MAR 12 1990

MEMO FOR LINEHAN

MEMORANDUM FOR: John L. Linehan, Director
Repository Licensing and Quality Assurance
Project Directorate, HLWM

Joseph O. Bunting, Chief
Engineering Branch, HLWM

Ronald L. Ballard, Chief
Geosciences & Systems Performance Branch, HLWM

FROM: Robert E. Browning, Director
Division of High-Level Waste Management, NMSS

SUBJECT: REQUEST FOR ACCELERATED REVIEW OF SANDIA DRAFT REPORT (FIN A1165) FROM SUBTASK 2.4a

I am requesting your assistance in providing review comments from your branch on the enclosed draft report by close of business, Monday, March 19, 1990. The report is "Guidelines for Judging the Validity of Models for Performance Assessment," a deliverable under Subtask 2.4a of FIN A1165.

The report provides guidelines for model validation developed by Sandia National Laboratories (SNL) staff as performance assessment tasks were done. The report assumes that validation of conceptual models is the central issue in model validation. It forms part of a series of reports which are intended to transfer SNL performance assessment technology to the staff and the Center for Nuclear Waste Regulatory Analyses. In this report, guidelines for reducing conceptual uncertainty are treated in some depth. A more general uncertainties report, now being completed, discusses the relationship between conceptual model uncertainty and other uncertainties.

As stated in my memorandum to your dated December 27, 1989, it is necessary to accelerate the reviews of reports from Sandia Contract FIN A1165 because this long-term contract with SNL will soon be coming to a close. This is the second of five draft reports and five final reports that are expected within a three month period. The report is short and can be reviewed readily within five working days.

9003150073 900312
PDR WMRES EXISANL
A-1165 PDC

Delete: ACNW

WM-11
A-1165
NH18

The PPSAS number for this review is 411224 and the TACS number is L62213, LAM: Review SNL Subtask 2.4. Please send your comments to Pauline Brooks (extension 20404).

/S/

Robert E. Browning, Director
Division of High-Level Waste Management, NMSS

Enclosure:
As stated

cc: D. Tiktinsky, PI/NMSS

DISTRIBUTION

Central Files w/ENCL.	HLGP r/f w/ENCL.	NMSS r/f w/ENCL.
RBrowning, HLWM	BJYoungblood, HLWM	JBunting, HLEN
RBallard, HLGP	JLinehan, HLPD	DChery, HLGP
PJustus, HLGP	SCoplan, HLGP	RJohnson, HLPD
JKennedy, HLPD	RWeller, HLEN	MNataraja, HLEN
PBrooks, HLGP		

OFC :	HLGP	:	HLGP	:	HLGP	:	HLWM	:		:
NAME:	PBrooks/cj	:	SCoplan	:	RLBallard	:	REBrowning	:		:
Date:	3/9/90	:	3/9/90	:	03/12/90	:	3/12/90	:	/ /90	: / /90

OFFICIAL RECORD COPY

NUREG/CR-xxxx
SAND90-xxxx

GUIDELINES FOR JUDGING THE VALIDITY OF MODELS FOR PERFORMANCE ASSESSMENT OF HLW REPOSITORIES

Paul A. Davis

February 1990

Sandia National Laboratories
Albuquerque, NM 87185
Operated by Sandia Corporation
for the
U.S. Department of Energy

Prepared for
Division of High-Level Waste Management
Office of Nuclear Materials Safety and Safeguards
U.S. Nuclear Regulatory Commission
Washington, DC 20555
NRC FIN A1165

ABSTRACT

The purpose of this report is to provide general guidelines to be applied in judging the validity of models used in performance assessment of HLW repositories. The guidelines are based on a validation strategy that Sandia National Laboratories (SNL) has proposed to the Nuclear Regulatory Commission's Office of Nuclear Regulatory Research. This strategy focuses on the demonstration that performance assessment models are adequate representations of the real systems they are intended to represent given the pertinent regulatory requirements rather than proving correctness from the purely scientific point of view. This report discusses (1) the definition of validation in the context of performance assessment for HLW repositories, (2) the need for validation, (3) a validation strategy, and (4) an approach to validation proposed by SNL.

FOREWORD

This report presents the views of the author and Sandia National Laboratories on the validation of models for performance assessment of HLW repositories. These views are not necessarily those of the Nuclear Regulatory Commission (NRC). Therefore, it should not be construed that following the guidelines presented in this report will result in acceptance of specific performance-assessment models by the NRC in the licensing process.

CONTENTS

		Page
1.	INTRODUCTION	1
2.	DEFINITION OF VALIDATION	3
3.	NEED FOR VALIDATION	7
4.	VALIDATION STRATEGY	8
	4.1 Use of Laboratory Experiments	8
	4.2 Use of Field Experiments	9
	4.3 Use of Natural Analogs	9
5.	PROPOSED APPROACH TO VALIDATION	10
	5.1 Definition and Prioritization of Models to be Validated	10
	5.2 Identification and/or Definition of Relevant Experiments	11
	5.3 Definition of Performance Measures that are Based on the Intended Use of the Model	13
	5.4 Quantification of the Uncertainty Associated with the Input Data and the Data Available for Comparison with the Model Output	14
	5.5 Definition of the Acceptance Criteria or Acceptable Model Error Based on Regulatory Requirements and Uncertainty in Experimental Results	14
	5.6 Simulation of the Experiment	19
	5.7 Experimentation	20
	5.8 Evaluation of Model Results Based on the Acceptance Criteria	20
6.0	SITE-SPECIFIC MODEL VALIDATION	22
	6.1 Agreement of the Model with Site-Specific Information	22
	6.2 Justification of Assumptions	23
	6.3 Treatment of Multiple Conceptual Models	23
7.0	CONCLUSIONS	25
	REFERENCES	

SUBTASK 2.4(A) FORMAL REPORT

GUIDELINES FOR JUDGING THE VALIDITY OF PERFORMANCE ASSESSMENT MODELS

1.0 INTRODUCTION

In order to obtain a license for a high-level radioactive waste (HLW) repository, the Department of Energy (DOE) will carry out a program of site characterization and performance assessment analyses to demonstrate with "reasonable assurance" that burial of HLW wastes will pose no undue risk to the public health and safety. No direct means of assessing the behavior of a repository system exists due to the size of the area of regulatory concern (100 square kilometers) and time scales involved (10,000 to 100,000 years). Therefore, it is expected that DOE will use a wide variety of mathematical and numerical models, and associated computer codes, to complete the required analyses. Rendering the analyses defensible will require the models and data used for the performance assessments to be representative of the repository system, including the engineered barrier and the natural barrier. DOE is responsible for ensuring the validity of the models and codes used, the quality of data, and the adequacy of the overall analyses.

The Nuclear Regulatory Commission (NRC), on the other hand, is responsible for evaluating DOE's license application and deciding whether or not to grant such license. Therefore, the NRC must be able to judge whether or not DOE has demonstrated, with reasonable assurance, that the

DOE models are representative of the repository system conditions. Sandia National Laboratories (SNL) has been responsible for developing and testing models that the NRC could use in an independent evaluation of a HLW repository. Therefore, the NRC requested that SNL provide them with this document on model validation. The intended use of this document is to provide the NRC with the author's philosophy on model validation and insights into the validation process.

2.0 DEFINITION OF VALIDATION

There are several definitions for the term "validation;" following are several examples. NRC (1984) defined validation as the process of obtaining "assurance that a model, as embodied in a computer code, is a correct representation of the process or system for which it is intended". DOE (1986) defines validation as "a process whose objective is to ascertain that the code or model indeed reflects the behavior of the real world". Others (e.g., Borgorinski and others, 1988) follow the International Atomic Energy Agency's definition that validation is confirmed when the model and computer code "provide a good representation of the actual processes occurring in the real system". All the definitions are consistent, although derived from different perspectives. That is, they all are concerned with providing assurance that a model represents reality. This is different from the classical scientific approach which consists of proposing a hypothesis and then designing tests to disprove that hypothesis. Thus, science is not concerned with validating models but, rather, invalidating them. A model can be declared "invalid" or "not invalid"; however, it is the premise of this report that one can never say for sure that a HLW performance assessment model is "valid". More than 30 years ago Popper (1958) perceived that "...whenever we propose a solution to a problem, we ought to try as hard as we can to overthrow our solution, rather than defend it." This idea also has a basis in ordinary statistics whereby one proposes an idea, then attempts to disprove it by testing the null hypothesis. This philosophy has a counterpoint in judicial history where the null

hypothesis is that one is "not guilty". A jury must either fail to reject the null hypothesis (the individual is "not guilty"), or reject the null hypothesis (not "not guilty"). Clearly, just because an individual is judged not guilty does not mean that the person is innocent. It simply means that there was enough evidence for the jury to fail, beyond a reasonable doubt, to reject the null hypothesis. Hence, from a regulatory perspective, declaring a model not invalid does not guarantee validity. It only provides a means for the modeler to demonstrate reasonable assurance that the model is not incorrect.

The desire for validated models arises from a decision-making framework, either for designing a repository or for providing assurance that the assessment of long-term repository performance is meaningful. In licensing a HLW repository, the distinction between a scientific approach to developing and testing models and a regulatory approach to validating models is critical. For one, the scientific approach generally would ask for a complete and detailed explanation for all observed phenomena, whereas the regulatory approach would ask only for an adequate description of the phenomena for a given purpose (e.g., for the licensing of a repository). Thus, a bounding or conservative model may be valid for regulatory purposes but, by definition, not provide a detailed description of all phenomena. This is not to say that the scientific processes is not followed in the validation of models used in the regulatory arena. However, the final decision as to the validity of a model will be based on regulatory, not scientific considerations.

Performance assessment models are a combination of the site-specific data, that describe a particular geologic setting or experiment, and the physical process models contained in the relevant computer codes. Hence, the term model, as defined herein, includes the conceptual model, the mathematical model, the computer code and its associated input data. Therefore, validation of performance assessment models is ultimately a site-specific issue. However, the testing of models of various geologic media will aid in gaining (or losing) confidence that the process models contained in the codes are adequate representations of what is occurring in nature.

Both NRC and DOE define validation as a process and not as a fixed product. Thus, the goal of a model-validation exercise should not be viewed as providing a set of "validated" models. Rather, the goal should be obtaining added assurance that the models are able to simulate the behavior of real system. Thus, the regulatory focus should be on: 1) well-defined problems that test key processes, with less concern placed on the type and location of the specific media involved in a given experiment and; 2) site-specific application of the models to the proposed repository system.

The final point to make about the definition of validation is that NRC and DOE are not asking for absolute proof that models are perfect representations of reality. Instead, they are asking only that assurance be provided that the models are adequate representations of the real

system (i.e., that compliance with current regulations can be shown). The decision of model validity should be based on:

(1) whether or not the types of validation tests are relevant to the intended use of the model;

(2) how well the models were able to simulate the validation tests;

(3) how many validation tests are sufficient before the models can be applied to a particular site; and

(4) how well the site-specific information conforms with the model's description of the site.

It cannot be overemphasized that model validation is a site-specific issue based on the intended use of the model. Moreover, the modeler and regulator alike should realize that the ultimate decision on model validity will be, to a large degree, subjective and will be based on an acceptable level of uncertainty.

3.0 NEED FOR VALIDATION

The need for model validation in the U.S. HLW program arises from at least two sources. First, NRC and DOE have interpreted the need for validation to be implicit in the U.S. Environmental Protection Agency (EPA) and NRC regulations (40 CFR Part 191 and 10 CFR Part 60, respectively) that require DOE to provide assurance that the proposed site will meet the stated performance criteria. In fact, it is stated explicitly in 10 CFR Part 60.21(c)(1)(ii)(F) that "Analyses and models that will be used to predict future conditions and changes in the geologic setting shall be supported by using an appropriate combination of such methods as field tests, in-situ tests, laboratory tests which are representative of field conditions, monitoring data, and natural analogue studies." Second, intervenors could use validation as an issue in litigation against either DOE (before a license application for construction of a repository) or NRC (after a license is granted). In fact, the issue of validation was the basis for the decision in a court case involving the State of Ohio and the EPA (23 ERC 2091). In that case, the Sixth Circuit of the U.S. Court of Appeals ruled that EPA had acted arbitrarily in using the CRSTER computer code (EPA, 1977) as a basis for establishing limitations on sulfur dioxide emissions from two electric utility plants. The Court decided that EPA had failed to establish the accuracy or trustworthiness of the model as compared with the actual discharge from the plants. In other words, EPA did not perform a site-specific validation of the model.

4.0 VALIDATION STRATEGY

Given enough time and resources, all of the models used to assess the performance of a HLW repository could be validated using site-specific data. However, limitations in time and resources lead to alternative approaches to validation. Foremost, they lead to the use of as many relevant experiments as possible to test the models. Thus, a validation strategy should include the use of so-called generic validation experiments as well as experiments performed at the proposed repository location. In general, both generic and site-specific experiments will include laboratory tests, field tests, and natural analogs.

4.1 Use of Laboratory Experiments

Laboratory experiments are useful in testing the processes controlling the behavior of the repository system because (1) they are performed in a controlled environment that minimizes uncertainty in initial and boundary conditions, and (2) the experiments can be performed on samples that exhibit relatively little geometric variability (i.e., homogenous). However, the use of laboratory experiments in validation efforts is limited due to (1) the inability to perform tests on either long time scales or large spatial scales experiments, (2) the difficulty in testing coupled processes, and (3) the possibility that the samples used are not representative of field conditions.

4.2 Use of Field Experiments

Field experiments overcome the problem of representativeness of data and, to a degree, the spatial-scale problem that plague laboratory experiments. To a certain degree, 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 the field conditions.

4.3 Use of Natural Analogs

In some sense, nature could be considered to have initiated experiments that could be used for validation. Transport of radionuclides from uranium deposits and transport and deposition of minerals along fractures are a few examples. These "experiments" have the advantage of having taken place on temporal and spatial scales that are comparable to the HLW repository system scales. In addition, coupled processes are often involved that are difficult to produce 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 analogs in validating models.

5.0 PROPOSED APPROACH TO VALIDATION

The following steps outline a generic approach to model validation that is consistent with the philosophy discussed above.

5.1 Definition and Prioritization of Models to be Validated

Licensing of a HLW repository requires that certain performance standards be met. The regulations do not explicitly provide specifications as to what, if any, models must be used; nor are there unique models for a given HLW application. On the contrary, many models exist that purport to be capable of simulating some aspect of repository system. The choice of which models to use is up to the applicant (DOE) for the license application and up to the regulator (NRC) for the license evaluation. Therefore mapping between the performance standards and models that could be used to evaluate the repository system is the first step in model validation. This process involves defining a strategy for demonstrating compliance with a given standard and then developing conceptual models; choosing, developing, or modifying a computer code to implement the conceptual model; and obtaining the data required by the model. A conceptual model is simply the combination of all of the assumptions about the repository system that are required to adequately simulate the behavior of the system. A mathematical model is a numerical expression of the conceptual model. It consists of a set of equations that describe the initial and boundary conditions, the governing processes and the solution of the equations. The mathematical model may or may not be

implemented in a computer code. As stated previously, the combination of all of these is what is herein referred to as a model. Once the models are chosen they can be tested against the relevant experiments.

Limitations of resources and time do not permit testing of all possible models required in HLW performance assessments. One possible approach to deciding the priorities of which models to test first would be to follow the strategy given by Price and others (1990) on the review of performance assessment models. That strategy focuses first on processes that could directly lead to a release of radionuclides and proceeds to processes that reduce the chance that radionuclides would be released. Following a similar strategy, model validation studies could be forced to test the most important models first.

5.2 Identification and/or Definition of Relevant Experiments

Once the required models are identified, a combination of laboratory, field, and natural analog experiments should be identified that test the range of expected repository conditions (both current and future). It is likely that a sufficient range of experiments are currently available. Therefore, experiments should be performed that test the remaining processes or conditions not covered with existing experiments.

It cannot be overemphasized that the chosen experiments should be relevant to repository performance irrespective of the model's reported capabilities. In other words, model purpose should be the focus of

validation efforts not model capability. Too many past validation efforts (experiments and modeling) have not addressed licensing requirements but have focused instead on testing a given model. In fact many "generic" validation experiments have tested models or parts of models that will never be used to assess compliance with the relevant regulations. The pitfall of such an approach is beyond merely wasting resources. First, the potential exists co develop false confidence in the model because the experiment is not relevant. Second, equal importance is the possibility of invalidating a model that is acceptable for repository performance but is unable to reproduce results of an experiment that tests conditions not found in a repository environment.

A point needs to be made here concerning experiments that have been performed prior to the validation effort. First, and perhaps most important, there is a great tendency to turn the validation effort into a calibration effort. This is because the data describing the results of the experiment are available prior to, during, and following the modeling effort. Thus, the modeler has a tendency to fit (or calibrate) their results to the experimental results instead of using the input data to "predict" the experimental results. This type of an effort affords some degree of confidence in the model if the results agree well. However, it is by no means validation because the degrees of freedom for model input are generally so large that a "reasonable" fit can be obtained with many very different models. The reason that so many degrees of freedom exist is not a criticism of the experimental program but results instead from the reason for the experiment in the first place. Namely, most past

experiments were not specifically designed to test a validation issue. Thus, it is likely that not all of the information necessary to "validate" a model was obtained. Ideally, the validation effort would rely only on specific validation experiments that are part of the repository program. However, this is not possible given the limited resources and time, and for the case of natural analogs, obviously, not possible at all.

5.3 Definition of Performance Measures that are Based on the Intended Use of the Model

In order to judge the performance of the model relative to the experimental results, some representative measure(s) of the system response must be defined. A performance measure must be a quantity that is of regulatory interest or directly related to a quantity of regulatory interest. For the licensing of a HLW repository, the regulatory interests are such factors as concentrations of contaminants, integrated quantities of contaminants, and/or travel time of contaminants through the geosphere. Regulatory analyses require these parameters to be predicted over distances of kilometers and times of 1,000's to 10,000's of years. Because this is not possible, validation studies must rely on indirect measures of repository performance. Indirect performance measures are quantities like ground-water velocities and fluxes. Other performance measures that are even more removed from regulatory requirements (e.g., pressure and moisture content distributions) should only be used as a last resort because of the potential for

misinterpretation of the model validation results. For example, a model may be judged invalid if it cannot reproduce a three-dimensional distribution of moisture content. However, the same model may provide a perfectly adequate prediction of regulatory quantities, such as groundwater flux or contaminant concentrations.

5.4 Quantification of the Uncertainty Associated with the Input Data and the Data Available for Comparison with the Model Output

The ability of a model to adequately predict the performance measures cannot be judged without consideration of the uncertainty of the input parameters and the accuracy of the experimental results. There are several types of error that could be associated with the data. These include data collection errors (e.g., instrument error, human error) and interpretation errors (e.g., fitting a Theis curve). Some types of errors are quantifiable while others can only be estimated subjectively by the experimentalists. In any case, the responsibility of quantifying the errors associated with input parameters and experimental results lies with the experimentalists and not with the modelers.

5.5 Definition of Acceptance Criteria or Acceptable Model Error Based on Regulatory Requirements and Uncertainty in Experimental Results

This is one of the most difficult steps in the validation process because the question of the adequacy of the model predictions relative to the experimental results is ultimately subjective. In judging the validity

of model results, the uncertainty in experimental results should be taken into account. That is, the model can only judged relative to the accuracy of the measurements. Accounting for errors in the experimental results can be accomplished directly by defining acceptance limits which correspond to a band defined by the experimental measurements, plus and minus their associated error.

Accounting for the uncertainty in input parameters is more difficult than accounting for measurement errors in the experimental results. To illustrate this difficulty, consider the following two extremes. First, the input parameters could be known with complete certainty. Given that the model is an accurate representation of the real system (i.e., the conceptual, mathematical, and computerized aspects of the model are valid), then the model results should agree with the experimental results within the accuracy of the experimental measurement error. However, if the model does not agree with the experimental results, the model is invalid. Second, consider the other extreme, that is, the input parameters are completely uncertain. This situation can occur when a particular parameter is not measured during an experiment. In this case, the modeler is free to assign any value to this parameter in an attempt to obtain agreement between model and experimental results. In the event that the modeler is unable to obtain agreement in this manner, the model is invalid. Normally, however, the modeler is able to achieve some degree of agreement between the model and the experimental results. This does not mean that the model is valid; it only indicates that the model is not invalid. Thus, uncertainty in the model input must be taken into

account but the judgment about the validity of the model should be tempered by this uncertainty. In other words, the more data uncertainty with which the modeler is faced, the more difficult it is to conclude that the model is valid.

In addition to the data uncertainty described above, the other main source of uncertainty is in the model structure. The model structure is the inherent structure of the conceptual, mathematical, and computerized portion of the overall model; that is, all aspects of the model except the input data. Model-structure uncertainty arises from an incomplete knowledge of the driving processes, a limited knowledge of complex natural geometries, and mathematical and computational limitations given the current state of the art.

Consideration of these two types of uncertainty, model structure and model input data, leads to the definition of at least two acceptance criteria for a given case. The first one is a measure of the accuracy of the model input parameters relative to the experimental results, while the second is a measure of the adequacy of the model structure in describing the system behavior. To illustrate the two types of criteria consider a simple tracer tests through a column with a steady-state uniform velocity field and a constant inlet concentration. In this example, the conceptual model to be tested is a homogeneous isotropic porous medium, the behavior of which is assumed to be governed by the classic convective-dispersion equation.

To elucidate the criterion used to judge the adequacy of the model input data, assume that the underlying model structure is correct. For the example described above, this means that the model and experimental results both display the classic S-shaped contaminant breakthrough curve. Now consider the possibility that the model predicts that the contaminant arrives earlier than it really did. Depending on how close the two curves are, the model results could still be acceptable. The proposed acceptance criterion is based on a combination of the distance between the curves, the uncertainty in input values, and the uncertainty in experimental results. Given these results, several possibilities arise.

First, it is possible that there is no overlap between the model-predicted values and the experimental results (taking into account the uncertainty in the experimental results) regardless of what combinations of model input parameters are used. In this case, the model is invalid and the invalidity arises out of errors in or incomplete knowledge of the values of the input parameters. If this happens the experimentalists should attempt to make additional parameter measurements. The next possible condition for this acceptance criterion is that the model results agree with the experimental results. That is, some combination of input parameters results in model predictions that lie within the band created by the plotting of the experimental results and the associated experimental error. In this case, one cannot state that the model is valid, rather that it is not invalid. The statement that the model is valid cannot be made unless the full range of experimental conditions has been tested. For example, the hypothetical tracer test would have to be

repeated under different experimental conditions (for instance, the experiment would have to be repeated under a range of imposed velocities). To illustrate this point, consider a model for tracer movement that treats retardation with a constant retardation factor. Conceivably, the modeler could find a value of the retardation factor that allowed the model to reproduce the experimental results. However, if the retardation factor was actually a function of velocity, then the model would be invalid for all velocities other than the one used in the experiment. Thus, without performing additional experiments one cannot know if the model is truly valid, only that it is not invalid.

A second acceptance criterion arises out of the need to determine whether or not the model structure (i.e., the underlying conceptual, mathematical, and computer model) is adequate. This criterion is based on the concept that the model will produce a unique form of model results regardless of the model input. For the example described above, the model output is always the classical S-shaped concentration breakthrough curve. If this equation is an adequate representation of the physics governing a particular experiment, then the experimental results should be of the S-shaped form. Deviations from this form indicate an incorrect model structure. For example, the tracer test described above could be affected by diffusion of tracer into dead-end pores. The experimental results, in this case, would display a concentration breakthrough curve which falls underneath the classical S-shaped curve at late times. This phenomenon of dead-end pore diffusion was not included in the model and the acceptance criterion should be designed to test for the differences

it produces. One possible approach is to start with the "best fit" model obtained in the data-input uncertainty analysis described above. That is, start with the combination of data input that results in the closest fit to the experimental results. Using the best fit results as the representation of the behavior of the model, the approach could be to search for systematic differences between the model results and the experimental results. The basis for this is the assumption that random errors about the model results could be caused by measurement error but systematic errors result from inherent errors in the model structure. This search could be accomplished by subtracting the conceptual model results from the experimental results and testing these so-called residuals against a model for white noise (i.e., no trend in the residuals). Thus, if the residuals do not reveal any detectable trend, the conceptual model adequately represents the experimental results. On the other hand, the conceptual model is in error if a trend exists in the residuals.

5.6 Simulation of the Experiment

Simulation of the experiment may be done in so-called back-of-the-envelope fashion or on a computer. Most of the HLW performance assessment calculations faced by modelers today are so numerically intensive that a computer code is the only reasonable way to arrive at a solution. However, this is not always true and a simple calculation may be adequate, depending on the purpose of the model. Regardless of how the calculations are done, they should include an analysis of the

uncertainty of the input data. Several uncertainty analysis techniques are available for this purpose (Zimmerman and others, 1990; Doctor, 1989). One popular method is to use a Monte Carlo approach based on the Latin Hypercube Sampling (LHS) scheme (Iman and Shortencarier, 1984). Using this approach, one can insure that the entire range of parameters are tested. However, it does not guarantee that the best fit model will be obtained. Inverse techniques do provide a best fit but generally do not provide information about other combinations of parameters that could work equally well.

5.7 Experimentation

The field or laboratory experiment should follow the model simulations to assure that the modeling effort is not simply a calibration effort based on the experimental results. This is not possible for many laboratory and field experiments and not possible at all for natural analogs. At any rate, a serious attempt should always be made to simulate the experiments without reliance on the experimental results.

5.8 Evaluation of Model Results based on the Acceptance Criteria

In this step, the performance measure (for example, radionuclide concentrations) predicted by the model is compared to the experimental results using the acceptance criteria defined above as the measures of fitness. This allows one to conclude whether or not the model structure is correct and whether or not the model input data have been adequately

estimated. In the event that the model structure and the input data are adequate, two questions remain. First, would the same model be able to adequately predict different experimental conditions and, second, are other model structures (conceptual models) also able to adequately simulate the experimental results. Both of these questions can only be answered with new experiments under different conditions. Understanding the differences in conceptual models should allow the modelers to recommend experiments that can only be adequately simulated by one of the models.

Comparison of model predictions and experimental results may lead one to conclude that the input data were not adequately estimated. Again, the modeler then should be able to provide the experimentalist with a list of the desired data, including some indication as to which data are most important. The most important data could be determined by using sensitivity analysis of the model results.

Finally, comparison of model predictions and experimental results may lead to the conclusion that the model structure is inadequate for describing the experiment. The modeler would then have to propose a new model structure and repeat the simulations and model comparisons. However, this approach will be just another form of calibration unless a new experiment is performed that follows the final model simulations.

6.0 SITE-SPECIFIC MODEL VALIDATION

The approach outlined above deals with experiments that test only certain models or parts of models and generally over relatively small spatial and temporal scales. However, the site-specific repository performance assessments that are required for licensing will be produced using a wide variety of models. As stated previously, this system of models can not be directly tested because they predict conditions for 1,000's of years over scales of kilometers. While these models can not be directly tested, certain factors can be evaluated to gain confidence that the models are valid. These factors are outlined below.

6.1 Agreement of the Model with Site-Specific Information

At a minimum the model input data must be site-specific. However, information other than input data is available to check the model validity. Consider a ground-water flow model for example. Not only should the model be based on site-specific hydraulic conductivities but the model results should agree with other site-specific hydrologic information such as hydraulic heads, ground-water chemistry, and isotopic ages of ground-water. Questions will remain, however, as to the sufficiency of the data used for comparison and the degree to which the model must be consistent with other site information. In addition, the check for consistency generally can only be based on comparison with current conditions. Repository conditions (both natural and man-made) may change significantly over time of interest for regulatory compliance

(i.e., 10,000 years). Thus, a major effort should be the identification of sites similar to the repository location that are currently under conditions different from the repository site. For example, it may be possible to find a site with similar geology to the repository site but that is currently under much wetter climatic conditions.

6.2 Justification of Assumptions

The foundation of any model is its assumptions. Therefore, considerable attention should be paid to the assumptions that are invoked in simulating the performance of the repository system. If the generic validation tests had covered all possible conditions of model use, then this task would involve only choosing those assumptions used with experiments that were identical to repository conditions. Unfortunately, experiments of this type are not possible. Therefore, untestable assumptions will have to be based to a large degree on expert judgement (Bonano and Cranwell, 1988; Bonano and others, 1990). In no case should they be based on lack of model input data.

6.3 Treatment of Multiple Conceptual Models

No matter how well characterized a site may be, multiple conceptual models (all consistent with available data) are possible. In fact, the use of multiple conceptual models in HLW repository performance assessments should be encouraged to gain reasonable assurance that the results encompass the true system behavior. Care should be taken,

however, that the necessary data were gathered during site characterization to exercise the models as well as test the validity of the models.

7.0 CONCLUSIONS

HLW performance assessment models can not be proven to be valid due to the temporal and spatial scales of regulatory concern. However, confidence in the models can be gained through testing model performance against a combination of laboratory, field, and natural analogs as well as assuring that the models are consistent with site-specific information and that all equally plausible models are used in the final assessment. This confidence arises out of showing that the models are not invalid over a wide range of conditions and by testing the validity of model input data separately from model structure.

References

Bonano, E.J. and R.M. Cranwell, 1988. "Treatment of Uncertainties in the Performance Assessment of Geologic High-Level Radioactive Waste Repositories," Math. Geology, 20, pp. 543-565.

Bonano, E.J., S.C. Hora, R.L. Keeney, and D. von Winterfeldt, 1990. Elicitation and Use of Expert Judgment in Performance Assessment for High-Level Radioactive Waste Repositories, NUREG/CR-5411, SAND89-1821, Sandia National Laboratories, Albuquerque, NM.

Borgorinski, P., B. Baltes, J. Larue, and K.-H. Martens, 1988. The Role of Transport Code Verification and Validation Studies in Licensing Nuclear Waste Repositories in the FR of Germany, Radiochimica Acta, 44/45, pp. 367-372.

Doctor, P.G., 1989. "Sensitivity and Uncertainty Analysis for Performance Assessment Modeling," Eng. Geology, 26, pp. 411-429.

DOE, 1986. Environmental Assessment - Yucca Mountain Site, Nevada Research and Development Area, Nevada. DOE/RW-0073, Vol. 2, U.S. Department of Energy, Office of Civilian Radioactive Waste Management, Washington, D.C.

EPA, 1977. User's Manual for Single Source (CRSTER) Model, EPA 450/2-77-013, U.S. Environmental Protection Agency, Washington, D.C.

Iman, R.L. 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, NM.

NRC, 1984. A Revised Modelling Strategy Document for High-Level Waste Performance Assessment, U.S. Nuclear Regulatory Commission, Washington, D.C.

NRC, 1988. NRC Model Simulations in Support of the Hydrologic Code Intercomparison Study (HYDROCOIN), Level 1 - Code Verification, NUREG-1249, Vol. 1, U.S. Nuclear Regulatory Commission, Washington, D.C.

Popper, K., 1958. The Logic of Scientific Discovery, Harper and Row, New York, NY, 479 pp.

Price, L.L., K.K. Wahi, D.P. Gallegos, M.T. Goodrich, N.E. Olague, and D.A. Brosseau, 1990. Technical Basis for NRC Review of High-Level Waste Repository Modeling, NUREG/CR-5398, SAND89-1557, Sandia National Laboratories, Albuquerque, NM.

Zimmerman, D.A., K.K. Wahi, A.L. Gutjahr, and P.A. Davis, 1990. A Review of Techniques for Propagating Data and Parameter Uncertainties in

High-Level Radioactive Waste Performance Assessment Models, NUREG/CR-5393, SAND89-1432, Sandia National Laboratories, Albuquerque, NM.