RESEARCH ARTICLE
ScienceAsia 35 (2009): **372–380**
doi: 10.2306/scienceasia1513-1874.2009.35.372

A statistical evaluation of collapsing criteria and related parameters of collapsing soils

M. Mohammad Ali[a,*], **F. Ahmad**[a], **A.S. Yahaya**[a], **Edward A. Nowatzki**[b], **Donald E. Myers**[b]

[a] School of Civil Engineering, University Science Malaysia, 14300 Nibong Tebal, P. Pinang, Malaysia
[b] University of Arizona, Tucson, AZ 85721, USA

*Corresponding author, e-mail: mohammad59@hotmail.com

Received 13 Dec 2008
Accepted 4 Oct 2009

ABSTRACT: An analytical investigation was conducted in which about 1000 sample points from over 400 borehole locations throughout the Tucson Basin, Arizona were used to determine the nature and extent of the variability of selected collapse criteria and collapse-related soil parameters both spatially and with depth. Analysis of seven data sets corresponding to different depth increments below the surface showed high dispersion tendencies as expressed by the value of coefficient of variation (CoV). The value of CoV was found to increase linearly with depth. All the collapse criteria and collapse-related soil parameters used in the study were found to follow a Gamma distribution except for two collapse-related soil parameters which were found to follow a Weibull distribution. A polynomial regression model was developed for the collapse criterion, C_{p}, which is defined as the percentage volumetric strain occurring in a soil sample when saturated under constant load. The model showed that C_{p} varies almost linearly with depth. A stepwise regression analysis revealed that C_{p} is strongly correlated with two collapse-related soil parameters, namely, in situ dry density (γ_{d}) and in situ moisture content (w_0). Factor analysis validated this finding by producing two strong factors which described almost 80% of the total variance. These factors were closely related to γ_{d} and in situ degree of saturation, which is directly related to w_0, the in situ void ratio, and the specific gravity of solids.

KEYWORDS: collapse susceptibility, factor analysis, regression model, coefficient of variation, probability distribution

INTRODUCTION

A considerable portion of the geotechnical engineer's effort is devoted to the identification of soils and the evaluation of their properties for use in a particular analysis. Because of the variation of soil properties, the uncertainty in determining soil parameters can be very high. Even within a limited area, soil properties may vary due to inherent variation or heterogeneity of the soil strata[1]. Differences in testing techniques, equipment, and overall human factors also add to the difficulty in evaluating these parameters deterministically. A statistical approach to geotechnical engineering problems provides a rational basis to achieve a more economical solution by avoiding the use of extreme values and by quantifying the uncertainty in a solution[2]. In general, a soil deposit in a region may be either residual or transported. Also a transported soil may be alluvial (stream borne), aeolian (wind borne), or colluvial (gravity transported). When alluvial soils are deposited in an arid or a semi-arid environment, they develop larger voids within their structure due to the high evaporation rate during consolidation process. Such soils undergo a large decrease in bulk volume virtually instantaneously upon saturation or load application and are known as collapsing soils. However, even within the context of this definition, it is difficult to identify collapse susceptible soils due to the existence of many different types of clay minerals and many other factors that contribute to the collapse phenomenon. Therefore, application of statistical and probabilistic methods in analysing collapsing soil parameter would provide an optimum solution.

In this study statistical techniques were applied to selected collapse criteria and collapse-related soil parameters for soil in Tucson, Arizona, where the presence of collapse-susceptible soils is well documented[3,4]. Previous work on this topic was limited only to studies involving either specific areas or specific soil parameters. The purpose of this study was to gather as much information as possible from reliable sources and to use this data with statistical techniques, such as regression and factor analysis, to determine the variation of selected collapse criteria and collapse-related soil parameters in three dimensions. Only the variation of these criteria with depth will be considered here. Other geostatistical aspects of the problem were also studied by the authors and are presented elsewhere[5–9].

COLLAPSE CRITERIA AND RELATED PARAMETERS

When soils are deposited in an arid or semi-arid environment, there is insufficient time for them to consolidate under their own weight due to high evaporation rates. They become partially saturated with large voids. Application of typical foundation loads on such soils causes only minor deformation as long as the degree of saturation remains low. As soon as the soil becomes saturated, large deformations take place due to the reduction of volume and collapse of the intergranular structure. If water is readily available, the subsequent volume change and deformation are rapid and the phenomenon is referred to as collapse. In general, collapse-susceptible soils can be identified by a dry density criterion. If the dry density of the soil is sufficiently low to give a void space larger than that required to hold the liquid limit water content, then collapse upon saturation is likely. Otherwise collapse generally occurs only when the soil is loaded. In some cases, collapse susceptible soils are also found in residual soils. In general, these soils are terrace sediments consisting of low density, organic-rich silts, sands, clays of varying percentage and traces of gravels. Their approximate distribution varies. Fig. 1 shows grain size distribution curves of typical collapsing soil. However, regardless of the formation process and grain size distribution, most collapsible soils are geologically young[10].

The influence of clay fraction on collapse has been studied (see Refs. 14, 15). The studies show that collapse potential is negligible when the clay content is greater than 30%. If the clay content is below 5%, a collapse settlement, which remains small, is likely to take place, whereas maximum collapse is reached at clay contents of about 15%. This result conforms to the interval established by Lawton et al[16] who indicated that maximum collapse potential for the natural soils studied is obtained when the clay content is in the range of 10–40%.

Collapsing soils have been recognized throughout the world; particularly in Africa, part of Asia, Europe, as well as in the US. In the US the severity of the problem has been observed for well over two decades in the Midwestern and Western US, where soil deposits are generally either aeolian or alluvial.

There are altogether about ten criteria for predicting the collapsing potential of a soil. Some of the criteria are empirical. Others are derived theoretically from consolidation test results. The methods for evaluating collapse susceptibility vary from simple to very complex. Considerable effort has been given to establish criteria for predicting the collapse potential and the critical values for severity of a soil. The more commonly used criteria are described in Ref. 11. The two criteria most widely used in the US are the Gibb's collapse parameter (R)[12] and the percentage collapse (C_p) as obtained from the double oedometer test[13]. A third parameter A developed in Ref. 10 was also included. The parameters are defined by

$$C_\mathrm{p} = \frac{\Delta e_\mathrm{c}}{1+e_0} = \frac{\Delta H_\mathrm{c}}{H_0},$$

$$R = \frac{\dfrac{\gamma_\mathrm{w}}{\gamma_\mathrm{d}} - \dfrac{1}{G_\mathrm{s}}}{w_\mathrm{l}},$$

$$A = \frac{(e_0 - e_\mathrm{l})\gamma_\mathrm{w}}{(1+e_0)w_0\gamma_\mathrm{d}},$$

where Δe_c and ΔH_c are changes in void ratio and sample height, respectively, after saturation under a pressure of 200 kPa, e_l is the void ratio at liquid limit, and H_0 is the initial height of sample. Other related parameters are the initial dry unit weight (γ_d), initial moisture content (w_0), initial void ratio (e_0), initial porosity (n_0), initial degree of saturation (s_0), and plastic limit (PL).

Specific cut-off values for collapse susceptibility of each parameter for each collapse criterion (R, C_p, A) and collapse-related parameters (γ_d, e_0, n_0, s_0, PL) are given below. The critical values for parameters R, C_p, and n_0 obtained are given in Table 1. Other critical values were derived from a conventional volumetric-gravimetric relationship among the parameters.

Table 1 Critical values for non-collapsing (NC), medium collapsing (MC), and high collapse (HC) soil parameters.

Parameter	HC	NC	MC
R	$\geqslant 1.4$	< 1.0	$1.0 \leqslant R < 1.4$
C_p (%)	> 5	$\leqslant 2$	$2 < C_\mathrm{p} \leqslant 5$
n_0 (%)	$\geqslant 45$	< 40	$40 \leqslant n_0 < 45$
e_0	$\geqslant 0.82$	< 0.67	$0.67 \leqslant e_0 < 0.82$
γ_d,(pcf)	$\leqslant 91.0$	> 99.0	$91.0 < \gamma_\mathrm{d} \leqslant 99.0$
s_0 (w_0/s_0)	$\geqslant 0.308$	< 0.253	$0.253 < s_0 \leqslant 0.308$
PL	$\geqslant 23$	< 19	$19 \leqslant \mathrm{PL} < 23$
A	$e_0 > 0.67$, $A > -0.67$	$e_0 < 0.67$, $A < -0.67$	-

For this study, field and laboratory test data were collected from local consulting engineers' offices and from the reports of previous researchers[4]. In all, data for 992 sample points were collected from 411 different locations within the city of Tucson and its surroundings. The raw data were reduced to obtain

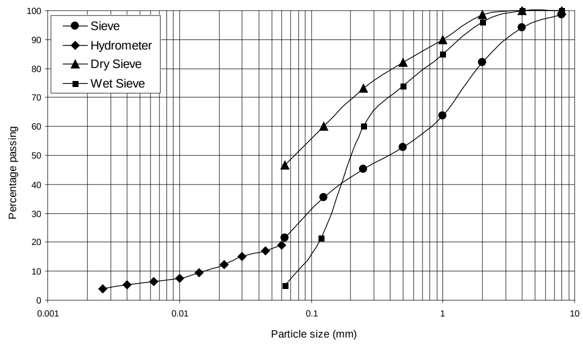

Fig. 1 Typical grain size distribution curve of collapsing soil.

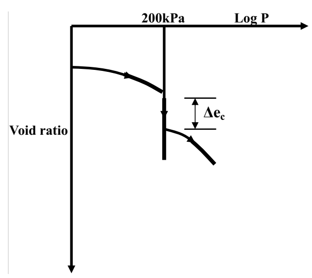

Fig. 2 Typical collapse potential test result[13].

parameters in two categories: established criteria (C_{p}, R, A) and collapse-related soil parameters (γ_{d}, w_0, e_0, n_0, s_0, and PL).

The collapse parameter, C_{p}, is obtained from a pseudo-consolidation test[13] and represents the volumetric strain the sample undergoes after it is saturated with water while under a pressure of 200 kPa as shown in Fig. 2. Some of the parameters listed previously are redundant since they can be calculated from a combination of the others. However, the objective was to investigate the effect of each commonly used collapse-related soil parameters on the collapse phenomenon.

Summary statistics

A total of 7 to 10 soil parameters were determined for each of the 992 sample points. Sample depths ranged from the surface to about 12.2 m. Table 2 shows the data sets for the range of depths considered and the total number of sample points (N) for each set. Data sets 1–6 each contain 7 parameters. Table 3 provides the descriptive statistics for each of the 6 parameters. Data set 7 contains the additional parameters R, A, and PL and the descriptive statistics of these are presented in Table 5.

Table 2 Data sets used in the analysis.

Data set	Range of depths (ft)	N
1	0–1	125
2	1.1–2	286
3	2.1–3	254
4	3.1–4	100
5	4.1–6	104
6	6.1–40	123
7	0–40	219

REGRESSION ANALYSIS

For the purpose of obtaining general descriptive parameters, primary attention was given to univariate data sets, i.e., numerical values obtained for a single characteristic of a sample. Linear regression and multiple linear regression analyses were used for all data sets in order to investigate possible functional relationships among the variables. Polynomial regression analyses were conducted on each data set to model the variation of the parameters with respect to depth, d.

The functional relationship between the dependent variables (y) and the independent variables (x) was found by the method of least squares. The goodness-of-fit or strength of this relationship was expressed in terms of percentage reduction of variance, which is defined as 100 times the sum of the squares of the computed values of y divided by the sum of the squares of the observed values of y.

Table 6 gives the results of the polynomial regression analyses for all data sets in terms of depth expressed in units of feet. The relationships apply for depths varying from 0.5 to 40 feet. Also included in the table is p-value used to test the null hypothesis that the multiple correlation is zero. Table 7 gives the results of a similar analysis for data set 7 which contains values for all collapse criteria and collapse related soil parameters. The regression equations for C_{p}, s_0, d, and w_0 given in Table 6 and Table 7 are different because of differences in the number of data points in each of the data sets considered and differences in their dispersion tendencies. The results in Table 6 were derived from 992 data points, whereas those in Table 7 were derived from 219 data points.

A stepwise linear regression analysis was used to express the collapse criterion C_{p} as a function of the collapse related soil parameters. A summary of the results of this analysis is given in Table 8. In general, the analysis shows that the parameters C_{p} is strongly related to the dry unit weight, γ_{d}, and the natural moisture content, w_0. It is weakly related to natural

Table 3 Descriptive statistics for collapse parameters for the 7 data sets (DS). CoV given as a percentage.

	C_p			e_0			s_0			γ_d			w_0			n_0		
DS	Mean	SD	CoV	Mean	SD	CoV	Mean	SD	CoV	Mean	SD	CoV	Mean	SD	CoV	Mean	SD	CoV
1	6.7	7.06	105.8	0.92	0.042	53.8	0.078	0.26	27.2	22.5	11.21	49.7	94.2	10.07	11.4	47.0	6.6	13.5
2	7.1	7.73	109.1	0.84	0.047	58.0	0.081	0.22	26.5	24.9	11.29	45.4	98.0	9.6	9.8	45.0	6.4	14.1
3	5.2	5.90	114.4	0.82	0.052	65.0	0.080	0.25	30.3	25.1	12.55	50.0	99.4	10.8	10.9	44.1	7.2	16.4
4	4.3	4.83	111.6	0.85	0.071	83.5	0.085	0.34	39.1	24.5	12.94	52.7	98.5	10.5	10.6	44.7	7.7	17.3
5	3.6	4.06	111.8	0.78	0.049	68.1	0.072	0.23	30.2	23.4	12.29	52.6	100.7	10.0	9.9	43.0	6.9	16.0
6	2.0	2.33	117.0	0.79	0.083	96.5	0.086	0.34	42.4	26.3	17.60	66.8	102.5	11.7	11.4	42.5	8.6	20.3
7	5.7	6.01	104.3	0.94	0.068	77.2	0.094	0.26	28.3	26.3	13.87	52.7	94.8	10.7	11.3	47.5	6.8	14.3
All	5.2	6.32	120.1	0.84	0.06	74.1	0.08	0.27	32.2	24.8	13.04	52.7	98.8	10.7	10.8	44.5	7.3	16.3

Table 4 Regression equation of CoV with depth.

Regression equation	R^2
$\mathrm{CoV}(C_\mathrm{p}) = 107.11 + 1.433d$	0.396
$\mathrm{CoV}(e_0) = 25.55 + 2.21d$	0.336
$\mathrm{CoV}(s_0) = 48.56 + 0.0099d^4$	0.825
$\mathrm{CoV}(\gamma_\mathrm{d}) = 10.5 + 0.02d$	0.960
$\mathrm{CoV}(w_0) = 50.5 + 6.25d$	0.545
$\mathrm{CoV}(n_0) = 13.14 + 0.966d$	0.624

Table 5 Descriptive statistics for R, A, and PL.

Parameter	Mean	SD	Skew, β_1	Skew, β_2	CoV	N
R	1.12	0.40	0.67	0.34	29.7	219
A	0.79	1.81	−1.75	12.0	28.6	219
PL	0.27	0.07	1.13	1.24	25.6	219

void ratio, e_0, porosity, n_0, degree of saturation, s_0, and the depth, d. As has been mentioned previously, the importance of a variable may increase or decrease when it is taken in combination with other variables. Moreover, a linear regression model is not the only model that can be used to explain the data set adequately.

Table 9 contains the results of a similar analysis for R, C_p, and A (data set 7). The collapse parameter R is related most strongly to A, PL, and d as seen in Step 5. The parameter C_p, as seen in Step 3, is strongly related to s_0, A, and d. The parameter A, on the other hand, is related most parameters and is strongly related to R, PL, and w_0.

Table 6 Polynomial regression equations for all data sets (d = depth in feet).

Regression equation	F_c	p-value
$C_\mathrm{p} = 7.89 - 1.035d + 0.0488d^2 - 0.00067d^3$	23.7	< 0.001
$e_0 = 0.09 - 0.0305d + 0.0016d^2$	6.8	0.0014
$n_0 = 46.67 - 0.19d + 0.043d^2 - 0.00002d^4$	9.9	< 0.001
$s_0 = 24.05 + 0.018d^2$	23.6	< 0.001
$\gamma_\mathrm{d} = 95.33 + 1.39d - 0.076d^2 + 0.0014d^3$	13.3	< 0.001
$w_0 = 0.078 + 0.00006d^2$	14.4	< 0.001

Table 7 Polynomial regression equations for data set 7 (d = depth in feet).

Regression equation	F_c	p-value
$C_\mathrm{p} = 7.42 - 0.063d + 0.014d^2$	8.01	< 0.001
$\mathrm{PL} = 0.26 - 0.05d$	4.9	0.028
$A = 1.21 - 0.16d + 0.005d^2$	4.3	0.015
$s_0 = 21.8 + 1.14d$	48.3	< 0.001
$\gamma_\mathrm{d} = 93.1 + 0.05d$	11.0	0.001
$w_0 = 0.077 + 0.0035d$	25.7	< 0.001

Table 8 Stepwise regression equations for C_p.

DS	Regression equation	PVE
1	$8.03 - 0.195s_0 + 5.06d$	10.8
2	$54.59 - 0.094 - s_0 - 0.436\gamma_\mathrm{d} - 87.8w_0$	26.4
3	$37.41 - 0.29\gamma_\mathrm{d} - 39.29w_0$	29.2
4	$29.8 - 0.026s_0 - 0.22\gamma_\mathrm{d} - 29.26w_0$	24.6
5	$73.22 - 0.05\gamma_\mathrm{d} - 17.1e_0$	32.5
6	$-19.14 + 0.125\gamma_\mathrm{d} - 47.23w_0 + 13.4e_0 + 0.072s_0$	27.2

PVE = % variation explained

Correlation coefficient matrix

The correlation coefficient matrices, which show the interrelationships among all the variables, were computed for each of the data sets X_{ij} considered in this study. Table 10, which applies to data set 7, illustrates these matrices. The correlation coefficient matrix and deviation matrix for a given data set are obtained from

$$Z_{ij} = (X_{ij} - X_j)/X_{ij},$$
$$\mathbf{R} = (1/n)\mathbf{Z}^\mathrm{T}\mathbf{Z},$$

where $\mathbf{Z}$ is the deviation matrix, $\mathbf{R}$ is the correlation matrix, and n is the number of observations in the data set.

Since the correlation coefficient is a measure of the linear association among variables, or how well one variable predicts another variable, it is an example of simple linear regression. The value of the correlation coefficient ranges between -1 and $+1$. If the value of the correlation coefficient is zero, no success in prediction is indicated. Perfect prediction is indicated by either $+1$ or -1. Based on the

Table 9 Stepwise regression equations for R, C_p and A.

Parameter	Regression equation (γ_d in pcf)	% variation explained
R	$R = 2.27 - 0.0005A - 0.009\gamma_\mathrm{d} - 3.51\mathrm{PL} - 0.9e_0 - 1.5w_0$	93.9
C_p	$C_\mathrm{p} = 3.94 - 0.138s_0 + 0.437A - 0.109d$	16.0
A	$A = -6.77 - 0.31R + 0.02C_\mathrm{p} + 0.065s_0 - 22.96w_0 + 0.25n_0 - 14.7\mathrm{PL}$	69.3

Table 10 Correlation coefficient matrix for data set 7.

	d	R	C_p	e_0	A	n_0	s_0	PL	γ_d	w_0
d	1.00	−0.24	−0.22	−0.06	−0.16	−0.08	0.43	0.15	0.22	0.33
R	−0.24	1.00	0.36	0.63	0.77	0.64	−0.43	−0.62	−0.75	−0.10
C_p	−0.22	0.36	1.00	0.27	0.38	0.31	−0.31	−0.14	−0.40	−0.17
e_0	−0.06	0.63	0.27	1.00	0.52	0.98	−0.01	0.10	−0.91	0.47
A	−0.16	0.77	0.38	0.52	1.00	0.55	−0.25	−0.52	−0.62	−0.04
n_0	−0.08	0.64	0.31	0.98	0.55	1.00	−0.02	0.09	−0.93	0.45
s_0	0.43	−0.43	−0.31	−0.01	−0.25	−0.02	1.00	0.27	0.37	0.83
PL	0.15	−0.62	−0.14	0.10	−0.52	−0.09	0.27	1.00	0.02	0.30
γ_d	0.22	−0.75	−0.40	−0.91	−0.62	−0.93	0.37	0.02	1.00	−0.11
w_0	0.33	−0.10	−0.17	0.47	−0.04	0.45	0.83	0.30	−0.11	1.00

information provided by the correlation matrices and the results of several regression analyses for each data set, the following conclusions are made.

No significant correlation was observed between depth and the other parameters for data sets 1–5 since values of the correlation coefficients were close to zero for these cases. This is to be expected because for each of these data sets the depth increment is small (1.0 or 2.0 feet) as shown in Table 2. The lack of correlation between C_p and depth and the other parameters is also clear from Table 7 where the parameter d does not appear in the regression equation for data sets 2–5. For data set 7, the depth varies from 6.0 to 40.0 feet. The first row of Table 10 gives values of the correlation coefficient sufficiently greater or less than zero to indicate a stronger correlation of d with other parameters than was obtained for any of the other data sets. This conclusion is also supported by the regression equation obtained for C_p from data set 7 as presented in Table 7.

The parameter C_p has a moderate negative correlation with γ_d (-0.40) and w_0 (-0.17), as was also seen by the use of regression analysis.

The natural void ratio, e_0, exhibits a strong negative correlation with γ_d (-0.91) and a strong positive correlation with n_0 (0.98).

The degree of saturation s_0 has a strong positive correlation with w_0 (0.83) and a moderate negative correlation with C_p (-0.31) as expected.

As shown in Table 10, the Gibbs collapse parameter R has strong correlation with e_0 (0.63), A (0.77), PL (-0.62) and γ_d (-0.75). Since both R and A are functions of PL, the high correlation is justified and expected. A comparison of the regression equations for R presented in Table 9 with the values of the coefficient of determination (R^2) for R presented in Table 10 verifies this observation. The comparison also shows that w_0 has only a small effect on explaining variations when it is included in the regression equation. One possible explanation is that when two or more variables are considered together, their effect on a third parameter is markedly different than if either of the two variables is considered alone. Moreover, when there is an extreme value for any one of the parameters, an unusual value in the deviation matrix may occur. This can result in a erroneously very low or high coefficient in the matrix.

The parameter A was found to be moderately to strongly correlated to R (0.77), C_p (0.33), n_0 (0.55), e_0 (0.52), s_0 (-0.25), PL (-0.52), and γ_d (-0.62). The parameter y_d apparently did not enter into the regression equation. It is often difficult to select the one or two strongest variables for a regression analysis on the basis of the correlation matrices or vice versa, since the importance of a given variable may change when it is entered into the regression analysis in combination with other variables.

FACTOR ANALYSIS

The various statistical techniques described in the previous section provide for rapid data evaluation, determination of statistical significance, evaluation of parameters, and the fitting of linear models. However, neither the predictor equations nor the correlation coefficients provide a complete characterization of the variability in collapse properties. For this reason, a higher order statistical technique called factor analysis was applied to each of the seven data sets. Factor analysis[15] is a powerful multivariate analytical tool which is used to reduce a large number of variables to a relatively small numbers of factors. The technique is useful in screening large data sets and helpful in formulating a hypothesis for an observed phenomenon.

The mathematical theory for the development of factor analysis is described in Ref. 17. Several factor solutions have been developed and procedures by which the factor coefficients are computed have been presented. The most general and widely used procedure is the principal-factor solution[17] which is also known as the 'method of axes'. In this method, the first axis in an ellipsoid is selected so that the sum of the squares of the distances of points from the axis is minimized. Successive axes, each orthogonal to the preceding ones, are chosen in order to minimize the squares of the distances of the points from the new axis. As a result of this process, the factor loadings, which are the correlations between the factors and the variables on each successive axis, become smaller and smaller until the number of factors reaches the rank of the correlation matrix satisfactorily.

The factor analysis presented here is based on principal components. Principal component analysis transforms a given set of variables into a new set of variables which are either orthogonal or uncorrelated. There is an essential difference between principal component analysis and factor analysis. In factor analysis, a variable is influenced by certain determinants, some of which are shared by other variables while others are not shared by any other variable. The main purpose of factor analysis is to define a minimum number of hypothetical variables or factors with which the correlation can be re-analysed.

In factor analysis, the data matrix **Z**, which is an $n \times m$ matrix containing the values of m measurements on n objects, is used to find m new measurements with zero as mean and identity matrix, **I**, as variance. These measurements, called factors, are linear functions of z scores. With the new measurements, the correlation matrix is recalculated and re-analysed.

The z scores expressed as a linear function of factors are

$$Z_i = f_1 U_{i1}\sqrt{\lambda_1} + f_2 U_{i2}\sqrt{\lambda_2} + \cdots + f_m U_{im}\sqrt{\lambda_m}, \quad (1)$$

where $i = 1, \ldots, m$. The correlation coefficient of z_i with f_m is given by $L = U\sqrt{\lambda}$ and is known as the loading.

With the availability of SPSS and other similar programs, lengthy routine calculations can be shortened considerably. This is particularly important when a large matrix is to be inverted or when the solutions of the characteristic equations used in factor analysis need to be obtained.

The factor model used here is based on the least squares method of the form $Z_i = a_{i1}f_1 + \cdots + a_{im}f_m$, where a_{ij} are factor coefficients (loadings) $= u_{ij}\sqrt{\lambda_i}$ ($i = 1 \ldots, m$; $j = 1, 2, 3, \ldots, m$) as shown in (1).

DISCUSSION

Factor analysis was applied to all seven data sets for collapse and collapse-related soil parameters. The analysis was also applied to a data set consisting of all the data. The results are presented in the form of rotated factor matrices in Table 11. In grouping the variables with a given factor, factor loadings less than 0.3 were arbitrarily considered to be zero and were reported in order to facilitate reading. An explanation of the significance of the factors for each of the data set follows.

Data set 1: Variables for 125 observations were analysed. The results in the form of a rotated factor matrix are shown in Table 11. Four factors were extracted, but only two are reported. The first two factors accounted for 76% of the total variance. The two factors extracted for the various parameters were F_1: C_p, e_0, n_0, γ_d, w_0 and F_2: C_p, s_0, γ_d, w_0. Factor 1 is seen to depend on the unit weight and other natural properties of the soil deposit. For each factor the variables associated with the factor are quite strong with respect to that factor and negligible for other factors. The inverse relationship between γ_d ($-$) and e_0 ($+$) is confirmation of the relationship $\gamma_d = G_s\gamma_w/(1+e_0)$. Factor 2 mainly depends on s_0. The C_p value was found to depend moderately on the degree of saturation in regression analysis. A strong correlation with water content is justified, i.e., $Se = wG_s$.

Data set 2: For this data set, the same variables were considered as for data set 1 and 286 observations were analysed. As shown in Table 11, the first two factors account for about 76% of the total variance. The two factors are quite similar to those derived for data set 1 except for C_p, which did not come out as a strong a factor as was previously obtained for data set 1.

Data set 3: Four factors were extracted and only two are reported in Table 11. The C_p value is found to have a strong negative correlation with the degree of saturation s_0 within Factor 1. This unusual relationship is probably due to an extreme value in the

Table 11 Rotated Factor Matrices of all Data Sets.

Variables	DS-1		DS-2		DS-3		DS-4		DS-5		DS-6		DS-7		All	
	F_1	F_2	F_1	F_2	F_1	F_2	F_1	F_2	F_1	F_2	F_1	F_2	F_1	F_2	F_1	F_2
d	-	-	-	-	-	0.227	-	-0.40	-	-	-	0.73	-	0.464	-	0.40
C_p	0.580	-	-	-	0.362	-	0.32	-0.75	0.52	-0.49	0.37	-	0.512	-	0.41	-0.62
e_0	0.994	0.516	0.987	0.738	0.986	-	0.97	-	0.97	-	0.97	0.75	0.847	-	0.98	-
n_0	0.997	-	0.983	-	0.989	-	0.98	-	0.98	-	0.97	-	0.863	0.472	0.99	-
s_0	-	-	0.268	-	-	0.918	0.52	0.76	-0.41	0.88	0.54	-	-0.38	0.779	0.29	0.88
γ_d	-	0.963	-	0.898	-	0.373	-0.88	0.37	-0.88	0.41	-0.83	0.70	-0.94	-	-0.90	0.31
w_0	0.941	0.291	0.887	0.443	0.905	0.707	0.86	0.44	0.73	0.66	0.84	0.46	-	0.931	0.70	0.68
R													0.892	-		
A													0.976	-		
PL													-0.31	0.55		
EV	3.299	2.000	3.404	2.089	3.408	1.994	3.299	1.633	3.667	1.664	3.716	1.72	4.362	2.600	3.493	1.920
CPTV	47.1	75.7	48.6	78.5	48.7	77.2	47.1	77.8	52.4	76.2	53.1	77.2	43.6	69.6	49.9	77.4

EV = eigenvalue; CPTV = cumulative percentage of total variance; DS = data set

deviation matrix.

Data set 4: As shown in Table 11, three factors accounted for 91.2% of the total variance, but only two are presented. This is essentially the same as for data set 2, except that a weak loading on d is combined with Factor 2.

Data set 5: As shown in Table 11 two factors accounted for 78% of the total variance. The factor loadings and variables are similar to those of data set 1.

Data set 6: As shown in Table 11, two factors accounted for 77% of the total variance. Factor 1 is the same as was found in all the other sets. Factor 2 appears here as a combination of the other factors. Data sets 3, 4, 5, and 6 have the same seven variables and 254, 100, 104, and 123 observations respectively. From the rotated factor matrices of these sets, as given in Table 11, it is observed that the factors extracted for each data set are virtually the same. This suggests that all of the collapse-related soil parameters are from the same population.

Data set 7: A separate analysis was done with the three collapse criteria and seven collapse-related soil parameters, i.e. ten variables altogether for 219 observations. Five factors were found to account for about 96% of the total variance, but only two are shown in Table 11. The factors extracted were: F_1: R, C_p, e_0, A, n_0, γ_d; F_2: d, e_0, n_0, s_0, PL, w_0; F_3: d, A; F_4: d, C_p; F_5: d, C_p. Factor 1 is clearly a factor of collapse criteria A and R. The strong negative correlation with γ_d was expected from their theoretical relationships. Factor 2 is determined by water content, w_0, which includes PL and s_0. Factor 3 is determined by PL. Factors 4 and 5 are the same as for other data sets.

All data sets: An additional run was made for all data sets with seven variables for 992 observations. The resulting rotated factor matrix is presented in Table 11. Four factors were extracted accounting for 98.7% of total variance, but only two are presented. It is seen that the strongest factor in the factor analysis of all the collapse-related soil parameter data is F_1 since it accounts for almost 50% of the variability. It appears most strongly related to e_0 and γ_d, therefore it can be considered as the "unit weight factor". The second strongest factor appears to be F_2 since it accounts for an additional 27.5% of the variance. It is most strongly related to s_0 can therefore be called the "degree of saturation factor". The third factor accounted for an additional 13% of the variance and was solely related to depth. The fourth factor accounted for an additional 8% variance and was related to C_p.

Table 12 presents a semi-quantitative summary of the results of the factor analysis. It is observed that the cumulative proportion of total variance among the eight sets of data is fairly uniform. There is little difference between the proportions of variance extracted from ten variables versus the proportion of variance extracted from seven variables.

In order to obtain meaningful associations from a factor analysis in this case, two conditions are required. The first is that the number of variables that characterize the collapse susceptibility should be the same for all strata analysed. The second is that the sample source from different locations and different depths should be identical. In other words, the deposit should be homogeneous.

In this analysis the most stable factors appear to be unit weight (F_1) and degree of saturation (F_2) factors. The introduction of unique factors such as C_p and d into the analysis reflects the independent character of these quantities and does not influence the deduced variables significantly. The parameters R and A may be strong factors as obtained from data set 7, but this cannot be investigated fully since they were not included as part of the other data sets due to the

Table 12 Summary of factor analysis.

Data Sets	N	No. of Variables	F_1	F_2	F_3	F_4	F_5	Cum % of Total variance
1	125	7	γ_d	s_0	d	C_p	-	99.4
2	286	7	γ_d	s_0	d	-	-	92.3
3	254	7	γ_d	s_0	d	C_p	-	99.2
4	100	7	γ_d	s_0	d	-	-	91.2
5	104	7	γ_d	s_0	d	C_p	-	99.2
6	123	7	γ_d	d	C_p	-	-	89.0
7	219	10	γ_d, R, A	s_0	PL	C_p	d	96.2
All	992	7	γ_d	s_0	d	C_p	-	98.7

unavailability of sufficient data at most depths.

Factor analysis was used in this study as a supplement to multiple regression analysis. In regression analysis, the significance of the regression and correlations can be tested. In factor analysis, there is no significance test for factor loadings. However, factor analysis provides more insight into the understanding of theoretical concepts and procedures than does the multiple regression analysis. In some cases, however, rotation in factor analysis may yield meaningless factors[18]. The factor analysis was performed by using the fact that γ_d and s_0 were the parameters most closely related to the two strongest factors. In this regard, factor analysis proved to be a sophisticated data reduction technique that confirmed the results of more conventional statistical analysis.

CONCLUSIONS

Values of selected collapse criteria and collapse-related soil parameters obtained from tests performed on about 1000 samples obtained from over 400 borehole locations throughout the Tucson Basin were analysed statistically to determine the nature and extent of their variability both spatially and with depth. The results of those analyses lead to the following observations and conclusions with respect to variability with depth.

(1) The statistical parameters for all the variables considered in the analysis manifest high dispersion tendencies as evidenced by the high values of the coefficient of variation (CoV). The value of the CoV was found to increase with depth as indicated in Table 8. The variations of CoV with depth were best modelled linearly for C_p, e_0, γ_d, w_0, and n_0 and nonlinearly for s_0. (2) In an attempt to fit the theoretical probability distribution function to each parameter, it was found that the distributions for all parameters except γ_d and n_0 were closely approximated by the Gamma distribution. The distributions for the parameters γ_d and n_0 were found to follow the Weibull distribution function. (3) All parameters were regressed with depth using a polynomial model in order to take nonlinearity within the profile into account. As indicated in Table 5, higher-order terms made a very insignificant contribution to the variation and can be ignored. (4) The results of a stepwise linear regression analysis of collapse parameter C_p presented in Table 8 reveal it to be significantly correlated with γ_d and w_0. The results of a similar analysis performed on the Gibbs Parameter R contained in Table 9 show a strong correlation between it and A, L, γ_d. (5) Factor analysis enabled the number of variables to be reduced to two independent parameters, γ_d and s_0, that were found to describe approximately 80% of the variation encountered in the data. This suggests that a good estimate of the variation of collapse susceptibility can be obtained from the two collapse-related soil parameters, dry unit weight, γ_d, and degree of saturation, s_0. It also validates the earlier findings revealed by regression analysis. (6) The database created with collapse and collapse-related soil parameters can be used for information regarding the severity of collapsing problems in a particular location of the city, and for further analysis.

Acknowledgements: The authors wish to acknowledge Western Technologies, Inc., Tucson, Arizona for allowing the author access to their job files to obtain data for this study at the University of Arizona.

REFERENCES

1. Baecher GB, Christian JT (2003) *Reliability and Statistics in Geotechnical Engineering*, Wiley, New York.
2. Christian JT (2004) Geotechnical engineering reliability: How well do we know what we are doing? *J Geotech Geoenviron Eng* **130**, 985–1003.
3. Crossley RW (1968) A geological investigation of foundation failures in small buildings in Tucson, Arizona. MSc thesis, Univ of Arizona.
4. Sabbagh AO (1982) Collapsing soil and their clay mineralogy in Tucson, Arizona. MSc thesis, Univ of Arizona.

5. Ali MM (1990) Statistical evaluation of Bangkok clay shear strength parameters. *Geotech Test J* **13**, 250–6.
6. Ali MM, Nowatzki EA, Myers DE (1989) Geostatistical techniques to estimate collapse-related soil parameters, Proceedings of the 25th Symposium on Engineering, Geology, and Geotechnical Engineering, Reno, Nevada, pp 289–96.
7. Ali MM, Nowatzki EA, Myers DE (1989) Geostatistical methods to predict collapsing soil. In: Proceedings of XII ICSMFE, Rio de Janeiro, Brazil, vol VII, pp 567–70.
8. Ali MM, Nowatzki EA, Myers DE (1990) Probabilistic analysis of collapsing soil by indicator kriging. *Math Geol* **22**, 15–38.
9. Nowatzki EA, Ali MM, Myers DE (1989) A use of geostatistics to predict the occurrence of collapsing soil. In: Proceedings of the Congress on Foundation Engineering, ASCE, vol 1, pp 176–90.
10. Alfi AAS (1984) Mechanical and electron optical properties of a stabilized collapsible soil in Tucson, Arizona. PhD thesis, Univ of Arizona.
11. Sultan HA (1971) Foundation failures on collapsing soils in the Tucson, Arizona area. In: International Residual and Engineering Conference on Expansive Clays, Texas A&M Press, College Station, pp 394–8.
12. Gibbs HJ (1961) Properties which divide loose and dense un-cemented soil. Report EM-608, US Bureau of Reclamation, Denver.
13. Jennings JE, Night K (1957) The additional settlement of foundations due to collapse of structure of sandy sub-soils on wetting. In: Proceedings 4th ICSMFE, London, vol 1, pp 316–9.
14. Khelifa A, Farid H, Tahar A (2007) Influence of relative density and clay fraction on soils collapse. In: Schanz T (ed) *Experimental Unsaturated Soil Mechanics*, Springer, Berlin, pp 3–9.
15. Kottegoda NT, Rosso R (1998) *Statistics, Probability, and Reliability for Civil and Environmental Engineers*, McGraw-Hill, New York.
16. Lawton EC, Fragaszy RJ, Hetherington MD (1992) Review of wetting-induced collapse in compacted soil. *J Geotech Eng* **118**, 1376–94.
17. Harman ME (1967) *Factor Analysis*, Wiley, New York.
18. Haldar A, Mahadevan S (2000) *Probability, Reliability and Statistical Methods in Engineering Design*, Wiley, New York.