Performance Assessments of Nuclear Waste Repositories: A Dialogue on Their Value and Limitations

Rodney C. Ewing,[1] Martin S. Tierney,[2] Leonard F. Konikow,[3] and Rob P. Rechard[2]

Performance Assessment (PA) is the use of mathematical models to simulate the long-term behavior of engineered and geologic barriers in a nuclear waste repository; methods of uncertainty analysis are used to assess effects of parametric and conceptual uncertainties associated with the model system upon the uncertainty in outcomes of the simulation. PA is required by the U.S. Environmental Protection Agency as part of its certification process for geologic repositories for nuclear waste. This paper is a dialogue to explore the value and limitations of PA. Two "skeptics" acknowledge the utility of PA in organizing the scientific investigations that are necessary for confident siting and licensing of a repository; however, they maintain that the PA process, at least as it is currently implemented, is an essentially unscientific process with shortcomings that may provide results of limited use in evaluating *actual* effects on public health and safety. Conceptual uncertainties in a PA analysis can be so great that results can be confidently applied only over short time ranges, the antithesis of the purpose behind long-term, geologic disposal. Two "proponents" of PA agree that performance assessment is unscientific, but only in the sense that PA is an engineering analysis that uses existing scientific knowledge to support public policy decisions, rather than an investigation intended to increase fundamental knowledge of nature; PA has different goals and constraints than a typical scientific study. The "proponents" describe an ideal, six-step process for conducting generalized PA, here called probabilistic systems analysis (PSA); they note that virtually all scientific content of a PA is introduced during the model-building steps of a PSA; they contend that a PA based on simple but scientifically acceptable mathematical models can provide useful and objective input to regulatory decision makers. The value of the results of any PA must lie between these two views and will depend on the level of knowledge of the site, the degree to which models capture actual physical and chemical processes, the time over which extrapolations are made, and the proper evaluation of health risks attending implementation of the repository. The challenge is in evaluating whether the quality of the PA matches the needs of decision makers charged with protecting the health and safety of the public.

[1] Nuclear Engineering and Radiological Sciences, University of Michigan, Ann Arbor, Michigan 48109-2104.
[2] Sandia National Laboratories, P.O. Box 5800, Albuquerque, New Mexico 87185-1328.
[3] U.S. Geological Survey, 431 National Center, Reston, Virginia 20192.

1. INTRODUCTION

Probabilistic systems analysis, exemplified by a procedure called performance assessment (PA), has become the standard regulatory approach in analyzing and evaluating the long-term performance of nuclear waste repositories in the United States. Uncertainties in site characterization, the extrapolation of models over time, and the complexity of the analysis have led to concerns about the usefulness and limitations of PA in evaluating actual health and safety impacts.

This paper examines and discusses some important issues that have persisted for over 20 years related to the use of performance assessment of nuclear waste repositories. We have cast this paper in the form of a dialogue in order to provide the reader with different views of these issues. The intent is not to win an argument, but rather to find common ground and carefully identify issues that require further attention.

In the analysis of complex systems, PA provides a systematic and disciplined means of evaluating the extent to which one reaches an endpoint (e.g., compliance with a regulation). Once the PA has been implemented, its results can then be used to (1) identify major sources of uncertainty and deficiencies in understanding, (2) focus resources on the scientific and technical issues that have the most effect on performance, and (3) evaluate alternative designs and materials. Finally, to the extent that the analysis is "traceable" and "transparent," PA can be a powerful tool in communication with the public and can provide a valuable, largely objective input to public policy decisions. Although performance assessment is required as a condition for licensing certain waste-disposal sites, there is no a priori reason that the public or scientific community should uncritically accept a PA analysis; on the contrary, any PA analysis should be carefully scrutinized and improvements will inevitably result.

Two "proponents" (M.S.T. and R.P.R.) describe PA, its history, and its strengths and limitations (Section 2). Two "skeptics" (R.E. and L.K.) describe their concerns and try to illustrate the limitations of PA when applied to geologic systems (Section 3). The "dialogue" is a series of questions about PA methodology that a knowledgeable person might ask, followed by answers provided by the PA proponents (Section 4). Not every important question has been raised (e.g., Is there a better way?), nor can all questions be fully answered within the limitations of this

paper. We hope that this discussion is useful to the scientific and engineering communities that are involved in conducting PAs, as well as to the public and government institutions that use PA results to set public policy. The opinions expressed in this paper are solely those of the individual authors and in no way reflect the policies or official positions of the institutions with which they are affiliated.

2. WHAT IS PERFORMANCE ASSESSMENT?

2.1. Brief History of PA in the United States

An in-depth history of PA and its relation to other types of probabilistic system studies is given in another paper in this issue of Risk Analysis (Rechard, "Historical Relationship Between Performance Assessment and Other Types of Risk Assessment").

Probabilistic risk analysis and geological waste disposal systems were first brought together in 1976 at the Bishop's Lodge Workshop on Geologic Data Requirements for Radioactive Waste Management Assessment Models.[1] Participants in the workshop included team leaders and scientists from waste management programs sponsored by the U.S. Energy Research and Development Administration (ERDA), the U.S. Nuclear Regulatory Commission (NRC), and the U.S. Environmental Protection Agency (EPA). The points of view were equally diverse: NRC team members, mainly nuclear engineers and mathematicians, advocated methods developed for the safety analysis of engineered systems, in particular, the methods employed in the "Rasmussen Study" of the safety of nuclear power plants,[2] as tools for assessing the long-term safety of geologic waste disposal. ERDA team members, mainly geoscientists and geotechnical engineers, were skeptical of the use of mathematical models for assessing risks of disposal of radioactive wastes in a geologic repository; they suggested a qualitative assessment of the safety of proposed waste repositories based on systematic identification of those natural and anthropogenic "features, events, and processes" (FEPs) that could cause or contribute significantly to failure of the repository to meet design or regulatory standards. EPA team members generally supported the views of the ERDA representatives; later, in developing standards for a high-level radioactive waste repository (see below, Section 2.2), EPA clearly indicated that it recognized the limitations of mathematical models in a regulatory process.

Different agency teams took different paths in their work on PA methodologies after 1976. Sandia National Laboratories (SNL), funded by NRC, developed a methodology for licensing of geologic repositories for commercial high-level waste and spent fuel that included many features of modern PAs: lists of natural and anthropogenic phenomenon for the screening of FEPs, probability models of FEPs, models of transport of radionuclides with dissolved compounds in groundwater, environmental pathway and health effects models, even "simulation analysis applied to questions of geological stability."[3] But the ERDA team at Pacific Northwest Laboratories (PNL) was first to propose simulation as a method for identifying and evaluating potential loss-of-isolation scenarios, and also first to remark on the unsuitability of fault-tree methods for quantifying risks attending geologic waste disposal systems.[4] Scientists from the U.S. Geological Survey also urged caution in the application of mathematical models to predicting long-term behavior of geologic waste repositories.[5]

The PA methodology eventually used by U.S. agencies was developed during the decade 1978–1988. Reports on major features of that methodology, e.g., scenario selection procedures, groundwater flow and contaminant transport codes, pathways-to-man models, and sensitivity analysis techniques, are cited in Cranwell et al.[6] Use of Monte Carlo integration, in particular the Monte Carlo variant called Latin hypercube sampling (LHS),[7] to propagate uncertainty through mathematical models of phenomena associated with geologic waste disposal was first proposed in 1981.[8] In spite of these early innovations, a coherent and well-documented collection of methods for conducting a PA did not take form until 1987 when methods proposed by Cranwell et al.[6] were adopted and modified by the Performance Assessment Group for the Waste Isolation Pilot Plant (WIPP). By 1985, "performance assessment" had become the term of choice to denote those geologic waste repository studies formerly called "risk analyses" or "probabilistic safety analyses."

The WIPP Group's 1989 approach to PA, total-system simulation, is similar to methods first proposed by Bartlett et al.[4] and later by the Sandia PA Group for the Yucca Mountain Project.[9] The central feature of total-system simulation is a set of coupled computer models under the control of a single "driver" code, with each model simulating one or more of the processes or events that are judged to affect releases of waste from the repository. Another characteristic of this method is use of event trees to

identify scenarios, and probability distributions of model parameters to calculate scenario probabilities.[10,11] Some European and Canadian programs for assessing the safety and regulatory compliance of geologic waste repositories use methods similar to total-system simulation.[12,13]

Applications of total-system simulation to PAs since 1989 include three preliminary assessments for the WIPP,[14–16] total-system performance assessments (TSPAs) for the Yucca Mountain project,[17] and the 1996 Compliance Certification Application (CCA) for the WIPP.[18] Clear and relatively brief reports of the CCA work are Helton et al.[19] and the article by Helton et al., "Performance Assessment in Support of the 1996 Compliance Certification Application for the Waste Isolation Pilot Plant," in this issue of *Risk Analysis.*

2.2. The Origin of Performance Assessment in EPA Standards

The term "performance assessment" (PA) was first defined in the context of a standard promulgated by the U.S. Environmental Protection Agency[20]: 40 CFR Part 191.13, "Containment Requirements," states

(a) Disposal systems for spent nuclear fuel or high-level or transuranic radioactive wastes shall be designed to provide a reasonable expectation, based on *performance assessments* [emphasis added], that the cumulative releases of radionuclides to the accessible environment for 10,000 years after disposal from all significant processes and events that may affect the disposal system shall:
  (1) Have a likelihood of less than one chance in 10 of exceeding the quantities calculated according to Table 1 (Appendix A); and
  (2) Have a likelihood of less than one chance in 1,000 of exceeding ten times the quantities calculated according to Table 1 (Appendix A).[20]

Performance Assessment is defined in paragraph 191.12(q) to be

. . . an analysis that: (1) Identifies the processes and events that might affect the disposal system; (2) examines the effects of these processes and events on the performance of the disposal system; and (3) estimates the cumulative releases of radionuclides, considering the associated *uncertainties* [emphasis added], caused by all significant processes and events. These estimates shall be incorporated into an overall probability distribution of cumulative releases to the extent practicable.[20]

The EPA does not rely entirely upon the PA process to determine the actual safety of a site:

> (b) Performance assessments need not provide complete assurance that the requirements of 191.13(a) will be met. Because of the long time period involved and the nature of the events and the processes of interest. . . . Proof of the future performance of a disposal system is not to be had in the ordinary sense of the word in situations that deal with much shorter time frames.[20]

These "Containment Requirements" are unique for several reasons. The requirements are among the first federal standards and regulations to mandate use of numerical models of a proposed technological system in a formal determination of system acceptability. The requirements also express numerical performance standards in terms of allowable limits on the *exceedance probability* of the performance measure (here, cumulative release of radionuclides to the accessible environment over a 10,000-year period); usually, limits are placed on the magnitude of the performance measure itself. Lastly, the requirements are the first federal standard to require an explicit and quantitative treatment of uncertainty. Mathematical procedures called "The Propagation and Analysis of Uncertainty"[21] are now central features of the EPA's approach to setting performance standards for geologic waste repositories.

Numerical but nonprobabilistic performance criteria for geologic waste repositories were also set out in NRC rules 10 CFR Part 60[22]; these criteria antedated the EPA standards and were in the form of performance objectives for different parts of the waste disposal system, e.g., the geologic setting and the engineered-barrier systems. However, since EPA standard 40 CFR 191 was remanded in terms of its application to the proposed Yucca Mountain repository for high-level waste, the NRC is currently developing new regulations for this facility. These new regulations will be issued as 10 CFR Part 63.

2.3. Performance Assessment and Probabilistic Systems Analysis (PSA)

The foregoing history (Section 2.1) suggests that in the United States, methods for studying long-term performance of geologic waste disposal systems grew out of techniques employed in earlier probabilistic risk analyses (PRAs) for nuclear power plants.[2] PAs and PRAs therefore have many "probabilistic" methods in common and can be regarded as instances of a wider category of investigations that will here be called "Probabilistic Systems Analysis" (PSA; Section 2.3.1). PRAs, however, usually apply to man-made systems, whereas PAs may treat systems that are partly engineered and partly natural. The natural parts of a PA system are usually heterogeneous, physically inaccessible, and therefore hard to map and characterize in the detail required for effective study of the system via mathematical models. Natural subsystems such as rock strata or aquifers are most effectively modeled as systems with continuously distributed variables and parameters, i.e., modeled by partial differential equations, in contrast with the discrete-element, lumped-parameter, ordinary differential equation approach usually taken in safety studies of engineered systems. This continuous versus discrete model dichotomy mainly affects the way in which parametric uncertainty is treated in the two kinds of systems studies, but also may affect the practicability of computations (Section 2.3.1.2). Finally, the fact that natural features may act as "barriers" to the release of waste from a geologic repository dictates that system failure will be a process occurring over times comparable to the expected operational life of the system. In contrast, engineered systems, such as nuclear power plants or light bulbs, usually fail in time intervals that are very short compared with their expected operational lifetimes. This difference in time scales for system failure determines the kinds of mathematical tools that are most effective in analyzing the performance of the two types of systems.

2.3.1. Structure of an Ideal Probabilistic Systems Analysis

The main distinction between "ordinary" and "probabilistic" systems analyses (PSA) is that, in ordinary systems analysis, point estimates of model system behavior are used to draw conclusions about the behavior of the real system. In a probabilistic system analysis, the system is imperfectly characterized from the investigators' points of view and attempts are made to quantify uncertainties in model input parameters and propagate them through the models to obtain probability distributions that quantify uncertainty in model system outcomes. A presentation of quantitative uncertainty in model results can give investigators and project sponsors a much better idea of how closely the real system would come to meeting performance standards. The methods of PSA can be

applied to the study of any technological system that is capable of being mathematically modeled, is imperfectly characterized, and for which numerical standards of performance can be stated. PSA may also be applied to assessments of the effects of proposed social programs or legislation (e.g., land-use planning or air-quality standards). Morgan and Henrion call these kinds of applications "policy analyses."[21]

There are six essential steps in conducting the ideal probabilistic systems analysis:

1. Identify numerical measures of performance for the system.
2. Characterize the system, i.e., identify, understand, and (if necessary) make site-specific measurements of those phenomena, features, events, and processes that may affect the system's performance.
3. Using knowledge of the factors identified in Step 2, build mathematical models of the system whose dependent variables, or "output," are the performance measures identified in Step 1.
4. Quantify uncertainties in modeling relationships and model parameters by the assignment of weights or probability distributions.
5. Analyze uncertainty; i.e., propagate the uncertainties assigned in Step 4 through the system model to predict uncertainty in the performance measures.
6. Compare predicted performance measures, including measures of uncertainty developed in Step 5, with operational or regulatory criteria for the performance of the real system. Interpret results to sponsors and regulators.

Choices of quantitative measures of performance (Step 1) are guided by operational or regulatory constraints placed upon the real system. The chosen measures of performance are expressed as an M-tuplet of numbers, $\mathbf{y} = (y_1, y_2, y_3, \ldots, y_M)$, where each number y_m denotes some quantity that we should like to compare with a predetermined value that is derived from some goal or standard placed on the real system's performance. For example, there are two numerical measures of repository performance in the EPA's long-term containment standards[20]: two exceedance probabilities for cumulative releases of radionuclides to the accessible environment in a 10,000-year period, with one probability measured at unit normalized release, and another at ten times the unit normalized release (Section 2.2).

2.3.1.1. The Model-Building Process. Steps 2 and

3 taken together can be called the "model-building process." What is done during Steps 2 and 3 almost entirely determines the empirical foundations and scientific content of a PSA and consequently the degree of reliability of the knowledge about the real system that is produced by an analysis with models that mimic the real system. The ideal model-building process can be summarized as follows. A "first-order" mathematical model of the system is created and used to make testable predictions of properties associated with the system. The predictions are then compared with actual observations of associated properties, and discrepancies between model predictions and observations are noted. Discrepancies are then examined and interpreted with the purpose of finding ways in which the model can be modified to more adequately reflect empirical evidence. The model is then modified to accommodate observations, and new predictions of system behavior are made—the process just described begins all over again and is repeated until it stops when investigators judge that agreement between empirical evidence and model results is satisfactory.

The final result of the model-building process in a PA is a computational model of the total system, usually composed of many coupled submodels, that is capable of generating values of the performance measures $(y_1, y_2, y_3, \ldots, y_M)$ given values of the model's *input parameters* $\mathbf{x} = (x_1, x_2, x_3, \ldots, x_N)$. It is convenient to view the total-system model as a *function* $\mathbf{y} = \mathbf{f}(\mathbf{x})$ that maps a point in the N-dimensional parameter space Σ_N to a point $\mathbf{y}$ in the M-dimensional performance measure space Σ_M. Thinking of the total-system model as a function suggests an unconventional but revealing interpretation of the standard approach to the "probabilistic" part of a probabilistic systems analysis[23]: Observe that the vector function $\mathbf{f}(\mathbf{x})$ traces out M distinct surfaces in an $(N + 1)$-dimensional space, $y_m = f_m(\mathbf{x})$, $m = 1, 2, 3, \ldots, M$, as the vector of parameters $\mathbf{x}$ take all values in the space Σ_N. These surfaces, sometimes called "response surfaces," represent *all possible information* concerning the performance measures $\mathbf{y}$ that is contained in the total-system model.

Model building for performance assessments does not usually achieve the ideal. In recent PAs,[16,17] the total-system model is built up by stringing together submodels of widely different levels of detail and whose solutions require widely different degrees of computational power. The most complicated of these submodels can involve numerical solutions of time-dependent, nonlinear partial differential equa-

tions in two spatial dimensions; this type of model is often called a "mechanistic" or "phenomenological" model and may have been built up through several cycles of the ideal model-building process. The simplest submodels involve solutions of algebraic equations and are ad hoc constructions based on an investigator's knowledge and interpretation of the physical principles that govern the particular phenomenon being studied. Because of the long time scales and the inaccessible and heterogeneous media involved in PAs for geologic waste repositories, none of these submodels can be validated or verified in the usual sense of these terms.

All submodels in a total-system model need not have the same level of phenomenological detail. The capacity to prioritize submodels in terms of their relative contributions to total-system's performance is a strength of the total-system simulation approach to PA. Resources for conducting site characterization and laboratory studies are inevitably limited, and it is appropriate to allocate resources to experimental and analytical studies that reduce the uncertainty in, and improve scientific quality of, those submodels that, as a result of *sensitivity analyses,* appear to dominate total-system performance. Some examples: Complicated models of flow and transport in the Rustler formation of the WIPP site would not be necessary if analyses of scenarios for inadvertent exploratory drilling through the WIPP repository were not required by the EPA (Section 2.3.1.2). Similarly, models of vadose-zone percolation of water and air at the proposed Yucca Mountain repository site could be much simpler than they are if there was less uncertainty concerning long-term waste-canister degradation mechanisms in vadose-zone mines. Requirements of basic realism can usually be met for the less important submodels, or the "first cut" of the model-building process, by basing initial models on physical and chemical principles in textbooks, e.g., Freeze and Cherry[24] for groundwater flow and solute transport problems.

2.3.1.2. Uncertainty and Sensitivity Analysis. It is seldom possible to specify every model parameter precisely, particularly if the model is of some heterogeneous natural system. Usually, all that is known confidently about some parameters is that their true values lie between two numbers that are constrained only by experimental evidence or the accepted physics of the situation (e.g., the speed of every material body lies between zero and the speed of light). If there are many such imprecisely specified parameters, say N_u of them, it becomes necessary to view

system performance no longer as a point in M-dimensional space, but again as surfaces, now imbedded in an $(N_u + 1)$-dimensional space (assuming of course that $N - N_u$ of the parameters are precisely specified). Since the topography of these surfaces could vary considerably as the uncertain parameters take all allowable values in their ranges, specification of system performance on the basis of only a few point estimates could lead to erroneous conclusions about the system's behavior. But one need not search through the entire space of uncertain parameters to map the system's possible behaviors. Investigators can use empirical evidence, data from field measurements, and values from compilations in handbooks to bound the ranges on the N_u uncertain parameters, thereby limiting the extent of the M response surfaces in $(N_u + 1)$-dimensional space.

Step 4 of the ideal PSA process becomes necessary when there is uncertainty on the part of investigators concerning values to be assigned to some model parameters (a situation called "parametric uncertainty") or when empirical evidence supports more than one mathematical expression of some phenomenon in the model (in PA, a situation called "alternative conceptual models" or "model uncertainty"). We here assume that model uncertainty can be treated in the same way as parametric uncertainty: that is, by assignment of a *probability distribution* to the indices (usually 1, 2, 3, . . .) that label each alternative conceptual model. In our unconventional view of PSA, the purpose of introducing probability distributions is to limit the size of the region in N_u-dimensional space that must be studied in order to gain insight into the range of model system behavior[23]; probability distributions are chosen so as to limit the ranges of the N_u uncertain parameters to regions of Σ_N where investigators have reason to believe that true values of the parameters x will lie.

The conventional view of an uncertainty analysis (Step 5) is that of a process of quantifying uncertainty in the model's dependent variables that results from uncertainty in the model's independent variables.[21] An equally correct view of uncertainty analysis regards that subject's associated mathematical techniques as tools for statistically mapping the topography of the M response surfaces in $(N_u + 1)$-dimensional space described above.[23] In either view, computational uncertainty and sensitivity analyses always reduce to numerical calculations of some statistical property of these surfaces—the average elevation of a surface's topography, the variance in the heights of the topography—or in sensitivity analyses,

the average slopes of the response surfaces along the coordinate "axis" of some parameter. The most general statistical property of these surfaces is the joint cumulative probability distribution (CDF) of the performance measures,

$$G(\mathbf{u}) = \Pr\{Y_1 \le u_1, Y_2 \le u_2, Y_3 \le u_3, \ldots, Y_M \le u_M\}$$

where $\Pr\{A\}$ stands for the probability that statement A is true (in the example above, A is the statement "Y_1 is less than or equal to u_1, and Y_2 is less than or equal to u_2, and Y_3 is less than or equal to u_3, etc.").

Estimating $G(\mathbf{u})$ requires numerical evaluation of a multidimensional integral of the form

$$G(\vec{u}) = \int_{\Sigma N_u} \left\{ \prod_{i=1}^{M} U[u_i - f_i(\vec{x})] \right\} dF(\vec{x}) \qquad (1)$$

where $U(\ldots)$ is the unit step function [$U(s) = 1$ if $s > 0$, $U(s) = 0$ if $s \le 0$], and $F(\mathbf{x})$ is the joint probability distribution of imprecisely known model parameters that was prescribed in Step 4. Popular methods for numerically evaluating multidimensional integrals such as Equation (1) are the Monte Carlo methods, in particular, crude Monte Carlo[25] or Latin hypercube sampling (LHS).[7]

Monte Carlo methods have benefits and possibly prohibitive costs associated with their effective use. Benefits arise from the intentional exploitation of the aleatory uncertainty implicit in random number generators to infer properties of response-surface topography using the tools of classical statistics. By making independent, identically distributed estimates of the value of a multidimensional integral such as Equation (1) investigators can place numerical confidence intervals on predicted response-surface properties that are being compared with performance standards. The costs of Monte Carlo methods become apparent when limitations on the time and funds that can be devoted to a calculation preclude the taking of a sufficiently large (>10,000) number of samples in the integration process. In such a case, when small sample sizes are necessarily used with complicated, nonlinear, and long-running computer models, the resulting confidence intervals on critical performance measures may be so wide that no meaningful comparisons can be made with performance criteria or standards.

We end this section with an illustration of the way sensitivity and uncertainty analysis can be used to establish relative importance of submodels in the total-system approach to PA. The evolution of predicted performance measures for two submodels of

the WIPP total-system model over the years 1990–1996 is depicted in Fig. 1. Performance measures in this figure are represented by the complementary cumulative distribution function (CCDF) for normalized release of radionuclides to the accessible environment over a 10,000-year period (Section 2.2). Note that a CCDF is one way of displaying the outcome of an uncertainty analysis; it is simply the probability that some prescribed value of a quantity relevant to system performance (in this case, cumulative release of radionuclides) will be exceeded during a realization of the particular release scenario. Figure 1 applies to scenarios in which the WIPP repository is penetrated by exploratory drillers at least once in the 10,000-year period following closure; Fig. 1a shows CCDFs for direct releases to the surface during the drilling events (radioactive material is brought to the surface in the drill cores and cuttings); Fig. 1b shows CCDFs for indirect releases to the environment that occur by way of the groundwater pathway as a consequence of exploratory drilling. We note that past calculations have consistently shown that there is virtually zero release to the accessible environment along groundwater pathways in the absence of exploratory drilling.

It is seen that, apart from 1990 results, the estimates of maximum direct release to the surface during drilling (Fig. 1a) did not change appreciably between 1991 and 1996 and were always at least a factor of 10 larger than the releases via the groundwater pathway; these curves are stable because the model for direct release of radionuclides through exploratory drilling, and the values assigned to that model's most sensitive parameter, remained almost the same throughout the study period. On the other hand, estimates of maximum releases via the groundwater pathway (Fig. 1b) changed by many factors of 10 throughout the study period, decreasing by more than four orders of magnitude between 1991 and 1996. The reasons for these decreases lie in the incorporation in 1991, 1992, and particularly in 1996 of increasingly detailed models of flow of gas and brine, and of corrosion chemistry, within the repository. Also, better, empirically based estimates of model parameters were made during this period of time. The effect of these model improvements was to reduce drastically estimates of the amount of contaminated brine that could reach the groundwater pathways starting at the points where hypothetical boreholes penetrated the Culebra (one of the aquifer-bearing formations overlaying the WIPP repository). Thus, parallel improvements made between 1992 and 1996 in the WIPP
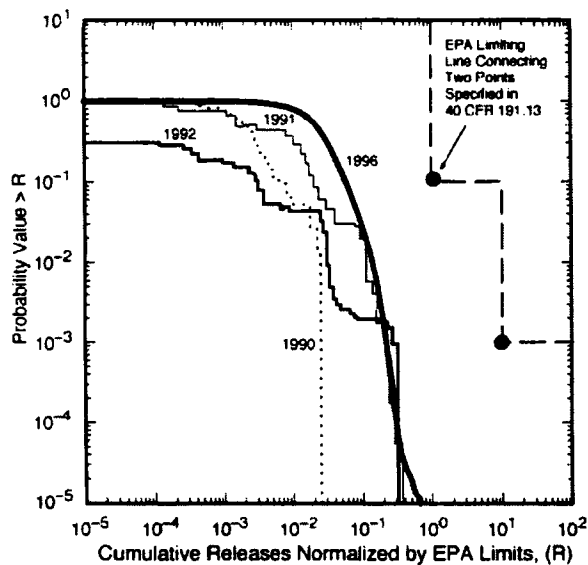
(a) Direct Releases to the Surface during Drilling
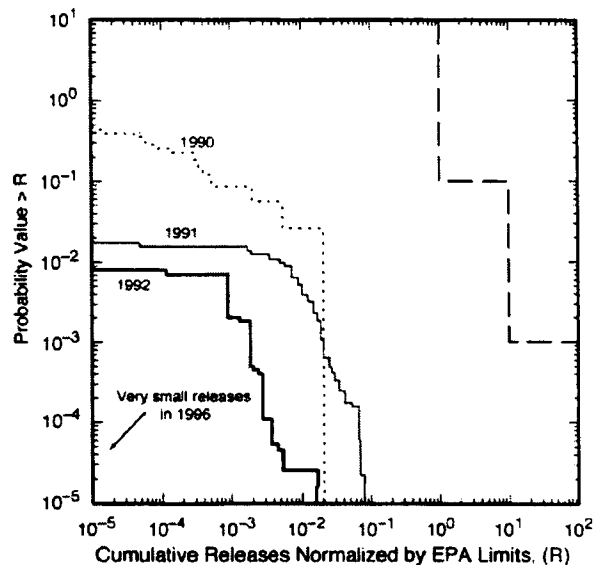
(b) Releases via Groundwater Pathway through Culebra

Fig. 1. Changes in the distribution of cumulative releases (normalized by the EPA limits) for (a) direct releases to the surface during drilling and (b) release via groundwater pathways through Culebra. See text for further explanations.

model of contaminant transport in the Culebra proved to be unnecessary, since the source term for this model turned out to be so small. See Section 3.2.2.2 for further discussion of contaminant transport models.

3. ISSUES IN PERFORMANCE ASSESSMENT

To the extent that PA is an organized and logical analysis of a complex system providing the basis for regulatory decisions, it makes sense; however, many remain skeptical of the methodology,[26] particularly the probabilistic analysis of geologic systems over extended periods. As performance and risk assessments become the basis for decisions of public policy, the limitations of these tools become important. In an assessment of the environmental movement's attitude toward risk assessment, Tal notes that, "Without enhanced scientific validity, they [environmental groups] will oppose the growing influence of risk numbers over environmental health decisions."[27] It is the issue of the "scientific validity" of probabilistic risk analyses that we examine in this section.

3.1. Philosophical Issues

One of the unique and daunting challenges of PA is that the effort is almost unprecedented in the

range of temporal and spatial scales that are modeled and finally coupled in order to provide a total system PA. This is illustrated by a plot of the characteristic space and time scales in Earth system processes[28] (Fig. 2). A typical PA includes models of molecular-scale corrosion of waste forms and canisters, complex interactions of fluids, waste forms, and rock in the near-field, transport (dilution, dispersion, and sorption) of radionuclides through the far-field, probabilistic analyses of seismic and volcanic events, and climate change effects on precipitation. Each submodel may be extremely sensitive to assumed boundary conditions, is coupled to other submodels, and the behavior of the total system is extrapolated over long periods. Output of one model becomes either input (e.g., radionuclide concentration in solution) or a boundary condition for other models (e.g., percolation rate and thermal field). Each submodel (e.g., spent fuel corrosion, climate change, fluid flow, thermal-hydrologic-chemical-mechanical interactions, dose-to-person calculations) represents a major effort and challenge within its own subdiscipline.

Much has been written, and need not be summarized here, on the scientific method, predictive capabilities of science, and need for testable hypotheses. Much less has been written on the unique aspects of a historical science (PA looking backward) such as geology. The projected performance of a geologic system over time (past or future) is the challenge of
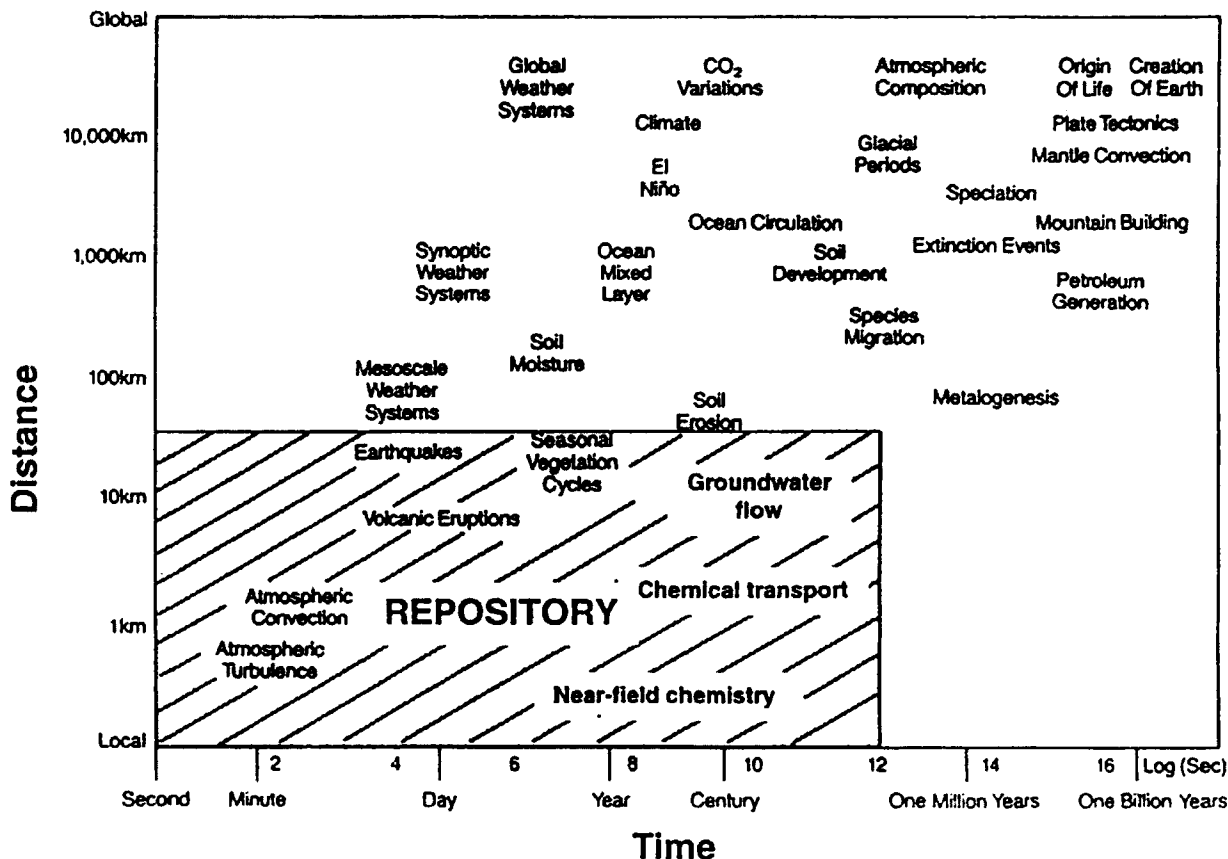
Fig. 2. Characteristic space and time scales of Earth system processes (modified from a report of the National Research Council[28]).

an evaluation of the performance of a nuclear waste repository. This paper has already summarized the history of PA and the methods of PSA, both growing out of a need to describe and analyze the performance of engineered systems, such as nuclear power plants. *The crux of this discussion is to determine the extent to which the PA analysis of engineered systems can be extended to natural, geologic systems over extended intervals of time, in some cases, geologic time scales (10^4 to 10^6 years). No question exists that mathematical models can be extrapolated over time; however, can these extrapolations actually capture the physical and chemical behavior of the geologic systems over extended time?*

Although seldom appreciated in discussions of PA, analysis of large-scale, complex systems over time is an old issue for geologists, who deal mainly with reconstructing the Earth's past history. Similar issues arise when describing the future behavior of large-scale, natural systems over time, such as the performance of a nuclear waste repository. The geo-

sciences not only can provide the required raw data for a PA, but should also contribute to the conceptual approaches used in the analysis. Early and important examples of such discussions include G. K. Gilbert's 1886 "The Inculcation of the Scientific Method by Example,"[29] T. C. Chamberlin's 1890 "The Method of Multiple Working Hypotheses,"[30] and D. Johnson's 1933 "Role of Analysis in Scientific Investigation."[31]

The value of alternative conceptual models was clearly articulated in Chamberlain's paper, and as early as 1933, J. Hoover Mackin had already contrasted the "rational" method, as it had developed in the geologic sciences, with "engineering" or "empirical" methods that were closely tied to the growing capacity to develop quantified models.[32] Randomness and uncertainty have also received considerable attention in the geosciences. Krauskopf,[33] in his Presidential Address to the Geological Society of America (GSA), discussed, as an example, the difficulty of characterizing igneous rocks of the Inyo

Batholith, a satellite suite of rocks on the east side of the Sierra Nevada Batholith. He discussed the difficulties of describing this small part of the Earth's crust and answering even simple questions about the origin of these plutons and their sequence of emplacement. Krauskopf argued that there is a limit to what can be known about this geologic system and that this limit is associated with uncertainty in the description of the rocks and random interactions in their history. This Presidential Address caused considerable discussion[34] and a rare reply from a retiring GSA president[35] in which he emphasized that by "random" he means "beyond the limit of possible measurement." In subsequent years, the discussion has continued[36-39] with recent exchanges on these same 10 plutons.[40-42] We recommend this series of papers to the reader because we believe that they demonstrate the difficult issues of describing and predicting the behavior of geologic systems.

The noted paleontologist, G. G. Simpson,[37] further distinguishes among deterministic phenomena that are called "random." There are immanent processes (governed by scientific laws) that provide the basis for deterministic models (process-based models), as contrasted with the state of a system (configuration) which is the sum of its previous history. The geologic history of a portion of the Earth's crust or the future behavior of a repository are both defined by "configurational changes through time, i.e., a sequence of real, individual but interrelated events."[43] In a PA, these are features, events, and processes (FEPs) that may affect the repository, and the PA analysis captures the sequence of configurations by conditional probabilities and explicit assumptions about time-dependent boundary conditions on the system. Outcomes become predictable, in principle, if all necessary immanent processes are known, but as soon as configuration is included as "one of the necessary causes which must always be done in historical science [or, the skeptics might add, the prediction of future repository performance], the situation may become extremely, often quite impossibly complicated. Prediction is possible only to the extent that correlation can be established with pertinent, abstracted and generalized, recurrent elements in configurations."[43] Thus, geologic systems are extremely difficult to model over extended periods, and uncertainties are so large that careful attention has to be paid to possibilities of alternative conceptual models, incomplete descriptions of natural systems, and difficulty of describing future configurational patterns correctly. The challenge to PA is to demonstrate the

practical extent to which an analysis can capture this sequence of configurations and provide useful information on long-term behavior.

The geosciences do not have unique intellectual tools with which to address problems of analysis of repository systems, but experience in the geosciences does provide a unique and extremely useful perspective on this problem. Note that Krauskopf's and Simpson's discussions of randomness[33,35,37] are based on actual examples, real experience in describing geologic systems. In contrast, the formulation of PA methodology is a prescribed sequence of steps.[44,45] The only possibility of an interplay between inductive and deductive methods is the recurring iterative analysis of a PA, and yet this iterative analysis requires time and associated increases in cost and delays in licensing. The failure to bridge the gap between the elegance of the mathematical formulation of PA and actual, demonstrable practice of describing geologic systems is one reason many remain skeptical of PA results. At the very minimum, PA must provide a demonstrable basis for its acceptance as a methodology for describing the long-term behavior of a geologic system. This might be done by designing laboratory or field-scale experiments to challenge the coupled submodels used in a PA. E. Bullen, a mathematician, in his acceptance of the Day Medal of the Geological Society of America (given for the application of physics and chemistry to geologic problems) said, "a mathematical model of the natural world has value to the extent that it represents the observational data and has the capacity to inspire new observational tests."[46] This should be a requirement of all models used in PA.

Geology provides many examples of the results of failing to use the correct conceptual model. At the end of the nineteenth century, a classic scientific controversy occurred over the age of the Earth.[47] In the second half of the nineteenth century, geologists and biologists were reaching the end of nearly 300 years of effort, culminating in Darwin's *Origin of Species* and the concept of natural selection operating over great periods of time. At the same time, physicists, led by Lord Kelvin, were establishing the fundamental laws of thermodynamics. Lord Kelvin applied energy laws to the Sun as an energy source and to the cooling of the Earth, and calculated a relatively short age of the Earth, tens of millions of years, with great confidence and mathematical skill. The modeled results were not consistent with observational evidence in the biological and geological sciences that suggested much longer times; at least hundreds of

millions of year were required. However, Kelvin so dominated the discussion that Darwin retreated in his advocacy of a great age of the Earth. Kelvin was wrong because he used the wrong conceptual model. He failed to consider the heat generated by radioactive decay. Failure to establish firmly the basis of the conceptual models (there will be hundreds in a single PA, each coupled to the next) and failure to challenge these conceptual models can have a profound effect on the results.

A source of data and the opportunity to test conceptual models can be provided by "natural analogue" studies. There is an abundance of literature on the use of natural analogues,[48-52] and yet little is actually used in the probabilistic PA, either for submodel testing or as a source of fundamental data for the PA. The failure to utilize fully these data is a failure of the PA probabilistic methodology; that is, the PA methodology has such a structured format for input parameters that simple, often qualitative but relevant observations cannot easily be included in the analysis.

A final set of philosophical issues has to do with proper use of models in general and the specific limitations of their use in describing geologic systems. A useful and provocative summary of these issues has been given by Oreskes et al.[53] Their paper deals directly with issues of verification and validation with special attention to use of numerical models in earth sciences. Their paper generated an immediate response from the nuclear waste community[54,55] and a continuing response (hundreds of citations) across a wide range of disciplines. Their most cogent point is that "In areas where public policy and public safety are at stake, the burden is on the modeler to demonstrate the degree of correspondence between the model and the material world it seeks to represent and to delineate the limits of that correspondence."[53]

3.2. Practical Limitations and Experience

Earth science provides the most fertile ground for determining the extent to which one should expect to successfully model large-scale, complex geologic systems over extended periods of time. This subject has a long history in the geosciences and continues to be the subject of active discussion. As recently as the fall of 1997, the Geological Society of America (GSA) sponsored a special symposium entitled, "Predictive Modeling in the Earth Sciences: Application and Misapplication to Environmental Prob-

lems." A dialogue between geoscientists can be found in the "Forum" discussion in GSA Today.[56] One of the speakers, Victor Baker, captures the central issue of the discussion: "Given the established modes of engineering and scientific reasoning . . . , it is clear that the model-centered viewpoint of Earth-system science embodies a highly questionable hybridization of engineering and scientific reasoning."[57]

At the scientific level, significant errors in PA may occur because of the selection of the wrong deterministic model for a phenomenon, an incorrect mathematical solution for the model, an incomplete description of the system to be modeled, or because the "abstractions" of the models may not capture the essential behavior of the phenomena in the system. Moreover, complicated coupled systems may behave nonlinearly. In the PA of a nuclear waste repository, two essential elements of the system are (1) the chemical interactions between fluids, the waste package, and surrounding rock and (2) the transport of fluids and radionuclides to the accessible environment. Thus, we illustrate and discuss the limitations of PAs by consideration of two important disciplines in the PA of a repository: geochemistry and hydrology. Both of these disciplines already have developed rather sophisticated, deterministic-based models. The models have a strong experimental and observational database. Still, even when experimental and field data are abundant and the present state of knowledge is utilized, capturing the long-term behavior of a geologic system has proven to be difficult. These geochemical and hydrologic models are more sophisticated than those commonly included in a PA analysis, for which the geochemical and hydrologic models are only a small part of the total system analysis. The following sections discuss some specific examples of the difficulties encountered in geochemical and hydrologic systems.

3.2.1. Geochemistry

Although there is a substantial understanding of the behavior of various chemical elements, one must compare the expectations of the PA against what has actually been demonstrated to be possible with geochemical models of geologic systems. Geochemical models have had limited, although increasing success in describing water–rock interactions. Good examples of this work can be found in the series Reviews in Mineralogy: Chemical Weathering Rates of Silicate Minerals[58] and Reactive Transport in Po-

rous Media.[59] Still, the evolution of groundwater compositions over time is difficult to predict, as are the phase assemblages formed during the alteration and weathering of even common minerals; particularly difficult to model are groundwater trace element compositions and their host phases. Further, geochemical models of even simple systems (e.g., O_2 fugacity set by sulfide equilibria) may not have unique solutions[60]; and despite impressive progress in quantitative analysis of the time-space transport of solutes and their reaction with minerals,[61] the limiting conditions of such calculations make them difficult to apply with confidence. Steady—even spectacular—progress has been made in predicting the behavior of geochemical systems; the progress is most often made, however, by systematic studies of well-controlled systems.[62]

Nordstrom,[63] Bethke,[64] and Read[65] provide examples and discussions of the limits of geochemical and geochemical-transport models. These include:

1. Lack of adequate characterization of the system and the identification of chemical components that have the primary effect on systems.
2. Lack of knowledge of species in solution or changes in speciation with evolving solution composition and temperature.
3. Lack of fundamental data (e.g., thermodynamic data for relevant actinide phases, e.g., uranium[66] has proven to be an enormous undertaking and many gaps and inconsistencies in the data remain).
4. Selection of the correct conceptual model (kinetic vs. equilibrium models; transport as molecular species in solution vs. colloids).
5. Use of the wrong kinetic rate constants because the experimentally determined rate laws do not apply to the natural system (e.g., due to a change in reaction mechanism).
6. Difficulties in scaling from well-constrained laboratory experiments to field-scale interpretations.
7. Extrapolation of short-term laboratory data to long periods of time.

Finally, inadequacies in the conceptual models or associated databases cannot be entirely overcome by the use of elicited expert opinion because the expert opinion itself relies on a knowledge and appreciation of the conceptual models and the database.

The quality of the thermodynamic data used in a geochemical model can seriously affect model output. As an example, under oxidizing conditions, second-

ary phases formed during corrosion of UO_2 are dominated by U^{6+} phases.[67,68] However, thermodynamic data for these uranyl phases are extremely limited, and rather small variations in the measured or estimated Gibbs free energy may cause appreciable changes in the modeled results. Consider the important system SiO_2–CaO–UO_3–H_2O. Stability relations among the uranyl phases that are found frequently in oxidized uraninite ore deposits and which are expected to form during oxidative corrosion of spent nuclear fuel are shown in Fig. 3. The Gibbs free energies for metaschoepite, soddyite, and becquerelite used in constructing Figs. 3a and 3b are identical (Table I). Other thermodynamic data necessary to construct Figs. 3a and 3b are from Grenthe et al.[66] However, the Gibbs free energy for uranophane is
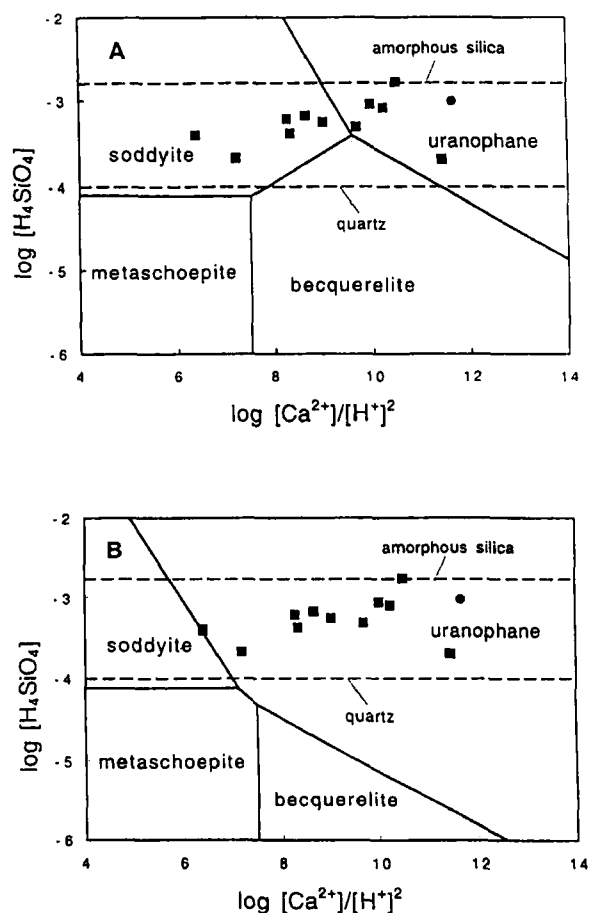

Fig. 3. Activity-activity diagrams for the system SiO_2–CaO–UO_3–H_2O constructed using different $\Delta G^0_{f,298}$ values of uranophane (Table I) with compositions of typical groundwaters from crystalline rocks plotted as squares[24] and the Yucca Mountain ground water composition as a circle. Horizontal dashed lines represent quartz and amorphous silica equilibria.

Table I. $\Delta G^0_{f,298}$ for Uranyl Phases used in Constructing Figs. 3a and 3b (kJ/mol)[a]

Uranyl phases	Fig. 3a	Fig. 3b	M/E
metaschoepite	$-13,092.0$[66]	$-13,092.0$[66]	M
becquerelite	$-10,329.1$[69]	$-10,329.1$[69]	E
soddyite	$-3,655.7$[69]	$-3,655.7$[69]	M
uranophane	$-6,192.3$[b]	$-6,210.6$[c]	M

[a] M, measured; E, estimated.
[b] Calculated by Chen et al.[69] based on the solubility data of Casas et al.[70]
[c] Calculated by Chen et al.[69] based on the solubility data of Nguyen et al.[71]

taken from two different studies. In Fig. 3a, δG of uranophane is $-6,192.3$ kJ/mol, as calculated by Chen et al.[69] based on the solubility data of Casas et al.[70] Figure 3b uses a δG of $-6,210.6$ kJ/mol based on experimentally determined solubility data.[71] Although both values were obtained from solubility experiments and the difference between these two values is only 0.3%, the stability relations among the phases (Figs. 3a and 3b) are significantly different. The stability field for becquerelite is much larger in Fig. 3a. Representative groundwaters from crystalline rocks plot in either the soddyite or uranophane stability fields and are close to becquerelite, indicating that uranophane and soddyite should be the most common uranyl phases in nature and as the reaction proceeds, becquerelite may be associated with soddyite or uranophane or both. In contrast (Fig. 3b), soddyite cannot equilibrate with becquerelite in aqueous solutions because all groundwater compositions are in the uranophane stability field, far from the bequerelite field. The stability fields for solid phases shift dramatically even for small variations in measured or estimated Gibbs free energies. If the conceptual model for radionuclide retardation involves the incorporation of radionuclides into the alteration phases (e.g., the formation of becquerelite), then the conceptual model may change greatly due to very small variations (in this case 3%) in experimentally determined data. This is not a new observation. In 1992, the International Association for Mathematical Geology compiled a monograph of contributed chapters on the status and limitations of models of geologic systems.[72] In the chapter, "Geochemistry," Siegel et al. provide this concluding evaluation of geochemical systems:

"The problem is exacerbated by the current lack of good thermodynamic and kinetic data for the chemistry of some radionuclides. In currently available

methods, the most serious limitations are the use of radionuclide solubilities or sorption distribution constants whose values are not coupled to changes in solution chemistry during transport. If all significant interactions among geochemical and hydrologic parameters are not included in the model, the probabilistic risk assessments and sensitivity analyses may be misleading."[73]

3.2.2. Hydrology

If a geologic repository were to violate compliance and radioactive contaminants in excess of what is permitted by regulation were to reach the accessible environment, the most likely pathway would be through water-saturated rocks, and the release mechanism would include the processes of advective and dispersive transport of constituents in flowing groundwater. Thus, PA must incorporate models of groundwater flow and solute transport through rocks and aquifers between the repository and the accessible environment. Even though the overall PA modeling procedure is probabilistic in nature, it is based on the use of deterministic groundwater flow and transport models. Can such deterministic groundwater models reliably predict future conditions? The PA process itself must be linked adequately with deterministic groundwater models, and neither the PA sampling procedure nor the coupling of groundwater models with other subsystem models should induce significant errors or bias into the predictive process.

Six essential steps in conducting the ideal probabilistic systems study were outlined previously. But what if the actual implementation deviates significantly from the ideal? The models will still work and generate numbers that will provide a basis for estimating risks.

A major concern is that fundamental differences exist in the characterization of engineered versus natural systems. In engineered systems, geometry, initial conditions, boundary conditions, and most properties of components are fairly well known, and the degree of uncertainty also can be fairly well characterized. System properties are generally homogeneous for given materials. Most operating processes are also well known. In modeling natural hydrogeologic systems, however, significant uncertainty typically exists about which variables and processes should even be incorporated into the model. That is, significant uncertainty may be associated with the conceptual model represented by the partial differential equa-

tions that are being solved. Thus, if the PA models generate mathematically accurate solutions to the wrong equations (i.e., wrong conceptual model), this "accuracy" does not mean that the PA analysis is reliable or meaningful.

For natural systems, we expect the "model-building process" (Steps 2 and 3 in Section 2.3.1) to be imperfect to some degree. Because natural systems are heterogeneous, and because costs of field testing typically leave us with only a very limited number of measurements and samples, accurate and precise characterization of system properties, their spatial variability, and their uncertainty is problematic.

The ideal model-building process also is characterized by using initial models to make testable predictions, then making measurements to test predictions, then improving the model on the basis of these tests and observations. The PA process is not typically implemented this way. Also, in model application, it is rare to make a true prediction and then wait to observe the outcome; the usual order is to observe the system, then calibrate the model to match the observed history. Although this may be called a prediction, it is the known past that is being "predicted." A groundwater model may adequately reproduce historical data, but fail to predict future responses under a new or extended set of stresses; a good match with historical behavior does not validate the model.[74]

3.2.2.1. Predictive Accuracy of Groundwater Models.
Predictive applications of deterministic groundwater models have become fairly common during recent years, and their goals include assessments of groundwater availability, assessments of subsurface contaminant migration at toxic waste sites, design of cleanup schemes for contaminated aquifers, prediction of saltwater intrusion in response to withdrawals from supply wells, and PA for proposed radioactive waste repositories.[75] The latter might represent the most extreme example of reliance on groundwater model predictions because regulations may require the assurance of site compliance for 10,000 years into the future.

Because of natural heterogeneity of geologic systems, hydrogeologic properties and boundary conditions of an aquifer system cannot be defined accurately and uniquely. Therefore, parameter estimates in groundwater models usually are calibrated through history matching for a specific site before the model is used to predict the future state of that system.

One method of assessing the predictive accuracy of groundwater models is to compare the actual response of a groundwater system with that predicted by a calibrated model and to perform such a comparison for a sufficiently long time after the prediction was made so that the state of the system at the time of evaluation will not be dominated by its "memory" of conditions during the calibration period for the model (typically, this may require several years).[75] This type of assessment of model reliability has been called a postaudit.[76–80]

One of the few documented postaudits was for a deterministic distributed-parameter model developed to predict regional water-level changes in the alluvial aquifer system of the Salt River Valley and the lower Santa Cruz River basin near Phoenix, Arizona. The model included a calibration period of more than 40 years.[81] Water-level declines over the next 10 years were predicted using the calibrated model. Predicted and observed water-level changes in 77 wells (Fig. 4) indicate poor predictive accuracy and the presence of a bias in the model predictions.[76] Also, data show a relatively wide scatter, indicating that model prediction was imprecise.

This one example is not atypical of results seen in other published postaudits. The errors in a 10-year prediction of water-level changes based on a calibration period of 40+ years can be contrasted with expectations for a 10,000-year prediction of transport (a more complex process) based on a model having a negligible calibration period. In general, results from all postaudits show that predictive accuracy is typically poor.[75]

When model parameters have been adjusted during calibration to obtain a best fit to historical data, a bias toward extrapolating existing trends is present when predicting future conditions because predictions of future stresses often are based on existing trends. Thus, although one advantage of deterministic models is that they represent processes and so have cause-and-effect relations built into them, estimating future stresses can be the major source of error in predictions of future effects (system responses). Furthermore, concepts inherent in a given model (e.g., two-dimensional flow and vertically averaged parameter values) may be adequate over the observed range of stresses, but may prove to be oversimplified or invalid approximations under a new type of stress or for a larger magnitude of stress or over a significantly longer period of time.

Discrepancies between observed and predicted responses of a system such as were noted in published postaudits are a manifestation of errors in the mathe-
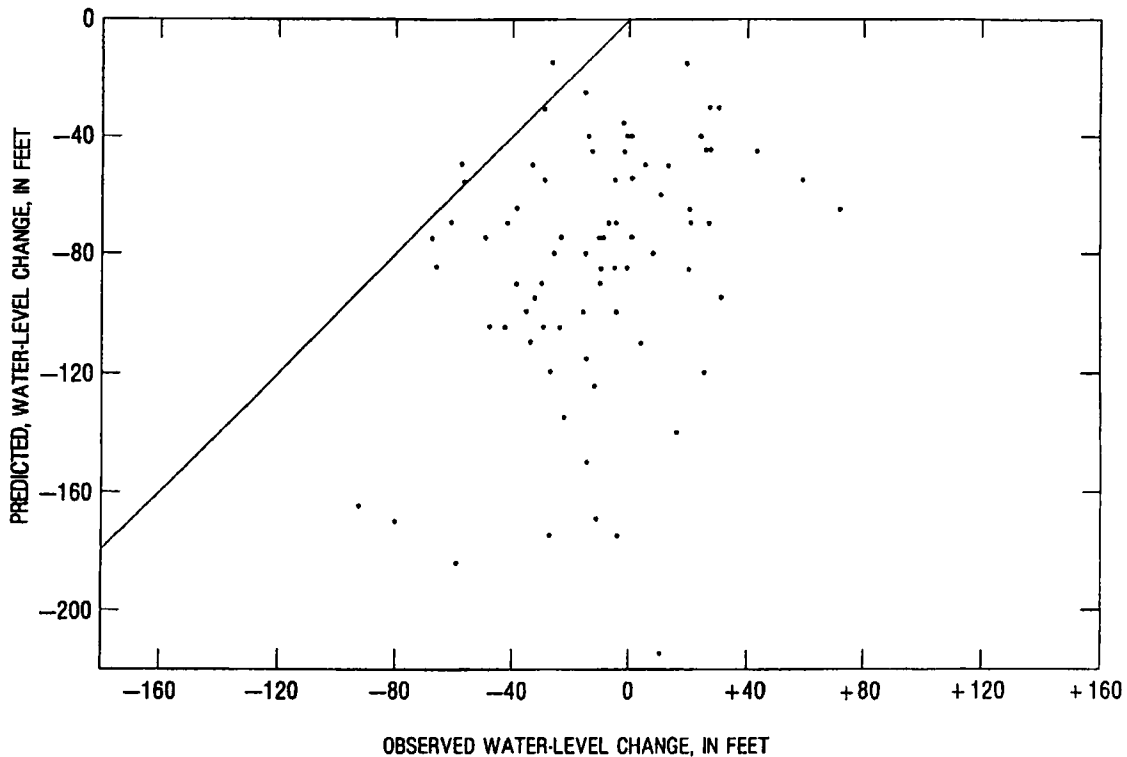
Fig. 4. Relation between predicted and observed changes in water level in the Tempe-Mesa-Chandler area of the Salt River Basin, Arizona, 1964–1974. Solid line shows where predicted equals observed values.[76]

matical model. In applying groundwater models to field problems, there are three sources of error[74]: (1) conceptual errors or theoretical misconceptions about the basic processes that are incorporated in the model, (2) numerical errors, and (3) uncertainties in input data. Conceptual errors include both neglecting relevant processes and representing inappropriate processes (e.g., application of a model based upon Darcy's law to materials where Darcy's law is inappropriate). Numerical errors include truncation errors and numerical dispersion, and usually are negligible in solving the flow equation, but are often of concern in solving the solute-transport equation. Uncertainties and inadequacies in input data reflect an inability to describe aquifer properties, stresses, and boundaries, and, together with conceptualization problems, are the most common sources of error.[74]

A key assumption in PA appears to be that uncertainty in parameters and boundary conditions can be adequately characterized so that the statistical distribution of the outcomes of a large enough number of samples will yield meaningful risk values. Hydrogeologists with extensive field experience, however, may remain skeptical because they think that the

degree and nature of uncertainties are themselves highly uncertain and difficult to characterize, and this does not appear to be accounted for in the PA analysis.

Uncertainty in conceptual models of the system and of governing processes may be incorrectly taken into account or neglected. Sensitivity analyses performed with the PA models can examine sensitivities of parameters only within the framework of the chosen models. PA models themselves cannot evaluate their sensitivity to other concepts or processes not incorporated into the model. Sampling over a large range in values of coefficients of governing equations does not correct this deficiency. Solving the wrong equations (i.e., using a flawed or erroneous conceptual model) many times, based on a wide range of parameter values, will not necessarily yield a meaningful probabilistic distribution of outcomes. If a flawed groundwater model is incorporated into the PA process, then results obtained by applying the model must be flawed. The seriousness of the "flaw" is difficult to assess.

3.2.2.2. Groundwater Models in Performance Assessment. A basic tenet of the PA process appears to

be that making many simulations based on realizations sampled from a wide range of parameter values can generate a meaningful statistical distribution of outcomes in which sampling procedure inherently compensates for (1) ignorance and uncertainty in parameter values, boundary conditions, and future stresses, (2) simplifications and linearizations of complex processes, and (3) uncertainty in underlying conceptual models. *This is a major and critical underpinning that is virtually unprovable and must be accepted on faith.* One could argue that if distributions to be sampled are wide enough, then all possible outcomes will be encompassed. However, if sampled distributions are very wide because of ignorance rather than because of physical or chemical heterogeneity, the statistical properties assumed for those parameters will be in error.

The incorporation of groundwater models into PA can introduce subtle errors in PA results from a variety of sources, including the PA parameter sampling procedure itself, inappropriate scaling of model parameters, numerical errors in transport models, and inconsistent linkages between different models. As with many complex modeling procedures, "the devil is in the details." For example, to generate statistical distributions from which risks are calculated, many simulations of hydrogeologic processes are performed to generate an adequate sample size. The approach to varying values of many parameters in multiple realizations can introduce errors into the final analysis. In particular, if hydrogeologic variables that are correlated strongly are sampled independently and the correlations are ignored, then some realizations may be based on unreasonable or unlikely combinations of parameters; such individual simulations will skew the statistical results. The performance assessment for the Waste Isolation Pilot Plant (WIPP) separately sampled and independently varied aquifer transmissivity, fracture spacing, and porosity, although good reasons exist to suspect that these variables are interrelated.

To illustrate, suppose that a strong positive correlation exists between two critical model parameters (Fig. 5a) and that all data points will fall within the indicated bounds. Furthermore, assume that a safety failure for the geologic repository will occur only if values for both parameters are large (Fig. 5a). If the PA sampling procedure generates values for both parameters independently from uniform distributions for a sample size of 25, for example, then the 25 realizations can be expected to be based on parameter values distributed uniformly in the sample space (Fig. 5b). A

plot of joint values of the two parameters should be expected to yield one point in each of the 25 squares in the grid. In this case, only one of the 25 realizations (representing the value in the upper right corner of the grid) will yield a failure, and the probability of failure would appear to be 0.04. Twelve of the samples, however, were obtained from outside of the bounds of the feasible set of values. If these 12 are discarded, the failure probability is only 1 of 13 (or about 0.077), or nearly twice as great as when the correlation between these two variables is ignored.

Much recent hydrogeologic research has clarified the importance of aquifer heterogeneity in controlling solute transport. Heterogeneity at a particular scale causes greater variation in solute concentration than in hydraulic head. Thus, an adequate definition of formation heterogeneity for a flow model may be inadequate for solving the equation for transport in the same formation. Konikow[82] presented results indicating that the 1996 WIPP PA consistently underpredicted migration distance of a plume emanating from a human intrusion borehole. Errors arising from several sources, including numerical dispersion and spatial truncation errors in the transport code, poor resolution from using a grid that is too coarse for the scale of the problem, and overestimates of the size of a solute source area, cause an artificial spreading of the calculated width of the plume, which is offset by an underestimate of the length of the plume. An alternative analysis using a model in which transmissivity variations are represented on a smaller scale, using a 2-m grid spacing rather than the original 50-m grid spacing, results in a much longer, but narrower, plume that has a significantly shorter travel time to the regulatory boundary.[82]

As noted in the discussion in Section 2.3.1.2 related to Fig. 1b, concern about errors in the WIPP Culebra model are made moot by changing the assumptions in other subsystem models that calculate releases to the Culebra and result in negligible quantities of contaminated brine being released to the Culebra. In the specific case of WIPP, the soundness of some of these new assumptions (especially relative to the interconnectedness of the repository, the volume of brine likely to escape from deeper geologic formations, and the need to consider potential incursions from waterflooding operations in nearby hydrocarbon production areas) is still questioned. Regardless, groundwater models used in a PA must be based on sound conceptual models and reliable parameter identification.
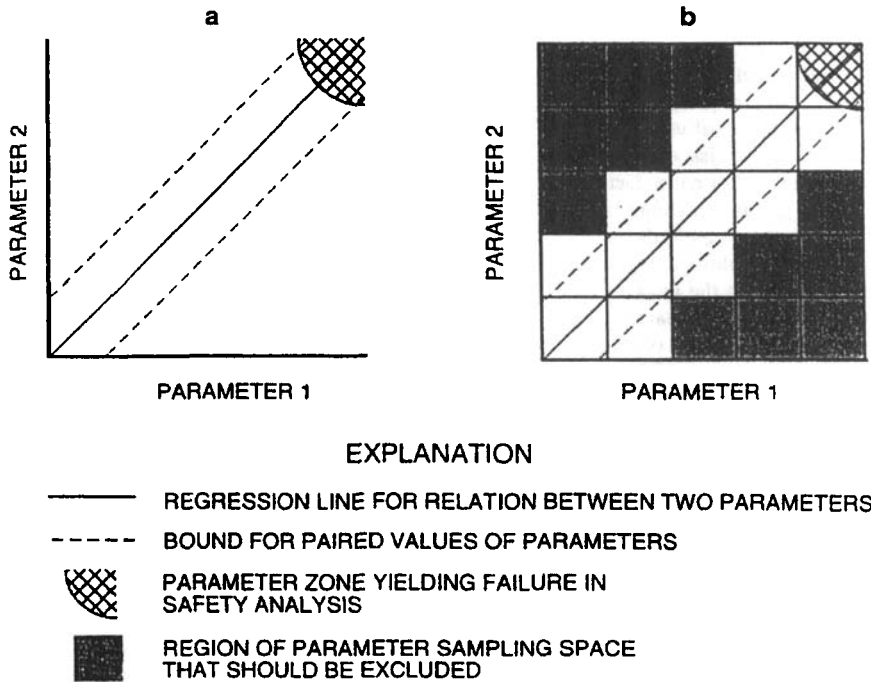
EXPLANATION

——————— REGRESSION LINE FOR RELATION BETWEEN TWO PARAMETERS

- - - - - BOUND FOR PAIRED VALUES OF PARAMETERS

PARAMETER ZONE YIELDING FAILURE IN
SAFETY ANALYSIS

REGION OF PARAMETER SAMPLING SPACE
THAT SHOULD BE EXCLUDED

Fig. 5. Example to illustrate effect on calculated risks of independently sampling values of two parameters that are strongly correlated. (a) All data points (paired values) fall within indicated bounds (or confidence intervals). (b) Independent sampling can yield paired values that are "out of bounds" and should be excluded from the PA.

3.3. Effect of Probabilistic Analysis on Regulatory Time Frame

Another important result and limitation of using probabilistic analysis may be the limited time frames proposed or accepted by most regulatory requirements. Time frames that are commonly discussed are in the range of tens of thousands of years. Two reasons for arriving at such geologically short periods of time are: (1) During the first 10,000 years, there is a substantial decrease in the radiotoxicity of the spent nuclear fuel because of decay of short-lived fission products; however, total activities because of actinides and their daughter products remain elevated (as compared with natural uranium ores) for periods extending beyond one million years.[83] (2) Beyond 10,000 years, uncertainties in the probabilistic analysis increase substantially. The Advisory Committee on Nuclear Waste[84] has noted:

The dilemma in developing a TOC [time of compliance] is that the time span must be sufficiently long to permit evaluation of potential processes and events leading to the loss of integrity of the repository and transport of radionuclides to the critical population. Yet the period must be short enough that inherent

uncertainties in processes and events and in the biosphere and critical population group, which will increase with time, will not invalidate the results of the evaluation. *Reasonable confidence must exist that the uncertainties in the reference calculation for the time span can be identified and quantified in a probabilistic format.*[84] [Italics added]

The ACNW[85] has proposed a two-step analysis: (1) An explicit regulatory time requirement for a shorter period such that "extrapolations of significant processes and their rates can be made robustly with reasonably modest uncertainties."[85] (2) A second part of compliance would be based on assessments extending from the shorter specific compliance period to the calculated time of the peak risk to a critical group.[85]

The National Academy of Sciences (NAS) committee has noted that the peak risks will probably occur after a 10,000-year period, and thus "the basis for the standard should be the peak risk, whenever it occurs [within the limits imposed by the long-term stability of the geologic environment]."[86] The ACNW finessed the obvious difficulties that arise from the NAS recommendation by noting that, for the second step of the analysis:

"There is no definitive measure of compliance in the sense of a numeric match between a standard and the calculated peak risk, and this second part should not be allowed to become a de facto regulation. A comparison between the standard used in the first part and the calculated peak risk should lead to identification of important performance factors that define risk to the critical group. Depending upon the extent to which the peak risk exceeds the standard, ameliorating actions to reduce this difference should be initiated, such as increasing the integrity of the engineered barriers, improving site characterization to more closely bound uncertainties, or, in the extreme, abandoning the candidate site."[85]

Thus, what is the final regulatory result of a probabilistic analysis?

1. The regulatory time for geologic disposal becomes remarkably short as compared to geologic periods. A 10,000-year period barely includes the time frames over which geologic systems operate. One of the ironies of the concept of geologic disposal at Yucca Mountain is that the rocks themselves are so young, approximately 10 million years, located in the Basin and Range Province, a broad region of the western United States that from a geologic perspective, is a region of active tectonism and vulcanism. The fundamental geologic characteristics of the region may not be captured in this 10,000-year "snapshot."

2. The probabilistic analysis is still carried forward for extended time periods (but not as a "de facto regulation") despite the inevitable propagation of large uncertainties in the analysis. Ironically, the more effective the repository is in extending the time of peak risk to a critical population, the further out in time the probabilistic analysis must be extended.

3. Abandoning a demonstrably poor site becomes an "extreme" possibility after having undertaken an expensive and time consuming process.

We suggest that the unsatisfying inconsistencies in the present situation have much to do with the reliance on probabilistic analysis.

4. DIALOGUE

In order to focus the dialogue on a few specific issues, the "skeptics" have posed a series of questions in **bold type** to be answered by the "proponents." Background for some of these questions can be found

in an early paper on the geologic disposal of nuclear wastes by Bredehoeft et al.[5] and a more recent paper by Oreskes et al.[53]

1. **What is the basis for selecting a conceptual model? Is it the ability of a model to capture the fundamental physical and chemical processes of the system the main criterion, or is it computational simplicity? We understand that both criteria must be used, but where is the balance? This is an important issue because the computational demands of a PA of the entire system are such that submodels in the total-system model must often be "abstracted," that is, reduced to computationally efficient modules that are supposed to capture the physical and chemical behavior of the submodel system. Is the abstraction compared with the model, or is it compared with actual experimental data or field observations? In what sense can the abstracted models used in a PA be tested? Are they tested?**

The only rational bases for selecting conceptual models of a system are that they fully capture those features and physical processes occurring in the real system that are believed to be essential to assessing the ability of the system to perform as intended, and they are not inconsistent with observations or general physical and chemical principles. Conceptual models need not be unique. A given conceptual model can support many mathematical models at different levels of detail; the computational simplicity (or complexity) inherent in a mathematical model should not be an initial consideration in determining the acceptability of its associated conceptual model. The balance struck between computational simplicity and a model's ability to capture essential effects is ideally determined by sensitivity analyses and comparisons of results of runs with alternative mathematical models; if two distinct models produce virtually the same results, the computationally simpler of the two will of course be preferred; if two different but physically acceptable models produce radically different results, the one that produces the more conservative results (conservative in the sense of moving performance measures toward violation of standards) will be selected, regardless of its computational complexity.

Two kinds of "abstraction" are used in current PAs. First, there is the process of replacing a complicated and computationally cumbersome model having, say, N uncertain parameters with a much simpler function that is constructed by least-

square fitting a limited number of outcomes of the complicated model to an analytically convenient formula for a surface in $(M + 1)$-dimensional space. Usually, $M \ll N$, and the M independent variables of the formula are ideally chosen from the most sensitive of the complicated model's parameters. This trick for simplifying complicated models works well when the complicated model's response surface in $(N + 1)$-dimensional space is known to be nearly "flat" everywhere throughout the domain of the uncertain variables, i.e., the complicated model's dependent variables are very nearly linear functions of the N_u uncertain parameters (see Section 2.3.1.2). But the technique should be applied with caution when working with submodels that are known to be highly nonlinear or whose sensitive parameters have not been fully explored: In these instances, blind application of this kind of abstraction can lead to the erasure of essential features on the complicated model's response surface and, consequently, erroneous conclusions about the ranges of the performance measures. It makes no sense to compare this kind of abstracted model with its more complicated "parent" or with empirical evidence that could be used to verify the parent model.

A second kind of abstraction involves volume-averaging a system of partial differential equations (PDEs) over each material domain in the computational grid to obtain a much more computationally tractable system of ordinary differential equations. The abstracted model that results from volume-averaging globally conserves the same quantities that are locally conserved in the original model, but now all knowledge of spatial variation in locally conserved quantities (e.g., solute mass densities in liquid and solid phases) is lost. But knowledge of spatial variation may be inessential in some applications, e.g., estimating the rate of natural bioattenuation of groundwater contaminants from a localized spill.[87] Results of these volume-averaged abstractions are often easier to compare with field and laboratory data than the voluminous results of solving their parent PDEs.

Abstracted models used in a PA may sometimes be compared with field and laboratory data, as indicated above; more often, because of extreme expenses of useful confirmatory experiments, they are only amenable to "testing" through peer review. Are they tested? In the sense of "testing by peer review" just mentioned, and in an ideal PA, the answer is yes.

2. To the extent that the hypotheses concerning the selection of the abstracted models or the databases used in the models are not testable, then is PA fundamentally unscientific?

The PA process could be said to be fundamentally unscientific, but only in the sense that a PA is *not* intended to be a scientific investigation. Scientific investigations seek to establish "reliable knowledge" of some part of the universe.[88] In contrast, a PA is a type of *engineering analysis* that uses *existing scientific knowledge* to assess, in as rational and objective manner as possible, the ability of a proposed geologic waste repository to meet the standards and regulations that apply to it. A performance assessment is simply a way of providing objective, science-based input to public-policy makers. Of course, use of the methodology of scientific investigation, including the hypothetico-deductive method mentioned by the skeptics, is mandatory in field and laboratory studies that support an engineering analysis and provide an empirical basis for model parameters.

But, does the fact that the PA process does not in every respect conform to the hypothetico-deductive paradigm for scientific investigation necessarily imply that performance assessment is always "bad science" and is therefore useless, perhaps harmful, as input to public-policy making? We think not. Over 20 years ago, concerns were voiced about the effects of bad science on decision making in the environment and health areas.[89] In reply to these concerns, M. Granger Morgan[90] acknowledged the need for good science, but noted that "Good science and good policy analysis are not the same thing and do not serve the same ends. . . . Good policy analysis recognizes that physical truth may be poorly or incompletely known. Its objective is to evaluate, order, and structure incomplete knowledge so as to allow decisions to be made with as complete an understanding as possible of the current state of knowledge, its limitations, and its implications. . . . But with or without full understanding, society and its policy makers make decisions."[90]

Later, this point was made again in answer to the question, "Is Risk Assessment A Science?": "Risk assessment cannot be intensive nor selective in the way traditional science is. It must deal with questions as they arise without regard to their disciplinary assignment or to the quality and completeness of the data that are obtainable or at hand. . . . The risk assessment process is an interface between science and the society that created it and looks to it for advice."[91]

3. **A common feature of research programs designed in support of repository licensing is the development of experimental programs that "confirm" or "validate" models and databases. Research that challenges the conceptual or abstracted models used in the PA (and thus may move the program further away from compliance) are given a low priority. This is certainly not the scientific method, which is fundamentally a testing process designed to invalidate hypotheses. By what method does PA preserve the scientific approach of invalidating hypotheses?**

Ideally, research programs in support of repository licensing have the purpose of reducing *uncertainty* in those alternative conceptual models, mathematical model formulations, and parameters that have been proposed to describe the real system in the PA. The "interior analyses" of these research programs may indeed consist of hypothesis invalidations. We have already remarked on the impossibility of strict-sense confirmation or validation of most mathematical models used in PA.

We have also remarked on the fact that PA methodology need not follow the hypothetico-deductive doctrine of scientific investigation to produce results useful to decision makers. For example, very different roles are played by the concept of alternative conceptual model in a scientific investigation and in a performance assessment. In scientific investigations, alternative conceptual models are sets of hypotheses that are tested against empirical evidence; the hypothesis that best accounts for all evidence is temporarily accepted as the "truth" until further experimental evidence forces its replacement with a more explanatory hypothesis. In PA, considerable uncertainty may exist concerning which of several, equally valid conceptual models of a subsystem best mimics the real-world situation; in such a case, and pending further observations that could reduce uncertainty, the conceptual model that produces the most conservative effects on total system's performance (i.e., "conservative" in the sense of consciously erring on the side of safety) is accepted as the truth.

4. **When is the description of the system (e.g., site characterization) sufficiently complete to provide a satisfactory analysis? How is this demonstrated? Ideally, site characterization and the model-building process proceed simultaneously. Model-based estimates of un-**

certainty could be compared to actual uncertainty as data on the site become available. Does this happen? If not, why not? If such comparisons are made, what is the basis of considering the predicted values to be in agreement with measured values?

Ideally, site characterization is complete when all features, events, and processes (FEPs) associated with the site that could significantly affect waste containment have been identified, these FEPs have been incorporated into mechanistic models of system behavior, and data have been obtained through field or laboratory measurements which are sufficient to establish values of the mechanistic models' parameters. How does one know that all significant FEPs have been identified? One is really never sure, but failure to discern essential processes or detect potentially disruptive features seems unlikely given modern remote-sensing techniques and international databases on geophysical processes. Also, if the site has been selected for development as a waste repository as a result of a rational site-selection process, significant features of the site that relate to its ability to contain waste will already have been identified by literature searches or preliminary field investigations; it is unlikely that a rational site-selection process will lead to selection of a waste repository site in an unmapped and poorly understood geologic domain.

We agree that submodels of a total-system PA should ideally be developed interactively and in parallel with site characterization activities (Section 2.3.1.1). We know of only one instance where model-based estimates of uncertainty were compared with actual uncertainty in site-specific data: Distributions of average far-field permeability and average far-field pore pressure for halite and anhydrite at the WIPP site were predicted with a bootstrap model using small-scale measurements of these quantities taken near walls of openings[16]; the predicted distributions of permeability were later found to compare favorably with empirical distributions of intermediate-scale measurements of halite and anhydrite permeability. In general, the "goodness of fit" of a measured distribution to a predicted distribution can be tested by application of the chi-square test or the Kolmogorov–Smirnov test.[92]

5. **What is the basis for extrapolating models beyond the temporal and spatial scales of experimental data, field observations, or human experience? Relatively simple but coupled systems may exhibit nonlinear behavior in**

which experimental and mathematical discontinuities confound accurate and useful descriptions of the system. Why should we expect greater success in accurately describing the behavior of more complex geologic systems?

Extrapolation of results of a mechanistic model to temporal and spatial intervals well beyond those used in formulating the model is often justified by the fundamental belief that the *physical laws* on which a model is based are invariant under translations in time and space. Note that this is not to say that *rates* of physical and chemical processes are invariant under translations in time and space, or that the phenomena being studied are only amenable to a deterministic description.

Relatively simple *real systems* and the nonlinear equations used to describe them may indeed exhibit complex, even chaotic behavior (Question 6); but an honestly chaotic model, one whose equations are not underdetermined or numerically unstable, does not necessarily preclude a quantitative and useful description of the phenomenon being studied; indeed, chaotic behavior might be the essence of that phenomena. If a complex geologic system actually behaves in an intrinsically stochastic manner (see the "10 plutons" controversy mentioned in Section 3.1) and a properly posed model of that system also turns out to be intrinsically stochastic, then the modeling effort should be viewed as a success, rather than a failure. It is true, however, that PA may not be able to make direct use of intrinsically stochastic models. How to conduct sensitivity/uncertainty analyses of intrinsically stochastic models is, in our opinion, an open and interesting research problem.

In general, physical scientists can expect "greater success" with deterministic models soundly based on macroscopic, classical physics (as are most submodels of a PA) than with models based on microscopic statistical physics (e.g., models of plasma turbulence) because classical physics predicts phenomenon on spatial and temporal scales that humans find most comfortable for experimentally testing results of model predictions.

6. **Is it possible to distinguish between uncertainty that cannot be reduced by additional experiments or data acquisition and the uncertainty that is inherent in the system? This is an important distinction that may be forgotten once compliance is demonstrated.**

Two kinds of uncertainty can arise in studies of any real system. The first kind is called epistemic uncertainty because it arises from the investigator's imperfect knowledge of some system characteristics. Some examples of epistemic uncertainty: One knows that there must be a formula expressing the relationship between mass flux through a porous medium and fluid pressure in that medium, but does not know the precise mathematical form of that relationship. One knows that the spatially averaged effective porosity of a particular 1000-m^3 volume of rock is a number between zero and one, but does not know the precise value of that number. In each of these examples, the degree of uncertainty could, in principle, be reduced by performing certain experiments. Almost all uncertainty encountered in modeling engineered systems is epistemic. In the skeptics' discussion of uncertainty (Section 3.1), it appears that G. G. Simpson was concerned with Laplacean epistemic uncertainty in geologic and evolutionary systems.[37,43]

Note that within the category of epistemic uncertainty one must include instances of ignorance: One is simply unaware of the existence of a thing or phenomenon. A good example of "uncertainty as ignorance," i.e., the Lord Kelvin versus Charles Darwin debate, is cited earlier in this paper (Section 3.1).

The second kind of uncertainty is called aleatory uncertainty. This kind of uncertainty is inherent in the outcomes of trials with chance devices: e.g., a throw of dice, a spin of a roulette wheel, an hour's worth of counts of a Geiger counter. In all of these examples, uncertainty cannot be reduced—even in principle—by experiments performed in advance of making the trial throw, spin, or count. Until recently, under the mechanistic and deterministic Laplacean view of the universe, it was believed that aleatory uncertainty is ultimately epistemic in nature.[93] The discovery of chaotic dynamical systems by Henri Poincare and development of the concepts of forced and intrinsic stochasticity[94] have strengthened the idea that aleatory uncertainty has an objective existence independent of epistemic uncertainty. The case for believing that "chance is real" receives further support from results of studies showing that many natural systems and phenomenon, from flour-beetle populations[95] to global atmospheric circulation,[96] behave as chaotic dynamical systems. Currently, aleatory uncertainty enters PA models only through the intentional employment of "random number" generators to generate sample values of imprecisely specified parameters necessary for the Monte Carlo integrations of sensitivity/uncertainty analyses.

7. **Expert opinion is an important part of a PA. A difficulty arises when expert opinion becomes the substitute for data that could otherwise be obtained in reasonable amounts of time. Occasionally, the expert elicitation is described as a "placeholder" until experiments can be completed, but the use of expert opinion in the PA analysis is such an ingrained part of the process that the placeholders may become permanent. A solicitation of expert opinion is a much faster and cheaper way of obtaining parameter values than actually doing the required site characterization or laboratory experiments. Is the use of expert opinion clearly identified in a PA? Can the analysis distinguish between data that are elicited from experts and actual data? What would be the typical proportions of each in a PA?**

We agree that expert opinion should not be the sole basis for assigning values to *sensitive model parameters* in an *uncertainty analysis* whose results will be used in an application for certification of compliance; and of course, there should be reasonable connections between values assigned to less-sensitive parameters and real conditions prevailing at the proposed repository site. Ideally, the most sensitive model parameters in a phenomenological model are identified by sensitivity analyses conducted during site characterization; results of these analyses are then used to set priorities on further experimental programs that could provide an empirical basis for, and reduce uncertainty in, the critical parameters.

On the other hand, since "The purpose of computing is insight, not numbers,"[97] there can be no harm in exclusive use of expert opinion to assign ranges and distributions of parameters for purposes of *sensitivity analyses*. The purpose of a sensitivity analysis is to rank parameters in the order of their importance for determining model outcomes; and the model's investigators need only choose parameter ranges within which they believe the true value of the parameters will lie (Section 2.3.1.2).

Use of expert opinion to formulate models and assign values to model parameters should always be clearly identified in reports of work supporting a PA. Of course, computational models "see" only numbers, and cannot distinguish between numbers that are derived from empirical data and numbers based on considered, but subjective judgments of

experts; thus, it is impossible to discern the effects of, say, just the expert-judgment parameters alone in postanalysis results of a sensitivity or uncertainty study.

Of more than 1,500 parameters employed in the performance assessment for the WIPP CCA,[18] roughly 40% were precisely known (i.e., physical constants, radionuclide half-lives, etc.) or were based on empirical data from field or laboratory experiments. The remaining 60% of parameters could be said to be based on expert judgement, but this would be misleading since many of these were reasonable analogues of empirically based quantities, or were reference values for waste properties set by the CCA's implementing agency, or were "model configuration parameters," numbers that must be assigned by the investigator to configure scenarios or to make computational models run efficiently.

8. **A typical PA analysis may well be beyond the comprehension of a group of knowledgeable experts. The PA for the Waste Isolation Pilot Plant in New Mexico has over 1,500 parameters. The PA for Yucca Mountain will be even more complex. The complexity increases because of an apparent desire to demonstrate completeness, but this may have little to do with the real, expected performance of the repository. Does the increased complexity improve the analysis or obscure a thoughtful consideration of the result?**

The sheer size of reports and apparent complexity of technical analyses that support development and licensing of geologic waste disposal systems do not, we think, arise from investigators' inherent desires to demonstrate completeness or technical prowess. More likely, the complexity of the process for licensing a waste repository is a response to the public's hypercritical assessment of nuclear power and the requirements of administrative and environmental law in an increasingly litigious society. The frustrations expressed in this question are understandable; but opinions differ as to whether "increased complexity improves the analysis or obscures thoughtful consideration of results." The regulatory agencies, EPA and NRC, apparently believe that a thorough and painfully detailed reporting of PA results is necessary. Regulatory agencies may also strongly constrain the acceptable forms of analysis that can be used to test model systems against their own standards and regulations (Section 2.2).[98]

While recognizing the necessity of following regulatory agency guidelines for reporting a compliance application, at the technical level we proponents believe that in building quantitative models of large and complicated technological systems, "the object should not be prediction but, rather, insight that can guide the development of heuristic policy strategies."[21] We also know that simple, sometimes "back-of-the-envelope," calculations often yield greater insight into certain technical issues than could immediately be gained by calculations with an elaborate computer model. But some sponsors and regulators will not readily accept insights that are so simply obtained.

9. **How does one grade or score a PA? Given the complexity of a PA of a repository, what are the characteristics of a well done PA? What are the characteristics of a poorly done PA?**

We know of no objective and practicable way to score a PA. Cost/benefit analysis cannot be used to determine the value of a PA because the costs involved in conducting an engineering analysis of a large, technological system are usually a tiny fraction (in the "noise") of the costs of implementing that system. PAs are like other engineering analyses in that they are ultimately scored by the people who sponsor, produce, review, and use them; the scores come in the form of subjective opinions, and a consensus on a final grade (i.e., well done or poorly done) is not to be expected. The sponsor of an engineering analysis will give the study a high score depending upon whether the study meets the commissioned objectives and is submitted on time and within budget. The analysts who produce the study are likely to grade their own work on the basis of the number of compromises in technical quality they had to make in the course of the study in order to meet sponsors' deadlines and stay within funding constraints. Technical reviewers will give the study a passing grade depending upon whether knowledge in the reviewers' specialty area is properly used in formulating conceptual models and in drawing conclusions from the associated computational models. The users of a PA might include those study sponsors who use study results to support licensing actions with regulatory agencies. After taking into account their own technical reviewers' opinions of the study, regulators will give the work a good grade according to the degree to which the report of work does adhere to their

agency's form-and-content guidelines for a license application.

5. CONCLUSIONS

Performance assessment offers a reasonable approach, in theory, to evaluating risks of geologic waste disposal when faced with uncertainties. A method closely related to PA, Reliability Analysis, has proved successful in evaluating and predicting risks of some engineered systems. A major concern expressed in this dialogue focuses on the question of whether uncertainties in natural geologic systems over long periods, both in defining critical processes and in quantifying model parameters, are so large or difficult to characterize that the methodology fails. That is, if underlying conceptual models are weak or unproved or both and numerical models that are components of the PA process consequently contain significant but unmeasurable errors, are calculated risks accurate or meaningful? The authors of this paper agree that a need exists for additional research into this issue.

The total system being simulated in a series of PA models includes the natural geologic, geochemical, and hydrogeologic environment that contains the engineered and constructed repository. The hard-to-quantify characteristics of a geologic system provide the multiple natural barriers to contaminant migration that underlie the concept of geologic disposal. This makes it even more important to test scientifically, evaluate critically, and improve wherever possible the models and data representing the natural system.

PA methodology inherently requires the use of simplified models. Model simplification must, however, be justified and tested to avoid oversimplification. This critical testing appears to the skeptics to be missing in some cases, while it appears to proponents to be generally adequate. Loss of inessential details during model simplification is obviously not critical; one must be concerned about (1) loss of essential details, (2) lack of proper evaluation of effects of using oversimplified conceptual models, and (3) lack of data to support assumed statistical distributions of parameters and correlations among parameters. Such critical testing can lead to more reliable and defensible levels of simplification in models and (or) to simple methods to compensate for recognized errors or bias induced by the simplifications.

Skeptics and proponents agree that the real

value of the results of any performance assessment will ultimately depend on (1) the degree of understanding of site-specific relevant physical and chemical processes, (2) the capability of the chosen models to simulate those processes, (3) the degree to which model parameters can be associated with site-specific conditions, (4) the length of real time for which simulations are required, and (5) the care taken in implementing and documenting the results of the PA.

ACKNOWLEDGMENTS

REFERENCES

1. S. E. Logan (Chairman), *Workshop on Geologic Data Requirements for Radioactive Waste Management Assessment Models,* UNM Report No. NE-27(76), Union Carbide 297-1, ERDA Document No. Y/OWI/SUB-76/81726 (September 1976).
2. U.S. Nuclear Regulatory Commission, *Reactor Safety Study—An Assessment of Accident Risks in U.S. Commercial Nuclear Power Plants,* WASH-1400, NUREG-75/014 (1975).
3. J. E. Campbell *et al., Risk Methodology for Geologic Disposal of Radioactive Waste: Interim Report,* NUREG/CR-0458, SAND78-0029 (Sandia National Laboratories, Albuquerque, New Mexico, 1978).
4. J. W. Bartlett, H. C. Burkholder, and W. K. Winegardner, "Safety Assessment of Geologic Repositories for Nuclear Waste," in J.B. Fussell and G. R. Burdick (eds.), *Nuclear Systems Reliability Engineering and Risk Assessment,* (Society for Industrial and Applied Mathematics, Philadelphia, Pennsylvania, 1977).
5. J. D. Bredehoeft, A. W. England, D. B. Stewart, N. J. Trask, and I. J. Winograd, *Geologic Disposal of High-Level Radioactive Wastes—Earth-Science Perspectives,* Geological Survey Circular 779 (1978).
6. R. M. Cranwell *et al., Risk Methodology for Geologic Disposal of Radioactive Waste: Final Report,* NUREG/CR-2452, SAND81-2573 (Sandia National Laboratories: Albuquerque, New Mexico, 1987).
7. R. L. Iman and M. J. Shortencarier. 1984. *A FORTRAN 77 Program and User's Guide for the Generation of Latin Hypercube and Random Samples for Use with Computer Models,* NUREG/CR-3624, SAND83-2365 (Sandia National Laboratories, Albuquerque).
8. R. L. Iman, "Statistical Methods for Including Uncertainties Associated with the Geologic Isolation of Radioactive Waste Which Allow for a Comparison with Licensing Criteria," in *Proceedings of the Symposium on Uncertainties Associated with the Regulation of the Geologic Disposal of High-Level Radioactive Waste. Gatlinburg, Tennessee, March 9-13, 1981* (NUREG/CP-0022, CONF-810372, 1982), pp 145-157.
9. U.S. Department of Energy, "Total System Performance," Section 8.3.5.13, Chapter 8, Volume VII, Part B of Nuclear Waste Policy Act (Section 113), Site Characterization Plan: Yucca Mountain Site, Nevada Research and Development Area, Nevada (DOE/RW-0199, 1988).
10. M. S. Tierney, *Combining Scenarios in a Calculation of the Overall Probability Distribution of Cumulative Releases of Radioactivity from the Waste Isolation Pilot Plant, Southeastern New Mexico,* SAND90-0838 (Sandia National Laboratories, Albuquerque, New Mexico, 1991).
11. M. S. Tierney, "PA Methodology Overview," in R. P. Rechard (ed.), *Initial Performance Assessment of the Disposal of Spent Nuclear Fuel and High-Level Waste Stored at Idaho National Engineering Laboratory. Volume 1: Methodology and Results.* SAND93-233011 (Sandia National Laboratories, New Mexico, 1993), pp. 3-1-3-28.
12. B. G. Y. Thompson and B. Sagar, "The Development and Application of Integrated Procedures for Post-closure Assessment, Based upon Monte Carlo Simulation: The Probabilistic Systems Assessment (PSA) Approach," *Reliability Eng. Syst. Safety* 42, 125-160 (1993).
13. K. W. Dormuth and G. R. Sherman, *SYVAC—A Computer Program for Assessment of Nuclear Fuel Waste Management Systems, Incorporating Parameter Variability,* AECL 6814 (Whiteshell Nuclear Research Establishment, Atomic Energy of Canada Limited, 1981).
14. S. G. Bertram-Howery, M. G. Marietta, R. P. Rechard, P. N. Swift, D. R. Anderson, B. L. Baker, J. E. Bean, Jr., W. Beyeler, K. F. Brinster, R. V. Guzowski, J. C. Helton, R. D. McCurley, D. K. Rudeen, J. D. Schreiber, and P. Vaughn, *Preliminary Comparison with 40 CFR Part 191, Subpart B for the Waste Isolation Pilot Plant, December 1990* SAND90-234 (Sandia National Laboratories, Albuquerque, New Mexico, 1990).
15. WIPP PA (Performance Assessment) Division, *Preliminary Comparison with 40 CFR Part 191, Subpart B for the Waste Isolation Pilot Plant. December 1991,* SAND91-089 Vol. 1-4 (Sandia National Laboratories, Albuquerque, New Mexico, 1991-1992).
16. WIPP PA (Performance Assessment) Department, *Preliminary Performance Assessment for the Waste Isolation Pilot Plant, December 1992,* Vol. 1-5 SAND92-0700 (Sandia National Laboratories, Albuquerque, New Mexico, 1992-1993).
17. M. L. Wilson, J. H. Gauthier, R. W. Barnard, G. E. Barr, H. A. Dockery, E. Dunn, R. R. Eaton, D. C. Guerin, Ning Lu, M. J. Martinez, R. Nilson, C. A. Rautman, T. H. Robey, B. Ross, E. E. Ryder, A. R. Schenker, S. A. Shannon, L. H. Skinner, W. G. Halsey, J. D. Gansemer, L. C. Lewis, A. D. Lamont, I. R. Triay, A. Meijer, and D. E. Morris, *Total System Performance Assessment for Yucca Mountain-SNL Second Iteration (TSPA-1993)* SAND93-2675 (Sandia National Laboratories, Albuquerque, New Mexico, 1994).
18. U.S. DOE (U.S. Department of Energy), *Title 40 CFR Part 191 Compliance Certification Application for the Waste Isolation Pilot Plant,* DOE/CAO-1996-2184 (U.S. Department of Energy, Carlsbad Area Office, Carlsbad, New Mexico, 1996).
19. J. C. Helton, J. E. Bean, J. W. Berglund, F. J. Davis, K. Economy, J. W. Garner, J. D. Johnson, R. J. MacKinnon, J. Miller, D. G. O'Brien, J. L. Ramsey, J. D. Schreiber, A. Shinta, L. N. Smith, D. M. Stoelzel, C. Stockman, and P. Vaughn, *Uncertainty and Sensitivity Analysis Results Obtained in the 1996 Performance Assessment for the Waste Isolation Pilot Plant* SAND98-0363 (Sandia National Laboratories, Albuquerque, New Mexico, in preparation.)
20. U.S. Environmental Protection Agency, "Environmental Standards for the Management and Disposal of Spent Nuclear Fuel, High-Level and Transuranic Radioactive Wastes; Final Rule, 40 CFR Part 191," *Fed. Reg.* 50, 38066-38089 (1985).

21. M. G. Morgan and M., Henrion, *Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis*, (Cambridge University Press, Cambridge, 1990), pp. 172–216.

22. U.S. Nuclear Regulatory Commission, "Disposal of High-Level Radioactive Wastes in Geologic Repositories: Technical Criteria," *Fed. Reg.* **48**, (120), 28194–28230 (1983).

23. M. S. Tierney and R. P. Rechard, "Assessing Performance of Imprecisely Characterized Systems: A Mathematical Perspective," in *Proceedings of International Conference on Safety and Reliability* **1**, S69–S76 (Lisbon, Portugal, 1997).

24. R. A. Freeze and J. Cherry, *Groundwater* (Prentice-Hall, Englewood Cliffs, New Jersey, 1979).

25. J. M. Hammersley and D. C. Handscomb, *Monte Carlo Methods* (Wiley, New York, 1964).

26. "Risk: A Suitable Case for Analysis?" *Nature* **380**, 10–14 (1996) [and associated news articles].

27. A. Tal, "Assessing the Environmental Movement's Attitudes Toward Risk Assessment," *Environ. Sci. Technol.* **31**(10), 470A–476 (1997).

28. National Research Council, *Solid-Earth Sciences and Society* (National Academy of Sciences Press, Washington, D.C., 1993).

29. G. K. Gilbert, "The Inculcation of the Scientific Method by Example, With an Illustration Drawn from the Quaternary Geology of Utah," *Am. J. Sci.* **31**, 284–299 (1886).

30. T. C. Chamberlin, "The Method of Multiple Working Hypotheses," *Science* **148**, 754–759 (1965) [reprint of original (1890) article].

31. D. Johnson, "Role of Analysis in Scientific Investigation," *Geol. Soc. Am. B* **44**, 461–494 (1933).

32. J. Hoover Mackin, "Rational and Empirical Methods of Investigation in Geology," in C. C. Albritton, Jr. (ed.), *The Fabric of Geology*, (Freeman, Cooper & Company, 1963), pp. 135–163.

33. K. B. Krauskopf, "A Tale of Ten Plutons," *Geol. Soc. Am. Bull.* **79**, 1–18 (1968).

34. J. H. Shea, "A Tale of Ten Plutons: Discussion," *Geol. Soc. Am. Bull.* **79**, 1243–1244 (1968).

35. K. B. Krauskopf, "A Tale of Ten Plutons: Reply," *Geol. Soc. Am. Bull.* **109**, 1245–1246 (1968).

36. C. J. Mann, "Randomness in Nature," *Geol. Soc. Am. Bull.* **81**, 95–104 (1970).

37. G. G. Simpson, "On Randomness and Determinism: Discussion," *Geol. Soc. Am. Bull.* **81**, 3185–3186 (1970).

38. I. J. Smalley, "Randomness in Nature: Discussion," *Geol. Soc. Am. Bull.* **81**, 3191–3194 (1970).

39. C. J. Mann, "On Randomness and Determinism: Reply," *Geol. Soc. Am. Bull.* **81**, 3187–3190 (1970).

40. E. H. McKee and J. E. Conrad, "A Tale of 10 Plutons—Revisited: Age of Granitic Rocks in the White Mountains, California and Nevada," *Geol. Soc. Am. Bull.* **108**, 1515–1527 (1996).

41. B. Hanson, "A Tale of 10 Plutons—Revisited: Age of Granitic Rocks in the White Mountains, California and Nevada: Discussion," *Geol. Soc. Am. Bull.* **109**, 1631 (1997).

42. J. E. Conrad and E. H. McKee, "A Tale of 10 Plutons—Revisited: Age of Granitic Rocks in the White Mountains, California and Nevada: Reply," *Geol. Soc. Am. Bull.* **109**, 1631–1632 (1997).

43. G. G. Simpson, "Historical Science," in C. C. Albritton, Jr. (ed.), *The Fabric of Geology* (Freeman, Cooper & Company, 1963), pp. 24–48.

44. J. E. Campbell and R. M. Cranwell, "Performance Assessment of Radioactive Waste Repositories," *Science* **239**, 1389–1392 (1988).

45. G. Apostolakis, "The Concept of Probability in Safety Assessments of Technological Systems," *Science* **250**, 1359–1364 (1990).

46. K. E. Bullen, "Acceptance by Keith Edward Bullen, Arthur L. Day Medal," *Geol. Soc. Am. Bull.* **74**(2), 161–162 (1963).

47. J. D. Burchfield, *Lord Kelvin and the Age of the Earth* (Macmillan, New York, 1975).

48. W. M. Murphy, "Natural Analog Studies for Geologic Disposal of Nuclear Waste," *Technol. Today*, **1992** (June), 2–7 (1992).

49. R. C. Ewing. "The Role of Natural Analogues in Performance Assessment: Applications and Limitations," in J. S. Tulenko (ed.), *Proceedings of the Third International Conference on High Level Radioactive Waste Management* (American Nuclear Society, 1992), pp. 1429–1436.

50. W. Miller, R. Alexander, N. Chapman, I. McKinley, and J. Smellie, *Natural Analogue Studies in the Geological Disposal of Radioactive Wastes* (Elsevier, Amsterdam, 1994).

51. J. Janeczek, R. C. Ewing, V. M. Oversby and L. O. Werme "Uraninite and UO_2 in Spent Nuclear Fuel: A Comparison," *J. Nucl. Materials* **238**, 121–130 (1996).

52. J. A. T. Smellie, F. Karlsson, and W. Russell Alexander, "Natural Analogue Studies: Present Status and Performance Assessment Implications," *Contaminant Hydrol.* **26**, 3–17 (1997).

53. N. Oreskes, K. Shrader-Frechette, and K. Belitz, "Verification, Validation, and Confirmation of Numerical Models in the Earth Sciences," *Science* **263**, 641–646 (1994).

54. J. L. Younker and J. M. Boak, "Geological Models," *Science* **264**, 1065 (1994).

55. N. Oreskes, K. Belitz, and K. Shrader-Frechette, "The Meaning of Models," *Science* **264**, 331 (1994).

56. B. F. Molina, "Modeling Geology—The Ideal World vs. the Real World," *GSA Today* **6**(5), 8–13 (1996).

57. V. Baker, "Modeling Global Change: Why Geologists Should Not Let 'System' Come Between Earth and Science," *GSA Today* **6**(5), 8–11 (1996).

58. A. F. White and S. L. Brantley (eds.), *Chemical Weathering Rates of Silicate Minerals* (Mineralogical Society of America, Washington, D.C., 1995).

59. P. C. Lichtner, C. I. Steefel, and E. H. Oelkers (eds), *Reactive Transport in Porous Media* (Mineralogical Society of America, Washington, D.C., 1996).

60. C. M. Bethke, "The Question of Uniqueness in Geochemical Modeling," Geochim. Cosmochim. Acta **56**, 4315–4320 (1992).

61. P. C. Lichtner, "Scaling Properties of Time-Space Kinetic Mass-Transport Equations and the Local Equilibrium Limit," *Am. J. Sci.* **293**, 257–296 (1993).

62. C. Ludwig, W. H. Casey, and P. A. Rock, "Prediction of Ligand-Promoted Dissolution Rates from the Reactivities of Aqueous Complexes," *Nature* **375**, 44–47 (1995).

63. D. K. Nordstom, "Geochemical Transport Modeling," in *Sixth CEC Natural Analogue Working Group Meeting Proceedings, EUR 16761* (1996), pp. 243–254.

64. C. M. Bethke, *Geochemical Reaction Modeling Concepts and Applications* (Oxford University Press, Oxford, 1996).

65. D. Read, "Conceptualisation of Hydrogeochemical Systems Uncertainty, Bias and Subjectivity," in *Sixth CEC Natural Analogue Working Group Meeting Proceedings, EUR 16761* (1996), pp. 255–269.

66. I. Grenthe *et al.*, *Chemical Thermodynamics of Uranium* (North-Holland, Amsterdam, 1992).

67. C. Frondel, *Systematic Mineralogy of Uranium and thorium.* U.S. Geological Survey Bulletin 1064 (U.S. Government Printing Office, Washington, D.C., 1958).

68. R. J. Finch and R. C. Ewing, "Corrosion of Uraninite Under Oxidizing Conditions," *J. Nucl. Materials* **190**, 133–156.

69. F. Chen, R. C. Ewing, and S. B. Clark, "The Gibbs Free Energies and Enthalpies of Formation of Uranium (VI) Phases: An Empirical Method of Prediction," *Am. Mineral.* **84**, 650–664 (1999).

70. I. Casas, I. Perez, E. Torrero, J. Bruno, E. Cera, and L. Duro, "Dissolution Studies of Synthetic Soddyite and Uranophane," *SKB Technical Report* 97–15 (1997).

71. S. N. Nguyen, R. J. Silva, H. C. Weed, and J. E. Andrews, Jr.,

"Standard Gibbs Free Energies of Formation at 303.15 K of Four Uranyl Silicates: Soddyite, Uranophane, Sodium Boltwoodite, and Sodium Weeksite," *J. Chem. Thermodynam.* **24**, 259–276 (1992).

72. R. L. Hunter and C. J. Mann (eds.), *Techniques for Determining Probabilities of Geologic Events and Processes* (Oxford University Press, Oxford, 1992).

73. M. D. Siegel, H. D. Holland, and C. Feakes, *Geochemistry in Techniques for Determining Probabilities of Geologic Events and Processes* (Oxford University Press, New York) 185–206 (1992).

74. L. F. Konikow and J. D. Bredehoeft, "Groundwater Models Cannot Be Validated," *Advan. Water Resources* **15**(1), 75–83 (1992).

75. L. F. Konikow, "The Value of Postaudits in Groundwater Model Applications," in A. I. El-Kadi (ed), *Groundwater Models for Resources Analysis and Management* (Lewis Publishers, Boca Raton, Florida, 1995), pp. 59–78.

76. L. F. Konikow, "Predictive Accuracy of a Ground-water Model—Lessons from a Postaudit," *Ground Water* **24**(2), 173–184 (1986).

77. W. M. Alley and P. A. Emery, "Groundwater Model of the Blue River Basin, Nebraska—Twenty Years Later," *J. Hydrol.* **85**, 225–249.

78. L. F. Konikow and L. A. Swain, "Assessment of Predictive Accuracy of a Model of Artificial Recharge Effects in the Upper Coachella Valley, California," in E. S. Simpson and J. M. Sharp, Jr. (eds.), *Selected Papers on Hydrogeology* (International Association of Hydrogeologists, Washington, D.C., 1990), Vol. 1, pp. 433–449 (1990).

79. E. G. Reichard and J. K. Meadows, "Evaluation of a Groundwater Flow and Transport Model of the Upper Coachella Valley, California," *U.S. Geological Survey Water-Resources Investigations Report* 91-4142 (1992).

80. M. P. Anderson and W. W. Woessner, *Applied Groundwater Modeling: Simulation of Flow and Advective Transport* (Academic Press, San Diego, California, 1992).

81. T. W. Anderson, "Electric Analog Analysis of Ground-water Depletion in Central Arizona," *U.S. Geological Survey Water-Supply Paper* 1860 (1968).

82. L. F. Konikow, "Numerical Errors Associated with Modeling Transport and Matrix Diffusion" [Abstract], in *Eos Transactions AGU,* **78**(17), Spring Meeting Supplement S138 (1997).

83. A. Hedin, "Spent Nuclear Fuel—How Dangerous Is It? A Report from the Project Description of Risk," *SKB Technical Report 97-13* (1997).

84. Advisory Committee on Nuclear Waste, Nuclear Regulatory Commission, "A 'Road Map' to the ACNW's Recommendation for Time Span for Compliance of the Proposed High-Level Waste Repository at Yucca Mountain, Nevada," Letter Report to The Honorable Shirley Ann Jackson, Chairman, U.S. Nuclear Regulatory Commission, November 14, 1996.

85. Advisory Committee for Nuclear Waste, Nuclear Regulatory Commission, "Time Span for Compliance of the Proposed High-Level Waste Repository at Yucca Mountain, Nevada," Letter Report to The Honorable Shirley Ann Jackson, Chairman, U.S. Nuclear Regulatory Commission, June 7, 1996.

86. NAS, *Technical Bases for Yucca Mountain Standards* (National Academy Press, Washington, D.C., 1995).

87. J. W. Weaver and F. Porter-Locklear, "Estimating the Rate of Natural Bioattenuation of Ground Water Contaminants by a Mass Conservation Approach," *SIAM Rev.* **40**(1), 113–117 (1998).

88. J. Ziman, *Reliable Knowledge—An Exploration of the Grounds for Belief in Science* (Cambridge University Press, Cambridge, 1978).

89. C. Comar, "Bad Science and Social Penalties," *Science* **200**(16 June) (1978).

90. M. Granger Morgan, "Bad Science and Good Policy Analysis," *Science,* **201**(15 September) (1978).

91. R. B. Cumming, "Is Risk Assessment a Science?" *Risk Anal.* **1**(1) 1–3 (1981).

92. H. F. Martz and R. A. Waller, *Bayesian Reliability Analysis* (Wiley, New York, 1982).

93. M. Stone, "Chaos, Prediction, and Laplacean Determinism," *Am. Philos. Q.* **26**, 123–131 (1989).

94. J. Glimm, "Nonlinear and Stochastic Phenomena: The Grand Challenge for Partial Differential Equations," *SIAM Rev.* **33**(4), 626–643 (1991).

95. R. F. Costantino, R. A. Desharnais, J. M. Cushing, and B. Dennis, "Chaotic Dynamics in an Insect Population," *Science* **275**, 389–391 (1997).

96. E. Lorenz, "Deterministic Non Periodic Flow," *J. Atmos. Sci.* **20**, 130–141 (1963).

97. R. W. Hamming, *Numerical Methods for Scientists and Engineers* (McGraw-Hill, New York, 1962).

98. U.S. Environmental Protection Agency, "40 CFR Part 194: Criteria for the Certification and ReCertification of the Waste Isolation Pilot Plant's Compliance With the 40 CFR Part 191 Disposal Regulations; Final Rule," *Fed. Reg.* **61**, 5224–5245 (1996).