Methods for Developing Defensible Subjective Probability Assessments

WILLIAM J. ROBERDS

Typically, some degree of uncertainty exists in the value of geotechnical parameters at any site (e.g., owing to insufficient data, natural spatial variability, or possible changes with time). Often, this uncertainty must be quantified (e.g., in terms of probability distributions that express the relative likelihood of any value). Because of inevitable data base deficiencies, those probability distributions must be based to some degree on subjective assessments, reflecting personal opinions and judgment, consistent with all available information (site-specific and generic), and recognizing the entire range of possible values. Subjectively derived probability distributions can represent the opinions of individuals or of groups. There are problems associated with either, which, if uncorrected, render the results suspect and difficult to defend. Various techniques have been developed to conduct subjective probability assessments with varying effort and success in mitigating such problems. Thus, the appropriate technique is that which provides the desired level of defensibility at least cost.

Geotechnical parameters are often complex, varying spatially and, in some cases, with time, as a function of scale and possibly other factors. The values of those parameters must often be estimated (e.g., for analysis and design). However, data bases regarding those parameters in many cases will contain a small number of samples and inexact representation of the true conditions because of the cost involved in gathering representative data. Parameter values cannot be determined accurately in such cases where the data base is statistically insufficient. Instead, the parameter values must be estimated on the basis of whatever information is available, including generic as well as site-specific data. Those estimates, depending on their application, may represent conservative assumptions (e.g., to demonstrate compliance with some criteria) or best guesses (e.g., to predict actual performance of alternatives as input to decision making). Such estimates must necessarily incorporate interpretations and judgments regarding the data base, which are subjective and in many cases nonunique, and thus may be open to controversy. Controversy can significantly delay a project and cause unnecessary expense, especially if an ultra-conservative assumption results. Hence, the objective is to produce cost effectively appropriately defensible estimates of parameter values where significant uncertainty exists and must be subjectively assessed.

Many examples of such uncertainty analyses exist (*1*). This paper is derived from a detailed manual developed by the author specifically for conducting subjective probability assessments (*2*).

Golder Associates, Inc., 4101 148th Avenue, N.E., Redmond, Wash. 98052

UNCERTAINTY/PROBABILITY CONCEPTS

Variables can represent (a) the state at a particular place and time or (b) where the state may vary spatially or temporally or both, a group statistic (e.g., mean or variance of the population). The variables, in either case, would have a unique value. Moreover, such variables may be "continuous" (i.e., it may have an infinite number of possible states) or may be "discrete" (i.e., it may have a finite number of possible states).

Often, the state of a variable has not been directly and accurately observed, and there will generally be some uncertainty as to what the state of that variable actually is, was, or will be. The possible sources of this uncertainty can be summarized as follows (*2*):

- Statistically insufficient data: In direct observations of a variable state, measurement errors (random or systematic) and accuracy limitations may exist. Where the state has not been directly observed, it must be inferred (e.g., interpolated, extrapolated, or analytically derived) from other information. In analytically deriving a variable state from other site- and time-specific measurements, there may be imperfect understanding regarding the processes involved and approximations and simplifications in the analytical procedure. The applicability of indirect observations in the inference of the variable state must be considered. In assessing group statistics there may not be enough data to be statistically significant or the data may not accurately represent the population (i.e., biased sampling).
- Natural spatial or temporal variability or both: The variable state may vary spatially or temporally (i.e., change with time) or both. For example, the space may not be homogeneous and uniform and may instead have heterogeneities or the variable state may be affected by future events that cannot be predicted with certainty. In interpolating or extrapolating from observations (direct or indirect) elsewhere or at other times or both, this spatial or temporal variability and the effects of heterogeneities and of events and processes (both past and future) must be considered.

The uncertainty in the actual state of a variable can be quantitatively expressed in various related ways (e.g., ranges, accuracy measures, confidence levels, or probability distributions) (*3*). As is illustrated in Figure 1, probability distributions can be defined for

- Discrete variables (Figure 1a), in terms of a probability mass function (pmf), which expresses the probability of each possible variable state;

- Continuous variables (Figure 1b), in terms of (a) probability density function (pdf), which expresses the relative likelihood of each possible variable state, (b) cumulative distribution function (cdf), which expresses the probability that the variable state will be less than or equal to each possible variable state;
- Group statistics (Figure 1c), in terms of pdfs or cdfs or both for the mean (μ_x, first moment about zero), variance (σ_x^2, second moment about the mean), standard deviation (σ_x, square root of the variance), or other higher moments of a distribution; and
- Multiple variables (Figure 1d), in terms of (a) joint pmf/pdf, which expresses the probability or relative likelihood of each possible combination of discrete or continuous variable states actually occurring, (b) marginal pmf/pdf, which expresses the probability or relative likelihood of each possible state of one variable actually occurring, regardless of the state of the other variable, (c) conditional pmf/pdf, which expresses the probability or relative likelihood of each possible state of one variable actually occurring given the state of another variable and (d) covariance function/correlation coefficient, which expresses the relationship of the state of one variable to the state of another variable (including spatial and temporal correlation).

INDIVIDUAL PROBABILITY ASSESSMENT TECHNIQUES

Potential problems associated with a single individual subjectively developing probability assessments include the following ones (*2*):

1. Poor quantification of uncertainty: The assessor might not express uncertainty in a self-consistent or proper fashion. It has been shown (*4*) that people not trained in probabilistic analysis typically have problems in quantifying their uncertainty accurately. For example, if someone expresses a 90 percent probability or level of confidence that something will happen, it should happen nine out of 10 times on the average. Typically when verified, however, the event happens much less than nine (more like five) out of 10 times.
2. Poor problem definition: The parameter where the value is to be assessed might have been ambiguously defined so that the basis of the assessment might not be correct. For example, in assessing the value for shear strength within a specific slope, with respect to slope stability, the scale (large-scale average over the entire failure surface versus small-scale laboratory values) might not have been specified.
3. Unspecified assumptions: The assessor might not specify (or even be aware of) the assumptions that underlie this assessment so that the conditional nature of the assessment might not be apparent. For example, the assessor may have assumed circular arc failure through the rock mass rather than considering wedge failures along intersecting fractures in assessing the value for shear strength within a specific slope with respect to slope stability.
4. Uncorrected biases: The assessor might not specify (or even be aware of) biases that underlie the assessment so that the assessment does not accurately reflect the assessor's knowledge. Biases fall within various categories:

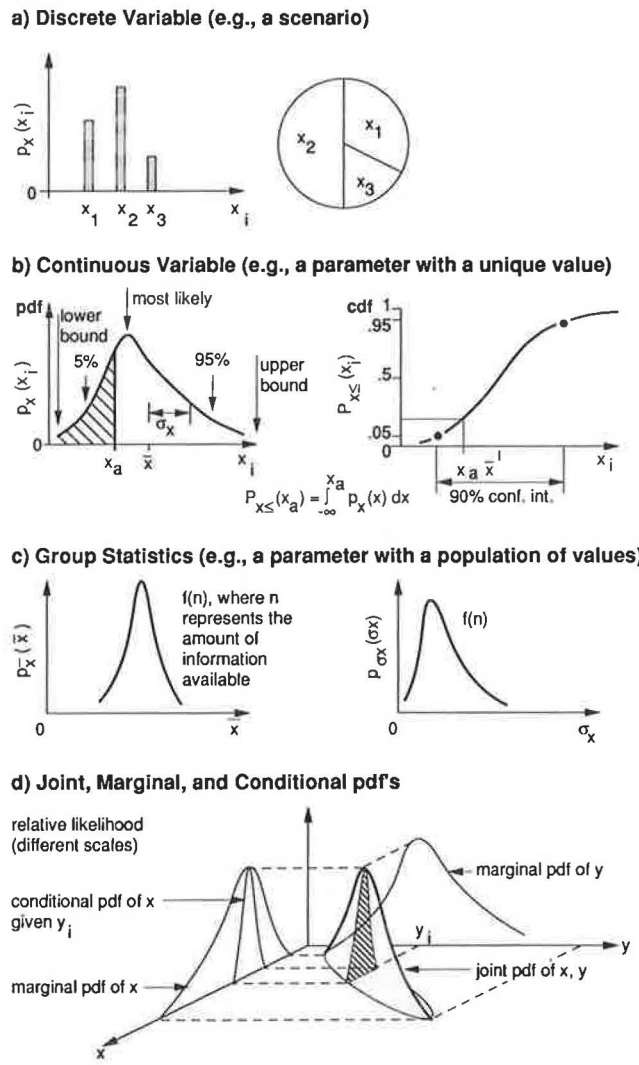

FIGURE 1 Probability distributions.

a. "Motivational," where the assessor's statements and conscious beliefs are inconsistent. Motivational biases, in turn, can be categorized as follows:
   (1) "Management" bias refers to the assessor's possible view of an uncertain variable (e.g., as an objective rather than an uncertainty). For example, if the objective is to achieve a high factor of safety with respect to slope stability, then the shear strengths may be overstated.
   (2) "Expert" bias refers to a possible reaction that the assessor may have to being considered as an expert. The assessor may feel that experts are expected to be certain of things. This bias tends to promote central bias (i.e., a tendency for the assessor to understate uncertainty). For example, the assessor may understate the range in the shear strength within a specific slope.
   (3) "Conflict" bias refers to a reward structure that might encourage the assessor to bias the estimates high or low. For example, an unethical assessor might overstate the value of a significant param-

eter (e.g., shear strengths) if it were personally beneficial (e.g., to make a project appear feasible).

(4) "Conservative" bias refers to the assessor's desire to err on the safe side. For example, if an event has an adverse impact, then the assessor may want to avoid underestimating the probability of that event (e.g., by consciously overstating its probability), thereby bounding the assessment rather than truthfully estimating it.

b. "Cognitive," where the assessor's conscious beliefs do not reflect the available information. Cognitive biases, in turn, can be categorized as follows:

(1) "Anchoring" refers to the tendency of individuals to produce estimates by starting with an initial value (suggested perhaps by the formulation of the problem) and then by adjusting the initial value to yield the final answer. The adjustment is typically insufficient. For example, the assessor might estimate the most likely value first and then the range in possible values where this estimated range would probably be larger if assessed first.

(2) "Availability" (or incompleteness) bias refers to the fact that if it is easy to recall instances of an event's occurrence (e.g., the event had some personal significance to the subject) then that event tends to be incorrectly assigned a higher probability. For example, if the assessor had been involved previously with a slope failure, then the resulting assessment of the shear strength would tend to be lower than without this experience.

(3) "Base rate" bias (or lack of moderation, law of small numbers) refers to the tendency of the assessor to focus only on specific information. Empirical evidence indicates that assessors often tend to attach less importance to general information. For example, if the specific information is some recent data (e.g., the results of recent field tests), then the importance of that information might be overrated in the assessor's mind.

(4) "Coherence and conjunctive distortions" refer to the tendency of an assessor to not account for and combine all of the components of a problem properly. For example, in assessing slope stability where various parameters (e.g., shear strength, pore pressures, ground accelerations) must all be within specific bounds for the slope to be stable, people seem especially prone to overestimate the probability that the slope will be stable.

(5) "Representativeness" refers to the tendency of an assessor to treat all information equally even though it may not be statistically representative. For example, intact rock (with high shear strengths) may be more easily sampled than highly weathered rock (with low shear strengths) so that there is a larger percentage of high shear strengths in the data base than there is in reality. If this sampling bias were not recognized, then the value of the parameter might be overestimated.

(6) "Overconfidence" refers to the tendency of an assessor to underestimate the uncertainty about the value of a parameter. For example, the assessor might not recognize and properly account for other possible values of the parameter.

5. Imprecision: The assessor may be indifferent over a specific range of values so that some "fuzziness" exists in the assessments. For example, an assessment of 20 to 30 percent probability that something will happen should be able to be refined further with additional consideration.

6. Lack of credibility: If the assessor cannot be considered an expert in the technical field, then the assessment (regardless of the other limitations) may lack credibility. Such an assessment would not be defensible to other experts or to the public. For example, a recent graduate engineer with little experience should not be making critical assessments alone.

As is summarized in Table 1, the techniques available for eliminating or mitigating the potential problems associated with developing individual subjective probability assessments include the following ones (*2*):

1. Self-assessment: The simplest approach to developing an individual subjective probability assessment is "self-assessment" (*5,6*) where the analyst interprets the available information and then quantifies an assessment of the likely value and its uncertainty. The rationale behind the assessment should be well documented, including a description of the available information and an evaluation of that information, to enhance defensibility of such subjective probability assessments. Although attractive because of its obvious simplicity, this method has significant limitations: (a) poor quantification of uncertainty; (b) uncorrected biases or unspecified assumptions or both, possibly in spite of documentation; (c) imprecision; and (d) lack of credibility if the analyst cannot be considered an expert in the technical field.

2. Informal solicitation of expert opinion: One of the most common methods of developing an individual subjective probability assessment consists of "informal solicitation of expert opinion" (*7,8*) where the analyst asks an "expert" to interpret the available information and quantify an assessment of the likely value of a parameter and its uncertainty. The defensibility of such assessments is increased over self-assessment techniques owing primarily to the increased credibility of the expert involved. As for self-assessment, the expert's rationale for the assessment should be well documented, including a description of the information available to the expert and the expert's evaluation of that information, to enhance defensibility of subjective assessments further. Although generally an improvement over self-assessment techniques, owing to increased credibility, informal solicitation of expert opinion has similar significant limitations and increased cost and potentially poor problem definition.

3. Calibrated assessment: A systematic approach to developing an individual subjective probability assessment is through the use of "calibrated assessments" (*9–11*) where the assessor's biases are identified and calibrated and the assessments are adjusted to correct for such biases. Hence, two sets of assessments are required: (a) the assessor's assessment (e.g., through the informal solicitation of expert opinion) and (b) an assessment of the assessor's biases. The assessment of the assessor's biases can be done either subjectively by peers (i.e.,

TABLE 1 EVALUATION OF SUBJECTIVE ASSESSMENT TECHNIQUES

TECHNIQUE	POTENTIAL PROBLEMS: Poor Quantification of Uncertainty	Poor Problem Definition	Uncorrected Biases/ Unspecified Assumptions	Imprecision	Lack of Credibility	Group Dynamics	Expense
INDIVIDUAL							
Self Assessment	●	○	●	●	●	NA	○
Informal Solicitation of Expert Opinion	●	●	●	●	◑	NA	◑
Calibrated Assessment	◑	●	◑	●	◑	NA	◑
Probability Encoding	○	○	○	◑	◑	NA	◑
GROUP (BEHAVIORAL)							
Open Forum	●	◑	●	●	◑	●	●
Delphi Panel	●	◑	◑	●	◑	○	●
Group Probability Encoding	○	○	○	◑	◑	◑	●
Formal Group Evaluation	○	○	○	◑	◑	◑	●

● Technique does not significantly mitigate potential problem

◑ Technique partially mitigates potential problem

○ Technique effectively mitigates potential problem

in the same way as other subjective assessments) or objectively through a set of experiments or questionnaires. The objective approach typically consists of asking the assessor a series of questions for which the true answer is available but unknown to the assessor. For example, the assessor may be given a set of relevant data that do not include the direct measurement of the parameter of interest even though such a measurement exists, the assessor estimates the parameter value based on the available data, the assessor's estimate is compared with the true value as given by the measurement, and a correction or calibration factor is determined for the assessor that when applied to the assessor's estimate results in the true value. In this way, the assessor's identified biases can be corrected. Although a general improvement over self-assessment or informal solicitation of expert opinion techniques exists, owing to the mitigation of some biases, calibrated assessments entail similar significant limitations (even after calibration) and increased costs and inherent difficulties in objectively determining calibration factors for many of the parameters of interest, because direct measurements might never be available for verification and the calibration factor may not be constant in any case.

4. Probability encoding: The most systematic and defensible approach to developing individual subjective probability assessments, but also the most expensive, is "probability encoding" (*1,12–15*). In probability encoding, analysts trained in probability theory elicit in a proper and self-consistent manner a technical expert's assessment of the pdf of a parameter value, which expresses that expert's uncertainty in the value in quantified terms. This is done in a formalized way (*2*) in five stages (i.e., motivating, structuring, conditioning, encoding, and verifying), during which the analyst attempts to (a) train the subject to properly quantify uncertainty, (b) identify and minimize the subject's bias tendencies, (c) define (and document) the item to be assessed in an unambiguous manner, (d) elicit and document the subject's rationale (including the available information) for assessment, (e) elicit (directly or indirectly) and document the subject's quantitative assessment of uncertainty and check for self-consistency, and (f) verify the assessment with the subject and repeating the process if necessary. As is illustrated by the example given in Figure 2, the subject's quantitative assessment of uncertainty can be elicited indirectly by determining the probability of various states through comparison with familiar reference events (e.g., poker hands) or by choosing between two lotteries (e.g., probability wheel or intervals, Figure 2a and b) until indifference is achieved. A cdf can then be defined, consistent with the various assessments (Figure 2c). Although a general improvement over other available methods (owing to mitigation of most of the potential problems), some imprecision may remain and probability encoding is relatively costly because it is labor intensive.

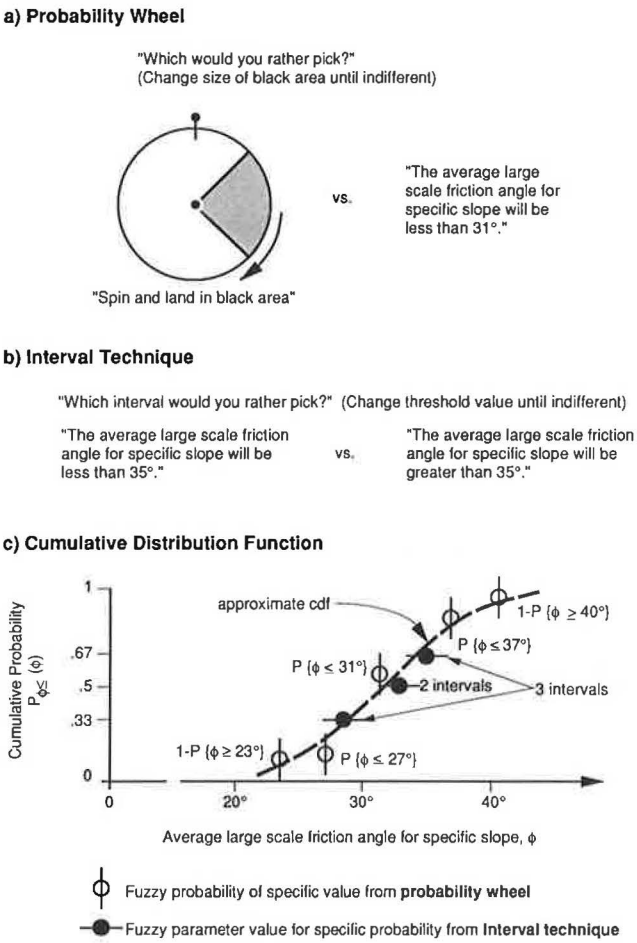

FIGURE 2 Probability encoding examples.

CONSENSUS PROBABILITY ASSESSMENT TECHNIQUES

Differences may exist in the assessment of individuals comprising a group that may arise from a number of sources (*2*):

- Disagreement on the assumptions or definitions that underlie assessments: Individual assessments are based on specific assumptions and definitions. If those assumptions or definitions differ between individuals, then the individual assessments may differ. For example, one individual may have ruled out a specific case that another individual assumes likely or one individual may have defined the parameter of interest at a different scale than did another individual.
- Failure to overcome assessment errors and biases: In conducting the individual assessments a key objective is to eliminate anchoring, availability, overconfidence, and other common distortions. Training individuals and then allowing them to practice making probability judgments prior to the individual assessment help to overcome biases, but such errors may persist. For example, overconfidence may have been mitigated to a large extent in an individual's assessment but not in another's so that although the means of their probability distributions may be similar the variances may be significantly different.
- Judgments based on differing information sources: Both specific data and general knowledge are relevant to the encoding process. Such knowledge varies even among highly specialized experts. Specific information may vary in quantity and quality and general information may vary owing to differences in training and experience. For example, an individual may have based an assessment on a specific data set in conjunction with general experience and another individual may have used a different specific data set in conjunction with different general experience.
- Disagreement on how to interpret available information: The available information must be interpreted by the individuals. For example, individuals may, in this interpretation, disagree on the methods used to obtain data, the relevance of such data to the quantity being assessed, or on the appropriateness of a particular theory or model. For example, individuals may disagree on how to interpret triaxial test results and their validity to assessing such large-scale shear strengths in assessing shear strengths within a slope with respect to slope stability.
- Different opinions or beliefs about the quantity of concern: Even after agreeing on the basis for the assessment, the information available, and how to interpret this information, individuals may still have a difference of opinion. For example, individuals may arrive at different pdfs for shear strength even after agreeing on all the preliminary aspects.

It is typically desirable to attempt to resolve those differences of opinion with the following possible outcomes:

- Convergence: A single assessment is determined that expresses the common belief of all individuals in the group as expressly agreed to by the group members.
- Consensus: A single assessment is determined, although the assessment may not reflect the beliefs of each individual. The consensus assessment may be derived from the individual assessments without the express agreement of the individuals (forced) or it may be expressly agreed to by the group for a particular purpose (agreed).
- Disagreement: Multiple assessments are determined where convergence or consensus on a single assessment is not possible (e.g., owing to major differences of opinion).

In general, convergence is generally desirable, as it is defensible, but may be difficult to achieve. Agreed consensus (i.e., with the concurrence of the group) is slightly less defensible but also less difficult to achieve. Forced consensus, without concurrence of the group, may be difficult to defend but is very simple. Disagreement may be difficult to use, because it is nonunique but is defensible.

Techniques available for resolving differences of opinion among a group of individual assessors can be categorized in terms of "mechanical aggregation" and "behavioral procedures."

Mechanical aggregation of individual assessments is a relatively simple approach to achieve at least forced consensus and involves applying a mathematical formula or procedure to combine the various individual probability distributions (*16,17*). If the individuals in the group agree to the resulting distribution, then agreed consensus (and possibly convergence) can be achieved. In general, mechanical aggregation

techniques are useful when the means, rather than the variances, of the individual probability distributions differ. Also, mechanical aggregation techniques can be used when a single distribution is required, but the parameter in question is not significant enough to warrant large amounts of effort to achieve convergence or agreed consensus.

The various forms of mechanical aggregation include the following ones (*2*):

- Averaging, the simplest mechanical aggregation technique, involves simply averaging the individuals' probabilities for each possible value. Several empirical studies (*18,19*) have indicated that averaged probabilities are often superior to individual assessments. As an example, if an individual assessed an 80 percent probability of the average large-scale friction angle for a cohesionless soil within a slope as being less than 35 degrees and the other individual in the group assessed a 60 percent probability, then the group average would be a 70 percent probability.
- Group statistics, a somewhat more rigorous treatment, involves determining the group's distribution of opinions regarding the probability for each value, thereby developing a "fuzzy" assessment or an assessment that corresponds to a given level of conservatism for the group. More complete statistical methods are available that incorporate dependence among variables and experts (*9*) and least squares or partitioning methods (*20*). As an example, the statistics of the group member opinions of the probability of the friction angle as being less than 35 degrees could be determined and used.
- Weighting methods, elaborations on either averaging or group statistics, involve the weighting of individual assessments by an external procedure to incorporate biases or differing levels of expertise among the individual assessors, similar to individual calibrated assessments. There are essentially two weighting procedures: (a) "calibration exercise," where the natural biases and tendencies of the individual assessors are evaluated and mitigated through the administration of a series of general questions to determine each assessor's ability to make correct assessments and the determination and application of weighting factors for each assessor to reflect that assessor's ability (relative to the other assessors) to make correct assessments, and (b) "peer ratings," where each of the individual assessor's relative ability to make correct assessments is assessed subjectively by peers, although such a subjective assessment may itself introduce additional biases.

Behavioral procedures can be used to attempt to develop convergence or at least agreed consensus and involve interaction among the individuals in the group and allows for the explicit identification and resolution of differences of opinion. Although evidence exists that such interaction results in better assessments (*21,22*) and that the results are generally more defensible because the group agrees on a given distribution, behavioral procedures tend to entail significantly more effort because the various individual assessors must be involved. Such behavioral procedures are necessary when at least agreed consensus (or disagreement) is required (i.e., for significant parameters) and are especially useful when the differences between the individual assessments are large.

As is summarized in Table 1, the various forms of behavioral procedures include the following ones (*2*):

- Open forum is a very informal means of achieving consensus and does not require prior individual assessments. The group attempts to achieve convergence or agreed consensus by open discussion of whatever each individual deems important to resolving the problem. However, the result can be distorted by the dynamics of the group, such as domination by an individual because of status or personality (*23*). For example, the persuasiveness of a vocal individual or the desire of some individuals to avoid dissension may distort the results. Other potential limitations to this method are the same as for the development of individual assessments through the informal solicitation of expert opinion (i.e., poor quantification of uncertainty, uncorrected biases, unspecified assumptions, and poor problem definition). The method is also limited by the credibility of the group members.
- Delphi panel is a systematic and iterative approach to achieve consensus and has been shown to generally produce results reasonably reproducible across independent groups (*22,24–27*). Each individual in a well-defined group is provided with the same set of background information and is asked to conduct and document (in writing) a self-assessment. Those assessments are then provided anonymously to each of the other assessors, who are encouraged to adjust their assessments in light of the peer assessments. Typically, the individual assessments tend to converge. Such iterations are continued until either consensus is achieved or the results stabilize otherwise (i.e., disagreement). Because the Delphi technique maintains anonymity and independence of thought through physical separation of the panelists, it precludes the possibility that any one member of the panel may unduly influence any other owing to actual or perceived personality dominance. Otherwise, it tends to have limitations similar to those for open forum.
- Group probability encoding is a formal process where a single probability distribution is assessed directly from a group of individuals, such as for the development of individual assessments by probability encoding (*1*). However, this requires the group to reach agreement on each question posed during the encoding process and would be a difficult and tiresome procedure. As is for the open forum, face-to-face interaction among participants can create destructive pressures within the group and distort the results.
- Formal group evaluation is a formal process of resolving differences between previously developed individual assessments (*1*). This process is similar to probability encoding in that it is a joint undertaking between a trained analyst and, in this case, a group that has completed individual assessments. It consists of six steps [motivating, identifying differences in the individual assessments, discussing the basis for each individual assessment, discussing information sources and interpretations, re-encoding (if warranted), and reconciling differences] where the analyst fulfills an essential role in questioning and probing the group, helping the group understand the differences, and guiding the group through the resolution process, often conducting group reassessments. This sharing of knowledge tends to produce a commonality (i.e., in definitions, assumptions, information bases, and interpretations) that is a key step in reducing the differences between individual assessments. As is for open forums, face-to-face interaction among participants can create destructive pressures within the group and distort the results. However,

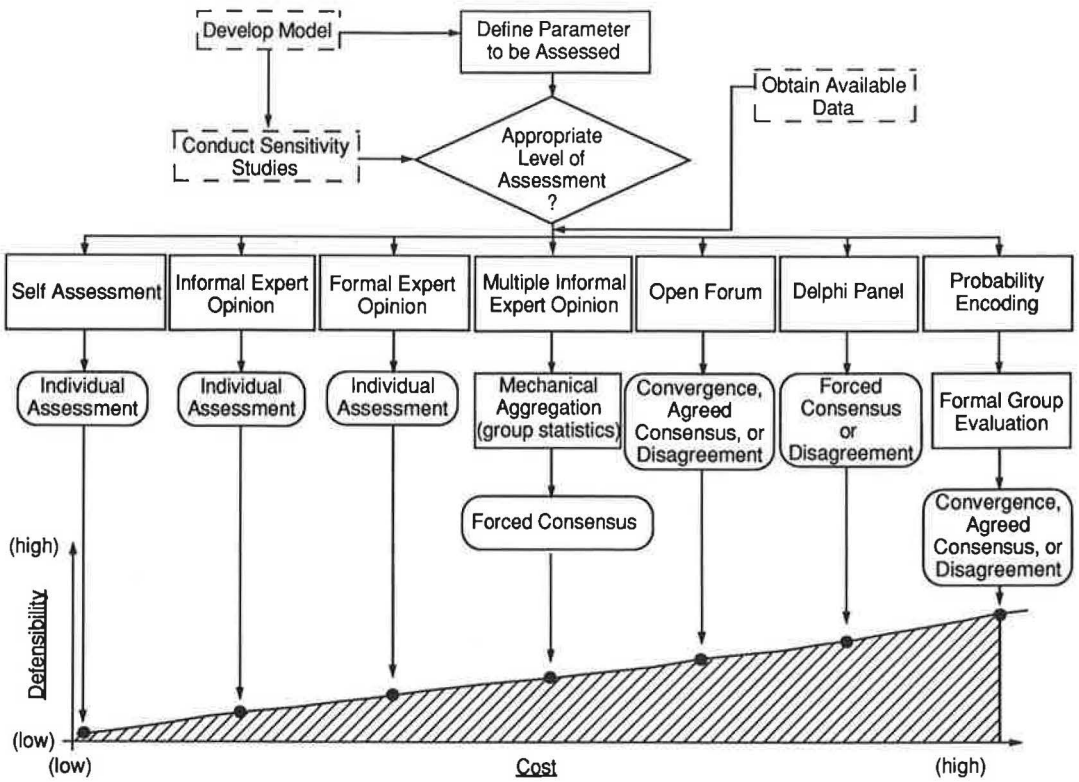

FIGURE 3 Procedure for selecting technique.

the analyst can be alert to such pressures and mitigate the effects to a large extent.

RECOMMENDED PROCEDURES

As is summarized in Figure 3, the recommended procedure for selecting the appropriate subjective probability assessment technique consists of the following steps (*2*):

1. Before conducting subjective probability assessments, (a) develop the model(s) for the system of interest (for example, a model would be needed to determine the factor of safety with respect to slope instability), (b) conduct sensitivity studies to determine the relative significance of each of the model parameters (for example, sensitivity studies on the model might indicate that slope instability is very sensitive to the average large-scale shear strength within the slope), and (c) obtain the available data regarding the various parameters, where the relative significance of the parameter will determine the appropriate level of effort in gathering data (for example, the data on shear strength might be limited to inference from measured physical properties and from generic information).

2. Each parameter to be assessed, on the basis of the model, must be defined unambiguously (considering temporal and spatial variability and conditional factors such as scale). Also, it may be useful to decompose a parameter into more elemental variables for assessment. For example, shear strength might be defined as being large scale (i.e., averaged over tens of meters), recognizing that the value may vary spatially within one geologic unit and within the time frame of interest (i.e., tens of years). Shear strength could be defined separately for the rock mass (e.g., for circular arc analyses) or for joints (e.g., for wedge analyses). Shear strength could be decomposed into cohesion and friction angle.

3. The appropriate level of assessment must be determined on the basis of relative significance of each parameter to be assessed (from sensitivity studies). For example, if a parameter is relatively insignificant (e.g., density), then a low level assessment (with corresponding low costs and low defensibility) will be appropriate. However, if a parameter is relatively significant (e.g., shear strength), a high level assessment (with corresponding high costs and high defensibility) will be appropriate. For cost efficiency, high level assessments should only be used for the most significant parameters where high defensibility is required, thus justifying their high costs.

4. The most cost-effective assessment technique is chosen (Figure 3) on the basis of necessary level of assessment and, in conjunction with the data base, implemented for each parameter [e.g., by using specific procedures (*2*)].

SUMMARY AND CONCLUSIONS

Subjective probability assessments must often be made (e.g., to predict performance accurately or to make decisions among alternatives or both) wherever the data are not statistically sufficient to make objective assessments. Such subjective probability assessments must be defensible enough to resolve potential controversies adequately. The required defensibility of such assessments is proportional to the significance of each

parameter that is assessed (e.g., as determined by sensitivity studies).

Potential problems have been identified that are associated with developing individual subjective probability assessments and with developing consensus subjective probability assessments among a group that if uncorrected can affect defensibility of the results. The available techniques for addressing those potential problems, with varying success and effort, have been presented. Procedures for cost-effectively conducting appropriately defensible subjective probability assessments have been developed.

ACKNOWLEDGMENT

The author is indebted to L. Merkhofer and R. Schwartz of Applied Decision Analysis, Menlo Park, Calif., for their significant contributions on past projects, from which some of the concepts presented in this paper have evolved.

REFERENCES

1. Golder Associates, *Groundwater and Methane Inflow Study for the BWIP ESF at Hanford, Washington*. SD-BWI-DER-001, Rev. 0, in conjunction with Applied Decision Analysis, Inc., Basalt Waste Isolation Project, Westinghouse Hanford Company, Richland, Wash., 1987.
2. W. J. Roberds. *Manual for Conducting Subjective Probability Assessments*. Draft report, Golder Associates Inc., Redmond, Wash., 1988.
3. A. Ang and W. H. Tang. *Probability Concepts in Engineering Planning and Design, 2, Decision, Risk, and Reliability*. John Wiley, New York, 1984.
4. E. C. Capen. The Difficulty of Assessing Uncertainty. *Journal of Petroleum Technology*, Vol. 28, 1976, pp. 843–849.
5. I. J. Good. *The Estimation of Probabilities*. Research Monograph 30, The M.I.T. Press, Cambridge, Mass., 1965.
6. C.-A. S. von Holstein. *Assessment and Evaluation of Subjective Probability Distributions*. Economic Research Institute, Stockholm, Sweden, 1970.
7. M. G. Morgan, M. Hennion, and S. C. Morris. *Expert Judgments for Policy Analysis*. BNL 51358, Brookhaven National Laboratory, Upton, N.Y., 1979.
8. D. L. Bernreuter. *Seismic Hazard Analysis: Solicitation of Expert Opinion*. NUREG/CR-1582, TERA Corp., Berkeley, Calif., 1980.
9. C. E. Agnew. Multiple Probability Assessments by Dependent Experts. *Journal of the American Statistical Society*, Vol. 80, No. 390, 1985, pp. 343–347.
10. A. P. Dawid. The Well Calibrated Bayesian. *Journal of the American Statistical Association*, Vol. 77, No. 379, Sept. 1982, pp. 605–613.
11. R. L. Winkler. Scoring Rules and the Evaluation of Probability Assessors. *Journal of the American Statistical Association*, Vol. 64, 1969, pp. 1071–1078.
12. C. S. Spetzler and C.-A. S. von Holstein. Probability Encoding in Decision Analysis. In *Example of a Probability Encoding Technique*. K-258802, Stanford Research Institute, ORSA-TIMS-AIEEE Joint National Meeting, Atlantic City, N.J., 1972.
13. R. M. Zamora. *ONR Probability Encoding Task*. Working paper, SRI Project 4030, Office of Naval Research, Operational Decision Aids Project, Stanford Research Institute, Menlo Park, Calif., 1975.
14. C.-A. S. von Holstein and J. E. Matheson. *A Manual for Encoding Probability Distributions*. Final Report, SRI International, Menlo Park, Calif., 1979.
15. M. W. Merkhofer and P. McNamee. *The SRI Probability Encoding Process: Experiences and Insights*. Technical report, SRI International, Menlo Park, Calif., 1982.
16. A. J. Harman and S. J. Press. *Collecting and Analyzing Expert Group Judgement Data*. The Rand Corporation, Santa Monica, Calif., 1975.
17. H. A. Ashton and R. H. Ashton. Aggregating Subjective Forecasts: Some Empirical Results. *Management Science*, Vol. 31, No. 12, 1985, pp. 1499–1508.
18. C.-A. S. von Holstein. Probabilistic Forecasting: An Experiment Related to the Stock Market. *Organizational Behavior and Human Performance*, Vol. 8, 1972, pp. 139–158.
19. R. V. Brown. *An Experiment in Probabilistic Forecasting*. R-944-ARPA, The Rand Corporation, Santa Monica, Calif., 1973.
20. D. V. Lindley, A. Tversky, and R. V. Brown. On the Reconciliation of Probability Assessments. *Journal of the Royal Statistical Society A*, Vol. 142, 1979, pp. 146–180.
21. D. A. Seaver. *Assessment of Group Preferences and Group Uncertainty for Decision Making*. SSRI 76-4, Social Science Research Institute, University of Southern California, Los Angeles, 1976.
22. J. Rohrbaugh. Improving the Quality of Group Judgement: Social Judgment Analysis and the Delphi Technique. *Organizational Behavior and Human Performance*, Vol. 24, 1979, pp. 73–92.
23. A. Delbecq, A. Van de Ven, and D. Gustafson. *Group Techniques for Program Planning*, Scott-Foresman, Glenview, Ill., 1975.
24. N. Dalkey and O. Helmer. *An Experimental Application of the Delphi Method to the Use of Experts*. AD A081 351, Memorandum RM-727-PR (Abridged), The Rand Corporation, Santa Monica, Calif., 1962.
25. O. Helmer. *The Delphi Method for Systematizing Judgments About the Future*. MR-61, University of California at Los Angeles, Los Angeles, 1966.
26. O. Helmer. *Systematic Use of Expert Opinions*. Technical paper, The Rand Corporation, Santa Monica, Calif., 1967.
27. H. A. Linstone and M. Truoff, eds., *The Delphi Method: Techniques and Applications*, Addison-Wesley, Reading, Mass., 1975.

Publication of this paper sponsored by Committee on Exploration and Classification of Earth Materials.