Model Validation From A Regulatory Perspective: A Summary

N. Eisenberg and M. Federline
U.S. Nuclear Regulatory Commission
Washington, DC 20555

B. Sagar and G. Wittmeyer
Center for Nuclear Waste Regulatory Analyses
6220 Culebra Rd.
San Antonio, TX 78238-5166

(U.S.A)

J. Andersson and S. Wingefors
Swedish Nuclear Power Inspectorate
Box 27106
S-102 52 Stockholm, Sweden

(Sweden)

Abstract

Building confidence in the mathematical models used in safety assessments of HLW repositories is necessary if such models are to be applied successfully to the support of regulatory decisions. Conventional comparisons of model predictions to actual performance are not possible for such models over the temporal and spatial scales of interest. Uncertainties inherent in describing and modeling complex natural and engineering systems, as well as their interactions, make it difficult to discriminate between inadequacies in the mathematical models and the inadequacies of input data. A successful regulatory strategy for model validation, therefore, must attempt to address and carefully document these difficulties. Regulatory model validation efforts should seek to provide a documented enhancement of confidence in the model in so far as the model is necessary to support regulatory decisions. This document summarizes an approach jointly developed by members of the staff of the U.S. Nuclear Regulatory Commission and of the Swedish Nuclear Power Inspectorate for the validation of models used to assess the performance of geologic repositories for high-level waste disposal. The terms "validation" and "confidence building" are used interchangeably throughout this summary. "Confidence building" reflects a recognition that full scientific validation of models for performance assessment of the repository system may be impossible and that the acceptance of mathematical models for regulatory purposes should be based on appropriate testing which will lead to a reasonable assurance that their results are acceptable.

INTRODUCTION

Performance assessment (PA) using mathematical models is a key component of the evaluation of the long-term radiologic safety provided by a geologic repository. There is general agreement in the international community that a repository for high-level radioactive waste (HLW), including spent nuclear fuel, will consist of multiple barriers, and that these barriers will each contribute to the safety of the overall system by providing some level of redundancy. The long-term performance of the overall repository system and its components will be demonstrated by using mathematical models. Before such models can be used in the regulatory process to demonstrate compliance with safety standards, some measure of credibility and confidence in these models must be demonstrated in a process transparent to the parties to the regulatory process. The terms "validation" and "confidence building" are used interchangeably throughout this summary. "Confidence building" reflects a recognition that full scientific validation of models for PA of the repository system may be impossible and that the acceptance of mathematical models for regulatory purposes should be based on appropriate testing which will lead to a reasonable assurance that their results are acceptable. In the regulatory context, the level of confidence required for a given model is necessarily a function of the importance of that model's results to safety decisions.

The literature on the validation of models used to assess the performance of HLW repositories is vast and often contradictory, clearly reflecting the disparate views of those scientists, engineers, and policymakers involved in HLW management. This document summarizes a proposed regulatory approach, jointly developed by staff members of the U.S. Nuclear Regulatory Commission (NRC) and the Swedish Nuclear Power Inspectorate (SKI), to the validation of models used to assess the performance of geologic repositories for HLW disposal. The results of this collaborative effort represent the views of the authors and will be presented in more detail in a separate report (U.S. NRC and SKI, in preparation). Repository developers should not view this summary or the more detailed document as guidance, but merely as a report on current thinking with a view to developing future guidance.

REGULATORY CONTEXT

Although the United States and Swedish regulatory structures are different, both rely on predictive models to demonstrate compliance, and both nations' regulatory programs share a common interest in the development of procedures for demonstrating the validity of PA models.

Under current NRC regulations[1], models used to predict future conditions and changes in the geologic setting of a repository must be supported by a combination of field and representative laboratory tests, along with monitoring data and natural analogue studies. NRC regulations also require a performance confirmation program, through which the adequacy of modeling assumptions and performance predictions is to be verified, to the extent possible. Compliance must be demonstrated for both the overall environmental standard established by the U.S.

---

[1] In the U.S., the Environmental Protection Agency (EPA) is responsible for establishing generally applicable environmental standards for the management and disposal of high-level radioactive wastes including spent nuclear fuel. Under the Energy Policy Act of 1992, the U.S. Congress directed EPA to promulgate health-based environmental standards applicable to a proposed repository at Yucca Mountain, Nevada, consistent with the recommendations of a National Academy of Sciences' study currently underway. The NRC is required to conform its technical regulations to final EPA standards.

Environmental Protection Agency (EPA), and with the quantitative criteria established by the NRC for the performance of key repository subsystems.

As yet, similar regulations have not been issued in Sweden. However, criteria under development are expected to follow the recommendations presented jointly by the nuclear safety and radiation protection authorities in the Nordic countries (Nordic, 1993). These recommendations call for models used in safety assessments to be validated as far as practicable based on laboratory data and field measurements from the HLW repository site and natural analogue experiments.

It is anticipated that U.S. EPA standards applicable to disposal of HLW at Yucca Mountain will be probabilistic in nature and require that uncertainties be considered explicitly in performance estimates. But, even when regulatory standards are deterministic (e.g., the Swedish standard is expected to be expressed in terms of dose without attached probability; similarly the NRC's subsystem performance requirements are deterministic), regulators require that the uncertainty in model predictions be estimated and presented either as a range or, more commonly, as a probability distribution. This means that even where conservatism is invoked, it is necessary that estimates be made of uncertainties introduced due to model structure and assignments of certain preferred values to parameters.

Under both U.S. and Swedish regulatory regimes, the long-term performance of a repository will be assessed using quantitative PA and modeling techniques. Evaluation of the adequacy of these assessments will not only check whether estimated radioactive releases comply with specified criteria, but must also ascertain whether essential physical and chemical processes and their interactions have been identified, described adequately, and addressed.

COMPONENTS OF PA MODELING

Modeling for assessing repository performance is closely tied to site characterization and repository design. Data from site characterization and design features are crucial, not only to the development of appropriate conceptual models, but also to the extraction of parameter values that are employed to obtain numerical estimates of repository performance. Development of the conceptual model or models is the first step in PA modeling. Conceptual model development includes determining the governing equations, the geometry of the system, initial conditions, appropriate boundary conditions, and level of detail. For most natural and many engineered systems, formulation of a single, acceptable conceptual model is difficult, if not impossible, to achieve. In most cases, several classes of conceptual models are derived that satisfy the known constraints to varying degrees. Formulation of conceptual models for the natural system introduces problems that may not be encountered for engineered systems. Engineered systems, within limits, can be designed to meet prescribed performance criteria; whereas, the geologic system may only be explored and characterized. Complete characterization, however, is never possible for large and complex natural systems. Because tests may perturb the very properties being measured and because of the possibility that destructive testing could impair the barrier properties of the site, site testing may be further constrained. The site conceptual model, therefore, is based on considerable extrapolation of sparse quantitative and qualitative data, which can give rise to large conceptual and parameter uncertainty. In view of this, it is imperative that alternate site models be formulated and tested to account for possible biases in conceptual model formulation.

Once the requisite conceptual models have been developed, mathematical models representative of each conceptual model must then be formulated and usually are implemented using computer codes. Mathematical models of the total system, and their corresponding computer codes, which include realistic details of all system components and which treat parameter and future states uncertainty, can become very complex and computationally onerous. Under such circumstances, it is logical and appropriate to perform modeling using a hierarchy of models (Figure 1). The very detailed, and more realistic, models of individual processes comprise the first level of this hierarchy and are useful for understanding the sensitivity of a process to parameter variations and external forces. These first-level models also may be used to demonstrate the conservatism of assumptions and to provide a basis for second-level models in the hierarchy. In the second level, a limited number of the detailed models, with some simplifications, are coupled with one another, to gain some understanding of the interfaces among processes. In the third and final level, all component models are further simplified and coupled to formulate a "total system PA (TSPA) model." These are the fast, efficient models required for a probabilistic treatment of performance. However, it must be kept in mind, that if the coupling among the detailed models is strongly nonlinear, then it may not be easy to ascertain whether assumptions for conservatism are valid for the coupled system. In addition, not all processes are reduced to the third level of simplicity for inclusion in the system model; some processes have such a strong effect on the final result that they must be included in full detail. When such a hierarchy of models is used in demonstrating compliance with performance requirements, then all parts of the hierarchy need to be evaluated to build the required confidence, even though the type and amount of testing for each level in the hierarchy will be different.

The NRC (U.S. Nuclear Regulatory Commission, 1992) and SKI (Swedish Nuclear Power Inspectorate, 1991) have conducted PA iterations applicable to the current disposal concepts under consideration in their respective countries. In SKI's Project-90 and NRC's Iterative Performance Assessment (IPA), the repository and the neighboring rock are termed the "process system" or "repository system." Classes of events and processes external to the repository system are referred to as "external events" or "external environment." External events or the external environment acting on the repository system gives rise to scenarios. The initial system description, at the time the repository is sealed, is called the undisturbed-, base-, or nominal-case. Future disturbances of this system as a result of either natural or human-initiated external events or ongoing processes may alter the boundary conditions on the system or may be so profound as to modify the system description and necessitate corresponding changes to the underlying conceptual model. Scenarios may be defined as "physically plausible sequences of events and processes [occurring in the external environment] that could lead to release and transport of radionuclides from a repository" [Nuclear Regulatory Commission, 1992]. However, all HLW programs do not define scenarios in the same manner. For purposes of model validation, a uniform definition of scenarios is less important than the fact that each scenario is associated with a conceptual model. Each of these conceptual models will require some degree of validation. Because the effects of each disruptive scenario on the repository may be manifest in a number of different ways, it may be prudent to thoroughly exercise the models that implement these scenarios to ensure that the resulting total system performance assessments are reasonable.

REGULATORY APPROACH TO MODEL VALIDATION

The primary objective of this proposed approach is to provide a means to establish the adequacy of a model for its intended use in the regulatory context. Although it has been prepared by regulators, it is intended primarily for application by repository developers.

HLW regulators will be responsible for determining compliance of a proposed repository with environmental standards and implementing criteria. For regulators in both the United States and Sweden, the test of compliance with regulatory criteria is "reasonable assurance." This concept recognizes that absolute assurance of compliance is neither possible nor required. Instead, an applicant must provide such information as may be necessary to convince a reasonable decision-

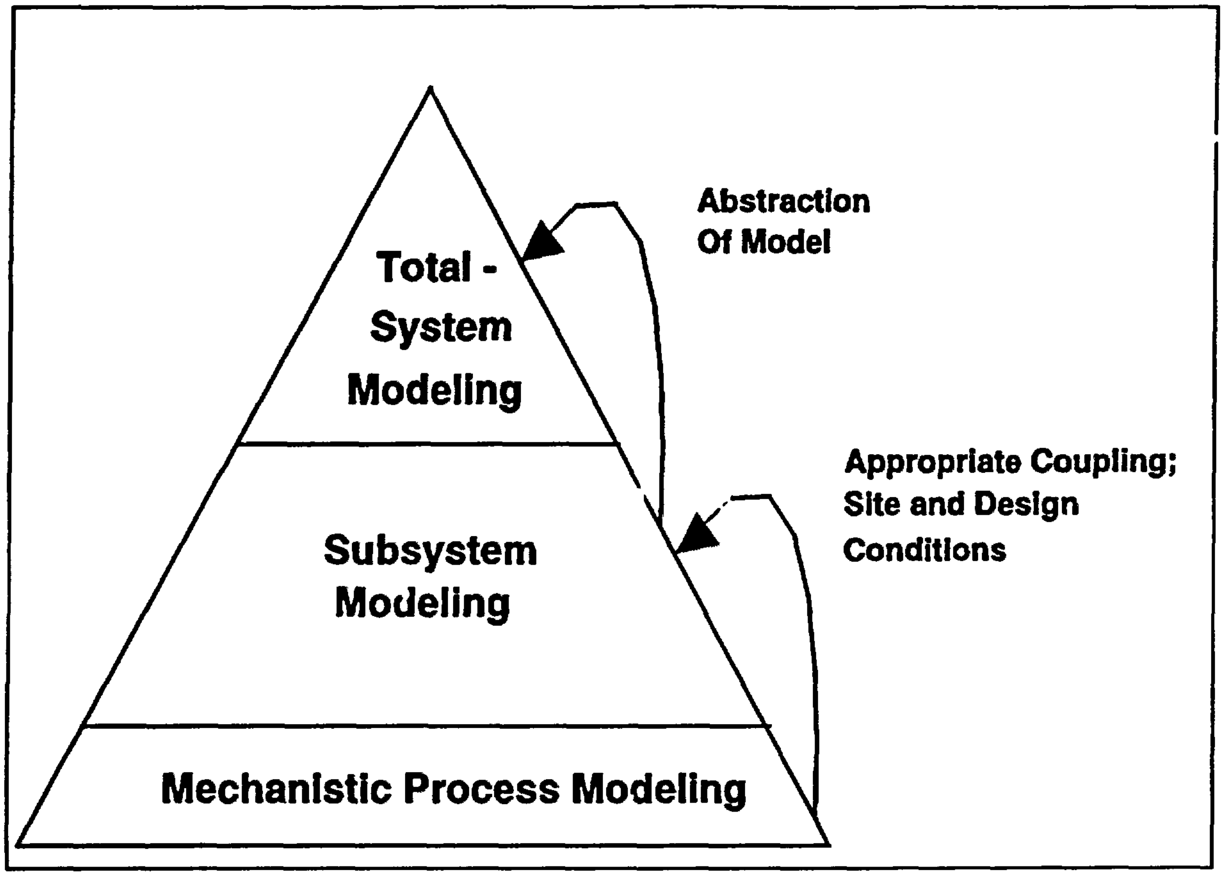

Figure 1. Hierarchy of performance assessment models

maker that compliance with regulatory criteria will be achieved. For these reasons, regulatory expectations for model validation are based on an applied science approach and differ from those appropriate to a purely scientific approach to developing and testing models. For example, a purely scientific approach compels pursuit of complete and detailed explanations for all observed phenomena independent of any particular model application. The regulatory approach requires only an adequate description of the phenomena for a given purpose (e.g., for the licensing of a repository) (Davis et al., 1991).

If, in the regulatory context, one assumes "validation" to mean demonstration that a model is sufficiently accurate for the purpose for which the model is used, there can be no standard answer to the question "How much validation is enough?" Rather, the answer will depend on the model itself and on its specific application. This does not imply that regulatory validation is entirely subjective. It is possible to envision a process or strategy whereby a repository developer and the regulator could reach an agreement on the degree of validation needed for each model used in a repository performance assessment.

Goals of a Validation Strategy

In devising such a strategy for validation of PA models, it should be made clear that the overall goals of validation are to: (i) establish whether the scientific basis is adequate for each model, and

(ii) demonstrate that each model is sufficiently accurate for the purpose for which the model is used by assuring that the scientific basis is correctly applied.

If a model is to be used as part of a demonstration of repository safety (or to challenge the projections of a model used for that purpose) it must be shown to have an adequate scientific basis. Speculative or conjectural models that have no plausible theoretical foundation or empirical basis will not demonstrate, with reasonable assurance, that repository performance will meet regulatory criteria. Thus, the minimum threshold to be achieved in validating PA models is to establish their scientific credibility.

Additionally, it must be demonstrated that any model used in a safety assessment is sufficiently accurate for the purpose for which the model is used. Implicit in this second goal is the need to validate each application of the model. The validity of a model prediction depends not only on the validity of the model, but also on the validity of the input parameters used with the model, the validity of any numerical implementation of the model, and the validity of interpretation of model projections. However, the strategy presented here focusses primarily on validating the model itself.

The repository developer should prepare a validation strategy describing the plans for validation of each model to be used as part of a repository performance assessment. A principal goal of this validation plan is to establish in a transparent fashion the process by which the repository developer will demonstrate a level of confidence in models consistent with their contribution to repository performance. In addition, the validation strategy will help to guide or focus the repository developer on formulating site characterization plans and in determining the performance goals for the components of the overall repository system. Validation strategies should be established in the early phases of a disposal program and updated as warranted. For programs well under way, this issue should be assigned a high priority.

For certain components of the repository system, a regulator may elect to develop his own PA models and may, therefore, need to establish an independent strategy for their validation. However, since the purpose of such models is not to demonstrate the safety of the repository system, but to probe and evaluate the projections of the developer's models, the goals of the regulator's validation strategy may be less ambitious than those of the developer. In many cases, the regulator only needs to establish the scientific credibility of its models so that the projections of those models can be compared to the projections of the developer's models. It should be recognized, however, that the regulator, in addition, will have to develop competence and procedures for review of the licensees compliance demonstration in this area. Any guidelines or rules for a validation strategy will have to be decided by the regulating body before the last phase of the licensing procedure.

A Strategy for Developing Confidence in Models

Shown in Figure 2, is a strategy for developing confidence in models that are used in HLW PA to demonstrate compliance with safety criteria. Implementation of this approach requires certain decisions, feedback, and iteration between a number of steps as follows:

(1) Define a compliance demonstration strategy that allocates performance of modeled systems and defines goals for model validation

(2) Determine the documented support that exists for each model necessary to demonstrate compliance

(3) Compare the validation goals to the existing support

(4) Decide whether to revise or retain the compliance strategy

(5) Obtain additional support for the model

Each of these steps is discussed briefly below and developed more fully in the a joint NRC/SKI white paper currently in preparation.

Define Compliance Demonstration Strategy

The confidence building process should begin with the development of an overall strategy for demonstrating compliance with the regulations. This strategy should identify the quantitative post-closure performance objectives and include plans for demonstrating that a repository will meet these objectives. The strategy is developed by taking into account repository design information and site information.

The quantitative performance objectives for a repository should be stated in terms of specific performance measures. A performance measure is a physical quantity, which depends on the long-term behavior of the repository and which indicates how well the repository isolates the radioactive waste from the environment or how well the environment is protected. Examples of performance measures include the concentration of radionuclides in groundwater or the dose to the maximally exposed individual. Minimum or maximum allowable values of the performance measure are identified as performance limits. Sometimes the performance measure may be estimated by a suite of linked computer codes, which represent models for various components of the repository or the environment. The resulting estimate of a given performance measure can then be compared to the performance limit for that component.

The repository developer may choose to include only certain components in a demonstration of compliance, either because (i) the components are very robust and drive the system performance measure well below the maximum limit allowed, or (ii), the components do not significantly add to the uncertainty in the estimate of performance. These choices of which components to include in models for demonstrating compliance comprise the compliance demonstration strategy. The degree to which each component is necessary for demonstrating compliance constitutes the allocation of performance to that component. This allocation of performance determines the level of validation required for a particular model. Models representing those components central to a compliance demonstration necessarily will be subject to a greater degree of validation than those representing more peripheral components. Although performance allocation can be described quantitatively, it is doubtful that a universally applicable, quantitative measure of model validity can be devised. Performance allocation should be used to establish semi-quantitative goals for the desired level of validity of various models. These goals might be expressed as the rank ordering of importance of the models or as a small number (2-5) of categories representing smaller or greater need for validation.

An iterative process is best suited for determining acceptable levels of performance for each component of an overall repository system. Based on the safety standards for the performance of the entire repository system, an initial design is made before site characterization begins. This initial design should describe: (i) the level of performance expected for the natural barriers, (ii) the level of performance to be obtained from engineered barriers, (iii) the level of confidence anticipated for each projection of performance, and (iv) the safety factors, margins for error, or redundancy among barriers (if any) to be incorporated into the design of the repository system. Descriptions of the desired level of confidence for projecting performance should include a

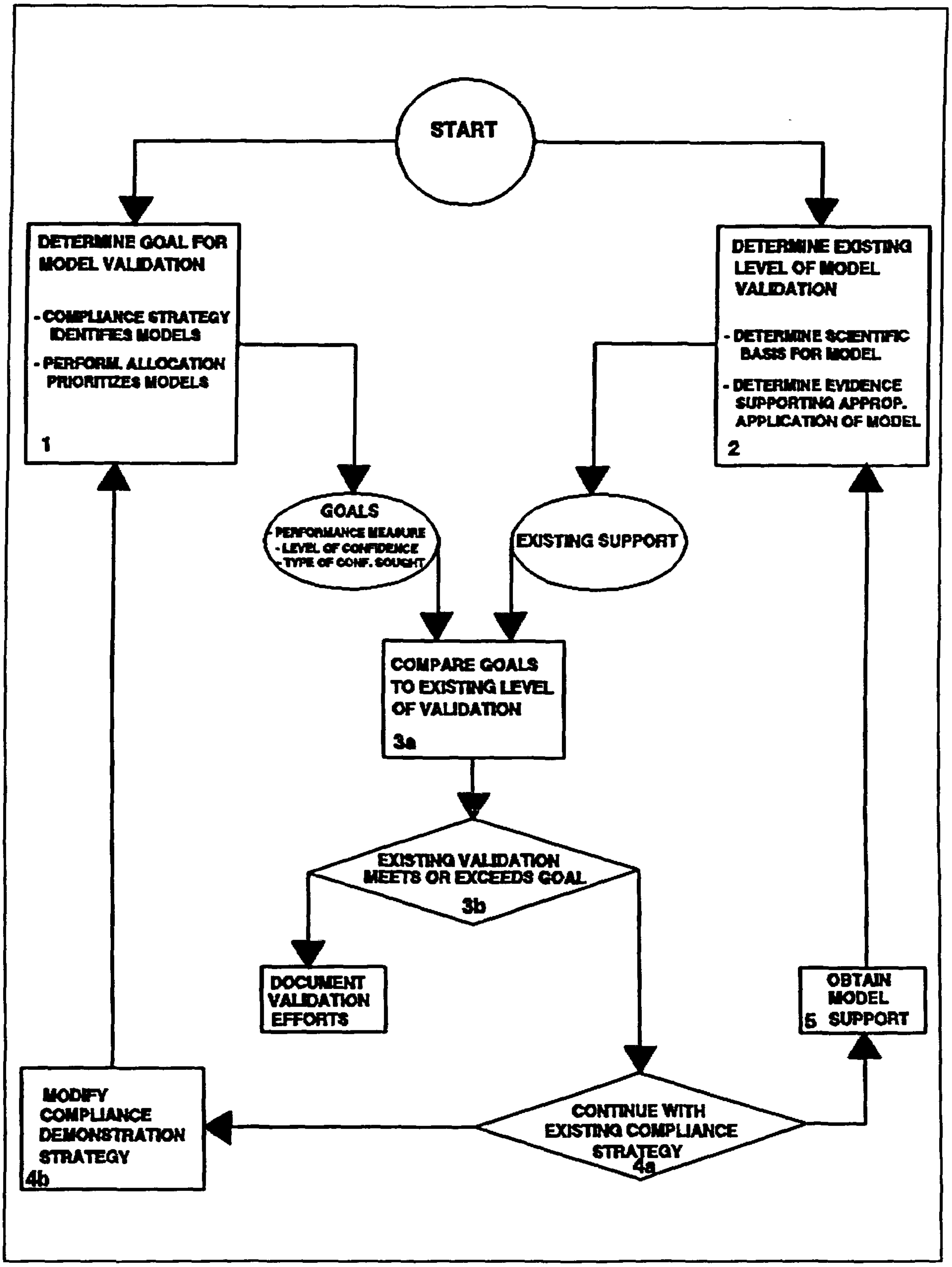

Figure 2. Regulatory strategy for developing confidence in models

discussion of the repository developer's plans for validating any models employed as well as for obtaining peer and regulator review of the results. Each IPA should be followed by a systematic review resulting in updated judgments of the relative importance of various submodels and assumptions.

Determine the Existing Level of Support for the Model

Models used in PA must be supported by a sound and well-documented scientific basis if regulators and the public are to have confidence in modeling results. It is essential to identify which aspects of a given model are based on accepted science, and which lack credibility or are potential sources of uncertainty. Having identified the models and performance measures of interest, an effort should be undertaken to review and assimilate the scientific literature relevant to the use of the models for the applications of interest. Because many studies might have employed a particular model, but may not have applied it in the same manner or for the same purpose, it may be necessary to reevaluate and/or recompute experimental or theoretical results, in order to apply the model to the compliance measure in question and subsequently arrive at a judgment as to what level of model support is exists.

In order to establish scientific support for a given model, it is necessary to both examine its theoretical basis and to evaluate the application of scientific principles in the model to assure the application is appropriate. Normally, the validity of the application of principles is achieved by comparing the model predictions against empirical information. For models used in HLW disposal, the evaluation of the application of scientific principles and the comparison of model results to empirical data is limited, because each site is unique. This limits the modelers' ability to extrapolate from one site to another. Another limitation on the use of empirical data is that experimental data collected over the time scales of interest are not available. Evidence from field data and natural analogue studies are available for longer time and spatial scales, but there is considerable uncertainty about the environmental conditions prevailing for the system over these long times.

Compare the Validation Goals to the Existing Support

If the existing support for the model exceeds that required to demonstrate compliance, then no further validation activities are needed. If the existing support is insufficient, then it is either necessary to reduce the reliance placed upon the modelled component and revise the compliance strategy accordingly, or to acquire additional support for the model. Reducing reliance on one component, in general, can be expected to require increased reliance on one or more other components.

Decide Whether to Revise Compliance Strategy

Most proposed repository systems should have sufficient margins-of-safety or redundancies such that several different combinations of system components assigned different relative priorities may be included in a successful demonstration of compliance. Overall safety criteria may be fixed, but there may be several legitimate means of demonstrating safety with those criteria, each of which may be associated with some degree of uncertainty. If it is decided to revise the compliance demonstration strategy, a new set of components must be selected, or new priorities assigned to the old ones. This will require, in turn, that new models with different performance measures be identified, or different priorities assigned to the old models. In either case, the evaluation and comparison with the existing level of support will need to be revisited. If the original compliance strategy is to be retained, then additional technical support must be acquired.

For example, if the Swedish HLW program were to decide that descriptions of far-field migration are very uncertain, it could elect to place greatest reliance on the long-term stability of the canister. However, if subsequent study of the validity of canister stability models cast doubt on their validity, more confidence might then be placed on the retarding mechanisms of the far-field (i.e., change the compliance strategy). Alternatively, it might be decided to conduct further research on canister corrosion phenomena (i.e., obtain further model support).

A similar hypothetical example, applicable to the proposed U.S. repository at Yucca Mountain, could be postulated such that only transport in the unsaturated zone would be modeled because the cumulative releases of radionuclides from the unsaturated zone to the saturated zone can be shown to meet the regulations. Conversely, if water transport in the saturated zone was shown to be slow, it could be assumed that radionuclides are released directly into the saturated zone, and the more complex modeling of the unsaturated zone transport could be avoided.

Modification of a compliance demonstration strategy may reduce the need for additional model support. However, in most cases, additional support will be required for adequate decision making.

Obtain Additional Model Support

The support for a given model may be bolstered by theoretical advances, the acquisition of additional laboratory or field data, or from further study of appropriate natural analogues. Generally speaking, some measure of theoretical support is required for all models used in PA and, in some cases, theoretical support may be used to substitute for experimental evidence. Virtually all the models used in PA are based on well-established scientific principles, such as conservation of mass, momentum, and energy. Difficulties arise in applying these principles to complex situations, such as the flow of water in heterogeneous, partially-saturated, fractured rock. Nevertheless, extensive theoretical analyses, with evaluative experimental studies, are available on topics and systems relevant to nuclear waste disposal. To the extent that the scientific basis is well-established for both the fundamental theory and the application of that theory to processes, phenomena, or systems related to nuclear waste, this information (which has incorporated previously obtained empirical results) may be substituted for experimental support for validation. This type of evidence, when presented in a logical fashion, may be especially useful in supporting claims that a particular model is conservative in a given application. Some models, for example, very simple "models," that consist solely of correlations of variables in experiments of limited scope, require further theoretical substantiation, if model results are to be extrapolated to times or conditions not encompassed by the original data.

Progress in validation may also require additional experimental evidence. This can be the case for generic issues, like coupled near-field phenomena, as well as for cases where the validation issue is to show that a particular process or structure is applicable to a specific site. Additional experimental evidence may be used to add confidence to either the scientific basis for the model or the particular application of the model. In the latter case, it may be necessary to show how well characterization measurements actually characterize the site. This will generally require experiments both at the actual site, as well as at other sites, in order to confirm the reliability of the site characterization techniques. In addition, confirmation of the application of established principles may be directed to the interpretation of site characterization data, especially as they pertain to conceptual models. Experimental evidence is more likely to be needed to support the scientific basis for a model for areas in which the theoretical and empirical bases are incomplete or still evolving (e.g., models of the formation and migration of colloidal contaminants in the geosphere).

When planning experiments or field studies to support a model, it should be stressed that the experiments should be planned based on a systematic analysis of their potential for resolving the identified problems. In such planning there are some "good practices" which may prove helpful. These practices, which can be applied to the collection of laboratory, site, and natural analogue data, include the following:

(1) Identify potential alternative models and then design tests that will discriminate among the various alternatives and the preferred model, if there is one.

(2) Design experiments that will enhance the fundamental understanding of important processes included in a model. A suite of experiments carried out on different scales, if achievable, will add confidence.

(3) To the extent practical, design experiments that test models over the type and range of conditions for which the models will be used. When it is impractical to test over the full range, as will usually be the case for repository models, means to expand the data base (accelerated testing) or to scale the data (e.g., by using dimensionless numbers) should be used with great caution. Since many of the phenomena of interest are dependent on scale and/or experimental conditions, simple relationships for scaling or extension of data may be unusable. Tests should be designed to identify the conditions for which model results will be invalid.

(4) Various scenarios may be represented by different boundary conditions applied to existing models or, in the case of profound changes produced by the scenario, completely different models. A complete program for building confidence in models must treat these different conditions and/or models to the extent that they contribute to the overall measure of safety of the repository.

(5) If a model intended for predictions over long times and large spatial scales fails to predict processes accurately over shorter time and smaller spatial scales, this is strong evidence that the model may not be valid.

(6) When analyzing the results of an experiment it is recommended that subsets of data not be excluded for arbitrary reasons; all relevant data should be used to evaluate the accuracy of the model and to evaluate the potential for errors and biases introduced by the experimental technique. However, only the data relevant to the predictive model under study need be used; and only the phenomena and variables of interest need be explained by the model.

(7) Agreement with an experiment is insufficient, alone, to validate a model; the scientific basis for the model must be supported and scrutable. One must ensure that generally accepted scientific principles (e.g., those describing flow of groundwater through a porous matrix) are applied to model the conditions anticipated for a specific repository.

(8) In general, if a single model is divided into two or more submodels, the degree of confidence imparted by evaluating the submodels individually will not be as great as the degree of confidence achieved by evaluating the submodels linked together. Therefore, it is desirable to perform additional tests designed to validate the combination of submodels, as the combination will be used for repository performance assessments. In the absence of the practical ability to perform tests for combinations of submodels, careful theoretical evaluation is called for.

Experimental data may not be necessary for validation of PA models in all cases. Applying the concept of model hierarchy discussed earlier, the theoretical and empirical basis for more detailed models may be sufficiently well-established, such that one may be able to use theoretical arguments, based on the detailed models, to confirm the validity of the abstracted models. This is limited by the degree to which the model of interest is coupled with other models. Such an approach may be especially useful for conservative or bounding assumptions used for the modeling of engineered components of the repository. In such cases, the support sought may be gained by development of a deeper understanding of the process in question than is actually needed in the PA model. Such an understanding can be attained simply by more detailed and careful modeling of important processes.

Due to scale and time limitations associated with conducting laboratory and field experiments, their usefulness in model validation is limited mainly to understanding the processes at work in the real system; however, if a model fails to agree with experiments conducted over limited scales, chances are small that it will be satisfactory at larger scales. Comparing PA subsystem and complete system models to laboratory and field test data is a much more difficult task, since the function of these models is to predict system performance over large spatial and time scales. In general, there will be more confidence in a model which compares favorably to several types of evidence covering a range of spatial and time scales, such as laboratory tests and natural analogues.

Laboratory experiments are useful because they can be performed in a controlled environment that minimizes uncertainty in initial and boundary conditions, and experiments can utilize samples that exhibit relatively little geometric variability or whose variability can be measured. However, the use of laboratory experiments in validation efforts is limited by the inability to perform tests on the long time and large spatial scales representative of a HLW repository, difficulties in testing some coupled processes, and the possibility that the systems used are not representative of *in situ* conditions (e.g., samples damaged in collection, not enough samples collected to characterize spatial variability, laboratory conditions inconsistent with field conditions which may produce phenomena that do not actually occur in nature) (Davis et al. 1991).

Field experiments overcome, to a degree, the problem of whether data is representative and some of the scale problems that plague laboratory experiments. To a certain extent, field experiments can be direct surrogates of repository performance (e.g., field heater tests and tracer tests). However, the usefulness of field experiments is limited by uncertainties in initial and boundary conditions and, to a large degree, by the possible conceptual misunderstanding of field conditions (Davis et al. 1991). Nevertheless, field experiments are necessary tools for site characterization and a thorough understanding of their potential usefulness and limitations is certainly warranted.

Lastly, evidence from appropriate natural analogues have value for increasing the temporal and spatial scales available for experimental study. In some sense, nature could be considered to have initiated experiments that could be used for validation. The transport of radionuclides from uranium deposits and the transport and deposition of minerals along fractures are two examples. These "experiments" have the advantage of having taken place on temporal and spatial scales that are comparable to those relevant to HLW repository systems. In addition, coupled processes are often involved that are difficult to reproduce in either the laboratory or the field. Uncertainty in initial conditions, boundary conditions, and the temporal evolution of the physical system, however, limit the usefulness of natural analogues in validating models (Davis et al. 1991). Another potential drawback of natural analogues is their complexity. It is difficult to demonstrate a sufficient understanding of an analogue, with inferred historic evolution, in order to support claims that the studied analogue is directly relevant to the system or subsystem modelled.

(9) Data used to develop or calibrate a model cannot be used alone to validate that model. Model calibration is performed to demonstrate that the model is consistent with the system being modeled. Validation on the other hand, is the testing of the models ability to simulate the same system under different conditions. Thus, at least two data sets (or a partitioned set of data) are required for model validation.

(10) If a model is intended to be conservative rather than realistic (i.e., to overestimate potential repository impacts), tests, or proofs, should be designed to verify that the model is, in fact, conservative.

(11) Accurate records of model development and testing should be maintained, subject to periodic peer reviews during the development and testing process. These records should include the analyses and rationale supporting the decision to accept or reject the plausibility of various conceptual models.

With all their limitations, long-term field experiments and natural analogue studies nonetheless make it possible to study complex, coupled systems. This constitutes an invaluable check that no essential process or coupling effect has been omitted. Furthermore, even a qualitative fit between standard PA models and results from long-term experiments provides increased confidence in the model.

DOCUMENTING STATEMENTS OF VALIDITY

Experiments, derivation of models and assumptions, compliance demonstration strategies, and other reasoning should be openly and clearly documented in order to make possible a comprehensive review of the validation strategy. Thorough and systematic review, possibly including international cooperative efforts, is fundamental in judging validity. However, it should be emphasized that review efforts will be productive only when the material to be reviewed is supported by quantitative analyses of experiments or other proofs, or is derived from interpretation of natural analogues or more detailed modeling.

In a license application, the documentation of the validation strategy will be of great importance in judging the credibility of safety analysis calculations. The overall objective of this documentation should be development of a framework which facilitates the acceptance (or rejection) of models, based on transparent and logical reasoning.

SUMMARY

Model validation has been a topic of debate in HLW management for several years. There is considerable disagreement within the technical community about what constitutes model validation and how to achieve it. Furthermore, the process for validating PA models is likely to be a lengthy one, requiring laboratory and field experiments and natural analogue and theoretical investigations. It is important that PA be viewed as an applied science, rather than a pure science, and that the models used need only be valid enough to provide predictions useful for their intended application. An approach to model validation has been presented which outlines an iterative process of defining and redefining compliance demonstration strategies based upon the extent of scientific support available or reasonably attainable throughout the characterization of a potential repository. A number of good practices are advocated that may assist the repository developer in building confidence in PA models and in their interpretation of their results.

REFERENCES

Davis, P.A., N.E. Olague, and M.T. Goodrich. 1991. *Approaches for the Validation of Models Used for Performance Assessment of High-Level Nuclear Waste Repositories*. NUREG/CR-5537, SAND90-0575. Washington, DC: U.S. Nuclear Regulatory Commission.

Environmental Protection Agency. 1993. Environmental radiation protection standards for management and disposal of spent nuclear fuel, high-level, and transuranic radioactive wastes. Code of Federal Regulations 40, Part 191. Washington, DC: Environmental Protection Agency.

Nordic Document. 1993. *Disposal of high-level radioactive waste. Consideration of some basic criteria*. The Radiation Protection and Nuclear Safety Authorities in Denmark, Finland, Iceland, Norway, and Sweden.

Swedish Nuclear Power Inspectorate. 1991. *SKI Project-90*. Stockholm, Sweden: Swedish Nuclear Power Inspectorate: SKI TR 91:23.

U.S. Nuclear Regulatory Commission. 1983. *10 CFR Part 60; Disposal of High-Level Radioactive Wastes in Geologic Repositories—Technical Criteria*. Federal Regulation 48: Washington, DC: U.S. Nuclear Regulatory Commission: 28,194–28,229.

U.S. Nuclear Regulatory Commission. 1983. Final Technical Position on Documentation of Computer Codes for High-Level Waste Management. NUREG-0856. Washington, DC: U.S. Nuclear Regulatory Commission.

U.S. Nuclear Regulatory Commission. 1992. *Initial Demonstration of the NRC's Capability to Conduct a Performance Assessment for a High-Level Waste Repository*. NUREG-1327. Washington, DC: U.S. Nuclear Regulatory Commission.

U.S. Nuclear Regulatory Commission and Swedish Nuclear Power Inspectorate. *Model Validation from a Regulatory Perspective: A White Paper*, in preparation.