Factors of Safety and Reliability in Geotechnical Engineering

By J. Michael Duncan,[1] Honorary Member, ASCE

Abstract: Simple reliability analyses, involving neither complex theory nor unfamiliar terms, can be used in routine geotechnical engineering practice. These simple reliability analyses require little effort beyond that involved in conventional geotechnical analyses. They provide a means of evaluating the combined effects of uncertainties in the parameters involved in the calculations, and they offer a useful supplement to conventional analyses. The additional parameters needed for the reliability analyses—standard deviations of the parameters—can be evaluated using the same amount of data and types of correlations that are widely used in geotechnical engineering practice. Example applications to stability and settlement problems illustrate the simplicity and practical usefulness of the method.

INTRODUCTION

The factors of safety used in conventional geotechnical practice are based on experience, which is logical. However, it is common to use the same value of factor of safety for a given type of application, such as long-term slope stability, without regard to the degree of uncertainty involved in its calculation. Through regulation or tradition, the same value of safety factor is often applied to conditions that involve widely varying degrees of uncertainty. This is not logical.

Reliability calculations provide a means of evaluating the combined effects of uncertainties, and a means of distinguishing between conditions where uncertainties are particularly high or low. In spite of the fact that it has potential value, reliability theory has not been much used in routine geotechnical practice. There are two reasons for this. First, reliability theory involves terms and concepts that are not familiar to most geotechnical engineers. Second, it is commonly perceived that using reliability theory would require more data, time, and effort than are available in most circumstances.

Christian et al. (1994), Tang et al. (1999), and others have described excellent examples of use of reliability in geotechnical engineering, and clear expositions of the underlying theories. The purpose of this paper is to show that reliability concepts can be applied in simple ways, without more data, time, or effort than are commonly available in geotechnical engineering practice. Working with the same quantity and types of data, and the same types of engineering judgments that are used in conventional analyses, it is possible to make approximate but useful evaluations of reliability.

The results of simple reliability analyses, of the type described in this paper, will be neither more nor less accurate than conventional deterministic analyses that use the same types of data, judgments, and approximations. While neither deterministic nor reliability analyses are precise, they both have value and each enhances the value of the other.

It is not advocated here that factor of safety analyses be abandoned in favor of reliability analyses. Instead, it is suggested that factor of safety and reliability be used together, as complementary measures of acceptable design. The simple types of reliability analyses described in this paper require only modest extra effort as compared to that required to evaluate factors of safety, but they will add considerable value to the results of the analyses.

EXAMPLE—RETAINING WALL STABILITY

A cantilever retaining wall is shown in Fig. 1. It will be backfilled with compacted silty sand, and the concrete footing will be cast on a layer of silty sand. The backfill will be drained to prevent buildup of water pressures behind the wall.

Factor of Safety against Sliding

The factor of safety against sliding on the sand layer beneath the footing is given by the following formula:

$$F_{ss} = \frac{W \tan \delta}{E} \qquad (1)$$

in which W = weight of wall and backfill over the heel of the wall (lb/ft or kN/m); $\tan \delta$ = tangent of friction angle between base of wall and sand; and E = earth pressure force on vertical plane through heel of wall (lb/ft or kN/m).

For the conditions shown in Fig. 1, the value of F_{SS} is 1.50, as shown at the top of Table 1. This is called the most likely value of factor of safety, F_{MLV}.

Uncertainty in Factor of Safety against Sliding

The terms involved in computing F_{ss} [W, $\tan \delta$, and E in (1)] all involve some degree of uncertainty. Therefore the computed value of F_{ss} also involves some uncertainty. It is useful

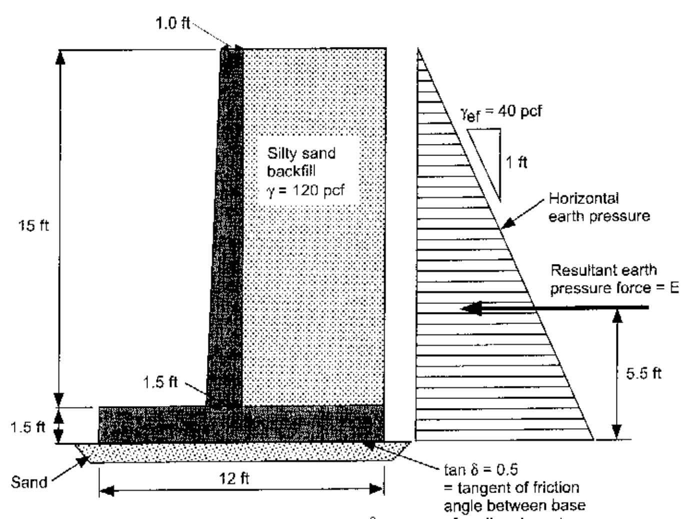


FIG. 1. Cantilever Retaining Wall with Silty Sand Backfill

---

[1]Univ. Distinguished Prof., Dept. of Civ. and Envir. Engrg., Virginia Tech, Blacksburg, VA 24061.

Note. Discussion open until September 1, 2000. To extend the closing date one month, a written request must be filed with the ASCE Manager of Journals. The manuscript for this paper was submitted for review and possible publication on May 13, 1999. This paper is part of the ***Journal of Geotechnical and Geoenvironmental Engineering***, Vol. 126, No. 4, April, 2000. ©ASCE, ISSN 1090-0241/00/0004-0307–0316/$8.00 + $.50 per page. Paper No. 20950.

TABLE 1. Taylor Series Reliability Analysis for Retaining Wall (with All Variables Assigned Their Most Likely Values, F_{ss} = 1.50)

Variable (1)	Values (2)	Factors of safety (3)	ΔF (4)
Equivalent fluid unit weight, γ_{ef}			
Most likely value plus σ	45 pcf	$F^+ = 1.33$	
Most likely value minus σ	35 pcf	$F^- = 1.71$	−0.38
Tangent of δ			
Most likely value plus σ	0.55	$F^+ = 1.65$	
Most likely value minus σ	0.45	$F^- = 1.35$	0.30
Backfill unit weight, γ_{bf}			
Most likely value plus σ	127 pcf	$F^+ = 1.56$	
Most likely value minus σ	113 pcf	$F^- = 1.44$	0.12
Concrete unit weight, γ_c			
Most likely value plus σ	152 pcf	$F^+ = 1.50$	
Most likely value minus σ	148 pcf	$F^- = 1.49$	0.01

Note: 1 pcf = 0.157 kN/m^3
standard deviation of factor of safety

$= \sqrt{(0.38/2)^2 + (0.30/2)^2 + (0.12/2)^2 + (0.01/2)^2} = 0.25$;

coefficient of variation of factor of safety = 0.25/1.50 = 17%.

to be able to assess the reliability of F_{ss}, as well as a best estimate of its value. This can be done using the Taylor series method, which involves these steps:

1. Estimate the standard deviations of the quantities involved in (1). Simple methods for estimating standard deviations are discussed in the next section of the paper. Using those methods, the following values of standard deviation of the parameters involved in this example have been estimated: σ_{efp} = standard deviation of the equivalent fluid pressure = 5 pcf (0.785 kN/m^3); $\sigma_{\tan\delta}$ = standard deviation of tan δ = 0.05; $\sigma_{\gamma bf}$ = standard deviation of the unit weight of backfill = 7 pcf (1.099 kN/m^3); and $\sigma_{\gamma c}$ = standard deviation of the unit weight of concrete = 2 pcf (0.314 kN/m^3).
2. Use the Taylor series technique (Wolff 1994; U.S. Army Corps of Engineers 1997, 1998) to estimate the standard deviation and the coefficient of variation of the factor of safety using these formulas:

$$\sigma_F = \sqrt{\left(\frac{\Delta F_1}{2}\right)^2 + \left(\frac{\Delta F_2}{2}\right)^2 + \left(\frac{\Delta F_3}{2}\right)^2 + \left(\frac{\Delta F_4}{2}\right)^2} \quad (2a)$$

$$V_F = \frac{\sigma_F}{F_{MLV}} \quad (2b)$$

in which $\Delta F_1 = (F_1^+ - F_1^-)$. F_l^+ is the factor of safety calculated with the value of the first parameter (in this case the equivalent fluid pressure) increased by one standard deviation from its best estimate value. F_1^- is the factor of safety calculated with the value of the first parameter decreased by one standard deviation.

In calculating F_1^+ and F_1^-, the values of all of the other variables are kept at their most likely values.

The values of ΔF_2, ΔF_3, and ΔF_4 are calculated by varying the values of the other three variables (footing/sand friction angle, backfill unit weight, and concrete unit weight) by plus and minus one standard deviation from their most likely values. The results of these calculations are shown in Table 1.

F_{MLV} = most likely value of factor of safety, computed using best estimate values for all of the parameters. For this example, F_{MLV} = 1.50.

3. Substituting the value of ΔF into (2*a*), the value of the standard deviation of the factor of safety (σ_F) is found to be 0.25, and the coefficient of variation of the factor of safety (V_F), calculated using (2*b*), is found to be 17%, as shown at the bottom of Table 1.
4. With both F_{MLV} and V_F known, the probability of failure and the reliability of the factor of safety can be determined using Table 2 or one of the methods described in Appendix I. Table 2 assumes a lognormal distribution of factor of safety values, which is usually a reasonable approximation. There is no "proof" that factors of safety are lognormally distributed, but the writer believes that it is a reasonable approximation.

The assumption of a lognormal distribution for factor of safety does not imply that the values of the individual variables (γ_{ef}, tan δ, γ_{bf}, and γ_c) must be distributed in the same way. As discussed below, it is not necessary to make any particular assumption concerning the distri-

TABLE 2. Probabilities That Factor of Safety Is Smaller than 1.0, Based on Lognormal Distribution of Factor of Safety

	Coefficient of Variation of Factor of Safety (V_F)														
F_{MLV}	2%	4%	6%	8%	10%	12%	14%	16%	20%	25%	30%	40%	50%	60%	80%
1.05	0.8%	12%	22%	28%	33%	36%	39%	41%	44%	47%	49%	53%	55%	58%	61%
1.10	0.00%	0.9%	6%	12%	18%	23%	27%	30%	35%	40%	43%	48%	51%	54%	59%
1.15	0.00%	0.03%	1.1%	4%	9%	13%	18%	21%	27%	33%	37%	43%	48%	51%	56%
1.16	0.00%	0.01%	0.7%	3%	8%	12%	16%	20%	26%	32%	36%	42%	47%	50%	56%
1.18	0.00%	0.00%	0.3%	2%	5%	9%	13%	17%	23%	29%	34%	41%	45%	49%	55%
1.20	0.00%	0.00%	0.13%	1.2%	4%	7%	11%	14%	21%	27%	32%	39%	44%	48%	54%
1.25	0.00%	0.00%	0.01%	0.3%	1.4%	4%	6%	9%	15%	22%	27%	35%	41%	45%	51%
1.30	0.00%	0.00%	0.00%	0.06%	0.5%	1.6%	3%	6%	11%	17%	23%	31%	37%	42%	49%
1.35	0.00%	0.00%	0.00%	0.01%	0.2%	0.7%	1.9%	4%	8%	14%	19%	28%	34%	40%	47%
1.40	0.00%	0.00%	0.00%	0.00%	0.04%	0.3%	1.0%	2%	5%	11%	16%	25%	32%	37%	45%
1.50	0.00%	0.00%	0.00%	0.00%	0.00%	0.04%	0.2%	0.7%	3%	6%	11%	19%	27%	32%	41%
1.60	0.00%	0.00%	0.00%	0.00%	0.00%	0.01%	0.05%	0.2%	1.1%	4%	7%	15%	22%	28%	38%
1.70	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.01%	0.06%	0.5%	2%	5%	12%	19%	25%	34%
1.80	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.01%	0.2%	1.2%	3%	9%	16%	22%	31%
1.90	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.08%	0.65%	2%	7%	13%	19%	29%
2.00	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.03%	0.36%	1.3%	5%	11%	17%	26%
2.20	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.01%	0.10%	0.56%	1.3%	8%	13%	22%
2.40	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.03%	0.23%	1.9%	5%	10%	19%
2.60	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.01%	0.09%	1.1%	4%	7%	16%
2.80	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.04%	0.66%	3%	6%	13%
3.00	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.02%	0.39%	1.8%	4%	11%

Note: F_{MLV} = factor of safety computed using most likely values of parameters.

butions of the variables to use the methods described here.

The retaining wall shown in Fig. 1 has a most likely value of factor of safety against sliding (F_{MLV}) equal to 1.50, and a coefficient of variation of factor of safety (V_F) equal to 17%. At the intersection of these values in Table 2, it can be seen that the probability of failure is about 1%. This equates to a reliability of about 99%.

Interpretation of "Probability of Failure"

The event whose probability is described as the "probability of failure" is not necessarily a catastrophic failure. In the case of retaining wall sliding, for example, "failure" would not be catastrophic. If the wall slid a small distance away from the backfill, the earth pressure on the wall would decrease, and sliding would stop. Subsequently, if the earth pressure increased again because of backfill creep, another episode of sliding might ensue. Eventually, if repeated episodes of sliding resulted in significant displacement of the wall, this behavior could constitute unsatisfactory performance of the wall, but not catastrophic failure.

In recognition of this important distinction between catastrophic failure and less significant performance problems, the Corps of Engineers uses the term "probability of unsatisfactory performance" (U.S. Army Corps of Engineers 1998). Whatever terminology is used, it is important to keep in mind the real consequences of the event analyzed and not to be blinded by the word "failure" where the term "probability of failure" is used.

SUMMARY OF TAYLOR SERIES METHOD

The steps involved in using the Taylor series method are these:

1. Determine the most likely values of the parameters involved and compute the factor of safety by the normal (deterministic) method. This is F_{MLV}.
2. Estimate the standard deviations of the parameters that involve uncertainty, using the methods discussed later in this paper.
3. Compute the factor of safety with each parameter increased by one standard deviation and then decreased by one standard deviation from its most likely value, with the values of the other parameters equal to their most likely values. This involves $2N$ calculations, where N is the number of parameters whose values are being varied. These calculations result in N values of F^+ and N values of F^-. Using these values of F^+ and F^-, compute the values of ΔF for each parameter and compute the standard deviation of the factor of safety (σ_F) using (2*a*) and the coefficient of variation of the factor of safety (V_F) using (2*b*).
4. Use the value of F_{MLV} from the first step and the value of V_F from the third step to determine the value of P_f, by means of Table 2 or one of the methods given in the Appendix I. Table 2 was developed using an Excel spreadsheet, following the method described in the appendix.

Today, when virtually all calculations of factor of safety are performed using spreadsheets or other computer programs, the $2N$ calculations in step 3 require little extra effort and little additional engineering time. These calculations can be done about as quickly as new parameter values can be entered into a spreadsheet or data file.

The bulk of the analysis effort is required in evaluating the data for the first calculation, in step 1. Thus, although additional calculations must be performed, they involve little time and effort beyond estimating values of the standard deviations of the parameters. As discussed in the following section, values of standard deviation of geotechnical parameters can be estimated using available data and applying engineering judgment. The use of prudent and informed judgment is as important in estimating values of standard deviations of parameters as it is in estimating most likely values of parameters.

The great advantage of computing P_f (the probability that the factor of safety could be less than 1.0) is that it provides an overall measure of the uncertainty in the results of the analysis. Computing both F_{MLV} and P_f, adds little to the time and effort required for the analysis, but adds greatly to the value of the result.

In addition, the values of ΔF computed for the different parameters afford a measure of their contributions to the probability that the factor of safety could be less than 1.0. For example, in Table 1 it can be seen that the unit weight of concrete has little effect on the result, and that γ_{ef} and tan δ have large effects. In this sense the Taylor's series method can be viewed as a structured sensitivity analysis or parametric study.

METHODS OF ESTIMATING STANDARD DEVIATION

An essential component of the art of geotechnical engineering is the ability to estimate reasonable values of parameters based on meager data, or based on correlations with results of in situ and index tests. In order to be able to estimate P_f, it is necessary to estimate the standard deviations of the parameters involved in computing the factor of safety. This can be done using the same types of judgment and experience used to estimate average values of parameters.

Depending on the amount of data available, various methods can be used to estimate the standard deviations of geotechnical parameters. Four methods that are applicable to various situations are described in the following paragraphs.

Computation from Data

If sufficient data are available, the formula definition of σ can be used to calculate its value:

$$\sigma = \sqrt{\frac{\sum [(x_i - \bar{x})^2]}{N - 1}} \tag{3}$$

in which σ = standard deviation; x_i = ith value of the parameter (x); $\bar{x}$ = average value of the parameter x; and N = number of values of x (the size of the sample).

Most scientific calculators and spreadsheet computer programs have facilities for calculating standard deviation using (3).

If the only method of determining values of standard deviation was (3), reliability analyses could not be used much in geotechnical engineering, because in most cases the amount of data is insufficient for use in (3). In order to be able to apply reliability analyses to the common situation, in which limited amounts of data are available and many properties are estimated using correlations, it is necessary to use other methods to estimate values of standard deviation. Three such methods are described in the following paragraphs.

Published Values

One approach to estimating values of standard deviation when sufficient data is not available to calculate σ using (3) is to use estimates based on published values, which are most conveniently expressed in terms of the coefficient of variation, V:

$$V = \frac{\text{Standard deviation}}{\text{Average value}} = \frac{\sigma}{\bar{x}} \quad (4a)$$

from which the standard deviation can be computed:

$$\sigma = (V)(\bar{x}) \quad (4b)$$

Values of V for a number of geotechnical engineering parameters and in situ tests, compiled by the writer and by Harr (1984), Kulhawy (1992), and Lacasse and Nadim (1997), are listed in Table 3. While the values in Table 3 represent a considerable number of tests, the value of V quoted in various references cover extremely wide ranges of values for the same parameter, and the conditions of sampling and testing are not specified. The values compiled in Table 3 therefore provide only a rough guide for estimating values of V for any given case. It is important to use judgment in applying values of V from published sources, and to consider as well as possible the degree of uncertainty in the particular case at hand.

"Three-Sigma Rule"

This rule of thumb, described by Dai and Wang (1992), uses the fact that 99.73% of all values of a normally distributed parameter fall within three standard deviations of the average. Therefore, if HCV = highest conceivable value of the parameter, and LCV = lowest conceivable value of the parameter, these are approximately three standard deviations above and below the average value.

The Three-Sigma Rule can be used to estimate a value of standard deviation by first estimating the highest and the lowest conceivable values of the parameter and then dividing the difference between them by six:

$$\sigma = \frac{HCV - LCV}{6} \quad (5)$$

in which HCV = highest conceivable value of the parameter, and LCV = lowest conceivable value of the parameter.

For example, the value of standard deviation of equivalent fluid unit weight (γ_{ef}) can be estimated as follows. First, the most likely value of γ_{ef} is estimated using experience, tables, or charts of the type found in Terzaghi et al. (1996). As shown in Fig. 1, the writer has estimated γ_{ef} = 40 pcf (6.28 kN/m^3) as the most likely value of γ_{ef} for the silty sand backfill. It seems reasonable that the highest conceivable value of γ_{ef} for this backfill might be about 55 pcf (8.635 kN/m^3), and the lowest conceivable value might be about 25 pcf (3.925 kN/m^3). These values are based on judgment. With HCV = 55 pcf, and LCV = 25 pcf, the standard deviation of γ_{ef} is computed as σ_{ef} = (55-25)/6 = 5 pcf (0.785 kN/m^3).

Studies have shown that there is a tendency to estimate a range of values between HCV and LCV that is too small. One such study, described by Folayen et al. (1970), involved asking a number of geotechnical engineers to estimate the possible range of values of $C_c/(1 + e)$ for San Francisco Bay mud, with which they all had experience. The results of this exercise are summarized in Table 4. It can be seen that, on average, these experienced engineers were able to estimate the value of $C_c/(1 + e)$ reasonably accurately but that they underestimated the standard deviation by about a factor of two as compared with the results of 45 laboratory tests.

The writer believes that the tendency to underestimate coefficients of variation results mainly from the fact that, while most geotechnical engineers have honed their ability to estimate average values of soil properties, they have had little experience in estimating coefficients of variation. With practice and experience, it should be possible to estimate coefficients of variation as accurately as average values of properties. When using the 3σ rule to estimate standard deviations and coefficients of variation, a conscious effort should be made to make the range between HCV and LCV as wide as seemingly possible, or even wider, to overcome the natural tendency to make the range too small.

With the 3σ rule it is possible to estimate values of standard deviation using the same amounts and types of data that are used for conventional geotechnical analyses. The three-sigma rule can be applied when only limited data are available and when no data is available. It can also be used to judge the reasonableness of values of coefficients of variation from published sources, considering that the lowest conceivable value would be three standard deviations below the mean and the highest conceivable value would be three standard deviations above the mean. If these values seem unreasonable some adjustment of the values is called for.

The 3σ rule uses the simple normal distribution as a basis for estimating that a range of three standard deviations covers virtually the entire population. However, the same is true of

TABLE 3. Values of Coefficient of Variation (V) for Geotechnical Properties and In Situ Tests

Property or in situ test result (1)	Coefficient of variation —V (%) (2)	Source (3)
Unit weight (γ)	3–7%	Harr (1984), Kulhawy (1992)
Buoyant unit weight (γ_b)	0–10%	Lacasse and Nadim (1997), Duncan (2000)[a]
Effective stress friction angle (ϕ')	2–13%	Harr (1984), Kulhawy (1992)
Undrained shear strength (S_u)	13–40%	Harr (1984), Kulhawy (1992), Lacasse and Nadim (1997), Duncan (2000)[a]
Undrained strength ratio (S_u/σ'_v)	5–15%	Lacasse and Nadim (1997), Duncan (2000)[a]
Compression index (C_c)	10–37%	Harr (1984), Kulhawy (1992), Duncan (2000)[a]
Preconsolidation pressure (p_p)	10–35%	Harr (1984), Lacasse and Nadim (1997), Duncan (2000)[a]
Coefficient of permeability of saturated clay (k)	68–90%	Harr (1984), Duncan (2000)[a]
Coefficient of permeability of partly saturated clay (k)	130–240%	Harr (1984), Benson et al. (1999)
Coefficient of consolidation (c_v)	33–68%	Duncan (2000)[a]
Standard penetration test blow count (N)	15–45%	Harr (1984), Kulhawy (1992)
Electric cone penetration test (q_c)	5–15%	Kulhawy (1992)
Mechanical cone penetration test (q_c)	15–37%	Harr (1984), Kulhawy (1992)
Dilatometer test tip resistance (q_{DMT})	5–15%	Kulhawy (1992)
Vane shear test undrained strength (S_v)	10–20%	Kulhawy (1992)

[a]Duncan (2000) refers to the present paper.

TABLE 4. Estimated and Measured Values of $C_c/(1 + e)$ and Its Coefficient of Variation, for San Francisco Bay Mud

Estimated by (1)	Estimated $C_c/(1 + e)$ (2)	Estimated V (3)
Geotechnical engineer #1	0.30	10%
Geotechnical engineer #2	0.275	5%
Geotechnical engineer #3	0.275	5.5%
Geotechnical engineer #4	0.30	10%
Average of #1–#4	0.29	8%
Measured	0.34	18%

other distributions (Harr 1987), and the 3σ rule is not rigidly tied to an assumed distribution of the variable.

Graphical Three-Sigma Rule

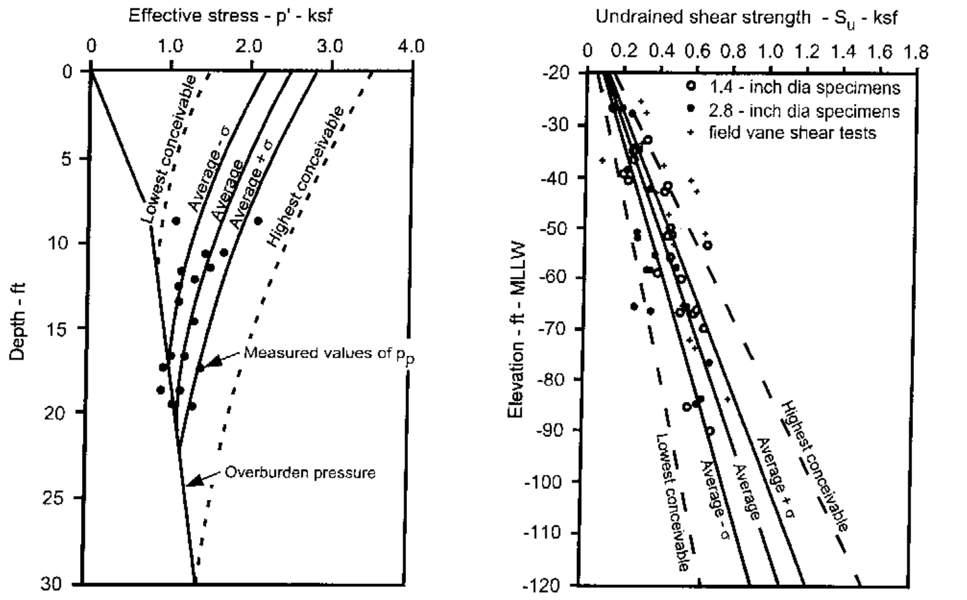
The concept behind the three-sigma rule of Dai and Wang (1992) can be extended to a graphical procedure that is applicable to many situations in geotechnical engineering, where the parameter of interest, such as preconsolidation pressure or undrained shear strength, varies with depth. Examples are shown in Fig. 2.

The graphical three-sigma rule is applied as follows:

1. Draw a straight line or curve through the data that represents the most likely average variation of the parameter with depth.
2. Draw straight lines or curves that represent the highest and lowest conceivable bounds on the data. These should be wide enough to include all valid data and an allowance for the fact that the natural tendency is to estimate such bounds too narrowly, as discussed previously. Note that some points in Fig. 2(b) are outside the estimated highest and lowest conceivable, indicating that these data points are believed to be erroneous.
3. Draw straight lines or curves that represent the average-plus-one standard deviation and the average-minus-one standard deviation. These are one-third of the distance from the average line to the highest and lowest conceivable bounds.

The average-plus-one-sigma and average-minus-one-sigma curves or lines, such as the preconsolidation pressure and undrained strength profiles in Fig. 2, are used in the Taylor series method in the same way as are parameters that can be represented by single values.

This same concept is useful in characterizing strength envelopes for soils. In this case the quantity (shear strength) var-

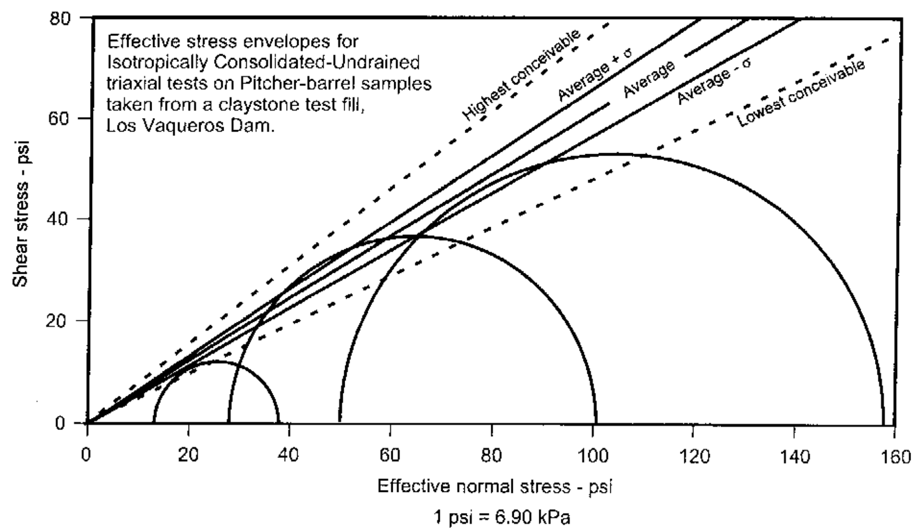


(a) Variation of preconsolidation pressures with depth for San Francisco Bay mud at Hamilton Air Force Base, Marin County, California (after Duncan 1965)

(b) Variation in undrained shear strength with depth for San Francisco Bay mud at the LASH Terminal site in San Francisco (after Duncan and Buchignani 1973)

1 ft = 0.305 m, 1 ksf = 47.9 kPa, 1 inch = 25.4 mm

FIG. 2. Examples of "Graphical Three-Sigma Rule" for Estimating Standard Deviation Limits for Parameters That Vary with Depth



1 psi = 6.90 kPa

FIG. 3. "Graphical Three-Sigma Rule" for Estimating Standard Deviation Limits for Strength Envelope

ies with normal stress rather than depth, but the procedure is the same. Strength envelopes are drawn that represent the average and the highest and lowest conceivable bounds on the data, as shown in Fig. 3. Then average-plus-one-sigma and average-minus-one-sigma envelopes are drawn one-third of the distance from the average envelope to the highest and lowest conceivable bounds. The average-plus-one-sigma envelope is used to compute the value of F^{+}, and the average-minus-one-sigma envelope is used to compute the value of F^{-}.

Using the graphical three-sigma rule to establish average-plus-one-sigma and average-minus-one-sigma strength envelopes is preferable to using separate standard deviations for the strength parameters c and ϕ. Strength parameters (c and ϕ) are useful empirical coefficients that characterize the variation of shear strength with normal stress, but they are not of fundamental significance or interest by themselves. The important parameter is shear strength, and the graphical three-sigma rule provides a straightforward means for characterizing the uncertainty in shear strength.

EXAMPLE—A SLOPE THAT FAILED

In August 1970, during construction of a new Lighter Aboard Ship (LASH) terminal at the Port of San Francisco, a 250-ft (75 m)-long portion of an underwater slope about 100 ft (30 m) high failed (Duncan and Buchignani 1973). A cross section through the failed area is shown in Fig. 4. The trench, which failed as it was being excavated, was to be filled with sand to provide a stability berm for the LASH Terminal.

The failure took place entirely within San Francisco Bay mud, a normally consolidated, slightly organic clayey silt or silty clay of marine origin. The clay at the site has moderate plasticity, with a Liquid Limit of about 50 and a Plastic Limit of about 30. The undrained shear strength of the Bay mud was measured using unconsolidated-undrained (UU) triaxial compression tests and in situ vane shear tests, the results of which are shown in Fig. 2(b).

Previous experience in the San Francisco Bay area indicated that underwater slopes in Bay mud could be excavated at least as steep as 1(*H*) on 1(*V*). Slope stability analyses performed during design of the LASH Terminal showed that, with the average strength profile shown in Fig. 2(b), the factor of safety of a 1(*H*) on 1(*V*) slope would have been 1.25.

Using steeper trench slopes would reduce the volume of excavation and fill and would reduce costs. It was estimated that, if the trench slopes could be excavated at 0.875(*H*) on 1(*V*), the cost of the stability trench would be reduced by about $200,000. Considerable effort was devoted to evaluating the undrained strength of the Bay mud, and the stability of the trench slope, as accurately as possible. Using the average strength profile shown in Fig. 2(b), it was found that the factor of safety of the slopes would be 1.17 if they were excavated at 0.875(*H*) on 1(*V*). Because the analyses were based on a considerable amount of high-quality data, it was decided to excavate the slopes at 0.875(*H*0 on 1(*V*), as shown in Fig. 4.

On August 20, 1970, after a section of the trench about 500 ft (150 m) long had been excavated, the dredge operator found that the clamshell bucket could not be lowered to the depth from which mud had been excavated only hours before. Using the side-scanning sonar with which the dredge was equipped, four cross sections were made within 2 h, which showed that a failure had occurred that involved a 250-ft (75 m) -long section of the trench. The cross section is shown in Fig. 4. Later a second failure occurred, involving an additional 200 ft (60 m) of length along the trench. The rest of the 2,000-ft (600 m) -long trench slope remained stable for about four months, at which time the trench was backfilled with sand.

The cost of excavating the mud that slid into the trench, plus the cost of extra sand backfill, was approximately the same as the savings resulting from the use of steeper slopes. Given the fact that the expected savings were not realized, that the failure caused great alarm among all concerned, and that the confidence of the owner was diminished as a result of the failure, it is now clear that using 0.875(*H*) on 1(*V*) slopes was not a good idea.

A detailed investigation after the failure indicated that the undrained strength of the Bay mud decreased, due to creep strength loss, to values smaller than measured in the laboratory UU tests and field vane shear tests, which were performed at normal rates of shearing and are quite rapid compared to rates of shearing in the field (Duncan and Buchignani 1973).

Reliability analyses of the slope have been performed recently using the Taylor series method, with the undrained shear strengths shown in Fig. 2(b). The results of this analysis are shown in Table 5. The greatest contributor to the standard deviation of the factor of safety (the largest value of ΔF) is the undrained shear strength of the Bay mud. However, the

TABLE 5. Taylor Series Reliability Analysis for LASH Terminal Cut Slope (with All Variables Assigned Their Most Likely Values, F_{ss} = 1.17)

Variable (1)	Values (2)	Factors of safety (3)	ΔF (4)
Bay mud undrained strength			
Most likely value plus σ	see Fig. 2	$F^{+} = 1.33$	
Most likely value minus σ	see Fig. 2	$F^{-} = 1.02$	0.31
Bay mud buoyant unit weight			
Most likely value plus σ	41.3 pcf	$F^{+} = 1.08$	
Most likely value minus σ	34.7 pcf	$F^{-} = 1.28$	0.20

Note:

standard deviation of factor of safety

$$= \sqrt{(0.31/2)^2 + (0.20/2)^2} = 0.18;$$

coefficient of variation of factor of safety = 0.18/1.17 = 16%.

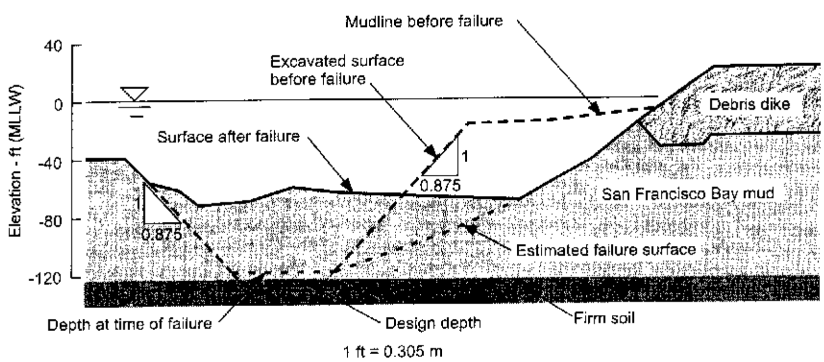


FIG. 4. Cross Section through Excavated Trench at LASH Terminal after Failure

buoyant unit weight of the Bay mud is also a significant factor. Although the standard deviation of the buoyant unit weight is only 3.3 pcf (0.518 kN/m^3), the average buoyant unit weight is only 38 pcf (5.97 kN/m^3) and the coefficient of variation is 9%. Coefficients of variation for buoyant unit weight are larger than for moist unit weight, because values of buoyant unit weight are smaller.

With a most likely value of factor of safety equal to 1.17 and a coefficient of variation of factor of safety equal to 0.16, as shown in Table 5, the probability of failure, found from Table 2, is 18%. The two sections that failed involved about 450 ft (140 m), or about 22% of the total 2,000 ft (600 m) length of cut slope. This close agreement between the computed probability of failure and the percentage of the slope that failed is probably fortuitous.

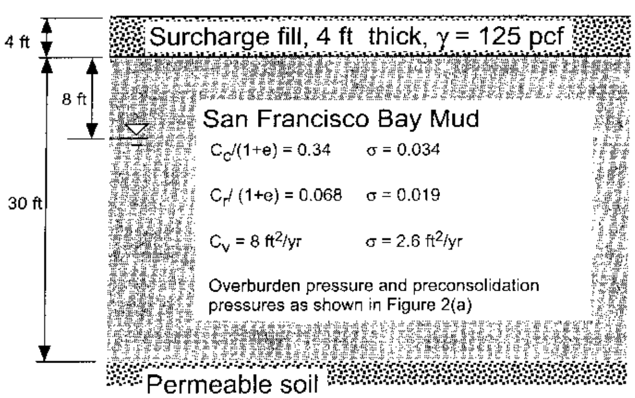
In retrospect, it is clear that a factor of safety equal to 1.17 was too low and a probability of failure equal to 18% was too high to be acceptable. At the time the slope was designed, however, we believed that $F = 1.17$ provided a real margin of safety. The failure showed that the real margin of safety was zero in some parts of the slope, and it must have been extremely small in the rest.

A reliability analysis was not performed when the slope was designed, and it is not possible to say 29 years later whether or not a calculated probability of failure of 18% would have altered our views about the advisability of excavating the slope at 0.875(H) on 1(V). However, it seems likely that knowing the probability of failure was 18% would have caused us to consider the steep slope to be less stable than we thought when we designed it, and we might well have changed the design had we evaluated it from the point of view of reliability. It is safe to say that our subjective perceptions at the time the slope was designed would have been that the probability of failure was considerably smaller than 18%.

EXAMPLE—CONSOLIDATION SETTLEMENT

Reliability theory can also be used to evaluate the effects of uncertainties in settlement calculations. As an example, consider the conditions shown in Fig. 5, where a 30-ft (9 m) -thick layer of San Francisco Bay mud is loaded by four feet of fill, imposing a surcharge of 500 psf (24 kPa). The overburden pressures and preconsolidation pressures for the layer are shown in Fig. 2(a).

For the conditions shown in Fig. 5, the computed most likely ultimate settlement is 1.07 ft (0.326 m). The results of a Taylor series reliability analysis of the settlement are shown in Table 6. Varying p_p, C_c, and C_r by plus and minus one standard deviation leads to the values of S^+ and S^- shown in Table 6. Although the coefficient of consolidation c_v also varies, it has no effect on the ultimate settlement. Using the values



1 ft = 0.305 m, 1 pcf = 0.157 kN/m^3, 1 ft^2/yr = 0.093 m^2/yr

FIG. 5. Consolidation Settlement Example

TABLE 6. Taylor Series Reliability Analysis for Ultimate Consolidation Settlement (with All Variables Assigned Their Most Likely Values, S = 1.07 ft)

Variable (1)	Values (2)	Settlement (3)	ΔS (4)
Preconsolidation pressure			
Most likely value plus σ	see Fig. 2	S^+ = 0.90 ft	
Most likely value minus σ	see Fig. 2	S^- = 1.21 ft	0.31 ft
$C_c/(1 + e)$			
Most likely value plus σ	0.374	S^+ = 1.18 ft	
Most likely value minus σ	0.306	S^- = 0.94 ft	0.24 ft
$C_r/(1 + e)$			
Most likely value plus σ	0.087	S^+ = 1.17 ft	
Most likely value minus σ	0.049	S^- = 0.97 ft	0.20 ft

Note: 1 ft = 0.305 m;

standard deviation of ultimate settlement

$= \sqrt{(0.31\ \text{ft}/2)^2 + (0.24\ \text{ft}/2)^2 + (0.20\ \text{ft}/2)^2} = 0.22$ ft;

coefficient of variation of ultimate settlement = 0.22 ft/1.07 ft = 21%.

of ΔS shown in Table 6, the standard deviation of ultimate settlement is 0.22 ft (0.067 m), and the coefficient of variation of ultimate settlement is 21%.

The probability that the settlement will be larger than some factor times the computed mostly likely settlement can be determined using Table 7. This table shows probabilities that the settlement ratio, SR, will be larger than the values listed in the left-hand column. SR is defined as

$$SR = \frac{\text{Possible settlement}}{\text{Most likely settlement}} \tag{6}$$

Table 7 provides a means of estimating an effective upper limit on settlement, based on the most likely settlement and the coefficient of variation. By choosing a small probability, the corresponding settlement ratio can be found in Table 7, and the possible settlement can be computed using the formula

$$\text{Possible settlement} = (SR) \times (\text{Most likely settlement}) \tag{7}$$

For example, Table 7 can be used to determine the settlement corresponding to a 1% probability of occurrence. With a coefficient of variation of settlement equal to 21%, the value of SR corresponding to 1% probability is about 1.6. In other words, there is a 1% chance that the ultimate settlement could be larger than (1.6) × (1.07 ft) = 1.7 ft, or, in metric units, the settlement would be larger than (1.6) × (0.326 m) = 0.522 m. Thus, 1.7 ft (or 0.522 m) can be viewed as an effective upper limit on settlement (with only 1% chance of being exceeded), considering possible variations in p_p, C_c, and C_r.

The same procedure can be used to determine the possible settlement at any time. For the conditions shown in Fig. 5, the computed most likely settlement at t = 2 years is 0.59 ft (0.18 m). A Taylor series reliability analysis of the settlement at 2 years is shown in Table 8. In this case four variables (p_p, C_c, and C_r, and c_v) influence the standard deviation and coefficient of variation of settlement. The standard deviation of settlement at two years is 0.12 ft (0.037 m), and the coefficient of variation is 21%.

As noted previously, a coefficient of variation equal to 21%, together with a 1% probability of being exceeded, corresponds to SR = 1.6. Thus the two-year settlement could possibly be as large as (1.6) × (0.59 ft) = 0.94 ft, or (1.6) × (0.18 m) = 0.288 m, with a probability of 1%. Thus 0.94 ft (0.288 m) can be viewed as an effective upper limit on settlement at two years (with a probability of 1%), considering variations in p_p, C_c, and C_r, and c_v.

Although in this case the coefficient of variation of settlement is equal to 21% for both the two-year settlement and the

TABLE 7. Probabilities That Settlement May Be Larger Than Computed Most Likely Settlement, Based on Lognormal Distribution of Settlement

	Coefficient of Variation of Settlement (V_s)											
SR	5%	10%	15%	20%	25%	30%	40%	50%	60%	**67%**[a]	70%	80%
1.10	3%	16%	24%	28%	30%	32%	33%	33%	33%	32%	32%	31%
1.20	0%	3%	10%	15%	19%	22%	25%	27%	27%	27%	27%	27%
1.30	0%	0%	3%	8%	12%	15%	19%	21%	23%	23%	23%	23%
1.40	0%	0%	1%	4%	7%	10%	14%	17%	19%	20%	20%	20%
1.50	0%	0%	0%	2%	4%	6%	11%	14%	16%	17%	17%	18%
1.60	0%	0%	0%	1%	2%	4%	8%	11%	13%	14%	14%	15%
1.70	0%	0%	0%	0%	1%	3%	6%	9%	11%	12%	12%	13%
1.80	0%	0%	0%	0%	1%	2%	4%	7%	9%	10%	11%	12%
1.90	0%	0%	0%	0%	0%	1%	3%	6%	8%	9%	9%	10%
2.00	0%	0%	0%	0%	0%	1%	2%	4%	6%	7%	8%	9%
2.20	0%	0%	0%	0%	0%	0%	1%	3%	4%	5%	6%	7%
2.50	0%	0%	0%	0%	0%	0%	1%	1%	3%	4%	4%	5%
3.00	0%	0%	0%	0%	0%	0%	0%	1%	1%	2%[a]	2%	3%

Note: *SR* = Settlement Ratio = Possible Settlement/Most Likely Settlement.
[a]Settlement of foundations on sand and gravel, computed using the method of Burland and Burbridge (1985), or Terzaghi et al. (1996), has a coefficient of variation of 67%.

TABLE 8. Taylor Series Reliability Analysis for Consolidation Settlement at $t = 2$ years (with All Variables Assigned Their Most Likely Values, $S = 0.59$ ft)

Variable (1)	Values (2)	Settlement (3)	ΔS (ft) (4)
Preconsolidation pressure			
Most likely value plus σ	see Fig. 2(a)	$S^+ = 0.58$ ft	
Most likely value minus σ	see Fig. 2(a)	$S^- = 0.60$ ft	0.02 ft
$C_c/(1 + e)$			
Most likely value plus σ	0.378	$S^+ = 0.65$ ft	
Most likely value minus σ	0.310	$S^- = 0.52$ ft	0.13 ft
$C_r/(1 + e)$			
Most likely value plus σ	0.087	$S^+ = 0.67$ ft	
Most likely value minus σ	0.049	$S^- = 0.50$ ft	0.17 ft
C_V (ft²/year)			
Most likely value plus σ	10.6 ft²/year	$S^+ = 0.64$ ft	
Most likely value minus σ	5.4 ft²/year	$S^- = 0.52$ ft	0.12 ft

Note: 1 ft = 0.305 m, 1 ft²/year = 0.093 m²/year;

standard deviation of settlement at 2 years

$= \sqrt{(0.02 \text{ ft}/2)^2 + (0.13 \text{ ft}/2)^2 + (0.17 \text{ ft}/2)^2 + (0.12 \text{ ft}/2)^2} = 0.12$ ft;

coefficient of variation of settlement at 2 years = 0.12 ft/0.59 ft = 21%.

ultimate settlement, this will not always be the case. The settlements during consolidation are influenced by c_v but the ultimate settlement is not. Therefore the coefficient of variation for during-consolidation and ultimate settlements will, in general, not be the same.

EXAMPLE—SETTLEMENT OF FOOTINGS ON SAND

Settlements of footings on sand can be estimated based on standard penetration blow count using the method developed by Burland and Burbridge (1985), which has been described in Terzaghi et al. (1996). Settlements are estimated using the formula

$$S_c = B^{0.75} \frac{1.7}{(\bar{N}_{60})^{1.4}} \left(q - \frac{2}{3} \sigma'_{vo}\right) \qquad (8)$$

in which S_c = computed settlement (immediate) in mm; B = footing width in meters; q = bearing pressure in kPa; σ'_{vo} = initial effective vertical stress at base of footing level in kPa; and $\bar{N}_{60}$ = average SPT blow count within a depth of $B^{0.75}$ m beneath the footing, corrected to 60% of the theoretical hammer energy (Terzaghi et al. 1996).

The relationship between settlements calculated using this formula and the measured settlements of 124 footings is shown in Fig. 6. It can be seen that the points scatter about the line of equality, indicating that (8) overestimates settlements in some cases and underestimates them in others.

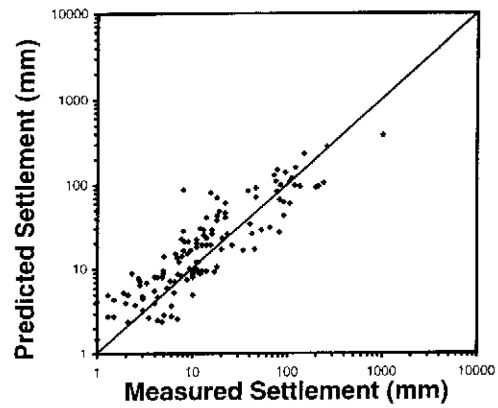


FIG. 6. Comparison of Measured Settlements at End of Construction with Settlements Predicted by Eq. (8)

Of greatest practical interest is the possibility that the actual settlement may be larger than the settlement computed using (8). The coefficient of variation associated with values computed from (8) was determined by computing the standard deviation of the data points around the $S_m = S_c$ line. Footings with measured settlements less than half an inch (13 mm) were not included in the analysis, because it was considered that settlements smaller than half an inch are usually of little practical significance. There were 54 footings with measured settlements greater than 13 mm. It was found that the expression $S_m = S_c$ has a coefficient of variation of 67% in these cases. Thus the column in Table 7 corresponding to V = 67% can be used to estimate how large the settlement might possibly be, based on a value calculated using (8).

As an example, consider the case of an 8 ft- (2.44 m)-square footing carrying a load of 320 kips (1,424 kN), founded 4 ft (1.22 m) below the surface at a site where $\bar{N}_{60}$ = 25. The bearing pressure, q = 320/64 = 5 ksf (240 kPa), and the initial effective vertical pressure, σ'_{vo} = 500 psf (24 kPa). The settlement computed using (8) is S_c = 8 mm, or 0.3 in.

Using the column in Table 7 for V = 67%, it can be seen that a probability of 2% corresponds to SR = 3.0. There is therefore a 2% chance that the settlement could be as large as (3.0) × (0.3 in.) = 0.9 in. Thus, with a probability of 2%, 0.9 in. can be viewed as an upper limit on the settlement of the footing.

Applying reliability concepts in this way provides a means of assessing the effects of the uncertainty involved in using (8), and a simple technique for relating the settlement estimate to the probability that the actual settlement could be larger than the estimated value.

SELECTING APPROPRIATE FACTORS OF SAFETY

Factors of safety provide a hedge against uncertainties in calculations, and the fact that it is never possible to compute with perfect accuracy. Through experience, conventions have developed with regard to what values of factor of safety are suitable for various situations. For example, the U.S. Army Corps of Engineers and many other agencies use $F = 1.5$ for long-term stability of slopes. Most geotechnical engineers use $F = 2.5$ to 3.0 for bearing capacity, and the same range of values for safety against erosion and piping.

Requiring the same factor of safety for all long-term slope stability applications, or all bearing capacity applications, is a "one size fits all" approach that is certain to result in inappropriate factors of safety in some cases. A more logical approach would consider

- The uncertainties in the quantities that enter the calculations
- The consequences of failure or unsatisfactory performance

This can be achieved, at least approximately, by choosing factors of safety such that the following relationship is satisfied:

$$\begin{pmatrix}\text{Reduction in } P_f\\ \text{associated with more}\\ \text{reliable design}\end{pmatrix} \times \begin{pmatrix}\text{Cost}\\ \text{of}\\ \text{failure}\end{pmatrix} < \begin{pmatrix}\text{Added cost}\\ \text{of more}\\ \text{reliable design}\end{pmatrix} \quad (9)$$

As discussed previously, not all "failures" are catastrophic. Some are better described as "unsatisfactory performance." Where the product of probability of failure times the cost of failure or unsatisfactory performance is small, it is justified to use smaller factors of safety. On the contrary, where the product of probability of failure times the cost of failure or unsatisfactory performance is large, higher factors of safety are logical.

The quantities in (9) cannot be evaluated precisely. Nevertheless, the relationship represented by this expression provides a basis for selecting appropriate factors of safety, even though approximations and judgment will be required in applying it to real conditions. In the case of the retaining wall in Fig. 1, for example, a 1% probability of unsatisfactory performance due to sliding would probably not justify the cost of increasing the factor of safety above 1.5 unless the consequences of sliding were unusually large. In the case of the LASH Terminal slope in Fig. 4, however, an 18% probability of failure of the slope, multiplied by the estimated cost of a failure, would have justified higher cost to increase the factor of safety and reduce the probability of failure.

SUMMARY AND CONCLUSIONS

Reliability theory can be applied to geotechnical engineering through simple procedures, and need not require more data than is required for conventional deterministic analyses. With a relatively small additional effort to perform reliability analyses, the value of analyses can be increased considerably.

It is proposed that probability of failure should not be viewed as a replacement for factor of safety, but as a supplement. Computing both factor of safety and probability of failure is better than computing either one alone. Although neither factor of safety nor probability of failure can be computed with high precision, both have value and each enhances the value of the other.

The word "failure," as used in the context of reliability theory, does not necessarily imply catastrophic failure. Some conditions—for example, sliding of a retaining wall—would be more aptly described as "unsatisfactory performance" than as "failure." Other conditions, e.g., slope failures that involve large movements, are appropriately described by the word "failure." It is important to bear in mind the likely consequences of the mode of "failure" being analyzed, and whether they are catastrophic or more benign.

Reliability concepts can be applied to settlement analyses as well as to factor of safety analyses. They can provide a measure of the accuracy of settlement computations and can be used to estimate the magnitude of settlement that has a very small probability of being exceeded.

Reliability analyses provide a logical framework for choosing factors of safety that are appropriate for the degree of uncertainty and the consequences of failure. While the factors that enter the relationship among probability of failure, consequences of failure, and the added cost of increased factor of safety cannot be evaluated with high accuracy, the relationship does serve to distinguish conditions where lower-than-normal factors of safety are appropriate or where higher-than-normal factors of safety are needed.

APPENDIX I. DETERMINING VALUES OF PROBABILITY OF FAILURE

Table 2 is convenient because it shows values of P_f related to F_{MLV} and V without further computation. Its shortcoming is that only approximate values of P_f can be determined for values of F_{MLV} and V that are intermediate between the values listed in the table. It may sometimes be desirable to have a means of computing more precise values of P_f.

The key to computing more precise values of P_f is to compute the value of the lognormal reliability index, β_{LN}, using the following formula:

$$\beta_{LN} = \frac{\ln\left(\frac{F_{MLV}}{\sqrt{1+V^2}}\right)}{\sqrt{\ln(1+V^2)}} \quad (10)$$

where β_{LN} = lognormal reliability index; V = coefficient of variation of factor of safety; and F_{MLV} = most likely value of factor of safety.

When β_{LN} has been computed, the value of P_f can be determined accurately in either of two ways:

1. Using tables of the standard cumulative normal distribution function, which can be found in many textbooks on probability and reliability. For example, the value of β_{LN} computed for sliding of the retaining wall in Fig. 1, with $F_{MLV} = 1.50$ and $V = 0.17$, is $\beta_{LN} = 2.32$. The standard cumulative normal distribution function (the reliability) corresponding to $\beta_{LN} = 2.32$ is 0.9898 (Dai and Wang 1992). The probability of failure is one minus the reliability, or $P_f = 1.0 - 0.9898 = 0.0102$.
2. Using the built-in function NORMSDIST in Excel. The argument of this function is the reliability index, β_{LN}. In Excel, under "Insert Function," "Statistical," choose "NORMSDIST," and type the value of β_{LN}. For $\beta_{LN} = 2.32$, the result is 0.9898, which corresponds to $P_f = 0.0102$. Table 2 was developed using this Excel function.

Table 7 has the same use with respect to probabilities of settlement. Values other than the values listed in Table 7 can be determined using the following expression for β_{LN}.

$$\beta_{LN} = \frac{\ln\{(SR)(\sqrt{1 + V^2})\}}{\sqrt{\ln(1 + V^2)}} \quad (11)$$

Table 7 was developed using this expression for β_{LN} as the argument for the NORMSDIST function in Excel.

ACKNOWLEDGMENTS

The assistance, guidance, and tutelage of many people have been of great assistance to me in writing this paper. My knowledge of reliability theory, while meager, has benefited from the patient instruction of Bill Houston, Greg Baecher, Don Javette, Rich Barker, Kamal Rojiani, Phillip Ooi, Chia Tan, John Sang Kim, Ed Demsky, John Christian, Tom Wolff, M. P. Singh, and Mike Navin. Tom Wolff wrote the publications that have been of greatest use to me, and he and M. P. Singh have patiently answered many questions regarding reliability theory and its practical use. Mike Navin, a Master's student at Virginia Tech, provided considerable assistance through his analyses of retaining wall stability, consolidation settlement, and settlement of footings on sand, and through many hours of discussion of these topics.

APPENDIX II. REFERENCES

Benson, C. H., Daniel, D. E., and Boutwell, G. P. (1999). "Field performance of compacted clay liners." *J. Geotech. and Geoenvir. Engrg.*, ASCE, 125(5), 390–403.

Burland, J. B., and Burbridge, M. C. (1985). "Settlement of foundations on sand and gravel." *Proc., Inst. of Civ. Engrs.*, Part 1, 78, 1325–1381.

Christian, J. T., Ladd, C. C., and Baecher, G. B. (1994). "Reliability applied to slope stability analysis." *J. Geotech. Engr.*, 20, Dec., 2180–2207.

Dai, S.-H., and Wang, M.-O. (1992). *Reliability analysis in engineering applications.* Van Nostrand Reinhold, New York.

Duncan, J. M., and Buchignani, A. L. (1973). "Failure of underwater slope in San Francisco Bay." *J. Soil Mech. and Found. Div.*, ASCE, 99(9), 687–703.

Folayan, J. I., Hoeg, K., and Benjamin, J. R. (1970). "Decision theory applied to settlement predictions." *J. Soil Mech. and Found. Div.*, ASCE, 96(4), 1127–1141.

Harr, M. E. (1984). "Reliability-based design in civil engineering." 1984 Henry M. Shaw Lecture, Dept. of Civil Engineering, North Carolina State University, Raleigh, N.C.

Harr, M. E. (1987). *Reliability-based design in civil engineering.* McGraw-Hill, New York.

Kulhawy, F. H. (1992). "On the evaluation of soil properties." ASCE Geotech. Spec. Publ. No. 31, 95–115.

Lacasse, S., and Nadim, F. (1997). "Uncertainties in characterizing soil properties." *Publ. No. 201*, Norwegian Geotechnical Institute, Oslo, Norway, 49–75.

Tang, W. H., Stark, T. D., and Angulo, M. (1999). "Reliability in back analysis of slope failures." *J. Soil Mech. and Found.*, Tokyo, October.

Terzaghi, K., Peck, R. B., and Mesri, G. (1996). *Soil mechanics in engineering practice*," 3rd Ed., Wiley, New York.

U.S. Army Corps of Engineers. (1997). "Engineering and design introduction to probability and reliability methods for use in geotechnical engineering." *Engr. Tech. Letter No. 1110-2-547*, Department of the Army, Washington, D.C., ⟨www.usace.army.mil/usace-docs⟩ (30 Sept. 1997).

U.S. Army Corps of Engineers. (1998). "Risk-based analysis in geotechnical engineering for support of planning studies." *Engrg. Circular No. 1110-2-554*, Department of the Army, Washington, D.C., ⟨www.usace.army.mil/usace-docs⟩ (27 Feb. 1998).

Wolff, T. F. (1994). "Evaluating the reliability of existing levees." *Rep., Res. Proj.: Reliability of existing levees*, prepared for U.S. Army Engineer Waterways Experiment Station Geotechnical Laboratory, Vicksburg, Miss.

Factor of Safety and Reliability in Geotechnical Engineering[a]

Discussion by Claudio Cherubini[2]

This paper is very interesting, highlighting the importance of the variability of soil properties in geotechnical calculations. The statistical and probabilistic techniques proposed are simple and very useful in current practice.

Some observations and considerations can, however, be made:

- Regarding the variability of soil properties, some further papers can be mentioned where coefficients of variation are reported and commented upon (Becker 1996; Cherubini 1997; Phoon and Kulhawy 1999).

  From Cherubini (1997) for example, the range of coefficients of variation (COV) of effective cohesion (between 13 and 70%) can be deduced wherever, from the paper of Becker (1996), among others, are reported coefficients of variation of design model uncertainty (5–15%), design decision uncertainty (15–45%), and construction variability (5–15%).

  Also, in their paper, Phoon and Kulhawy evidence some COV values regarding "inherent" or "intrinsic" soil variability, without any external influence (incorrect sampling, errors of calculation, and so on) that could cause adjunctive variability. From this paper and that by Cherubini (1997), it is evidenced that the COV of friction angle for clay is higher than the COV for sand. Dealing with standard deviation, all the variabilities for friction angle are in the range between 1.5 and 5°.
- Concerning the possibility to obtain coefficients of variation and means from estimated values of a property, we can use the following expressions (Cherubini and Orr 1999).

  For the mean:

$$x_m \cong \frac{a + 4b + c}{6}$$

  which, together with (5) from the original paper, gives the coefficient of variation

$$\mathrm{COV}_x \cong \frac{c - a}{a + 4b + c}$$

  where a = estimated minimum value; b = most likely value; and c = estimated maximum value.

  Alternatively, knowing the range (minimum minus maximum value) for a limited set of data, it is possible to use the table given by Snedecor and Cochran (1964) to obtain the standard deviation in a function of the number of values considered.

  Knowing the mean and standard deviation, it is possible to determine the so called "characteristic value" according to Eurocode 7 (Cherubini and Orr 1999; Orr and Farrell 1999).
- A comparison between calculated and measured values of —for example, the settlement of a shallow foundation, can be rationally performed (Cherubini and Orr 2000) evaluating

$$k_i = \frac{i\text{th calculated value}}{i\text{th measured value}}$$

  with i varying from 1 to n and calculating the mean of k_n values as a measure of "accuracy" and its standard deviation as a measure of "precision." Some indexes have been proposed, considering accuracy and precision lumped together, as the ranking index (Briaud and Tucker 1998) and the ranking distance (Cherubini and Orr 2000).
- The author's statement "that the probability of failure should not be viewed as a replacement for factor of safety but as a supplement" is acceptable at present, but the discusser thinks that, in the future, completely deterministic evaluations of safety will disappear. According to Lacasse (1994), "the fact that one finds difficult the quantifying of the uncertainties is not a valid reason to evade defining the uncertainties or establishing their significance on the results obtained. On the contrary, the greater the uncertainties, the more urgent is the need for reliability analysis."

For all these reasons, this paper represents a significant contribution towards reliability evaluation in geotechnics.

REFERENCES

Becker, D. E. (1996). "Eighteenth Canadian Geotechnical Colloqium: Limit states design for foundations. II: Development for the National Building Code of Canada." *Can. Geotech. J.*, Ottawa, 33, 984–1007.

Briaud, J. L., and Tucker, L. M. (1998). "Measured and predicted axial response of 98 piles." *J. Geotech. Engrg.*, ASCE, 114(9), 984–1001.

Cherubini, C. (1997). "Data and considerations on the variability of geotechnical properties of soils." *Proc., ESREL '97*, Lisbon, Portugal, 2, 1583–1591.

Cherubini, C., and Orr, T. L. L. (1997). "Considerations on the applicability of semi-probabilistic Bayesian methods to geotechnical design." *Proc., 20th Convegno Nazionale di Geotecnica*, Parma, Italy, 421–426.

Cherubini, C., and Orr, T. L. L. (2000). "A rational procedure to compare measured and calculated values in geotechnics." *Proc., IS-Yokohama*, Yokohama, Japan, in press.

Lacasse, S. (1994). "Reliability and probabilistic methods." *Proc., 13th Int. Conf. on Soil. Mech. and Found. Engrg.*, New Delhi, India, 225–227.

Orr, T. L. L., and Farrell, E. R. (1999). *Geotechnical design to Eurocode 7*, Springer, London.

Snedecor, G. W., and Cochran, W. G. (1964). *Statistical methods*, University of Iowa Press, Iowa City, Iowa.

Discussion by John T. Christian,[3] Fellow, ASCE, and Gregory B. Baecher,[4] Member ASCE

The discussers strongly agree with the author's position that the reliability of geotechnical systems can be evaluated by combining modest computational effort with data that are readily available on the engineering project. The discussers also support the conclusion that reliability analysis provides a framework for establishing appropriate factors of safety and other design targets and leads to a better appreciation of the relative importance of uncertainties in different parameters. Geotechnical engineers have long recognized that they deal with an uncertain world; reliability analysis is a rational way to grapple with it.

Daunting terminology and notation encumber the fields of reliability, probability, and statistics, and this has needlessly

[a]April 2000, Vol. 126, No. 4, by J. M. Duncan (Paper 20950).

[2]Full Prof. of Engrg. Geol.; formerly, Assoc. Prof. of Soil Mech., Tech. Univ. of Bari, 70125 Bari, Italy.

[3]Consulting Engr., Waban, MA.

[4]Prof. and Chair, Dept. of Civ. and Envir. Engrg., Univ. of Maryland, College Park, MD 20742.

prevented many practicing geotechnical engineers from taking advantage of these powerful tools. The discussers commend the author for his courage in presenting a straightforward exposition of reliability methods without mystification and hope the paper will encourage wider use of the techniques.

This discussion addresses issues raised by the paper in the hope that they might be subjects of further consideration within the geotechnical community. They are the implications of the three-sigma rule, the assumption that factors of safety are lognormally distributed, and the computational method expressed by (2*a*).

THREE-SIGMA RULE

The paper lists three methods for establishing standard deviations of uncertain variables in the order: (1) statistical computation from data; (2) reliance on published values; and (3) the three-sigma rule. The first two methods are clearly sound. However, it should be noted that many of the published values of variances in geotechnical properties are too large, because they include the measurement noise in addition to the actual variability in the properties. As regards the three-sigma rule, the concept of bracketing an uncertain quantity and then assuming that the range corresponds to some number of standard deviations is also sound. Certainly, when there are not enough data to compute the parametric values needed in an analysis, one must rely on approximations, and the three-sigma rule is one way to approximate standard deviations. However, we know from psychological studies that people have difficulty assessing the extent of their own uncertainty. When asked to state the greatest and least values an uncertain quantity can have, even statisticians report highly overconfident intervals —overconfident by manyfold. Morgan and Henrion (1990) summarize much of the work on this problem. The discussers suspect that, whereas highly experienced engineers who are also sophisticated in statistical issues might be able to interrogate themselves to obtain reasonable upper and lower bounds on a soil property, it is too much to hope that less experienced engineers could do so without systematic and unconservative bias. Thus, we are wary of the three-sigma rule in practice.

The literature on order statistics provides a useful first approximation for estimating standard deviations when some data are available. Consider that a sample of n observations is made from a normal distribution (other distributions could be used). One can calculate the expected values of the maximum and minimum of the n sample values and their difference, which is the range. This expected sample range is a function of the standard deviation. Table 9 shows the expected range in standard deviation units as a function of sample size n (Burlington and May 1970). So, if we make, say, ten tests and observe some range of values, we can estimate the standard deviation as that range divided by 3.078. Note that the three-sigma rule divides the range by 6.0.

TABLE 9. Number of Standard Deviations in Expected Sample Range after Burlington and May (1970) (σ = Range/N_n)

n	N_n	n	N_n
2	1.128	12	3.258
3	1.693	13	3.336
4	2.059	14	3.407
5	2.326	15	3.472
6	2.534	16	3.532
7	2.704	17	3.588
8	2.847	18	3.640
9	2.970	19	3.689
10	3.078	20	3.735
11	3.173	30	4.09

DISTRIBUTION OF FACTOR OF SAFETY

The factor of safety is often assumed to be lognormally distributed. Although arguments on this topic often seem to be like the warfare among the Lilliputians, for whom the issue was whether one should eat a boiled egg from the big end or the small end, consequences do follow from the lognormal assumption.

There seem to be three reasons for assuming a lognormal distribution. First, a lognormal distribution avoids negative values of factor of safety. In practice, the probability that a negative factor of safety will arise is insignificant. For example, in the LASH case, the mean and the standard deviation of the factor of safety are 1.17 and 0.18, respectively. If the factor of safety is normally distributed, the probability of a negative factor of safety is 4×10^{-11}. Second, there are many multiplications and divisions in geotechnical computations, and by the Central Limit Theorem, the lognormal distribution is a good representation for this situation. However, there are many additions and subtractions as well. Furthermore, for the argument to be valid for small numbers of variables, the individual variables should also be lognormally distributed. The three-sigma rule assumes that variables are normally distributed.

Third, it might be true that factors or margins of safety calculated in practice really do have a lognormal distribution. The consolidation problem in the paper has a complicated interaction of parameters. If the maximum past pressure is held constant, C_c and C_r contribute linearly; if they are normally distributed, so is the settlement. On the other hand, the maximum past pressure contributes in a nonlinear manner. It is not immediately clear what the distribution of the settlement should be, but Monte Carlo simulation indicates it is closer to a normal than a lognormal distribution. In the time-dependent analysis, c_v appears in an exponential; when it is normal and dominates, the settlement will closely approximate a lognormal distribution. Since the result of a realistically complicated calculation is not likely to follow a simple distribution, the discussers prefer to assess the probability of failure by: (1) assuming a normal distribution in the absence of other information; (2) using a margin of safety instead of a factor of safety as illustrated below; or (3) using the Hasofer-Lind (1974) approach to computing the reliability index without relying on the distribution of the margin or factor of safety.

COMPUTATIONAL METHOD

The derivation of (2*a*) starts with the first-order approximation for the variance of the factor of safety:

$$\sigma_F^2 \approx \sum_{i=1}^{n} \left(\frac{\partial F}{\partial x_i}\right)^2 \cdot \sigma_{x_i}^2$$

When the partial derivatives cannot be evaluated analytically, a central difference approximation is used:

$$\frac{\partial F}{\partial x_i} \approx \frac{F(x + \Delta x_i) - F(x - \Delta x_i)}{2 \cdot \Delta x_i}$$

When the increments in each of the variables are made equal to the standard deviations, (2*a*) results. The arguments for using this form of the first-order second-moment (FOSM) method include that it is simple, that it eliminates multiplication by the variances, and that it ensures that the evaluation is carried out over a significant range of values of the variables. The discussers prefer to use small values of the variables to compute the central differences in order to obtain the best estimate of the derivatives. The additional effort is small, and the technique avoids errors that arise from taking a secant approximation to a tangent. In most cases the differences between the two approaches are significant, but the following example shows that this is not always so.

EXAMPLE

The failure load P on a mine pillar is described by Karzulovic (1999) as

$$P = k \cdot \frac{W_{\text{eff}}^{0.5}}{H^{0.7}}$$

where k = design rock mechanics strength (DRMS) in MPa; H = height of the pillar in m; and W_{eff} = effective width of the pillar in m, defined as four times the cross-sectional area divided by the perimeter. The formula is used when $W_{\text{eff}} \leq 4.5H$. Since the formula is empirical, it is likely to have bias and uncertainty, and the exponents also have uncertainty. These uncertainties are ignored in the following calculations.

Let the load per unit area on the pillar be L. The safety of the pillar could be described by the factor of safety F (=P/L) or by the margin of safety M (=$P - L$). Obviously, the failure condition can be expressed equally well as $F = 1$ or $M = 0$. Data for pillars in several mines provided to the discussers by Dr. Antonio Karzulovic (personal communication, 1999) yield the statistics presented in Table 10. All four variables are reasonably presented by normal distributions. For present purposes, the correlations among the variables are ignored.

The partial derivatives of F or M can be evaluated directly, so the FOSM results can be computed exactly. With the failure criterion expressed in terms of M, the probability of failure was computed five ways: (1) FOSM with the exact partial derivatives; (2) FOSM using (2*a*); (3) the Rosenblueth (1975, 1981) point-estimate method; (4) the Hasofer-Lind first-order reliability method; and (5) Monte Carlo simulation with importance sampling. The results are shown in Table 11. In the first three cases, the probability of failure was computed from β with the assumption that M was normally distributed. All methods give substantially the same results, and (2*a*) is as accurate as any of them.

With the failure criterion expressed in terms of F, six computations of probability of failure were made using: (1) FOSM with the exact partial derivatives; (2) FOSM with (2*a*); (3) FOSM with small increments in the variables; (4) the Rosenblueth point-estimate method; (5) the Hasofer-Lind method; and (6) Monte Carlo simulation. In the third case, the increments were 1.0 for k, 0.5 for W, 0.1 for H, and 1.0 for L. In the first four cases, the probability of failure was computed from β with the assumption that F was normally distributed. The results are in Table 12. It is clear that, except for the Hasofer-Lind and Monte Carlo results, the methods based on the factor of safety disagree with each other and from those based on the margin of safety. Since the latter are consistent for all methods, including Monte Carlo simulation, the errors must lie in the methods based on the factor of safety. Furthermore, the most inaccurate method is that using the factor of safety and (2*a*).

It might also be assumed that the factor of safety is lognormally distributed. The mean and standard deviation of the logarithm of F can be calculated from the mean and standard deviation of F itself. The failure criterion is then $\ln F = 0$, and the value of β for $\ln F$ leads to a revised probability of failure. This procedure is codified in Table 2 of the original paper. For the present problem, β based on $\ln F$ is 1.123, and the probability of failure is 0.131.

A strong argument for a lognormal distribution of F is that it is computed by multiplying and dividing four variables. For a very large number of variables, the Central Limit Theorem implies that this is true for almost any distributions of the individual variables, but, when the number of variables is small, each variable should also be lognormally distributed. If each variable is lognormally distributed with the same means and standard deviations used before, the expression for $\ln F$ becomes a linear combination of normally distributed variables. The mean and standard deviation of F can be calculated directly, leading to $\beta = 1.429$ and $p_f = 0.077$.

TABLE 10. Means and Standard Deviations of Parameters for Mine Pillar Analysis

Variable	Mean	Standard deviation	Units
k	49.13	12.21	MPa
W_{eff}	13.85	2.91	m
H	4.00	0.20	m
L	33.66	16.44	MPa

TABLE 11. Results for Mine Pillar Problem Using Margin of Safety

Method	μ_M	σ_M	β	p_f
FOSM—direct	35.623	25.012	1.424	0.077
FOSM—(2*a*)	35.623	25.012	1.424	0.077
P-E	35.236	24.906	1.415	0.079
H-L			1.454	0.073
M-C				0.076

TABLE 12. Results for Mine Pillar Problem Using Factor of Safety

Method	μ_M	σ_M	β	p_f
FOSM—direct	2.058	1.151	0.920	0.179
FOSM—(2*a*)	2.058	1.434	0.738	0.230
FOSM—small increments	2.058	1.151	0.919	0.179
P-E	2.688	1.544	1.093	0.137
H-L			1.454	0.073
M-C				0.076

CONCLUSIONS

The author has presented a cogent argument for more widespread use of reliability methods in geotechnical engineering and has demonstrated that it can be applied easily to a variety of geotechnical problems. The discussers agree with this position and hope that the paper encourages wider use of reliability methods.

The discussers caution against uncritical use of the three-sigma approach. Experts of all sorts have great difficulty estimating the ranges of uncertain parameters. An alternative based on a limited number of data has been discussed.

Analysis of the safety of mine pillars suggest that, when one is dealing with a failure criterion such as F that is calculated primarily by multiplications and divisions, the FOSM method with F normally distributed may be in error, and the simplified version expressed by (2*a*) is least accurate. Assuming that F is lognormally distributed improves the accuracy but still gives results that differ significantly from the correct values. Accurate results based on the factor of safety require either: (1) using the Hasofer-Lind procedure; (2) using Monte Carlo simulation; or (3) describing each of the variables as well as F by a lognormal distribution. Failure estimates based on the margin of safety are mutually consistent and accurate.

REFERENCES

Burington, R. S., and May, D. C., Jr. (1970). *Handbook of probability and statistics*, McGraw-Hill, New York.

Hasofer, A. M., and Lind, N. (1974). "Exact and invariant second-moment code format." *J. Engrg. Mech.*, ASCE, 100(1), 111–121.

Morgan, M. G., and Henrion, M. (1990). *Uncertainty: a guide to dealing with uncertainty in quantitative risk and policy analysis*, Cambridge University Press, Cambridge, U.K.

Rosenblueth, E. (1975). "Point estimates for probability moments." *Proc., Nat. Academy of Sci.*, 72(10), 3812–3814.

Rosenblueth, E. (1981). "Two-point estimates in probabilities." *Appl. Mathematical Modelling*, 5(2), 329–335.

Discussion by Roger A. Failmezger,[5] Member, ASCE

The author has presented a practical approach for using probability in geotechnical engineering design. The numeric examples are very beneficial for understanding probability concepts.

The discusser believes there is an error in some of the values contained in Table 2. The probability of failure should not exceed 50% for an F_{MLV} that exceeds 1.0. For a uniform probability distribution (the distribution with the highest coefficient of variation and all possibilities equally likely) with an F_{MLV} = 1.05 and limits from 0 to 2.1, the probability of failure = 1.0/2.1, or 48%. In Table 7, the discusser feels that it is not possible for the coefficient of variation to increase and the probability of failure to decrease as is the case for SR = 1.10.

The discusser also questions the use of the lognormal probability distribution for geotechnical design applications. The lognormal distribution has limits of zero and positive infinity, and thus the distribution is always skewed to the left. With probability design, the engineer evaluates the area beneath the probability distribution function at the tail ends. The failure zone will be the area beneath the left tail below 1.00 for factor of safety based designs (the factor of safety is the abscissa). For settlement based design, the failure zone will be the area beneath the right tail above a maximum settlement threshold value (the settlement is the abscissa). Because of the left skewness of the lognormal distribution, designs where the failure zone is along the left tail will tend to be conservative and those with the failure zone along the right tail will tend to be unconservative.

The normal probability distribution function is symmetrical about its mean and has limits from negative infinity to positive infinity. These limits are not realistic and probably cause some error when evaluating the failure zone. The discusser suggests that a beta probability distribution be used because its limits can be realistically chosen by the engineer (Harr 1977). (The normal distribution is a subset of the beta distribution.) Where the average value occurs with respect to those limits will determine the skewness of the beta distribution.

In the example for determining the probability of unsatisfactory performance for settlement of footing on sand, the author uses settlement predictions based on SPT N_{60} values. He evaluates the coefficient of variation of how well the model predicts what has been measured based on case study data as 67%. This high coefficient of variation is partly due to using a dynamic penetration test to predict the static deformation properties of sand. In addition to the uncertainty of the model, there is uncertainty from measurement noise (test repeatability) and the spatial (subsurface) variability of the site. The discusser believes that these sources of uncertainty are independent and should be summed using the following equation:

[5]Pres. In-Situ Soil Testing, L.C., 2762 White Chapel Rd., Lancaster, VA 22503. E-mail: insitusoil@prodigy.net

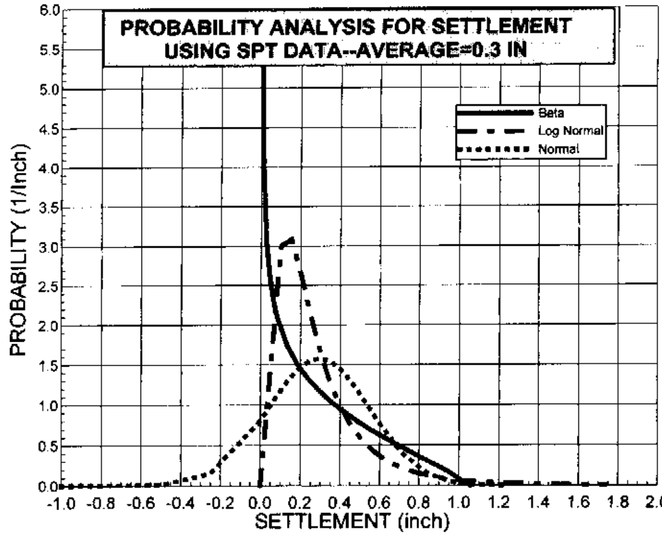


FIG. 7. Probability Analysis for Settlement Using SPT Data: Average = 0.3 in.

$$\sigma_{\text{overall}} = \sqrt{[(\sigma_{\text{model}})^2 + (\sigma_{\text{noise}})^2 + (\sigma_{\text{spatial}})^2]} \qquad (12)$$

where σ_{overall} = overall standard deviation; σ_{model} = standard deviation from model uncertainty; σ_{noise} = standard deviation from measurement noise; and σ_{spatial} = standard deviation from spatial variability.

The uncertainty from measurement noise for SPT can be as high as 45–100% (Schmertmann 1978; Kuhawy 1996). Wickremesinghe (1989) showed that measurement noise for piezocone (CPTU) equaled 5% and dilatometer tests (DMT) equaled 6% at the McDonald Farm test site in Vancouver. From case study data (Schmertmann 1986), the coefficient of variation for the DMT model for predicting settlement was 21% when the dilatometer is pushed and excluding quick clayey silts (Failmezger et al. 1999).

As shown in Table 13 and Figs. 7 and 8, the discusser analyzed the different probability distributions and the test and

TABLE 13. Analysis of Probability Distributions and Test and Analysis Methods

Probability distribution function	Test and analysis method	Average settlement (in.)	Standard Deviation From			Overall standard deviation[a]	Threshold settlement (in.)	Probability of unsatisfactory performance	Probability of success
			Spatial variability	Measurement noise	Model error				
Beta	SPT	0.30	0.059	0.148	0.198	0.254	0.50	0.21	0.79
Lognormal	SPT	0.30	0.059	0.148	0.198	0.254	0.50	0.14	0.86
Normal	SPT	0.30	0.059	0.148	0.198	0.254	0.50	0.21	0.79
Beta	SPT	0.30	0.059	0.148	0.198	0.254	0.90	0.02	0.98
Lognormal	SPT	0.30	0.059	0.148	0.198	0.254	0.90	0.03	0.97
Normal	SPT	0.30	0.059	0.148	0.198	0.254	0.90	0.01	0.99
Beta	DMT	0.30	0.059	0.018	0.062	0.087	0.50	0.00	1.00
Lognormal	DMT	0.30	0.059	0.018	0.062	0.087	0.50	0.02	0.98
Normal	DMT	0.30	0.059	0.018	0.062	0.087	0.50	0.01	0.99
Beta	DMT	0.30	0.059	0.018	0.062	0.087	0.90	0.00	1.00
Lognormal	DMT	0.30	0.059	0.018	0.062	0.087	0.90	0.00	1.00
Normal	DMT	0.30	0.059	0.018	0.062	0.087	0.90	0.00	1.00

[a]See Eq. (12).

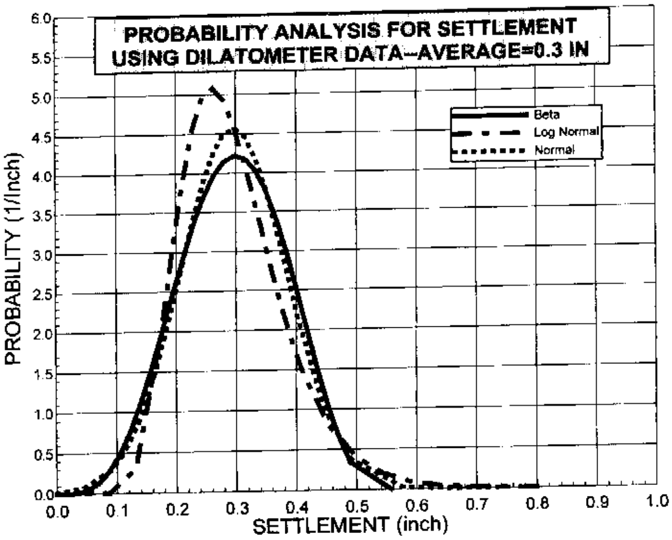


FIG. 8. Probability Analysis for Settlement Using Dilatometer Data: Average = 0.3 in.

analysis methods to determine their effects on the probability of unsatisfactory performance of exceeding a threshold settlement. The probability of success equals 1.0 minus the probability of unsatisfactory performance. Because the overall standard deviation was so high in comparison to the average value for the SPT case, the left side of all three distributions was distorted (Fig. 7). The probability analysis for settlement, however, focuses on the right side of the distribution curve. Because of its left skewness, the lognormal distribution for SPT with a threshold settlement of 0.5 in. gave a higher probability of success (86%) (unconservative) than beta or normal distributions (79%). A higher threshold settlement lessens the effect of the probability distribution.

However, the choice of test and analysis method in Table 13 had a much more significant effect than the probability distribution. The standard deviation from spatial variability was assumed to be equal to 20% of the average settlement value for all the SPT and DMT cases. The standard deviations from measurement noise and model uncertainty from SPT were much larger than those from DMT. In fact, they were huge! The overall standard deviation for the SPT was 86% of the average value, as compared with only 29% for the DMT. The discusser questions the value of using the SPT as a method to compute settlement altogether.

The high SPT variability shown above emphasizes that, for geotechnical design, the engineer should select the best available test and analysis method and attempt to minimize model uncertainty and measurement noise. The engineer should then focus on and quantify the spatial variability of the site, which is often beyond his or her control. Probabilistic design methods provide a good means to address variability. The probability distribution chosen for analyses should provide an appropriate result. The more heterogeneous the site is, the more uncertainty there is, the flatter the probability distribution will be, and the more conservative the design should be. The reverse is also true.

REFERENCES

Failmezger, R. A., Rom, D., and Ziegler, S. B. (1999). "SPT? A better approach to site characterization of residual soils using other in-situ tests." *Behavioral characteristics of residual soils*, B. Edelen, ed., ASCE, Reston, Va., 158–175.

Harr, M. E. (1977). *Mechanics of particulate media: a probabilistic approach*, McGraw-Hill, New York.

Kulhawy, F. H., and Trautmann, C. H. (1996). "Estimation of in-situ test uncertainty." *Uncertainty in the geologic environment: from theory to practice*, Vol. 1, C. D. Shackelford, P. P. Nelson, and M. J. S. Roth, eds., ASCE, New York, 269–286.

Schmertman, J. H. (1978). "Use the SPT to measure dynamic soil properties?—Yes, but . . . !" *Dynamic geotechnical testing*, American Society for Testing and Materials, West Conshohocken, Pa., 341–355.

Schmertmann, J. H. (1986). "Dilatometer to compute foundation settlement." *Proc., In Situ '86: ASCE Specialty Conf. on Use of In Situ Tests and Geotech. Engrg.*, ASCE, Reston, Va., 303–321.

Wickremesinghe, D. S. (1989). "Statistical characterization of soil profiles using in-situ tests." PhD thesis, Dept. of Civ. Engrg., The University of British Columbia, Vancouver.

Discussion by John A. Focht Jr.,[6] Fellow, ASCE, and John A. Focht III,[7] Member, ASCE

The author is to be commended for developing a rational technique for incorporating reliability into routine factor of safety analyses that can be understood and effectively utilized by a geotechnical engineering practitioner. Most practitioners, including the discussers, do not have enough confidence in "reliability based design" (RBD) to substitute it for their more conventional deterministic approaches. Most RBD papers suggest the blind application of statistical analyses of data without much engineering judgment regarding individual data points, trends in data, the type of design problem, or spatial variations within the data. The discussers believe that the application of RBD-based design approaches does not eliminate the need for sound engineering judgment. The author's proposed approach will certainly enhance the value of problem solutions for the engineering practitioner. The author also assumed that sound engineering judgment would be applied to both the data and the engineering problem, but still seemed to use the numerical average as the "most likely value." The discussers concur with the author's belief that the use of sound engineering judgment is always a criteria for properly evaluating engineering problems. This view is neither new nor unique; Karl Terzaghi very pointedly addressed the importance of sound engineering judgment in his May 1936 Presidential Address to the First International Conference on Soil Mechanics and Foundation Engineering (ICSMFE):

> The major part of the college training of civil engineers consists in the absorption of the laws and rules which apply to relatively simple and well-defined materials, such as steel or concrete. This type of education breeds the illusion that everything connected with engineering should and can be computed on the basis of a priori assumptions. As a consequence, engineers imagined that the future science of foundations would consist in carrying out the following program: Drill a hole into the ground. Send the soil samples obtained from the hole through a laboratory with standardized apparatus served by conscientious human automatons. Collect the figures, introduce them into the equations, and compute the result. Since the thinking was already done by the man who derived the equation, the brains are merely required to secure the contract and to invest the money. The last remnants of this period of unwarranted optimism are still found in attempts to prescribe simple formulas for computing the settlement of buildings or of the safety factor of dams against piping. No such formulas can possibly be obtained except by ignoring a considerable number of vital factors.

[6]Sr. Consult., Focht Consultants, Inc., 12226 Perthshire, Houston, TX 77024.

[7]Chf. Engr., Focht Consultants, Inc., 12961 Park Central, Ste. 1390, San Antonio, TX 78216.

PROJECT TYPE

The author included three problem examples in his paper. The first problem concerned the actual failure of underwater slopes in San Francisco Bay, while the other problems concerned hypothetical problems involving predictions of consolidation settlement in soft clays and footing settlements in sands. The slope stability example is different from the other two examples, because the problem involves a linear project. Other types of linear projects include tunnels, many wharf and bulkhead projects, open-cut trenches for pipelines, and electric transmission lines. The discussers have considerable experience with linear projects, which have (in the discussers' experience) always encountered spatially diverse conditions. The discussers believe that the author's approach is particularly useful for these types of projects.

The discussers believe that the missed prediction of the slope failure illustrates the difficulty in selecting the proper strength for analysis despite a very comprehensive investigation. Part of the error may have been with the tacit assumption that the "most likely" strength is the average of all the strength data, 2.8 in. specimens, 1.4 in. specimens, and field vane. The discussers would not have included the field vane results in the averaging process, because past experience suggests they would be too large; however, we would have been inclined to rely more on the 1.4 in. data than that from 2.8 in. specimens. A more significant departure from simple averaging would have been to anticipate spatial variations in the 2,000 ft length and pick a "most likely" strength line for design less than the average for all the borings. The borings may have not been drilled at locations where lesser strengths were more predominant. Selection of a "most likely" profile should also take into account that the actual failure surface can deviate for its "theoretical" position to pass through weaker material so that the average mobilized strength is less than the "apparent average strength" at that location. The senior discusser, based on 50 plus years of experience, would today have picked Profile 1 in Fig. 9 as the "most likely" strength profile based on the 1.4 in. data. Fig. 9 shows the reported strength data from 1.4 in. specimens and five potential interpretations of the available data: (1) the senior discusser's "most likely" line for the 1.4 in. data; (2) the 1973 Duncan average profile for all of the strength data; (3) the 2000 Duncan average profile for all of the data; (4) the linear regression of the 1.4 in. data; and (5) the linear regression of the 2.8 in. data. Profile 2 was taken from Fig. 5 of Duncan and Buchignani (1973), and an equivalent approximation of it was used for the original stability analyses. Profile 3 was taken from Fig. 2(b) of Duncan (2000). It would have been interesting to compare Profile 1 with ones drawn for individual borings for "average" and "most likely" interpretations. It is also significant to note that Profile 1 (our "most likely") actually differs only slightly from the "average σ" profile shown in Fig. 2(b). Profile 1 would have yielded a lesser computed factor of safety than those computed for Profiles 2 and 3, but perhaps still more than 1.0. This comparison again supports the discussers' conclusion that, for linear projects, use of the average of all data from all borings can be an overly optimistic prediction of the "most likely" strength prevailing somewhere along the project length.

The discussers are strong proponents of using a "most likely" strength profile based on judgment rather than a simple numerical average or the computed regression line, and in most instances the most likely strength profile will fall under the average profile. At the same time, they recognize that there is not a simple, uniformly applicable way to select the "most likely" profile other than to consider the scatter of the data, the character of the site geology, the project, and the type of design problem.

Some other linear problems, such as electric transmission

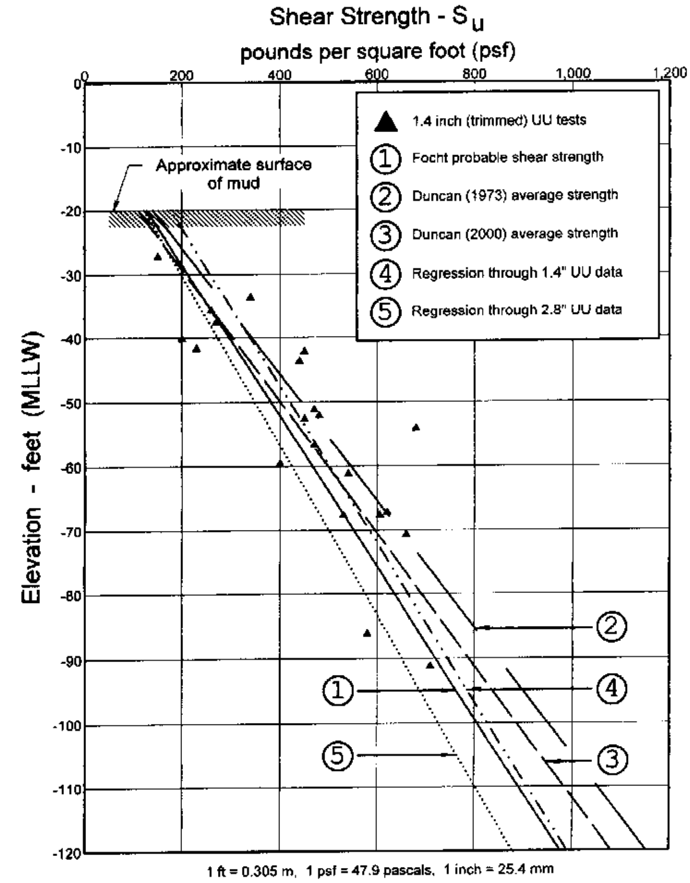


FIG. 9. Comparison of Various Shear Strength Profiles for San Francisco Bay Mud at LASH Terminal Site in San Francisco [after Duncan and Buchignani (1973) and Duncan (1999)]

lines, may cross a number of geologic units. Frequently, there might be only a few borings in each unit, so that "averaging" involves limited data at widely dispersed locations within a distinct unit. The junior discusser has been involved in the design and construction of a number of electric transmission lines in south and central Texas. The most recent long transmission line was a 138 kV line about 90 mi long, and the alignment traversed at least four distinct geologic units. Some of the borings were dry; others were near rivers and were probably wet for at least a portion of the year. The soil borings were typically a mile apart with towers about 700 ft apart. Even with the use of aerial photography, geologic maps, and ground reconnaissance, the choice of design parameters was difficult. The majority of structures were steel monopoles placed in drilled holes with concrete backfill. The client was a regional electric utility company with in-house design capabilities, and the client's representative indicated the utility recognized that the risk of failure of a given structure increases as the design factor of safety decreases. The client was willing to accept "some" risk of failure; the acceptable risk of failure was (in retrospect) not really defined. The design procedure for the monopoles, chosen jointly by the client's representative and the junior discusser, used a combination of expected average soil parameters at the monopole site, performance limits at full design loading (by limiting calculated ground-line deflection, pier-tip deflection, near-surface bearing pressures and pier rotation), and rotational stability with the design loads doubled (for an effective minimum design factor of safety against catastrophic failure of 2). The procedure utilized a modified soft clay *p-y* criteria [proposed by Evans and Duncan (1982), which permits the use of both cohesion and ϕ in calculating p], which resulted in some conservatism with respect to deflection and rotation predictions. However, the true risk of failure was (and still is) unclear. The use of the author's procedure would have allowed the junior discusser to provide the client with clearer choices regarding the economic tradeoff between initial construction cost and risk of failure of a portion of the structures. The junior discusser is unaware of any performance problems with any of the structures built as a part of this project; however, it is also doubtful that the structures have experienced the design loads.

HISTORICAL PERSPECTIVE

A historical recounting of Arthur Casagrande's procedure to recognize probability or analysis reliability in earth dam stability is appropriate for this discussion. Livingston Dam, designed in the 1960s on the Trinity River not far from Houston, has a 14,000-ft-long, 90-ft-high earth embankment and a 584-ft-long gated spillway in the embankment. The geotechnical consultant was Associated Soil Engineers, a joint venture of McClelland Engineers and National Soil Services, both of Houston. Ralph Reuss and the senior discusser were the principals. The critical foundation stratum, which is overlain by the usual alluvial deposits of clay over sand, is a stiff to very stiff, slickensided, highly plastic, heavily overconsolidated Miocene clay. As a result of the Waco Dam failure, we were concerned about the implications of residual strength of the Miocene clay in reference to embankment stability. Dr. Arthur Casagrande was retained as a special consultant, and he focused on the strength of the slickensided surfaces in the Miocene clay. We had already run a large number of consolidated-drained direct shear tests to determine both peak and residual strengths. At Arthur's suggestion, we ran three tests on specimens carefully trimmed to have a slickensided surface in the center of each. Because the stratum was deep, the tests were run under a vertical load of either 3.0 or 6.0 ksf and the strength envelopes were drawn through the origin ($c = 0$). We jointly agreed, based on those test results, that a strength of $\varphi = 17°$, $c = 0$ could be assigned to the Miocene clay. This value lies near the lower boundary of the peak strength envelopes ($\varphi = 14°$, $c = 0$) from the routine tests. The average peak strength parameters were $\varphi = 24°$, $c = 0.4$ ksf. Residual strengths had been found to be in the range of 10–13°. Based on triaxial tests, the pore pressure response in the Miocene clay to embankment loading was predicted to be 50% (which was later confirmed by piezometers during construction).

Because he believed the combination of assumptions was conservative, Arthur recommended a factor of safety of 1.4 for design. However, he added a second criterion for acceptability. While the results would not be reported, the proposed embankment must have a factor of safety of 1.0 computed for a residual strength of $\varphi = 10°$, $c = 0$. Arthur indicated that it was a routine procedure for him to seek a factor of safety of at least 1.0 for a least likely strength. His approach to selection of design strengths and acceptable factors of safety convinced the senior discusser that he understood at that time the concept that the author is proposing. The senior discusser further believes that Arthur Casagrande would have remained a proponent of relying heavily on judgment rather than just sophisticated computer analyses and would have endorsed the proposed concept.

SUMMARY

The discussers believe that the author's approach, when applied with good judgment, will improve the value of geotechnical analyses for many practitioners. The proposed approach provides the practitioner with the simple, direct tools needed to evaluate the risk of failure for a wide variety of geotechnical problems. The method is rational, integrates easily with existing design approaches, and provides the geotechnical engineering practitioner with the tools he/she needs to effectively communicate relative risk—and benefit—to the client. The discussers hope that the author's modified RBD approach will become widely used in geotechnical engineering practices for a wide variety of design problems.

REFERENCES

Evans, L. T., and Duncan, J. M. (1982). "Simplified analysis of laterally loaded piles." *Rep. No. UCB/GT/82-04*, Dept. of Civ. Engrg., University of California at Berkeley, Berkeley, Calif.

Terzaghi, K. (1936). "Presidential address: Relation between soil mechanics and foundation engineering." *Proc., 1st Int. Conf. on Soil Mech. and Found. Engrg.*, Cambridge, Mass.

Discussion by Demetrious C. Koutsoftas,[8] Member, ASCE

The author presents interesting applications of statistics and reliability theory to problems involving stability and settlement analyses in geotechnical engineering. The basic concepts presented in the paper provide a useful tool for geotechnical engineers in evaluating the risk associated with their design recommendations. However, the discusser is concerned that some engineers may be attempted to use this tool as a substitute for thorough investigations with high quality data that are essential in assessing risk and reliability for important projects. The following two examples illustrate this concern.

STRESS HISTORY OF BAY MUD

One of the examples used by the author to illustrate the use of the proposed simplified reliability analysis methodology involves the stress history of a deposit of soft Bay Mud at the Hamilton Air Force Base in Marin County, California. In Fig. 2(a), the author uses the results of a limited number of consolidation tests to assess the variability of the preconsolidation stress and estimated settlements.

Several years ago, the discusser's firm performed a series of Geonor vane shear tests at the Hamilton Air Force Base, which are summarized in Fig. 10. The results of the tests are divided into two groups to illustrate likely variations in the undrained strength and stress history over the site. On the left side are the results of four strength profiles that seem to be fairly consistent in terms of strength variations with depth as well as thickness of the layer. On the right side are the results of two other profiles where the Bay Mud is deeper, and where the strength is fairly constant within the depth interval of 10–40 ft (3.05–12.19 m).

The site has been reclaimed by placing approximately 4 ft (1.22 m) of fill after the author performed his investigation (Duncan 1965). The stress profiles in Fig. 10 have been modified to reflect the increase in in situ overburden stresses due to the fill.

The discusser has found from numerous cases in the Bay Area, where vane shear tests and consolidation tests are available, that the preconsolidation stress σ'_p, can be approximately estimated as four times the undrained vane shear strength after it has been corrected to make the strengths correspond to the direct simple shear. This is reasonably consistent with Mesri's (1989) empirical equation of

$$(S_u)_{mob} = 0.22\sigma'_p$$

[8]Prin., URS/Dames & Moore, 221 Main St., Ste. 600, San Francisco, CA 94105.

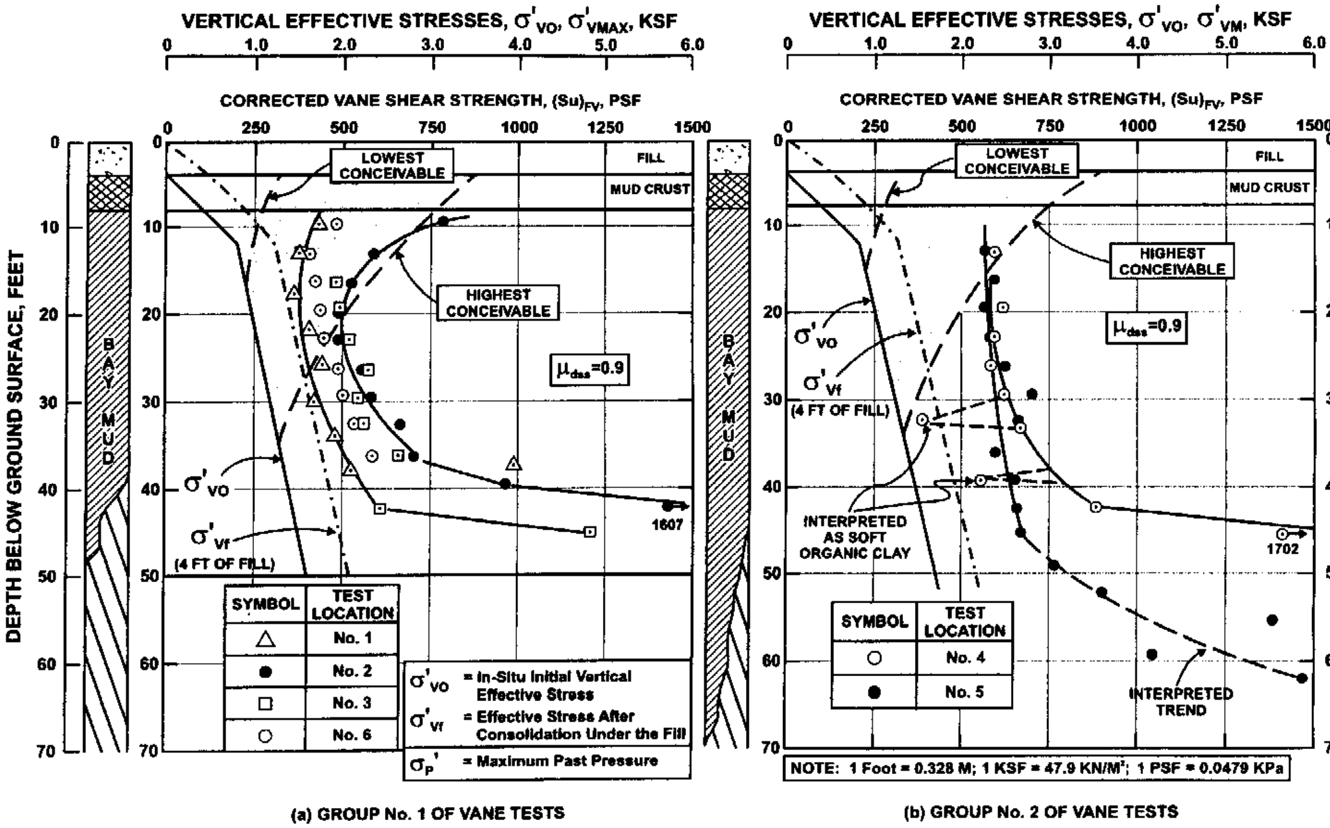


FIG. 10. Vane Shear Tests and Interpreted Stress History of Bay Mud: Hamilton Air Force Base

Fig. 10(a) shows the profiles of initial vertical effective stress, the vertical effective stress after full consolidation under 4 ft (1.22 m) of fill, and the profiles of the highest and lowest conceivable preconsolidation stress estimated by the author. It is clear that within the depth interval of 8–25 ft (2.44–7.62 m), the highest conceivable preconsolidation stress profile is consistent with the upper limit indicated by the vane shear data. However, at depths greater than 25 ft (7.62 m), the vane shear data indicate that the preconsolidation stresses are significantly higher than the highest conceivable values assumed by the author. The same general behavior is evident in Fig. 10(b), although the differences are now much greater than shown in Fig. 10(a).

The vane shear data also reveal some details that cannot be detected from conventional investigations. At test location number 4, there appear to be lenses of softer material that are believed to be highly organic clays. Explorations conducted by the discusser in a nearby site show that random layers of organic clay are sandwiched between the Bay Mud. Also, the data show clearly a transition zone near the base of the Bay Mud layer where strengths increase rapidly with depth and the soil is heavily preconsolidated. The author's data extend down to a depth of only 20 ft (6.1 m), even though Figs. 2(a) and 5 indicate that the mud extends at least to a depth of 30 ft (9.14 m). The transition zone at the base of the layer was not detected by the author's investigation. The presence of this transition zone is important in evaluating both the magnitude and the rate of settlements. The stiff clay at the base of the mud forms an impermeable drainage boundary, which is critical in assessing time rates.

It is evident from the data presented in Fig. 10 that

1. Vane shear data provide very valuable information from which the stress history of the Bay Mud can be evaluated. Even if one does not want to accept the numerical values of the σ_p' calculated from the vane strengths, the data certainly provide a much more reliable guide for estimating upper and lower limits for the preconsolidation stress than those obtained from limited laboratory testing.
2. The vane shear data reveal variations in the strength and stress history of the Bay Mud that conventional investigations are likely to miss, such as the potential presence of lenses of soft organic clay that may be more compressible than Bay Mud, as well as the presence of transition zones near the base of the Bay Mud, which affect both the magnitude and the rate of settlements.
3. Vane shear tests provide an economical and practical method for evaluating substantial variations in stress history and undrained shear strength across a site. Hence, they constitute a very valuable supplement to laboratory testing.

The above example illustrates the usefulness of the vane shear test and points out that even the most experienced geotechnical engineer may not be able to assess reliably the characteristics of soft clay deposits from "meager data." In this case, a much more thorough site investigation would be required to improve the reliability of the geotechnical analyses and recommendations.

UNDERWATER SLOPE FAILURE

Another case used to illustrate the application of reliability analysis involved a deep cut in soft San Francisco Bay Mud at the LASH terminal in San Francisco, which failed during excavation, as detailed by Duncan and Buchignani (1973). Their explanation was that the slope failed primarily because of loss of strength due to undrained creep. The author reexamines this case and concludes that the estimated variations in the undrained shear strength and buoyant unit weight of the Bay Mud resulted in a probability of failure of 18% associated with the calculated factor of safety of 1.17. The author concludes that a probability of failure of 18% was "too high to be acceptable," and, perhaps, if it was realized that the probability of failure was that high, the design might have been changed to reduce the probability of failure.

This is a very interesting case, because, in spite of the fact that the 1973 paper presented a comprehensive assessment of a number of factors affecting the undrained strength of the mud, there are several other factors not discussed in the paper that might have affected the observed failure, and which are relevant to the reliability analysis presented by the author. These factors are discussed below.

Stress History and Undrained Strength Parameters

The slope stability analyses were based on undrained shear strengths measured from unconsolidated, undrained (UU) triaxial compression tests and from field vane shear tests. The interpreted strengths from UU and field vane shear tests were within 8% of each other, which, given the scatter in the data, is interpreted by the discusser to suggest that the UU strengths and vane shear strengths are essentially the same. The results of test comparisons on many projects performed by the discusser in the last 16 years, involving San Francisco Bay Mud, are consistent with this finding. However, the agreement in strengths is entirely fortuitous, for the following reasons: (1) Undrained strengths of Bay Mud measured from UU tests, even on high quality samples, where strains to failure are on the order of 3% or less, underestimate the in situ strength in compression. The primary reason for this is that the effective stresses, even in high quality samples, are significantly lower than the in situ effective stresses, as explained by Ladd and Lambe (1963). The discusser's experience has been that UU strengths underestimate the in situ strength of Bay Mud in compression by about 30%. (2) Vane shear strengths adjusted using a factor 0.9 are on the average equal to the undrained shear strengths determined from direct simple shear tests, and they also underestimate the strength of the mud in compression, because of anisotropic effects, by about 30%. This relationship between vane shear strengths and strengths in compression is consistent, although somewhat lower than the correction factor of 1.35 recommended for Bay Mud by Bjerrum et al. (1972). Fig. 11 shows the results of three vane shear soundings performed along the Islais Creek, not far from the LASH terminal, and they are compared with the undrained strength profiles in compression, extension, and simple shear estimated based on SHANSEP (Ladd and Foott 1974) principles. It is evident that the vane shear strengths adjusted by a factor of 0.9 agree reasonably well with the strengths for direct simple shear. As seen in Fig. 11, and as discussed in more detail by Koutsoftas et al. (2000), the Bay Mud is anisotropic with respect to its undrained shear strength. It is therefore necessary to carefully consider how the results of the UU strength tests and the field vane shear tests apply to the stability of the cut slope.

Another important fact revealed from the data shown in Fig. 11 is that the mud appears to be lightly preconsolidated down to a depth of 35 ft (10.67 m), even though the mud is under 20 ft (6.1 m) of fill placed in the 1920s. This suggests that the mud must have been preconsolidated prior to the placement of the fill, probably due to aging. The three vane soundings shown in Fig. 11 were performed right along the north bank of Islais Creek, where the mud extends down to 100–120 ft (30.48–36.58 m) deep. The stress history of the mud, prior to the placement of the fill at this location, could not have been much different from that of the mud at the LASH terminal. Therefore, based on the results shown in Fig. 11, it could be concluded that the mud at the LASH terminal must have been preconsolidated also. Close examination of the consolidation and vane shear data included in the 1973 paper by Duncan

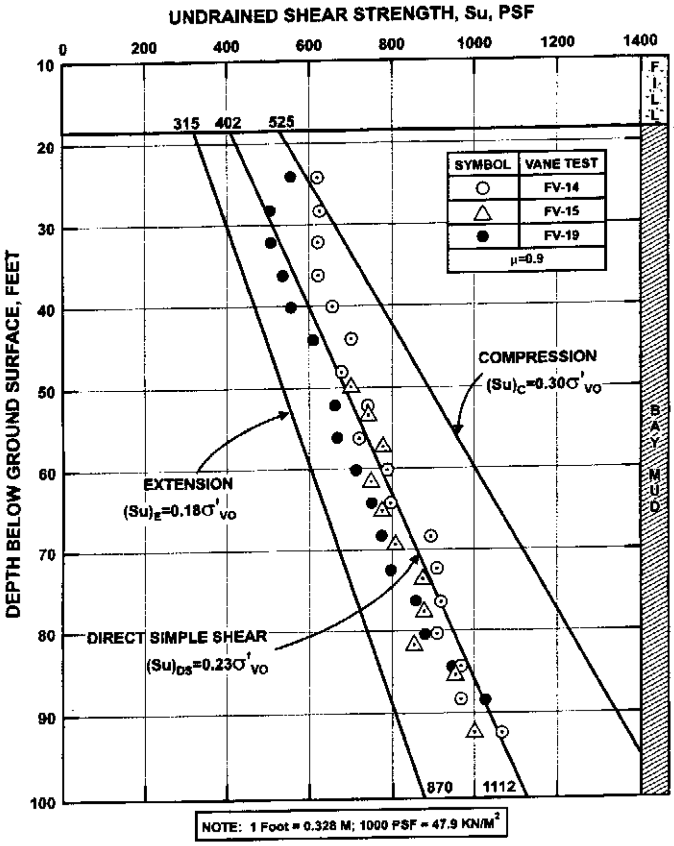


FIG. 11. Undrained Shear Strength Profiles: San Francisco Bay Mud at Islais Creek Contract D Site

and Buchignani also suggest that the mud is preconsolidated. The uncorrected vane shear strengths (figure 6 of the 1973 paper) indicate approximate undrained strength ratios (S_u/σ'_{vo}) of 0.66, 0.46, 0.41, and 0.38, respectively, at elevations of −30 ft (−9.14 m), −40 ft (−12.19 m), −50 ft (−15.24 m), and −60 ft (−18.29 m). These values are significantly higher than the corresponding undrained strength ratio of 0.25 (uncorrected) for normally consolidated Bay Mud and lead to the conclusion that the mud at the LASH terminal was overconsolidated, as suggested by the data in Fig. 11. Given the age of the Holocene mud, and that the mud at the LASH terminal site is not likely to have been exposed to desiccation or other externally applied stresses from human activity, it would be reasonable to conclude that the mud in the Bay would be overconsolidated due to aging (Bjerrum 1973), with probable overconsolidation ratios in the range of 1.6–2.0. Fig. 12 shows the index properties and stress history of Bay Mud at a site in the Bay located just outboard of the end of the runways at San Francisco Airport. The mud is below water and the site is far enough from the old shoreline that it is unlikely it had been exposed to desiccation or had been subjected to external stresses from human activities. The results of five consolidation tests performed on undisturbed samples obtained within the upper 30 ft (9.14 m) of the mud indicate preconsolidation stresses that range between 1.6 and 2.0 times the in situ vertical effective stresses. The results shown in Fig. 12, particularly the trend for increasing preconsolidation stress with depth, lead to the conclusion that the preconsolidation is the result of aging. Given the geologic history of the Bay Area, it is the discusser's assessment that the stress history of the mud at the airport site would be quite similar to the stress history of the mud at the LASH terminal. The results shown in Fig. 12 reinforce the conclusion that the mud in the Bay, including the LASH terminal site, is overconsolidated, most likely due to aging.

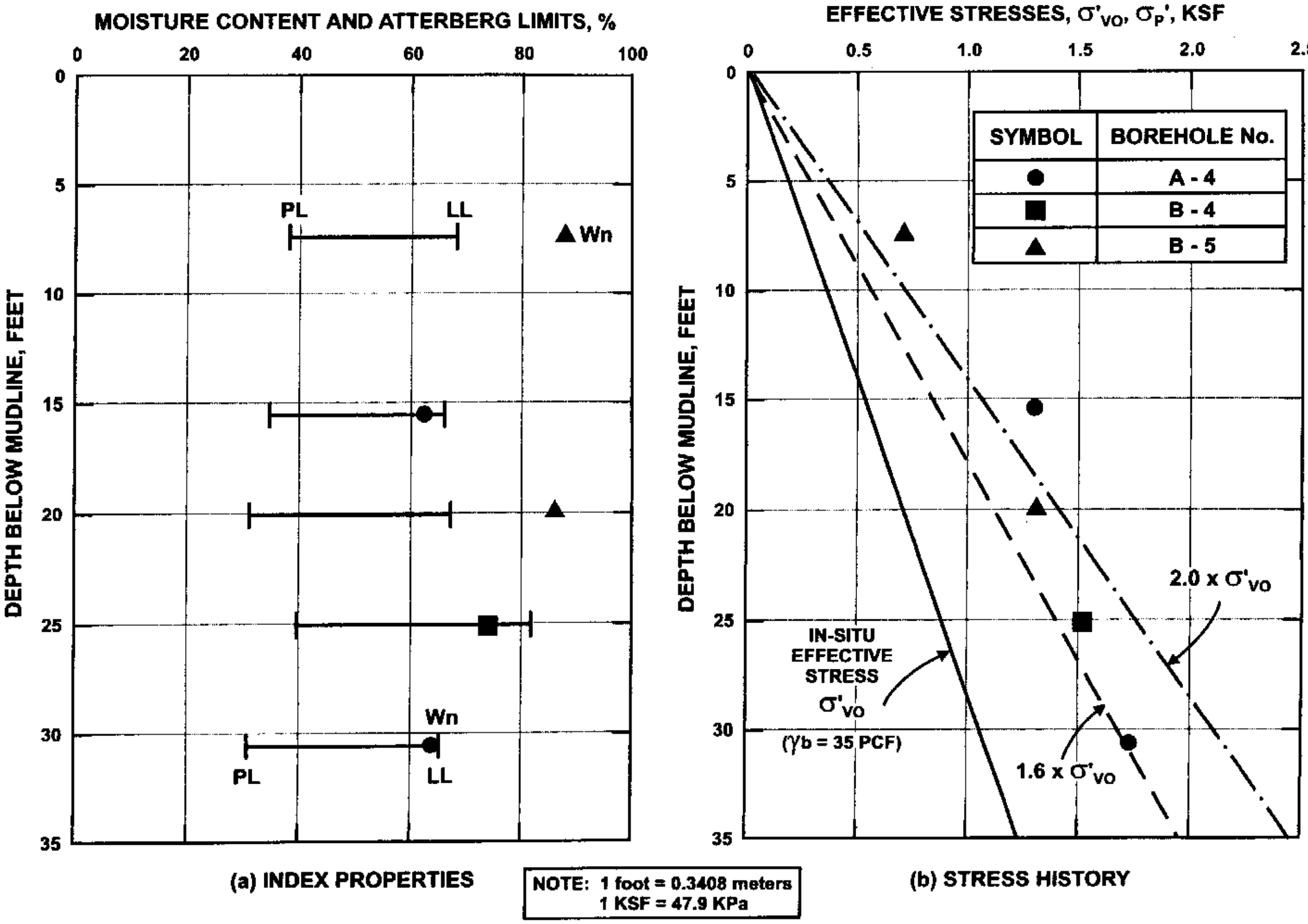


FIG. 12. Index Properties and Stress History of Bay Mud from Site Offshore San Francisco Airport Runway No. 28

If the mud was indeed overconsolidated, the dissipation of negative excess pore pressures, generated by the excavation process, could have been faster than anticipated by the author and could have contributed to loss of strength due to swelling.

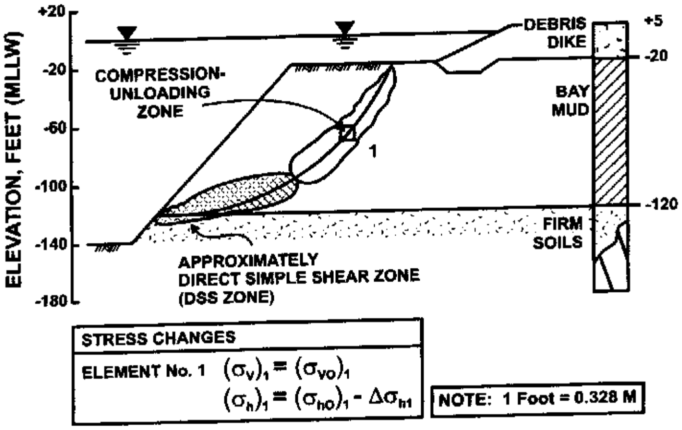
Stability Analysis

Fig. 13 shows the critical failure as it was depicted in figure 9 of the paper by Duncan and Buchignani (1973). Much of the critical failure surface shown in Fig. 12 represents shear in compression (designated the "compression unloading" zone), while shear along the rest of critical failure surface can be approximated by simple shear. As indicated by the stress paths in Fig. 13(b), negative excess pore pressures are generated during shear in the compression unloading zone (Fig. 13). It is not intuitively obvious that the loss of strength due to creep, under conditions that involve generation of positive pore pressures (in triaxial compression tests like those performed by the author), is applicable under conditions that involve generation of negative excess pore pressures. Hence, the assumption of loss of strength due to creep is questionable.

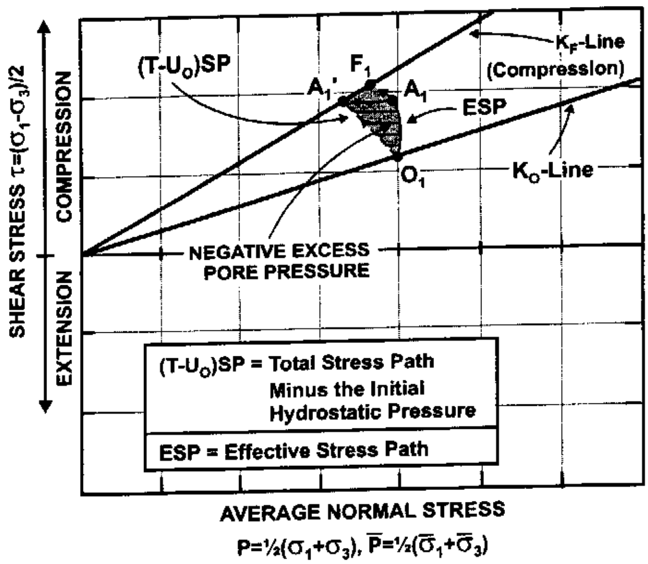
If the vane shear strengths were corrected for anisotropy (as per Fig. 11) before being used to represent the in situ strengths along the portion of the critical failure surface that passes through the "compression-unloading" zone, the calculated factor of safety would be significantly higher than the value of 1.17 calculated by the author. If that was the case, then the mud must have experienced a significant loss of strength during excavation of the trench to cause the failure. The failure could not be explained either by the statistical variations in strength and buoyant unit weight, or by the strength reduction due to creep. The required reduction in undrained strength to cause the failure could have occurred only if there was significant dissipation of negative excess pore pressures. As described below, the geologic conditions at the LASH site might have been favorable, at least over some portions of the site, for rapid dissipation of negative excess pore pressures and might have caused the failure.

The discusser has conducted numerous explorations within the Islais Creek basin and has reviewed borehole records from many previous explorations conducted by Dames and Moore and others in the area, including Piers 80, 90, and 92 in the vicinity of the LASH terminal. In a large number of cases, a marine sand layer was encountered directly below the mud. The marine sand varies from fairly clean fine sand to clayey sand with stringers of Bay Mud. In some instances the marine sand is less than 5 ft (1.52 m) thick and is underlain either by permeable dense (Colma) sand or Old Bay Clay. Often, the presence of the thin marine sand layer may not be detected, especially when sampling at depth intervals of 5 ft (1.52 m) or more, which is fairly standard in the Bay Area at depths of 100 ft (30.48 m) or more. It is less common to find Old Bay Clay directly below the Bay Mud. However, in some of the boreholes drilled at Pier 80, which is located just north of the Islais Creek Channel near the LASH terminal, Old Bay Clay was found directly below the mud, acting as an impermeable drainage boundary at the base.

If a layer of reasonably permeable sand (i.e., much more permeable than Bay Mud) was present below the Bay Mud, as many of the boreholes show, the negative excess pore pressures near the base of the mud could dissipate very rapidly. Approximately 30% of the critical failure surface (and perhaps also of the actual failure surface) passes directly over or within



(a) APPROXIMATE STRESS STATES ALONG CRITICAL FAILURE SURFACE



(b) EFFECTIVE STRESS PATH ELEMENT No. 1

FIG. 13. Approximate Stress States and Stress Paths along Failure Surface

a few feet of the interface with the underlying "firm" soils. Dissipation of negative excess pore pressures along the base of the mud could have caused loss of strength over a significant portion of the failure surface to cause the observed failure.

Based on the above observations, it is the discusser's conclusion that the stability of the underwater slopes at the LASH terminal was critically dependent on the drainage conditions at the base of the mud. Where the marine sand layer was present, rapid drainage must have occurred, leading to the observed failures. Where Old Bay Clay, or clayey sand, were present below the mud, drainage must have been inhibited, thus avoiding loss of strength due to dissipation of excess pore pressures, and the slopes remained stable. Given the general stratigraphic conditions that prevail in the Islais Creek area, it is perhaps fortunate that slope failures were not more prevalent.

CONCLUSION

The two cases reviewed in this discussion lead to the conclusion that reliability analyses can be of little benefit to the geotechnical engineer, unless such analyses are based on a realistic assessment of field conditions and employ design parameters that truly represent the in situ behavior of the soil. Undrained strengths determined from UU tests rarely provide realistic assessments of in situ strengths because of sample disturbance, strength anisotropy, and other factors, as discussed by Germaine and Ladd (1988) and Ladd et al. (1998). Likewise, vane shear strengths need to be corrected for the effects of rate of shear and anisotropy before they are used for stability analysis, as discussed by Bjerrum (1972, 1973) and Ladd et al. (1977).

Finally, the discusser would modify the author's claim that "an essential component of the art of geotechnical engineering is the ability to estimate reasonable values of parameters based on meager data." It appears from the two cases reviewed in this discussion that using meager data often involves significant risks that one might reach the wrong conclusions. If meager data are used for design, one should also recognize the inherent risks and utilize higher factors of safety to account for the uncertainties. The discusser believes that the true art of geotechnical engineering should also include knowing when a situation demands comprehensive exploration and testing to correctly characterize a given field situation. When this is accomplished systematically, we can then expect a real improvement in the practice of geotechnical engineering and improved reliability of our designs.

REFERENCES

Bjerrum, L., Frimman Claussen, C. J., and Duncan, J. M. (1972). "Earth pressures on flexible structures—a state-of-the-art report." *Proc., 5th European Conf. on Soil Mech. and Found. Engrg.*, 2, 169–196.

Bjerrum, L. (1973). "State-of-the-art report: problems of soil mechanics and construction on soft clays and structurally unstable soils." *Proc., 8th Int. Conf. on Soil Mech. and Found. Engrg.*, Moscow, 3, 111–159.

Germaine, J. T., and Ladd, C. C. (1988). "State-of-the-art: triaxial testing of saturated cohesive soils." *Advanced triaxial testing of soil and rock*, ASTM, West Conshohocken, Pa., 421–459.

Koutsoftas, D. C., Frobenius, P., Wu, C. L., Meyersohn, D., and Kulesza, R. (2000). "Deformations during cut-and-cover construction of MUNI Metro Turnback Project." *J. Geotech. and Geoenvir. Engrg.*, ASCE, 126(4), 344–359.

Ladd, C. C., and Foott, R. (1974). "New design procedure for stability of soft clays." *J. Geotech. Engrg. Div.*, ASCE, 100(7), 763–786.

Ladd, C. C., Foott, R., Ishihara, K., Schlosser, F., and Poulos, H. G. (1977). "Stress, deformation, and strength characteristics." *Proc., 9th Int. Conf. on Soil Mech. and Found. Engrg.*, Tokyo, 2, 421–494.

Ladd, C. C., and Lambe, T. W. (1963). "The strength of undisturbed clay determined from undrained tests." *Laboratory shear testing of soils*, ASTM, West Conshohocken, Pa., 342–371.

Ladd, C. C., Young, G. A., Kraemer, S. R., and Burke, D. M. (1998). "Engineering properties of Boston blue clay from special testing program." *Proc., Geo-Congr. 98*, ASCE, Reston, Va., 1–24.

Lambe, T. W. (1967). "Stress path method." *Soil Mech. and Found. Engrg.*, ASCE, 93(6), 309–331.

Mesri, G. (1989). "A reevaluation of $S_{u(mob)} = 0.22\sigma_p'$ using laboratory shear tests." *Can. Geotech. J.*, Ottawa, 29(1), 162–164.

Discussion by Charles C. Ladd,[9] Honorary Member, ASCE, and Gregory Da Re,[10] Associate Member, ASCE

The author presents convincing arguments for using reliability analyses to estimate the probability of failure (P_f) associated with factors of safety (F) obtained from conventional stability analyses. The discussers especially endorse his view that selection of design values of F should reflect the uncertainties in the stability analyses and the consequences of failure, rather than the "one size fits all" approach commonly

[9]Prof., Dept. of Civ. and Envir. Engrg., MIT, Cambridge, MA 02139.

[10]Postdoct. Assoc., Dept. of Civ. and Envir. Engrg., MIT, Cambridge, MA 02139.

specified in regulations or adopted by practitioners. Unfortunately, many engineers do not appreciate the fact that a computed factor of safety per se has little physical meaning, unless one also has some measure of its accuracy. As clearly stated in the paper, a reliability analysis provides a systematic method for evaluating the combined influence of uncertainties in the parameters affecting F, so that one can assess the actual degree of safety (at least in relative terms) via the corresponding probability of failure.

The author recommends using published values (Table 3) and/or the three-sigma rule for estimating the standard deviation (σ) of parameters for "the common situation in which limited amounts of data are available and many properties are estimated using correlations," i.e., in lieu of using (3), which defines σ. These values of σ are used to compute the resulting standard deviation in the factor of safety (σ_F). The values of σ_F and F_{MLV} (the most likely value of F based on best estimates for all parameters) are then used with Table 2 to obtain the probability of failure (P_f). The discussers feel that this simple approach may lead to misleading results unless the user also understands the different sources of uncertainty, which can affect selection of both the best estimate of parameters and reasonable values for their standard deviation. These different sources also affect the physical meaning of P_f.

The discussion focuses on undrained slope stability analyses and starts with a description of the different sources of uncertainty in undrained shear strength (S_u). The discussers then apply this approach to the soils data for the LASH case history, leading to estimates of P_f based on statistical analyses of the data rather than the three-sigma rule. The effect of "model error" is also introduced.

SOURCES OF UNCERTAINTY IN S_u AND MEANING OF P_f

A reliability analysis should distinguish between the two basic types of uncertainty in a soil property such as S_u (e.g., Christian et al. 1994). One type is the uncertainty in the best estimate value or mean trend with depth of S_u, which is called the systematic uncertainty (denoted by σ_{sy} or the coefficient of variation, COV_{sy}). The second type of uncertainty is the spatial variation in S_u about the mean trend and, in theory, should reflect both the magnitude of the fluctuation about the mean trend (σ_{sp} or COV_{sp}) and the scale (distance) over which this fluctuation occurs. Note that identifiable zones of stronger or weaker soil should be assigned different mean trends.

Suppose that a reliability analysis produces a best estimate factor of safety of $F_{MLV} = 1.25$ and $\sigma_F = 0.25$, which leads to $P_f = 15\%$ from Table 2. If $\sigma_F = \sigma_{sy} = 0.25$ (i.e., all due to uncertainty in the best estimate of S_u), this implies that there is a 15% chance that the entire length of slope will fail. But if $\sigma_F = \sigma_{sp} = 0.25$ (i.e., all due to spatial variation in S_u about a mean trend having no error), this implies that, depending on the scale of the fluctuations, up to 15% of the length of the slope will fail.

Selection of the best estimate of S_u and analyses to obtain the systematic and spatial uncertainty must address four problems in evaluating strength data. Using S_u from field vane tests, $S_u(FV)$, as the example, they are

1. Scatter in $S_u(FV)$ due to real spatial variability
2. Scatter in $S_u(FV)$ due to random testing errors (noise)
3. Error in the mean $S_u(FV)$ due to the limited number of tests, called the statistical uncertainty
4. Error in the mean $S_u(FV)$ due to measurement bias

Items 3 and 4 produce the systematic uncertainty (i.e., the potential error in the selected best estimate of S_u): $\sigma_{sy}^2 = \sigma_{st}^2 + \sigma_{bias}^2$, where σ_{st} and σ_{bias} denote the statistical and bias components. σ_{st} is easily obtained from conventional statistics, even though the data may be meager. For a layer with a constant S_u, $\sigma_{st}^2 = \sigma_{su}^2/n$, where σ_{su} = standard deviation of n values of the measured S_u. Estimates of σ_{bias} are generally far more difficult. However, when using field vane data, one would usually base the best estimate on values of $\mu S_u(FV)$, where μ is Bjerrum's (1972) empirical correction factor. For this case, the uncertainty (bias error) comes from scatter about the recommended μ versus PI correlation (as discussed later).

For items 1 and 2, it is difficult to separate real spatial variability from random scatter, and even more difficult to estimate the scale of spatial fluctuations about the mean. For simplicity, the first discusser often assumes that σ_{sp}^2 equals one-half of the total scatter, σ_{su}^2. The total uncertainty in S_u then becomes $\sigma_T^2 = \sigma_{sy}^2 + 0.5\sigma_{su}^2$ and values of P_f using σ_T represent the likelihood of having "small" failures within the unknown scale of spatial fluctuations. In any case, the discussers believe that situations in which the systematic uncertainty dominates will generally be of greater concern than those involving mainly spatial variations in S_u. In other words, serious errors in the best estimate of S_u are generally more important than fluctuations about the selected mean trend.

The above discussion raises questions about the coefficients of variation cited in Table 3. Are the ranges meant to reflect the total uncertainty ($\sigma_T \div$ best estimate) or only the systematic error in the mean? Also, should the quoted values for the range in S_u depend on the type of test and heterogeneity of the soil and that for S_u/σ_v' also depend on the overconsolidation ratio (OCR)?

LASH UNDERWATER SLOPE FAILURE

Available Soils Data

Fig. 14 plots measured strengths from 18 field vane tests and 21 UUC tests run on 1.4 in. (36 mm) diameter specimens. These data were scaled from Duncan and Buchignani (1973) and exclude results from 2.8 in. (71 mm) diameter UUC tests that were considered "somewhat disturbed." Line 1 was selected by these authors to represent the UUC strengths and is very similar to a linear regression (LR) line that excludes the two deepest tests. Fig. 15 plots stress history data and shows

- The effective overburden stress, σ_{vo}', based on buoyant unit weights having an average equal to that selected by the author.
- Values of minimum-maximum and average preconsolidation pressure (σ_p') obtained via Casagrande from conventional "one-day" incremental oedometer tests. A linear regression line is also plotted (it excluded two of the 11 tests), which indicates that the deposit is actually slightly overconsolidated.
- Values of σ_p' predicted from the field vane tests using Chandler's (1988) equation:

$$OCR = \left(\frac{S_u(FV)/\sigma_{vo}'}{S_{FV}}\right)^{1.05} \quad (13)$$

with $S_{FV} = 0.20$ for PI = 20%. Linear regression on these data produce a line parallel to, but significantly higher than, the lab σ_p' line.

Reliability Analysis

Table 14 summarizes the results of three reliability analyses for the LASH underwater slope having a height of 100 ft (30.5 m). The author's is listed first. It predicts that most of the P_f = 18% comes from the estimated uncertainty in S_u as compared with the estimated uncertainty in buoyant unit weight. Because one does not know the systematic versus spatial uncertainty,

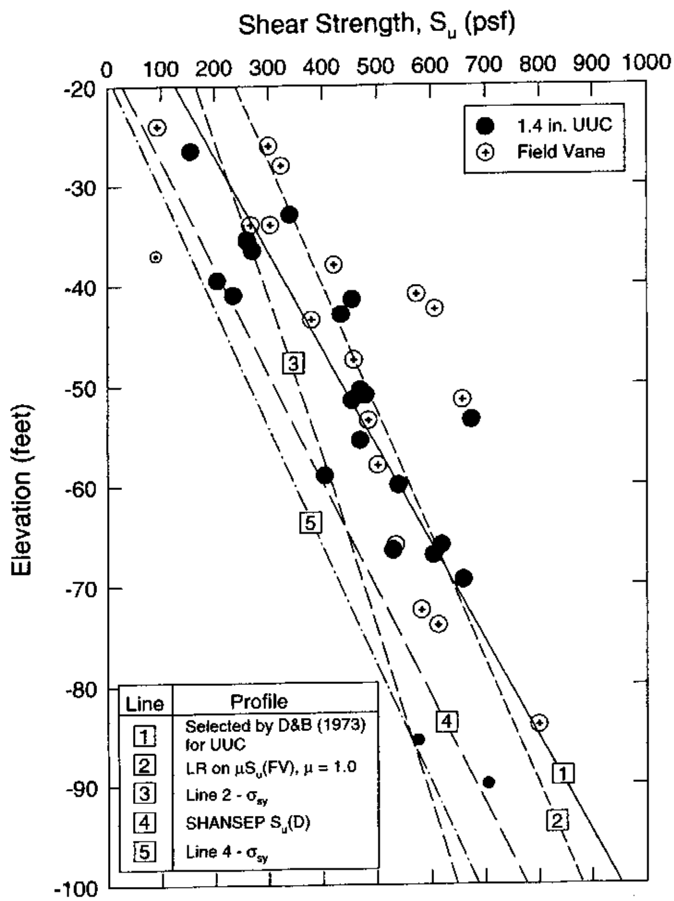


FIG. 14. Undrained Strengths from 1.4 in. UUC and Field Vane Tests and SHANSEP. Note: Small Symbols Excluded for LR Analysis (1 in. = 25.4 mm; 1 ft = 0.305 m; 1 psf = 0.0479 kPa)

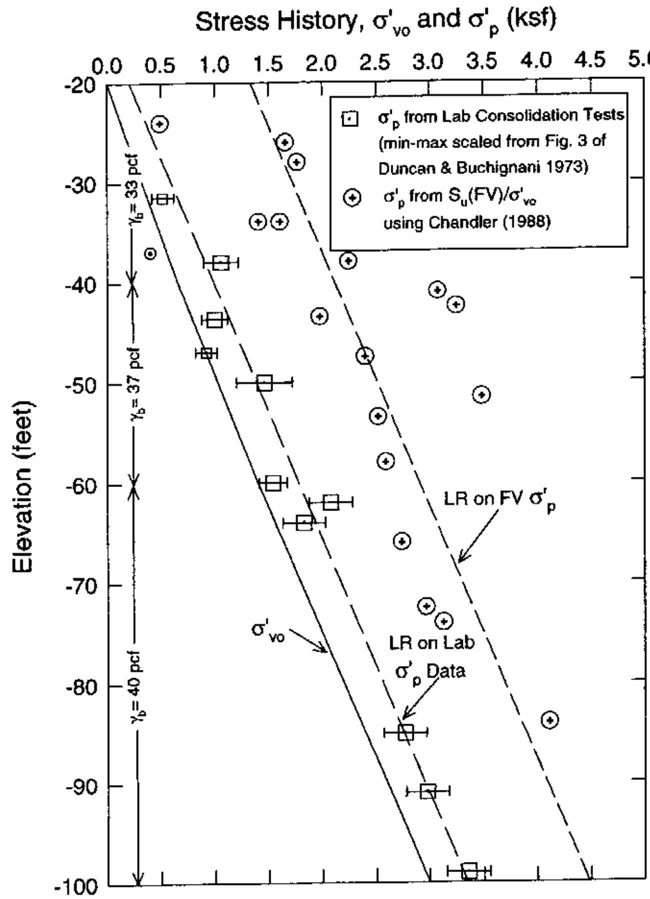


FIG. 15. Stress History from Laboratory Consolidation and Field Vane Tests. Note: Small Symbols Excluded for LR Analysis (1 ft = 0.305 m; 1 ksf = 47.9 kPa; 1 pcf = 0.157 kN/m³)

the fact that 22% of the slope length failed is indeed fortuitous. Also, Line 1 in Fig. 14 produces $F = 1.25$ using the UTEXAS3 computer program (Wright 1991), as compared with $F = 1.17$ reported by Duncan and Buchignani (1973), who replaced the line by 20 ft (6.1 m) thick layers of constant S_u material.

Analysis Using Field Vane Strengths

The second analysis is based on statistical treatment of the field vane (*FV*) data using a correction factor of $\mu = 1.00$ based on the reported PI = 20% for Bay Mud. Hence, the linear regression Line 2 in Fig. 14 represents the best estimate, which produces $F_{MLV} = 1.19$. The statistical uncertainty in the location of Line 2 reflects the vertical distribution and scatter in $S_u(FV)$ and equals

$$\sigma_{st}(\text{psf}) = \left(646\left[1 + \frac{(\text{El.} - 48)^2}{318}\right]\right)^{0.5} \qquad (14)$$

which varies with depth between about 25 and 105 psf (1.2 and 5.0 kPa). Based on the data in Figure 51 of Ladd et al. (1977), the uncertainty in μ at PI = 20% has a coefficient of variation of $COV[\mu] = 20\%$. Thus, the total systematic uncertainty at any depth equals

$$\sigma_{sy}(\text{psf}) = [\sigma_{st}^2 + (0.20 \cdot S_u \text{ on Line } 2)^2]^{0.5} \qquad (15)$$

The mean minus σ_{sy} profile produces $F = 0.93$, $\sigma_F = 0.26$, and

TABLE 14. Results from Reliability Analysis

Basis for analysis	Mean S_u and factor of safety	Source of uncertainty	Partial σ_F (COV, %)	Total σ_F (COV, %)	P_f from Table 2 %	Remarks
Author's Table 5	Fig. 2, $F_{MLV} = 1.17$	$COV[S_u] = 13\%$	—	0.155 (13.2)	13	S_u contributes most of P_f
		Plus $COV[\gamma_b] = 8.7\%$	0.10 (8.5)	0.185 (15.8)	18	
Corrected field vane for μ = 1.00	Fig. 14, Line 2, F_{MLV} = 1.192	Statistical and $COV[\mu] = 20\%$	—	0.262 (22.0)	24	Probability of "large" failure
		Statistical and[a] $COV[\mu] = 25\%$	—	0.317 (26.6)	29	
SHANSEP DSS $S = 0.22$, $m = 0.8$	Fig. 14, Line 4, F_{MLV} = (0.997)(1.10) = 1.097	Systematic uncertainty[b] in S_u	—	0.133 (12.1)	24	Probability of "large" failure
		Plus model error, $COV[3D]$ = 5%	0.055 (5.0)	0.144 (13.1)	26	
		Plus spatial uncertainty in S_u	0.116[c] (10.6)	0.185 (16.9)	32	Probability of "small" failure

[a] S_u = Fig. 14, Line 3.
[b] S_u = Fig. 14, Line 5, F = (0.876)(1.10) = 0.964.
[c] $0.116 = (0.185^2 - 0.144^2)^{0.5}$.

P_f = 24%. Line 3 in Fig. 14 equals the mean minus $\sigma_{sy}S_u$ computed with $COV[\mu]$ = 25% and gives F = 0.875 and P_f = 29%. The actual μ could be significantly less than 0.75, based on the high σ'_p predicted from field vane tests relative to the laboratory (Fig. 15). A lower than usual μ would be expected if the field vane tests were run with a torque wrench rather than a gear system (failure in seconds rather than minutes). In any case, this analysis predicts an unacceptable level of risk, even without consideration of the spatial uncertainty in S_u. The discussers also believe that this approach is more realistic and less dependent on judgment than the one based on the three-sigma rule lines in Fig. 2(b). Does the author agree that there were sufficient field vane data for statistical analyses?

Analysis Using SHANSEP

The third analysis used the SHANSEP equation

$$\frac{S_u}{\sigma'_{vo}} = S(OCR)^m; \quad OCR = \frac{\sigma'_p}{\sigma'_{vo}} \tag{16}$$

Parameters for the best estimate of the direct simple shear (DSS) strength [$S_u(D)$ = Line 4 in Fig. 14] are:

- S = 0.220 from three CK_0UDSS tests run by MIT on normally consolidated Bay Mud having a PI = 25% (DeGroot et al. 1992)
- m = 0.8, recommended by Ladd (1991), p. 584
- σ'_{vo} = line in Fig. 15
- σ'_p = LR line in Fig. 15, which is then increased by 10% to convert the one-day compression curve to an estimated end-of-primary curve

The UTEXAS3 program with $S_u(D)$ gives F = 0.997. This value is then increased by 10% to model three-dimensional end-effects based on the mean value from case histories of failures in Azzouz et al. (1983). Hence, the best estimate becomes F_{MLV} = 1.097. (Note: Four significant figures are shown for numerical consistency, not for implied accuracy. Also, since Bjerrum's μ factor already includes end-effects, this 10% correction was not applied to the FV factor of safety.)

Baecher and Ladd (1997) show that the uncertainty in the SHANSEP S_u profile, assuming no error in σ'_{vo}, equals

$$COV^2[S_u] = COV^2[S] + m^2COV^2[\sigma'_p] + \ln^2(OCR)\cdot\sigma_m^2 \tag{17}$$

Values used to compute the systematic uncertainty in $S_u(D)$ were:

1. $COV[S]$ = 10% estimated for bias (the three DSS tests produced a negligible statistical error of less than 2%)
2. $\sigma[m] = \sigma_m$ = 0.05 estimated for bias
3. $COV[\sigma'_p]$ decreased from 29 to 7% with depth, mainly due to the increasing mean σ'_p and includes one statistical and two bias components:
   - σ_{st} from LR analysis, which varies with depth in a fashion similar to (14) and ranges between 0.05 and 0.15 ksf (2.4 and 7.2 kPa)
   - σ = 0.15 ksf (7.2 kPa) for error in the Casagrande estimates of σ'_p (reflects min-max range in Fig. 15)
   - COV = 5% for error in converting from the one-day to end-of-primary curve.

The resulting $S_u(D)$ minus σ_{sy} profile (Line 5 in Fig. 14) gives F = (0.876)(1.10) = 0.964, σ_F = 0.133, and P_f = 24%, due to uncertainty in the best estimate of the SHANSEP strength profile.

Azzouz et al. (1983) indicate a COV[3D] of about 5% for the increase in factor of safety due to three-dimensional end-effects. This model error of σ_{3D} = (0.05)(1.097) = 0.055 increases the total standard deviation in F to $\sigma_F = (0.133^2 + 0.055^2)^{0.5}$ = 0.144, and P_f increases (slightly) to 26%. (Note: For more complex stability analyses, such as embankments with staged construction, the total model error may increase to $COV \cong$ 10–15%.)

The last SHANSEP analysis in Table 14 includes a rather subjective estimate of the spatial variations about the best estimate $S_u(D)$ line. This uncertainty was obtained by increasing $COV[S]$ from 10 to 14% and adding a spatial fluctuation in σ'_p of $\sigma_{sp}[\sigma'_p]$ = 0.125 ksf (6.0 kPa), which equals one-half of the total scatter in the data about the σ'_p LR line. The resulting P_f implies a 32% probability of having a "small" slope failure, as compared with 25% for a "large" failure.

The SHANSEP reliability analysis appears rather complicated because the discussers purposely included a rather comprehensive set of factors to illustrate different sources of errors and how they contribute to the overall uncertainty in the stability analyses. Nevertheless, after entering the soils data in a spreadsheet, the calculations of best estimates and uncertainties in S_u, plus the stability analyses for each profile, should not require more than a day or so.

CONCLUSIONS

The reliability analyses using the field vane data with Bjerrum's (1972) correction and using SHANSEP, both of which relied heavily on statistics, predict a 25–30% probability of a large failure, such as actually occurred, due to uncertainty in the undrained shear strength of the Bay Mud. In contrast, the author's approach predicts a P_f of less than 15% due to the uncertainty in S_u, which appears to be too low. This apparent underestimate based on the three-sigma rule may have occurred because the $\pm 3\sigma$ lines in Fig. 2(b) ignored the $S_u(FV)$ data above El. − 50 (values "believed to be erroneous") and focused on the UUC strengths. It is preferable to start with objective statistical analysis techniques to compute mean values and amount of scatter, then use judgment in selecting best estimates of parameters and their uncertainties. Sole reliance on the three-sigma rule should be discouraged.

If one has only UUC strength data, the discussers believe that estimates of uncertainty are extremely difficult, if not impossible, unless one has extensive prior experience with the same soil. This problem occurs because estimates of the S_u from UUC strengths appropriate for stability analyses depend on a totally fortuitous cancellation of three major errors: the increased S_u due to rapid shearing (strain rate effects) and due to failure in the vertical direction (ignores anisotropy) must be offset by an S_u reduction due to significant sample disturbance. The first discusser's experience shows that UUC strengths can range from being several times too high [e.g., Table 7 of Germaine and Ladd (1988)] to being well less than one-half of an appropriate design S_u.

REFERENCES

Azzouz, A. S., Baligh, M. M., and Ladd, C. C. (1983). "Corrected field vane strength for embankment design." *J. Geotech. Engrg.*, ASCE, 109(5), 730–734.

Baecher, G. B., and Ladd, C. C. (1997). "Formal observational approach to staged loading." *Transp. Res. Rec. No. 1582*, Transportation Research Board, Washington, D.C., 49–52.

Bjerrum, L. (1972). "Embankments on soft ground: SOA Report." *Proc., Specialty Conf. on Performance of Earth and Earth-Supported Struct.*, ASCE, New York, 2, 1–54.

Chandler, R. J. (1988). "The in-situ measurement of the undrained shear strength of clays using the field vane: SOA paper." *Vane shear strength testing in soils, field and laboratory studies, STP 1014*, ASTM, West Conshohocken, Pa., 13–44.

DeGroot, D. J., Ladd, C. C., and Germaine, J. T. (1992). "Direct simple shear testing of cohesive soils." *Res. Rep. R92-18*, Dept. of Civ. Engrg., MIT, Cambridge, Mass.

Germaine, J. T., and Ladd, C. C. (1988). "Triaxial testing of saturated cohesive soils: SOA paper." *Advanced triaxial testing of soil and rock, STP 977*, ASTM, West Conshohocken, Pa., 421–459.
Ladd, C. C. (1991). "Stability evaluation during staged construction." *J. Geotech. Engrg.*, ASCE, 117(4), 540–615.
Wright, S. G. (1991). *UTEXAS3: a computer program for slope stability calculations*, Shinoak Software, Austin, Tex.

Discussion by K. S. Li,[11] Member, ASCE, and Joley Lam[12]

Use of probabilistic methods in geotechnical engineering has been advocated by many researchers since the 1960s, most notably by Lumb (1966, 1967, 1968). It is gratifying to see the author, a long-time prominent figure in the determinist's camp, publish a paper to promote the wider use of probabilistic methods in geotechnical engineering.

Probabilistic methods are often regarded as being quite mathematical and difficult to learn by determinists who are used to the simple concept of factor of safety. Despite this, probabilists seem to enjoy developing increasingly complex mathematical techniques of probabilistic analyses. Consequently, geotechnical engineers would find it difficult to come to grips with difficult probabilistic concepts and jargons such as normal-tail approximation, Rosenblatt's transformation, and zero-upcrossing, to name a few. The end-result is obvious—geotechnical practitioners are disinterested in using probabilistic methods.

The author should be commended for presenting a methodology for quick assessment of the probability failure of geotechnical structures, which is simple to understand and be used by practicing geotechnical engineers, although it may be regarded by some pure probabilists as being not very rigorous mathematically. As discussed by Li and Lumb (1987) and Nadim and Lacasse (1999), the factor of safety is not a good measure of risk of failure. A geotechnical structure with a higher factor of safety can have a higher risk of failure than a similar structure with a lower factor of safety, depending on the accuracy of the model used for analysis and the uncertainties of the input parameters. The discussers consider that a crude probabilistic analysis would be far more fruitful in giving the designer an idea of the uncertainties and risks associated with a geotechnical design than a very precise calculation of the factor of safety. The three-sigma rule proposed by the author is thus regarded as being good enough for application in geotechnical engineering for most practical cases.

The failure probability given by the three-sigma rule depends significantly on the values of sigma to be used in the analysis. A soil property is regarded as a spatially random variability, as the actual values (i.e., realizations) of the soil properties can vary from one point to another within an apparently homogeneous soil. The statistical properties of a soil property at any particular point [which is called the point property by Li and Lumb (1987)] can be characterized by constant mean value μ and standard deviation σ.

The sigma of a soil property is often taken to be standard deviation of the point property, as illustrated in Fig. 2 of the author's paper. As discussed by Li and Lumb (1987), Li (1991, 1992), and Li and Lee (1991), the appropriate value of sigma to be used for probabilistic analyses may not necessarily be the standard deviation of the point property. The discussers would like to take this opportunity to discuss some concepts of soil variability using a simple language. It is hoped that the discussion could give some guidance to practicing engineers in choosing suitable values of sigma when using the three-sigma rule for probabilistic analyses. To enable a better understanding of the concepts, we will consider the simple case of a homogeneous slope.

CONCEPT 1: POINT PROPERTY VERSUS SPATIAL AVERAGE

For a nonbrittle soil, yielding or failure of soil at a single point along a potential failure slip surface will not lead to an ultimate failure. Failure will occur when the sum of the resistances of all individual soil elements along the entire slip surface is less than disturbing force. Therefore, the probability of failure tends to be controlled by the variability of the average resistance along the slip surface (i.e., spatially averaged property) rather than the variability of the soil resistance at any particular point (i.e., the point property). The variability of the spatially averaged property is usually significantly less than the point property as a result of variance reduction due to spatial averaging (Varmarcke 1977). We can illustrate this using a simple example. For sake of discussion, let us assume that there are only two soil elements along the slip surface. If we denote the soil resistance as X, using a deterministic approach, the total resistance X_T can be expressed as $2X$ in a deterministic approach. A mistake is commonly made in assuming that the statistical properties of X are the same as those of the point property. If such a mistake is made, the standard deviation of the total resistance X_T will become 2σ.

In reality, the soil resistances along the slip surface will vary from point to point. The two soil elements may have different values of soil resistance, and they should be treated as two separate random variables, to be denoted as X_1 and X_2. Assuming a uniform soil in a statistical sense, the standard deviation of X_1 and X_2 can each be taken to be equal to σ. The total resistance X_T along the slip surface should be given by $X_T = X_1 + X_2 = 2X$, where X is the spatially averaged soil resistance along the slip surface. If the soil resistances of the two soil elements are independent of each other, the sigma value for X is equal to $\sigma/\sqrt{2}$, as is discussed in many elementary textbooks in probability. Therefore, the sigma value of X_T becomes $\sqrt{2}\sigma$ instead of 2σ. In general, the standard deviation of a spatially averaged soil property can be written as $\sigma\Gamma(L)$, where $\Gamma(L)$ is a reduction factor that depends on the dimension of the domain of the spatial average. For most geotechnical structures, the variance reduction is very significant and the total uncertainty will be dominated by other factors, as will be discussed below.

If the sigma value of X_T is incorrectly taken to be equal to the larger standard deviation of the point property σ, the failure probability of the slope will be overestimated. The error can be very significant, particularly when the dimension of the slip surface is large relative to the scale of fluctuation of the soil property (Li and Lumb 1987).

CONCEPT 2: SAMPLE UNCERTAINTY VERSUS INNATE VARIABILITY

In the preceding section, it is assumed that the true mean value of the soil property is known. In practice, the true mean value of a soil property is never known, and it can only be estimated by, for example, the sample mean value, m, of the soil property determined from soil measurements. The variance of a spatially averaged soil property estimated by the sample mean value is given (Li and Lumb 1987; Li 1989) by

$$\mathrm{var}\{\bar{X}\} = \mathrm{var}\{m\} + \sigma^2\Gamma^2(L) \tag{18}$$

The first term on the right-hand side of (18) is associated with the sampling uncertainty, while the second term is associated

[11]Dir., Victor Li and Assoc., Ltd., and Assoc. Prof., Dept. of Civ. Engrg., Univ. of Hong Kong, Hong Kong, China.
[12]Engr., Victor Li and Assoc., Ltd., Hong Kong, China.

with the spatial variability of the soil property. Taking a simple example, if the average soil strength is estimated from the sample mean value calculated from a total of n independent measurements, the sampling variance var$\{m\}$ becomes σ^2/n. The standard deviation of the sample mean is then given by $\sigma/\sqrt{n}$.

The first term of (18) would not be affected by the number of soil measurements, but not the dimension of the domain of spatial average. The concept of variance reduction due to spatial averaging, therefore, does not apply for sampling uncertainty. For a large spatial domain, the variance reduction due to spatial averaging is significant and the total uncertainty of the spatially averaged soil property will then be dominated by the sampling uncertainty [i.e., the first term of (18)].

The sample uncertainty can be reduced by more soil measurements. However, when the sampling uncertainty [i.e., the first term of (18)] becomes comparable with the innate variability of the soil property [i.e., the second term of (19)], it will not be cost effective to further reduce the sampling uncertainty by taking more measurements. In most geotechnical designs, soil data is usually limited and the effect of spatial averaging is significant, with the result that the total uncertainty is normally dominated by the sampling uncertainty, i.e., the first term of (18). The sigma value to be used for the three-sigma rule is therefore simply the standard deviation of the sample mean.

CONCEPT 3: JUDGMENTAL VALUES

Sometimes, very limited or virtually no data exists for a particular random soil parameter. Certain judgmental values may then be assigned by the designers for the mean value and standard deviation of the random soil parameter to enable a probabilistic analysis to be performed. The uncertainty associated with the judgmental value is similar in concept to the sampling uncertainty associated with the estimation of the mean value of the soil property. There is no variance reduction due to spatial averaging for judgmental design values. One can therefore treat the uncertainties for judgmental values in a similar way as the sampling uncertainty discussed in (18). The sigma value to be used for the three-sigma rule is then simply the judgmental standard deviation of the soil property.

CONCEPT 4: MODEL ERROR

Measurements of soil properties are often subjected to bias. The corrected soil property X' is often expressed in terms of the measured soil property X using the simple relationship of $X' = NX$, where N is the correction factor. The same correction factor can be applied to obtain the "corrected" spatially averaged soil property. If the contribution from spatial variability can be ignored, the variance of corrected spatial average X'' is given as follows (Li 1989):

$$\mathrm{var}\{X''\} = N^2\,\mathrm{var}\{m\} + m^2\,\mathrm{var}\{N\} \qquad (19)$$

The expression in (19) covers both the sampling uncertainty and the uncertainty associated with the determination of the correction factor. The sigma value to be used in the three-sigma rule is then the square root of the expression on the right-hand side of (19).

REFERENCES

Li, K. S. (1989). "Discussion of 'Stability of gravity platform: reliability analysis.' " *Géotechnique*, London, 39(4), 561–562.

Li, K. S. (1991). "Discussion of 'Embankment reliability versus factor of safety: before and after slide repair.' " *Int. J. for Numer. and Analytical Methods in Geomech.*, 15(12), 893–895.

Li, K. S. (1992). "Some common mistakes in probabilistic analysis of slopes." *Proc., 6th Int. Symp. on Landslides*, Christchurch, New Zealand, 1, 475–480.

Li, K. S., and Lee, I. K. (1991). "The assessment of geotechnical safety." *Selected topics in geotechnical engineering—Lumb volume*, K. S. Li, ed., University of New South Wales, Kensington, Australia, 195–229.

Li, K. S., and Lumb, P. (1987). "Probabilistic design of slopes." *Can. Geotech. J.*, Ottawa, 24(4), 520–535.

Lumb, P. (1966). "The variability of natural soils." *Can. Geotech. J.*, Ottawa, 3, 74–97.

Lumb, P. (1967). "Statistical methods in soil investigations." *Proc., 5th Australian–New Zealand Conf. on Soil Mech. and Found. Engrg.*, 26–33.

Lumb, P. (1968). "Statistical aspects of soil measurements." *Proc., 4th Australian Road Res. Conf.*, 4, 1761–1770.

Nadim, F., and Lacasse, S. (1999). "Probabilistic slope stability evaluation." *Geotech. Risk Mgmt., Proc., Ann. Seminar of Geotech. Div.*, Hong Kong Institution of Engineers, Hong Kong, 179–186.

Vanmarcke, E. H. (1977). "Probabilistic modeling of soil profile." *J. Geotech. Engrg.*, ASCE, 103(11), 1227–1246.

Discussion by Yoshi Moriwaki[13] and John A. Barneich,[14] Members, ASCE

The author has presented a simplified approach to facilitate the application of reliability analysis for a variety of geotechnical problems. The discussers believe that the geotechnical engineering community needs to start embracing such a reliability-based approach, which explicitly recognizes the uncertainties so inherent in geotechnical engineering. Because the author's paper is easy to follow, the discussers sincerely hope that it would facilitate this process.

The author's approach was recently applied to a slope stability problem for which the discussers were commissioned to perform a third-party review. Through this case history, the discussers would like to provide some discussions on issues involved in applying the author's simplified approach in practice, restricting the comments on slope stability problems involving back-calculation of bedding plane materials.

A slope in Southern California had been cut and buttressed adjacent to a roadway with the resulting critical factor of safety of about 1.3. The geotechnical engineer of record had argued that this factor of safety was adequate, rather than the generally required factor of safety of 1.5. The reasons given were that the slope had been thoroughly investigated and tested and that a small landslide had occurred during construction from which the effective friction angle of the bedding plane materials, considered to be the most critical material in the slope, was back-calculated.

In situations like this, the discussers believe that application of the reliability-based approach is very useful. The slope stability analysis can be performed using any method that satisfies all the equilibrium conditions; the Spencer's method was used here. Beyond that, the discussers found it vital to think about each parameter affecting the slope stability as applied to the particular case before starting the analysis. One convenient way is to think in terms of epistemic and aleatory uncertainties (as used in the probabilistic seismic hazard analysis). The former uncertainty could, in theory, be reduced by more studies; shear strength may belong to this category. The latter is random uncertainty, not amenable to reduction through additional studies; future pore water pressure conditions may belong more to this category. It is noted that one of the benefits of reliability-based slope stability analysis is the ability to quantify the contribution of various parameters to the computed factor of safety, allowing for rational choice on the most productive aspects of investigation to focus. The discussers found discussions on the uncertainty considerations associated

[13]Prin., GeoPentech, Santa Ana, CA 92705.

[14]Prin., GeoPentech, Santa Ana, CA 92705.

with parameters presented by Christian et al. (1994) to be very useful in thinking about these issues.

As part of this third-party review, the discussers first repeated the back-calculation using the Spencer's method to confirm the general reasonableness of the back-calculated effective friction angle of 17°. In Southern California, the shear strength associated with many landslides involving bedding plane and presheared materials can often be represented by an effective friction angle without an artificial effective cohesion intercept. Though not used in this case, the discussers believe that back-calculation should in general result in a distribution of back-calculated friction angle. Because adequate details of actual landslide are rarely known, one should apply the author's approach in back-calculation as well: forcing the back-calculated factor of safety of one with the best estimate and plus and minus one standard deviation of parameters that are not well known at the time of landslide to develop this distribution. The resulting range of back-calculated shear strength can be used with presumably greater, future ranges of other parameters to perform forward reliability-based slope stability analysis using the author's approach.

One caution regarding the use of the author's approach to slope stability problems is the lack of spatial variability considerations in the author's approach. This may result in overestimation of failure probability. Although discussions presented in Christian et al. (1994) should be helpful in this regard, one can bound the potential problem by assuming the two extremes of perfect correlation and no correlation of properties over space without getting into analysis involving autocorrelation distance.

Using the author's approach together with the ranges in material properties and geologic conditions, the discussers performed parametric slope stability analyses of the slopes at several sections along the 1,000 ft long slope to compute a best estimate overall static factor of safety of about 1.3 and a 90–92% reliability that the actual factor of safety would be greater than 1. The results were not very dependent on the assumed distributions (lognormal or normal relations), as expected for the range of the reliability values involved. Further, it was found that, if the computed factor of safety had been 1.5 (for the same ranges of material parameters and geologic conditions), this reliability would have increased to 96–98%. If, however, less were known about the geologic conditions and material properties, the reliability likely could have been in the 90–92% range, even with a computed best-estimate factor of safety of 1.5. Thus, in essence, the author's approach was used to quantify the benefit of detailed investigation and back-calculation as applied to this particular program. The computed factor of safety of 1.3 in this case was equivalent to the computed factor of safety of 1.5 without such additional information.

The above evaluations were computed assuming that the water table stayed below the base of the critical slide surface, which was considered reasonable due to restriction to development upslope of the critical slide surface (no significant watering) and drainage provisions incorporated into the slope. However, because so many landslides in Southern California are caused by unexpectedly adverse water conditions and because the future water conditions tend to be "random," an additional analysis was performed using extremely high groundwater conditions as the best estimate; it was found that the computed factor of safety would be slightly greater than 1.0, reducing the reliability of the actual factor of safety to be greater than 1.0 to 50–60%. Such conditions, though rather unlikely, would not be acceptable in general. However, if the likelihood of such conditions from ever developing were reduced further by a water monitoring/observation system, the acceptance would be easier. Such a monitoring/observation system was suggested.

Based on the results of this third-party review and consideration that the consequence of the failure of slope would not be life threatening, the owner decided to accept the slope with the provision that the design build contractor provide a 10-year insurance policy covering stability of the slope.

In closing, the discussers believe that the author has done the profession a great service. This easy-to-use procedure allows the practicing geotechnical engineers and engineering geologists to quickly put factor of safety estimates in perspective with respect to uncertainty in analysis parameters. At the same time, the results would provide a means of communicating this perspective to client and possibly regulatory agencies, so that informed discussion can be started. Though, as applied to slope stability problems, a more complete reliability approach such as the one by Christian et al. (1994) may be preferable for some problems, it should be easier to apply the author's approach. With the emphasis in practice being more on the relative reliability (as was the case in the example use of the author's approach presented herein), the lower-bound estimate of the probability of failure provided by these simplified procedures should not be critical. For significant projects, these limitations can be overcome by an even more formal approach (e.g., Whitman 1984; Barneich et al. 1996) using fault trees, etc. In whatever ways possible, the discussers would like to see the geotechnical engineering profession move more in the direction of reliability-based approach, at least for certain types of problems.

REFERENCES

Barneich, J., Majors, D., Moriwaki, Y., Kulkarni, R., and Davidson, R. (1996). "The reliability analysis of a major dam project." *Proc., Uncertainty '96*, ASCE, Reston, Va., 1367–1382.

Christian, J. T., Ladd, C. C., and Baecher, G. B. (1994). "Reliability applied to slope stability analysis." *J. Geotech. Engrg.*, ASCE, 120(12), 2180–2207.

Whitman, R. V. (1984). "Evaluating calculated risk in geotechnical engineering." *J. Geotech. Engrg.*, ASCE, 110(2), 145–188.

Discussion by John H. Schmertmann,[15] Fellow, ASCE

The author has made an excellent and welcome contribution towards simplifying and thus encouraging the use of the interrelationship between factor of safety (F) and reliability (R), which depends on the variability of F. Tables 2 and 7 should prove very useful for evaluating low-end-tail and high-end-tail reliability, respectively, for the lognormal distribution of F considered reasonable and used by the author.

At the expense of added complexity, which hopefully will not discourage use of the methods described in this paper, some warning seems appropriate to help users of the proposed methods avoid, or at least recognize, some items that can produce unconservative results (R lower than expected for a given F).

The potentially unconservative items briefly discussed herein include the distribution assumed for the variability in F, unconservative bias, and three items that can increase V_F—non-independent parameters, soil variability, and model uncertainty. The following uses the author's $\sigma_F = 0.25$, $F = 1.50$, $R = 99\%$ example [(1), Fig. 1, and Table 1] for a convenient demonstration of the effect of these items. Although the author's example uses the low-end tail of a lognormal distribution, the items discussed remain appropriate in principle for

[15]Schmertmann & Crapps, Inc., 4509 W 23rd Ave., Ste. 19, Gainesville, FL 32606.

both tails of any distribution. The following considers the effect of each item separately.

DISTRIBUTION OF *F*

The F required for a given R depends on the shape of the tail of the distribution used. For example, using the normal versus the lognormal increases the required F to 1.58 for R = 99%. The increase becomes more dramatic as V_F increases. Doubtless, nature provides many distributions that our idealized ones can only approximate. Perhaps we can include the approximation as part of the model uncertainty noted subsequently?

BIAS

Users should also consider the possibility of unconservative bias in one or more of the parameters used to calculate F. For example, if the test for tan δ in (1) produced values consistently 10% too high because of undetected or uncorrected apparatus friction, then the required F increases to 1.65.

NON-INDEPENDENCE OF PARAMETERS

The use of the Taylor series to obtain (2*a*) assumes that each of the parameters in Table 1 varies independently of the others (often questionable; for example, γ_{ef} and γ_{bf} may correlate with each other). Eq. (2*a*) provides one approximate value for σ_F. The upper bound for σ_F results form adding the absolute values of ($\Delta F/2$) in Table 1—in this case, giving $\sigma_F = 0.405$. This would require $F = 1.75$ for $R = 99\%$. How best to combine the parameter standard deviations depends on the judgment of the engineer. Perhaps the author can comment and provide some guidance.

SOIL VARIABILITY

The author presents Table 3 as a guide for choosing V for various geotechnical parameters, as obtained from the literature and his experience. If the author knows, it would help to know which Vs in Table 3 include soil (material) as well as test variability. Would the author suggest that the low V values in each range apply primarily to sites judged to have relatively low soil variability and the high V values to those with high soil variability? When previously considering such questions in Schmertmann (1989, see Tables 2 and 3), this discusser applied an approximate 1.5 factor to V_{tests} to estimate $V_{(\text{tests+soil})}$ for all relative site soil variabilities (low, average, high). Assuming Table 1 includes only test variability, applying a 1.5 factor to the standard deviation in Table 1 gives a new σ_F = 0.375, which gives a new required $F = 1.70$ for $R = 99\%$.

MODEL UNCERTAINTY

Every mathematical model, including (1), presents a simplified picture of reality. $[F_{\text{actual}} - F_{(1)}]$ varies from site to site and, thus, the model of reality itself includes uncertainty. Denote the standard deviation of this uncertainty as σ_M. One determines σ_M from comparisons of F_a and F_1, with F_a generally only known when equal to 1 at some defined failure. Or, the engineer can estimate σ_M from the literature and personal experience. Perhaps Table 1 already includes model uncertainty? Perhaps the author can suggest typical values for σ_M or V_M? For the purpose of this discussion, assume no bias and σ_M = 0.20 when $F = 1$, or $V_M = 20\%$. Further assuming the model uncertainty independent of the Table 1 parameter uncertainties, and using the Taylor series approximate value for a combined V_F, gives $V_F = 26\%$ and $F = 1.88$ for $R = 99\%$.

COMBINATIONS

Each of the above items produces a relatively small, but possibly important, change in the F required to maintain R = 99%. But they can occur in combination. If they all occurred together, namely [normal distribution and 10% bias and (a non-independent upper bound σ_F and a 1.5 soil variability factor) and (an independent V_m = 20%)], then the required F increases to 3.28 for $R = 99\%$. Using the lognormal distribution reduces the required F to 2.48. Also, using the assumption of parameter independence further reduces the required F to 2.15. As shown, the items presented can increase F substantially from the author's 1.50 and may require consideration in his and other applications of the proposed methods.

REFERENCE

Schmertmann, J. H. (1989). "Density tests above zero air voids line." *J. Geotech. Engrg.*, ASCE, 115(7), 1003–1018.

Closure by J. M. Duncan[16]

INTRODUCTION

The writer is very pleased that the paper has drawn such extensive discussion, reflecting a high degree of interest in the topic, and is grateful to all the discussers for their valuable comments. Their views of the benefits of reliability analysis in geotechnical engineering practice are enlightening and welcome. Their discussions of the necessary prerequisites for deterministic and probabilistic analyses will be of lasting value. Finally, their discussions of various methods for calculating probability of failure and unsatisfactory performance provide a more comprehensive view of these techniques than was contained in the paper. The writer appreciates each of these contributions and extend his sincere thanks to all of the discussers.

BENEFITS OF RELIABILITY ANALYSES

The discussions describe a number of benefits of reliability analyses in geotechnical engineering applications. These are summarized briefly in Table 15.

APPROPRIATE USE OF RELIABILITY METHODS

The discussions offer useful insights into appropriate (and inappropriate) use of reliability analyses with regard to the need for good data, choices of suitable test methods and interpretation techniques, and the need for sound judgment in many aspects of geotechnical analyses.

Need for Good Data

Koutsoftas expressed the concern that some engineers may be tempted to use reliability analyses as a substitute for thorough investigations with high quality data. The writer hopes that this will not be the case. It would be inappropriate to use reliability analyses to justify inferior investigations or substandard data. Quite the contrary, the type of reliability analyses described in the paper are most appropriately used to evaluate the combined effects of uncertainties in analyses, to identify those aspects of analyses where uncertainties are most significant, where further investigation would be most valuable, and where further investigation would or would not significantly reduce the degree of uncertainty involved in the final result.

Test Methods and Interpretation

Although there is undoubtedly uniform agreement that good data is needed for either deterministic or probabilistic analyses, preferences for different test methods to obtain "good data"

[16]Univ. Distinguished Prof., Dept. of Civ. and Envir. Engrg., Virginia Tech., Blacksburg, VA 24061.

TABLE 15. Benefits of Reliability Analyses

Discussers	Noted benefits of reliability analysis for geotechnical engineering
Christian and Baecher	• Provides framework for establishing appropriate factors of safety and leads to better understanding of relative importance of uncertainties.
Ladd and Da Re	• Provides systematic method for evaluating combined influence of uncertainties in parameters affecting factor of safety. • Provides systemic method of assessing degree of safety, at least in relative terms.
Moriwaki and Barneich	• Quantifies contributions to overall uncertainty of each parameter. • Provide means of determining consequences of uncertainty, benefit of detailed investigation, and benefits of back analysis. • Puts computed values of factor of safety into perspective with regard to uncertainties in analysis parameters. • Provide means of communicating reliability of factor of safety to clients and perhaps regulators, so that informed discussion can begin.
Koutsoftas	• Provides useful tool for evaluating risk associated with design recommendations.
Focht and Focht	• Provides tools for communicating relative risk and benefit to client, and provides client with clearer choices regarding economic tradeoff between initial construction costs and risk of failure. • Method described in paper integrates easily with conventional approaches. • When used with judgment, method described in paper will improve value of geotechnical analyses.
Li and Lam	• Method described in paper is easy to understand and apply, "although it may be regarded by some . . . as being not very rigorous mathematically." • Even crude probabilistic analysis is more useful for developing idea of uncertainties and risks than very precise deterministic calculation of factor of safety.

vary widely, based on an individual's experience. As noted in the next section, some engineers eschew SPT data and prefer the DMT, some reject UU triaxial tests and favor field vane shear, and some reject vane shear and prefer UU triaxial tests, based on their individual experiences.

It is well to remember that all of the methods of collecting and interpreting data that are used in geotechnical engineering analyses are semiempirical, and that their use is justified by favorable experience. It is doubtful that any type of geotechnical investigation and analysis can be found where approximations and assumptions are not required to link the results of field investigations to conclusions regarding likely performance. Experience with the use of particular tests and methods of interpretation is an important component of a geotechnical engineer's personal art and forms the necessary core for making the important judgments that are inevitably required.

Need for Judgment

As noted by Focht and Focht, the use of reliability techniques does not eliminate the need for sound engineering judgment. In fact, reliability methods cannot be applied without making judgments regarding both the techniques and the data on which the analyses are based. The necessity to use sound engineering judgment in applying reliability analysis pervades all of the discussions of the paper and is clearly of paramount importance.

The discussions include many specific examples of the use of judgments to interpret and supplement available test data. It is worthwhile to summarize a few of them here, to illustrate their importance and their pervasive nature:

- Ladd and Da Re employed engineering judgment in several aspects of their reanalyses of the LASH Terminal slope failure. They used SHANSEP parameter correlations, developed for other clays, which they judged to be applicable to San Francisco Bay Mud. They adjusted values of OCR based on previous experience. They increased their computed two-dimensional factor of safety as an allowance for end effects. And they used a subjective estimate of spatial variation in undrained strength to supplement the available data. It is always necessary to use judgments and adjustments such as these to fill gaps in data and to improve the reasonableness of parameter values for geotechnical analyses.
- Christian and Baecher and Li and Lam noted that one must rely on approximations to estimate values of needed parameters when insufficient data are available. Examples in this category would include parameters such as friction angles for sands and gravels (no undisturbed samples, and therefore no laboratory test data), and equivalent fluid unit weights for use in wall design (no data because values are based entirely on experience and judgment). Use of judgments, approximations, and correlations is essential in many circumstances and should not be construed as a sign of poor engineering practice.
- Focht and Focht noted that, for linear projects that cross a number of geologic units or spatially diverse conditions, judgment is necessary to choose conditions that are deemed to be representative for each segment of the project, based on an understanding of the geologic conditions and information from widely spaded borings.
- Koutsoftas used correlations between vane shear strength and preconsolidation pressure developed from other sites in the Bay Area to estimate preconsolidation pressures at Hamilton Air Force Base. Although it has been determined that the site at Hamilton described by Koutsoftas is not the one where the data shown in Fig. 2(a) of the paper were obtained (Koutsoftas, personal communication, 2001), the use of vane shear correlations as a tool for evaluating stress history is nevertheless of interest and illustrates the use of judgment and correlations to supplement site-specific data.
- Failmezger uses judgment and experience to discount the use of SPT data for estimating settlement of foundations on sand.
- Ladd and Da Re and Koutsoftas use judgment and experience to rule out the use of UU triaxial compression tests to evaluate undrained shear strengths of clays.
- Focht and Focht use judgment and experience to reject the use of field vane shear tests to evaluate undrained shear strengths of clays.
- Ladd and Da Re used judgment and experience to conclude that three CK_0UDSS tests run by MIT on normally consolidated Bay Mud from another site in the Bay Area provided a sufficient basis for reanalysis of stability of the LASH Terminal slope.
- Moriwaki and Barneich used judgment and experience as the basis for including assumed extremely high groundwater as a condition in their analysis of the stability of a slope in southern California.
- Koutsoftas used the results of his reanalysis of the LASH Terminal slope to conclude that the undrained factor of safety would have been significantly higher than the value calculated in the paper, and therefore the strength of the mud must have been reduced during excavation of the trench, most likely by drainage at the base of the deposit.
- Schmertmann used his experience with reliability analyses to identify several items that could lead to higher estimates of the probability of failure for the retaining wall example in the paper.

Thus, although reliability analysis can provide a valuable supplement to conventional deterministic analyses of stability and settlement, neither deterministic nor probabilistic analyses can be applied without experience and judgment.

METHODS AND ASSUMPTIONS FOR CALCULATING P_f

Many of the discussions focused on reliability analysis techniques and suggested enhancements and alternatives to the method described in the paper. These comments concern sources of uncertainty, the three-sigma rule, model uncertainty, and the lognormal distribution.

Sources of Uncertainty, Types of Uncertainty, Effects on Computed P_f

Evaluation of geotechnical parameters is fraught with many types of uncertainties. Given a collection of test data, geotechnical engineers are faced with a number of important questions. Among them are

- Does the scatter in the data represent variations in the properties of the ground, variations due to random test errors, or both?
- Should the data be corrected for systematic effects before they can be used in analyses to estimate performance?
- Would the uncertainties in the value of a needed parameter be significantly reduced if more tests were performed?
- Is the data possibly deficient because the sampling or test locations missed some important feature of the site—some important geologic detail?
- Is it appropriate to select the "most likely value" of a needed parameter as the average measured value, or should the "most likely value" be higher or lower than the average?
- Do the results of the analysis to be performed depend on quantities not reflected in the data, such as possible future changes in groundwater levels?

Ladd and Da Re noted that it is important to consider the sources of uncertainty involved, or else the results of reliability analyses will be misleading. They suggest that uncertainties be considered in the following categories, which provide a useful framework for evaluation and judgment:

1. Systematic variations of properties from one location to another
2. Scatter due to spatial variations about the mean trend, and the scale over which this fluctuation occurs
3. Scatter due to random test errors
4. Statistical uncertainty due to limited numbers of tests
5. Error in the mean due to measurement bias

While it is useful to realize that these types of uncertainty all affect test data, it is important to understand that these affects can only be identified, separated, and evaluated through the exercise of judgment.

One issue that emerges clearly from the discussions by Ladd and Da Re and by Li and Lam is that the variability of the average value of a property for a large mass of soil is significantly less than the variability of the "point values" measured in a series of tests. Li and Lam pointed out that, all other things being equal, using the standard deviation of the point value as the standard deviation of the spatial average will result in a value of probability of failure that is too large.

Focht and Focht pointed out that an additional uncertainty should be considered when selecting an appropriate "most likely" spatial average value of a property such as shear strength, because the failure mechanism will find and tend to pass through the weakest material.

Moriwaki and Barneich prefer to group uncertainties in two categories, following the convention in probabilistic seismic hazard analysis:

1. Epistemic uncertainties are those that are reducible through better or more extensive testing. Hypothetically, at least, uncertainties about properties such as the residual friction angle of a clay deposit would fall in this category. There are, however, practical limits on the extent of investigation, with the result that epistemic uncertainties cannot be completely eliminated and must be dealt with through use of theory and judgmental evaluation.
2. Aleatory uncertainties are those that cannot be evaluated through investigation, no matter how thorough. An example in this category would be the highest possible future water level within a slope.

Moriwaki and Barneich suggested that the emphasis in practice is on values of relative uncertainty, rather than absolute uncertainty. Useful results for such comparative analyses can be achieved using approximate values of properties and their standard deviations and using simplified analyses, provided they are used consistently.

Coefficients of Variation and Three-Sigma Rule

Schmertmann raised a number of questions about the coefficients of variation listed in Table 3: Which of the values include both soil and test variability? Is it correct that the smaller values pertain to sites with less variability and the larger values to sites with more variability? The values of coefficient of consolidation from the writer's own files represent statistics for C_v evaluated using Casagrande's method and Taylor's method for several different clays and therefore represent both soil and test variability. Likewise, where values are attributed to more than a single source, they would reflect more than one soil, and test variability would also inevitably be involved.

Ladd and Da Re asked if the coefficient of variation of S_u/p should depend on OCR, and if the coefficient of variation of S_u should depend on the type of test and the heterogeneity of the deposit? The writer believes the answer to both questions is yes, but has no data to support that judgment.

As noted in the paper, the values of coefficient of variation listed in Table 3 cover extremely wide ranges of values for the same parameter, and the conditions of sampling and testing are not specified. The values in Table 3, therefore, provide only a rough guide for any given case. It is important to use judgment in applying coefficients of variation from published sources.

Christian and Baecher provided very useful information relating to the use of the three-sigma rule, showing that people (experienced engineers included) tend to be overconfident about their ability to estimate values, and estimate possible ranges of values that are narrower than the actual range. If the range between the highest conceivable value (HCV) and the lowest conceivable value (LCV) is too small, values of coefficient of variation estimated using the three-sigma rule will also be too small, introducing an unconservative bias in reliability analysis.

Their Table 9 is very interesting in this regard. Based on statistics, it shows that the expected range of values in a sample of 20 values is 3.7 times the standard deviation, and the expected range of values in a sample of 30 to 4.1 times the standard deviation. This information could be used to improve estimated values of standard deviation by modifying the three-sigma rule. If the experience of the person making the estimate encompasses examination of sample sizes in the range of 20–30 values, a better estimate of standard deviation would be made by dividing the range between HCV and LCV by 4 rather than 6:

$$\sigma = \frac{\mathrm{HCV} - \mathrm{LCV}}{4} \tag{20}$$

Based on Christian and Baecher's discussion, and the idea that experience will often be limited to examination of samples including no more than 30 values, this "two-sigma rule" would be an improvement on the three-sigma rule discussed in the paper. It could be extended easily to a "graphical two-sigma rule," placing the average minus sigma and average plus sigma lines halfway between the most likely variation and the extremes.

Model Uncertainty

Model uncertainty was included in only one of the four examples in the paper—the one involving estimation of the settlement of a footing on sand based on SPT blow count. The data shown in Fig. 6 were used to estimate the coefficient of variation = 67% associated with the use of (8), which in this case is the "model." The uncertainty is due to scatter in SPT results and approximations made in deriving (8).

Model uncertainty was not included in the other examples in the paper, but it probably should have been. Li and Lam, Failmezger, and Schmertmann all mentioned model uncertainty (or model error) in their discussions and suggested ways of including it in analyses. Schmertmann asked if the writer could suggest coefficients of variation for model uncertainty for the retaining wall example, and if the assumed lognormal distribution of factor of safety would be considered to be an aspect of model uncertainty. In the writer's opinion, if movement of the retaining wall is not resisted by passive pressure on the front of the footing, the equation of horizontal equilibrium on which (1) is based provides a highly reliable model, which would need no additional allowance for uncertainty beyond the coefficients of variation assigned to the earth pressure, the base friction, and the unit weights of the concrete and the backfill. The effect of the assumed lognormal distribution, while deserving of consideration, is not part of the deterministic calculation, which the writer considers constitutes the model. The writer believes it would be preferable to examine the effect of the assumed distribution by computing probabilities of failure using more than one assumed distribution. The following paragraphs provide an easy means for doing this.

Assumed Distributions of *F*

Christian and Baecher discuss the characteristics of the lognormal distribution and indicate their preference for the normal distribution. There is no way of determining which of these is best in a given case, as is clear from their discussion. It may often be useful, therefore, to be able to compute P_f using both the normal and the lognormal distributions.

Table 16 can be used to determine P_f based on an assumed normal distribution in the same way that Table 2 is used to determine P_f based on an assumed lognormal distribution. By comparing the tables, it can be seen that values of P_f based on a lognormal distribution, and values of P_f based on a normal distribution, are considerably different in some regions of the tables. In Table 16, the boldface values of P_f based on an assumed normal distribution are greater than those based on an assumed lognormal distribution. In the lightface values the reverse is true.

As examples of the effect on P_f involved in assuming that factor of safety is normally or lognormally distributed, consider the values from the paper and the discussions given in Table 17.

Failmezger suggested that, because the lognormal distribution is skewed to the left, estimates of the probability that the factor of safety could be less than 1.0 will tend to be conservative, and estimates of the probability that settlements could be larger than computed will tend to be unconservative, as compared with an assumed normal distribution. Table 16 shows that this is not the case. Assuming a lognormal distribution of factor of safety can be either less conservative or more conservative than assuming a normal distribution, de-

TABLE 16. Probabilities that Factor of Safety Is Smaller than 1.0, Based on Normal Distribution of Factor of Safety

	Coefficient of Variation of Factor of Safety														
F_{MLV}	2%	4%	6%	8%	10%	12%	14%	16%	20%	25%	30%	40%	50%	60%	80%
1.05	**0.9%**	**11.7%**	21.4%	27.6%	31.7%	34.6%	36.7%	38.3%	40.6%	42.4%	43.7%	45.3%	46.2%	46.8%	47.6%
1.10	**0.0%**	**1.2%**	**6.5%**	**12.8%**	18.2%	22.4%	25.8%	28.5%	32.5%	35.8%	38.1%	41.0%	42.8%	44.0%	45.5%
1.15	**0.0%**	**0.1%**	**1.5%**	**5.2%**	**9.6%**	**13.9%**	17.6%	20.7%	25.7%	30.1%	33.2%	37.2%	39.7%	41.4%	43.5%
1.16	**0.0%**	**0.0%**	**1.1%**	**4.2%**	**8.4%**	**12.5%**	**16.2%**	19.4%	24.5%	29.1%	32.3%	36.5%	39.1%	40.9%	43.2%
1.18	**0.0%**	**0.0%**	**0.6%**	**2.8%**	**6.4%**	**10.2%**	**13.8%**	**17.0%**	22.3%	27.1%	30.6%	35.1%	38.0%	40.0%	42.4%
1.20	**0.0%**	**0.0%**	**0.3%**	**1.9%**	**4.8%**	**8.2%**	**11.7%**	**14.9%**	20.2%	25.2%	28.9%	33.8%	36.9%	39.1%	41.7%
1.25	**0.0%**	**0.0%**	**0.0%**	**0.6%**	**2.3%**	**4.8%**	**7.7%**	**10.6%**	**15.9%**	21.2%	25.2%	30.9%	34.5%	36.9%	40.1%
1.30	**0.0%**	**0.0%**	**0.0%**	**0.2%**	**1.1%**	**2.7%**	**5.0%**	**7.5%**	**12.4%**	**17.8%**	22.1%	28.2%	32.2%	35.0%	38.6%
1.35	**0.0%**	**0.0%**	**0.0%**	**0.1%**	**0.5%**	**1.5%**	**3.2%**	**5.3%**	**9.7%**	**15.0%**	**19.4%**	25.8%	30.2%	33.3%	37.3%
1.40	**0.0%**	**0.0%**	**0.0%**	**0.0%**	**0.2%**	**0.9%**	**2.1%**	**3.7%**	**7.7%**	**12.7%**	**17.0%**	23.8%	28.4%	31.7%	36.0%
1.50	**0.0%**	**0.0%**	**0.0%**	**0.0%**	**0.0%**	**0.3%**	**0.9%**	**1.9%**	**4.8%**	**9.1%**	**13.3%**	**20.2%**	25.2%	28.9%	33.8%
1.60	**0.0%**	**0.0%**	**0.0%**	**0.0%**	**0.0%**	**0.1%**	**0.4%**	**1.0%**	**3.0%**	**6.7%**	**10.6%**	**17.4%**	**22.7%**	26.6%	32.0%
1.70	**0.0%**	**0.0%**	**0.0%**	**0.0%**	**0.0%**	**0.0%**	**0.2%**	**0.5%**	**2.0%**	**5.0%**	**8.5%**	**15.2%**	**20.5%**	24.6%	30.3%
1.80	**0.0%**	**0.0%**	**0.0%**	**0.0%**	**0.0%**	**0.0%**	**0.1%**	**0.3%**	**1.3%**	**3.8%**	**6.9%**	**13.3%**	**18.7%**	**22.9%**	28.9%
1.90	**0.0%**	**0.0%**	**0.0%**	**0.0%**	**0.0%**	**0.0%**	**0.0%**	**0.2%**	**0.9%**	**2.9%**	**5.7%**	**11.8%**	**17.2%**	**21.5%**	27.7%
2.00	**0.0%**	**0.0%**	**0.0%**	**0.0%**	**0.0%**	**0.0%**	**0.0%**	**0.1%**	**0.6%**	**2.3%**	**4.8%**	**10.6%**	**15.9%**	**20.2%**	**26.6%**
2.20	**0.0%**	**0.0%**	**0.0%**	**0.0%**	**0.0%**	**0.0%**	**0.0%**	**0.0%**	**0.3%**	**1.5%**	**3.5%**	**8.6%**	**13.8%**	**18.2%**	**24.8%**
2.40	**0.0%**	**0.0%**	**0.0%**	**0.0%**	**0.0%**	**0.0%**	**0.0%**	**0.0%**	**0.2%**	**1.0%**	**2.6%**	**7.2%**	**12.2%**	**16.5%**	**23.3%**
2.60	**0.0%**	**0.0%**	**0.0%**	**0.0%**	**0.0%**	**0.0%**	**0.0%**	**0.0%**	**0.1%**	**0.7%**	**2.0%**	**6.2%**	**10.9%**	**15.3%**	**22.1%**
2.80	**0.0%**	**0.0%**	**0.0%**	**0.0%**	**0.0%**	**0.0%**	**0.0%**	**0.0%**	**0.1%**	**0.5%**	**1.6%**	**5.4%**	**9.9%**	**14.2%**	**21.1%**
3.00	**0.0%**	**0.0%**	**0.0%**	**0.0%**	**0.0%**	**0.0%**	**0.0%**	**0.0%**	**0.0%**	**0.4%**	**1.3%**	**4.8%**	**9.1%**	**13.3%**	**20.2%**

Note: F_{MLV} = factor of safety computed using most likely values of parameters. Where table values are boldface, value of P_f computed assuming normal distribution is greater than value of P_f computed assuming lognormal distribution. Where table values are lightface, normal P_f is smaller than the lognormal P_f.

TABLE 17. Effect on P_f Involved in Assuming Factor of Safety is Normally or Lognormally Distributed

Case	Factor of safety, F	Coefficient of variation of F	P_f based on assumed lognormal distribution of F	P_f based on assumed normal distribution of F
Retaining wall sliding	1.50	17%	1.0%	1.9%
LASH slope	1.17	16%	18%	18%
Mine pillar problem (Christian and Baecher)	2.06	70%	20%	23%

pending on the most likely value of factor of safety and the coefficient of variation.

Failmezger also questioned values of P_f in Table 2 that exceed 50% for values of F_{MLV} greater than 1.0. The values in Table 2 have been checked and found to be correctly calculated using Excel. The writer believes that the values of P_f greater than 50% are due to the fact that the lognormal distribution is not symmetrical.

Failmezger also questioned the values in Table 7, which, for $SR = 1.10$, increase and then decrease as the coefficient of variation increases. These values were also checked and found to be correctly calculated using Excel. However, as Failmezger points out, it is not logical that the values should first increase and then decrease. The writer thinks this is probably due to a limitation on the numerical accuracy of the Excel algorithm. In any case, the variation is not likely to be significant from a practical point of view.

Accuracy of P_f

Christian and Baecher indicate a preference for computing probabilities of failure by assuming a normal distribution in the absence of other information, using a margin of safety instead of a factor of safety, or using the Hasofer-Lind (1974) approach to computing the reliability index without relying on the distribution of the margin or factor of safety, and showed an example to illustrate the benefits of their approach. These alternatives to the method proposed in the paper may well be desirable in cases where they are applicable. Table 16 provides a simple tool for using normal rather than lognormal distribution. However, many problems—for example slope stability—are not readily formulated in terms of margin of safety. The writer thinks factor of safety is preferable because of its widespread use in practice. Similarly, while the Hasofer-Lind method may have advantages, a simple explanation of it will be necessary before it can be applied widely in routine practice.

In view of the uncertainties involved in tests, test interpretation, deterministic models, and probabilistic analyses, values of geotechnical probability, no matter how computed, will seldom be highly accurate. This was well-stated by Dr. Ralph Peck at a recent Corps of Engineers workshop on risk analysis (Duncan and Smith 2000). In response to the question "What approach would you recommend to obtain the final results (i.e., probability of failure = 4.65×10^{-4})?" Dr. Peck replied, "I would disabuse myself of the idea of reporting a probability of failure to three significant figures, especially as the exponent might even be in question!"

CONCLUSION

Given the uncertainties inherent in geotechnical engineering, neither factors of safety nor probabilities of failure can be viewed as being highly accurate measures of safety. Circumstances are rare where factors of safety can be computed with an accuracy better than $\pm 15\%$, or probabilities of failure can be evaluated with an accuracy better than factor of two.

Although probabilities of failure or probabilities of actual settlement exceeding computed settlement can only be estimated approximately, they nevertheless provide valuable insight regarding the importance of the uncertainties that pervade geotechnical engineering analyses. As is clear from the example described by Moriwaki and Barneich, even approximate values of probability of failure are useful in assessing the consequences of various types of uncertainties. While probabilities of failure cannot be calculated without the exercise of judgment, the exercise itself is useful and sheds additional light on the reliability of the analysis process.

As stated in the paper, probability of failure should be viewed as a complement to factor of safety, not a substitute for it. Knowing both an approximate value of factor of safety and an approximate value of probability of failure is better than knowing either one alone.

REFERENCE

Duncan, J. M., and Smith, C. J. (2000). *Proc., Workshop on Risk Assessment for Seepage and Piping in Dams and Found.*, Ctr. for Geotech. Practice and Res., Virginia Tech, Blacksburg, Va.