Unresolved Problems in Geotechnical Risk and Reliability

John T. Christian[1] and Gregory B. Baecher[2]

[1] Corresponding author. Consulting Engineer, 89 Kodiak Way, Ste. 2312, Waltham MA 04451, (781) 209-2799, christian2@rcn.com
[2] Glenn L. Martin Institute Professor of Engineering, University of Maryland, College Park, MD 20742.

ABSTRACT

The geotechnical engineering and research communities have made significant progress in understanding how concepts of probabilistically based reliability analysis interact with the unique features of the geotechnical environment and in developing procedures to apply probabilistic methods to practical geotechnical problems. This conference presents evidence of both the scope of current activities and the success of recent applications. However, many issues remain unresolved. This paper identifies ten problems, issues, or areas of activity in which major questions are still open. It is to be hoped that they will form the agenda for the next generation of research and development at the interface of two disciplines that address uncertainty every day – geotechnical engineering and reliability analysis.

INTRODUCTION

One of the major attractions of geotechnical engineering for civil engineering students has always been that it deals with a world in which many important matters are not well known or understood, a world in which cut-and-dried solutions seldom apply. At about the same time as the pioneers in geotechnical engineering were establishing the foundations of their discipline, another group of researchers were applying probabilistic concepts to develop rational ways to deal with uncertainty in what came to be reliability analysis. It is remarkable that there was little interaction between the geotechnical and reliability efforts.

However, in the 1970s the pioneering efforts of people such as T. H. Wu, Peter Lumb, Allin Cornell, and Robert Whitman showed that the reliability and geotechnical worlds really did have something to say to each other. Recent years have seen great advances in applying probabilistic reliability ideas to geotechnical problems, reflected both in increasing numbers of papers and in greater interest on the part of clients and practical engineers in expressing reliability in quantitative terms. The profession has learned a lot in the process, but it has also discovered that some issues remain stubbornly intractable. This paper identifies ten problems that have not been resolved satisfactorily – at least to the authors' satisfaction – and whose clarification would greatly benefit the practice of geotechnical engineering.

On August 8, 1900, speaking at the International Congress of Mathematicians at the Sorbonne, the great mathematician David Hilbert presented a set of ten important problems in mathematics. These were unsolved at the time, and some remain so today. Finding solutions to these problems strongly influenced the development of mathematics throughout the 20thC. The complete list of 23 problems that Hilbert eventually proposed became available in English translation in the *Bulletin of the American Mathematical Society* (1902). It is interesting to consider a similar list of important and unresolved problems in the application of risk, reliability, and probabilistic methods to geotechnical engineering. Although there is no reason there could not be more or fewer important unresolved problems, ten is a convenient number. These problems range from statistically-informed guidance for practical site characterization, to dealing with correlated failure modes in systems risk assessments, to risk communication and active risk management

The subsequent sections contain brief statements of the unresolved problems and some comments on where the efforts to resolve them now stand. The available space and time are not adequate for a full discussion of any one problem, each of which could easily occupy a full professional paper. The order of presentation of the problems is independent of their relative importance or likelihood of successful solution. Also, these problems are not necessarily mutually independent; progress toward resolving one of them may affect the state of others.

TEN UNRESOLVED PROBLEMS

1. Why are failures less frequent than our reliability studies predict? The reliability studies carried out over the past few decades generally give probabilities of failure on the order of several percent or more for the usual range of uncertainties in soil properties and analytical tools. We do not observe this frequency of failures. Why not?

The typical coefficients of variation (standard deviation as a fraction of the mean) reported from soil engineering property testing are on the order of 20-30% (Phoon and Kulhawy 1996; Baecher and Christian 2003). Presuming a mean factor of safety of E[FS]=1.5, the corresponding reliability indices (β) are about 1.67 to 1.1, implying probabilities of failure in the vicinity of 0.1 for Normally distributed uncertainties. These are at least an order of magnitude larger than the observed frequency of adverse performance. They are two orders of magnitude larger than the frequency of all-modes failures of earth dams (Baecher *et al*. 1979).

One answer to this riddle is that the uncertainty in soil properties is being overestimated. If the coefficient of variation in the field is actually smaller than the value used to estimate the probability of failure, the probability of failure will be overestimated. However, extensive studies of the measured variability of soil properties in the laboratory and *in situ*, (e. g., Phoon and Kulhawy 1999a and 1999b) have shown that, although different properties have different variabilities, there does not seem to be much difference between the laboratory and the field.

A second answer is that we do not apply factors of safety to mean property values but to some conservative fraction of the mean. US Army Corps of Engineers practice, as an example, is to use a 1/3-rule in choosing engineering properties: the design property is taken at that value which is larger than 1/3 of the observations and smaller than 2/3 (USACE 2003). For Normally distributed data, that is approximately the mean less 0.4 standard deviations. This implies a reliability index with respect to the mean of 1.5 to 2.1, or corresponding probabilities of failure in the vicinity of 0.05. These are more in keeping with observed rates, but still too high. There is a tendency for engineers to be conservative in estimating soil properties, underestimating strength and overestimating compressibility, even when trying to identify the best estimates. It is hard to overcome the habits of years.

Another answer is that the variation we observe in test data includes both actual variations of *in situ* properties and measurement error (*i.e.,* noise). That measurement error can be large. In the post-Katrina risk analysis studies of the New Orleans levee system, the noise in undrained strength data was estimated to be about 75% of the total data variance (USACE 2009). This is hardly surprising, especially for a measurement as prone to disturbance and measure error as unconsolidated-undrained strength. The lesson to be learned from these empirical studies is that care must be exercised in assigning coefficients of variation to soil engineering properties, and that we likely over-estimate the uncertainties in soil properties, maybe by a great deal.

Finally, many analytical models used to calculated factors of safety and probabilities of failure are not accurate. Because slope stability analyses usually involve assuming general patterns of failure, the analyst may miss a critical mode, and there can be an unconservative bias. However, most analytical models used in geotechnical engineering are conservatively biased, the bearing capacity equations being an especially good example. A further complication is the large number of multiplied factors used to account for deviations from the basic case in, for example, bearing capacity and liquefaction analyses. It is far from clear that these factors apply across the full range of parameters or even that they should be combined by multiplication. Christian and Carrier (1978) demonstrated that, even in the relatively simple case of a foundation load on a linearly elastic, isotropic medium with only two factors, one for shape and one for embedment, multiplying the factors gave less accurate results than simply ignoring the embedment factor.

It is also possible that the actual rate of unsatisfactory performance may be under-reported. Unless the failure is spectacular, it is unlikely that every case of foundation failure, excessive settlement, slope distress, and so on is included in the data repositories for geotechnical facilities.

Resolution of some of the other issues in this list may have an impact on the understanding of the infrequency of failures. In any case, the credibility of reliability analysis requires that there be a demonstrable relation between observed behavior and behavior predicted by analytical models.

2. What is the actual variability of soil and rock properties? Several studies have been published on the variability of the properties of soils and rocks (*e. g.*, Phoon and Kulhawy 1999a 1999b). Sometimes the values in these publications are simply adopted into reliability calculations without further efforts to establish the variability for the particular project at hand. Despite the excellent work that has been done on this issue, it is not a closed matter, and more work needs to be done. In particular there needs to be more work on how much effort is required to improve the estimates for a particular site.

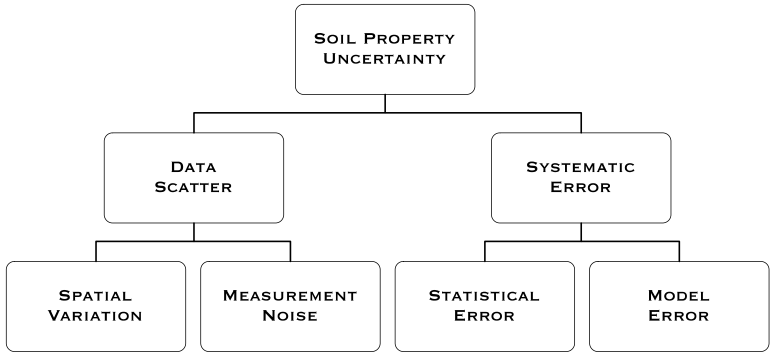

Figure 1. Contributions to soil engineering property uncertainty

This issue follows on the last concerning our over estimation of probabilities of failure. Variations in soil engineering data involve at least two things: (1) actual spatial variability from one point to another in the soil mass, and (2) noise introduced by our methods of measurement. The latter can be large. In addition, however, there are at least two bias (*i.e.*, systematic) errors that creep into our assessments: (3) statistical error due to limited numbers of observations, and (4) model error due to the approximate nature of our mathematical descriptions of the physics of soil behavior.

The systematic errors do not appear in the data scatter since they are biases on the mean. Yet, they affect our reliability calculations in profound ways, because they do not average out with scaling. For example, if we are told, “the probability of failure of a long embankment is 0.1,” does this mean that, 10% of the length is expected to fail, or there is a 1-in-10 chance that the embankment fails in its entirety, or something between these two? The answer depends on the source of uncertainty underlying the probability. If the uncertainty derives entirely from spatial variation of strengths or loads, then the first statement is correct. If the uncertainty derives entirely from a systematic error, like the model used to predict stability, then the second statement is correct. But if the uncertainty derives from a mixture of sources, some spatial and some systematic, then the third statement is correct. The third is almost always the case in practice.

One can anticipate some of the general trends of variability in soil and rock. It is to be expected that the coefficient of variation would be larger for undrained strength than for friction angle, and even larger for the coefficient of hydraulic conductivity. The variances of properties of rocks are clearly dependent on the size of the test specimen—larger specimens have smaller variance—and this raises questions of how to extrapolate to field conditions. In other words, establishing the variability of a property requires different approaches for different properties, different materials, and different ways of testing.

One complication is that the uncertainty in properties may be larger than expected and may have unforeseen consequences. For example, although one would expect that the unit weight of soil in a fill would have a relatively low coefficient of variation, actual experience has shown that field measurements of density yield surprising ranges of values. Furthermore, the variance in the unit weight of fill can have a large effect on the uncertainty in the factor of safety in slope stability calculations.

A fundamental problem is the lack of data. Field exploration is expensive, so there are seldom enough data to support meaningful statistical analyses. Similarly, there is a tradition of basing conclusions in laboratory testing programs on small numbers of tests. It is quite common to encounter laboratory programs consisting of three or four consolidation tests or triaxial strength tests. These are not enough to support broad conclusions about the statistics of soil properties, and run afoul of Tversky and Khanaman's (1971) "law of small numbers" biases.

Phoon and Kulhawy developed their statistical results by scouring the literature to find enough data. However, this should not be the end of the matter. Others should be extending the available data, looking for more data, and refining the statistical studies. One hopeful development is that the ASCE now encourages authors of papers in the Journal of Geotechnical and Geoenvironmental Engineering to submit supplemental data, which become available when the electronic form of the paper is accessed.

3. The effects of spatial correlation and how to deal with them. Geological materials arrive at their present configurations by either an orderly construction process operated by humans or a geologic process that follows physical principles. Therefore, their physical properties exhibit spatial correlation to a greater or lesser extent. While there has been some success in describing spatial correlation by techniques such as auto-correlation and geostatistics, and spatially correlated random finite element methods have been employed to analyze their effects, the techniques for dealing with spatial correlation are difficult to implement, they are poorly understood by the practice, and their consequences are often ignored.

Fenton and Griffiths (2008) have used stochastic finite element methods to investigate the effects of spatial correlation in classical geotechnical problems such as flow through dams, slope stability, and settlement of shallow foundations. Their book summarizes research results that were previously published individually. The results show that correlation can have significant effects, even on problems that are thought to be well understood. Although Fenton and Griffith's work comprises by far the most extensive investigation of correlation effects, more limited studies by others have shown similar effects.

Engineers often deal with phenomena that are difficult to analyze by considering the limiting cases. For correlation effects, these are the cases of perfect correlation ($\delta = \infty$) and no correlation ($\delta = 0$). In the case of perfect correlation the values of the uncertain parameters are the same for all distances, and correlation effects need not be explicitly considered. In the uncorrelated case, random variations about the expected value average out in any analysis with fine enough discretization. For a problem such as settlement of shallow foundations or stability of a slope, for which the relevant properties are averaged over a broad extent, one would expect the cases of perfect correlation and no correlation to give the same expected results. However, intermediate values of correlation may give larger values of, say, relative settlement, and the maximum effect in some cases seems to occur for values of correlation coefficient approximately equal to the distance for which the difference in performance is sought. Where correlation effects are concerned, the conventional approach of considering limiting cases does not work, and the analysis of the actual correlation effects is difficult.

A further concern is that determining correlation patterns in the field is not easy. Auto-correlation methods and geostatistics require large amounts of data taken over a broad range of distances. In an actual project it is rarely feasible to collect the data using conventional exploration tools. Can the exploration techniques be improved to provide adequate descriptions of correlation? How can correlation best be described? Are there general patterns that can be used without the need for a full-blown exploration program for every project? How can correlation effects be conveniently incorporated in analyses without recourse to stochastic simulations in every case?

4. Scale effects in geotechnical reliability. Much of geotechnical engineering involves scaling properties from small laboratory samples or from field tests on limited volumes. We know that scaling to full scale properties is a statistical problem, but it has not been fully addressed or assimilated by the profession.

Some geotechnical problems are governed by average properties; others are dominated by the behavior of local seams or discontinuities. The settlement of a shallow foundation is an example of the former, and the stability of a rock slope is typical of the latter. Sometimes, as in the case of compaction testing, the properties of a small sample are reasonably comparable to those in the field, but often, as in the testing of rock samples, the behavior of a small sample is notably different from that of the rock mass. The combination of the relative importance of local discontinuities and different effects of sample size creates conditions in which scaling the properties of soils and rocks becomes a difficult and often poorly understood problem.

Geotechnical engineers and researchers have long recognized that scaling problems present difficulties. There has also been some progress in addressing them rationally. Unfortunately, many of the problems remain unresolved, and practitioners are often not aware of what progress has been made. This is an area in which progress in research could have important impact on practice.

For example, levee systems comprise constructed embankments or walls extending tens of miles across ground that is poorly characterized from an engineering perspective. Levees fail, if they do, at locations where loads are high with respect to strength, or where seepage resistance is low. If these critical locations are identified ahead of time, traditional methods can be used to analyze stability and calculate factors of safety or probabilities of failure. In such situations, the overall length of levee is immaterial, because the weakest spots have been identified and dealt with. The probability that the levee fails is that of these weakest spots.

The more common situation is that the full length of the levee system is not characterized with enough detail for the engineer to know unambiguously where the weakest spots are. In this case, any section of the levee system has some probability of experiencing higher than average loads or lower than average characteristics, and as a result, of being a “weak spot.” Since this unfortunate combination cannot be uniquely identified before a failure occurs—there is not enough information to do so—the longer the total length of levee, the greater the chance that such an unfortunate combination exists somewhere, and thus the higher the probability of a failure somewhere.

This "probability of a failure somewhere" is the *probability of system failure*. Other things being equal, most people would agree that the longer the length of levee, the greater the chance of system failure. The probability of system failure is not a property of natural randomness in nature, but of the limited information available to the engineer with which to characterize the levee system, and consequently to know where the weakest link is and how weak it is.

For the purposes of reliability modeling, common failure modes can be categorized as one of three types with respect to the effect levee length has on the probability of system failure: (1) those depending on continuum properties of the levee structure or subsurface—examples are limiting equilibrium strength stability, or large high permeability zones in levee foundations; (2) those depending on unknown or undetected "flaws" in either the levee or subsurface—examples are buried channels, cracks in the levee structure, or animal burrows; and (3) those depending on known discrete features like through-going pipes, transitions between levee or wall sections, and gates.

The probability of failure for the first two categories is affected by the length of the levee. In both cases, longer levees, in principle, have higher probabilities of system failure, because the chance of encountering either a weak zone (category 1) or a flaw (category 2) increases with length. The probability of failure for the third category is unaffected by length, because critical locations are known.

In a seminal paper dealing with spatial variability, Vanmarcke (1977) showed that both the probable length of failure and the growth of systems failure probability with levee length could be related to the autocorrelation distance of soil strength properties. This leads to an equation for systems failure probability of the form,

$$P(one-or-more-failures) = 1 - (1 - p)^n \tag{1}$$

in which n = the number of equivalent independent reaches, which depends on the characteristic length of correlation, and p = the probability of failure of the individual reach.

The difficulty with this formula is, that while theoretically elegant, it seems to strongly over-estimate the probability of systems failure in practice. For a practical case such as New Orleans or Sacramento, the probabilities calculated with this formula far exceed the observed rates of levee failure. Why? One interesting direction of study is better to understand the positive correlations among the uncertainties affecting each levee reach, and thus lowering the probability of a first-excursion failure.

5. Can we develop reliability models for internal erosion? Internal erosion or piping is a well-known phenomenon, especially in dam and levee construction, and it clearly involves a stochastic process of successive removal of material by the forces of flowing water. However, the practical design methods are deterministic, and we do not have adequate methods to detect its occurrence in existing structures.

About one-third of the failures of modern earth dams occur because of internal erosion, and, as a result, there has been a great deal of attention paid to this failure mode and to how its probability of occurrence for a specific structure might be assigned. The major problem in rationally assigning these probabilities, however, is that we lack a physics-based engineering model for predicting internal erosion and consequent piping from first principles. The issue here is not one of applying probabilistic methods to a problem whose physics is well understood. It is a problem for which the basic mechanical processes have not been described adequately. Internal erosion or piping clearly involves a stochastic process of flowing water removing particles one by one from the erodible zone. However, the criteria used for design of inverted filters are expressed in terms of the grains sizes of the different layer and do not incorporate the thicknesses of the layers. Workers on dam safety recognize that current methods do a poor job of predicting the formation of a solution cavity or even of identifying it after it has developed.

Our current reliability model for internal erosion derives from pioneering work done at USBR in the 1990's (Von Thun 1996) and by Fell and his colleagues in Australia (Fell and Fry 2007). This work provides valuable insight into the causes and prediction of internal erosion. It decomposes the process of internal erosion into discrete steps: the initiation of erosion at a flaw, the subsequent progression of that erosion through the embankment, the continuation of the erosion to create a self-supported pipe, the failure of human intervention to prevent the development of a pipe, and finally the event of the pipe forming a breech (Figure **2**).

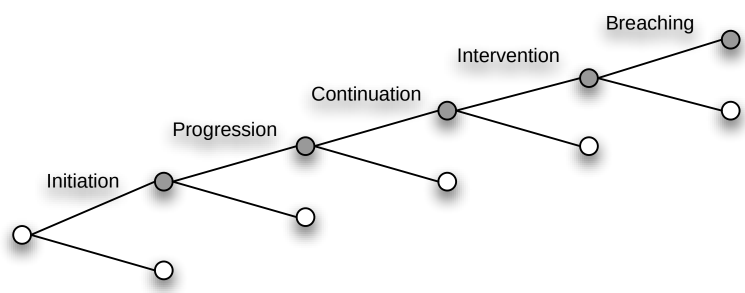

Figure 2. Event tree for internal erosion failure in an earth dam

The difficulty with this analytical approach is that the first three components of the event tree are essentially unobservable. The probability at each stage is subjectively assessed but there is no ground-truth by which to judge whether the resulting computed probability of piping failure makes sense. With no disrespect to the insightful work that has led to this reliability approach, the event tree analysis itself is a little like theology, but of course the present authors have nothing better to propose.

In view of the potentially catastrophic consequences of internal erosion, successful resolution of the probabilistic mechanics of internal erosion could have a major impact on the safety evaluation of existing facilities.

6. Connecting the observational method to Bayesian updating. Peck (1969) described the development of the observational method originally proposed by

Terzaghi and provided several examples of its practical use. In a paper near the end of his life, Terzaghi (1961) wrote, "Soil engineering projects […] require a vast amount of effort and labor securing only roughly approximate values for the physical constants that appear in the equations. The results of the computations are not more than working hypotheses, subject to confirmation or modification during construction. In the past, only two methods have been used for coping with the inevitable uncertainties: either adopt an excessively conservative factor of safety, or make assumptions in accordance with general, average experience. The first method is wasteful; the second is dangerous. A third method is provided that uses the experimental method. The elements of this method are 'learn-as-you-go:' Base the design on whatever information can be secured. Make a detailed inventory of all the possible differences between reality and the assumptions. Then compute, on the basis of the original assumptions, various quantities that can be measured in the field. On the basis of the results of such measurements, gradually close the gaps in knowledge, and if necessary modify the design during construction."

The observational method requires that information be collected as the project is carried out so that the design and construction strategy can be updated. It also requires that there be an initial plan for how to deal with changed conditions. Invoking the observational method to make the design more robust without making provisions for obtaining updated information and for changing the design in response is a prescription for unsatisfactory performance. In other words, the observational method requires rational consideration of the uncertainties in the parameters and a clear plan for dealing with events as they develop.

Bayesian thinking is essentially the same logic that Terzaghi describes, but based on mathematical logic, and as a result it is a powerful basis for inference. *Bayes' Rule* says that uncertainties expressed as probabilities can be modified (*i.e.,* updated) by observational information according to the conditional probability (*Likelihood*) of those observations were a certain hypothesis true or not,

$$P(hypothesis \mid data) \propto P(hypothesis) P(data \mid hypothesis) \quad (2)$$

in which $P(hypothesis \mid data)$ is the updated probability based on having made certain observations, $P(hypothesis)$ is the probability before seeing the data, and $P(data \mid hypothesis)$ is the Likelihood of the data. Wu (1974), Einstein *et al.* (1976), and others long ago drew the connection between the observational method and Bayesian thinking, but the two approaches—one growing out of geotechnical practice and the other out of statistics—have yet to be functionally combined except in special cases such as Einstein's Tunnel Cost Model. Yet, the benefits could be substantial.

Bayesian thinking has permeated a range of fields from artificial intelligence to criminal forensics to the attribution of disputed authorship. It served as an important piece of Alan Turning's approach to cracking the German *Enigma* codes during World War II. Why haven't we used it to bring the observational method into the 21st century?

7. Risk communication and active risk management. We have developed methods for estimating the risk of various types of failure. We have not developed ways to communicate the risks to owners and the public, despite wide use of F-N diagrams. Nor do we have methods for managing and controlling the risks of our projects while projects are underway.

There are actually two types of risk communication involved: communication with the owners and communication with the public. In both cases the basic problem is that the engineer must convert technical knowledge of risk into language that can be understood by persons who are not conversant with the technology used to develop the estimates of the risk. However, the public (or publics) has a much wider range of concerns and agendas and has great difficulty dealing with risks posed by low probability events that have large consequences.

The F-N diagram (Figure 3) is now often used to communicate relative risks. It shows the annual probability of exceedance for different levels of damage and is especially useful for conveying relative risks. The IPET studies of the effects of Hurricane Katrina developed color-coded contour plots of expected levels of inundation with annual probabilities of 2%, 1%, and 0.2%.

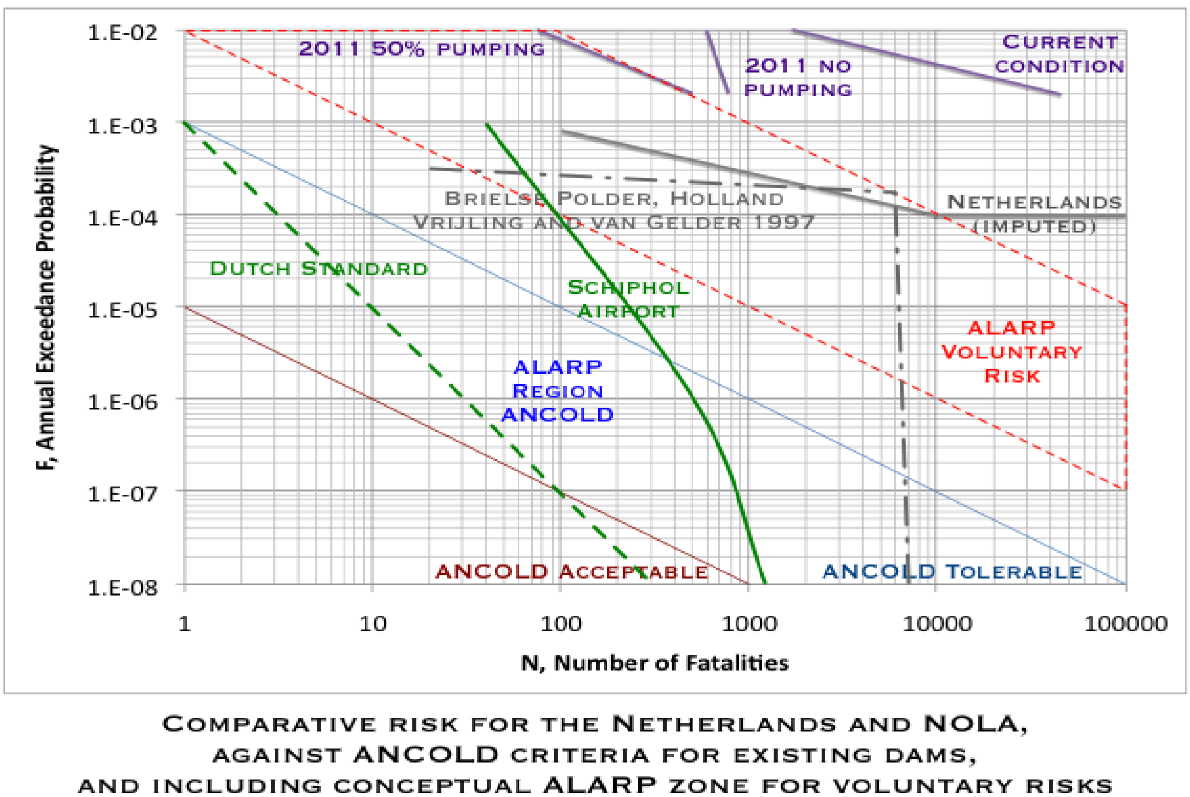

Figure 3. F-N plot for flooding risks, including ANCOLD criteria for dams.

Although computers now make it possible to prepare graphical descriptions of risk in formats that convey the essential information accurately both to owners and to the general public, the unthinking use of computerized graphics can also yield misleading and incomprehensible plots. It is now very easy to produce very bad graphics. Except for the F-N plot, we have not established standard ways to disply our results.

The underlying issue is that it is difficult for the public, owners, and even engineers to deal with and plan for extremely rare events that have severe consequences. How should resources be allocated? Which of many dire scenarios should be taken seriously? We would like to achieve some sort of balance between events with different probabilties of occurrence and diferent consequences, but we can be sure that we have not yet achieved it.

8. Applying LRFD when properties are determined by effective stress. Load and Resistance Factor Design (LRFD) was originally developed as a rational way to qualify the contributions of different loadings in the analysis of steel and concrete structures. Resistance was and remains a single variable in those applications. The factors are applied primarily to the loads. LRFD has been applied with some success to develop design procedures for piles, which are essentially structural members embedded in the ground. The approach has been somewhat less satisfactory when applied to retaining structures, and even then effects such as strength dependent on normal stress and varying water pressures are often ignored. Attempts to apply the LRFD philosophy to general geotechnical problems including effective stress, uncertain water levels, uncertain failure modes, and other common geotechnical phenomena have often devolved into efforts to make the new design procedures give the same results as the old ones. Can a method that has revolutionized structural engineering be applied rationally to geotechnical engineering? If so, how?

A second issue is that it is relatively easy to explain how the various load combinations ought to be chosen to be consistent with their contributions to the probability of failure. It is much harder to explain how this procedure works when the loads and resistance interact, as they do in frictional materials. There is a danger that the factors become mysterious parameters with no physical justification and that the designers therefore lose track of what they are doing.

The idea that different safety factors ought to apply for different parameters or different loading conditions has a long history in geotechnical engineering. Taylor (1948) discussed the concept in the context of slope stability analysis. Also, codes have been developed in which load and resistance factors were developed to be generally consistent with previous geotechnical practice and structural design methods. However, it does not appear that the research community has developed an approach that is consistent with the principles of LRFD and with the physical principles of soil and rick mechanics.

9. Guidance to practitioners on exploration strategies. There is a large literature on sampling strategy, but it has had little impact on the way we do explorations for new projects. Part of the problem is that the results are not very useful unless they are combined with Bayesian updating. There is a need for serious study of what can be done in a realistic environment and for communicating these insights to practitioners.

Much of current sampling strategy evolved from military applications such as searching for mines. The first result of the theory is that one needs to do a lot of sampling to reduce the probability of error to an acceptable level. This is not a very useful result. Bayesian updating can be used to improve the situation.

However, we have not developed sampling strategies that are consistent with modern Bayesian methods and that are understood by people who must make decisions in the field about borings and sampling. To complicate matters, it is clear that the appropriate strategies depend strongly on the geological settings and the nature of the facility to be built.

If a value of a soil property, such as shear strength, or of a field measurement, such as settlement, lies far from the previously observed general trend, what impact should this have on the design procedures? While there are statistical techniques for dealing with outliers, it is not clear how they can be applied in the actual practice of geotechnical engineering. This may be a modern extension and elaboration of Terzaghi's well-known emphasis on the importance of "minor geologic details."

10. Can we improve on "DeMorgan's Rule" for multiple failure modes? In situations such as dam safety the usual rules for calculating the failure probability of the system from the computed failure probabilities of the individual modes of performance often lead to results that seem unreasonably conservative. Current practice is to assume the extreme cases that the behaviors of the individual failure modes are either perfectly correlated or perfectly uncorrelated ("De Morgan's Rule"). Is this the best we can do? How can we improve on this approach?

In the situation where one does not know whether the failures of the modes are in fact probabilistically independent, the probability of system failure, presuming all the modes mutually independent, is an upper bound. The lower bound on the probability of system failure is given by the case of perfect dependence among all the probabilities $p_i = p_j = p$; that is, for the case in which, if one mode fails, they all fail, or if one mode does not fail, they all do not fail. In this case the probability of system failure is the same as any mode failing, p. Thus, the bounds on the probability of system failure are,

$$p \leq P(\text{system failure}) \leq 1 - (1 - p)^n \tag{3}$$

in which n = the number of modes. Ang and Tang (1984) have called these, the *uni-modal bounds* on the probability of systems failure. They are sometimes also called, the *DeMorgan bounds*, after the logician Augustus DeMorgan (1806-1871). It should be noted that these bounds are very wide if n is large and p is small.

There has been considerable discussion over how to select a unique probability of system failure within these bounds, but no satistactory solution appears at hand. Bromwell, Dean, and Vick (2006) proposed taking the mid-point, but that is an arbitrary choice. Why the mid point? Why not the mid point of the logarithms of the probabilities? The nuclear safety industry adopts a simple "beta factor" or weighted average of the two bounds, typically about 25% of the independent case (Ali Mosleh, personnal communication, 2010).

The problem with this formulation is that the probability of system failure rises quickly with n. Note, this is almost identical to the "length effect" problem. If we are considering piping forming at the interface between core and foundation, there could be an essentially infinite number of possible channels, in which case $P \rightarrow 1$.

In practice, the uncertainties in the failure modes are likely not to be probabilistically independent, and thus DeMorgan's Law is only an upper bound, and may in fact be quite a distant upper bound, especially for large *n*, implying failure geometries that are spatially adjacent or even overlapping. The problem is that we don't have a good way of estimating the degree of probabilistic dependence among the separate failure modes.

SUMMARY

The ten issues described above range from technical, mechanical problems that have defied solution to matters of communication between researchers, designers, and constructors. It is certainly not an exhaustive list, but we present it so that it will challenge the geotechnical reliability community in its efforts in the coming years. It will be interesting to return to these issues in the future to see how much progress we have made.

ACKNOWLEDGEMENTS

The authors are indebted to the helpful comments of many colleagues on an earlier draft of this paper, but especially to (in alphabetical order): J. Michael Duncan of Virginia Tech; W. Allen Marr of Geocomp Corporation; Therese McAllister of the National Institute of Standards and Technology; Thomas O'Rourke of Cornell University; Robert Patev of the US Army Corps of Engineers; and K.K. Phoon of the National University of Singapore.

REFERENCES

Ang, A. H. S., and W. H. Tang (1984) *Probability Concepts in Engineering Planning and Design, Vol. 2 – Decision, Risk, and Reliability*, New York, J. Wiley.

Baecher, G. B. and J. T. Christian (2003). *Reliability and Statistics in Geotechnical Engineering.* Chichester, West Sussex, England ; Hoboken, NJ, J. Wiley.

Bromwell, L.G., R. G. Dean, and S. G. Vick (2006) Report of Expert Review Panel Technical Evaluation of Herbert Hoover Dike Lake Okeechobee, Florida. Prepared for South Florida Water Management District. BCI Engineers and Scientists, Lakeland, Florida.

Christian, J. T., and W. D. Carrier III (1978) "Janbu, Bjerrum, and Kjaernsli's chart reinterpreted," *Canadian Geotechnical Journal*, 15(1), 123-128.

Einstein, H. H., D. A. Labreche, *et al*. (1976). "Decision analysis applied to rock tunnel exploration." *Engineering Geology*. 12(2), 143-161.

Fell, R. and J. J. Fry (2007). *Internal erosion of dams and their foundations*. London ; New York, Taylor & Francis.

Fenton, G. A., and D. V. Griffths (2008) *Risk Assessment in Geotechnical Engineering*, New York, J. Wiley.

Hilbert, D. (1902) "Mathematical problems." *Bulletin of the American Math.ematical Society*, 8(10), 437-479.

Hill, P., D. Bowles, *et al.* (2003). “Estimating overall risk of dam failure: Practical considerations in combining failure probabilitries.” *ANCOLD 2003 Risk Workshop*. Hobart, Tasmania, Australian National Committee on Large Dams.

Peck, R. B. (1969) “Advantages and limitations of the observational method in applied soil mechanics.” *Geotechnique,* 19(1), 171-187.

Phoon, K. K. and F. H. Kulhawy (1996). “On quantifying inherent soil variability.” *Uncertainty in the Geologic Environment*, Madison, WI, ASCE.

Phoon, K. K., and F. H. Kulhawy (1999a) “Characterization of geotechnical variability.” *Canadian Geotechnical Journal*, 35(4), 612-624.

Phoon, K. K., and F. H. Kulhawy (1999b). “Evaluation of geotechnical property variability.” *Canadian Geotechnical Journal*, 35(4), 625-639.

Tversky, A. and D. Kahneman (1971). “Belief in the law of small numbers,” *Psychological Bulletin, 76*(2): 105-110.

USACE (2003). Slope Stability. U. S. Army Corps of Engineers. Washington, USACE.

USACE (2009). Performance Evaluation of hte New Orleans and Southeast Louisiana Hurricane Protection System. Washington, U. S. Army Corps of Engineers. Volume VIII - Engineering and operaoinal risk and reliability alanysis.

Vanmarcke, E. (1977). "Probabilistic modeling of soil profiles." *Journal of the Geotechnical Engineering Division, ASCE* 103(11), 1227-1246.

Von Thun, J. L. (1996). “Understanding seepage and piping failures – the No.1 dam safety problem in the west.” *ASDSO Western Regional Conference*. Lake Tahoe, Nevada,, Association of State Dam Safety Officials.

Wu, T. H. (1974). "Uncertainty, Safety, and Decision in Soil Engineering." *Journal Geotechnical Engineering Division, ASCE,* 100(3), 329-348.