

Geostatistical methods to predict collapsing soils

Les méthodes géostatistiques de prévision des sols collapsibles

M.M.ALI, Staff Engineer, Western Technologies, Inc., Phoenix, Arizona, USA

E.A.NOWATZKI, Associate Professor, Department of Civil Engineering and Engineering Mechanics, University of Arizona, Tucson, Arizona, USA

D.E.MYERS, Professor, Department of Mathematics, University of Arizona, Tucson, Arizona, USA

SYNOPSIS: An application of the theory and methods of geostatistics for predicting the distribution of collapse-susceptible soils in Tucson, Arizona is described. Selected collapse-related soil parameters were derived from laboratory test results for 992 samples retrieved at various depths from 411 different locations in the city. These parameters and sample location grid coordinates were organized into seven different data sets on the basis of depth to form a massive data base. Conventional statistical analyses consisting of descriptive statistics, probability distribution analyses, regression analyses, and factor analyses were performed on the data in each set to establish the reliability of each parameter as an indicator of collapse susceptibility. The methods of ordinary and indicator kriging were then used to estimate values of each parameter in areas where no test data were available, and to develop contour plots showing: 1) the estimated probability that the value of the parameter under consideration was above or below pre-defined cutoff values corresponding to various degrees of collapse susceptibility, and 2) the associated kriging variance. Results are presented in the paper for the collapse criterion C_p , which is defined as the percent collapse a soil sample undergoes in an oedometer following saturation under load.

1. INTRODUCTION

The virtually limitless variety of soil materials encountered in nature leads to uncertainty in defining their engineering properties. A statistical approach to defining soil properties provides a rational basis to achieve more economical solutions to geotechnical engineering problems by avoiding the use of extreme values in analysis and design and by quantifying the uncertainty involved.

The properties of collapse susceptible soils and their relation to foundation problems in Tucson, Arizona, have been investigated extensively (e.g., Ali, 1987; Anderson, 1968; Crossley, 1969; Abdullatif, 1969; Nowatzki, 1980; Sabbagh, 1982; Sultan, 1969). Previous studies, however, were limited either to specific areas within the City or to specific soil parameters. The purpose of this study was first to apply statistical techniques to a wide range of collapse criteria and collapse-related soil parameters in order to determine which criteria and parameters are the most reliable indicators of collapse susceptibility and then to obtain the probability distribution for each of those indicators within the Tucson Basin without bias and with known variance.

In order to accomplish these goals, field and laboratory test data for 922 sample points at 411 different locations within the City were collected from the job files of local consulting engineers and from the reports of previous researchers. These data were reduced and organized into a massive data base containing the following information for each sample:

1. Coordinates of the sampling location on a grid corresponding to the street plan of the City of Tucson.
2. Sample depth (D) from surface to 12 m.
3. Collapse-related soil parameters:
 - a. Insitu dry unit weight (γ_d)
 - b. Insitu moisture content (w_o)
 - c. Insitu void ratio (e_o)
 - d. Insitu porosity (n_o)
 - e. Insitu degree of saturation (S_o)
 - f. Plastic limit (PL)
4. Collapse criteria:
 - a. Gibbs' (1961) collapse ratio (R)
 - b. Alfi's (1984) collapse parameter (A)
 - c. Percent Collapse (Jennings and Knight, 1957) following saturation under load (C_p).

The data were arranged into seven sets according to depth. The range of depths considered and the total number of sample points for each set are shown in Table I. Data Set 1 through 6 each contains values for the parameters D, C_p , e_o , n_o , S_o , γ_d , and w_o . Data Set 7 contains values for three additional parameters, R, A, and PL. Because of space limitations, only a brief description of the geostatistical techniques followed in evaluating these data sets will be given, and only the results for C_p will be presented.

2. CONVENTIONAL STATISTICAL ANALYSES

In order to establish the reliability of each parameter as an indicator of soil collapse susceptibility, conventional statistical analyses consisting of descriptive statistics, probability analyses, regression

TABLE I
Data Set Characteristics

Data Set	Depth Range (m)	Number of Observations
1	0 - 0.30	125
2	0.30 - 0.61	286
3	0.61 - 0.91	254
4	0.91 - 1.22	100
5	1.22 - 1.83	104
6	1.83 - 12.2	123
7	0 - 12.2	219*

*Data from other sets containing values for three additional parameters.

analyses, and factor analyses were performed on the data in each data set. A brief description of the results of these analyses follows.

Descriptive Statistics: The mean value (μ), standard deviation (σ), coefficient of skewness (β_1), coefficient of kurtosis (β_2), and coefficient of variation (COV) were determined for each parameter in each of the seven data sets. The COV, which is a commonly used relative measure of the degree of uncertainty associated with a random variable, was found to increase linearly with depth for all parameters. This increase is attributed to the greater chance for a higher degree of sample disturbance at increased depths.

Probability Distribution: The development of a probability distribution for a random variable is a simple way to verify the homogeneity of its spatial distribution and to identify extreme or suspect data. The adequacy of a proposed theoretical distribution to describe the empirical distribution determined for each parameter was evaluated by the Kolmogorov-Smirnov (KS) and the Chi-square (CS) goodness of fit tests. The theoretical distribution represents what would be expected under the "null hypothesis." All parameters of each data set were tested against the Normal, Lognormal, and Gamma or Weibull distributions. The results of the KS and CS goodness of fit tests showed that all parameters followed the Gamma distribution, except γ_d which was found to follow the Weibull distribution. Therefore, in this study the Gamma distribution was used to describe the probability distribution of C_p .

Regression Analysis: Stepwise linear regression analyses were performed on all data sets for each of the variables in order to investigate possible functional relationships among the variables and to identify the "strongest" variables from a statistical perspective; i.e., those that yield the largest percent reduction of variance in the functional relationship being considered (Holtz and Krizek, 1971). Expressions were derived for each variable from all possible combinations of the parameters of interest.

The results of the stepwise linear regression analyses for C_p are presented in Table II. The associated correlation coefficients

TABLE II
Stepwise Regression Equations for C_p (%)
(All dimensionless variables expressed as decimals)

Data Set	Regression Equations $D=m; \gamma_d=kN/m^3$	% Variation Explained
1	$C_p = 8.03 - 19.5 S_o + 1.54D$	10.8
2	$C_p = 54.59 - 9.4 S_o - 0.068 \gamma_d - 87.8 W_o$	26.4
3	$C_p = 37.41 - 0.046 \gamma_d - 39.29 W_o$	29.2
4	$C_p = 29.80 - 2.6 S_o - 0.035 \gamma_d - 29.26 W_o$	24.6
5	$C_p = 73.22 - 0.086 \gamma_d - 17.1 e_o$	32.5
6	$C_p = -19.14 + 7.2 S_o + 0.02 \gamma_d - 47.23 W_o + 13.4 e_o$	27.2

are shown in Table III. Values of the correlation coefficient range between ± 1 , with "perfect prediction" indicated by either $+1$ or -1 and "no association" by zero. Although the strongest variables are seen to differ among data sets, in general, C_p is most strongly related directly to γ_d and indirectly to w_o through moderately strong correlations with e_o and S_o [$S_o e_o = w_o G_s$ where G_s = specific gravity of solids and is assumed constant].

Factor Analysis: In order to obtain a complete characterization of the variability of the collapse criteria and the collapse-related parameters, a higher order statistical technique called "factor analysis" was applied to each of the data sets (Harman, 1967). The main purpose of factor analysis is to define a minimum number of hypothetical variables or "factors" with which the correlations determined previously can be reanalyzed.

Factor analyses for each data set and for the combined data reveal that the most stable factors are those associated with γ_d and S_o . For most data sets unique factors were extracted for C_p and D as sole variables. In summary, the factor analyses validated all of the findings of the previous statistical analyses and lent confidence to the selection of C_p as a valid measure of the collapse susceptibility of soils in the Tucson Basin.

TABLE III
Correlation Coefficients for C_p

Data Set	D	Associated Variable e_o	S_o	γ_d	w_o
1	0.13	0.37	-0.29	-0.14	-0.24
2	-0.10	0.17	-0.42	-0.36	-0.26
3	-0.06	0.30	-0.36	-0.46	-0.15
4	0.10	0.33	-0.28	-0.46	-0.01
5	-0.05	0.44	-0.08	-0.52	0.10
6	-0.22	0.29	0.10	-0.37	0.10
7	-0.22	0.27	-0.43	-0.75	-0.10

3. KRIGING

A natural way to compare the values of a soil parameter $[Z(x) \text{ and } Z(x+h)]$ at two points in space $[x \text{ and } (x+h)]$ is to consider the difference in the values. For a set of pairs of sample points a certain separation distance apart the absolute average of the difference $[Avg|Z(x) - Z(x+h)|]$ can be obtained easily. For mathematical reasons, the squared differences are considered instead and the dissimilarity function is chosen as

$$2\gamma(h) = Avg [Z(x) - Z(x+h)]^2 \quad (1)$$

The term $2\gamma(h)$ is known as the "variogram." Being a function of the distance vector, it expresses how the average value of a parameter varies with distance in a given direction. If the data for the parameter demonstrate directional anisotropy, then the $\gamma(h)$ function will also depend on direction as well as separation distance and should be written as $\gamma(h, \theta)$. The variogram can also be interpreted as the elementary estimation variance of a variable $Z(x)$ by another variable $Z(x+h)$ at a distance h units from x . As such it can be expressed as:

$$2\gamma(h) = E \{ [Z(x+h) - Z(x)]^2 \} \quad (2)$$

The variogram can be estimated by

$$2\gamma^*k(h) = 1/N(h) \sum [Z(x_i) - Z(x_i+h)]^2 \quad (3)$$

where $[(x_1, x_1+h); \dots; (x_N(h), x_N(h) + h)]$ are $N(h)$ pairs of samples separated by the distance vector h . In this form, (h) is called the "semi-variogram." When it is estimated by $\gamma^*(h)$ it is called the "experimental semi-variogram." This parameter bears the same relationship to γ that a histogram does to a probability distribution (Clark, 1979).

If a semi-variogram for a given soil parameter indicates that it has a spatial structure over an area of interest, it may be advantageous to consider that spatial dependence to describe the distribution of observed values over that area. In such cases the variogram may be used to estimate values for the parameter by interpolating between observed values. The geostatistical estimation technique which provides the "best linear unbiased estimation" of the values for the parameter is known as "kriging."

There are three types of kriging: punctual, block, and universal. Punctual kriging, which provides estimates for values of a random variable at points where there is no drift, has two forms: simple kriging if the mean value of the variable is known, and ordinary kriging if the mean value is not known. Drift is defined as a nonstationary expectation of a random function. Block kriging is used when an estimation of the spatial average is required over a volume or an area. Universal kriging is an optimal method of interpolation that applies in all cases where drift must be taken into account because of lack of data to make stationary or quasi-stationary estimates (Matheron, 1963, 1969).

A major advantage of kriging over other interpolation methods is that the estimation variances can be calculated before actual sampling is made. The estimation variances depend on the semi-variogram and the configuration of the data points relative to the origin. They do not depend on the observed values of the parameter (Burgess and Webster, 1980). Therefore, if the variogram is known for a given soil parameter in a region, sampling intervals for the desired variance of estimation can be selected before actual samples are taken at the site.

As indicated previously, the main goal of this study was to obtain the probability distribution of each collapse criterion and collapse-related soil parameter at its three pre-defined cut-off levels relative to collapse susceptibility: high, medium, and low. There are several geostatistical methods for estimating probability distributions: multi-gaussian kriging, disjunctive kriging, indicator kriging, and probability kriging. The former two are parametric estimation techniques in which some assumptions regarding the distribution of the variable under study must be made. Their mathematical complexity makes them harder to apply than the latter two methods which, in addition to being distribution-free, are also nonlinear.

In this study, variograms were estimated for each collapse criterion and collapse-related soil parameter in each data set by using a discrete number of values obtained from test data at incremental distances corresponding to sampling locations throughout Tucson. These variograms were then used in conjunction with ordinary kriging to estimate values of the parameters at unsampled locations. Indicator kriging (Journel, 1983; Journel and Huijbregts, 1978) was then utilized to produce contour plots of estimated probability and associated kriging variance for each parameter in each data set. All of the geostatistical computations, including variogram estimation and kriging, were performed by using the computer program BLUEPACK developed by the Centre de Geostatistique, Fontainebleau, France.

4. RESULTS AND DISCUSSION

Typical results of kriging analyses are shown for C_p in Fig. 1. The solid line contours represent the estimated probability that soils having a "high" collapse potential will be encountered within 0.3 m of the surface. The cutoff values used to delimit "high" ($C_p > 5\%$), "moderate" ($2\% < C_p < 5\%$), and "low" ($C_p < 2\%$) collapse potential are those proposed by Sabbagh (1982) based on the work of Jennings and Knight (1975). The shaded zones in the figure indicate areas within Tucson where there is a 60%-80% probability of encountering collapse susceptible soils within 0.3 m of the surface, i.e., within the founding elevation of footings and floor slabs for most residential structures built in the area.

The broken line contours represent the estimation variance for the probabilities shown in the figure. The variance is seen to lie within a narrow range of moderate values between approximately 0.5 and 0.6. Although not presented here, similar contour plots were

developed for "moderate" and "low" collapse potential based on C_p . When the three plots are combined, a clear picture of the distribution of collapse susceptible soils in Tucson within 0.3 m of the surface emerges. Although not presented here, similar plots were developed for all of the other collapse criteria and collapse-related soil parameters for each of the seven data sets.

5. CONCLUSIONS

(1) The geostatistical analyses described in this paper predict a high probability of occurrence of soils having a high collapse potential in areas where the geomorphological features also favor the presence of such soils, i.e., predominantly within the flood planes of the ephemeral rivers and streams in the area. This supports the validity of the geostatistical approach used in this study.

(2) The geostatistical techniques described in this paper can be used to predict values of other soil properties in unsampled areas without bias and with known variance provided there is a statistically significant amount of reliable data available from tests performed on soils from sampled areas.

(3) Kriging methods can be used to develop probability contour plots for selected soil parameters at predetermined depths. These plots can then be combined to provide a unique and comprehensive picture of the three-dimensional spatial distribution of the parameter within an area of interest. This information is extremely valuable to planners, developers, and practicing geotechnical engineers. Its reliability can be improved with time if the data base is continuously updated.

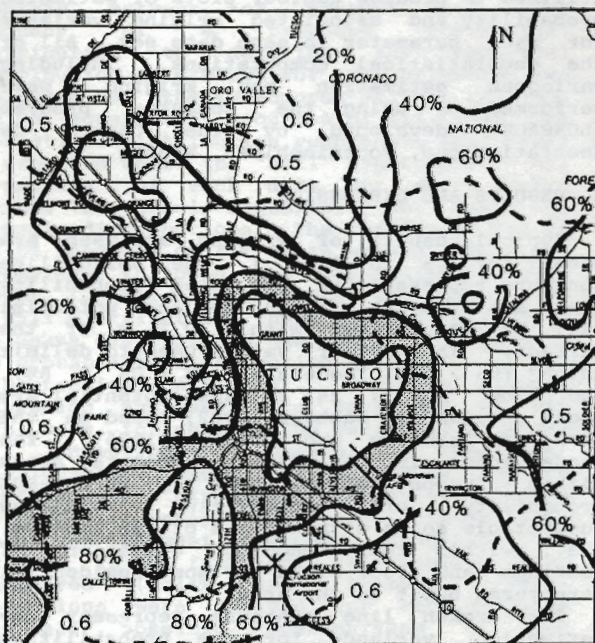


FIGURE 1
Probability Contours (solid) and Estimation
Variance (dashed) for Occurrence and of
"Highly" Collapsing Soils in Tucson, Arizona.

REFERENCES

- Abdullatif, A.A. (1969). Physical Testing of Engineering Properties of Soil in the City of Tucson, Arizona. M.S. Thesis, Univ. of Ariz. Tucson, Ariz.
- Alfi, A.A.S. (1984). Mechanical and electron optical properties of a stabilized collapsible soil in Tucson, Arizona. Ph.D. Dissertation, Univ. of Ariz. Tucson, Ariz.
- Ali, M.M. (1987). A probabilistic analysis of the distribution of collapsing soil in Tucson using kriging methods. Ph.D. Dissertation, Univ. of Ariz. Tucson, Ariz.
- Anderson, F. (1968). Collapsing soils and their basic parameters in an area in the Tucson, Arizona Vicinity. M.S. Thesis, Univ. of Ariz. Tucson, Ariz.
- Burgess, T.M. and Webster, R. (1980). Optimum interpolation and isarithmic mapping of soil properties, the variogram and punctual kriging. *Jour. of Soil Sci.* 31, 315-331.
- Clark, I. (1979). *Practical Geostatistics*, Applied Science Publishing, Ltd. Essex, London.
- Crossley, R.W. (1969). A geologic investigation of foundation failures in small buildings in Tucson, Arizona. M.S. Thesis, Univ. of Ariz. Tucson, Ariz.
- Gibbs, H.J. (1961). Properties which divide loose and dense uncemented soils. U.S. Bureau of Reclamation, Report No. EM-608, Denver, Colorado.
- Harman, H.H. (1967). *Modern factor analysis*, 2nd. edition. University of Chicago Press. Chicago, Illinois.
- Holtz, R.D. and Krizek, R.J. (1971). Statistical evaluation of soil test data. *Proc. 1st Intl. Conf. on Applc. of Stat. & Prob. to Soil and Struc. Engrg.* 230-266. Hong Kong.
- Jennings, J.E. and Knight, K. (1957). The additional settlement of foundations due to collapse of sandy subsoils on wetting. *Proc. 4th ICSMFE* 1, 316-319.
- Journel, A.J. (1983). Non-parametric estimation of spatial distribution. *Jour. Intl. Assc. for Math. Geol.* 15:3, 445-465.
- Journel, A.J. and Huijbregts, C.J. (1978). *Mining geostatistics*. Academic Press. New York.
- Matheron, G. (1963). Principles of Geostatistics. *Econ. Geol.* 58, 1246-1266.
- Matheron, G. (1969). *Le krigeage universel, les cahiers du centre morphologie mathématique*. Fasc. No. 1, ENSMP, Paris.
- Nowatzki, E.A. (1980). Settlement estimates in collapse susceptible soils. *Jour. Civ. Engrg. Desg.* 2:2, 121-148.
- Sabbagh, A.O. (1982). Collapsing soils and their clay mineralogy in Tucson, Arizona. M.S. Thesis, Univ. of Ariz. Tucson, Ariz.
- Sultan, H.A. (1969). Foundation failures on Collapsing soils in the Tucson, Arizona area. *Proc. 2nd Conf. on Expan. & Coll. Soils.* 394-398. Texas A & M Press, College Station.