf{x}) - \mu(\mathbf{x})]\,[\hat{\mu}(\mathbf{x} + \mathbf{s}) - \mu(\mathbf{x} + \mathbf{s})]\}$, of the corresponding estimation error, $\hat{\mu}(\mathbf{x}) - \mu(\mathbf{x})$, is given by

$$C_\mu(\mathbf{x}, \mathbf{x} + \mathbf{s}) = \sum_{i=1}^{J_p} \sum_{j=1}^{J_p} f_i(\mathbf{x})\, V_{aij} f_j(\mathbf{x} + \mathbf{s}) \tag{24}$$

where V_{aij} is the (i, j)th component of $\boldsymbol{V}_a$.

Since $\gamma_R(s)$ is initially unknown, $\boldsymbol{V}$ is unknown, and the above regression model cannot be used. We therefore proceed in two stages. In Stage 1, the residuals are treated as if they were uncorrelated, and $\boldsymbol{V}$ is replaced by the identity matrix, $\boldsymbol{I}$. This is equivalent to ordinary, as opposed to generalized, least squares. Starting with $p = 1$, we compute $\hat{\mathbf{R}}$, and use these residuals to obtain a semivariogram, $\gamma_{\hat{R}}(s)$, in a standard manner. If $\gamma_{\hat{R}}(s)$ shows a distinct sill and no marked anisotropy, it is adopted for Stage 2 of the analysis. Otherwise, p is incremented by 1, and the procedure is repeated until $\gamma_{\hat{R}}(s)$ attains the desired properties.

Stage 2 starts by computing $\hat{\boldsymbol{V}}$ from

$$\hat{V}_{ij} = \rho_{\hat{R}}(o) - \gamma_{\hat{R}}(s_{ij}) \tag{25}$$

in analogy to (7). Upon replacing $\boldsymbol{V}$ in the regression model by $\hat{\boldsymbol{V}}$, we proceed to evaluate $\hat{\mathbf{R}}$, which leads to a new $\gamma_{\hat{R}}(s)$, which in turn yields a new $\hat{\boldsymbol{V}}$ matrix, and so on. This iterative process continues until the computed drift, $\hat{\boldsymbol{\mu}}$, and the semivariogram, $\gamma_{\hat{R}}(s)$, attain stable values.

The above procedure worked well in all the three case studies we have mentioned in the introduction. In principle, however, Stage 1 should work only if the ordinary least-squares estimator of the drift is efficient with respect to the (optimal) generalized least-squares estimator. When this is the case, the improvement attained at Stage 2 will be small compared to that attained at Stage 1, although not necessarily insignificant (as we find in all three case studies). If Stage 1 does not show a marked improvement within a few steps, it may be necessary to repeat it by iterating after each step, that is, applying Stage 2 after each step of Stage 1.

Thus far, $\mu_p(\mathbf{x})$ has been taken to be linear in the regression parameters, a_j. Extension to the case where $\mu_p(\mathbf{x})$ is nonlinear in $\mathbf{a}$ (as in the work of Gambolati and Volpi, 1979) is relatively straightforward. Assume that, instead of (9), we have

$$\mu_p(\mathbf{x}) = \sum_{j=1}^{J_p} f_j(\mathbf{a}, \mathbf{x}) \tag{26}$$

where $\mathbf{a}$ is a vector of K_p coefficients, a_k, $k = 1, 2, \ldots, K_p$, $K_p \geqslant J_p$. Then (11) takes the form

$$\Omega(\hat{\mathbf{a}}) = (\mathbf{Z} - \hat{\boldsymbol{\mu}})^T \boldsymbol{V}^{-1} (\mathbf{Z} - \hat{\boldsymbol{\mu}}) \tag{27}$$

which must be minimized by nonlinear mathematical programming to yield an estimate of $\mathbf{a}$, $\hat{\mathbf{a}}$. Let $\boldsymbol{G}$ be the $I \times K_p$ Jacobian (or sensitivity) matrix whose (i, k)th component is defined as

$$G_{ik} = \sum_{j=1}^{J_p} \frac{\partial f_j(\hat{\mathbf{a}}, \mathbf{x})}{\partial \hat{a}_k} \tag{28}$$

Then, to a second order of approximation [Neuman and Yakowitz, 1979]

$$\boldsymbol{V}_a \approx (\boldsymbol{G}^T \boldsymbol{V}^{-1} \boldsymbol{G})^{-1} \tag{29}$$

$$\boldsymbol{V}_v \approx \boldsymbol{G} \boldsymbol{V}_a \boldsymbol{G}^T \quad \text{and} \tag{30}$$

$$C_\mu(\mathbf{x}, \mathbf{x}+\mathbf{s}) \approx \sum_{k=1}^{K_p} \sum_{l=1}^{K_p} g_k(\mathbf{x})\, V_{akl}\, g_l(\mathbf{x}+\mathbf{s}) \quad \text{where} \tag{31}$$

$$g_k(\mathbf{x}) = \sum_{j=1}^{J_p} \frac{\partial f_j(\hat{\mathbf{a}}, \mathbf{x})}{\partial \hat{a}_k}$$

Otherwise, the linear and nonlinear cases are treated in identical manners.

ESTIMATION OF Z BY RESIDUAL KRIGING

Suppose that we are interested in estimating the spatial average of $Z(\mathbf{x})$,

$$Z_n = \frac{1}{\Gamma_n} \int_{\Gamma_n} Z(\mathbf{x})\, d\mathbf{x} \tag{32}$$

over some finite domain, Γ_n. We can accomplish this by kriging the residuals and adding the result to the drift. Let μ_n and R_n be the spatial averages of $\mu(\mathbf{x})$ and $R(\mathbf{x})$, respectively, over Γ_n, and let $\hat{\mu}_n$ and $\hat{R}_n$ be the corresponding generalized least-squares estimates. These quantities are related through

$$Z_n = \mu_n + R_n = \hat{\mu}_n + \hat{R}_n \tag{33}$$

which can be viewed as a definition of $\hat{R}_n$.

We define the kriging estimate of Z_n, Z_n^*, as

$$Z_n^* = \hat{\mu}_n + R_n^* \tag{34}$$

where R_n^*, the kriging estimate of the unknown generalized least-squares residual, $\hat{R}_n$, is given by

$$R_n^* = \sum_{i=1}^{I_n} \lambda_{ni} \hat{R}_{ni} \tag{35}$$

Here λ_{ni} are "kriging coefficients," and $\hat{R}_{ni}$ are components of $\mathbf{R}$ corresponding to $I_n (\leqslant I)$ measurement points of $Z(\mathbf{x})$, $\mathbf{x}_{ni}$, $i = 1, 2, \ldots, I_n$, inside and/or in the neighborhood of Γ_n. By virtue of (18)

$$E[Z_n^*] = \mu_n \tag{36}$$

and the kriging estimate of Z_n is, thus, unbiased for any choice of λ_{ni}.

Defining the kriging estimation error as

$$\epsilon_n = Z_n^* - Z_n = R_n^* - \hat{R}_n \tag{37}$$

the covariance of this error, $V_{nm}^* = E[\epsilon_n \epsilon_m]$, becomes (Appendix B)

$$V_{nm}^* = -\gamma_{\hat{R}}(\Gamma_n, \Gamma_m) + \sum_{i=1}^{I_n} \lambda_{ni} \gamma_{\hat{R}}(\mathbf{x}_{ni}, \Gamma_m) + \sum_{j=1}^{I_m} \lambda_{mj} \gamma_{\hat{R}}(\mathbf{x}_{mj}, \Gamma_n) - \sum_{i=1}^{I_n} \sum_{j=1}^{I_m} \lambda_{ni} \lambda_{mj} \gamma_{\hat{R}}(\mathbf{x}_{ni}, \mathbf{x}_{mj}) \tag{38}$$

where $\gamma_{\hat{R}}(\mathbf{x}, \mathbf{y}) = \gamma_{\hat{R}}(|\mathbf{x} - \mathbf{y}|)$ and $\gamma_{\hat{R}}(\mathbf{x}, \Gamma)$ is the spatial average of $\gamma_{\hat{R}}(\mathbf{x}, \mathbf{y})$ over all $\mathbf{y} \in \Gamma$. This expression holds true provided that

$$\sum_{i=1}^{I_n} \lambda_{ni} = \sum_{j=1}^{I_m} \lambda_{mj} = 1 \tag{39}$$

The kriging coefficients, λ_{ni}, are determined so as to minimize V_{nm}^* subject to (39). Appendix B shows that this leads to the $(I_n + 1)$ "residual kriging equations"

$$\sum_{j=1}^{I_n} \lambda_{nj} \gamma_{\hat{R}}(\mathbf{x}_{ni}, \mathbf{x}_{nj}) + \beta_n = \gamma_{\hat{R}}(\mathbf{x}_{ni}, \Gamma_n) \qquad i = 1, 2, \ldots, I_n \tag{40a}$$

$$\sum_{j=1}^{I_n} \lambda_{nj} = 1 \tag{40b}$$

where β_n is a Lagrange multiplier. Upon solving these equations for λ_{nj} and β_n, the "kriging variance," V_{nn}^*, can be computed from

$$V_{nn}^* = -\gamma_{\hat{R}}(\Gamma_n, \Gamma_n) + \sum_{i=1}^{I_n} \lambda_{ni} \gamma_{\hat{R}}(\mathbf{x}_{ni}, \Gamma_n) + \beta_n \tag{41}$$

which is obtained by substituting (40a) into (38) when $m = n$. The "kriging error" is defined as $(V_{nn}^*)^{1/2}$.

The above kriging equations differ from their conventional counterparts in that they involve $\gamma_{\hat{R}}$ instead of γ_R. Since these equations are exact, knowing the semivariogram of the true residuals, γ_R, would be of little help unless μ was also known exactly.

RELATIONSHIP TO UNIVERSAL KRIGING

In universal kriging the estimate of Z_n, Z_n^U, is written as

$$Z_n^U = \sum_{n=1}^{I_n} \lambda_{ni} Z_{ni} \tag{42}$$

where Z_{ni} are the values of $Z(\mathbf{x})$ at the I_n measurement points, $\mathbf{x}_{ni}$, $i = 1, 2, \ldots, I_n$. Since the method does not rely on a prior estimate of the drift, the latter is filtered out by requiring that the kriging coefficients satisfy

$$f_j(\Gamma_n) = \sum_{i=1}^{I_n} \lambda_{ni} f_j(\mathbf{x}_{ni}) \tag{43}$$

for all $j = 1, 2, \ldots, J_p$, where

$$f_j(\Gamma_n) = \frac{1}{\Gamma_n} \int_{\Gamma_n} f_j(\mathbf{x})\, d\mathbf{x}$$

This, by virtue of (23), implies that

$$\mu_n = \sum_{i=1}^{I_n} \lambda_{ni} \mu(\mathbf{x}_{ni}) \qquad \text{and} \tag{44}$$

$$\hat{\mu}_n = \sum_{i=1}^{I_n} \lambda_{ni} \hat{\mu}(\mathbf{x}_{ni}) \tag{45}$$

As shown in Appendix B, (44) leads to the universal kriging equations

$$\sum_{j=1}^{I_n} \lambda_{nj} \gamma_R(\mathbf{x}_{ni}, \mathbf{x}_{nj}) + \sum_{l=1}^{J_p} \beta_{nl} f_l(\mathbf{x}_{ni}) = \gamma_R(\mathbf{x}_{ni}, \Gamma_n) \qquad i = 1, 2, \ldots, I_n \tag{46a}$$

$$\sum_{i=1}^{I_n} \lambda_{ni} f_l(\mathbf{x}_{ni}) = f_l(\Gamma_n) \qquad l = 1, 2, \ldots, J_p \tag{46b}$$

where β_{nl} are Lagrange multipliers and it is understood that $f_1(\mathbf{x}) \equiv 1$. Similarly, (45) leads to an equivalent set of equations in terms of $\gamma_{\hat{R}}$. Both sets of universal kriging equations exceed the number of residual kriging equations in (40) by $(J_p - 1)$.

The increase in the number of equations and unknowns in universal kriging,

as compared with residual kriging, stems from the attempt of the former method to circumvent the need for estimating the drift coefficients, **a**. In addition to increasing problem size, this attempt leads to ambiguities in the determination of the semivariogram and the choice of the basis functions for the drift.

AVRA VALLEY CASE STUDY

We illustrate our method by considering groundwater level data from the Avra Valley aquifer in southern Arizona (Fig. 1). A detailed description of the

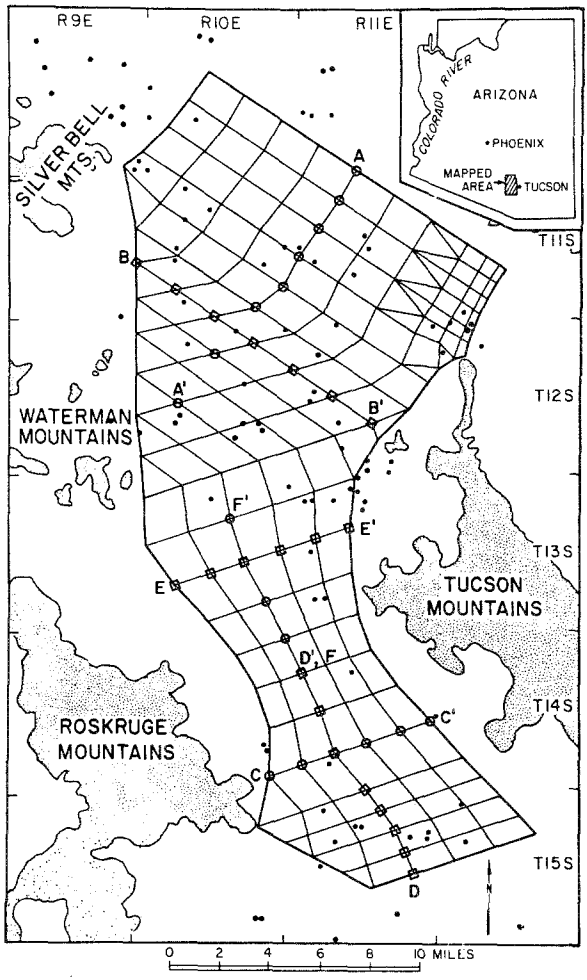

Fig. 1. Location of Avra Valley study area and finite element grid. Dots indicate location of wells in which water levels were measured.

valley and aquifer can be found in Clifton (1981) and Clifton and Neuman (1982). Figure 1 shows the location of 99 wells at which quasi steady-state water level data are available. Superimposed on the figure is a finite element grid used by Clifton and Neuman to simulate groundwater flow in the aquifer. Our objective is to estimate the steady-state water levels at all the finite element grid points.

Let $Z(\mathbf{x})$ represent water levels or heads in the aquifer. Figure 2 shows an average sample semivariogram, γ_Z, based on the 99 available Z data. The semivariogram is seen to increase with s at a rate higher than quadratic, suggesting the presence of a drift.

To deal with this problem the basis functions, $f_j(\mathbf{x})$, in (9) were chosen to be monomials of up to 4th order. In Stage 1 of the analysis, p varied from 1 to 4 so that $\mu_p(\mathbf{x})$ represented complete polynomials ranging from 1st to 4th order (with $J_1 = 3$, $J_2 = 6$, $J_3 = 10$, and $J_4 = 15$). For each value of p the coefficients of the polynomial, a_j, were evaluated by an ordinary least-squares fit to the data. Figure 3 shows directional and average sample residual semivariograms, $\gamma_{\hat{R}}(s)$, for all four cases. One can see that as p increases from 1 to 4, the spread of the directional semivariograms gradually decreases, the presence of a sill becomes more distinct, and the magnitude of the sill diminishes (note the gradual reduction in vertical scale from Fig. 3a to 3d). Based on these results, we decided to

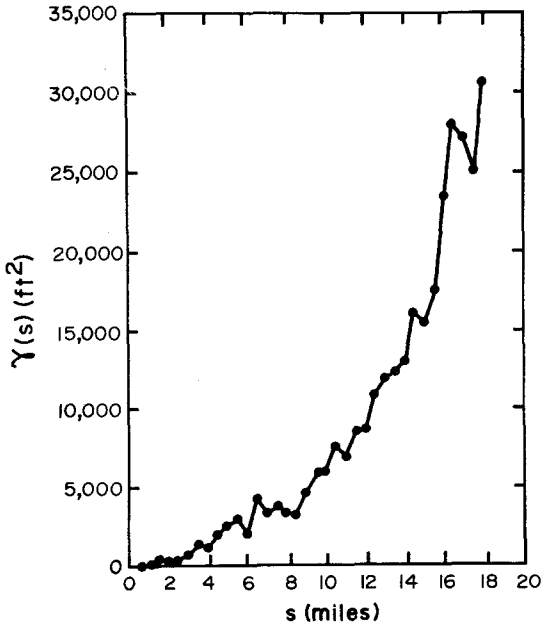

Fig. 2. Average sample semivariogram of water levels for Avra Valley.

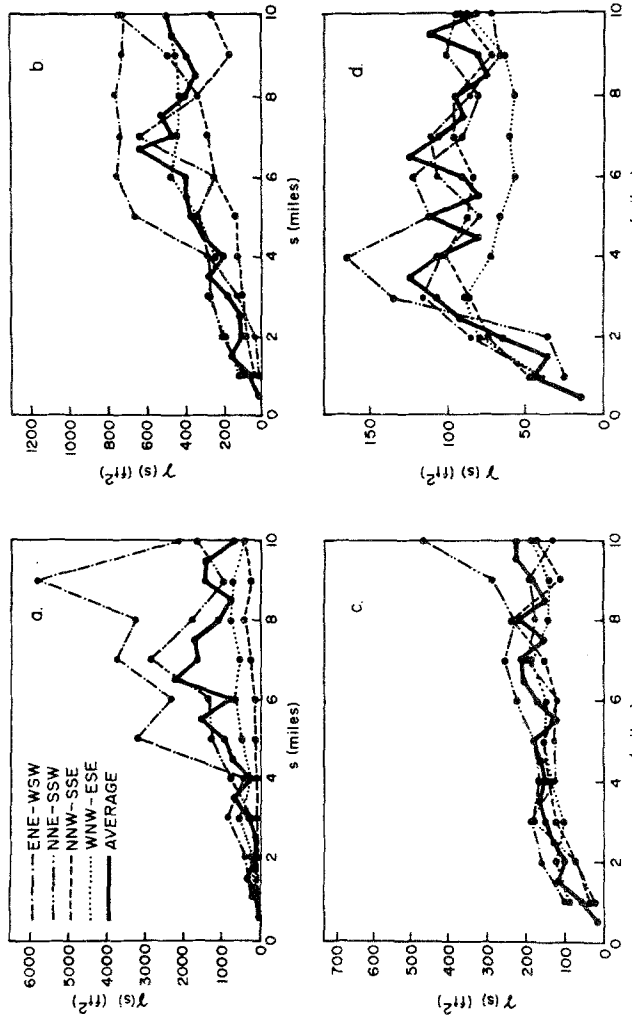

Fig. 3. Directional and average sample semivariograms of residuals for Avra Valley during Stage 1, based on (a) linear drift, $p = 1$, (b) quadratic drift, $p = 2$, (c) cubic drift, $p = 3$, and (d) 4th-order polynomial drift, $p = 4$.

adopt a 4th-order polynomial as an acceptable representation of the global drift, thereby completing Stage 1.

Stage 2 is a refinement accounting for the correlation structure of the residuals. The effect of this refinement on the average sample residual semivariogram is seen in Fig. 4. After only two generalized least-squares iterations, the sample residual semivariogram appears to have converged to a stable shape. Figure 4 shows the spherical model fitted to the final version of the sample semivariogram and used in all subsequent calculations.

The polynomial drift $\hat{\mu}_4(\mathbf{x})$, associated with the same spherical semivariogram model, is contoured in Fig. 5. The standard deviations of the corresponding drift estimation errors, computed with the aid of (24), are presented in Fig. 6.

Having estimated both the semivariogram and the drift, the next step is to obtain kriging estimates of the residuals at the finite element grid points and add them to the drift estimates according to (34). The resulting water level estimates are shown by the solid contours in Fig. 7. To compare our method with the generalized covariance technique described earlier, we applied BLUEPACK-3D (a package developed at the Ecole des Mines in Fontainebleau, France) to our data. The results are shown by the dashed contours in Fig. 7. Except for areas where data are missing (see Fig. 1), the results agree reasonably well.

The kriging errors associated with our water level estimates are plotted in Fig. 8. These are computed with the aid of (41) and represent estimation errors of the solid contours in Fig. 7. Figure 9 shows contours of kriging errors as computed by BLUEPACK-3D. The general configuration of the contours in the two figures is similar in many areas. The largest differences between the two maps occur near the bulge in the western boundary and the southern and north-

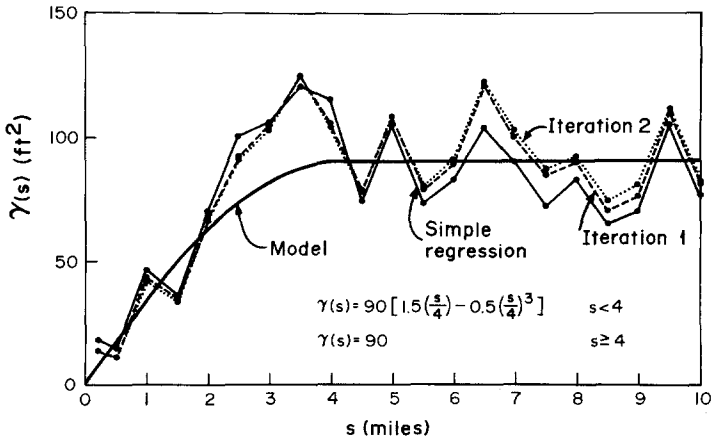

Fig. 4. Effect of iterative regression on the sample residual semivariogram for the Avra Valley with 4th-order polynomial drift. The solid curve shows a spherical model fitted to a sample semivariogram from the 2nd iteration.

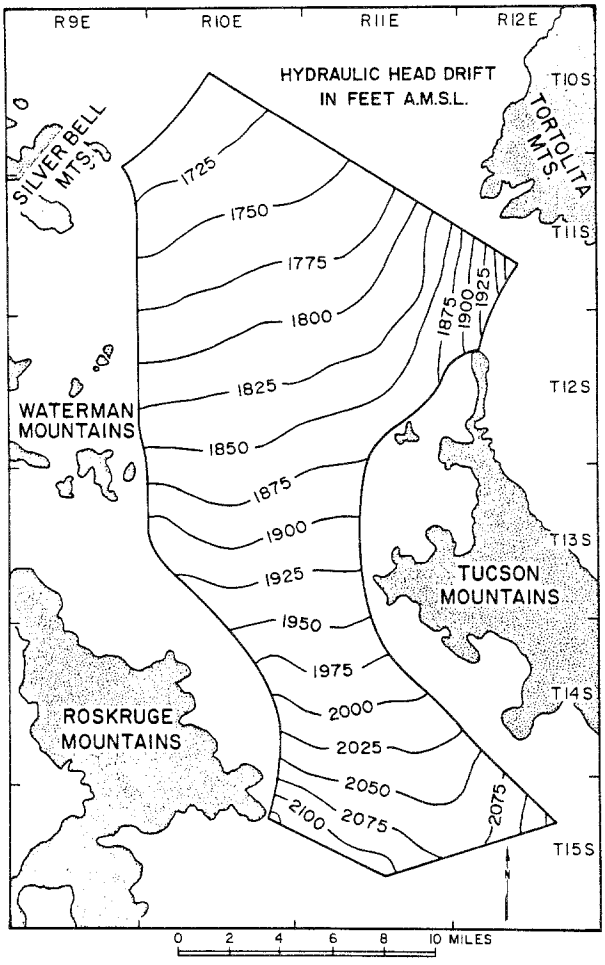

Fig. 5. Drift associated with the spherical semivariogram model for Avra Valley shown in Fig. 4.

eastern corners where the BLUEPACK-3D contours show steeper gradients and, therefore, larger errors than our contours. These areas include very few data points (see Fig. 1) and it is, therefore, natural that a local interpolator such as BLUEPACK-3D would associate them with large estimation errors. A global estimator such as ours interpolates across gaps in the data and thus tends to be more optimistic.

A comparison of Figs. 6 and 8 will reveal that the kriging errors in the southern and northeastern corners of the modeled area are smaller than the corresponding drift estimation errors. This simply shows that in these corners there

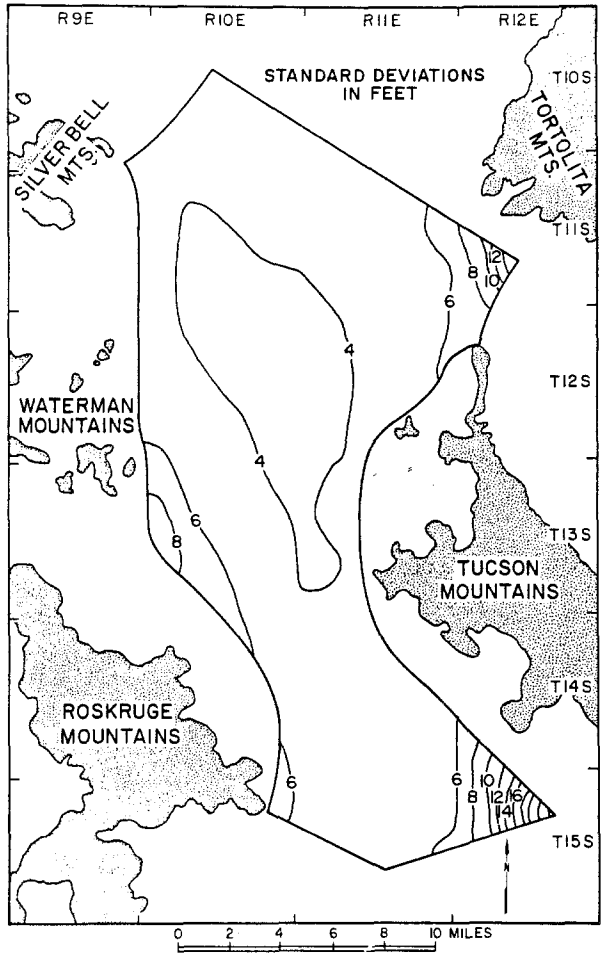

Fig. 6. Standard deviations of drift estimation errors for Avra Valley.

is less uncertainty in the estimation of Z than in the estimation of its components, μ and R.

To study the autocorrelation of our kriging errors, we consider the six traverses AA′, BB′, CC′, DD′, EE′, and FF′ shown in Fig. 1. Fig. 10 shows the autocorrelation along each of these traverses. The distance along which the errors are positively correlated varies between 1 and 2 miles, as compared with the 4-mile range of the spherical semivariogram model. Negative correlations are small or nonexistent. Unfortunately, our results cannot be compared with BLUEPACK-3D because autocovariances and autocorrelations are not included in its output.

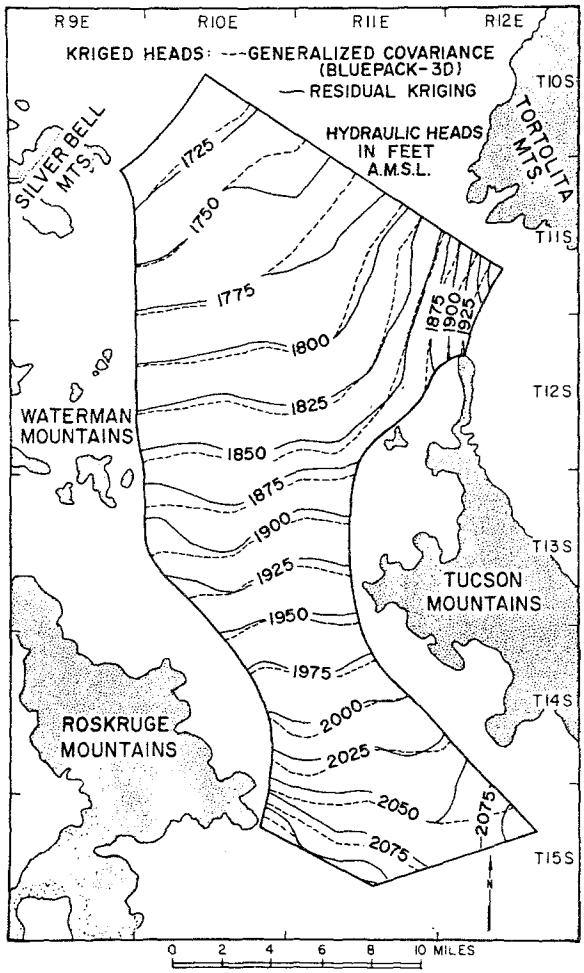

Fig. 7. Kriged water levels for Avra Valley. Dashed contours show estimates obtained with BLUEPACK-3D.

CONCLUSIONS

The following conclusions can be drawn from this paper:

(1) The drift and residual semivariogram of a nonintrinsic function can be determined simultaneously by stepwise iterative regression. The method is easy to implement and appears to converge rapidly.

(2) Once a prior estimate of the drift has been obtained, the function of interest can be estimated by residual kriging. The residual kriging equations differ

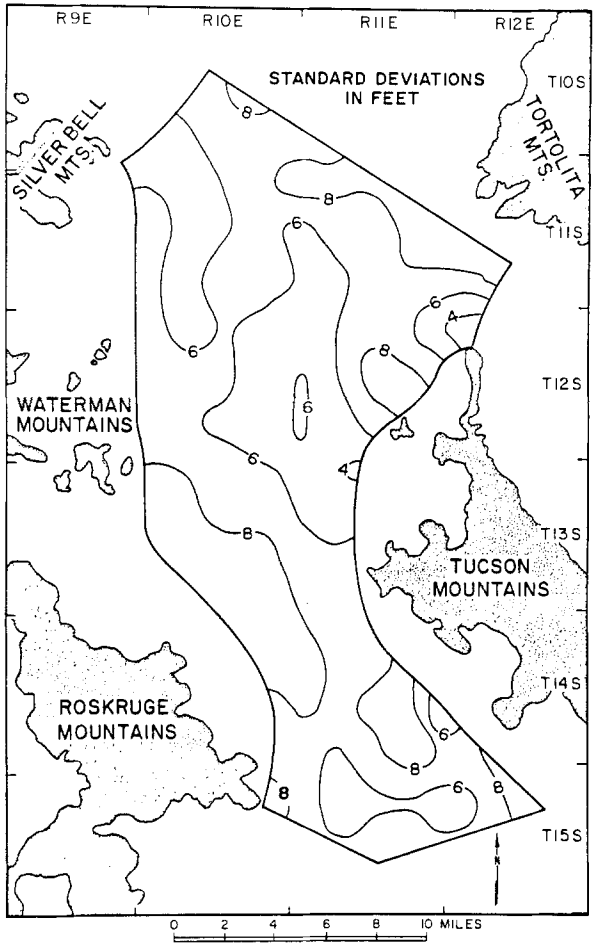

Fig. 8. Kriging errors for Avra Valley.

from their conventional counterparts in that their exact form involves the semivariogram of the residuals estimated by regression, rather than the unknown true residuals.

(3) Our estimates and estimation variances compare favorably with those obtained from the same data by means of the generalized covariance package BLUEPACK-3D. However, we were not able to compare the estimation error autocovariance and autocorrelation structures because these are not included in the output of BLUEPACK-3D.

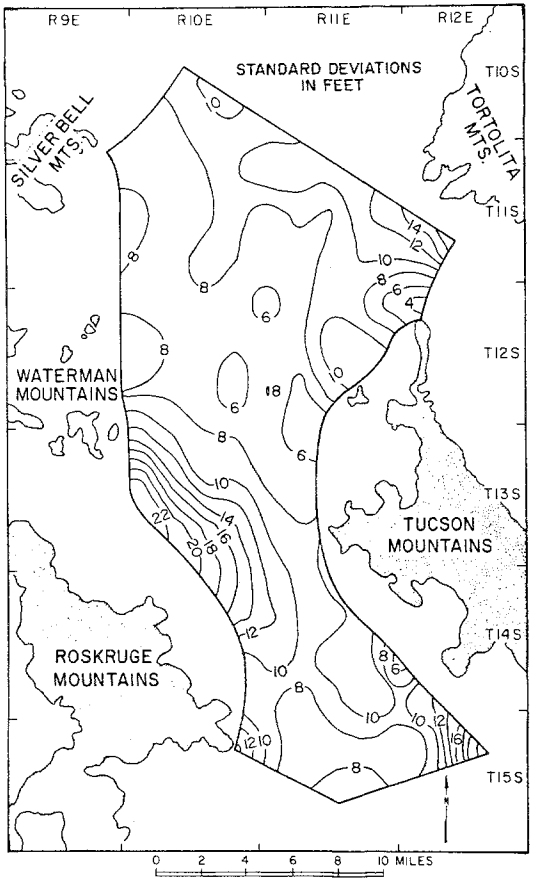

Fig. 9. Kriging errors for Avra Valley obtained with BLUEPACK-3D.

APPENDIX A

Let $\boldsymbol{P} = \boldsymbol{F}\boldsymbol{V}_a\boldsymbol{F}^T\boldsymbol{V}^{-1}$. Then $\boldsymbol{P}\boldsymbol{F} = \boldsymbol{F}\boldsymbol{V}_a\boldsymbol{F}^T\boldsymbol{V}^{-1}\boldsymbol{F} = \boldsymbol{F}\boldsymbol{V}_a\boldsymbol{V}_a^{-1} = \boldsymbol{F}$ and, thus

$$(\boldsymbol{I} - \boldsymbol{P})\boldsymbol{F} = \boldsymbol{0} \tag{A1}$$

A lower bound for the rank of $(\boldsymbol{I} - \boldsymbol{P})$ is established by noting that (Pearson, 1974, p. 903)

$$\begin{aligned} \text{rank}(\boldsymbol{P}) &= \text{rank}(\boldsymbol{F}\boldsymbol{V}_a\boldsymbol{F}^T\boldsymbol{V}^{-1}) \\ &\leqslant \min\,[\text{rank}(\boldsymbol{F}), \text{rank}(\boldsymbol{V}_a), \text{rank}(\boldsymbol{F}^T), \text{rank}(\boldsymbol{V}^{-1})] \\ &= J_p \end{aligned}$$

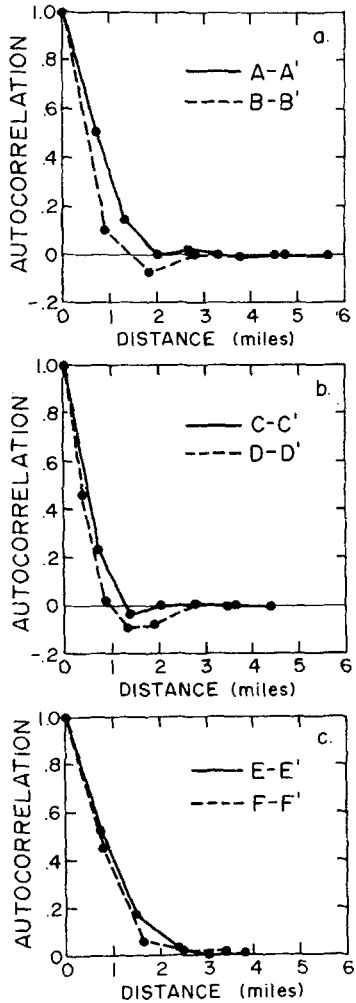

Fig. 10. Sample autocorrelation of estimation errors along selected traverses in Avra Valley (refer to Fig. 1).

since the columns of $\boldsymbol{F}$ are linearly independent. Thus, for $I \geqslant J_p$

$$\text{rank}(\boldsymbol{I} - \boldsymbol{P}) \geqslant \text{rank}(\boldsymbol{I}) - \text{rank}(\boldsymbol{P}) \geqslant I - J_p \tag{A2}$$

On the other hand

$$\begin{aligned} \boldsymbol{PVP}^T &= \boldsymbol{FV}_a\boldsymbol{F}^T\boldsymbol{V}^{-1}\boldsymbol{VV}^{-1}\boldsymbol{FV}_a\boldsymbol{F}^T = \boldsymbol{FV}_a\boldsymbol{F}^T \\ &= \boldsymbol{PV} \end{aligned} \tag{A3}$$

and since $\boldsymbol{PV}$ and $\boldsymbol{V}$ are both symmetric

$$\boldsymbol{PV} = (\boldsymbol{PV})^T = \boldsymbol{VP}^T \tag{A4}$$

APPENDIX B

Let $\epsilon_n = R_n^* - \hat{R}_n$. Then (35) implies that the covariance of ϵ is

$$E[\epsilon_n \epsilon_m] = E\left[\left(\sum_{i=1}^{I_n} \lambda_{ni}\hat{R}_{ni} - \hat{R}_n\right)\left(\sum_{j=1}^{I_m} \lambda_{mj}\hat{R}_{mj} - \hat{R}_m\right)\right]$$

$$= \sum_{i=1}^{I_n}\sum_{j=1}^{I_m} \lambda_{ni}\lambda_{mj}E[\hat{R}_{ni}\hat{R}_{mj}] - \sum_{i=1}^{I_n} \lambda_{ni}E[\hat{R}_{ni}\hat{R}_m]$$

$$- \sum_{j=1}^{I_m} \lambda_{mj}E[\hat{R}_{mj}\hat{R}_n] + E[\hat{R}_n\hat{R}_m] \tag{B1}$$

If the kriging coefficients satisfy (39) then, by virtue of (25), (B1) reduces to (38). The kriging equations (40) are obtained by minimizing

$$E[\epsilon_n^2] + \beta_n\left(\sum_{i=1}^{I_n} \lambda_{ni} - 1\right)$$

with respect to λ_{ni} and the Lagrange multiplier β_n.

In the case of universal kriging, the kriging variance can be expressed by virtue of (44) as

$$E[(Z_n^U - Z_n)^2] = E\left[\left(\sum_{i=1}^{I_n} \lambda_{ni}Z_{ni} - Z_n\right)^2\right]$$

$$= E\left[\left(\sum_{i=1}^{I_n} \lambda_{ni}Z_{ni} - \sum_{i=1}^{I_n} \lambda_{ni}\mu(\mathbf{x}_{ni}) - Z_n + \mu_n\right)^2\right]$$

$$= E\left[\left(\sum_{i=1}^{I_n} \lambda_{ni}R_{ni} - R_n\right)^2\right] = E[\epsilon_n^2] \tag{B2}$$

The universal kriging equations (46) are obtained by minimizing

$$E[\epsilon_n^2] + \sum_{j=1}^{J_p} \beta_{nj}\left[\sum_{i=1}^{I_n} \lambda_{ni}f_j(\mathbf{x}_{ni}) - f_j(\Gamma_n)\right]$$

with respect to λ_{ni} and the Lagrange multipliers β_{nj}. If, instead of (44), one uses (45) in (B2), the result is a set of universal kriging equations equivalent to (46) but expressed in terms of $\gamma_{\hat{R}}$ instead of γ_R.

ACKNOWLEDGMENT

The BLUEPACK-3D results in this paper were obtained during a brief service of the senior author as visiting Directeur de Recherche at the Centre d'Informatique Geologique, Ecole des Mines de Paris, Fontainebleau, France, in the summer of 1981. These results could not have been obtained without the active cooperation of Mrs. H. Beucher and the support of Dr. G. de Marsily, the Centre's director, to both of whom we are deeply indebted. We are also grateful to Dr. F. Szidarovszky of the University of Agriculture in Budapest, Hungary, and Mr. Jesús Carrera of the Department of Hydrology and Water Resources at the University of Arizona, Tucson, for their valuable comments.

REFERENCES

Beucher, H., Delhomme, J. P., and de Marsily, G., 1981, Analyse stochastique des proprietes des systems poreux naturels heterogenes: Centre d'Informatique Geologique, Ecole des Mines de Paris, Fontainebleau, France, 115 p.

Binsariti, A. A., 1980, Statistical analysis and stochastic modeling of the Cortaro aquifer in southern Arizona: Unpublished Ph.D. dissertation, Department of Hydrology and Water Resources, University of Arizona, Tucson, Arizona, 243 p.

Chiles, J. P., 1977, Geostatistique des phenomenes non stationnaires: Unpublished Docteur-Ingenieur thesis, University of Nancy, France.

Clifton, P. M., 1981, Statistical inverse modelling and geostatistical analysis of the Avra Valley aquifer: Unpublished M.S. thesis, Department of Hydrology and Water Resources, University of Arizona, Tucson, Arizona, 190 p.

Clifton, P. M. and Neuman, S. P., 1982, Effects of kriging and inverse modeling on conditional simulation of the Avra Valley aquifer in southern Arizona: Water Resour. Res., v. 18, n. 4, p. 1215–1234.

David, M., 1977, Geostatistical ore reserve estimation: Elsevier, New York, 364 p.

Delfiner, P., 1976, Linear estimation of non-stationary spatial phenomena, *in*, Guarascio, M., David, M., and Huijbregts, C. J., (Eds.), Advanced geostatistics in the mining industry: Proceedings of the NATO Advanced Study Institute: P. Reidel Publishing Company, Dordrecht, The Netherlands, p. 49–68.

Delhomme, J. P., 1976, Application de la theorie des variables regionalisees dan les sciences de l'eau: Unpublished Docteur-Ingenieur thesis, Universite Pierre et Marie Curie, Paris, France, 130 p.

Fennessy, P. J., 1982, Geostatistical analysis and stochastic modeling of the Tajo Basin aquifer, Spain: Unpublished M.S. thesis, Department of Hydrology and Water Resources, University of Arizona, Tucson, Arizona, 124 p.

Gambolati, G. and Volpi, G., 1979, A conceptual deterministic analysis of the kriging technique in hydrology: Water Resour. Res., v. 15, n. 3, p. 625–629.

Haas, A. and Jousselin, C., 1976, Geostatistics in petroleum industry, *in* Guarascio, M., David, M., and Huijbregts, C. J., (Eds.), Advanced geostatistics in the mining industry, Proceedings of the NATO Advanced Study Institute: P. Reidel Publishing Company, Dordrecht, The Netherlands, p. 333–347.

Huijbregts, C. and Matheron, G., 1971, Universal kriging (an optimal method for estimating and contouring in trend surface analyses): Can. Inst. Min. Metall., Spec. v. 12, p. 159–169.

Journel, A. G. and Huijbregts, C. J., 1978, Mining geostatistic: Academic Press, New York, 600 p.

Matheron, G., 1969, Le krigeage universel: Le cahiers du Centre de Morphologie Mathematique, Fasc. No. 1: Ecole des Mines de Paris, Fontainebleau, France.

Matheron, G., 1971, La theorie des F.A.I.G.: Note Geostatistique u°117: Centre de Morphologie Mathematique, Ecole des Mines de Paris, Fontainebleau, France.

Matheron, G., 1973, The intrinsic random functions and their applications: Adv. Appl. Prob., v. 5, p. 439–468.

Neuman, S. P. and Yakowitz, S., 1979, A statistical approach to the inverse problem of aquifer hydrology, 1, Theory: Water Resour. Res., v. 15, n. 4, p. 845–860.

Pearson, C. E., 1974, Handbook of Applied Mathematics: Van Nostrand Reinhold Company, New York, 1265 p.

Ripley, B. D., 1981, Spatial statistics: John Wiley & Sons, New York, 252 p.

Sabourin, R., 1976, Application of two methods for the interpretation of the underlying variogram, *in* Guarascio, M., David, M., and Huijbregts, C. J., (Eds.), Advanced geostatistics in the mining industry, Proceedings of the NATO Advanced Study Institute: P. Reidel Publ. Co., Dordrecht, The Netherlands, p. 101–109.

Schweppe, F. C., 1973, Uncertain dynamic systems: Prentice-Hall, Englewood Cliffs, New Jersey, 563 p.

Seber, G. A. F., 1977, Linear regression analysis: John Wiley & Sons, New York, 465 p.