Engineering Geology and Geotechnical Engineering, Watters (ed.),
© 1989 Balkema, Rotterdam. ISBN 90 6191 878 2

Geostatistical techniques to estimate collapse-related soil parameters

Molla M.Ali
Western Technologies, Inc., Phoenix, Ariz., USA

Edward A.Nowatzki & Donald E.Myers
University of Arizona, Tucson, Ariz., USA

ABSTRACT: Application of the theory and concepts of geostatistics to geotechnical problems is usually difficult because of inadequate sample size. A successful application is demonstrated here in which approximately 1,000 sample values for collapse-related soil parameters obtained from reliable sources were used. Geostatistical methods were applied to seven sets of data obtained from over 400 locations within the Tucson Basin in Arizona. The variables were found to have an isotropic spatial structure which could be fitted by a spherical model. Kriging, which is a local averaging method of spatial prediction, was applied to estimate values for the collapse-related soil parameters in unsampled locations. Kriging is an optimal method which provides estimates for unsampled locations without bias and with known minimum variance. Results are presented in the form of contour plots of values for two parameters; contour plots of the associated variance are also presented.

1 INTRODUCTION

The use of probabilistic models and statistical techniques in various disciplines of geotechnical engineering has increased rapidly in the recent past. These techniques are not limited to data analysis alone but include reliability assessment of earth structures and other constructed facilities, risk assessment for regulatory control, economic optimization and project feasibility determination (Beacher, 1984). In performing these tasks, geotechnical engineers use theories of statistics and probability for an assessment of the uncertainty in the prediction of the performance of the structures which have been designed on the basis of the average values obtained for design parameters from a limited number of laboratory and/or field test results. As a result of small sample sizes, the statistical estimation error plays an important role in geotechnical reliability.

In order to examine the variability of a given parameter over an area, classical statistical methods rely upon an analysis of the variance which quantifies the variability without regard to direction. Generally, the sample size necessary to estimate the mean within some specified confidence interval is calculated on the basis of a frequency distribution of the observations (Vieira et al., 1981). Regardless of the sampling plan, these methods provide an incomplete description of the variability of the property since there is no link between the calculated variances and the distance between observations. In other words, a knowledge of the frequency distribution of the observations alone does not provide any information about the variability of the observations with respect to spatial locations. The spatial variability of a soil parameter given in terms of the position coordinates provides additional characterization not derived from the frequency distribution of observations alone. The statistical treatment known as "geostatistics" provides the tools necessary for an adequate description of the spatial variability of a parameter as well as an unbiased estimation of its value at unsampled locations. Only the coordinates of the sample locations are utilized and no knowledge of the theoretical frequency distribution is required.

The theory of regionalized variables provides the foundation of the entire field of geostatistics (Journel and

Huijbergts, 1978). Modelling of variograms is the first and most important step in applying the technique of Kriging, which is the method used here for obtaining unbiased estimates of parameters in unsampled locations. A considerable amount of computation is necessary to obtain an adequate estimate of the variogram because of the empirical and subjective nature of the estimation process.

Theoretically, the variogram of Z(x) is defined by Knudsen and Kim (1978) as:

$$\gamma(h) = \frac{1}{2}\mathrm{Var}[Z(x+h) - Z(x)] \quad (1)$$

where Z(x) is a particular parameter value at point x and Z(x + h) is the parameter value at a point located a distance h from x. The formal definition of the variogram is given by:

$$\gamma(h) = \frac{1}{2V}\int_V [Z(x+h) - Z(x)]^2\, dx \quad (2)$$

The expression for $\gamma(h)$ applies to one-, two-, or three-dimensional space. In practice, variograms are computed from a discrete number of points obtained using an incremental distance. Therefore, Equation 2 can be written as:

$$\gamma(h) = \frac{1}{2N_h}\sum_{i=1}^{N_h}[Z(x_i) - Z(x_i+h)]^2 \quad (3)$$

for N_h pairs of samples, each being separated by a distance h. It is assumed that all samples lie on a straight line along which the computation is being performed. The basic unit, h, used for the interval in Equation 3 is known as the "class size." A tolerance on the class size is also required so that data points closer to one class interval than another can be considered during computation.

2 MODELING OF VARIOGRAMS

The first step in variogram modeling is to define the purpose of the evaluation. This must be done in order to decide which variables are of interest. In this study, the purpose as mentioned earlier, was to determine whether or not selected collapse-related soil parameters had any particular structure. The variogram was utilized to estimate values of the parameters at unsampled locations and to produce a probability contour plot using the method of Indicator Kriging.

The parameters of interest are listed in Table 1. The various data sets containing values of these parameters are listed in Table 2. Representative variograms were obtained for each of the parameters in each of the seven data sets, but only a few will be presented here. All of the geostatistical computations including variogram estimation and Kriging were performed by using the computer program BLUEPACK developed by Centre de Geostatistique, Fontainebleau, France. Since the

Table 1. Parameters of interest in this study.

Symbol	Definition	Type*
C_p	Percent collapse (Jennings and Knight, 1957)	CC
γ_d	Dry unit weight	CP
n_o	In situ porosity	CP
s_o	In situ degree of saturation	CP
w_o	In situ moisture content	CP
e_o	In situ void ratio	CP
PL	Plastic limit	CP
R	Gibbs' parameter (Gibbs, 1961)	CC
A	Alfi parameter (Alfi, 1984)	CC

* CC = collapse criterion
CP = collapse-related parameter

Table 2. Data sets used in the analysis.

Data Set Number	Range of Depths, ft	Number of Data N
1	0.0-1	125
2	1.0-2	286
3	2.0-3	254
4	3.0-4	100
5	4.0-6	104
6	6.0-40	123
7	0.0-40	219*

* Data from other sets also containing values for three additional parameters: R, A, and PL.

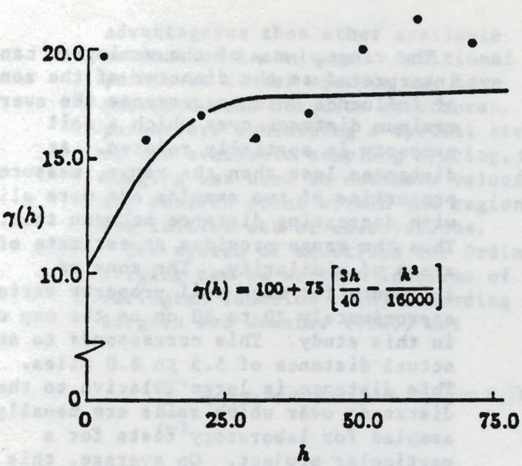

Figure 1. Semi-variogram and fitted equations for C_p of Data Set 5.

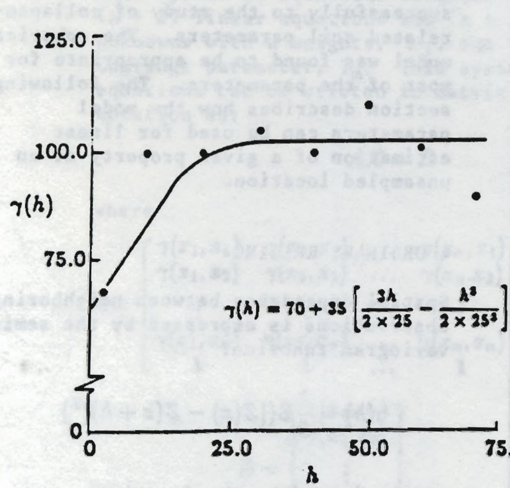

Figure 2. Semi-variogram and fitted equation for γ_d of Data Set 5.

modeling of a variogram is, in part, an art requiring some subjective judgment, multiple trials are usually necessary in order to obtain a satisfactory variogram. The important parameters for a variogram are the range of influence, a, and the sill C. Figures 1 and 2 show the variograms for C_p and γ_d of Data Set 5, respectively. In several cases a pure "nugget effect" model was obtained, indicating a complete lack of geologic structure. Figure 3 shows such a model for the parameter e_o of Data Set 6.

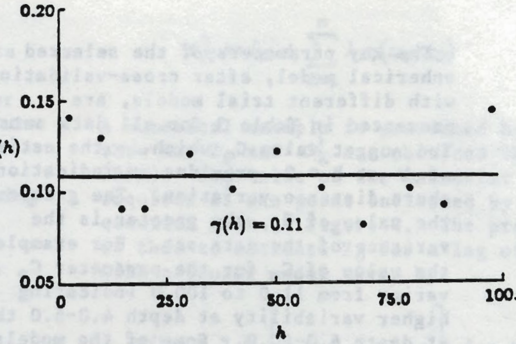

Figure 3. Semi-variogram and fitted pure nugget model for e_o of Data Set 6.

3 FITTING A THEORETICAL VARIOGRAM MODEL

The final step in constructing a variogram is to find the theoretical model that best describes the experimental variogram. Choosing a theoretical model and estimating its parameters are generally done subjectively. The choice is often limited to linear or spherical models, with a spherical model being the most common.

Fitting and subsequently using a model are usually done arbitrarily. This does not alter the results of the geostatistical calculations. What is important is that the chosen theoretical model $\gamma(h)$ should be a good fit to the experimental semi-variogram within the model's limits of reliability (Journel and Huijbregts, 1978). The choice of a theoretical model is generally made by examining the experimental variogram and taking into account the fact that variograms are subject to significant fluctuations at large distances.

It must be realized that the true variograms of a soil deposit can never be known, but it is expected that the experimental variograms will reflect the underlying theoretical variogram. Since most of the experimental variograms could be approximated by a spherical model, such a model was fitted to all the computed variograms in this study. Theoretically, the selected model can be fitted using a least-squares method, but this does not ensure an optimal selection. According to David (1977), a visual fit by hand is usually sufficient. If least-squares fitting is used, weightings must be used.

The key parameters of the selected spherical model, after cross-validation with different trial models, are presented in Table 3 for all data sets. The nugget value C_o which is the estimate of γ at h = 0, provides an indication of short distance variation. The greater the value of C_o, the greater is the variance of the data set. For example, the value of C_o for the parameter C_p varies from 11.0 to 100.0 indicating higher variability at depth 4.0-6.0 than at depth 6.0-40.0. Some of the models (such as that for e_o in Data Set 1) show a "pure nugget" effect indicating lack of spatial correlation.

Table 3. Parameters of spherical models fitted to each data set.

Data Set	Parameter Z	Nugget C_0	Range a	Sill C	Mean Error $(Z^* - Z)$
	C_p	30.0	20.0	45.0	0.0045
	γ_d	100.0	20.0	110.0	−0.0049
	n_0	38.0	25.0	45.0	−0.0061
1	s_0	185.0	12.5	120.0	−0.0157
	w_0	0.0015	20.0	0.0018	0.0003
	e_0	0.073	—	—	—
	C_p	18.5	—	—	—
	γ_d	80.0	35.0	100.0	0.0057
2	n_0	32.4	35.0	45.0	0.0125
	s_0	102.0	30.0	148.0	0.0576
	w_0	0.002	25.0	0.0025	0.0002
	e_0	0.042	30.0	0.053	0.0000
	C_p	13.0	25.0	16.0	−0.0039
	γ_d	85.0	45.0	140.0	0.0356
	n_0	40.0	35.0	57.5	0.0246
3	s_0	155.0	—	—	—
	w_0	0.0021	40.0	0.0029	0.0011
	e_0	0.047	30,0	0.065	0.0006
	C_p	14.0	20.0	17.0	0.0356
	γ_d	85.0	27.0	120.0	−0.0793
4	n_0	40.0	27.5	65.0	0.0915
	s_0	140.0	12.0	170.0	0.0896
	w_0	0.005	—	—	—
	e_0	0.09	17.0	0.12	0.0070
	C_p	100.0	20.0	175.0	0.0557
	γ_d	70.0	25.0	105.0	−0.3190
	n_0	35.0	30.0	50.0	0.1769
5	s_0	130.0	20.0	162.0	0.0293
	w_0	0.0016	25.0	0.0028	0.0002
	e_0	0.04	25.0	0.06	0.0061
	C_p	11.0	30.0	16.5	0.0970
	γ_d	150.0	—	—	—
6	n_0	75.0	—	—	—
	s_0	150.0	20.0	350.0	−0.1433
	w_0	0.05	25.0	0.08	−0.0015
	e_0	0.11	—	—	—
	C_p	23.0	20.0	27.0	0.1363
	γ_d	110.0	—	—	—
	n_0	45.0	—	—	—
	s_0	140.0	35.0	200.0	0.0258
7	w_0	0.005	—	—	—
	e_0	0.07	—	—	—
	PL	27.5	25.0	82.0	0.0007
	R	0.11	12.0	0.16	−0.0037
	A	21.0	20.0	35.0	−0.0041

The range, a , of the variogram can be interpreted as the diameter of the zone of influence which represents the average maximum distance over which a soil property is spatially related. At distances less than the range, measured properties of two samples are more alike with decreasing distance between them. Thus the range provides an estimate of areas of similarity. The zone of influence for each soil property varies approximately 20 to 30 cm on the map used in this study. This corresponds to an actual distance of 5.5 to 8.0 miles. This distance is large relative to the distances over which soils are usually sampled for laboratory tests for a particular project. On average, this distance also represents the minimum distance at which maximum variation occurs.

The above discussion suggests that geostatistical concepts can be applied successfully to the study of collapse-related soil parameters. The spherical model was found to be appropriate for most of the parameters. The following section describes how the model parameters can be used for linear estimation of a given property at an unsampled location.

4 ORDINARY KRIGING

Spatial dependence between neighboring observations is expressed by the semi-variogram function:

$$\gamma(h) = \frac{1}{2}\mathrm{E}\{[Z(x) - Z(x+h)]^2\} \tag{4}$$

and estimated by the function:

$$\gamma^*(h) = \frac{1}{2N_h}\sum_{i=1}^{N_h}[Z(x+h) - Z(x)]^2 \tag{5}$$

These functions were investigated for all parameters. The experimental variogram of most of the parameters was fitted by a spherical model with range a, sill C, and a nugget value C_o. Measurements separated by distances smaller than the range were considered spatially correlated. For some parameters the variogram showed complete discontinuity at the origin resulting in a pure nugget effect which indicated a total lack of any geologic structure over the spacing interval measured. For these parameters Kriging techniques are not more advantageous than other available estimation techniques. Additional data points at closer spacing may have provided some adequate structures. For parameters exhibiting a spatial structure at the available sampling spacing, Kriging was used to estimate values at any desired point within the region using the initial set of observations.

The system of equations for Ordinary Kriging can be written in terms of the variogram function $\gamma(h)$ according to Burgess and Webster (1980) as:

$$\sum_{i=1}^{n} \lambda_i \gamma(x_i - x_j) + l_m = \gamma(x_0 - x_j) \qquad (6a)$$

$$\sum_{i=1}^{n} \lambda_i = 1 \qquad (6b)$$

This system of equations consists of (n + 1) linear equations and (n + 1) unknowns with n weights, λ_i, and the Lagrange parameter, l_m. This system of equations can be written in matrix notation as:

$$AX = B \qquad (7)$$

where

$$A = \begin{bmatrix} \gamma(x_1,x_1) & \gamma(x_2,x_1) & \dots & \gamma(x_n,x_1) & 1 \\ \gamma(x_1,x_2) & \gamma(x_2,x_2) & \dots & \gamma(x_n,x_2) & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \gamma(x_1,x_n) & \gamma(x_2,x_n) & \dots & \gamma(x_n,x_n) & 1 \\ 1 & 1 & \dots & 1 & 0 \end{bmatrix} \qquad (7a)$$

$$B = \begin{bmatrix} \gamma(x_1,x_0) \\ \gamma(x_2,x_0) \\ \vdots \\ \gamma(x_n,x_0) \\ 1 \end{bmatrix} \qquad (7b)$$

and

$$X = \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \\ l_m \end{bmatrix} \qquad (7c)$$

The solution of the system of equations AX = B is of the form:

$$X = A^{-1}B$$

The estimation at an unknown location is

$$Z_0^* = \sum \lambda_i Z_i \qquad (8)$$

$$\sigma_K^2 = l_m + \sum_{i=1}^{n} \lambda_i \gamma(x_1, x_0) \qquad (9)$$

A numerical example is presented here to show how Z_0 and σ_K^2 are obtained (Myers, 1985). The value of the parameter Z is required at the point indicated by the question mark in Figure 4. The problem is then to estimate Z_0 for a lag of 20 with a linear model given by:

$$\gamma(h) = 7.8 \left[0.4 + 0.6 \frac{h}{400} \right] \quad \text{for } h < 400$$

and

$$\gamma(h) = 0 \quad \text{for } h \geq 400$$

From the lag distance and the given model, the elements of matrix A can be calculated as follows:

$$\gamma(x_1,x_1) = 7.8(0.4) = 3.12 = \gamma(x_2,x_2)$$

$$\gamma(x_1,x_2) = 7.8 \left[0.4 + 0.6 \frac{40}{400} \right]$$

$$= 7.8(0.46)$$

$$= 3.588 = \gamma(x_2,x_1)$$

$$\gamma(x_1,x_0) = 7.8 \left[0.4 + 0.6 \frac{20}{400} \right]$$

$$= 7.8(0.43)$$

$$= 3.354 = \gamma(x_2,x_0)$$

Then

$$AX = B$$

$$\begin{bmatrix} 3.120 & 3.588 & 1.000 \\ 3.588 & 3.120 & 1.000 \\ 1.000 & 1.000 & 0.000 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ l_m \end{bmatrix} = \begin{bmatrix} 3.354 \\ 3.354 \\ 1.000 \end{bmatrix}$$

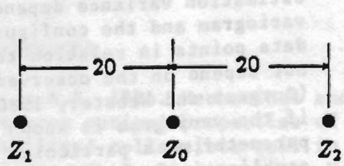

Figure 4. Geometry of a hypothetical problem to estimate unknown value by Kriging.

The solution to these equations can be obtained by inspection as

$$\lambda_1 = 0.5$$

$$\lambda_2 = 0.5$$

$$l_m = 0.0$$

Therefore from Equation 8

$$Z_o^* = \sum \lambda_i A_i = 0.5Z_1 + 0.5Z_2$$

and from Equation 9

$$\sigma_K^2 = l_m + \sum \lambda_i \gamma(x_1, x_0)$$

$$= 0.0 + 0.5 \times 3.354 + 0.5 \times 3.354$$

$$= 3.354$$

Thus the estimation at the required point is obtained as Z_o with known estimation variance equal to 3.354. Therefore, the estimated value can be applied to practical problems with known confidence.

When the number of available measurements n is not too large, matrix A is the same regardless of the position of the point, x_0, to be estimated. However, the matrix B changes for every location. Therefore, only one inversion of matrix A is necessary to obtain weights λ_i and l_m in order to estimate any number of values within the area.

In cases where the number of available measurements is large, and the distance beyond which the variogram is not known is less than one-half the largest distance of the sampled area, an estimation neighborhood must be selected. The radius of the neighborhood is varied until enough sampled points are included to provide an acceptable variogram.

One advantage of Kriging over other interpolation methods is that the estimation variance can be calculated before the actual sampling is made. The estimation variance depends on the semi-variogram and the configuration of the data points in relation to x_0. It does not depend on the observed value Z (Burgess and Webster, 1980). Therefore, if the variogram is known for a given parameter at a particular location, sampling intervals for the desired variance of estimation can be selected before actual samples are taken at the site.

5 PRESENTATION OF RESULTS

The mechanism by which an interpolated surface is displayed is essentially a separate problem from that of the interpolation itself, especially if the display is to be produced on an automatic plotter. In most automatic contouring systems a fine grid of values is generated to represent the surface, usually by methods that are not optimal. For example, in the program SURFACE II (Sampson, 1975) contour lines are developed by linear interpolation between grid nodes. The points where a contour line of specified value crosses the edge of a grid cell are located in this way. The string of successive X and Y coordinates of these intersections define the contour line to be drawn. Figure 5 shows contour plots of γ_d for Data Set 2 as obtained from program SURFACE II. Areas where no data points are available are easily detected by discontinuities in the contour line.

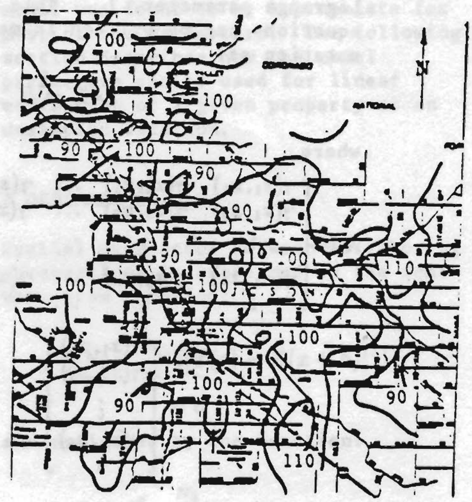

Figure 5. Contour plots of available values for γ_d (pcf) of Data Set 2.

Results of ordinary Kriging with known variance of estimation have been obtained by the use of the program BLUEPACK. The contour plots were generated from 1,763 points estimated by Kriging on an arbitrarily chosen grid size that was found to produce satisfactory contour plots for the scale used. Contour plots are largely dependent on the size of the contour interval which is also dependent on the actual scaled dimension of the ordinates. Figures 6 and 7 show the contour plots of the estimated values and associated Kriging variance respectively of the parameter γ_d of Data Set 2. If the critical values of γ_d are known, areas containing high, medium or low collapse-susceptible soils can be obtained from these contour plots with known confidence.

6 SUMMARY AND CONCLUSION

Geostatistical techniques were applied to collapse-related soil parameters determined by laboratory tests on soils from Tucson, Arizona. The purpose of the study was to model associations among the variables, to investigate the structure of spatial dependency, and to estimate the probability that the value of a certain parameter at a given location is above or below a critical value that defines collapse susceptibility. Based on the results of this investigation, the following conclusions can be drawn:

1. The principles of geostatistics can be applied successfully to geotechnical engineering problems where large amounts of data are available from a reliable source.
2. Geostatistics was found to be a valuable tool for characterizing and modeling the spatial variability of geotechnical parameters.
3. Collapse-related soil parameters can be considered as regionalized variables having spatial structures that can best be fitted by a spherical model variogram. In this study, the range of influence of the structure varied from 5.5 to 8.0 miles. Since this distance is large relative to the usual distances over which soils are sampled, the application of geostatistical concepts for estimation of collapse-related soil parameters at unsampled locations is justified.
4. The method of Ordinary Kriging provides a means for estimating soil properties at an unsampled loction with known variance of estimation. Therefore, the method can be applied to estimate any collapsing soil parameter in Tucson with a known degree of confidence.

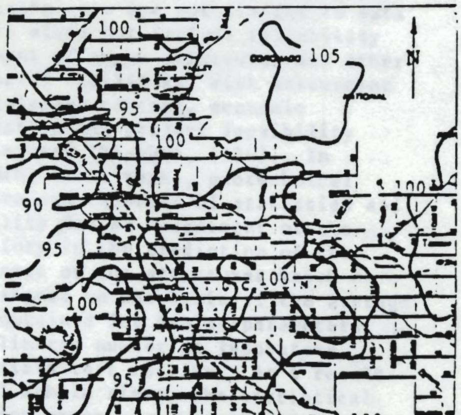

Figure 6. Contour plots of estimated values for γ_d (pcf) of Data Set 2 by ordinary Kriging.

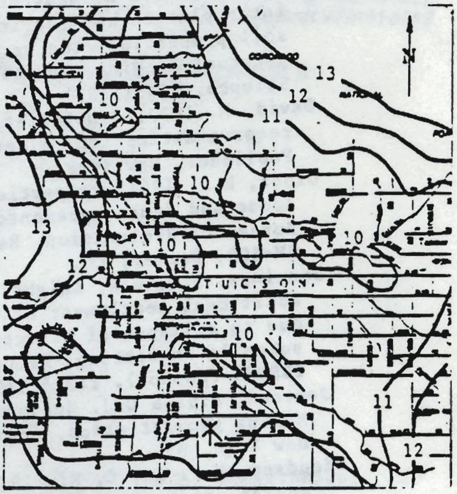

Figure 7. Contour plots of associated Kriging variance for γ_d (pcf) of Data Set 2 by ordinary Kriging.

7 REFERENCES

Alfi, A.A.S. 1984. Mechanical and electron optical properties of a stabilized collapsible soil in Tucson, Arizona, Ph.D. Dissertation, University of Arizona, Tucson, AZ.

Beacher, G.B. 1984. Geostatistics, reliability, and risk assessment in geotechnical engineering, Geostatistics for Natural Resource Characterization, G. Verly et al (eds.), D. Reidel, pp. 731-734.

Burgess, T.M. & R. Webster 1980. Optimum interpolation and isarithmic mapping of soil properties, the variogram and punctual Kriging, Journal of Soil Science, 31:315-331.

David, M. 1977. Geostatistical ore reserve estimation, Elsevier Science Publisher, New York.

Gibbs, H.J. 1961. Properties which divide loose and dense uncemented soils, U.S. Bureau of Reclamation, Report No. EM-608, Denver, CO.

Jennings, J.E. & K. Knight 1957. The additional settlement of foundations due to collapse of structure of sandy subsoils on wetting, Proceedings, 4th ICSMFE (London), 1:316-319.

Journel, A.J. & C.J. Huijbregts 1978. Mining geostatistics, Academic Press, New York, NY.

Knudsen, H.P. & Y.C. Kim 1978. A short course on geostatistical ore reserve estimation, Department of Mining and Geological Engineering, University of Arizona, Tucson, AZ.

Myers, D.E. 1985. Advanced geostatistics, unpublished class notes, Department of Mathematics, University of Arizona, Tucson, AZ.

Sampson, P.J. 1975. Series on spatial analysis, Kansas Geological Survey, Lawrence, KS.

Vieira, S.R., Nielson, D.R. & J.W. Bigger 1981. Spatial variability in field measured infiltraion rate, Journal of the Soil Science Society of America, 45:1040-1048.