The reliability of the geotechnical enterprise

John T. Christian
Consulting Engineer, Waban, Massachusetts 02468, USA

Keywords: geotechnical engineering, probability, reliability, risk, safety

ABSTRACT

Uncertainty is a central part of geotechnical engineering. Although in the past the profession has dealt with uncertainty by such procedures as the observational method, probabilistic methods are increasingly widely employed. In using these approaches it is important to understand the differences between aleatory and epistemic uncertainty and to distinguish between the views of probability as defined by relative frequency versus degree-of-belief. In particular, degree-of-belief approaches lend themselves to Bayesian updating, which is in turn consistent with the observational approach in geotechnical engineering. The engineer also needs to be sensitive to the different ways that spatial and systematic errors contribute to the final error. In most practical cases estimation of statistical properties relies to some extent on eliciting expert opinion, a process that can be quite difficult. In general, groups of experts tend to be good at estimating mean trends, but they also tend to be overconfident in their estimates of ranges. Examples or recent practical application of probabilistic methods include the stability of a slope in Karst terrain, the decision on how to protect a tank farm against earthquake induced liquefaction, probabilistic estimates of costs and schedules for highway projects, and risk exposure for future hurricanes in southern Louisiana. The current presentation draws heavily from the author's 2003 Terzaghi lecture (Christian 2004) with additional examples and comments based on recent experience.

1 INTRODUCTION

The world of the geotechnical engineer is an uncertain place. We have made progress in measuring the properties of geologic media and in solving analytical problems. However, we still have only an imperfect understanding of how reliable the results are. To put it another way, how much confidence can we place in the results of our efforts, and how can we describe the degree of our confidence?

Geotechnical engineers, like engineers in other disciplines, have developed several strategies for dealing with uncertainty. They include:

- Ignoring it. While on its face such a head-in-the-sand approach would seem insupportable, it is surprisingly widespread. There are many stories of agencies and corporations that wilfully ignored warnings that the assumptions underlying their decisions were fraught with uncertainty. Surely one of the earliest involved the British and Dutch governments, who in the 17th and 18th centuries sold annuities to finance their expenses (usually wars) and dismissed out of hand the objections of early statisticians that the annuities were actuarially unsound (Gigerenzer *et al.* 1989).
- Being conservative. This is an obvious and frequently sound approach. Rather than get involved in the details of how often undesirable things might happen and what their consequences might be, the engineer makes the structure or system so robust that it will resist anything. While this works in many cases, it is usually expensive, it may drag the project out to unacceptable completion times, and in some cases it may simply not be possible. Eventually one must ask how conservative is conservative enough.
- Using the observational method. The observational method has established itself as the preferred way for geotechnical engineers to deal with uncertainty in situations for which simple conservatism is unsatisfactory (Peck 1969). It involves (1) considering possible modes of unsatisfactory performance or other undesirable developments, (2) developing plans for dealing with each such development, (3) making field measurements during construction and operation to establish whether the developments are occurring, and (4) reacting to the observed behavior by changing the design or construction process. While the observational method has made it possible to carry out many projects that would have been impossible under conventional

conservative procedures, it has limitations. The engineer must have access to the decision maker if the design or construction sequence is to be changed in mid-project, the usual applications do not consider explicitly the relative likelihood of the undesirable occurrences, and field measurements are expensive. Another factor limiting wider use of both the observational method and reliability-base approaches is that some regulatory agencies and the public often demand what they consider certainty at the outset of a project.

- Quantifying uncertainty. This is the purpose of reliability approaches. Quantifying the uncertainty is consistent with the philosophy of the observational method; it might be considered a logical extension of the observational method that accommodates modern developments in probabilistic methods. It is central to this lecture. Although the geotechnical community long ago learned practical ways to deal with uncertainty, it has been reluctant to embrace the more formal and rational approaches of reliability theory while other fields of civil engineering have made major commitments to probabilistic approaches. Nevertheless, over the last ten or fifteen years several researchers have made major advances in applying probabilistic methods to geotechnical problems.

Other disciplines have developed techniques that closely resemble the observational method with or without probabilistic input. An especially notable case is the already mentioned adaptive management approach widely used in environmental and ecological management. Whatever they are called, such methods require carefully thought out programs for field measurements and explicit determination of what the measurements are going to achieve.

2 MEANING OF UNCERTAINTY AND PROBABILITY

2.1 The nature of uncertainty

What exactly do we mean when we say that something is uncertain? Do we mean that the thing occurs at random in some unpredictable way, like the roll of a set of dice? That is, is the thing so unpredictable that additional knowledge or analysis will not affect our ability to estimate it? This type of uncertainty is now known as *aleatory*, after the Latin word for gambler or dice thrower (Hacking 1975). Alternatively, we might mean that the thing is uncertain only in the sense that we do not know enough about it. For example, after a deck of playing cards is shuffled, the arrangement of the cards is fixed but unknown. We could discover the arrangement by simply examining each card in turn. However, that is precisely what we are not allowed to do, so the strategy in a game such as Bridge or Poker is to discover the arrangement by observation and induction. The uncertainty is due to lack of knowledge. This type of uncertainty is called *epistemic*, after the Greek word for knowledge (Hacking 1975). There are several other sets of words used in various disciplines to describe this difference, but the words aleatory and epistemic have achieved wide circulation and application, so they will be used here.

It will immediately be clear that the problem of establishing the geometry and properties of geologic deposits is closer to that of determining the arrangement of a deck of cards than it is to predicting the throw of a set of dice. Jensen (1997) was one of the first to point out the analogy between the configuration of geologic formations and the order of cards in a deck. In effect, the problem facing the geotechnical or geological engineer is epistemic rather than aleatory; it follows more from a lack of knowledge about materials and geometries than from inherent randomness in them.

Aleatory and epistemic uncertainties must be treated differently. If something is uncertain in the epistemic sense, the uncertainty may be reduced by additional information. Closer attention to the bidding and play of the hand in Bridge or additional exploration and testing in geotechnical engineering may reduce the epistemic uncertainty. It may not eliminate it, and the cost of reducing it below some level may not be worth it, but, in general, more information tends to reduce epistemic uncertainty. Conversely, more information will not reduce aleatory uncertainty, although it may establish more precisely the parameters governing that uncertainty.

2.2 The meaning of probability

The mathematical theory of probability is an algebra that can be derived from three simple axioms:

$$
\begin{aligned}
&P[A] \geq 0. \\
&P[A] = 1 \text{ means } A \text{ is certain.} \\
&P[A \cup B] = P[A] + P[B] \text{ if } A \text{ and } B \text{ are mutually exclusive.}
\end{aligned} \tag{1}
$$

However, none of this describes what probability is. Does it describe the relative frequency with which something happens? Or does it describe the degree of belief that something happens or exists? The relative frequency view implies that there is some underlying frequency with which things happen and that repeated trials or experiments will reveal it. The degree-of-belief view argues that most important questions do not admit of repeated trials and that most practical applications of probabilistic methods employ probability as a measure of confidence in an uncertain outcome. The frequentist argues that probability is inherent in the state of nature and that the analyst's job is to estimate it. The adherent to the degree-of-belief school argues that probability is in the mind of the individual and the analyst's job is to elicit it.

It should be noted that it is possible for the two approaches to apply to the same transaction. The insurance company prices its products as a frequentist. It employs actuaries to calculate the rates of occurrence of various events from observed frequencies. Indeed, it has great difficulty pricing insurance for an event for which it does not have much actuarial data. On the other hand, the purchaser of insurance buys it on the basis of his or her degree of belief. Each of us has one life and a limited number of houses, cars, businesses, and so on. Our decisions whether to buy insurance, how much, and what sort are informed by our own particular circumstances, the exposure we are willing to undertake, and the steps we have taken to minimize risk. Thus, the insurance company is a frequentist, and we are degree-of-beliefers.

When the geotechnical engineer processes laboratory data from many tests to obtain estimates of the properties of geological materials, the engineer is acting like a frequentist. The results are often expressed as means and standard deviations, and there is an implication that the distributions of properties observed in the laboratory apply in the field. However, when carrying out an exploration program, geotechnical engineers are trying to sharpen their degree of belief in a model of the geologic conditions at the site. The author would argue that, in geotechnical engineering, the most important issues involve the engineer's degree of belief, especially when engineering judgment is employed.

2.3 Classical and Bayesian statistics

One outgrowth of the historical arguments between frequentist and degree-of-belief schools of probability is the distinction between classical (or frequentist) and Bayesian statistics. (Bayesian methods can be applied to situations in which probability is defined by the degree of belief, and classical methods can be used with frequentist probabilities, but the usual distinction is that described here.) Frequentist or classical statistics are described in most statistics textbooks and college courses. The essential thrust of classical statistics is to answer the question, "If a particular hypothesis is true, what is the probability that the data I have before me could have been generated?" In other words, it addresses the probability of the data given the state of nature, or, in mathematical notation, $P[data | state\ of\ nature]$. Geostatistics, logistic regression, and discriminant analysis are examples of classical statistical methods. Figure 1 shows the 20% curves resulting from discriminant and logistic regression analysis of the modified ground acceleration and standard penetration data for liquefaction and non-liquefaction cases used by Christian and Swiger (1975; 1976). The results are nearly identical, as they should be. Users first coming across this type of plot are inclined to believe that a site whose data fall below and to the right of the line has a 20% probability of liquefaction. This is precisely the wrong interpretation. The actual meaning of the plot is that, if a new site were to liquefy during an earthquake, there is 20% probability that its data would fall below and to the right of the curve. Similarly the probabilities associated with locations of curves in a geostatistical analysis are not the probabilities that the lines are located correctly but the probabilities that the data used in the analysis would be observed if the lines were correct (Baecher and Christian 2003a).

Bayesian analysis addresses the converse question, "If I have before me a set of data, what is now the probability that my view of the subject is true?" That is, it gives the probability of the state of nature given the data, or, in mathematical notation, $P\left[\text{state of nature}\,|\,\text{data}\right]$. The approach was first proposed by the Reverend Thomas Bayes in 1763 and independently discovered by Marquis Pierre Simon de Laplace in 1782. Bayes received the credit, but the version of the theory now commonly used is due to Laplace.

Bayesian analysis starts with a prior probability or prior probability distribution; that is, the analyst must first estimate the state of nature before the new data are introduced. The data then provide an update to the probability of the state of nature. Additional data make possible further updates and better estimates of the state of nature. The procedure is most easily explained by a simple example that nonetheless illustrates some of the insights that arise from Bayesian analysis.

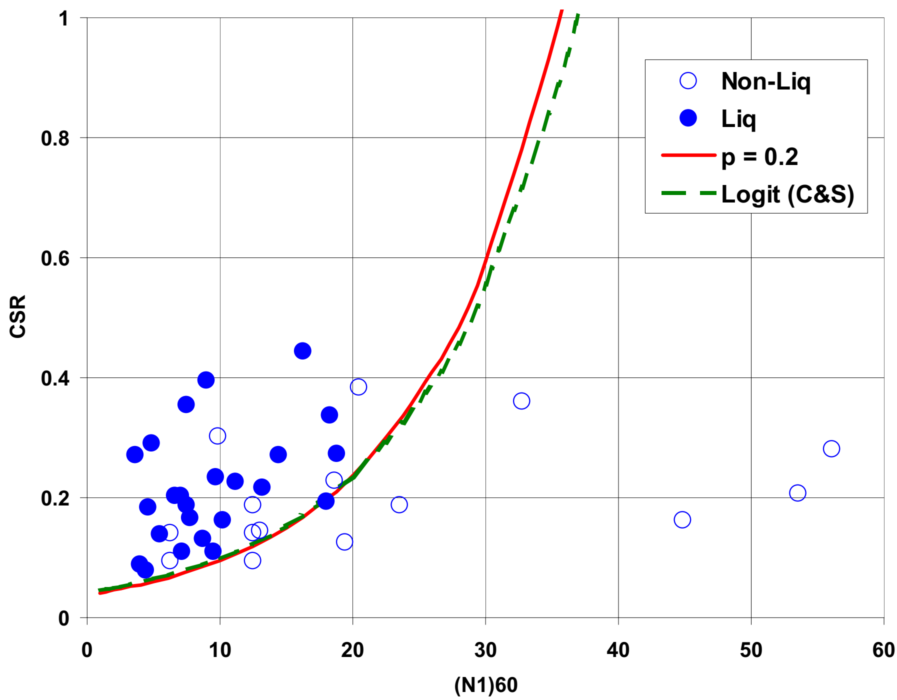

Figure 1. Cyclic Stress Ratio vs. normalized blow counts, instances of liquefaction and non-liquefaction, and 20% separation lines determined by discriminant analysis (solid line) and logistic regression (dashed line). Data are taken from the database used by Christian and Swiger (1975; 1976)

2.4 A simple illustrative example

Consider the problem of determining whether a liquefiable zone exists under a proposed facility. The field data are based on results of either the standard penetration test or cone penetration test, and the design earthquake has been specified in advance. The questions to be answered are:

- What is the probability that a liquefiable zone exists?
- How is this probability affected by the results of successive borings?
- Are more borings justified?

Let the probability of finding the zone, if it exists, be 0.3 for any one boring; hence the probability of not finding it, if it exists, is 0.7. Also, it is possible to get a false positive when no liquefiable zone exists, so let the probability of the false positive be 0.05. This implies that the probability of not finding it if it does not exist is 0.95. If *F* indicates that the zone is found, *E* indicates that the zone exists, and a superposed bar indicates the complement, then the conventional probability notation is

$$P\left[F\middle|E\right]=0.3 \qquad P\left[\bar{F}\middle|E\right]=0.7$$
$$P\left[F\middle|\bar{E}\right]=0.05 \qquad P\left[\bar{F}\middle|\bar{E}\right]=0.95 \tag{2}$$

The basic form of Bayes' Theorem states that, if there is some prior estimate of the probability that the zone exists, $P_0\left[E\right]$, the posterior probability that it exists if the zone is "found" in one boring, $P_1\left[E\middle|F\right]$, is

$$P_1\left[E\middle|F\right]=\frac{P\left[F\middle|E\right]P_0\left[E\right]}{P\left[F\middle|E\right]P_0\left[E\right]+P\left[F\middle|\bar{E}\right]P_0\left[\bar{E}\right]}. \tag{3}$$

The posterior probability that it exists if it is not found in one boring is

$$P_1\left[E\middle|\bar{F}\right]=\frac{P\left[\bar{F}\middle|E\right]P_0\left[E\right]}{P\left[\bar{F}\middle|E\right]P_0\left[E\right]+P\left[\bar{F}\middle|\bar{E}\right]P_0\left[\bar{E}\right]}. \tag{4}$$

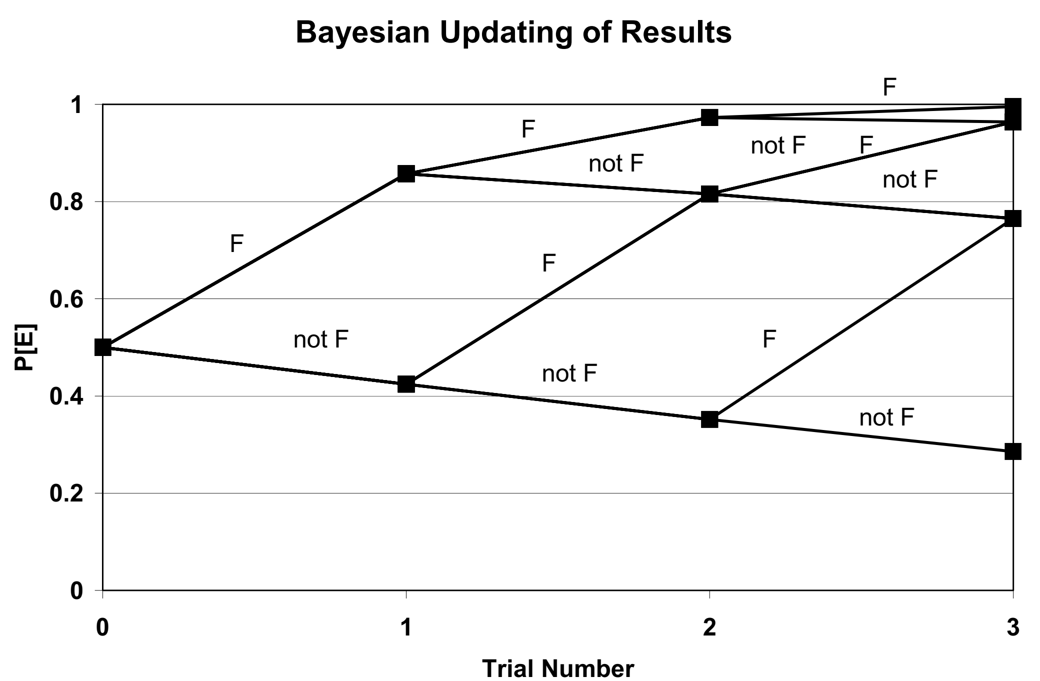

Figure 2. Posterior probability of existence of a liquefiable zone by Bayesian updating on the basis of three borings when the initial prior probability is 0.5. At each fork the upper branch corresponds to "find' and the lower to "not find."

Now, let us suppose that we are of two equal minds about whether or not the zone exists; we really do not know and would not be surprised to find that it does or does not exist. This is equivalent to

$$P[E] = P[\bar{E}] = 0.5. \tag{5}$$

Further, let the result of the first boring be that it "finds" the zone, but this could be a false positive. We want the probability that the zone exists if the boring seems to find it. Inserting the appropriate numbers into equation (3) gives

$$P_1[E|F] = \frac{(0.3)(0.5)}{(0.3)(0.5)+(0.05)(0.5)} = 0.86, \tag{6}$$

which indicates a sharp increase in the degree of belief that the zone exists. If the boring had not found the zone, equation (4) would give

$$P_1[E|\bar{F}] = \frac{(0.7)(0.5)}{(0.7)(0.5)+(0.95)(0.5)} = 0.42, \tag{7}$$

indicating that the degree of belief in the existence of the zone has decreased, but not by much.

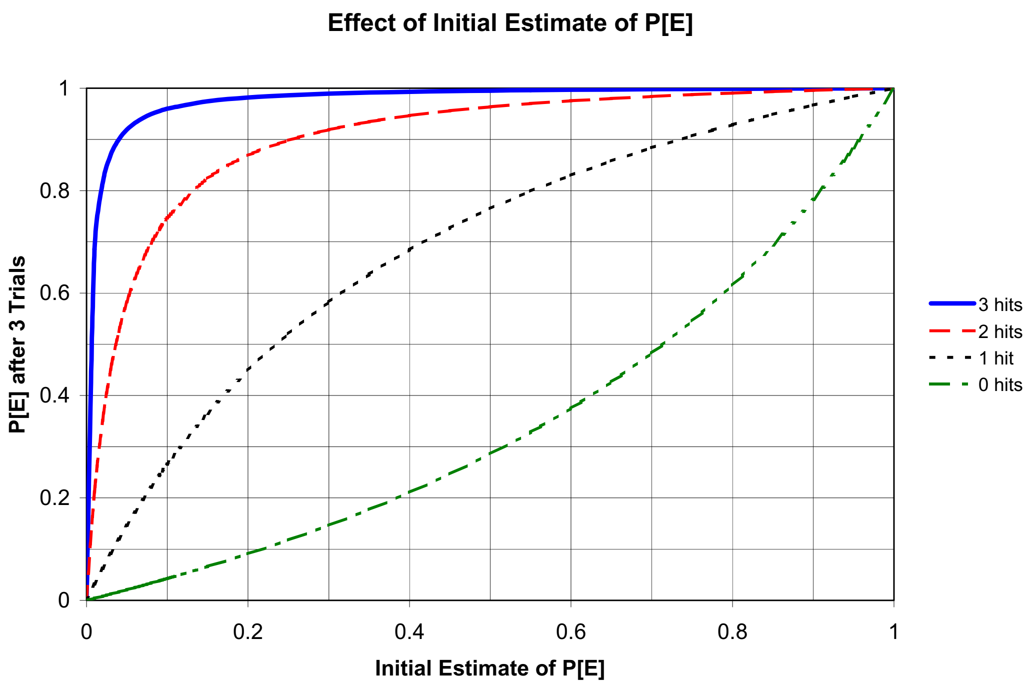

Figure 3. Posterior probabilities of existence of a liquefiable zone vs. initial prior probabilities after three borings for all possible initial priors and all outcomes of the boring program.

As results from additional borings are obtained, the probability of existence of the zone can be updated by treating the posterior result of the previous updating as the prior result for the next. Figure 2 shows all possible results for three borings when the initial prior probability of existence is 0.5. Four observations that conform to our intuitive experience are apparent:

- The order of the results makes no difference.
- Two or three positive results lead to near certainty that the zone exists for this set of parameters.
- Two or three negative results reduce the belief that the zone exists, but not by much.
- As more data accumulate, the probabilities move from the prior assignment to values that reflect the data more strongly.

The analysis can be repeated for a range of initial prior probabilities, with the results plotted in Figure 3. The horizontal axis represents the initial probability; the vertical axis is the ultimate posterior probability after three borings. The four lines correspond to the four possible outcomes: three, two, one, or no hits. Again, if there are two or three hits, the data overwhelm the prior probabilities. If there are one or no hits, there is an effect on the posterior probabilities, but it is not nearly so strong. In particular, failure to find the zone in three borings does not lend much support to the belief that the zone does not exist. This conforms to the not-uncommon experience of encountering undesirable conditions during construction despite the exploration programs carried out during design.

2.5 Types of error

A central problem facing the geotechnical engineer is to establish the properties of soils and rocks that will be used in analysis, whether that analysis is probabilistic or deterministic. An engineer wishing to estimate the vane shear strength of a Clay might be justified in choosing a value that was constant with depth and fell approximately at the mean of the measured data. The engineer would make such a choice regardless of whether the strength was to be used deterministically or probabilistically. There is some scatter about the mean in the data. On the other hand, the vane shear strength might vary with depth, so there would be substantial scatter about a constant mean value. The engineer might choose a description of the strength that varied linearly with depth, or maybe a more complicated trend line such as a sine wave would be appropriate. The data would fall closer to the trend line, so the scatter about the trend would be reduced. Unfortunately, the uncertainty in the location of the trend is correspondingly increased. The scatter about the trend line is called *data scatter*, and the uncertainty in the location of the trend is called *systematic error*. The choice of the shape and location of the trend line is *not* an artifact of nature; it is a modeling decision made by the engineer. Thus, the separation between data scatter and systematic error is also a modeling decision. This is true even if the results are used entirely deterministically. Put another way, the model for soil properties is a choice made by the modeler and is not a simple reflection of the realities of nature.

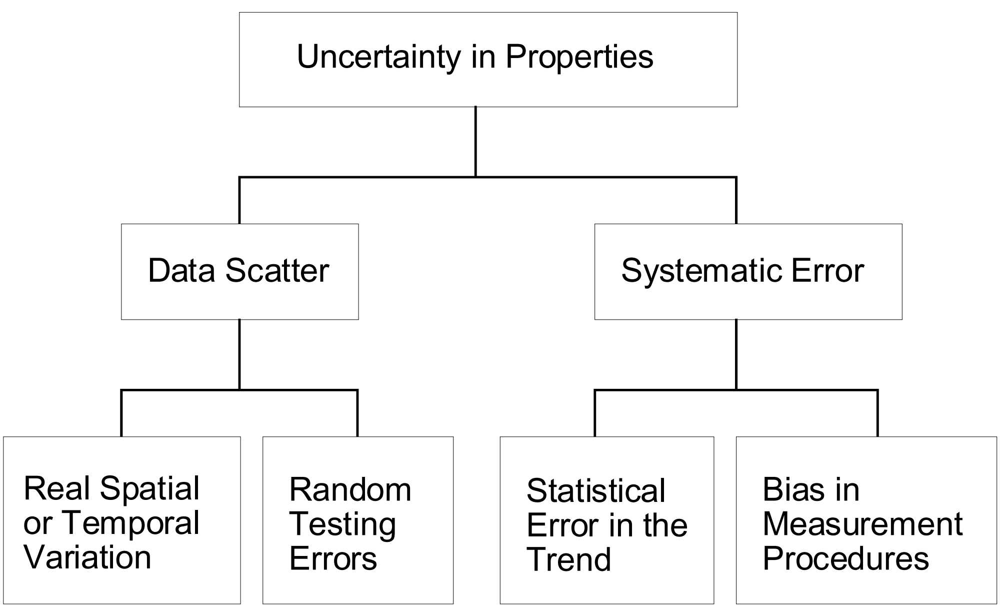

Figure 4. Conceptual separation of uncertainty into its components.

Figure 4 (Christian *et al*. 1994), depicts the dichotomy between data scatter and systematic error. It shows that data scatter can be further divided into actual spatial or temporal variation and random measurement error. It is desirable to remove the random measurement error from further analysis. The most common ways to do this are the method of moments and the method of maximum likelihood estimators, which have been described in detail by De Groot and Baecher (1993).

It is important to bear in mind that data scatter and systematic error have different effects on a reliability analysis. In many problems where the contributions of shear strength are summed along a failure surface the scatter in the value of the shear strength averages out, or nearly does so. The contribution of the scatter in the shear strength to the uncertainty of the result is thus greatly reduced as the geometry of the problem gets larger. On the other hand, the systematic error propagates throughout the analysis. There are other situations, such as those governed by a single plane of weakness or potential seepage path, in which the larger the volume the more likely the critical feature is to be found. In such cases the data scatter does not average out and is more important than the systematic error.

Another problem arises from the use of small numbers of test results. Much of statistical theory is based on the Law of Large Numbers, which can be summarized in a mathematically non-rigorous way by the statement that, if there is a large enough number of data points, statistical properties can be estimated with an arbitrary degree of accuracy. In the real world, and certainly in geotechnical engineering, there are often far from enough data to satisfy the conditions of the Law of Large Numbers. Consider a data set consisting of six values of shear wave velocity: 229, 224, 229, 217, 200, and 241 m/s. For these data the sample mean is 223 m/s, the standard deviation is 13.9 m/s, and the standard error of the mean is 5.7 m/s. In fact, these are not measured values, but the first six values created by a random number generator from an underlying normal distribution with mean of 240 m/s and standard deviation of 24 m/s. It is clear that the inferences drawn from the small sample of six values are not valid. Unfortunately, the same problem of inadequate numbers of data arises in many geotechnical problems, except that the underlying distribution is not known. Basing estimates of geotechnical properties on small numbers of data points, which is the case in many geotechnical projects, can lead to significant and unknown biases in those estimates. This is true regardless of whether the estimates are used in probabilistic or deterministic analyses.

2.6 Expert elicitation and engineering judgment

In view of the limited number of field and experimental data usually available, the geotechnical engineer often has to rely on the opinions of experts and engineering judgment to establish the values and ranges of engineering properties. Obtaining relevant information from experts, or, to use the technical term of art, "elicitation" of expert opinion, has been the subject of extensive study in the management and psychological communities. Morgan and Henrion (1990) and Vick (2002) provide accessible summaries of the issues. These can become quite complicated, so that elicitation of expert opinion is seldom the straightforward process imagined by those who have never worked on it.

The first problem is identifying an expert. Who is an expert, and how well qualified is the expert? Obviously, the expert's own opinion of his or her own worth may be too high or too low, so procedures have to be developed to establish the range of the experts' expertise. Furthermore, an expert trained in one discipline may not appreciate the statistical implications of an opinion. Some feedback and iteration is needed to address this problem.

The literature on eliciting expert opinion generally arrives at two conclusions. First, real experts tend to be good at estimating mean or median values or trends. That is, they get the expected values right. Furthermore, the average of the opinions of several experts tends to be even better. Second, experts are usually too confident in their estimates and tend to underestimate the uncertainty in their estimates.

The last points are illustrated nicely by results published by Hynes and Vanmarcke (1976). An embankment had been built north of Boston, Massachusetts. In 1974 a team from the Massachusetts Institute of Technology placed additional fill on the embankment to bring it to failure in conjunction with an international workshop at which seven acknowledged experts were invited to make predictions of the behavior of the embankment. Each was asked to predict how much additional fill it would take to cause the embankment to fail and to provide a range within which the expert's confidence of the failure was 50%, also known as the interquartile range The results, modified from Hynes and Vanmarcke's paper, are in shown Figure 5. The large square points are the experts' best estimates; the vertical lines are the interquartile ranges. The dashed line represents the actual amount of fill that caused failure: 18.7 ft. The average of the seven experts' best estimates was 15.6 ft. This is a good estimate of the actual event, especially since the critical parameter leading to failure is the total height of the embankment, not the last increment. However, the figure also shows that in no case did the actual amount of fill to cause failure fall within any expert's 50% confidence limits. Pure chance would predict that, if the 50% confidence estimates really represent the uncertainties in the experts' judgments, half the vertical lines (i. e., 3 or 4) would intersect the observed value of 18.7 ft. Thus, the experts performed well on the average, but each expert was too confident of his own estimate. When the audience was asked to estimate the required additional fill, twenty-six people submitted estimates. Once again, the best estimates were distributed approximately evenly about the actual result, but in this case sixteen of the 50% confidence estimates intersected the observed value. Thus, the audience, which had much less time to do its work, managed to include the correct value within the interquartile range 62% of the time. In this case, the experts performed less well than their audience. It is not clear why this is so. Another interesting result is that the seven experts were asked to provide, in addition to the interquartile range, the minimum and maximum values of the additional height of fill. In only three cases did the actual additional height of fill fall within an expert's minimum to maximum range. In many cases the minimum to maximum range is virtually identical to or less than the interquartile range. Hynes and Vanmarcke concluded, "It is clear that there are wide differences among engineers in the way they interpret the terms 'minimum' and 'maximum.' These widely used terms are essentially meaningless unless related to relative likelihood or probability."

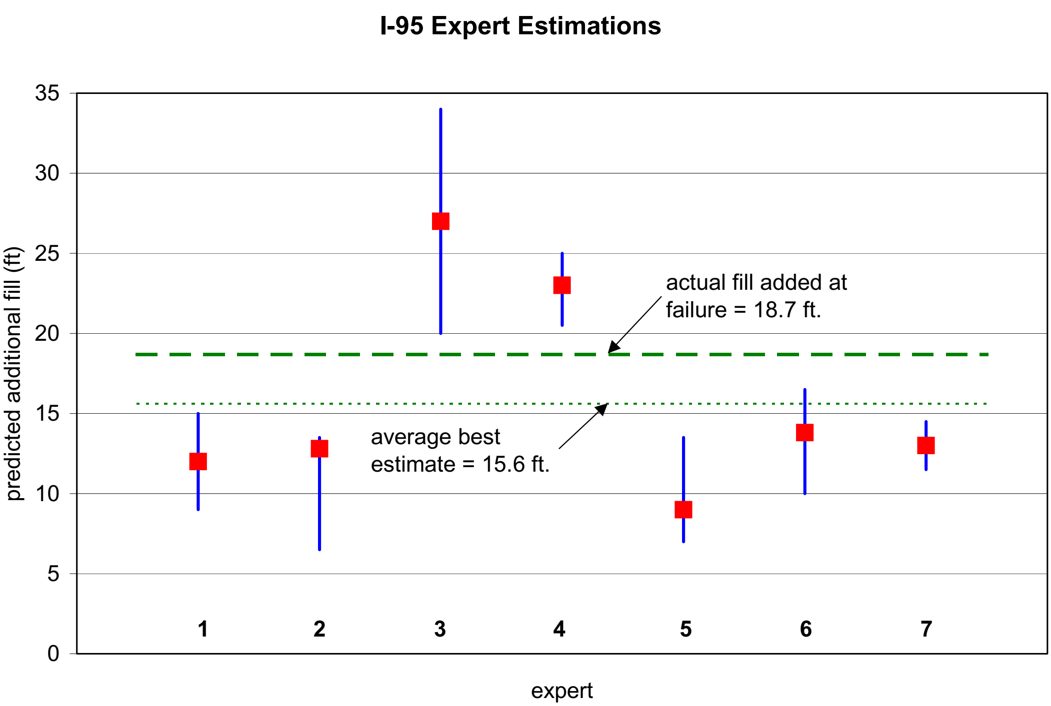

Figure 5. Seven experts' estimates of additional height of fill to cause failure of the I-95 embankment. Square points are the experts' best estimates, and the vertical bars are their 50% confidence bounds. (Hynes and Vanmarke 1976)

2.7 Selecting parametric values

The above examples suggest certain conclusions about selecting parametric values, whether they will be used in probabilistic or deterministic analyses. Among these are:

- Dividing uncertainty between spatial and systematic components is fundamentally a modeling choice and not a fact of nature.
- Spatial and systematic uncertainties contribute differently to uncertainty analyses; in particular, spatial uncertainties tend to average out.
- Values computed from small samples can be misleading.
- Experts tend to be more confident than they should be; that is, they underestimate the variances.
- Engineering judgment is invaluable if it is based on evaluated experience and a demonstrable chain of reasoning; it should never be used as a euphemism for speculation or guessing.

3 RELIABILITY ANALYSIS

The tools available to the engineer to perform probabilistically based reliability analysis fall into three categories: direct reliability analysis, logic trees, and other generalized methods of estimating probabilities of failure.

3.1 Reliability methods

Reliability methods start by assuming that there are a loading *Q* and a resistance *R*, the margin of safety *M* is defined as

$$M = R - Q \,. \tag{8}$$

If both *Q* and *R* are uncertain, so is *M*. Elementary probability theory then provides that the means (μ) and the standard deviations (σ) are related by

$$\begin{aligned} \mu_M &= \mu_R - \mu_Q \\ \sigma_M^2 &= \sigma_R^2 + \sigma_Q^2 - 2\rho_{QR}\sigma_R\sigma_Q, \end{aligned} \tag{9}$$

in which ρ_{QR} is the correlation between *Q* and *R*. If *Q* and *R* are not correlated, the last equation reduces to

$$\sigma_M^2 = \sigma_R^2 + \sigma_Q^2. \tag{10}$$

It is often more convenient to work with the logarithms of *Q* and *R*. Then the factor of safety *F* is the ratio *R*/*Q*, so

$$\ln F = \ln R - \ln Q. \tag{11}$$

Since this is similar to equation (8), we can work with the logarithms of the variables provided we use ln *F* and the logarithms of the variables. It is customary to define λ and ζ as the mean and the standard deviation of the *logarithms* of a variable. It follows that, for any distribution,

$$\begin{aligned} \zeta^2 &= \ln\left(1 + \frac{\sigma^2}{\mu^2}\right) \\ \lambda &= \ln \mu - \frac{1}{2}\zeta^2. \end{aligned} \tag{12}$$

The essence of reliability methods is to recognize that the condition *M* = 0 (or ln *F* = 0) corresponds to failure, so the problem is to find the probability that $M \le 0$. We now define a reliability index β as

$$\beta = \frac{\mu_M}{\sigma_M}. \tag{13}$$

It follows that, if we are working with uncorrelated variables,

$$\beta = \frac{\mu_R - \mu_Q}{\sqrt{\sigma_R^2 + \sigma_Q^2}}, \tag{14}$$

or, for the case of the logarithms of uncorrelated variables,

$$\begin{aligned}\beta &= \frac{\lambda_R - \mu\lambda_Q}{\sqrt{\zeta_R^2 + \zeta_Q^2}} \\ &= \frac{\ln\left[\left(\mu_R/\mu_Q\right)\sqrt{\left(1+\Omega_Q^2\right)/\left(1+\Omega_R^2\right)}\right]}{\sqrt{\ln\left[\left(1+\Omega_R^2\right)/\left(1+\Omega_Q^2\right)\right]}},\end{aligned} \tag{15}$$

in which Ω (= σ/μ) is the coefficient of variation. The probability of failure is the area under the probability density function of *M* lying to the left of *M* = 0.

Now, if *Q* and *R* are both Normally distributed, so is *M*. Then it follows that the probability of failure is

$$p_f = \Phi\left(-\beta\right), \tag{16}$$

where Φ is the cumulative distribution function (CDF) of the standard normal distribution and β is defined by equation (7). If Q and R are logNormally distributed, so is M, and equation (9) again applies, only with β defined by equation (15). In the past, evaluating the CDF required interpolation in tables, but today the CDF is a library function in spreadsheets and in mathematical software packages like Mathcad or Matlab.

While the model described by equations (14) through (16) is conceptually straightforward, calculating the various means and standard deviations is anything but simple. Furthermore, distributions other than Normal or logNormal arise often in practice. Methodologies based on Normal or lognormal distributions must be modified when other distributions exist, but the underlying theory remains similar even while the details become more complicated.

Several methods for performing the calculations have evolved over the years:

- First Order Second Moment (FOSM) methods. The idea here is that, if we know the means and the variances (the second moments) of the variables that enter into the evaluation of a function such as *M*, we can estimate the mean and variance of *M* using only first order terms in a Taylor expansion (Cornell 1969):

$$\begin{aligned}\mu_M &\approx M\left(\mu_{x_1}, \mu_{x_2}, \cdots, \mu_{x_n}\right) \\ \sigma_M^2 &\approx \sum_{i=1}^{n}\left(\frac{\partial M}{\partial x_i}\right)^2 \sigma_{x_i}^2,\end{aligned} \tag{17}$$

  in which the $\boldsymbol{x_i}$ are the uncertain variables. Equation (10) applies when the variables are uncorrelated; a somewhat more complicated expression is used when some of the variables are correlated.

- First Order Reliability Method (FORM). One shortcoming of the FOSM approach is that the results depend on the particular values of the variables x_i at which the partial derivatives are calculated. Hasofer and Lind (1974) proposed to resolve this difficulty by evaluating the derivatives at the critical point on the failure surface. Finding this point usually requires iteration, but the process tends to converge rapidly. If the variables are all normalized by dividing them by their respective standard deviations, the distance between the failure point and the point defined by their normalized means is the reliability index β. This method assumes Normal distributions and must be modified to accommodate other distributions.

- Point-Estimate (P-E) methods. The variance of a function – or any of its moments – is essentially the result of integration. Rosenblueth (1975) proposed that an accurate approximation is obtained by evaluating the function *M* at a set of discrete points and using those values to compute the desired moments. In practice, for uncorrelated variables, the points are usually taken at plus or minus one standard deviation from the mean of each of the variables. Other schemes can be used, especially when the variables are correlated or skewed. The method is a form of Gaussian quadrature.

- Monte Carlo (M-C) simulation. Monte Carlo simulation enjoys a long history and a rich literature. Each continuous variable is replaced by a large number of discrete values generated from the underlying distribution; these values are used to compute a large number of values of function *M* and its distribution. The large numbers of computations once presented a constraint on the use of this method, but cheap modern computers have largely removed this obstacle. Several so-called variance reduction schemes can be effective in improving convergence and reducing computational effort.

- Others. Perhaps the most significant methods other than those just described are the Second Order Second Moment (SOSM) and Second Order Reliability Method (SORM), which provide higher order approximations than those underlying FOSM and FORM. While these have found some applications in structural reliability studies, they have not found much application in geotechnical work.

3.2 Event Trees, Fault Trees, and Influence Diagrams

Event trees start with an initiating event, such as, say, the occurrence of an earthquake. Then the analyst develops a set of events that could follow; say the peak ground acceleration could fall within a certain range. Associated with each range is a conditional probability; for example, for the range 0.05g to 0.10g the conditional probability could be 25%. These events must be exhaustive – that is, all possible outcomes are included – and exclusive – that is, no possible result could fall within more than one outcome. The analysis then proceeds along each path to evaluate the next outcomes, and so on and so forth. At each stage the probabilities are *conditional*; that is, they are the probabilities of the current event if all preceding events in that branch have occurred. At the end of the tree, the probability of each outcome is simply the product of the conditional probabilities

Fault trees start with the failure and work backward. The tree contains the conditions that must be met for the failure to occur. There are two basic situations. If all the conditions must be met, they are connected to the event by an "and" gate; if the event will occur if one or more of the conditions are met, they are connected by an "or" gate. The analyst develops the tree from the top down, moving from condition to condition. In the usual formulation, the conditions at each stage must be independent and must encompass all the conditions that could lead to the next stage. To compute the probability of failure, the analyst works from the bottom up.

4 EXAMPLES

4.1 Slopes in Karst terrain

Several years ago the author was involved in the study of the stability of the slope behind a warehouse facility located in Karst terrain. It was immediately clear that the trad ional techniques of slope stability analysis were not applicable, but there were abundant data on the annual rate of small and large failures, ranging from general slope failures to the fall of pebbles. From these data it was possible to construct a recurrence relation for the failures, that is, a relation between the annual rate of observing a particular size of failure and the size of the failure. For each size of failure the expected damage could be estimated, ranging from dents in cars parked near the slope to collapse of the warehouse wall under the impact of a major failure. The product of the rate of a particular size of failure by the expected damage due to that type of failure gave the expected loss due to that type of failure. Adding over all the sizes of failure gave the expected annual loss. It was then a simple matter to perform parametric studies of different configurations of the slope, additional construction operations, and so on. This permitted the owners of the warehouse to decide on c course of action that was in their best interests. In turned out that the most expensive solution – cutting back the entire slope – was not economically justified, and a much more modest program of local remediation and observation was adopted. Presenting the alternatives in probabilistic terms was congenial to the thought processes of managers trained in modern financial methodologies.

4.2 Reliability of a tank far subject to liquefaction

In the 1980s the owners of a tank farm storing various petrochemical products on filled land in Tokyo Bay became concerned over their financial exposure. Many of the details of this work (in which the present author was not involved) are set forth by Baecher and Christian (2003b). The risk assessment consisted of five steps:

1. Identify the factors influencing seismic performance of the storage tanks: frequency of earthquakes, variability of soil properties, behavior of the tanks during earthquakes, performance of the foundation soils during earthquakes, performance of the fire walls and revetments during earthquakes, performance of other site facilities designed to prevent spills, and so forth.
2. Analyze earthquake source models to establish seismic hazard at the site and establish levels of ground shaking and their respective probabilities of being exceeded.
3. Develop an event tree for a typical patio; the event tree is used to illustrate how various facilities in a tank patio interact to lead to system failure.
4. Build a numerical (Monte Carlo) simulation of the entire site to analyze random behavior of each key component; run the model through many realizations to develop a statistical sample of the behavior of the site and facility.
5. Compare the predicted probability of an off-site spill and the corresponding consequences to other risk; use the comparison and the annual expected monetary loss to judge the acceptability of the risk faced at this site.

Many details are involved in each of these steps. For example, in evaluating the probability of spill from any tank it is necessary to account for the fact that the tanks are only partially full at any time and their fill levels can be described probabilistically.

The results were compared to the probability of failure of other facilities as described in the F-N plot of Figure 6. This demonstrated that the computed risks were far higher than those considered acceptable for facilities of this type, and remedial action. The remediation consisted of placing a slurry wall around the site and lowering the water table inside the patios by continuous pumping.

4.3 State of Washington highway costs and schedules

The State of Washington, in the north-western corner of the continental United States has instituted an innovative program for estimating the costs and schedules of its highway projects. In effect, the designers are required to provide best estimates of costs and schedules for the various portions of a project and their estimates of the ranges of uncertainty in those estimates. Then a Monte-Carlo simulation of all scenarios is performed to obtain a probabilistic estimate of the project's cost and time of completion. The officials making decisions about whether to proceed or how to fund the project are presented not with a single estimate but a range of expectations. Experience has shown that the officials are comfortable with this type of information, which recognizes the uncertainty that everyone knows exists in such estimates.

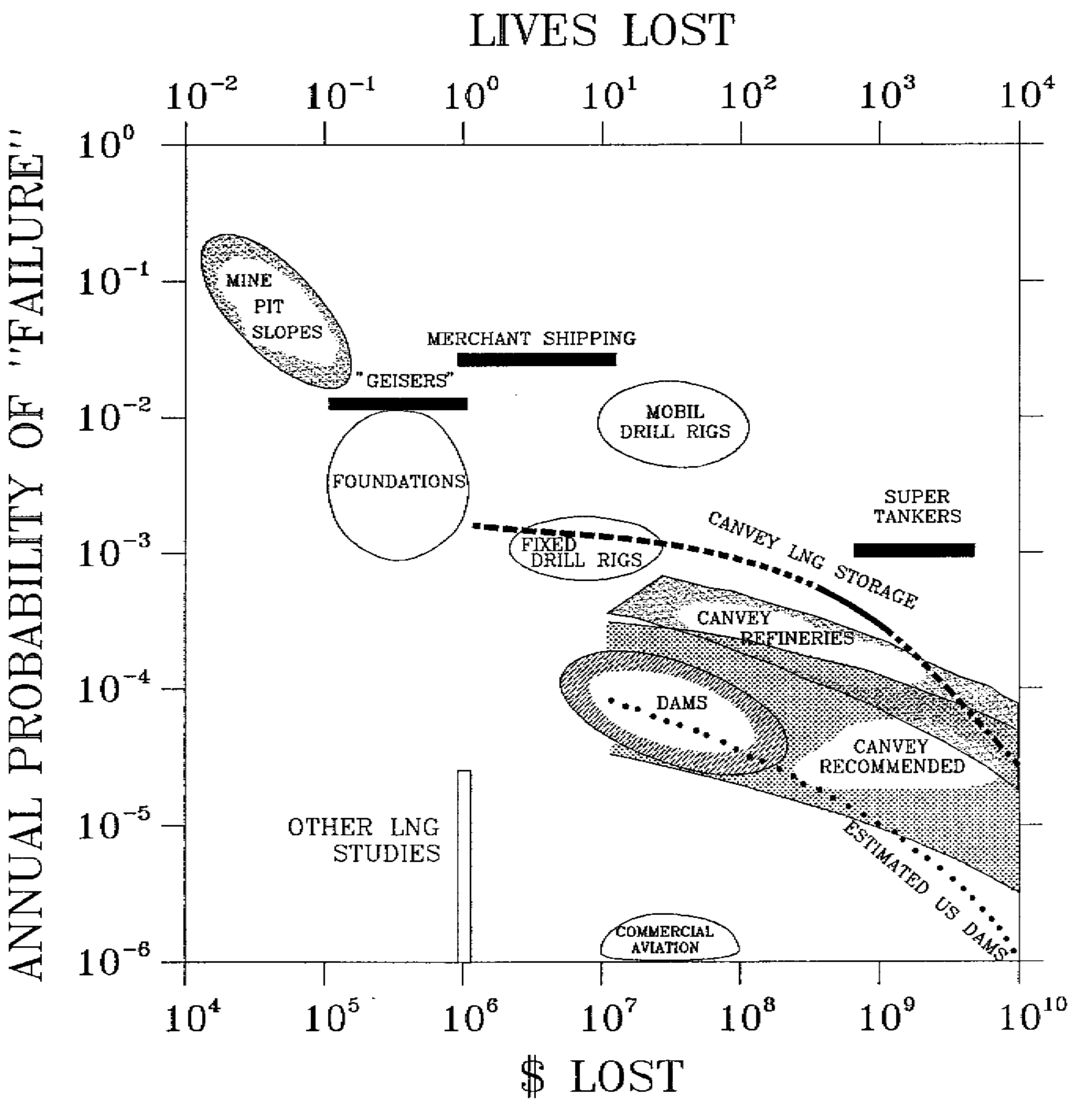

Figure 6. One version of the F-N plot of annual risk versus cost or number of lives. In this plot both cost and lives are shown; it is customary to use one or the other rather than both on the same plot.

4.4.1 New Orleans hurricane risk

The Interagency Performance Evaluation Taskforce (IPET) studying the events surrounding the August 2005 Hurricane Katrina damage to New Orleans has developed a detailed model of hurricane risk. This starts with models of many scenario hurricanes that follow probabilistically determined paths and proceeds through the various stages of hurricane attack on the southern Louisiana coast. Each step is governed by an event tree, and the results are computed by a large Monte-Carlo simulation. The results are a series of probabilistic risk evaluations for each section of New Orleans. The final version of the results is still not released at this writing, but it is expected that they will be available in the autumn. This will provide, for the first time, a rationally generated risk map for flooding of a major U. S. city.

5 CONCLUSIONS

In using probabilistic approaches it is important to understand the differences between aleatory and epistemic uncertainty and to distinguish between the views of probability as defined by relative frequency versus degree-of-belief. In particular, degree-of-belief approaches lend themselves to Bayesian updating, which is in turn consistent with the observational approach in geotechnical engineering. The engineer also needs to be sensitive to the different ways that spatial and systematic errors contribute to the final error. In most practical cases estimation of statistical properties relies to some extent on eliciting expert opinion, a process that can be quite difficult. In general, groups of experts tend to be good at estimating mean trends, but they also tend to be overconfident in their estimates of ranges. Examples or recent practical application of probabilistic methods include the stability of a slope in Karst terrain, the decision on how to protect a tank farm against earthquake induced liquefaction, probabilistic estimates of costs and schedules for highway projects, and risk exposure for future hurricanes in southern Louisiana.

REFERENCES

Baecher, G. B., and Christian, J. T. (2003a). *Discussion of 'Evaluating site investigation quality using GIS and Geostatistics by R. L. Parsons and J. D. Frost*. ASCE Journal of Geotechnical and Geoenvironmental Engineering, 129(9), 867-868.

Baecher, G. B., and Christian, J. T. (2003b). *Reliability and Statistics in Geotechnical Engineering*. John Wiley & Sons, Chichester, U. K.

Christian, J. T. (2004). *Geotechnical Engineering Reliability: How Well Do We Know What We Are Doing? The 39th Terzaghi Lecture*. ASCE Journal of Geotechnical and Geoenvironmental Engineering, 130(10), 985-1003.

Christian, J. T. and Swiger, W. F. (1975). *Statistics of Liquefaction and SPT Results*. ASCE Journal of the Geotechnical Engineering Division. 101(GT11). 1135-1150.

Christian, J. T. and Swiger, W. F. (1976). *Closure to Statistics of Liquefaction and SPT Results*.ASCE Journal of the Geotechnical Engineering Division. 102(GT12), 1279-1281.

Christian, J. T., Ladd, C. C. and Baecher, G. B. (1994). *Reliability Applied to Slope Stability Analysis*. ASCE Journal of Geotechnical Engineering. 120(12), 2180-2207.

Cornell, C. A. (1969). *Structural safety specifications based on second-moment reliability analysis*. Cambridge, Massachusetts, Massachusetts Institute of Technology, Department of Civil Engineering.

De Groot, D. J. and Baecher, G. B. (1993). *Estimating Autocovariances of in situ Soil Properties*. ASCE Journal of Geotechnical Engineering. 119(1), 147-166.

Gelman, A., Carlin, J. B., Stern, H. B. and Rubin, D. B. (1995). *Bayesian Data Analysis*. Chapman & Hall/CRC , Boca Raton, FL.

Gigerenzer, G., Swijtink, Z., Porter, T., Daston, L., Beatty, J., and Kruger, I. (1989). *The Empire of Chance: How probability changed science and everyday life*. Cambridge University Press, New York.

Hacking, I. (1975). *The Emergence of Probability*. Cambridge University Press, Cambridge, U. K.

Hasofer, A. M. and Lind, N. (1974). *An Exact and Invariant First-Order Reliability Format*. ASCE Journal of Engineering Mechanics. 100(EM1), 111-121.

Hynes, M. E. and Vanmarke, E. (1976). *Reliability of Embankment Performance Predictions*. Proceedings Engineering Mechanics Division Specialty Conference, University of Waterloo Press., ASCE:

Jensen, J. L. (1997). *Statistics for Petroleum Engineers and Geologists*. Prentice-Hall, Upper Saddle River, N. J.

Morgan, M. G. and Henrion, M. (1990). *Uncertainty: a Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis*. Cambridge University Press. Cambridge, New York.

Peck< R. B. (1969). *Ninth Rankine Lecture: Advantages and Limitations of the Observational Methodin Applied Soil Mechanics*. Geotechnique, 19(2), 171-187.

Rosenblueth, E. (1975). Point *Estimates for Probability Moments*. Proceedings, National Academy of Science. 72(10), 3812-3814.

Vick, S. G. (2002). *Degrees of Belief: Subjective Probability and Engineering Judgment*. Reston, VA, ASCE Press.