NUREG-1636

# Regulatory Perspectives on Model Validation in High-Level Radioactive Waste Management Programs: A Joint NRC/SKI White Paper

**Swedish Nuclear Power Inspectorate**

**Center for Nuclear Waste Regulatory Analyses**

**U.S. Nuclear Regulatory Commission**
**Office of Nuclear Material Safety and Safeguards**
**Washington, DC 20555-0001**

# AVAILABILITY NOTICE

## Availability of Reference Materials Cited in NRC Publications

NRC publications in the NUREG series, NRC regulations, and *Title 10, Energy,* of the *Code of Federal Regulations,* may be purchased from one of the following sources:

1. The Superintendent of Documents
   U.S. Government Printing Office
   P.O. Box 37082
   Washington, DC 20402–9328
   <http://www.access.gpo.gov/su_docs>
   202–512–1800

2. The National Technical Information Service
   Springfield, VA 22161–0002
   <http://www.ntis.gov/ordernow>
   703–487–4650

The NUREG series comprises (1) brochures (NUREG/BR-XXXX), (2) proceedings of conferences (NUREG/CP-XXXX), (3) reports resulting from international agreements (NUREG/IA-XXXX), (4) technical and administrative reports and books [(NUREG-XXXX) or (NUREG/CR-XXXX)], and (5) compilations of legal decisions and orders of the Commission and Atomic and Safety Licensing Boards and of Office Directors' decisions under Section 2.206 of NRC's regulations (NUREG-XXXX).

A single copy of each NRC draft report is available free, to the extent of supply, upon written request as follows:

Address: Office of the Chief Information Officer
Reproduction and Distribution
Services Section
U.S. Nuclear Regulatory Commission
Washington, DC 20555–0001
E–mail: <DISTRIBUTION@nrc.gov>
Facsimile: 301–415–2289

A portion of NRC regulatory and technical information is available at NRC's World Wide Web site:

<http://www.nrc.gov>

All NRC documents released to the public are available for inspection or copying for a fee, in paper, microfiche, or, in some cases, diskette, from the Public Document Room (PDR):

NRC Public Document Room
2120 L Street, N.W., Lower Level
Washington, DC 20555–0001
<http://www.nrc.gov/NRC/PDR/pdr1.htm>
1–800–397–4209 or locally 202–634–3273

Microfiche of most NRC documents made publicly available since January 1981 may be found in the Local Public Document Rooms (LPDRs) located in the vicinity of nuclear power plants. The locations of the LPDRs may be obtained from the PDR (see previous paragraph) or through:

<http://www.nrc.gov/NRC/NUREGS/SR1350/V9/lpdr/html>

Publicly released documents include, to name a few, NUREG-series reports; *Federal Register* notices; applicant, licensee, and vendor documents and correspondence; NRC correspondence and internal memoranda; bulletins and information notices; inspection and investigation reports; licensee event reports; and Commission papers and their attachments.

Documents available from public and special technical libraries include all open literature items, such as books, journal articles, and transactions, *Federal Register* notices, Federal and State legislation, and congressional reports. Such documents as theses, dissertations, foreign reports and translations, and non-NRC conference proceedings may be purchased from their sponsoring organization.

Copies of industry codes and standards used in a substantive manner in the NRC regulatory process are maintained at the NRC Library, Two White Flint North, 11545 Rockville Pike, Rockville, MD 20852–2738. These standards are available in the library for reference use by the public. Codes and standards are usually copyrighted and may be purchased from the originating organization or, if they are American National Standards, from—

American National Standards Institute
11 West 42nd Street
New York, NY 10036–8002
<http://www.ansi.org>
212–642–4900

NUREG-1636

# Regulatory Perspectives on Model Validation in High-Level Radioactive Waste Management Programs: A Joint NRC/SKI White Paper

Manuscript Completed: December 1998
Date Published: March 1999

Prepared by
N. A. Eisenberg, M. P. Lee, M. V. Federline[a]
S. Wingefors, J. Andersson, S. Norrby[b]
B. Sagar, G. W. Wittmeyer[c]

[a]Division of Waste Management
Office of Nuclear Material Safety and Safeguards
U.S. Nuclear Regulatory Commission
Washington, DC 20555-0001

[b]Office of Nuclear Waste Safety
Swedish Nuclear Power Inspectorate
(Statens Kärnkraftinspektion)
S-106 58 Stockholm, Sweden

[c]Center for Nuclear Waste Regulatory Analyses
6220 Culebra Road
San Antonio, TX 78238-5166

# ABSTRACT

Validation (or confidence building) should be an important aspect of the regulatory uses of mathematical models in the safety assessments of geologic repositories for the disposal of spent nuclear fuel and other high-level radioactive wastes (HLW). A substantial body of literature exists indicating the manner in which scientific validation of models is usually pursued. Because models for a geologic repository performance assessment cannot be tested over the spatial scales of interest and long time periods for which the models will make estimates of performance, the usual avenue for model validation—that is, comparison of model estimates with actual data at the space-time scales of interest—is precluded. Further complicating the model validation process in HLW programs are the uncertainties inherent in describing the geologic complexities of potential disposal sites, and their interactions with the engineered system, with a limited set of generally imprecise data, making it difficult to discriminate between model discrepancy and inadequacy of input data. A successful strategy for model validation, therefore, should attempt to recognize these difficulties, address their resolution, and document the resolution in a careful manner. The end result of validation efforts should be a documented enhancement of confidence in the model to an extent that the model's results can aid in regulatory decision-making. The level of validation needed should be determined by the intended uses of these models, rather than by the ideal of validation of a scientific theory. This White Paper presents a model validation strategy that can be implemented in a regulatory environment. It was prepared jointly by staff members of the U.S. Nuclear Regulatory Commission and the Swedish Nuclear Power Inspectorate—SKI. This document should not be viewed as, and is not intended to be formal guidance or as a staff position on this matter. Rather, based on a review of the literature and previous experience in this area, this White Paper presents the views of members of the two organizations regarding how, and to what degree, validation might be accomplished in the models used to estimate the performance of HLW repositories.

# CONTENTS



## FIGURES



## TABLES



## APPENDICES

# ACKNOWLEDGMENTS

The authors wish to convey their thanks to several individuals who provided useful information, comments, and reviews that have enhanced the substance of this White Paper as it was being developed. These individuals include:

Seth M. Coplan/Division of Waste Management (DWM)
Daniel J. Fehringer[1]/DWM
Timothy Margulies[2]/ Office of Nuclear Regulatory Research (RES)
Timothy J. McCartin/DWM
Keith I. McConnell/DWM
Christopher A. McKenney/DWM
Thomas J. Nicholson/RES
Jake Philip/RES
John D. Randall/RES
William C. Reamer/DWM

They do not necessarily approve, disapprove, or endorse the views expressed in this document. In addition, two DWM staff members, John T. Buckley and Janet P. Kotra, provided valuable assistance in compiling early versions of this White Paper and deserve special recognition. Also, Ellen Kraus, Office of Nuclear Material Safety and Safeguards, provided editorial guidance.

Finally, special thanks to G.S. ("Bo") Bodvarsson/Lawrence Berkeley National Laboratory (LBNL), Mark Cushey/LBNL, Dwight T. Hoxie/U.S. Geological Survey, Bruce A. Robinson/Los Alamos National Laboratory (LANL), Paul W. Reimus/LANL, and Abraham van Luik/U.S. Department of Energy (DOE), who provided an update of the description of DOE validation activities within the Yucca Mountain Site Characterization Project in Appendix B.

[1] Now with the U.S. Nuclear Waste Technical Review Board.

[2] Now with the U.S. Environmental Protection Agency.

# FOREWORD

During the past two decades, there have been a number of international efforts underway, both individually and cooperatively, to contribute to progress in the development of procedures to validate mathematical models used in safety assessments of geologic repositories for high-level radioactive waste (HLW). In parallel with these efforts, repository regulators have also given considerable thought to this issue. Based on a review of the literature and previous experience in this area, this White Paper presents the authors' views regarding what degree of validation might be desirable in models used to estimate the long-term performance of a geologic repository for HLW. The collaborative effort elaborates on these views, from a regulatory perspective, which were originally presented in a shorter paper on this subject at an international symposium sponsored in 1994 by the Organization for Economic Co-operation and Development/ Nuclear Energy Agency.

Overall, it is believed that the responsibility of validating models to be used in any potential licensing action rests primarily with the repository developer. In this regard, the degree of validation needed would be commensurate with the extent to which the safety case depends upon the model(s) in question. By contrast, the degree of validation of the regulator's models may be less rigorous since its models will be used to independently ensure that the developer has made an adequate fundamental determination of repository safety.

This document does not have the status of formal guidance nor does it represent a staff position on this matter. However, the two organizations may move jointly or individually to develop formal guidance or a staff position on this matter, at a later date. Until that time, the authors would welcome public feedback on the concepts being advanced in this White Paper. Finally, this White Paper will be published in parallel by the Swedish Nuclear Power Inspectorate (SKI) under its own cover as *SKI Report 99:2*.

# ACRONYMS AND ABBREVIATIONS

| | |
|---|---|
| ASTM | American Society for Testing and Materials |
| BIOMASS | **Bio**sphere **M**odeling and **Ass**essment program |
| BIOMOVS | **Bio**spheric **Mo**del **V**alidation **S**tudy |
| CCDF | complementary cumulative distribution function |
| CEC | Commission of the European Communities |
| CHEMVAL | validation and verification of geochemical models project |
| CNWRA | Center for Nuclear Waste Regulatory Analyses (in San Antonio, Texas) |
| DECOVALEX | **De**velopment of **Co**upled Models and their **Val**idation against **Ex**periments project |
| DOE | U.S. Department of Energy |
| EBS | engineered barrier system |
| EnPA | Energy Policy Act of 1992 |
| EPA | U.S. Environmental Protection Agency |
| GEOVAL | series of symposia on the verification and validation of geosphere performance assessment models |
| HLW | high-level radioactive wastes (including spent nuclear fuel) |
| HYDROCOIN | **Hydro**logic **Co**de **In**tercomparison project |
| IAEA | International Atomic Energy Agency |
| ICRP | International Commission on Radiation Protection |
| IPA | iterative performance assessment |
| INTRACOIN | **In**ternational **Tra**nsport **Co**de **In**tercomparison project |
| INTRAVAL | **In**ternational **Tra**nsport Model **Val**idation program |
| MIRAGE | **Mi**tigation of **Ra**dionuclides in the **Ge**osphere project |
| NAS | National Academy of Sciences (in the United States) |
| NRC | U.S. Nuclear Regulatory Commission |
| NEA | Nuclear Energy Agency |
| NWPA | Nuclear Waste Policy Act of 1982, as amended |
| OECD | Organization for Economic Co-operation and Development |
| PRA | probabilistic risk assessment |
| QA | quality assurance |
| R&D | research and development |
| SKB | Swedish Nuclear Fuel and Waste Management Company (or Svensk Kärnbränslehantering AB) |
| SKI | Swedish Nuclear Power Inspectorate (or Statens Kärnkraftinspektion) |
| SSI | Swedish Radiation Protection Institute (Statens Strålskyddsinstitut) |
| T–H–M | thermo-hydrologic-mechanical (coupling) |
| U.S. | United States |
| VOG | DOE's **V**alidation **O**versight **G**roup |

# 1 INTRODUCTION

The U.S. Nuclear Regulatory Commission (NRC), in the United States (U.S.), and the Swedish Nuclear Power Inspectorate (Statens Kärnkraftinspektion or SKI), in Sweden, are, respectively, the regulatory authorities responsible for licensing geologic repositories for the disposal of spent nuclear fuel and other high-level radioactive waste (HLW). In reaching the necessary licensing decisions, both NRC and SKI are concerned that their final judgments regarding compliance with their respective geologic repository disposal regulations be made with *reasonable assurance*.[1] To reach necessary licensing findings, both NRC and SKI staffs will need to do two things. First, they each will need to confirm that any relevant numerical performance standards have been met. This will be done independently, for example, by each using its own analytical capability to corroborate the repository developer's conclusions and supporting calculations. Second, the staff of each regulatory agency will need to satisfy itself that the repository developer's[2] analyses of the site and design are sufficiently conservative, that the limitations of its analyses are well-understood, and that appropriate allowances have been made for the time period, hazards, and *uncertainties* involved. To do this, each regulatory staff will selectively probe the developer's assessment for potential weaknesses, based on a familiarity with the methods, site data, and prevailing assumptions used.

One of NRC's and SKI's greatest challenges in making these determinations will be to understand and evaluate the repository developer's treatment of uncertainties in its analyses. Various methods may be used (e.g., probability distributions, conservative "bounding" analyses, etc.). Previous licensing practice suggests that the two regulatory authorities ultimately will have to seriously consider both quantitative and non-quantitative arguments, to ascertain the adequacy of handling of uncertainty by the repository developer.

A geologic repository for the disposal of HLW is a complex system. It is generally expected to consist of multiple barriers,[3] where each barrier will contribute to the overall safety of the system by providing some contribution to containment and waste isolation. Because the future performance of a geologic repository will be estimated for many thousands of years, the long-term contribution of each barrier class as well as the overall repository system itself is expected to be demonstrated, in a regulatory setting, through the use of conceptual models that can be mathematically expressed.[4] There is general consensus within the international community that to evaluate the safety of these facilities, before their implementation, repository developers and regulators will rely on current state-of-the-art mathematical models as part of *performance assessment* that is integral to an overall *safety assessment*. Performance assessment may thus be defined as the process of quantitatively evaluating the ability of a geologic repository to contain and isolate radioactive waste (Campbell and Cranwell, 1988). This quantitative evaluation, through the use of mathematical models, is a key component in the development of a geologic repository design and in the demonstration of compliance with the applicable safety standards and regulations. However, before such models can be used, for that purpose, some measure of credibility and confidence in these models

[1] Although both the U.S. and Sweden use the same term to describe the standard for meeting each country's licensing requirements, the criteria used to define reasonable assurance are different in the two countries.

[2] In the U.S., this is the U.S. Department of Energy (DOE); and in Sweden, this is the Swedish Nuclear Fuel and Waste Management Company (Svensk Kärnbränslehantering AB—SKB).

[3] For ease of discussion, two barrier classes are identified—engineered and natural—although there may be several individual barriers in each class.

[4] In addition, it is expected that numerical methods and computer codes will be used in the requisite compliance/safety demonstrations. However, the procedures for **verifying** the correctness of the numerical methods and computer codes are different from those used to **validate** conceptual models. Typically, the verification (including benchmarking) of numerical methods and computer codes is undertaken to establish the numerical correctness of the methods and codes to ensure that the numerical solutions are converging and adequately represent the conceptual model (see Eisenberg *et al.*, 1988).

should also be demonstrated and done so in a way that is transparent.

Finally, any fundamental understanding of the performance of a geologic repository, by the developer, will be based in large part on the manner in which the repository system has been *modeled*. For the purposes of this White Paper, a simplified approach to modeling is considered (e.g., Mercer and Faust, 1980; pp. 108-110). In general, the process begins with the formulation of a *conceptual model*, whose purpose is to describe the physical behavior of interest. The next step is to translate the conceptual understanding of these physical processes into mathematical terms (i.e., make simplifying assumptions and develop governing equations), consistent with established scientific theory. This step constitutes development of the *mathematical model*. Once a mathematical model has been prepared, it can be solved either *analytically* or *numerically*, to derive numerical estimates of performance. Because the issues of the correctness of the mathematical formulation and the numerical solution can be addressed by fairly standard methods [see Eisenberg *et al.* (1988, pp. 348-349)], the strategy proposed herein focuses principally on the *confidence* or *validity*[5] in *conceptual models*.

## 1.1 Background

In the context of radioactive waste disposal, the International Atomic Energy Agency (IAEA) defines performance assessment as "... an analysis to predict the performance of a system or subsystem, followed by a comparison of the results of such [an] analysis with appropriate [safety] standards and criteria...." [see Organization for Economic Co-operation and Development (OECD)/Nuclear Energy Agency (1991, p. 14)]. A performance assessment is a type of a *probabilistic risk assessment* (PRA)[6] requiring an analysis of repository performance similar to a PRA conducted for a nuclear power plant—e.g., NUREG-1150 (NRC, 1990). Performance assessment technology has evolved over the last few decades because of the national and international interest in the geologic disposal of HLW. The methodology and means for conducting performance assessments vary from country to country. However, three generic criteria have been suggested to judge the adequacy of any performance assessment on which a safety case is based (*Op cit.*, pp. 12-13):

- The need for an integrated assessment;
- The consideration of uncertainties in the assessment results; and
- The methods for building confidence in assessment results.

With respect to the last criterion, confidence building, this would include, among other things,[7] appropriate steps for assuring or *validating* that the predictive models and codes used in the safety assessments adequately represent the behavior of the disposal system.

The need for validation in some form is acknowledged as part of a safety assessment in NRC's regulations for geologic disposal—10 CFR Part 60 (see Appendix A)[8]—which currently state:

> ".... 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,

[5] In this White Paper, the terms "confidence," "confidence building," "validation," and "validation process" are used interchangeably.

[6] Also see DOE *et al.* (1992).

[7] The *Collective Opinion* also suggests that the use of quality assurance (QA) procedures; critical *peer review*, including [formal] expert judgment; and international cooperation contribute to the confidence-building process.

[8] Currently, a revised set of standards specific to the Yucca Mountain site is being developed in accordance with the provisions of the Energy Policy Act of 1992 (EnPA—Public Law 102-486). EnPA, directs NRC to promulgate a rule, modifying 10 CFR Part 60 of its regulations, so that these regulations are consistent with the U.S. Environmental Protection Agency's (EPA's) public health and safety standards for protection of the public from releases to the accessible environment from radioactive materials stored or disposed of at Yucca Mountain, Nevada, consistent with the findings and recommendations made by the National Academy of Sciences (NAS), to EPA, on issues relating to the environmental standards governing the Yucca Mountain repository. It is assumed that the revised EPA standards for the Yucca Mountain site will not be substantially different from those currently contained in 40 CFR Part 191, particularly as they pertain to the need to conduct a quantitative performance assessment as the means to estimate post-closure performance of the repository system.

> laboratory tests which are representative of field conditions, monitoring data, and natural analog studies" [(10 CFR 60.21(c)(1)(ii)(F)].

In addition, Subpart F of NRC's regulations require the establishment of a *Performance Confirmation Program*, during which the adequacy of data, parameters, modeling assumptions, and designs is to be confirmed, to the extent practicable.

As yet, no similar regulations have been issued in Sweden, although they are currently under development (Dverstorp *et al.*, 1997). However, recommendations on criteria for disposal of HLW in Sweden have been presented jointly[9] by the respective nuclear safety and radiation protection authorities in the Nordic countries (The Radiation Protection and Nuclear Safety Authorities in Denmark, Finland, Iceland, Norway, and Sweden, 1993). Future Swedish regulations are expected to follow these recommendations. Among other things, in the area of validation, these recommendations propose that (*Op cit.*, p. 34):

> "....Compliance of the overall disposal system with the radiation protection criteria shall be convincingly demonstrated through safety assessments which are based on qualitative judgement and quantitative results from models that are validated as far as practicable...."

Moreover, in the recommendations for the requisite safety (assessment) models it is stated (*Op cit.*, p. 35) that the:

> "...models to be used in safety assessments should be validated as far as is reasonable by evidence from laboratory tests and field observations, including natural analogues...."

Thus, under both the current U.S. and proposed Swedish regulatory regimes, the long-term performance of a geologic repository will be assessed, using quantitative modeling techniques that will rely on a reasonable degree of validation.

Evaluation of the adequacy of a developer's performance assessment will not only check whether estimated performance (e.g., dose) complies with specified performance criteria, but should also ascertain whether the essential physical and chemical processes and their interactions have been identified, adequately described, and addressed. However, the level of confidence building or validation in a safety/performance assessment, adequate for licensing decisions, remains to be defined. The notion of confidence building is used in general recognition of the fact that full scientific validation, in the conventional sense, of the mathematical models used in these assessments is a practical impossibility, and that the acceptance of mathematical models will be based on appropriate testing, which will lead to the expectation that their results are sufficiently supported for making the necessary licensing decisions. **Thus, in a regulatory context, it is expected that the level of confidence required for a particular performance assessment model will be tied directly to the importance of the model to the licensing decision it supports.** Within this context, it is permissible to use models in repository performance assessments that do not necessarily attempt to predict the exact outcome, but instead rely on appropriate assumptions that provide conservative estimates (or predictions) of performance. (It should be noted that the use of *predictions*, in the context of repository performance, is perhaps inappropriate because, in practice, the best that can be expected is an estimate of performance under stipulated future conditions under which a hypothetical geologic repository has to perform. Such models are considered *conservative* if they systematically estimate worse performance than actual performance. Since the actual performance may never be known, the overall degree of conservatism is generally established qualitatively.)

In light of these considerations, the question that needs to be addressed is what degree of conservatism is sufficient in a safety

[9] Hereafter referred to as the *Nordic Document*.

demonstration? As noted above, it is expected that the level of confidence required for a particular performance assessment model should be tied directly to the importance of the model to the particular licensing decision it supports. From the regulator's perspective, in the first instance, the implementer (i.e., the repository developer) is believed to be the party primarily responsible for deciding this because it is the most knowledgeable when it comes to understanding the limitations in site data and supporting analyses. In the presence of strong nonlinearities in geologic repository systems, the level of conservatism necessary for licensing may not always be obvious. Too much conservatism may render a performance assessment unacceptably unrealistic and thus ineffective for the purpose of making licensing decisions. At the other extreme, a lack of conservatism in a particular assessment would likely result in prejudicial treatment by the cognizant regulatory authorities. Thus, to support its compliance demonstrations, the repository developer needs to describe the extent to which its models and codes have been supported (i.e., validated).

## 1.2 Purpose of the White Paper

This document presents the regulatory perspective of the authors, who are members of the NRC's Division of Waste Management and SKI on the validation process of HLW performance assessments. However, current or potential repository developers or other regulators should recognize that this **document does not have the status of formal guidance nor does it represent a staff position on this matter**. (The two staffs are free to move jointly or individually to develop such formal guidance, at a later date, based on feedback on this White Paper.) Rather, based on a review of the literature and previous experience in this area, this White Paper presents the authors' views regarding the nature of confidence building desirable for models used to estimate the long-term performance of a geologic repository for HLW, as well as issues that might be considered in any future guidance. The collaborative effort elaborates on the views of the authors, which were presented earlier in a shorter paper on this subject (see Eisenberg *et al.*, 1995). To support these views, an overview is provided in Section 2 which describes earlier international efforts in the area of HLW model validation. Section 3, outlines one validation approach that may be acceptable, in the opinion of the authors, from a regulatory perspective. Summary recommendations are included in Section 4.

In Appendix A, the U.S. and Swedish regulatory performance requirements are first briefly described. Then, the nature of conceptual and mathematical models that are expected to be used in a performance assessment of a HLW repository is described. A selective review of literature on model validation is provided in Appendix B which indicates a wide divergence in thinking on this subject. A glossary defining selected terms used in this document is provided in an Appendix C.

A list of acronyms and abbreviations can be found before the "Introduction" to this White Paper.

# 2 INTERNATIONAL EXPERIENCES WITH MODEL VALIDATION

During the past two decades, there have been a number of international efforts underway, both individually and cooperatively, to contribute to progress in the development of procedures to validate mathematical models used in HLW repository safety assessments. Results/progress in many of the international efforts have been the focus of several meetings and symposia—see, for example, the *GEOVAL*[10] series of symposia, American Nuclear Society (1993), and Witherspoon (1991 and 1996).

In 1980, SKI took the initiative to organize several collaborative efforts relevant to validation issues related to the use of radionuclide transport models. INTRACOIN —**Int**ernational **Tr**ansport **Co**de **In**tercomparison—(1981-86)[11] was organized by SKI to study computer code verification procedures for radionuclide transport models. HYDROCOIN—**Hydro**logic **Co**de **In**tercomparison (1984-90),[12] also organized by SKI, studied code verification and, to some extent, validation procedures for ground-water flow models. SKI later initiated the INTRAVAL (the **In**ternational **Tr**ansport Model **Val**idation) program as a follow-up to the earlier INTRACOIN and HYDROCOIN efforts (see Larsson, 1992). INTRAVAL was directly concerned with validating geosphere flow and transport models used in safety assessments and placed less emphasis on code verification procedures. Initially, the goal of these studies was the intercomparision of computer codes with some attention to how well the models underlying the computer codes represented ground-water flow and geosphere transport. As experience was accumulated, it became clear how difficult it was to address the validity of these models, so emphasis shifted from code intercomparision to model validation, which was the primary focus of INTRAVAL.

In addition to the initiatives sponsored by SKI, there have been several other cooperative validation efforts conducted internationally. For example, the International Stripa Project was a cooperative research and development project among several members of the OECD Nuclear Energy Agency. It combined on-going site characterization of the Stripa research mine with two validation experiments of flow and transport models. BIOMOVS—the **Bio**spheric **Mo**del **V**alidation **S**tudy—(1986–96), coordinated by the Swedish Radiation Protection Institute (Statens Strålskyddsinstitut—SSI), intends to evaluate the uncertainty in models used to determine the environmental transfer and bioaccumulation of radionuclides. CHEMVAL (1987-90) was concerned with verifying and validating equilibrium speciation and chemical transport models. Finally, DECOVALEX (an acronym for the "**De**velopment of **Co**upled Models and their **Val**idation against **Ex**periments" project in nuclear waste isolation) addressed the validation of coupled thermo-mechanical-hydrological (T-M-H) models used in near-field repository safety assessments.

The following is a brief summary of five of the aforementioned international programs. This summary is not intended to be comprehensive nor complete, and has been provided as a way of illustrating the types of activities that have been undertaken, to develop an understanding of model validation procedures in an international setting. Certain aspects of these validation activities themselves, as well as some lessons-learned are useful to consider from a regulator's perspective. As noted above, SKI has figured in many of the international programs and also directed some of the validation studies. Unlike SKI, NRC has not taken the lead in international validation studies; however, the NRC staff has maintained cognizance of this work and in some cases has participated in certain projects HYDROCOIN (NRC, 1988), INTRAVAL,

[10]A series of symposia on the verification and validation of geosphere performance assessment models. See SKI (1988), SKI/OECD/NEA (1991), and OECD/NEA/SKI (1995).

[11]The work of the INTRACOIN project is summarized in SKI (1984 and 1986).

[12]The work of the HYDROCOIN project has been summarized in SKI (1992) and The Coordinating Committee of the HYDROCOIN Project (1992).

DECOVALEX—as well as sponsoring its own independent work related to validation [e.g., Davis *et al.* (1991), and Kozak and Olague (1995)].

## 2.1 The International Stripa Project

The goals of the International Stripa Project (1980-92), carried out in granitic rocks of the Stripa research mine in central Sweden, were to investigate several aspects of the technology concerned with the feasibility and safety of disposal of HLW. The activities and results of this project were documented in more than 170 technical reports, and are summarized in Fairhurst *et al.* (1993); Gnirk (1993); Gray (1993), and SKB (1993).

Part of this project concerned an evaluation of the validity of flow and transport models. To conduct the validation exercise, the site was first characterized with different measurement techniques. Based on this information, two validation experiments were designed and conducted (see Gnirk, 1992). Different modeling teams then tried to simulate these experiments. In each exercise, the modelers were asked to predict the results of the validation experiments conducted, without prior knowledge of the outcomes.

Recognizing that both definitions and requirements for model validation varied among participants and that the definition is still being discussed at the international level, the project selected an operational definition of validation. This was that a model was considered to be validated for use in a given application when the model had been determined, by appropriate measures, to provide a representation, of the process or system, that was acceptable to an assembled group of "knowledgeable/recognized experts" from the member countries. A set of validation criteria, for evaluating the validity of the modeling approach and its components, was formulated in the form of questions or criteria. The first set of questions addressed both the quantitative and qualitative features of the modeling approach:

- ***Quantitative:*** Do the predictive calculations adequately reflect the measured values? That is, are the predictions of the correct order of magnitude as compared with the measurements?
- ***Qualitative:*** For the purpose of the application, are the predicted distribution patterns sufficiently accurate as compared with the observations? That is, are the predictions of the patterns reasonable when compared to observations?

The second set of questions addressed the usefulness and feasibility of the modeling approach from the perspective applicability:

- ***Usefulness:*** Is the modeling approach useful for representing ground-water flow and transport in a hydrogeologic environment similar to that of the investigated site?
- ***Feasibility:*** Can the data required to support fully the modeling approach be collected in a feasible and timely manner?

These criteria were applied to evaluate the different modeling exercise outcomes, following a formal process (see Figure 1). In this manner, the group of experts found a structured approach to:

- Identify processes covered and processes that appeared not to be covered by the different models; and
- Compare the different modeling approaches (in a qualitative sense).

At the close of the project, the experts made documented judgments of the validity of various approaches for modeling ground-water flow and transport at the Stripa research site [see Hodgkinson (1992), and Hodgkinson and Cooper (1992a and b)]. This effort constitutes an example of a case history of a formal and deliberate approach to the evaluation of numerical models, for a specific application.

## 2.2 INTRAVAL

The goal of the INTRAVAL exercise (1987-93) was to advance the state of knowledge, regarding the practical use of qualitative and quantitative methods, to demonstrate the

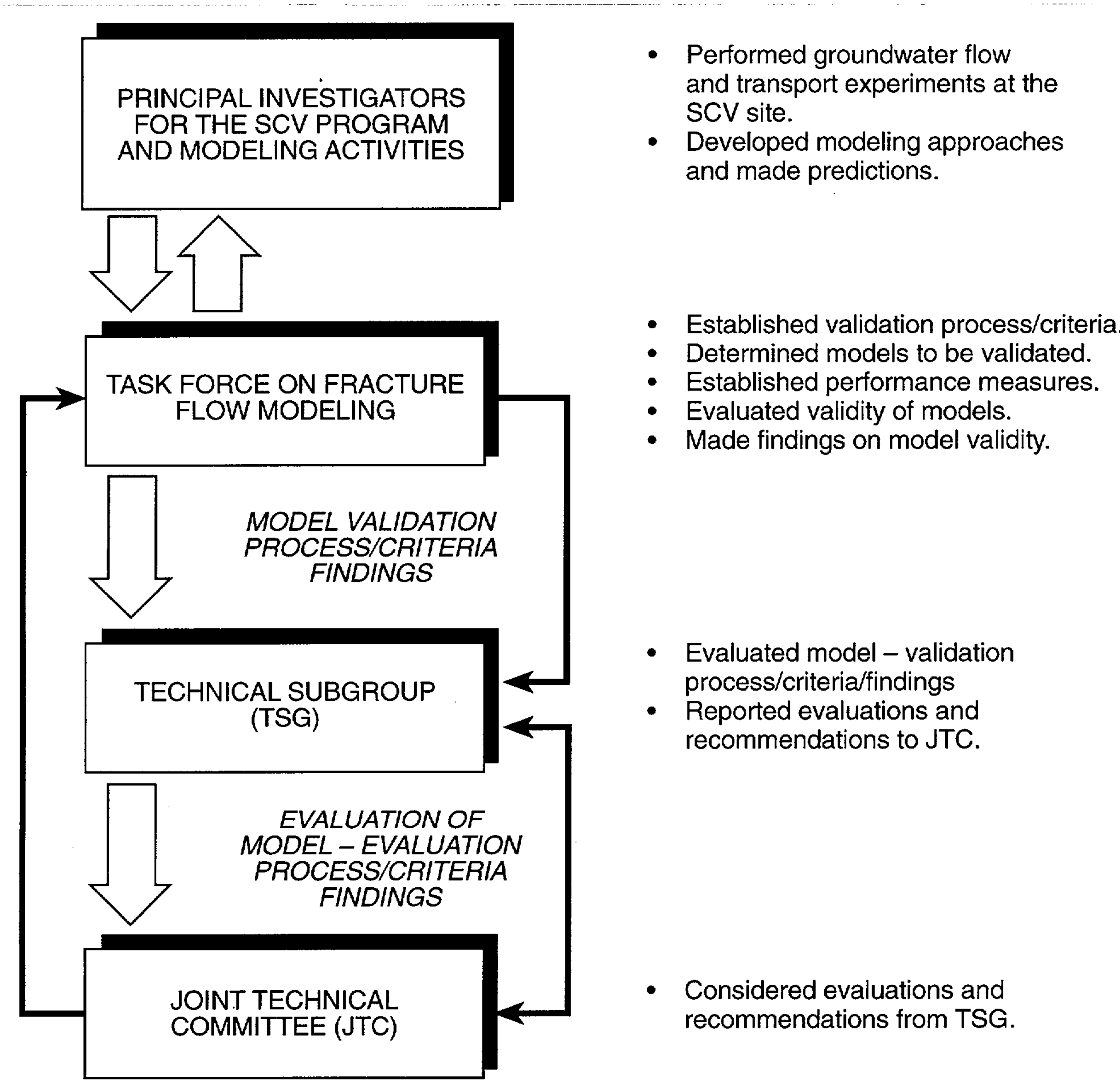


**Figure 1. Process used by the International Stripa Project for evaluating the validity of ground-water flow and transport models** (from Gnirk, 1992). "SCV" means site characterization and validation.

accuracy of geosphere transport codes used in performance assessments. To demonstrate confidence in these codes, one must document those aspects of performance assessment models that are based on accepted scientific principles and identify those aspects which are potential sources of uncertainty. Where the potential for uncertainty exists, comparison of experimental results with independently obtained model predictions provides a direct, quantitative measure of model error. The focus of INTRAVAL ultimately evolved into a study of the kind of experiments, as well as the type, quality, and quantity of data needed, to distinguish between alternative conceptual models. The validation procedure adopted for INTRAVAL is summarized in Table 1.

Eighteen laboratory, field, and natural analogue experiments, so-called *test cases*, were conducted as part of INTRAVAL. The test cases ranged from well-controlled, centimeter-scale laboratory experiments to field-scale work with less control and precision. The primary emphasis during the 13 test cases in Phase 1 was on experimental design, process identification, and model calibration, although some blind predictions of experimental outcomes were made using numerical models. Twenty-two organizations from 12 countries participated in the first phase of INTRAVAL, which began in 1987 and lasted for 3 years. Reports from INTRAVAL Phase 1 were published during the 1992-94 period [for example, see The Coordinating Group of the INTRAVAL Project (1990 and 1993)]. In analyzing the different Phase 1 Test Cases, many of the project teams reported that systematic evaluation of the experimental setup and data was required to detect unanticipated biases and artifacts introduced by errors in the experimental design.

Phase 2 was a continuation of six Phase 1 Test Cases and was designed to focus more closely on the development of validation procedures and devoted less time to optimizing the design of particular experiments to ensure the generation of data suitable for validating models. In Phase 2, which began in 1990 and concluded at the end of 1993 and involved 23 organizations from 13 countries, the remaining test cases were divided among four working groups, each of which was expected to develop practical validation strategies appropriate for its set of experiments. Overall, the activities and results of Phase 1 and Phase 2 of

| | |
|---|---|
| **Table 1.** | **INTRAVAL Validation Procedure**. Within the INTRAVAL project was a *Validation Overview and Integration Committee*, which established a validation procedure consisting of the following three major elements [adopted from The Coordinating Group of the INTRAVAL Project (1990, p. 19)]: |
| *1.* | ***Understanding and Research.*** *Without proper understanding of the processes and system structures involved, validation cannot be achieved. Thus, a thorough understanding of processes and system structures represents a major element of validation.* |
| *2.* | ***Comparisons of Theory and Modeling Calculations with Experiments.*** *This element is to study how well one is able to quantitatively predict or simulate experimental results. Discrepancies may be caused by parameter uncertainties, the statistical nature of the system, or lack of understanding. Improving our lack of understanding requires further efforts in Item 1 (above). It is important to be cautious and avoid curve-fitting without proper understanding and additional confirmatory results.* |
| *3.* | ***Peer Review and Public Scrutiny.*** *The last element involves publishing the work in the open literature, both to receive the benefits of anonymous technical review and to open the model and its validation to public scrutiny. A study whose results are in the open literature, and are examined and used by the general scientific community over long periods of time, stands a better chance of receiving the appropriate scrutiny and of being correct.* |

INTRAVAL were documented in hundreds of publications prepared by many of the 46 participating organizations. Sixteen *Test Case Reports* document the results from the various test cases. A series of annual *Progress Reports* (INTRAVAL Project Secretariat, 1988-94) provides an overview of this work as well as of the many supporting workshops and related coordinating meetings that took place. Integrated conclusions from both phases of the INTRAVAL Project have been published in several reports [see The Coordinating Group of the INTRAVAL Project (1990 and 1993); and Larsson *et al.* (1997)].

During the course of the INTRAVAL project, it was realized that a validation strategy should include more than a procedure for comparing model results with experimental data. Tsang (1991) suggests that validation is a process that should be carried out at every step of the modeling process. Among INTRAVAL participants, it has generally been agreed that a model cannot be validated in any generic sense. However, models and data may be considered to be validated with respect to a given process or a given site, implying that validation is closely related to site characterization.

Within the INTRAVAL project, the criteria used to judge model validity varied greatly from test case to test case. For INTRAVAL Test Case 1b, on uranium migration through a small crystalline rock core, for example, some modelers suggested that model validity may be assessed simply by evaluating the reasonableness of the parameters in the calibrated model (Haderman, 1992). Although close agreement of model and experimental results and the use of physically plausible parameter values do not constitute a proof of model validity, the appearance of overall consistency between model and data enhances one's confidence in the model.

A standard procedure for quantitatively validating a model is to split the experimental data into two sets, calibrate the model with one set, and compare model predictions with the second data set. The final assessment of model validity using this procedure still depends very much on the quantitative measures used to compare model predictions with experimental results and the criteria used to determine the acceptability of the fit. Some of the project teams developed statistical hypothesis testing procedures to apply quantitative criteria for accepting a model's predictions.

A validation strategy was suggested for INTRAVAL Test Case 1a, on radionuclide migration through clay cores, wherein the statistical structure of the residuals between predicted and observed breakthrough curves was examined. According to Davis *et al.* (1991), if examination of the residuals reveals little or no serial correlation, the model is presumed to adequately represent the experimental results and at least the model structure is deemed acceptable. If, on the other hand, the residuals are strongly serially correlated, the model structure is incorrect. Luis and McLaughlin (1992) described a series of statistical procedures to test the null hypothesis that model error is negligible and applied their methods to INTRAVAL Test Case 10, the Las Cruces Trench experiment. They noted that the probability of accepting a false model cannot be evaluated by their technique. For regulatory purposes, the objective to reduce the probability of accepting a false model may lead to the adoption of overly strict test case criteria that increase the likelihood of rejecting good models.

Based on their experience modeling INTRAVAL Test Case 10, Ababou *et al.* (1992) suggested that integrated performance measures, such as the first and second moments or total mass flux of a contaminant plume crossing a compliance plane, be used as acceptance criteria for model validation instead of a simple sum-of-squared residuals. In many cases, these integrated performance measures are similar to regulatory standards.

Carrera *et al.* (1990), for INTRAVAL Test Case 1a, on radionuclide migration in clay cores, and Usunoff *et al.* (1992), for INTRAVAL Test Case 1b, on uranium migration in crystalline rock cores, applied quantitative model identification methods to distinguish alternative conceptual transport models.

In concluding INTRAVAL Phase-1 (OECD/NEA/SKI, 1993), the following conclusions were reached regarding the possibilities of determining how well a model can describe an experiment and how much uncertainty is involved:

1. Careful evaluation of the experimental setup and data is needed. Biases or artifacts, that, if not explicitly accounted for, would be attributed to the medium or process measured, need to be taken into account.

2. Insight is gained by analyzing an experiment using several different conceptual models. Often the experimental data do not suffice to discriminate between these models. The spread of different models that could be fitted to an experiment gives information on the degree of uncertainty or non-uniqueness.

3. Calibration is often the only viable alternative for determining physical parameters to be used for long-term model predictions, but the resulting parameter value may depend on the calibration criterion chosen. If an automatic inverse method based on a statistical technique is used, it can be applied to rank models and to evaluate confidence intervals of the estimated parameters. The drawback is that this information is only valid under the tested hypothesis. The application of statistical inverse techniques gives no guarantee that the resulting model is a good description of reality, let alone resolving the question of whether it is the best description, given the data and the particular conceptual model analyzed.

4. Different suggested models can be compared by extrapolating the results of the models to situations relevant for waste disposal. It could certainly be argued that such a comparison of extrapolated results illustrates the (practical) degree of uncertainty related to the problem.

5. An important part of the validation work is to propose new experiments and see if these are more sensitive to the critical parameters and to the differences between alternative conceptual models. Such exercises are important in the actual design and planning of new experiments. In general, modeling before carrying out the experiment will contribute to optimizing the experimental design, to discriminate between models and to reduce the uncertainty in best-fit values of model parameters.

6. Enhanced confidence in a model is not only a matter of comparing (blind) predictions of the model with data. Other important aspects include assessing reasonableness of parameters, consistent explanation of all data, and consideration of alternative models. In assessing the quality of fits, systematic analysis of the origin of residuals, and other statistical techniques, all have their merits and pitfalls. If a prediction does not coincide with an experiment, it is necessary to explore why the experiment and the prediction differ and if the difference has any impact on the predictive power of the model.

In general, it is evident from the INTRAVAL studies that the adequacy of models predicting conditions into the far distant future cannot be proven. Ultimately, statements regarding the adequacy of predictions into the far distant future should be based on a combination of scientific reasoning and the outcome of studies such as those conducted in INTRAVAL.

### 2.3 BIOMOVS

BIOMOVS was launched in 1986 to test models designed to calculate the environmental transfer and bioaccumulation of radionuclides and other trace substances. The primary task has been to quantify the extent of uncertainty associated with model predictions as well as to identify means to reduce the uncertainty. The first phase of

BIOMOVS was completed in 1990 and focused principally on terrestrial and aquatic pathways (Hagg *et al.,* 1991; and BIOMOVS Steering Committee, 1993). However, it was not possible to consider all potentially important pathways and scenarios in the first phase of BIOMOVS. Moreover, differences among the predictions for the exposure scenarios considered were not fully resolved in the first phase. Accordingly, a second phase of this effort, designated BIOMOVS II, was initiated in 1991 and completed in 1996[13] (BIOMOVS II Steering Committee, 1996c).

The primary objectives of the two BIOMOVS studies were threefold:

1. To test the accuracy of the predictions of the environmental assessment models for selected contaminants and exposure scenarios.

2. To explain differences in model predictions caused by differences in model structure, modeling assumptions, and/or differences in selected input parameters.

3. To recommend priorities for future research to improve the accuracy of model predictions.

Although the first phase of BIOMOVS attempted to address the impact of uncertainty on the biosphere modeling exercises undertaken (see Figure 2), there was no explicit treatment of model validation (BIOMOVS Steering Committee, 1988; p. 5). More explicit treatment of the impact of model validation (and uncertainty) in the use of models was addressed in BIOMOVS II, where a formal working group, with associated sub-groups, was established. To aid in the evaluation of issues related to uncertainty and validation, the *Uncertainties and Validation Working Group* developed guidance, in the form of questions (supported by hypotheses and tests) and design criteria, that could be incorporated into the respective model tests and thus facilitate inter-comparison of model predictions by the other working groups (see BIOMOVS II Steering Committee, 1993). The activities of the respective Working Groups were reported in a series of *Progress Reports,* and study results and conclusions were documented in 16 technical reports, including a final report. At least four of these reports (BIOMOVS II Steering Committee, 1993, 1995, and 1996a and b) dealt directly with the treatment of model validation and uncertainty.

In its summary of the Phase 2 findings, the BIOMOVS II Steering Committee made a number of specific comments on ways to improve biosphere modeling exercises and assess the uncertainty associated with them [see BIOMOVS II Steering Committee (1996c, pp. 29-33)]. In addition to reporting the development of improved biosphere transport models for certain radionuclides, BIOMOVS II participants reported progress in the development of techniques for the evaluation of uncertainty types (parametric, model, scenario) and model validation. Ways to improve confidence and credibility in modeling exercises were also identified. These included, for example, the use of blind testing, improved data acquisition, and the use of guidelines in uncertainty analysis and comparison of model predictions. Finally, it was noted that the concept of forming multiple assessment groups, using (informal) expert judgment, played a significant role when interpreting the description of an exposure scenario, deriving relevant parameter values, and estimating the uncertainty, leading to large discrepancies among modeling groups. It was also observed that increased model complexity usually results in increased flexibility, making it possible to address a larger number of assessment questions. Increased model complexity, however, did not always lead to a decreased uncertainty because of the lack of site-specific data.

[13] In October 1996, the IAEA initiated the BIOMASS (**Bio**sphere **M**odeling and **Ass**essment) program as an expansion and continuation of BIOMOVS II (see IAEA, 1996).

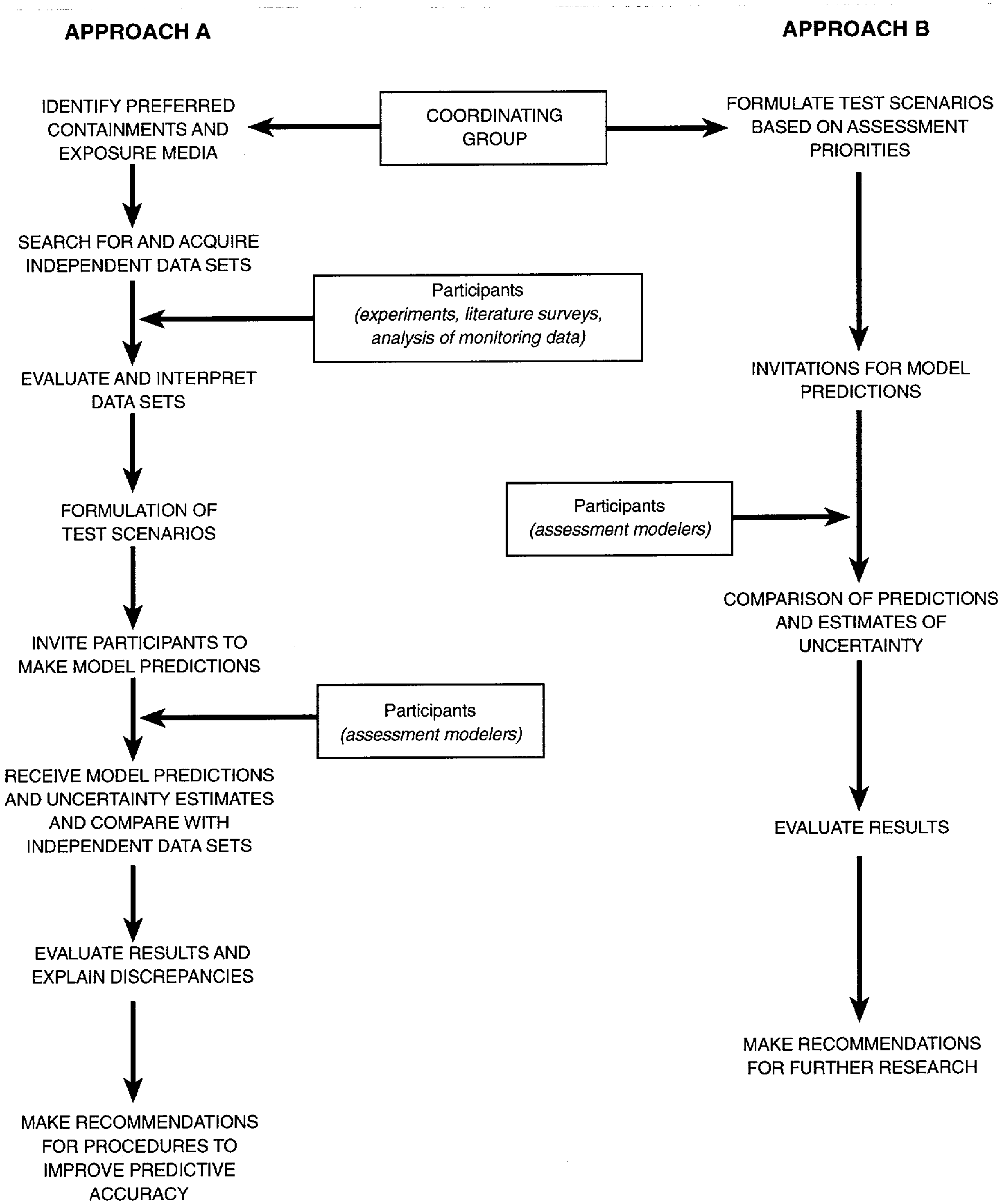


**Figure 2. BIOMOVS validation procedure** [see BIOMOVS Steering Committee (1987, p. 2)]. The BIOMOVS validation procedure relied on two parallel modeling activities. *Approach A* involved the formulation of test scenarios based on suitable data (sets) and a comparison of model predictions against the independent data sets. *Approach B* involved the comparison of model predictions and associated estimates of uncertainty for specific test scenarios selected on the basis of assessment priorities. The progress of the respective validation studies was reported in the annual *BIOMOVS Progress Reports*, and detailed information concerning the results and comparisons between the various scenarios was documented in the *BIOMOVS Technical Report Series.*

Overall, the BIOMOVS studies showed that there is a potential for very large uncertainty to be associated with any given prediction of transfer of radionuclides through the biosphere. The uncertainty in predictions in the far future, for long-lived nuclides, is many orders of magnitude. The confidence is much higher in near-future predictions of short-lived and well-studied nuclides, such as iodine-131 and cesium-137.

## 2.4 CHEMVAL

CHEMVAL loosely stands for the "validation and verification of geochemical models." The original CHEMVAL project (1987-90) was concerned with the verification (bench marking) and validation of computer-based equilibrium speciation and chemical transport models used to describe the chemistry of radioactive waste disposal systems, together with the establishment of a reviewed thermodynamic database. The project was initiated by the Commission of the European Communities (CEC) and the United Kingdom's H.M. Inspectorate of Pollution/ Department of the Environment as an extension of the CEC MIRAGE—**Mi**tigation of **Ra**dionuclides in the **Ge**osphere—project (CEC, 1984; Read and Broyd, 1987; and Côme, 1988). Seventeen organizations from eight countries participated in the initial phase of the project.

Before CHEMVAL, verification-validation procedures for chemical models varied widely among practitioners. Significant differences in model results had become apparent when different groups attempted similar problems. The differences in results were thought to be attributed to code performance, database compilation, user judgment, and conceptual model validation (Read and Broyd, 1992; p. 1422). To address these issues, the initial phase of the CHEMVAL project had the following objectives, which served to direct how the validation exercise would be conducted (*Op cit.*):

- To benchmark aqueous speciation and coupled chemical transport codes by applying them to a range of realistic waste disposal situations;
- To provide some degree of validation of aqueous speciation and coupled chemical transport computer codes; and
- To provide a "project-standard" thermodynamic reference database, tailored to the needs of radiological safety assessment.

To achieve these goals, the CHEMVAL project was divided into four consecutive stages: *Stage 1*—verification of chemical equilibrium models; *Stage 2*—attempted validation of speciation-solubility models by comparison with experimental, field, and laboratory data; *Stage 3*—verification of coupled chemical transport models; and *Stage 4*—attempted validation of coupled models.

*Stage 1* was focused primarily on activities to verify the accuracy of the thermodynamic databases contained in five chemical speciation computer codes.[14] Sufficient results were obtained to demonstrate that there was reasonable agreement among the respective computer codes and databases selected, although some discrepancies were identified and accounted for (see Read and Broyd, 1989). For *Stage 2*, the CHEMVAL exercise focused on the performance of equilibrium computer codes and databases when simulating field and laboratory data. Nineteen test problems of varying complexity were conducted at four locations (so-called *test case systems*)—the Mol (Belgium) HLW underground research facility; the proposed HLW repository in the Gorleben (Federal German Republic) salt dome; the Maxey Flats (U.S.) low-level radioactive waste disposal facility; and the Oman natural analogue complex. Overall, the equilibrium models produced reliable estimates of experimental data when used within their known frames of reference at three of the four test case systems. Moreover, Stage 2 served to highlight the limitations of current computer codes and data, and the

[14] These were *PHREEQE* (Parkhurst *et al.*, 1980); *MINEQL* (Westall *et al.*, 1976); *EQ3/6* (Wolery, 1979), *WHATIF* (Skytte-Jensen *et al.*, 1984); and CHIMERE (Coudrain-Ribstein and Jamet, 1988).

need to consider all processes of significance (Read, 1990). For example, equilibrium models were validated for actinide solubility in synthetic clay water at Mol; americium solubility and actinide complexation by EDTA[15] and citrate at Gorleben; and iron solubility and pH changes during controlled oxidation of anoxic trench leachates at Maxey Flats. [The Oman natural analogue study was less successful because of inaccuracies in field data and uncertainties surrounding the interpretation of the source mineralogy (*Op cit.*).]

In *Stage 3*, five speciation and six migration computer simulations were conducted and compared with predictions from several fully-coupled chemical speciation and transport computer codes used in *Stage 1 (MINEQL, EQ3/6*, and *CHIMERE*). The results were subsequently intercompared with three hypothetical test-case problems addressing cement dissolution, bentonite clay alteration, and sodium hydroxide injection into a siliceous aquifer. Overall, the principal investigators reported good agreement for the three test cases studied (see Read, 1991). Finally, in *Stage 4*, validation of fully-coupled models, model predictions were compared with measurements obtained from experiments involving: (i) neptunium migration through glauconitic Mol sand; and (ii) heating and acidification of Fountainbleu sands. In contrast to *Stage 2*, it was not possible to state, after *Stage 4*, that coupled chemical transport models had been validated to any extent (Read and Broyd, 1992; p. 1425). Although the column tests were reproduced, the solutions were obtained by back fitting rather than by prediction (*Op cit.*). Among other things, the first phase of the CHEMVAL project succeeded in verifying both equilibrium speciation and coupled transport codes (Read *et al.*, 1991; p. 412) as well in establishing a reviewed thermodynamic database (see Chandratillake *et al.*, 1992).

A second phase of the project, designated CHEMVAL2, was initiated toward the end of 1991 (Read and Broyd, 1992; p. 1426).

[15]ethylenediaminetetraacetic acid.

CHEMVAL2 is considered to be a logical extension of the original CHEMVAL project and is intended to build on its results. Although one of the overall aims of the CHEMVAL project was to identify the areas of greatest uncertainty in the use of predictive chemical models, CHEMVAL2 does differ somewhat from the original project by focusing on discrete technical areas that would complement the on-going MIRAGE program (*Op cit.*). Eighteen organizations from nine countries have participated in CHEMVAL2. Progress/results that have been published to date include Bruno *et al.* (1993); Read (1993); and Warrick *et al.* (1995).

## 2.5 DECOVALEX

The overall goals of the DECOVALEX can be viewed as twofold (Stephansson *et al.*, 1995; p. 350). First, the DECOVALEX studies seek to better understand the effects of T-H-M processes on the movement of radionuclides in various geologic media. Second, the project seeks to determine how coupled T-H-M processes could be described by mathematical models and computer codes. Using a variety of bench-mark tests and test cases, the respective DECOVALEX research teams seek to validate and improve the predictive capabilities of the mathematical models, numerical methods, and computer codes employed by the various research teams. For the purposes of determining when model validation had been achieved, the DECOVALEX project adopted the IAEA (1993, p. 48)[16] definition of model validation as the "... process carried out by [the] comparison of model predictions with field observations and experimental measurements. A model is considered validated when sufficient testing has been performed to ensure an acceptable level of predictive accuracy over the range of conditions over which the model may be applied...." (Stephansson, 1995; p. 387)

In the first stage of the project (between October 1991 and December 1994), designated DECOVALEX I, three phases of the project were successfully conducted. Research teams from 15 organizations

[16]Although the IAEA recognizes slightly different definitions [IAEA, 1985a (p. 26) and 1985b (p. 6)].

(so-called *national research teams*) in eight countries, participated, with funding from multiple international sources, including from NRC and SKI. Three bench-mark tests (hypothetical problems) and six test cases (actual laboratory or field experiments) were defined for parallel study by multiple research teams, at various phases. Analytical and semi-analytical solutions to coupled problems were developed whenever possible to assist in model verification and computer code validation. By design, the lessons-learned from the early phase of the analyses were to be used in subsequent phases of the project. Results from the respective research teams were presented and compared in regularly scheduled workshops, and the similarities and differences were discussed in detail. In this way, the technical soundness and scientific applicability of the models and results generated in each bench-mark test and test case underwent detailed peer-review. Results and progress of the respective research teams have been described in many reports and are summarized in Jing *et al.* (1993, 1994, 1995, and 1996).

Overall, it can be concluded from the DECOVALEX project that the capability of modeling T-H-M processes is in an early stage of development compared with that of geosphere transport models. Also, considerable work will be needed before computer codes can be developed that are capable of modeling coupled T-H-M processes realistically in geologic systems. With respect to validation and verification, another problem is the lack of applicable test cases for verification and validation purposes. The cooperative work completed thus far can be considered to have yielded three major benefits (Stephansson *et al.*, 1996; pp. xi-xii): (i) encouraging the development of coupled T-H-M codes by the national research teams and providing peer review and advice to each of them; (ii) defining both simple and realistic benchmark test problems, so that the national research teams could study and carry out code verification studies of these problems and compare computational results with those from other teams; and (iii) collecting and documenting major laboratory and field tests, so that the national research teams can use them to perform validation studies of their models and codes.

So far, the work has concentrated on different benchmark tests analyzed by different research teams. This work has led to increased understanding of algorithms suitable for T-H-M modeling and verification of the main features of T-H-M codes. Presently, the project plans to analyze a few experiments, with multiple research groups, using models in a fashion much similar to the INTRAVAL study. The duration of the DECOVALEX I project was extended into a second stage—designated DECOVALEX II—until 1998, for conduct of new experiments based on lessons learned and utilization of these new experiments for model validation. Current plans call for the integration of DECOVALEX II work with one or more on-going, large-scale, underground research projects (NIREX/Sellafield, England; and Kamaishi Mine, Japan) of coupled T-H-M processes in fractured rocks and buffer materials (Jing *et al.*, 1996; p. 40).

# 3 MODEL VALIDATION APPROACH FROM A REGULATORY PERSPECTIVE

In this section, an approach for validating HLW performance assessment models is outlined that may be suitable for HLW disposal decision-making. This approach is articulated in a form that may be useful for the repository developer to follow, to provide the needed degree of substantiation for the models used in a performance assessment submitted as part of an overall safety assessment. (However, the repository developer may need other model validation strategies for those models used for other aspects of the safety assessment.) Finally, this approach may also be of some usefulness to other regulatory authorities engaged in the development of an independent performance assessment capability.

From the regulator's perspective, the specific goals of the performance assessment model validation process outlined below should include: (i) establishing the adequacy of the model's scientific basis for its intended use; and (ii) demonstrating that the model is sufficiently accurate for its intended use. As a way of meeting these goals, a detailed step-by-step validation strategy is described (in Section 3.4). This strategy consists of: (i) defining a compliance demonstration strategy; (ii) determining the goals for model validation; (iii) determining the existing degree of validation for the model selected; (iv) comparing the validation goals with the existing degree of validation; (v) deciding whether to revise the compliance demonstration strategy; and (vi) obtaining additional information to support validation of the model, if needed.

As noted earlier, the conclusions regarding the validity of a model would be based on the various lines of evidence available and would, in the first instance, be subjectively made by the repository developer, and later corroborated by the cognizant regulatory authority. No quantitative criteria for model validation are suggested at this time.

As a potential licensee, the burden of proof regarding compliance with the applicable standards and criteria rests primarily with the repository developer rather than with the regulator. By contrast, the regulatory agency has no specific mandate for model and computer code development, nor their applications, although the regulatory agency may consider it appropriate to develop and apply models in selected or all areas as part of their oversight role. Model and code development conducted by the regulatory agency would be for the purpose of independently evaluating developer activities and plans as well as to ensure that the developer has made an adequate fundamental determination of repository safety. The regulator is not expected to remedy perceived deficiencies in the developer's programs. Thus, it is important to recognize that it should not be the responsibility of the regulatory agency to conduct independent numerical analyses. To the extent that the regulatory staff does undertake such independent analyses and any related validation activities, it is a matter of policy and technical judgment.

## 3.1 Regulatory Definition of Model Validation

At present, there is no internationally agreed-to definition of validation. As noted earlier in this White Paper, several definitions of the term "validation" have been coined in the past. However, for the purposes of the respective programs, both NRC and SKI have assumed operational definitions of validation. In the context of development of a performance assessment review capability in the U.S., the NRC staff has, for example, previously 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.... " [see Browning (1984, p. 68)].[17,18] By contrast, validation has also been proposed to be defined as "... the testing of a model in the real world...." (SKI *et al.*, 1990; p. 26).[19] Others [e.g., Bogorinski *et al.* (1988) follow IAEA's definition (1985a, p. 26) that validation is confirmed when the model and computer code "...provide a good representation of the actual processes occurring in the real system...." The problem with these and other definitions is that they are ambiguous. [For a lengthy discussion of the problems in terminology, see Oreskes *et al.* (1994).] Furthermore, since they are definitions, they do not provide any practical guidance for achieving validation, although some attempts have been made to do so in the past [see SKI *et al.* (1990, pp. 25-30)].

HLW regulators will be responsible for determining compliance of a proposed repository with the applicable environmental standards and implementing criteria. In both the U.S. and Sweden, the test of compliance with the standards and criteria is that of *reasonable assurance*. This concept recognizes that absolute assurance of compliance is neither possible nor required.[20] Instead, what is envisioned is that the repository developer should provide such information as may be necessary to convince the decision-maker that compliance with regulatory criteria will be achieved. It is important to recognize that this regulatory perspective should also hold for model validation. 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 (see Appendix B). 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 envisions only an adequate description of the phenomena for a given purpose [e.g., for reaching the necessary licensing decisions—see Davis *et al.* (1991, p. 2)]. Therefore, much greater uncertainties may be acceptable, depending on the importance of the model in the overall decision regarding repository acceptability. Thus, the distinction between a scientific approach to developing and testing models, and the regulatory approach for validating models is critical.

If, in the regulatory context, one assumes "validation" means 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's specific application. This does not imply that regulatory validation is entirely subjective. Furthermore, it is possible that the repository developer and regulator, consistent with their respective roles, could work together to define a mutually acceptable approach to validation. As described in the next section, it is possible to envisage a **process**, or, from the repository developer's point of view, a **strategy**, where both the repository developer and the regulator could reach agreement on the degree of validation needed for each model used in the repository performance assessment and how to achieve that degree of validation.

## 3.2 Goals of Model Validation in a Regulatory Setting

In formulating a strategy (or process) for validation of performance assessment models, it should be made clear that the overall goals of validation are twofold: first, establish the adequacy of the scientific basis for each model's intended use, and second, demonstrate that each model is sufficiently accurate for the purpose for which the model is used.

If a model is used, by the developer, to demonstrate repository safety or is used , by the regulator, to evaluate the developer's

---

[17] This definition was adopted from Silling (1983, p. 3).

[18] Similarly, DOE (1986) has defined validation as "...a process whose objective is to ascertain that the code or model indeed reflects the behavior of the real world...."

[19] According to this definition, model validation would be achieved by: (i) constructing a model that adequately describes the behavior of the system of interest; (ii) application of the model to predict quantities that can be observed or measured in the same or similar system; and (iii) verification that the predictions are correct (*Op cit.*).

[20] For a more detailed discussion of the Commission's views on the "reasonable assurance" concept, in the context of the generic geologic repository regulation, see NRC (1983 and 1986).

demonstration of safety, the model should be shown to have an adequate scientific basis. Speculative or conjectural models that have no plausible theoretical foundation or empirical basis will not be sufficient. Thus, the minimum threshold to be achieved in validating each performance assessment model is to establish an adequate scientific basis for regulatory credibility.

Additionally, it should be demonstrated that any model used in a safety assessment (i.e., the demonstration of compliance) 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, in a regulatory context. The validity of a model estimate 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 proposed strategy outlined here focuses 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 so-called *validation plan* is to establish, in a transparent fashion, the set of activities by which the repository developer will seek to demonstrate a level of confidence in models consistent with their importance to demonstrating compliance. In addition, the *validation strategy* will 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, and be updated, as warranted. Preferably, validation strategies should be established in the early phases of the program. For programs well underway, this issue should be assigned a high priority.

Any validation strategy should also recognize the various stages of the repository development process. In general, more confidence in models is expected in the later stages of the process owing to the collection of site characterization information throughout the process. As the repository development process progresses, new information is expected to be factored into the evaluation of model validity. The model validation approach articulated in this White Paper asks for an appraisal of the current level of scientific evidence relevant for the evaluation of a particular model given the particular stage in the development process.

For certain components of the repository system, the regulator may elect to independently develop its own performance assessment models, and may, therefore, need to establish an independent strategy for its validation. However, since the purpose of such models is not to demonstrate the safety of the repository system (nor its components) but to probe, evaluate, and corroborate the projections (and conclusions) of the repository developer's models, the goals of the regulator's validation strategy may be different than those of the developer. The regulator will need to establish the scientific credibility of its models so that the projections of its models can be compared with the projections of the repository developer's. 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.

### 3.3 Aspects of the Validation Strategy

A model validation strategy should consider two aspects of model validation in this context: (i) a description of the activities that will be implemented to gain confidence in those models used to demonstrate compliance; and (ii) documentation of the results of these activities and the logic by which the conclusions were drawn.

An important part of development of the validation strategy is the identification of the performance assessment models to be validated, considering their relative importance to the overall safety case. The primary means of determining the importance of models in the overall safety case is their *a priori* selection in the developer's *compliance demonstration strategy*. The compliance demonstration strategy indicates which components of the repository system will be relied on to isolate waste from the environment, and the degree of reliance to be placed on each component (i.e., *performance allocation*). The models associated with these components then take on the degree of reliance associated with the component. The component most relied-on should be represented by a model with a higher degree of confidence and thus a higher degree of validation. Conversely, components less-relied on may have models in which the confidence is less and therefore would need a lesser degree of validation. The *a priori* choices delineated in an initial compliance demonstration strategy should be reviewed periodically as part of the iterative process of conducting the total system performance assessment. As improved analyses (models), site data, and more complete designs become available, more robust performance assessments become possible. These more robust assessments may lead the developer to change the compliance demonstration strategy, thereby altering the level of confidence needed for the various models. The decisions regarding the importance of the particular models used should be transparent and documented for each iterative step.

Thus, an iterative process should be used for determining acceptable levels of performance for each component of an overall repository system. Based on the performance goals for the overall repository system, a conceptual system design is developed before site characterization begins. This conceptual design may describe: (i) the engineered and natural barriers to be relied on; (ii) the level of performance allocated to these 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 overall design. In the description of the level of confidence for barrier performance, the developer's plans (i.e., a *validation plan*) for validating the associated models should be referenced.

The validation plan for a model should move beyond the need for validation to include a description of the plans for implementing those specific experiments, tests, or other investigations needed to achieve the degree of validation described in the model validation strategy. Natural analogues may play an important part in the study of trace element behavior in the geological environment, for checking of model completeness and judging of process relevance, and for evaluating models for the repository system and its components [see Chapman *et al.* (1984); Murphy and Kovach (1993); and Miller *et al.* (1994)]. Moreover, both generic and site-specific tests and laboratory experiments should be considered in conjunction with natural analogue studies [see Davis *et al.*, (1991, pp. 6-7)]. As part of the validation process, formal review steps should be scheduled, for example, including *formal peer review* (Altman, 1988) of the program and the results. Because some types of validation activities may require long lead-times, it would be worthwhile to involve the regulator with respect to certain validation issues at any early time.

Still, a complete understanding of the behavior between the repository design and the site may never be fully attainable. Instead, iterations between performance assessment and **systematic review** of these assessments, leading to updated judgments of the relative importance of various sub-models and assumptions, appear to be the best approach. The more formal organization of the validation framework should also be considered, as well as the documentation. In both cases there are good reasons to consider the structure needed in the final licensing documents.

### 3.4 An Example of a Model Validation Strategy

In developing a model validation strategy, the developer will need to consider what level of validation in the modeling exercise is desirable? Although performance allocation is typically expressed *quantitatively,* it is doubtful (based on the previous validation experiences described in Section 3) that a universally applicable or *quantitative measure* of model validity can be devised and agreed to. Rather, in the process being outlined below, what is envisioned is the articulation and evaluation of *qualitative goals* and support for achieving model validation. Although *quantitative goals* would be desirable, from the regulator's perspective, confidence in a model, and its estimates, cannot generally be measured by simple quantification. Quantitative comparison of model estimates and experimental outcomes is possible, but is not the only, nor necessarily a good, measure of model validity. A model may provide acceptable estimates for a particular experiment, but by most standards would be considered invalid. For example, the Ptolemaic theory of astronomy provided fairly accurate predictions of the movement of stars, and to a degree, of the planets; but the theory was ultimately proven invalid by the theories of Kepler, Copernicus, and Newton. As an example closer to the area of HLW management, an acknowledged problem, in demonstrating the validity of ground-water flow and transport models, is that typically these models have a number of parameters that are determined empirically. Given a set of field data, these parameters, and their number, may be adjusted so that whatever degree of agreement desired can be obtained. This, of course, is not validation, but *calibration* (see Anderson and Woessner, 1992).[21] Unfortunately, many parameters describing the natural system in waste management models cannot be determined by first principles, but must be obtained by interpreting field data.[22]

The *performance allocation concept* [see DOE (1988, pp. 8.1-1–8.1-5); and Bailey (1998)] might prove useful in establishing *semi-quantitative* goals for the desired validity of the various performance assessment models. These goals might be expressed as the rank ordering of importance of the models or as a small number (two to five) of categories representing smaller or greater need for validation.

[21]In the context of ground-water modeling, calibration refers to the demonstration that a particular model is capable of producing field-measured heads and flows (i.e., the *calibration values*). Calibration is accomplished by finding a set of parameters, boundary conditions, and stresses that produce simulated heads and fluxes within a pre-established range of error (*Op cit.*, p. 223).

[22]Use of the "Ockham's Razor principle" may assist in selecting appropriate models for HLW systems. This principle may be simply stated as "...an explanation of the facts should be no more complicated than necessary...." A recent paper suggests that formal use of Bayes' theorem favors the validation of simpler models; that is, "...Bayesian analysis ... shows that a hypothesis with fewer adjustable parameters automatically has an enhanced posterior probability, because the predictions it makes are sharp...." (see Jefferys and Berger, 1992; p. 72).

An example of a validation strategy is shown in Figure 3. The steps in this example strategy are briefly described below (in italics) and each step is followed by a discussion. As shown in the figure, when implementing this strategy, what is envisioned is feedback and iteration between the various decision points.

> ***Step 1—Define a Compliance Demonstration Strategy.***
> *As a first step, the repository developer should prepare a* ***compliance demonstration strategy*** *that identifies the performance measure(s) of interest and the relationship of these measures among the various engineered and natural components of the repository. The* ***compliance demonstration strategy*** *should also include a* ***performance allocation*** *that describes which barriers and, importantly, which mathematical models (and the implementing computer codes) will be specifically relied on to demonstrate compliance.*

A prerequisite to implementing a model validation strategy is the specification of an overall strategy for demonstrating compliance with the applicable standards and regulatory criteria. This strategy should identify the quantitative post-closure performance objectives, and include plans for demonstrating that a repository site and design will meet these objectives. The strategy, therefore, would be developed by taking into account available information on the proposed repository site and design.

As discussed in Section 2, the performance objectives for a geologic repository are usually stated in terms of specific *performance indicators* (i.e., performance measures). A performance measure is a physical quantity

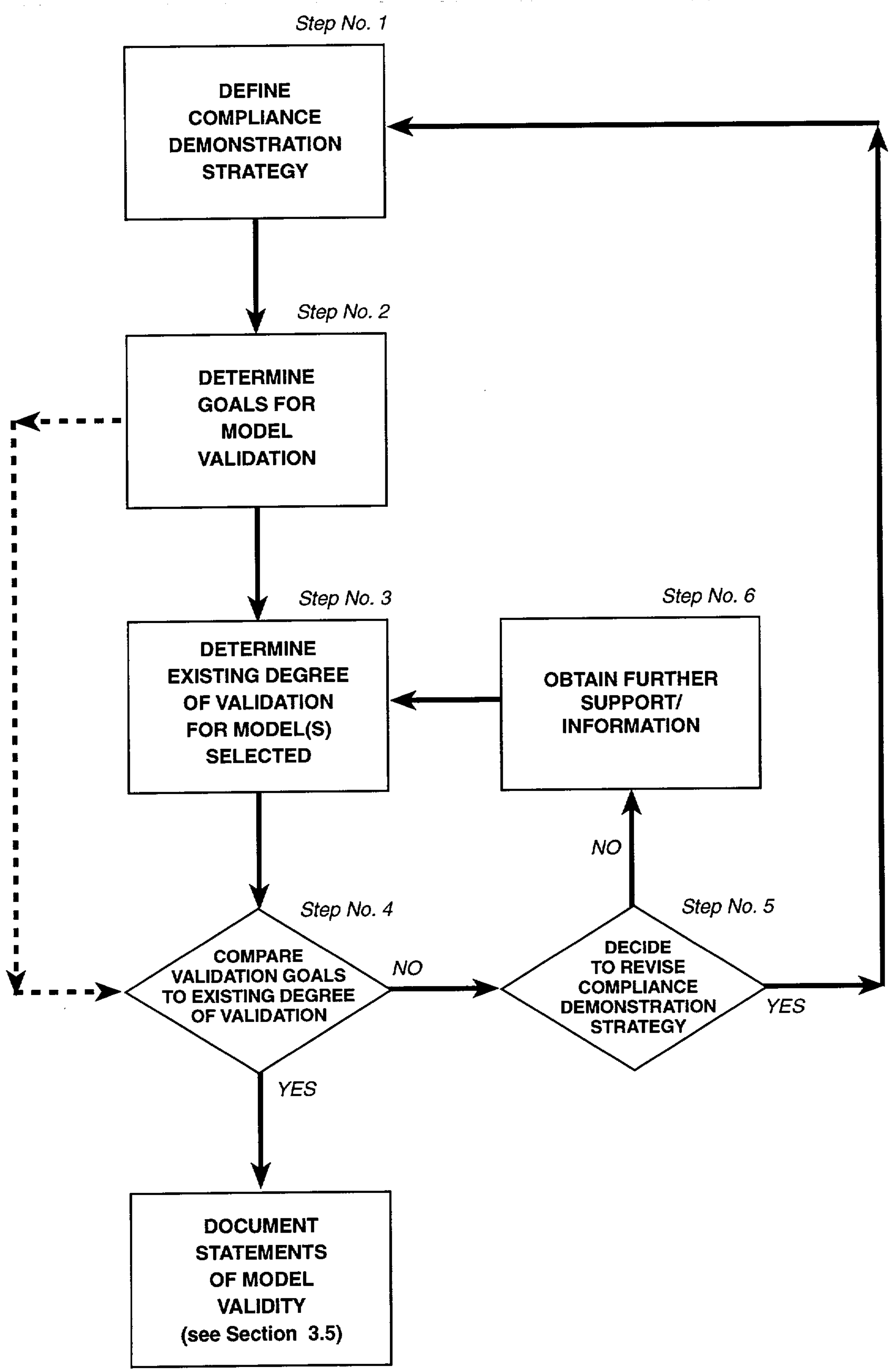


**Figure 3. Regulatory strategy for developing confidence in models.**

that 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. Performance measures may, for example, address the lifetime of a waste package or the dose to the maximally exposed individual. The minimum or maximum allowable value of the performance measure is referred to as the *performance criterion*. Such performance measures can be expressed deterministically, although the compliance demonstration would be probabilistic. Sometimes the performance measure may be estimated by a suite of computer codes, which represent models for various components of the repository or the surrounding environment. For a single set of input variables for these models, the resulting estimate of a given performance measure can then be compared with the performance limit for the particular component of interest. For multiple sets of realizations of inputs, a distribution of a given performance measure may be obtained.

Because the overall repository system (including engineered and natural barriers), is expected to be safe and, to some degree, redundant, not all the components of the repository system need to be included in the estimate of performance, to demonstrate that the performance limit is met. The repository developer may choose to include only certain components in a demonstration of compliance either: (i) because the components excluded are beneficial and drive the system performance measure toward better compliance, or (ii) because the excluded components do not significantly add to the uncertainty in the estimate of performance. The choice of which components to include in which specific models for demonstrating compliance comprises a *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 thus determines the level of validation required for a particular model (Step No. 2). That is, the models describing the component must be increasingly more valid, the more central a component is to the demonstration of compliance. *Performance allocation* can be used as the tool to identify the validation goals for various models. An iterative process is believed to be best suited for determining acceptable levels of performance for each component of an overall repository system. Each iteration, which includes a performance assessment, should be followed by a systematic review resulting in updated judgments of the relative importance of various sub-models and assumptions, based on the current assessment.

> ***Step 2—Determine the Goals for Model Validation.***
> *For each model important to demonstrating compliance, qualitatively define the goals of the validation exercise. For example, in defining these goals, one should: (i) identify the performance measure to be predicted by the model; (ii) describe the relative level of confidence required for a particular model; and (iii) define the type of confidence or supporting information desired to substantiate validation.*

The next step in the validation strategy is to define the goals of the validation effort. Specifically, what level of validation in the modeling exercise is desirable? In general, the overall goal of model validation is to remove the conceptual uncertainties associated with models and to demonstrate that the system being described by a particular model is sufficiently well understood to support the model's intended use. Thus, for each model important to demonstrating compliance, it will be necessary to define qualitatively the goals of the validation exercise.

> ***Step 3—Determine the Existing Degree of Validation for the Model(s) Selected.***
> *The literature should be reviewed to determine the extent to which the models selected have been previously used and validated. For newly developed models or for those models for which there is little relevant experience, determine the extent to which there is empirical support for the scientific basis or application of the proposed model.*

Models used in performance assessment should 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 scientific or engineering practice, and which lack technical credibility and thus are potential sources of uncertainty. Having identified the models and performance measures of interest (Step No. 1), an effort should be undertaken to review and assimilate the scientific and technical literature relevant to the model applications of interest. Many physical processes have been modeled previously and there is a large body of technical and scientific experience in this regard. However, because many studies might have previously 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, to apply the model to the compliance measure in question and subsequently arrive at a judgment as to what level of model support exists.

To establish scientific support for a given model, *one should both* examine its theoretical basis and 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 estimates against empirical information. For models used in HLW disposal, the evaluation of the application of scientific principles and the comparison of model results with empirical data are limited, because each disposal 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.

> ***Step 4—Compare the Validation Goals to the Existing Degree of Validation.***
> *If the existing degree of support for the model (based on previous experience or scientific evidence) exceeds the required degree of validation, no further validation activities are needed and the information-collecting activities should be documented as part of the compliance demonstrations. However, if the existing experience or scientific evidence does not satisfy the model validation needs, then decide on what further actions can be undertaken to achieve the model validation goals.*

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 a decision should be made either to reduce the reliance placed on the modeled 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.

> ***Step 5—Decide Whether to Revise the Compliance Demonstration Strategy.***
> *In the preceding step, it was determined that existing experience or scientific evidence was not sufficient for the validation needs of the preferred model. Thus, the repository developer is left with two choices: either obtain additional support for the preferred model or revise the* ***compliance demonstration strategy.*** *If it is decided to continue with the same* ***compliance demonstration strategy*** *(i.e., retain the same level of reliance on the preferred model), then the repository developer should obtain further support (i.e., evidence of validation) for the preferred model.*
>
> *However, the repository developer may find that it is not practical or possible to validate the preferred model with new or additional information. In this instance, the only recourse is to repeat Step Nos. 2 through 4, and revise the* ***compliance demonstration strategy*** *(Step No. 1), including redefined* ***model validation goals, performance allocation, etc.***

As stated above, most proposed repository systems appear to have sufficient margins of safety and/or redundancies to allow several different choices regarding which system components are to be included in a compliance demonstration and the relative importance of such components in making the demonstration (i.e., their respective contributions to performance).

Overall safety criteria, stated as *performance limits*, may be fixed, but the means of demonstrating safety for a HLW repository are judgmental and suffer from large uncertainties [see SKI *et al.* (1990; pp. 20-33); Fehringer (1991); and Ekberg (1995)]. An assessment may rely on an understanding believed to be better for some parts of the system than others—this may affect priorities and assumptions in the performance assessment. Poor understanding mandates the use of conservative assumptions. Supposedly better understanding allows the use of less conservative assumptions. If compliance is reached in a justifiable manner, for example, by consideration of relevant uncertainties, no further validation is needed. If not, there are two choices: (i) changing reliance on various models in the performance assessment (Step No. 1); or (ii) the search for more support for models and assumptions (Step No. 6).

A hypothetical Swedish example of performance allocation could be to suppose that the description of far-field migration is very uncertain, and to suppose that the waste package canister is very stable. However, scrutinizing the validity of waste package canister stability models may make it necessary to put more confidence into the retarding mechanisms of the far field (Step No. 1). Alternatively, one might put more efforts into further waste package canister corrosion research (Step No. 6).

For the proposed U.S. geologic repository at Yucca Mountain, a hypothetical example of performance allocation could be to model migration solely in the unsaturated zone with the object to show that the cumulative releases of radionuclides from the unsaturated zone to the saturated zone migration are sufficiently small to meet compliance with the regulatory limit. Similarly, if the saturated zone were thought to be robust enough, the geologic repository could be assumed to release radionuclides directly into the saturated zone, and modeling of the unsaturated zone migration could be avoided.

> ***Step 6—Obtain Additional Information to Support Validation of the Preferred Model.*** *This step presumes: (i) that the existing level of support needed for validation of the preferred model is insufficient based on previous experience or scientific evidence; and (ii) that it is practical and possible to obtain the needed experience or scientific evidence. As noted earlier, both generic and site-specific tests, laboratory experiments, and natural analogue studies are ways to achieve these goals. After the new information is obtained (from one or more of the approaches described below), it should be combined with the body of supporting evidence (identified earlier in Step No. 3). This revised, aggregated body of technical and scientific experience would then be compared with the initial model validation goals first identified in Step No. 2.*

The goal of model validation is to obtain sufficient confidence, commensurate with the models' intended use, that the models are able to describe the behavior of interest in the real system. Confidence is gained in two ways: (i) by examining the theoretical or scientific basis for the model to assure it is sufficient for the application of the model; and (ii) by evaluating the application of scientific principles in the model to assure the application is appropriate, which can be accomplished by reviewing application of the principles in similar circumstances.

Normally, the validity of the application of scientific principles is achieved by comparing the model estimates against empirical information. For models used in the HLW program, the evaluation of the application of scientific principles and the comparison of model results with empirical data are limited, because the data are collected for a specific site and each specific site has a unique set of natural/physical idiosyncrasies. This limits the ability to generalize. Another limitation on the use of empirical data is that data over the time scales of interest are not available for experiments. Evidence from field data and natural analogue studies are available for longer time and spatial scales, but there is generally greater uncertainty regarding the environmental conditions prevailing for the system over these long times. One potential way to deal with such difficulties is to employ the chosen site for a repository as a natural analogue itself. Thus, the ability to describe the past evolution of hydrology and geochemistry of a site might provide strong evidence of the ability to accurately predict its future evolution.

Some additional sources of information that might be useful in providing support (i.e., validation) for models are briefly described below.

***Theoretical support for models:*** Generally, theoretical support should exist for models used in performance assessment and, in some cases, this information may be used to substitute for experimental evidence. Virtually all the models used in performance assessment 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 (e.g., very simple "models" that consist solely of correlations of variables in experiments of limited scope) should have further theoretical substantiation, if the intention is to extrapolate the model results to times or conditions not encompassed by the original data.

***Need for experiments:*** Progress in validation may have to be based on additional experimental evidence. This can be the case both for non site-specific 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. In the parlance used in the preceding section, 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 the validation problem may be to show how well the experiment (measurements) made in the site characterization really characterize the site. This will generally involve experiments both at the actual site as well as at other sites, to confirm the reliability of the site characterization techniques. Alternatively, confirmation of the application of established principles may be directed to the interpretation of site characterization data, rather than to their representativeness or quality. Experimental evidence is more likely to be needed to support the scientific basis for a model for areas in which the theoretical and empirical scientific bases are incomplete and/or developing; for example, the formation and migration of colloidal contaminants in the geosphere.

On the other hand, experiments might not be necessary for validation of performance assessment models. In Section A.2 of Appendix A, the concept of a hierarchy of performance assessment models is introduced (see Figure A-1). In this hierarchy less detailed models are derived from more detailed models by assumptions that simplify the models, but retain their essential behaviors. If the theoretical and empirical bases for the detailed models are well-established, 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-after may be gained by development of a deeper understanding of the process in question than is actually needed in the performance assessment model. Such an understanding can be attained simply by more detailed and careful modeling of important processes.

***Planning and analysis of experiments:*** Performance assessment models can be compared with laboratory experiments, field tests, and/or natural analogues to add to the confidence that they are, over some time period and with some degree of accuracy, able to describe the relevant behavior of the real system—i.e., quantitatively estimate the performance measure. Because of 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 total system and subsystem performance assessment models with laboratory and field test data is a much more difficult task, since the function of these models is to estimate system performance over large spatial and time scales. In general, there will be more confidence in a model that compares favorably to several types of evidence ranging in spatial and time scales, such as laboratory experiments and natural analogues. The utility of these various sources of substantiating evidence is described below.

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. When conducting such planning, there are some "good practices" that may prove helpful. These include the following:

- Identify potential alternative conceptual models and then design tests that will discriminate among the various alternatives and a preferred model, if there is one.
- 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.
- To the extent practical, design experiments to 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 database (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.
- If a model intended to estimate for large times and spatial scales fails to estimate for smaller scales, to the degree required, this is strong evidence to invalidate the model.
- 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; unlike scientific validation, only the phenomena and variables of interest need be explained and related by the model.
- Agreement with an experiment is insufficient, alone, to validate a model; the scientific basis for the model should be supported and scrutable. Ensure that generally accepted scientific principles (e.g., those describing flow of ground

water through a porous matrix) apply to the actual conditions anticipated for a specific repository.

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

- Data used to develop or calibrate a model cannot be used 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 model's ability to simulate the same system under different conditions. Thus, at least two data sets (or a partitioned set of data) is required for model validation.

- 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.

- Maintain records of model development and testing, and subject these records 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.

These good practices may be applied to the various lines of evidence for building confidence in models. Further descriptions of the lines of evidence follow.

***Laboratory experiments:*** Laboratory experiments are useful because: (i) they are performed in a controlled environment that minimizes uncertainty in initial and boundary conditions; and (ii) the experiments can be performed on samples that exhibit relatively little geometric variability or whose variability can be measured. However, the use of laboratory experiments in validation efforts is limited because of: (i) the inability to perform tests on either long time scales or large spatial scales required for assessing the performance of a HLW repository; (ii) the difficulty in testing some coupled processes; and (iii) 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 are not equivalent to field conditions, which may produce phenomena that do not actually occur *in situ*).

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

***Natural analogues:*** To increase the temporal and spatial scales, evidence from studies of natural analogues could be used. In some sense, nature could be considered to have initiated experiments that could be used for validation. Transport of radionuclides from naturally-occurring uranium deposits, and 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 repository system scales. 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, and thus may account for the reported lack of natural analogue-derived data in direct use in various performance assessment projects [see OECD/Nuclear Energy Agency (1997, p. 26)].

A potential drawback of natural analogues is their complexity. The problem of demonstrating understanding of an analogue, with inferred historic evolution, is difficult. In addition, the question can always be raised as to whether the studied analogue has sufficient relation to the system or subsystem modeled. Still, long-term field tests and natural analogues make it possible to study coupled systems. This constitutes an invaluable check that no essential process or coupling effect has been omitted. Furthermore, even a qualitative fit between standard performance assessment models and results from long-term experiments provides increased confidence in the model. Above all, the possibility to employ the site of a repository as a natural analogue in its own right should be stressed again. By developing credible models able to describe the evolution of, for example, the hydrology and geochemistry of a site to conformity with observed conditions, the developer will have provided confidence in the ability of the same models to estimate the future evolution of the site.

## 3.5 Documenting Statements of Model Validity

In any potential license application, documentation of the validation strategy employed will be of great importance in judging the credibility of safety calculations supporting the compliance demonstrations . The overall objective of this documentation should be development of a framework that would facilitate the acceptance (or rejection) of models used, based on transparent and logical reasoning. In this regard, all steps associated with the implementation of the model validation strategy (Section 3.4 of this White Paper), including the *compliance demonstration strategy* and other reasoning employed, should be openly and transparently documented .

Moreover, peer review, including international cooperative efforts similar to those described in Section 2, is believed to be a fundamental contributor to judging the validity of performance assessment models. In this regard, the extent to which the performance assessment models used have undergone (or may undergo) peer review (e.g., OECD/ Nuclear Energy Agency Peer Review Team, 1997) should also be included in the documentation step. However, it should be emphasized that peer review efforts may be considered of little value in regulatory decision-making unless the material to be reviewed finds support from quantitative analyses of experiments or other proofs, for example, derived from interpretation of natural analogues or more detailed modeling.

# 4 SUMMARY

Model validation has been a topic of debate in the area of HLW management for several decades. There is considerable disagreement within the technical community about what constitutes model validation and how to achieve it. Furthermore, the process for validating performance assessment models is likely to be complex, including some combination of laboratory experiments, field tests, natural analogue studies, and other theoretical investigations. In a regulatory context, information needed to build confidence in the models used in the safety demonstrations should be developed in a timely fashion, pursuant to an acceptable model validation strategy. This White Paper has attempted to articulate one such strategy.

Key aspects of this validation strategy were described in the preceding pages. In describing this strategy, an applied—rather than a pure—science perspective was used; the model only needs to be valid enough to provide estimates useful for the particular application. From the regulator's perspective, the specific goals of the performance assessment model validation process outlined should include: (i) establishing the adequacy of the model's scientific basis, for its intended use; and (ii) demonstrating that the model is sufficiently accurate for its intended use. As a way of meeting these goals, a detailed step-by-step validation strategy was described. This strategy consists of: (i) defining a compliance demonstration strategy; (ii) determining the goals for model validation; (iii) determining the existing degree of validation for the model selected; (iv) comparing the validation goals with the existing degree of validation; (v) deciding whether to revise the compliance demonstration strategy; and (vi) obtaining additional information to support validation of the model, if needed.

The need for model validation depends on the importance of the model in the safety demonstration; the more reliance is placed on a component, the greater is the need for the implementor to provide information that the models describing it are valid. Performance allocation describes the relative importance of components. Certain good practices are advocated to assist in the development of activities to build confidence and in their interpretation. The repository developer is expected to focus on two essential elements: (i) procedures for the development of confidence in models; and (ii) documentation of the results from confidence-building activities. The need for model validation is expected to rest mainly with the repository developer; the regulator's program is expected to have a lesser need for model validation given the regulator's role.

# 5 REFERENCES

Altman, W.D., *et al.*, "Peer Review for High-Level Waste Repositories," U.S. Nuclear Regulatory Commission, Generic Technical Position, NUREG-1297, February 1988.

Ababou, R., B. Sagar, and G. Wittmeyer, "Testing Procedures for Spatially Distributed Flow Models," *Advances in Water Resources*, 15:181–198 [1992].

American Nuclear Society, *Proceedings of the Topical Meeting on Site Characterization and Model Validation (Focus '93), September 26-29, 1993, Las Vegas, Nevada*, 1993.

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

Bailey, J.N., "Post Closure Safety Case," in "U.S. Nuclear Waste Technical Review Board Summer Meeting," Las Vegas, Nevada, June 24, 1998. [Presentation made by TRW Environmental Safety Systems Inc. on behalf of the U.S. Department of Energy/Yucca Mountain Site Characterization Project.]

BIOMOVS Steering Committee, "BIOMOVS Progress Report 4—June 1987," Stockholm, Sweden, Swedish Radiation Protection Institute [Statens Strålskyddsinstitut], 1987.

BIOMOVS Steering Committee, "Technical Report for the B3 Scenario: Release of Radium-226 and Thorium-230 to a Lake," Stockholm, Sweden, Swedish Radiation Protection Institute, BIOMOVS II Technical Report 1, September 1988.

BIOMOVS Steering Committee, "Biospheric Model Validation Study [BIOMOVS]: Final Report," Stockholm, Sweden, Swedish Radiation Protection Institute, BIOMOVS Technical Report 15, 1993.

BIOMOVS II Steering Committee, "Guidelines for Uncertainty Analysis Developed by Participants in the BIOMOVS II Study (Developed by a Working Group of the BIOMOVS II Study)," Stockholm, Sweden, Swedish Radiation Protection Institute, BIOMOVS II Technical Report 1, July 1993. [Includes a bibliography of methodologies and case studies regarding the treatment of uncertainty and validation.]

BIOMOVS II Steering Committee, "Quantitative and Qualitative Guidelines for the Comparison of Environmental Model Predictions (Developed by a Working Group of the BIOMOVS II Study)," Stockholm, Sweden, Swedish Radiation Protection Institute, BIOMOVS II Technical Report 3, March 1995.

BIOMOVS II Steering Committee, "Using Validation Studies of Water Flow and Radionuclide Transport in Vegetated Soils Using Lysimeter Data (Developed by a Working Group of the BIOMOVS II Study)," Stockholm, Sweden, Swedish Radiation Protection Institute, BIOMOVS II Technical Report 15, September 1996a.

BIOMOVS II Steering Committee, "Uncertainty and Validation: Effect of User Interpretation on Uncertainty Estimates (Developed by a Working Group of the BIOMOVS II Study)," Stockholm, Sweden, Swedish Radiation Protection Institute, BIOMOVS II Technical Report 7, November 1996b.

BIOMOVS II Steering Committee, "An Overview of the BIOMOVS II Study and its Findings (Developed by a Working Group of the BIOMOVS II Study)," Stockholm, Sweden, Swedish Radiation Protection Institute, BIOMOVS II Technical Report 17, November 1996c.

Borgorinski, P., *et al.*, "The Role of Transport Code Verification and Validation Studies in Licensing Nuclear Waste Repositories in the Federal Republic of Germany," *Radiochimica Acta*, 44/45:367-372 [1988].

Browning, R.E., U.S. Nuclear Regulatory Commission/Division of Waste Management, Memorandum to J.G. Davis, U.S. Nuclear

Regulatory Commission/Office of Nuclear Material Safety and Safeguards [Subject: Revised Modeling Strategy Document for HLW Performance Assessment], Washington, D.C., July 23, 1984.

Bruno, J., *et al.*, "Status Review of CHEMVAL2 Technical Areas—June 1992," London, United Kingdom, Her Majesty's Inspectorate of Pollution/Department of Environment, DoE Report No. DOE/HMIP/RR/93.014, 1993.

Campbell, J.E., and R.M. Cranwell, "Performance Assessment of Radioactive Waste Repositories," *Science*, 239[4846]: 1389-1392 [1988].

Carrera, J., *et al.*, "An Approach to Process Identification: Application to Solute Transport through Clays," in K. Kovar (ed.), *Calibration and Reliability in Groundwater Modeling [Model-Care '90 Conference Proceedings], September 3–6, 1990, The Hague, The Netherlands,* International Association of Hydrological Sciences Publication No. 195, pp. 231–240.

Chandratillake, M., W.E. Falck, and D. Read, "CHEMVAL Project: Guide to the CHEMVAL Thermodynamic Data Base," London, United Kingdom, Her Majesty's Inspectorate of Pollution/Department of Environment, DoE Report No. DOE/HMIP/RR/92.094, 1992.

Chapman, N.A., I.G. McKinley, and J.A.T. Smellie, "The Potential of Natural Analogues in Assessing Systems for Deep Disposal of High-Level Radioactive Waste," Wettingen, Switzerland, Nationale Genossenschaft für die Lagerung radioaktiver Abfälle [National Cooperative for the Disposal of Radioactive Waste—Nagra], Nagra NTB 84-41, December 1984.

Côme, B., "An Overview of the CEC Project MIRAGE (Migration of Radionuclides in the Geosphere, 1985-1989)," *Radioactive Waste Management and the Nuclear Fuel Cycle,* 10[1-3]: 41-49 [1988].

Commission of the European Community, "CEC Co-ordinated Project MIRAGE," Brussels, Belgium, EUR 12053 EN, 1984.

Coudrain-Ribstein, A., and Ph. Jamet, "Le Modèle Géochimique CHIMERE—Priniciples et Notice d'Emploi," Centre d'Informatique Géologique, Ecole Nationale Supérieure des Mines de Paris, Rapport ENSMP-CIG LHM/RD/88/35, 1988.

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

Dverstorp, B. S. Norrby, and S. Wingefors, "Premises for Regulations Concerning Safety in Connection with the Final Disposal of Spent Nuclear Fuel, etc.," Stockholm, Sweden, Swedish Nuclear Power Inspectorate, SKI Memorandum 97017, March 26, 1997.

Ekberg, C., "Uncertainties in Safety Analysis: A Literature Review," Stockholm, Sweden, Swedish Nuclear Power Inspectorate, SKI Report 95:17, May 1995.

Eisenberg, N.A., *et al.*, "A Proposed Validation Strategy for the U.S. DOE Office of Civilian Radioactive Waste Management Geologic Repository Program," in Swedish Nuclear Power Inspectorate, *Symposium on Verification and Validation of Geosphere Performance Assessment Models Proceedings (GEOVAL '87), April 7-9, 1987, Stockholm, Sweden,* 2:341–374 [1988].

Eisenberg, N.A., *et al.*, "Model Validation from a Regulatory Perspective: A Summary," in Swedish Nuclear Power Inspectorate/OECD Nuclear Energy Agency, *International Symposium on Validation through Model Testing with Experiments (GEOVAL '94), October 11-14, 1994, Paris, France,* pp. 421–434 [1995].

Fairhurst, C., *et al.* [compilers], *OECD/NEA International Stripa Project—Overview Volume I: Executive Summary*, Stockholm, Sweden, Swedish Nuclear Fuel and Waste Management Co. [SKB], January 1993.

Fehringer, D.J., "Staff's Approach for Dealing with Uncertainties in Implementing the EPA HLW Standards," U.S. Nuclear Regulatory Commission, SECY-91-242 [Commission Paper], August 6, 1991.

Gnirk, P., "The Process for Validation Groundwater Flow and Transport Models in the SCV Program," Stockholm, Sweden, Swedish Nuclear Fuel and Waste Management Company, Stripa Project Technical Report 87-15, 1992.

Gnirk, P. [compiler], *OECD/NEA International Stripa Project—Overview Volume II: Natural Barriers*, Stockholm, Sweden, Swedish Nuclear Fuel and Waste Management Co., January 1993.

Gray, M.N. [compiler], *OECD/NEA International Stripa Project—Overview Volume III: Engineered Barriers,* Stockholm, Sweden, Swedish Nuclear Fuel and Waste Management Co., January 1993.

Haderman, J., "Phase 2, Case 1b, Uranium Migration in Crystalline Rock; Borecore Pressure Infiltration Experiments," Paris, France, Organization for Economic Cooperation and Development, 1992.

Hagg, C., K. Skagius, and N.G.B. Johansson, "Some Conclusions from a Biospheric Model Validation Study (BIOMOVS)," in American Nuclear Society, *International High-Level Radioactive Waste Management Conference [Proceedings], April 28–May 3, 1991, Las Vegas, Nevada,* pp. 902-907.

Hodgkinson, D., "A Compilation of Minutes for the Stripa Research Task Force on Fracture Flow Modeling," Stockholm, Sweden, Swedish Nuclear Fuel and Waste Management Company, Stripa Project Technical Report 92-09, 1992.

Hodgkinson, D. and N. Cooper, "A Comparison of Measurements and Calculations for the Stripa Validation Drift Inflow Experiment," Stockholm, Sweden, Swedish Nuclear Fuel and Waste Management Company, Stripa Project Technical Report TR 92-07, 1992a.

Hodgkinson, D. and N. Cooper, "A Comparison of Measurements and Calculations for the Stripa Tracer Experiments," Stockholm, Sweden, Swedish Nuclear Fuel and Waste Management Company, Stripa Project Technical Report TR 92-20, 1992b.

International Atomic Energy Agency, "Performance Assessment for Underground Radioactive Waste Disposal Systems," Vienna, Austria, Safety Series No. 68 (Procedures and Data), July 1985a.

International Atomic Energy Agency, "Safety Principles and Technical Criteria for the Underground Disposal of High-Level Radioactive Wastes," Vienna, Austria, Safety Series No. 99 (IAEA Safety Standards), November 1985b.

International Atomic Energy Agency, *Radioactive Waste Management Glossary (Third Edition)*, Vienna, Austria, 1993.

International Atomic Energy Agency, "BIOMASS Newsletter," Vienna, Austria, Waste Safety Section/Division of Radiation and Waste Safety, No. 2, December 1996.

INTRAVAL Project Secretariat, "INTRAVAL Progress Reports," Stockholm, Sweden, Swedish Nuclear Power Inspectorate, 10 vols., 1988-94.

Jefferys, W.H., and J.O. Berger, "Ockham's Razor and Bayesian Analysis," *American Scientist*, 80:64–72 [1992].

Jing, L., *et al.*, "DECOVALEX—Mathematical Models of Coupled T-H-M Processes for Nuclear Waste Repositories: Report of Phase I," Stockholm, Sweden, Swedish Nuclear Power Inspectorate, SKI Report 93:31, June 1993.

Jing, L., *et al.*, "DECOVALEX—Mathematical Models of Coupled T-H-M Processes for Nuclear Waste Repositories: Report of Phase II," Stockholm, Sweden, Swedish Nuclear Power Inspectorate, SKI Report 94:16, October 1994.

Jing, L., *et al.* (ed.), "DECOVALEX—Mathematical Models of Coupled T-H-M

Processes for Nuclear Waste Repositories: Report of Phase III," Stockholm, Sweden, Swedish Nuclear Power Inspectorate, SKI Report 95:80, December 1995.

Jing, L., *et al.*, "DECOVALEX—Mathematical Models of Coupled T-H-M Processes for Nuclear Waste Repositories: Executive Summary for Phases I, II, and III," Stockholm, Sweden, Swedish Nuclear Power Inspectorate, SKI Report 96:58, June 1996.

Konikow, L.F., "Discussion of the 'The Modeling Process and Model Validation,' by Chin-Fu Tsang" *Ground Water*, 30(4):622-623 [1992].

Kozak, M.W., and N.E. Olague, "Evaluation of a Performance Assessment Methodology for Low-Level Radioactive Waste Disposal Facilities: Validation Needs," U.S. Nuclear Regulatory Commission, NUREG/CR-5927, Vol. 2, February 1995. [Prepared by the Sandia National Laboratories.]

Larsson, A., "The International Projects INTRACOIN, HYDROCOIN, and INTRAVAL," *Advances in Water Resources:* 15:85-87 [1992].

Larsson, A., *et al.*, "The International INTRAVAL Project—Phase 2, Summary Report," The Coordinating Group of the INTRAVAL Project/SKI, Paris, France, Organization for Economic Cooperation and Development, 1997.

Luis, S., and D. Mclaughlin, "A Stochastic Approach to Model Validation," *Advances in Water Resources:* 15:15-32 [1992].

Mercer, J.W., and C.R. Faust, "Ground-Water Modeling: An Overview," *Ground Water*, 18(2):108-115 [1980].

Miller, W., *et al.*, "Natural Analogue Studies in the Geologic Disposal of Radioactive Wastes," Wettingen, Switzerland, Nationale Genossenschaft für die Lagerung radioaktiver Abfälle, Nargra Technical Report 93-03, January 1994.

Murphy, W.M., and L.A. Kovach (eds.), *The Role of Natural Analogues in Geologic Disposal of High-Level Nuclear Waste," [Proceedings of the July 23–25, 1991, Workshop held in San Antonio, Texas]*, Center for Nuclear Waste Regulatory Analyses, CNWRA 93-020, September 1993.

OECD/Nuclear Energy Agency, "Can Long-Term Safety be Evaluated?—An International Collective Opinion," Paris, France, Radioactive Waste Management Committee of the OECD, NEA, IAEA, and the CEC, Disposal of Radioactive Waste Series, 1991.

OECD/Nuclear Energy Agency, "Lessons Learnt from Ten Performance Assessment Studies," Paris, France, Working Group on Integrated Performance Assessments of Deep Repositories, Disposal of Radioactive Waste Series, 1997. [Includes case studies on performance assessments for Yucca Mountain conducted by the U.S. Department of Energy and NRC.]

OECD/Nuclear Energy Agency Peer Review Team, "The SKI SITE-94 Project: An International Peer Review Carried Out by an OECD/NEA Team of Experts," Stockholm, Sweden, Swedish Nuclear Power Inspectorate, SKI 97:41, October 1997.

OECD/Nuclear Energy Agency/Swedish Nuclear Power Inspectorate, "The International INTRAVAL Project: Phase-1, Summary Report (Executive Summary)," Stockholm, Sweden, Statens Kärnkraftinspektion, December 1993.

OECD/Nuclear Energy Agency/Swedish Nuclear Power Inspectorate, *GEOVAL'94—Validation through Model Testing Symposium Proceedings, October 11-14, 1994, Paris, France,* Organization for Economic Cooperation and Development, Safety Assessment of Radioactive Waste Repository Series, 1995.

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

Parkhurst, D.L., D.C. Thorstenson, and L.N. Plummer, "PHREEQE—A Computer

Program for Calculating Mass Transfer for Geochemical Reactions in Ground Water," U.S. Geological Survey, Water Resources Investigations Report, WRI-80-96, 1980. [Revised 1990.]

Read, D. (ed.), "CHEMVAL Project: Report on Stage 2—Application of Speciation Models to Laboratory and Field Data Sets," Brussels, Belgium, Commission of the European Communities, EUR 13124 EN, 1990.

Read, D. (ed.), "CHEMVAL Project: Report on Stages 3 and 4—Testing of Coupled Transport Models," Brussels, Belgium, Commission of the European Communities, EUR 13675 EN, 1991. [Also published by Her Majesty's Inspectorate of Pollution as DoE Report No. DOE/HMIP/RR/91.003.]

Read, D., "The CHEMVAL2 Project: Status Report," in H. von Maravic and J. Moreno, "Progress Meeting on Migration of Radionuclides in the Geosphere" *[Proceedings of the Workshop on Migration of Radionuclides in the Geosphere held in Brussels, Belgium, April 9-10, 1992], Brussels, Belgium*, Commission of the European Communities, EUR 14690 EN, pp. 79-90, 1993.

Read, D. and T.W. Broyd, "Verification and Validation of Predictive Computer Programs Describing the Near and Far Field Chemistry of Radioactive Waste Disposal Systems," *Radiochimica Acta*, 44/45:407-415 [1987].

Read, D. and T.W. Broyd, "Report on Stage 1 of Project CHEMVAL/MIRAGE: Verification of Speciation Codes," Brussels, Belgium, Commission of the European Communities, EUR 12237 EN, 1989.

Read, D. and T.W. Broyd, "The International CHEMVAL Project," in American Nuclear Society/American Society of Civil Engineers, *High-Level Radioactive Waste Management: Proceedings of the Third International Conference, April 12-16, 1992, Las Vegas, Nevada,* 2:1421-1428 [1992].

Read, D., T.W. Broyd, and B. Côme, "The CHEMVAL Project—An International Study Aimed at the Verification and Validation of Equilibrium Speciation and Chemical Transport Models," in Statens Kärnkraft-inspektion/OECD Nuclear Energy Agency, *Symposium on Validation of Geosphere Flow and Transport Models Proceedings (GEOVAL '90), May 14-17, 1990, Stockholm, Sweden,* Organization for Economic Cooperation and Development, pp. 417-424 [1991].

Silling, S.A., "Final Technical Position on Documentation of Computer Codes for High-Level Waste Management," U.S. Nuclear Regulatory Commission, NUREG-0856, June 1983.

Skytte-Jensen, B., H. Jensen, and O.J. Nielsen, "WHATIF-AQ: A Computer Program for Speciation Calculations," Roskilde, Denmark, Risø National Laboratory, Report 382-83-7 WASDK, 1984. [Also published by the CEC under the title "Geochemical Modeling—WHATIF-AQ: A Computer Programme for Speciation Calculations," EUR 10119 EN, 1985.]

Stephansson, O. (ed.), "Special Issue: Thermo-Hydro-Mechanical Coupling in Rock Mechanics," *International Journal of Rock Mechanics and Mining Sciences (and Geomechanics Abstracts)*, 32[5]:387-535 [1995].

Stephansson, O., *et al.*, "Development of Coupled Models and Their Validation against Experiments—DECOVALEX," in OECD/ Nuclear Energy Agency/Swedish Nuclear Power Inspectorate, *GEOVAL'94 Validation through Model Testing Symposium Proceedings, October 11-14, 1994, Paris, France,* Organization for Economic Cooperation and Development, Safety Assessment of Radioactive Waste Repository Series, pp. 349-361 [1995].

Stephansson, O., L. Jing, and C.-F. Tsang (eds.), *Coupled Thermo-Hydro-Mechanical Processes of Fractured Media: Mathematical and Experimental Studies—Recent Developments of [the] DECOVALEX Project for Radioactive Waste Repositories*, Amsterdam, Holland, Elsevier Science, Developments in Geotechnical Engineering, Series 79, 1996.

Swedish Nuclear Power Inspectorate, "INTRACOIN (International Nuclide

Transport Code Intercomparison Study)—Final Report Level 1 (Code Verification), Stockholm, Sweden, SKI 84:3, September 1984.

Swedish Nuclear Power Inspectorate, "INTRACOIN (International Nuclide Transport Code Intercomparison Study)—Final Report Levels 2 and 3 (Model Validation and Uncertainty Analysis)," Stockholm, Sweden, SKI 86:2, May 1986.

Swedish Nuclear Power Inspectorate, *Symposium on Verification and Validation of Geosphere Performance Assessment Models Proceedings (GEOVAL '87), April 7–9, 1987, Stockholm, Sweden,* 3 vols. [1988].

Swedish Nuclear Power Inspectorate, "The International HYDROCOIN Project: Groundwater Hydrology Strategies for Performance Assessment of Nuclear Waste Disposal—Summary Report," Paris, France, SKI Coordinating Group of the HYDROCOIN Project, OECD/Nuclear Energy Agency, 1992.

Swedish Nuclear Power Inspectorate, Swedish National Institute of Radiation Protection, and Swiss Nuclear Safety Inspectorate, "Regulatory Guidance for Radioactive Waste Disposal—An Advisory Document," Stockholm, Sweden, Swedish Nuclear Power Inspectorate, SKI Technical Report 90:15, 1990.

Swedish Nuclear Power Inspectorate/OECD/Nuclear Energy Agency, *International Symposium on Validation of Geosphere Flow and Transport Models (GEOVAL '90), May 14-17, 1990, Stockholm, Sweden,* Organization for Economic Cooperation and Development [1991].

Swedish Nuclear Fuel and Waste Management Co.,"Summary of the International Stripa Project 1980-92" in "SKB Annual Report 1992 (Including Summaries of Technical Reports Issued during 1992)," Stockholm, Sweden, SKB Technical Report TR 92-46, May 1993.

The Coordinating Group of the HYDROCOIN Project, "The International HYDROCOIN Project—Summary Report," Paris, France, HYDROCOIN Project Secretariat/SKI, Paris, France, Organization for Economic Cooperation and Development, 1992.

The Coordinating Group of the INTRAVAL Project, "The International INTRAVAL Project—Background and Results," Paris, France, INTRAVAL Project Secretariat/SKI, Organization for Economic Cooperation and Development, 1990.

The Coordinating Group of the INTRAVAL Project, "The International INTRAVAL Project—Phase 1, Summary Report," Paris, France, INTRAVAL Project Secretariat/SKI, Organization for Economic Cooperation and Development, December 1993.

The Radiation Protection and Nuclear Safety Authorities in Denmark, Finland, Iceland, Norway, and Sweden, *Disposal of High-Level Radioactive Waste—Consideration of Some Basic Criteria*, Stockholm, Sweden, Swedish Radiation Protection Institute, 1993. [The so-called *Nordic Document*. Available from SKI.]

Tsang, C.-F., "The Modeling Process and Model Validation," *Ground Water,* 29(6):825-831 [1991]. [Also see Konikow (1992), which includes Tsang's reply.]

U.S. Department of Energy, "Environmental Assessment: Yucca Mountain Site Nevada Research and Development Area, Nevada," Nevada Operations Office/Yucca Mountain Project Office, Nevada, DOE/RW-0073, Vol. II, May 1986.

U.S. Department of Energy, "Site Characterization Plan, Yucca Mountain Site, Nevada Research and Development Area, Nevada," Nevada Operations Office/Yucca Mountain Project Office, Nevada, DOE/RW-0199, 9 vols., December 1988.

U.S. Department of Energy, *et al.*, "Risk Assessment: A Survey of Characteristics, Applications, and Methods used by Federal Agencies for Engineered Systems (submitted to the Federal Coordinating Council for Science, Engineering, and Technology, Ad Hoc

Working Group of Risk Assessment), Washington, D.C., U.S. Nuclear Regulatory Commission, November 1992. [Includes contributions from eight Federal agencies, including NRC.]

U.S. Nuclear Regulatory Commission, "10 CFR Part 60, Disposal of High-Level Radioactive Wastes in Geologic Repositories, Technical Criteria [Final Rule]," *Federal Register*, Vol. 48, No. 120, June 21, 1983, pp. 28194–28229.

U.S. Nuclear Regulatory Commission, "10 CFR Part 60, Disposal of High-Level Radioactive Wastes in Geologic Repositories Technical Criteria [Conforming Amendments]," *Federal Register*, Vol. 51, No. 118, June 19, 1986, pp. 22286–222301.

U.S. Nuclear Regulatory Commission, "NRC Model Simulations in Support of the Hydrologic Code Intercomparision Study (HYDROCOIN): Level 1—Code Verification, Office of Nuclear Regulatory Research, NUREG-1249, vol. 1, March 1988.

U.S. Nuclear Regulatory Commission, "Severe Accident Risks: An Assessment for Five U.S. Nuclear Power Plants—Final Summary Report," Office of Nuclear Regulatory Research, NUREG-1150, 3 vols., December 1990.

Usunoff, E., *et al.*, "Part III—Interpretation of Tracer Experiments on Granite Samples (INTRAVAL Test Case 1b)," in Empresa Nacional de Residous Radiaactivos, S.A. [ENRESA], "Spanish Participation in the International INTRAVAL Project," Barcelona, España, ENRESA Publicacion Technica Numbre 03/92, February 1992, pp. 263-426. [Prepared by the Centro Internacional de Methodos Numericos en Ingeniería. English edition.]

Warrick, P., A. Hall, and D. Read, "Modeling Metal Humic and Fluvic Acid Interactions: CHEMVAL2," in *NEA Workshop on Binding Models Concerning Natural Organic Substances in Performance Assessment [Proceedings], September 14-16, 1994, Bad, Switzerland,* OECD/Nuclear Energy Agency, pp. 109-135, 1995.

Westall, J.C., J.L. Zachary, and F.M.M. Morel, "MINEQL: A Computer Program for the Calculation of Chemical Equilibrium Composition of Aqueous Systems," Cambridge, Massachusetts, Massachusetts Institute of Technology, Department of Civil Engineering, Technical Note No. 18, [1976].

Witherspoon, P.A. (ed.), "Geological Problems in Radioactive Waste Isolation: A World Wide Review" *[Proceedings of Workshop W3B, 28th International Geological Congress, Washington, D.C., July 15–16, 1989]*, Berkeley, California, Lawrence Berkeley Laboratory, LB-29703, January 1991. [Prepared for DOE.]

Witherspoon, P.A. (ed.), "Geological Problems in Radioactive Waste Isolation: Second Worldwide Review," Berkeley, California, Lawrence Berkeley Laboratory, LB-38915, September 1996. [Prepared for DOE.]

Wolery, J.T., "Calculation of Chemical Equilibrium Between Aqueous Solution and Minerals: The EQ3/6 Software Package," Livermore, California, Lawrence Livermore National Laboratory, UCRL–52658, February 1979. [Revised as Version 7.0, dated October 1992.]

# APPENDIX A
# U.S./SWEDISH PERFORMANCE ASSESSMENT MODELING FRAMEWORKS

In this section, overviews of the United States (U.S.) and Swedish regulatory frameworks governing the disposal of high-level radioactive waste (HLW) are first presented, followed by a brief description (outline) of the salient features of performance assessment modeling. There are significant differences in the regulatory philosophies of the U.S. and Sweden. However, despite these differences, both the U.S. and Sweden intend to employ predictive models to evaluate the long-term performance of their HLW repositories, and thus both countries are concerned with developing consensus on what degree of confidence is desirable in models that are used to demonstrate the long-term performance of a geologic repository for HLW.

## A.1 U.S. Regulatory Framework

Section 121(a) of the Nuclear Waste Policy Act (NWPA), as amended (Public Law 97-425), called for the U.S. Environmental Protection Agency (EPA) to promulgate generally applicable environmental standards for the management, storage, and disposal of HLW. In addition, NWPA prescribed (Section 121(b)) that the EPA standards be implemented by NRC as part of the procedural and technical regulations it was to promulgate for the licensing of geologic repositories for the disposal of HLW. EPA promulgated its standard as 40 CFR[1] Part 191 (EPA, 1985); the NRC standard is 10 CFR Part 60 (NRC, 1983 and 1986).

### A.1.1 The EPA Standard[2]

As currently written, the EPA standard seeks to protect public health and safety by limiting cumulative releases of radionuclides over the first 10,000 years after disposal (40 CFR 191.13). Additional requirements limit radiation exposure to individual members of the public for 10,000 years after disposal (40 CFR 191.15), and also limit concentrations of radionuclides in ground water for 10,000 years after disposal (40 CFR 191.16). The standard limits the cumulative release of radionuclides to the accessible environment (or biosphere) over the first 10,000 years after disposal.[3] The limit, which is expressed in probabilistic terms, must not be exceeded with a probability of 0.1, or 10 times the limit must not be exceeded with a probability of 0.001. This cumulative release limit is to be met at the accessible environment boundary, defined as the atmosphere, land surfaces, surface waters, oceans, and the lithosphere outside the controlled area of 100 square kilometers (39 square miles), or the lithosphere not more than 5 kilometers (8 miles) from a proposed repository. The limits on radiation protection and concentration in ground water are to be met for expected conditions and processes only.

The EPA Standard specifies that a performance assessment is to be done to provide reasonable expectation that there is compliance. The EPA Standard further states that the performance assessment should:

- Identify processes and events that might affect the disposal system;
- Evaluate the effects of the processes and events on the performance of the disposal system;
- Determine the cumulative release of radionuclides caused by all significant processes and events, considering uncertainties.

The results of such a performance assessment would yield a complementary cumulative

1 "CFR" stands for the *Code of Federal Regulations*, which is the comprehensive set of U.S. Federal regulations compiled in a multi-volume set that is published annually.

2 The EPA Standard described here applies to the disposal of transuranic wastes at the Waste Isolation Pilot Plant but does not apply to the proposed repository site of Yucca Mountain that is being evaluated in accordance with NWPA, as amended.

3 Cumulative release is defined as the time integral of release rate to the accessible environment, summed-up over each radionuclide, weighted according to its radiotoxicity.

distribution function (CCDF) of cumulative release for comparison with the limits.

### A.1.2 NRC's Geologic Repository Disposal Requirements

As noted above, 10 CFR Part 60 incorporates 40 CFR Part 191 as the overall performance requirement for a geologic repository. The requirements in 10 CFR 60.112 set an overall system performance objective that amounts to meeting EPA's containment requirements, whereas certain other sections (10 CFR 60.113) set forth subsystem performance objectives.[4]

### A.1.3 Recent Regulatory Developments[5]

Since the late 1970s, EPA has been engaged in setting standards to protect the public and environment from the potential hazards associated with the geologic disposal of HLW (see U.S. General Accounting Office, 1993). However, in July 1987, the U.S. Court of Appeals for the First Circuit in Boston vacated Subpart B of the HLW standards and remanded the rule to EPA for further consideration, as noted earlier. After the 1987 court decision, EPA was working to revise its environmental standards (*Op cit.*). However, before EPA could complete its work, Congress enacted the Energy Policy Act of 1992 (EnPA), which changed EPA's standard-setting authority.

Through EnPA, Congress directed EPA to promulgate new environmental standards specific to a potential geologic repository for HLW at Yucca Mountain, Nevada. EnPA stipulates that EPA's new standards are to be based on and consistent with the findings and recommendations of the NAS. Once the final standards are promulgated, EnPA directs the staff to modify its requirements to conform to the new EPA standards. Under EnPA, NRC has 1 year to make the necessary modifications.

In advising EPA on the technical bases for newly revised disposal standards for Yucca Mountain, Section 801(a)(2) of EnPA directed the NAS to provide EPA with recommendations on the following issues:

- Whether health-based standards based on doses to individual members of the public from releases to the accessible environment ... will provide a reasonable standard for protection of the health and safety of the general public?
- Whether it is reasonable to assume that a system of post-closure oversight of the repository can be developed, based on active institutional controls, that will prevent an unreasonable risk of breaching the repository's engineered or geologic barriers or increasing the exposure of individual members of the public to radiation beyond allowable limits?
- Whether it is possible to make scientifically supportable estimates of the probability that the repository's engineered or geologic barriers will be breached as a result of human intrusion, over a period of 10,000 years?

As noted above, the NAS has published its findings and recommendations for a site-specific environmental standard for Yucca Mountain. Among the NAS findings and recommendations was a key recommendation that the revised standard limit individual risk to a member of the public and abandon the existing quantitative release limit with its implied population-protection basis (National Research Council, 1995).[6] Specifically, the NAS has recommended that the level of protection provided for in the new environmental standard should be comparable to that level of risk that may be acceptable to society at large, given that society currently tolerates certain *involuntary* risks—e.g., in the

[4] In developing a Yucca Mountain-specific disposal regulation, the National Academy of Sciences (NAS) asked NRC to reconsider the role of numerical subsystem performance objectives. In its findings and recommendations, the NAS has indicated that the potential of such a requirement might lead to the sub-optimization of repository design (National Research Council, 1995; p. 126).

[5] As noted earlier, the need for future revision to 10 CFR Part 60 is currently under consideration. These revisions, however, are not expected to change the staff's views expressed earlier.

[6] The proposal for a *health-based standard* (e.g., limiting individual dose or risk) suggested by the EnPA is somewhat different from the existing regulatory structure.

range of $10^{-5}$ to $10^{-6}$ per (Smith, 1995). To demonstrate that a geologic repository can be designed to provide comparable protection to society, the NAS therefore recommended that assessments of individual risks be conducted for certain target populations, in the Yucca Mountain vicinity, using the approach specified by the International Commission on Radiological Protection—ICRP (1985)—the so-called "critical group" approach (Smith, 1995).[7] As EPA considers this particular recommendation, the NRC will need to explore how its existing performance assessment review capability might be revised to accommodate such a standard.

The forthcoming revisions to both EPA's standards and NRC's conforming regulations notwithstanding, it is expected that the NRC regulations will continue to require compliance with applicable EPA environmental standards as the overall system performance objective for the repository and that demonstration of compliance with that objective will continue to necessitate a quantitative performance assessment to estimate post-closure performance of the repository system (see Kotra *et al.*, 1998). Thus, because the revised EPA standard is expected to be probability-based, the demonstration of compliance will also be probability-based. However, a probabilistic evaluation is useful regardless of the nature of the standard because of the large uncertainties in predicting future geologic repository performance.

## A.2 Swedish Regulatory Framework

There are three principal organizations involved with the Swedish geologic repository program. The Swedish nuclear utilities have formed an operator-owned company—the Swedish Nuclear Fuel and Waste Management Company (or Svensk Kärnbränslehantering AB–SKB)—to manage and dispose of radioactive waste and spent nuclear fuel. In addition to building and operating the repositories and other waste and fuel management facilities, SKB has a mandate to conduct research and development and other activities required for the safe management and disposal of the wastes. The Swedish Radiation Protection Institute (or Statens Strålskyddsinstitut—SSI) and the Swedish Nuclear Power Inspectorate (or Statens Kärnkraftinspektion—SKI) are the authorities responsible for independently overseeing the work of SKB and the waste producers, as well as developing the necessary regulatory criteria: SSI develops the radiation protection standards needed to protect public health and the environment, whereas SKI develops safety requirements that ensure the that the facilities, including the geologic repositories, will maintain the level of protection against radiological consequences that SSI establishes in its regulations. For this purpose, SKI also develops requirements on how to demonstrate compliance with the SSI regulations.

A discussion of points of departures for the SKI regulations was presented in 1997 (Dverstorp *et al.*, 1997). The main principles build on earlier recommendations issued jointly by the Nordic safety authorities (the so called *Nordic Document*, see below) as well as other principles discussed and issued by the ICRP and the International Atomic Energy Agency (IAEA).

### A.2.1 The *Nordic Document*

Over the past three decades, the radiation protection and nuclear safety authorities in Denmark, Finland, Iceland, Norway, and Sweden have been working to develop common recommendations for nuclear safety. In the late 1980s, these authorities established a working group with the goal of preparing recommendations concerning basic criteria for the disposal of HLW (including spent nuclear fuel). As a result of their efforts, a draft report was issued in 1989 for comment both within and outside the Nordic countries (The Radiation Protection and Nuclear Safety Authorities in Denmark, Finland, Iceland, Norway, and Sweden, 1989).[8] Based on the comments received, the document was revised

[7] Also, the ICRP defines risk to a critical group in terms of dose. The term "dose" generically refers to the quantity of radiation absorbed by body organs or tissue. NRC defines "dose" for its regulatory purposes at 10 CFR Part 20. In its recommendations, the NAS adopts the ICRP terminology to its proposed risk-based framework.

[8] Hereafter referred to as the *Nordic Document*.

and re-issued in 1993. Together with other such international positions, the *Nordic Document* (1993) has constituted a common basis for the development of national HLW disposal policies and regulations within the respective Nordic countries.

In its current form, the *Nordic Document* (*Op cit.*, p. 30) proposes that:

> "...Up to reasonably predictable time periods, the radiation doses to individuals from the expected evolution of the disposal system shall be less than 0.1 millisievert (mSv)[9] per year. In addition, the probabilities and consequences of unlikely disruptive events shall be studied, discussed and presented in qualitative terms, and wherever practicable, assessed in quantitative terms in relation to the risk of death corresponding to a dose of 0.1 mSv per year...."

The direct application of this criterion was foreseen for times shorter than about 10,000 years. In addition, a release requirement is proposed for the (very) long term (*Op cit.*, p. 32):

> "The radionuclides released from the repository shall not lead to any significant changes in the radiation environment. This implies that the inflows of the disposed radionuclides into the biosphere, averaged over long time periods, shall be low in comparison with the respective inflows of natural alpha emitters."

It is also stated that the constraint on activity inflow should be such that: (i) the resulting peak individual doses are generally not in excess of the dose limit; (ii) the resulting activity concentrations in primary recipients at the site fall within the range of typical concentrations of long-lived alpha emitters in similar environments; and (iii) the activity inflow from all HLW to be disposed of globally is very low compared with the inflow of natural long-lived alpha emitters.

The *Nordic Document* also addresses validation in the requirements on safety assessments (*Op cit.*, p. 34):

> "Compliance of the overall disposal system with the radiation protection criteria shall be convincingly demonstrated through safety assessments which are based on qualitative judgement and quantitative results from models that are validated as far as practicable."

In addition, it is stated (*Op cit.*, p. 35) that both models and data to be used in safety assessments should be validated as far as is reasonable by evidence from laboratory tests and field observations, including natural analogues.

### A.2.2 Swedish Draft Regulations

After publication of the *1993 Nordic Document*, SSI and SKI worked to develop the necessary policies and regulations for radioactive waste disposal.

In 1995, SSI published its preliminary views on potential radiation protection criteria for geologic disposal (see SSI, 1995). These views were consistent with the *Nordic Document*, existing ICRP principles for protection against ionizing radiation (see ICRP, 1985), and Swedish statutes.[10] In light of these considerations, as well as SSI's evaluation of the on-going SKB research program (e.g., SKI, 1993), SSI published its proposed regulations for comments in 1997 (see SSI, 1997). In September 1998, SSI decided to issue a revised version of these regulations, which is planned to come into force by February 1999. This regulatory revision is expected to include:

- Provisions for the *optimization* and the use of *best available (technology) techniques* for limiting the release of radiative substances and the harmful effects of the releases on human health and the environment.

[9] One mSv equals approximately 10 mrem.

[10] The Act on Nuclear Activities (*ca.* 1984) and the Radiation Protection Act (*ca.* 1988).

- Quantitative requirements for the calculation of an annual global collective dose over a period of 10,000 years.

- Qualitative requirements to ensure preservation of biodiversity and the sustainable use of biological resources. Biological effects on habitats and ecosystems shall be described and reported.

- A constraint in the annual risk (of acquiring cancer or genetic damage) to a representative individual belonging to the most exposed group of $10^{-6}$.

- Specification of a 1000-year compliance (time) period in which repository performance must be analyzed based on the present biosphere and other conditions, and the most likely evolution of such conditions. For times after 1000 years, the analysis shall be based on different possible scenarios for evolution of the repository, its surroundings, and the biosphere. For all time-scales, the analysis shall include a case based on present biosphere conditions.

- Evaluation of repository performance after human intrusion.

SKI is developing regulations intended to give requirements on the construction and operation of a repository and on how to demonstrate compliance with the radiation protection criteria prescribed by SSI. In 1997, SKI proposed some general considerations and recommendations that would form the basis for subsequent safety regulations. In that document (Dverstorp *et al.*, 1997), SKI described the basic design goals for a geologic repository and the issues and areas to be addressed in the safety assessment. These ideas have been further developed, and presently (Autumn, 1998) it is expected that the regulations will cover the following items.

***Basic Safety Requirements:*** As noted earlier, the aim of SKI's safety requirements will be to ensure that facilities for the disposal of radioactive waste maintain the level of protection against radiological consequences established by SSI. With this goal in mind, SKI proposes that a geologic repository be designed and operated to fulfill the following basic safety requirements:

- The level of risk associated with geologic disposal shall be comparable to that of other nuclear activities.

- The assessment of safety shall consider the risks of the repository and be based on a performance assessment that includes relevant scenarios.

- The geologic repository shall be designed in a way that obviates the need for post-closure monitoring.

- Repository safety (performance) shall be based on several functions of technical and natural barriers, in which the failure of any one barrier function would not impair the overall performance of the repository.

- The results of on-going research programs, on the long-term safety of the repository, shall be reported regularly to SKI until the repository is sealed.

- A quality assurance (QA) process should be implemented for the operation phase, to ensure that the various barriers of the repository will perform as intended, and that the *Safety Analysis Report* (SAR) is properly updated.

***Design and Construction Requirements:*** The repository shall be designed and constructed to meet design basis requirements specified in the SAR. Best available techniques shall be considered, and it shall be possible to retrieve spent nuclear fuel without impairing safety (at least until the repository is sealed).

***Safety Assessments Requirements:*** The SAR will need to be renewed within 10 years. The SAR shall contain a description of the facility and its operation, as well as a performance assessment for both the operation phase and the long-term, post-operational phase. The long-term assessment will address the following:

- *Time scale of the analysis and scenario selection:* The performance of the repository should be assessed for as long as safety functions are required, for 10,000 years as a minimum, but not longer than 1 million years. If so needed, the performance assessment shall be based on a *main scenario*, covering the next glacial cycle of about 100,000 years. Additional *less-likely scenarios* will have to be analyzed separately or superimposed on the main scenario (e.g., climate variants, tectonic events, future human actions) as well as *residual scenarios* (e.g., extreme natural events, human intrusion, and "what-if" cases).

- *Safety and radiation protection indicators for various time scales:* A "multiple lines-of-reasoning" approach will be encouraged, possibly employing environmental concentrations and fluxes of radionuclides as safety indicators complementary to risk and dose for the long to very-long time scales (greater than 1000 years).

- *Scenarios, models, uncertainties*: A systematic approach should be adopted with regard to the identification of scenarios, processes, and uncertainties that could affect repository performance. In addition, a comprehensive documentation must be provided of how validation of models, assumptions, and data for the intended use has been achieved.

No future Swedish regulations are expected to require formal calculations of CCDFs for cumulative releases. Therefore, both deterministic analyses and sensitivity analyses performed probabilistically as well as through parameter variations, in a deterministic manner, are viewed to be important elements of performance assessment calculations. Model uncertainties should be analyzed by applying several alternative models. SKI also wishes to emphasize that the evaluation of safety assessments is not restricted to checking whether or not estimated radioactive releases comply with the criteria that have been specified. Most of the evaluation work focuses on investigating whether all essential processes and their inherent interactions have been included, or addressed, in the assessment and whether they have been correctly described from a technical/scientific perspective.

## A.3 Components of Performance Assessment Modeling

Performance assessment modeling is closely linked to the site characterization and engineering design process. Data acquired during site characterization combined with proposed repository design features are used in developing not only the conceptual model(s) to be used in the analysis but also in selecting the parameter values that would eventually be used to obtain numerical estimates of the performance. For most natural systems, and even for some engineered systems, formulation of a single unique conceptual model is an exception; more often several classes of conceptual models can be derived that satisfy the known (either observed or postulated by accepted theories) constraints to varying degrees. To ensure that model uncertainty is appropriately gauged, a useful approach is initially to formulate as many alternate conceptual models as are consistent with known information; this initial set of conceptual models is then screened as more data become available, thereby building confidence in those models remaining.

Formulation of the conceptual model(s) is the first and perhaps the most important step in the performance assessment modeling process. It is in the formulation of the conceptual model(s) that the level of detail to be incorporated in a model is decided. It includes decisions on: (i) governing equations (e.g., equations corresponding to the conservation of mass, momentum, and energy); (ii) constitutive equations (e.g., relations between stress and strain, between saturation and hydraulic conductivity, and between corrosion potential and corrosion rate, etc.); (iii) geometry of the system (e.g., one-dimensional, two-dimensional, or three-dimensional, single or double porosity, detail of heterogeneity, discrete fractures or equivalent continuum, faults, geologic structures, etc.); (iv) boundary conditions (e.g., fixed or time variable, spatially uniform or nonuniform, specified fluxes or specified values of variables, etc.); and (v) initial conditions (e.g., ambient distribution of pressures, temperatures, concentrations, etc).

Formulation of conceptual models for the natural system presents problems that may not be encountered for engineered systems. The primary reason for this is that unlike engineered components, the geology of the natural system cannot be designed—it can only be investigated and described. In this regard, because of the variability and heterogeneity of geologic systems in space and in time (e.g., Schumm, 1991), it is often difficult to determine the types, kinds, and amounts of data necessary to adequately characterize a candidate site. Moreover, because of limitations in current testing/exploration technology, the tests themselves can disturb the very physical properties being measured, and possibly impair the desirable barrier properties of the site. The conceptual model of the site, therefore, is often based on imperfect information resulting in considerable extrapolation of sparse quantitative data which, in turn, could possibly lead to large conceptual errors in the analysis. In view of this, it is especially important that alternate models be formulated and tested to account for possible biases in conceptual model formulation.

Next, mathematical models corresponding to each one of the conceptual models are formulated. Depending on what system-state parameters are selected (e.g., temperature, pressure, concentrations, current density, free energy, etc.), it is possible to develop alternate mathematical models for the same conceptual model. However, alternative mathematical models are normally equivalent, and the selection of system-state parameters (and hence the special form of the mathematical models) is based on the advantages (e.g., numerical stability, computational efficiency, desire for obtaining a closed-form analytic solution) a particular model provides in implementation. It is expected that the mathematical models will embody the accepted scientific principles such as the conservation of mass, momentum, and energy. Use of a purely empirical model, such as a regression model, may be acceptable if the results of such a model are applied only within the range of observations on which it was based.

Because of the complexity of the overall repository system, it is anticipated that the performance assessment models will be implemented on digital computers. Thus, the mathematical models will be implemented in a discrete form using computer codes. High-quality maintenance of these codes, through appropriate QA (Silling, 1983), verification and benchmark testing, and thorough documentation are essential to building confidence in the estimates of these codes and may help in building confidence in the model on which they are based. However, the testing of such computer codes, even though necessary, is not sufficient in and of itself for achieving model validation.

Models (and corresponding computer codes) for the overall system, which include realistic details of all system components, can become very complex and computationally impractical. In such cases, it is advantageous to perform modeling through the use of a hierarchy of models (see Figure A-1). The very detailed models of individual processes are the first level in this hierarchy and are normally used to understand the sensitivity of a process to parameter variations and external forces. Such models are necessary to demonstrate conservatism of assumptions and to provide a basis for the second-level models in the hierarchy. In the second level, a subset of the detailed models with some simplifications is coupled to study and understand the interfaces between processes. In the third and final level, all component models are further simplified and coupled to formulate a *total-system performance assessment (TSPA) model*. A caution to be kept in mind is that if the coupling between the detailed models is strong and nonlinear, then it may not be easy to

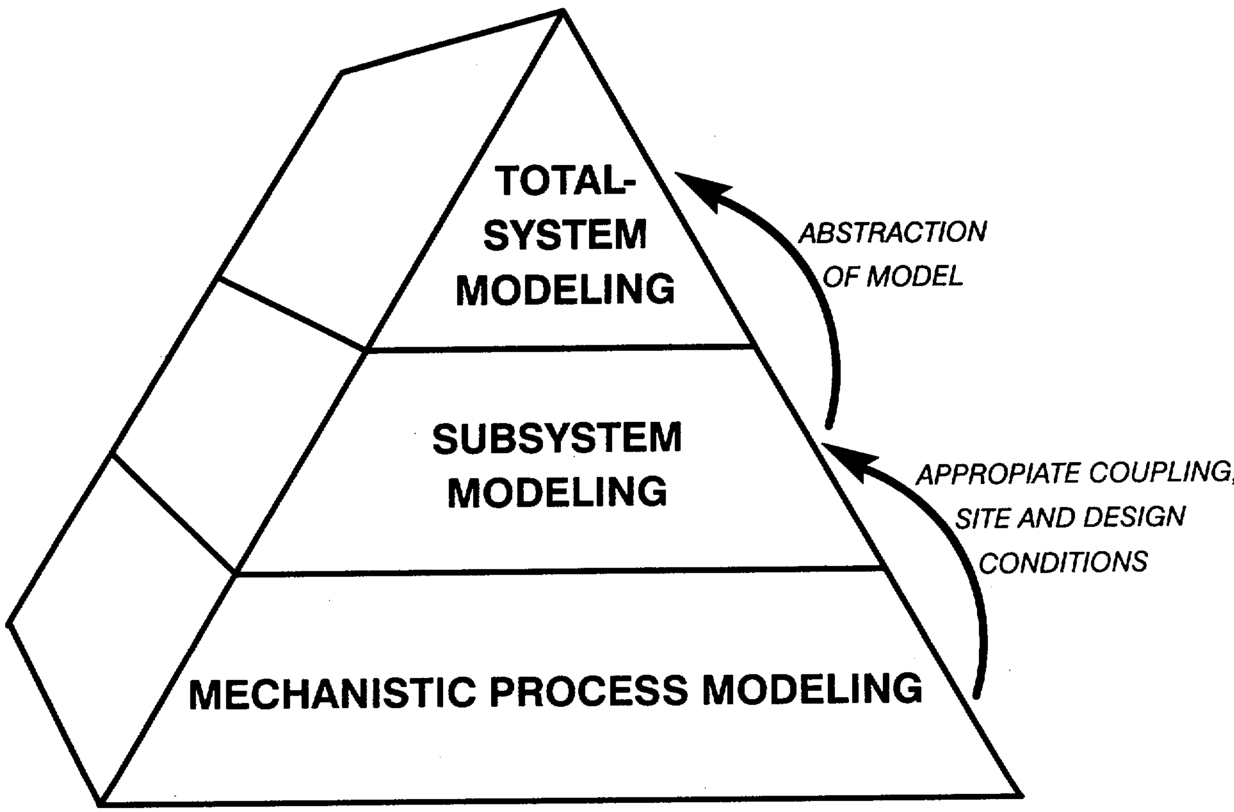


**Figure A-1. Hierarchy of performance assessment codes.**

determine whether assumptions for conservatism made in one model remain conservative when this model is coupled to another. Special model tests to check the hypothesis of conservatism may, therefore, be advisable. It may be noted that not all processes need to be reduced to their third level of simplicity for inclusion in the system model. Some of the processes may be so central to the final result that they have to be included in full detail.

When such a hierarchy-of-models approach is used, then all parts of the hierarchy need to be tested to build the required confidence, even though the type and amount of testing for each level in the hierarchy will be different. The measure of acceptability for models at different hierarchical levels will generally not be the same, and it is critical to identify these measures properly. Illustratively, estimation of the pressure distribution may be the performance measure for the first-level detailed hydrology models, whereas the total amount of flux may be the measure for the second-level simplified hydrologic models. The testing of the first-level models in this hierarchy, which can be assumed to be mechanistic in nature, comes close to scientific model validation, although such validation will be at temporal and spatial scales that are much shorter than those that characterize the repository system.

To provide confidence in the performance calculations, analyses may be conducted at two different levels. In the trial performance assessments conducted by both SKI and NRC, *auxiliary analyses* were performed to complement the overall or *total-system* performance calculations. Such auxiliary analyses provided important insights for proper interpretation of the total-system results, based on a complex, Monte Carlo-based analysis.

As noted earlier, the current U.S. standard applicable to disposal of HLW at sites other than Yucca Mountain is probabilistic in nature and requires that uncertainties be considered explicitly in any performance estimates. But, even when regulatory standards are deterministic (e.g., the Swedish standard is expected to be in terms of dose without an attached probability; similarly NRC's Part 60 subsystem performance requirements are deterministic), regulators require that an estimate of uncertainty (either as a range or more commonly as probability) in system performance be formed and presented. This means that even where conservatism is invoked, estimates of uncertainties introduced because of model structure and assignments of certain preferred values for parameters should be made. Sometimes, the models that are designed to estimate these uncertainties as an integral part of the overall model are called *probabilistic performance assessment models*. Testing of probabilistic models introduces additional requirements to obtain the necessary assurance that the uncertainty propagation in the models is such that the estimates of the uncertainties are either accurate or conservative.

In view of these considerations, both the NRC and SKI staffs have conducted TSPAs applicable to the current disposal concepts in their respective countries. In both NRC's iterative performance assessment program [see Codell *et al.* (1992), Wescott *et al.* (1995), and Mantuefel and Baca (1995)] and SKI's *Project-90* and *SITE-94* [see SKI (1991 and 1996), respectively], performance assessments, the repository, and the neighboring host rock are termed the *repository* or *process system*; natural or human-induced events or processes acting on the repository system, and originating outside of it, are considered *external events* or the *external environment*. The external event (environment) acting on the repository (process) system gives rise to *scenarios*.[11] The initial system description, at the time the repository is closed, is called the *undisturbed, base,* or *nominal case*. The disturbance of this system in the future by natural (e.g., seismic motion, tectonic events, magmatic eruptions, climatic change, etc.) or man-made/human-induced (e.g., exploratory drilling, mining, etc.) external causes may

[11] Not all HLW programs in the world define scenarios in exactly the same way [see OECD/Nuclear Energy Agency (1992); and Stenhouse *et al.* (1993)]. However, a strict definition of a scenario is not critical for this White Paper except to note that each scenario has a conceptual model associated with it.

change the system description and hence the underlying conceptual model.

Thus, to have the needed confidence that performance assessment models provide reasonable estimates of future performance of the repository, these models should be tested under as varied conditions as practical. The organization of performance assessment calculations is strongly dependent on the nature of the performance measures that are incorporated into regulations.

## A.5 References

Codell, R.B., *et al.*, "Initial Demonstration of the NRC's Capability to Conduct a Performance Assessment for a High-Level Waste Repository," U.S. Nuclear Regulatory Commission, NUREG-1327, May 1992.

Dverstorp, B., S. Norrby, and S. Wingefors, "Premises for Regulations Concerning Safety in Connection with the Final Disposal of Spent Nuclear Fuel, etc.," Stockholm, Sweden, Swedish Nuclear Power Inspectorate, SKI Memorandum 97017, March 26, 1997.

International Atomic Energy Agency, "Safety Principles and Technical Criteria for the Underground Disposal of High-Level Radioactive Wastes," Vienna, Austria, Safety Series No. 99 (IAEA Safety Standards), November 1985.

International Commission on Radiological Protection, "Radiological Protection Principles for the Disposal of Solid Radioactive Waste," Pergamon Press, Oxford, 1985. [ICRP Publication 46.]

Kotra, J. P., *et al.*, "Strategy for Development of NRC Regulations for the Proposed Repository at Yucca Mountain, Nevada," in American Nuclear Society, *High-Level Radioactive Waste Management: Proceedings of the Eighth International Conference, May 11-14, 1998, Las Vegas, Nevada,* pp. 361-363 [1998].

Mantuefel, R.D., and R.G. Baca, "Iterative Performance Assessment Phase 3: Status of Activities," San Antonio, Texas, Center for Nuclear Waste Regulatory Analyses, CNWRA 95-007, April 1995. [Prepared for the NRC.]

National Research Council, "Technical Bases for Yucca Mountain Standards," Washington, D.C., National Academy Press, Commission on Geosciences, Environment, and Resources, July 1995.

OECD/Nuclear Energy Agency, "Systematic Approaches to Scenario Development," Paris, France, NEA Working Group on the Identification and Selection of Scenarios for Performance Assessment of Radioactive Waste Repositories, Safety Assessment of Radioactive Waste Repositories Series, 1992.

Schumm, S.A., *To Interpret the Earth: Ten Ways to Be Wrong,* New York, Cambridge University Press, 1991.

Silling, S.A., "Final Technical Position on Documentation of Computer Codes for High-Level Waste Management," U.S. Nuclear Regulatory Commission, NUREG-0856, June 1983.

Smith, G.M., "Compilation and Analysis of Information on Critical Groups (Draft)," Oxfordshire, United Kingdom, Intera Information Technologies, August 1995. [Prepared for the Electric Power Research Institute.]

Stenhouse, M., N. Chapman, and T. Sumerling, "SITE-94 Scenario Development FEP Audit List Preparation: Methodology and Preparation," Stockholm, Sweden, Swedish Nuclear Power Inspectorate, SKI Technical Report 93:27, April 1993. [Prepared by Intera Information Technologies.]

Swedish Nuclear Power Inspectorate, "SKI Project-90," Stockholm, Sweden, SKI Technical Report 91:23, 2 vols. and "Summary," August 1991.

Swedish Nuclear Power Inspectorate, "SKI's Evaluation of SKB's RDD&D Programme 92: Review Report," Stockholm, Sweden, SKI Technical Report 93:30, July 1993.

Swedish Nuclear Power Inspectorate, "SKI SITE-94," Stockholm, Sweden, SKI Technical Report 96:36, 2 vols., December 1996.

Swedish Radiation Protection Institute, "The Swedish Radiation Protection Institute's Protection Criteria for the Disposal of Spent Nuclear Fuel," Stockholm, Sweden, Division of Waste Management and Environmental Protection, SSI-rapport 95-35, December 28, 1995.

Swedish Radiation Protection Institute, "The Swedish Radiation Protection Institute's Proposed Regulations Concerning the Final Management of Spent Nuclear Fuel or Nuclear Waste," Stockholm, Sweden, Division of Waste Management and Environmental Protection, SSI-rapport 97-07, May 1997.

The Radiation Protection and Nuclear Safety Authorities in Denmark, Finland, Iceland, Norway, and Sweden, *Disposal of High-Level Radioactive Waste—Consideration of Some Basic Criteria: A Consultative Document [Draft Report]*, Stockholm, Sweden, Swedish Radiation Protection Institute, 1989. [Available from SKI.]

The Radiation Protection and Nuclear Safety Authorities in Denmark, Finland, Iceland, Norway, and Sweden, *Disposal of High-Level Radioactive Waste—Consideration of Some Basic Criteria*, Stockholm, Sweden, Swedish Radiation Protection Institute, 1993. [The so-called *Nordic Document*. Available from SKI.]

U.S. Environmental Protection Agency, "Environmental Standards for the Management of Spent Nuclear Fuel, High-Level and Transuranic Wastes [Final Rule]," *Federal Register*, vol. 50, no. 182, September 19, 1985b, pp. 38066–38089. [Also see U.S. Environmental Protection Agency, "High-Level and Transuranic Radioactive Wastes: Background Information Document for Final Rule," Washington, D.C., EPA 520/1-85-023, August 1985.]

U.S. General Accounting Office, "EPA's Development of Radiation Protection Standards," Washington, D.C., GAO/RCED-93-126, June 3, 1993.

U.S. Nuclear Regulatory Commission, "10 CFR Part 60, Disposal of High-Level Radioactive Wastes in Geologic Repositories, Technical Criteria [Final Rule]," *Federal Register*, Vol. 48, No. 120, June 21, 1983, pp. 28194–28229.

U.S. Nuclear Regulatory Commission, "10 CFR Part 60, Disposal of High-Level Radioactive Wastes in Geologic Repositories Technical Criteria [Conforming Amendments]," *Federal Register*, Vol. 51, No. 118, June 19, 1986, pp. 22286-222301.

Wescott, R.G., *et al.* (eds.), "NRC Iterative Performance Assessment Phase 2: Development of Capabilities for Review of a Performance Assessment for a High-Level Waste Repository," U.S. Nuclear Regulatory Commission, NUREG-1464, October 1995.

# APPENDIX B
# REVIEW OF PERFORMANCE ASSESSMENT MODEL VALIDATION LITERATURE

This section contains brief reviews of the literature on the validation of scientific theories or models, and concepts of model validation as proposed by current practitioners. In the earth science area, for example, numerical models have been used extensively in the past for the assessment of fuel and non-fuel mineral resources [see Drew (1990); and Singer and Moiser (1981), respectively], and in predictive ground-water hydrology (National Research Council, 1990). More recently, though, model validation has come up in the context of safety assessments for geologic repositories. Thus, the volume of material that has been written on this subject is extensive and often contradictory, reflecting the disparate views of those scientists, engineers, and policy-makers involved in the respective assessments. Given these views, no attempt has been made to make the following review comprehensive. Instead, excerpts from works that span the philosophical spectrum are cited to demonstrate the variety of current views on model validation.

## B.1 Concepts of Model Validation

Public and private entities responsible for siting, constructing, licensing, or operating high-level radioactive waste (HLW) disposal facilities recognize that they must convince themselves and the public of the validity of the conceptual, mathematical, and numerical models used to assess the safety of these repositories. In the *Introduction* to the 1987 GEOVAL[1] symposium (Larsson *et al.*, 1988; p. 1), it is stated:

> "...In generic or site-specific evaluations of the safety of final repositories for spent nuclear fuel and high-level radioactive waste, the models used in performance analyses are of fundamental importance. It is essential that these models can gain a satisfactory level of confidence. Verification and validation efforts have an important role in this process...."

Validation and verification are necessary for instilling confidence in performance assessment models. Direct comparison of model predictions with experimental observations usually provides the most convincing evidence that a scientific theory or model is indeed correct. However, as noted by Kuhn (1970), acceptance of a scientific theory often depends more on achieving consensus among the scientific community than on a preponderance of objective evidence demonstrating quantitative agreement between theory and observation. Although many scientists involved in the design of geologic repositories have adopted the pragmatic, positivist approach to model validation suggested by Kuhn, others retain the more traditional and more restrictive negativist approach outlined by Popper (1959). The negativist viewpoint developed by Popper implies that a theory (or model) cannot be verified; theories can only be falsified. Moreover, it follows from Popper's philosophy that scientific theories of observable phenomena must be structured in such a manner as to be falsifiable or refutable, to be useful.

Konikow and Bredehoeft (1992) are among the staunchest proponents of the negativist approach to model validation. However, it appears that their assertion that ground-water models, in particular, cannot be validated, is based more on their strict definition of validation than on their adherence to Popper's view. Regarding the use of models in selecting a HLW repository, Konikow and Bredehoeft assert (*Op cit.*, p. 82) that "...[i]t is naive to believe that we will somehow validate a computer model so that it will make accurate predictions of system responses far into the future...." Konikow and Bredehoeft express the view that the emphasis being placed on

[1] One in a series of international symposia on the verification and validation of geosphere performance assessment models—see Swedish Nuclear Power Inspectorate—SKI (1988); SKI/Organization for Economic Co-operation and Development (OECD)/Nuclear Energy Agency (1991); and OECD/Nuclear Energy Agency/SKI (1995).

model validation may in fact lull the public into believing that all uncertainty associated with predicting performance of the repository can be eliminated. Although most practitioners of performance assessment modeling do not equate model validation with elimination of uncertainty, the possibility for such confusion suggests that prudent use of the term validation is warranted in this context.[2]

Davis *et al.* (1991) outline an approach for validating models, used in performance assessment, that is based on Popper's premise that a model cannot be declared to be valid simply because its predictions agree with experimental observation. The model will retain the status of being "not invalid" until experimental evidence is obtained that clearly rejects the validity of the model. Although they have adopted the negativist approach, Davis *et al.* contend that "...[s]howing that a model is not incorrect builds confidence in the model and acknowledges that perfection (i.e., validated performance assessment model) is not possible...." (*Op cit.*, p. 2). They stress that from the regulatory perspective "...[t]he goal of a model-validation exercise should not be viewed as providing a set of validated models...." (*Op cit.*, p. 3). Instead, the goal of the exercise is to "...[obtain] sufficient confidence that the models are able to simulate the behavior of the real system to accomplish the regulatory purpose." (*Op cit.*) It is interesting to note that although Konikow and Bredehoeft (1992), and Davis *et al.* (1991) are both proponents of the negativist approach, the former invoke Popper's argument to decry any attempt to validate models, whereas the latter appear to imply that repeated failures to invalidate a model may generate confidence that the model is "... an adequate representation of the real system...." (*Op cit.*)

In the GEOVAL *1987 Proceedings*, Niederer (1988, p. 16) first notes that "...the nature of a safety assessment is not that of a mathematical proof...." Indeed it may be argued that the nature of a safety assessment is not even that of a scientific proof. Niederer (1991, p. 32 ) later asserts that "...[i]t does not make sense to demand strict proof that a [performance assessment] model is correct...." However, Niederer does note that "...it makes a lot of sense...to promote consensus by providing ample evidence for the correctness of the [performance assessment] model...." (*Op cit.*) Overall, Niederer's views, in this regard, are generally similar to the Commission's perspective regarding the nature of the decision-making (e.g., the so-called *reasonable assurance determination*) necessary to support a construction authorization decision [see NRC (1983, pp. 28200-28201, 28204); and Schweitzer and Sastre (1987)].

Niederer (1991) and Neuman (1992) promote the philosophy of Kuhn (1970), and argue that "positive evidence" also contributes significantly to validation of models. In this context, "positive evidence" means that a model has met with repeated success in explaining pertinent observations and experimental data. Even if no competing models exist to be disproven, evidence for the validity of a model could be generated by demonstrating that the model adequately reproduces experimental observations. Neuman (1992, p. 1404) states that the process of model validation is the "...gradual building of confidence among scientists, and thereby among the public, that...understanding is being developed on the basis of a...research program...." Furthermore, Neuman asserts that "...[t]he best way to achieve...consensus [that confidence is warranted] is through a careful validation of all models that are used in isolation, and/or in tandem for safety assessment, regardless of how complex or simplified their components may be...." (*Op cit.*) For Neuman, Kuhn's assertion that consensus, and therefore validation, are most readily attained through comparison of model predictions to empirical evidence, demands that much greater emphasis must be placed on conducting those experiments needed to improve our understanding of large-scale, geosphere transport models.

[2] The publication of the Konikow and Bredhoeft paper was followed by a series of rebuttals—see de Marsily *et al.* (1992), Bredehoeft and Konikow (1992, 1993), McCombie and McKinley (1993), and Bair (1994).

A *Standard Practice* of the American Society for Testing and Materials (ASTM) distinguishes between "mechanistic" and "empirical" models (see ASTM, 1991). Mechanistic models are those based on a substantial understanding of the causal relationships between the dependent and independent variables of the model. In contrast, empirical models merely describe experimentally observed correlations between the variables without any significant understanding of the cause-and-effect relationships that might exist. The ASTM *Standard Practice* defines validation as meaning that "...the model can account for all the data available. It is preferred that models incorporate substantial mechanistic understanding of the [relevant] processes...." (*Op cit*., p. 732). The ASTM *Standard Practice* also notes that "...[t]he principal difficulty with empirical models is that the validity of extrapolations usually decreases rapidly the further one extends them beyond the original data. Thus, for the purposes of this practice, purely empirical models are considered unacceptable...." (*Op cit*., p. 731)

Tsang (1991, p. 829), citing Sargent (1984), lists eight types of information that could be used for model validation. These include: (i) event validity; (ii) face validity; (iii) traces; (iv) historical methods; (v) internal validity; (vi) historical data validation; (vii) predictive validation; and (viii) Turing tests. Some of these are more concerned with evaluating the internal structure and operations of a computer code than with model validation, as the term is used in this paper. However, Sargent's list introduces the notion of *peer review*, not only of model results, but also of the structure and development of the model itself. [Peer review is now frequently cited as a source of affirmation regarding the accuracy, validity, or relevance of technical information (see American Ceramic Society and the Conservation Foundation, 1985).]

Many authors recognize a distinction between the scientific goal of validation and developing confidence that a model is sufficiently valid for regulatory purposes. As Davis *et al.* (1991, p. 2) emphasized:

> ....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 the regulatory approach for validating models is critical. For one, the pure scientific approach, generally, would ask for a complete and detailed explanation for all observed phenomena and is not concerned with the specific application of science. 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 adequate for regulatory purposes but, by definition, not provide a detailed description of all phenomena....

Davis *et al.* (1991, p. 3) go on to argue that the adequacy of a model, for regulatory purposes, will be a subjective decision to be made by the regulator and suggest that the regulator base any decision on the following criteria:

- Whether the types of validation tests are relevant to the intended use of the model;
- How well the models are able to simulate the validation tests;
- How many validation tests are sufficient before the models can be applied to a particular site; and
- How well the site-specific information conforms with the model's description of the site.

On the other hand, some authors seem not to distinguish between scientific and regulatory validation. Niederer (1991, p. 31), citing Kuhn (1970), argues that:

> ....The [respective] roles of falsification and positive evidence are generally accepted, but neither is considered as sufficient to prove a theory. What, then, is the additional ingredient that establishes

> a theory as true? The most recent and, in our context, most interesting answer was given by Thomas Kuhn, and the answer is surprising and perhaps not very flattering for the make-believe strictness of the supposedly exact sciences. After analyzing many so-called scientific revolutions and taking into account psychological aspects —which cannot be neglected because science, after all, is made by human beings—Kuhn concludes that the proof of a scientific theory largely rests on consensus. In other words, and to put it simply and brutally, a scientific theory by definition is true if it has gained broad consensus among the experts of that particular science....

It should be noted that Kuhn's (1970) positive approach does not preclude incorrect popular opinion from being elevated to the level of scientific "truth."

Confidence in the ability of deep geologic repositories to safely contain HLW for thousands of years may be inferred by investigating the behavior of naturally occurring systems or analogs. Such consideration would be consistent with the requirements of NRC's 10 CFR 60.101(a)(2), where it is explicitly stated that "...[d]emonstration of compliance with such objectives and criteria will involve the use of data from...natural analog studies...." These analogs could include, for example, anomalous uranium (ores) deposited in environments that are geochemically similar to those expected to occur at the repository[3] or even man-made archeological sites.[4] Ewing (1993) outlines in detail the conceptual and philosophical issues that complicate the use of natural analogues and asserts that they will provide vital information for understanding the behavior of complex systems over long periods of time. Ewing ultimately concludes that natural analogues cannot be used to validate models; however, this conclusion appears to have less to do with any shortcomings in the value of analog data than with his apparent adherence to the negativist approach to model testing. Hoxie (1993, p. 104), on the other hand, states that "...natural analog systems are expected to be of indispensable use as part of the overall model-validation process, especially for those models invoked to predict the long-term performance of a repository system and its environment...." Sagar and Wittmeyer (1993, p. 27) are "...[s]omewhat pessimistic regarding the use of natural analog data for the specific purpose of [quantitative] model validation...." However, unlike Ewing (1993), their pessimism is based on the general lack of experimental control posed by natural analogues and not on a fundamental philosophical objection to their use in validation.

To address the requirement to study natural analogs, Alexander and van Luik (1990) note that DOE's (1988) *Site Characterization Plan* (SCP) explicitly discusses a number of natural and anthropogenic analog sites that may prove useful in assessing the long-term post-closure performance of the proposed geologic repository at Yucca Mountain, Nevada.[5]

---

[3] A number of reports and analyses have addressed the reduction of radioactivity and radiological hazards of HLW with time or made comparisons of its relative hazard to uranium ore. For example, see Hamstra (1975); Levi (1980); Wick and Cloninger (1980); Williams (1980); Cohen (1982); EPA (1982a, 1982b, 1985a); Elayi and Schapira (1987); Cohen *et al.* (1989); Mehta *et al.* (1991); and Tacca *et al.* (1991).

[4] Winograd (1986) suggests that the archeological record of man's past can provide an invaluable empirical data base to evaluate the reliability of long-term model predictions. He notes that some archeological records, dating as far back as the Late Paleolithic Age (ca. 40,000 to 10,000 B.C.), have been successfully preserved in the unsaturated zone of arid and semi-arid regions. Because the archeological record may be biased in favor of successful preservation at such sites, Winograd argues that it is important to synthesize the available physical, chemical, and biological information to better understand how the processes in place there might contribute to repository performance.

[5] Subsequent to the publication of the SCP, DOE was reported to have established a *Validation Oversight Group* (VOG), within the Office of Civilian Radioactive Waste Management, to implement a model validation methodology that would be used by the Department in preparing its licensing case for a geologic repository (see Voss, 1990). In proposing this draft methodology, VOG is reported to have adopted the following working definition of model validation (*Op cit*, p. 360):

> "...[T]hat the model is appropriate and adequate for the problem being addressed; is logically developed using the best available technology; is supported by experimental and observational data; the quality of the data is high; and the limitations of the model are understood...."

## B.2 References

Ahlers, C.F., *et al.*, "Calibration and Prediction of Pneumatic Response at Yucca Mountain, Nevada, using the LBNL/USGS Three-Dimensional, Site-Scale Model of the Unsaturated Zone," Berkeley, California, Lawrence Berkeley National Laboratory, Yucca Mountain Project Level 4 Milestone OB12M, 1996.

Ahlers, C.F., S. Finsterle, and G.S. Bodvarsson, "Characterization and Prediction of Subsurface Pneumatic Response at Yucca Mountain, Nevada," *Journal of Contaminant Hydrology*, 1999 [in press].

*footnote 5 continued*

Although the recommendations of VOG were never published, the influence of the VOG groundwork is reflected in DOE's current site characterization program for Yucca Mountain. For example, DOE has adopted internal requirements on model validation that adopt NRC's generic geologic disposal regulations' suggestion to support modeling activities with field tests and appropriate laboratory tests in addition to natural analogue studies [10 CFR 60.21(c)(1)(ii)(F)]. DOE's requirements for validation have been incorporated in the *Quality Assurance Requirements and Description* (QARD) document, which defines *model validation* "... as the process that demonstrates that the model is an acceptable representation of the process or system for which it is intended...." (DOE, 1998a; p. 3 of "Glossary") In turn, the *validation process* is defined as "... comparing analysis results against data acquired from laboratory, field experiments, natural analogue studies, or observations that were not used in the original development of the model...." (*Op cit.*, p. 3 of Supplement III) The QARD also specifies that "...validation is to be carried out to the extent practical to confirm that the mathematical representation appropriately depicts the natural phenomena...." (*Op cit.*) Supplement III of the QARD continues by indicating that if the type of data enumerated is not available, alternative approaches are to be documented and used, and if peer review is the alternative approach, it is to be carried out in accord with the guidance described in Section 2 of the QARD (DOE, 1998; p. 3 of Supplement III). These DOE requirements are self-explanatory, but do point to the need for proper and prompt documentation of confidence-building activities that subsequently lead to changes in the model(s) being used.

In December 1998, DOE published a 5-volume viability assessment, of the Yucca Mountain (Nevada) site, that presents progress and results from the scientific studies that have been conducted over the past 15 years. With its publication, the *Viability Assessment* effectively supersedes the *1988 Site Characterization Plan* (DOE, 1988) and represents the current and future framework of DOE's scientific programs at the site. (All 5 volumes of the *Viability Assessment*, including an *Overview*, are available on the following Internet location: *http:/www.ymp.gov/va.htm.*)

Within the *Viability Assessment*, there are several references to model validation. Under the heading "Increasing the Reliability of Performance Assessment Models," DOE's *Viability Assessment Overview* (DOE, 1998a; p. 31) states that:

> While forecasts of repository performance over thousands of years can never be proven, laboratory and field studies and experiments provide opportunities to validate the performance assessment models. By comparing the empirical results of the experiments with the predicted results of the models, analysts can assess how well their models represent the natural processes and engineered features of a repository. Validating the performance assessment models will
>
> reduce uncertainties and increase confidence that a repository will work as expected.

There is a discussion of model validation in Volume 3—"Total System Performance Assessment (TSPA)." In that volume, the following reference is made to validation (DOE, 1998b; p. 1–5):

> Although TSPAs can never be proven to be absolutely valid, many environmental problems require modeling of long-term interactions of man-made and geologic systems. Using the term "model" acknowledges that whether the descriptions of geologic features, events, and processes are unique and represent absolute reality will never be known. "Validation" of a long-term predictive model means that, on the basis of tests of the assumptions, inputs, outputs, and sensitivities, the model adequately reflects the recognized behavior of the portion of the system it is intended to represent. Adequacy is driven by the needs of the application for which the model is developed (Boak and Dockery 1998, p. 178–180).

The *Viability Assessment* discussion goes on to say that adequacy of models can be addressed through the judicious use of expert judgment, conservatism, and stochastic uncertainty studies, followed up with thorough documentation of the modeling and comprehensive external review.

One activity made possible, in a scientific program that has continuing exploratory and field-testing work, while models are evolving, such as those in DOE's Yucca Mountain program, is the "forward" prediction (e.g., forecasting, estimating) of physical conditions to be encountered and results to be obtained from on-going experiments (e.g., Bodvarsson *et al.*, 1994, p. 2039). As exploration and testing advance, the adequacy of a forward prediction can be evaluated, using new data. However, these new data are also intended to aid in the recalibration and improvement of the model, contemporaneous with its use. Thus, depending on the duration of the particular investigation, the window of opportunity for validation (confidence-building) could be a narrow one. Therefore, to be both credible and timely to the program, the results of these types of efforts must be documented in real time and made available to interested stakeholders, to establish that the forward prediction, itself, was indeed made well in advance of the work, and the comparison was objectively and correctly done before the adjustment of the model.

This approach of: (a) using forward model prediction to evaluate the model; and then (b) using the new data to improve the model, has been implemented in a number of instances for new boreholes and for the Exploratory Studies Facility, as well as for the Busted Butte unsaturated zone transport experiment [e.g., Wittwer *et al.* (1995); Ahlers *et al.* (1996, 1998); Wu *et al.* (1997); CRWMS M&O (1998a,b); Robinson (1994); Reimus *et al.* (1998); Bussod *et al.* (1998)]. This type of work suggests that validation efforts focus on the process-level modeling that is the first tier of models, based on the primary interpretation of field observation and measurement, laboratory and *in situ* data, and on accepted scientific interpretations of these observations and data. In compliance analyses, however, abstracted or simplified models may be used to represent complex processes in a system-level, multi-process, perhaps stochastic simulation. System-level models cannot be validated, but their credibility can be established by showing that single, or simple coupled processes, are properly accounted for in the modeling.

Alexander, D.H., and A. van Luik, "Natural Analogue Studies Useful in Validating Regulatory Compliance Analyses," in Swedish Nuclear Power Inspectorate/OECD Nuclear Energy Agency, *Symposium on Validation of Geosphere Flow and Transport Models Proceedings (GEOVAL '90), May 14–17, 1990, Stockholm, Sweden*, Organization for Economic Cooperation and Development, pp. 589–597 [1991].

American Ceramic Society and the Conservation Foundation, *Issues in Peer Review of the Scientific Basis for Regulatory Decisions*, Washington, D.C., American Ceramic Society, November 1985.

American Society for Testing and Materials, "Standard Practice for Prediction of the Long-Term Behavior of Waste Package Materials, Including Waste Forms Used in the Geologic Disposal of High-Level Nuclear Waste," in "1991 Annual Book of ASTM Standards—Nuclear, Solar, and Geothermal Energy," Philadelphia, PA, Section 12, Vol. 12.01 (Nuclear Energy), ASTM Standard C 1174-91, pp. 722-735, 1991.

Bair, E.S., "Model (In)Validation—A View from the Courtroom: Guest Editorial," *Ground Water*, 32[4]:530-531 [1994].

Boak, J.M. and H.A. Dockery, "Providing Valid Long-Term Projections of Geologic Systems for Policy Decisions: Can We Succeed? Should We Try?" in C.W. Welby, and M.E. Gowan (eds.), "A Paradox of Power Voices of Warning and Reason in the Geosciences," Geologic Society of America, *Reviews in Engineering in Geology*, Vol. XII, 1998.

Bodvarsson, G., G. Chen, and C. Wittwer, "Preliminary Analysis of Three-Dimensional Moisture Flow Within Yucca Mountain, Nevada," in American Nuclear Society/ American Society of Civil Engineers, *High-Level Radioactive Waste Management: Proceedings of the Fifth International Conference, May 22-24, 1994, Las Vegas, Nevada*, 4:2038-2047 [1994].

Bussod, G.Y., K. Coen, and R.C. Eckhardt (eds.), "Busted Butte Unsaturated Zone Transport Test," Los Alamos, New Mexico, Los Alamos National Laboratory, Milestone Deliverable No. SPU85M4, September 30, 1998.

Bredehoeft, J.D. and L.F. Konikow, "Reply to Comment [by de Marsily *et al.* (1992)]," *Advances in Water Resources*, 15:371-372 [1992].

Bredehoeft, J.D. and L.F. Konikow, "Ground-Water Models—Validate or Invalidate: Editorial," *Ground Water*, 32[2]:530-531 [1993].

Civilian Radioactive Waste Management System Management & Operating Contractor [CRWMS M&O], "ECRB [Enhanced Characterization of the Repository Block] Predictive Geotechnical Report," Las Vegas, Nevada, BABEA00001717-5705-00002, Rev. 01, May 1998.

Civilian Radioactive Waste Management System Management & Operating Contractor, "Predictions of Hydrologic, Geochemical, and Microbiological Properties and Conditions in the East-West Cross Drift (EWCD)," Las Vegas, Nevada, BABEA0000-01717-5705-0003, Rev. 00, June 1998.

Cohen, B.L., "Effects of ICRP Publication 30 and the 1980 BEIR Report on Hazard Assessments of High-Level Waste," *Health Physics*, 42:133-143 [1982].

Cohen, J.J., *et al.*, "An Assessment of Issues Related to Determination of Time Periods Required for Isolation of High Level Waste," in R.G. Post (ed.), *Waste Management '89: Proceedings of the Symposium on Waste Management, February 26–March 2, 1989, Tucson, Arizona*, 1:187-195 [1989].

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

de Marsily, G., P. Combes, and P. Goblet, "Comment on 'Ground-Water Models Cannot be Validated, by L.F. Konikow and J.D. Bredehoeft,' " *Advances in Water Resources,* 15:367-369 [1992].

Drew, L.J., *Oil and Gas Forecasting: Reflections of a Petroleum Geologist*, New York, Oxford University Press, International Association for Mathematical Geology/Studies in Mathematical Geology No. 2, 1990.

Elayi, A.G. and J.P. Schapira, "Impact of the Changes from ICRP-26 to ICRP-48 Recommendations on the Potential Radiotoxicity of the Discharged LWR, FBR and CANDU Fuels," *Radioactive Waste Management and the Nuclear Fuel Cycle*, 8(4):327-338 [1987].

Ewing, R.C., "Long-Term Predictions using Natural Analogues," in W.M. Murphy and L.A. Kovach (eds.), *The Role of Natural Analogues in Geologic Disposal of High-Level Nuclear Waste [Proceedings of the July 23–25, 1991, Workshop Held in San Antonio, Texas*], Center for Nuclear Waste Regulatory Analyses, CNWRA 93-020, September 1993, pp. 29–36.

Hamstra, J., "Radiotoxic Hazard Measure for Buried Solid Radioactive Waste," *Nuclear Safety*, 16(2):180-189 [March-April 1975].

Hoxie, D.T. "Natural Analogues for Far-Field Environment/Hydrology," in W.M. Murphy and L.A. Kovach (eds.), *The Role of Natural Analogues in Geologic Disposal of High-Level Nuclear Waste," [Proceedings of the July 23–25, 1991, Workshop Held in San Antonio, Texas]*, Center for Nuclear Waste Regulatory Analyses, CNWRA 93-020, September 1993, pp. 101-104.

Konikow, L.F., and J.D. Bredehoeft, "Ground-Water Models Cannot be Validated," *Advances in Water Resources,* 15:75–83 [1992].

Konikow, L.F., "Discussion of the 'The Modeling Process and Model Validation,' by Chin-Fu Tsang" *Ground Water*, 30(4):622-623 [1992].

Kuhn, T., *The Structure of Scientific Revolutions (2nd Edition)*, Chicago, The University of Chicago Press, 1970.

Larsson, A., K. Andersson, and C.–F. Tsang, "GEOVAL '87 [Introduction]: A Symposium on [the] Verification and Validation of Geosphere Performance Assessment Models" in Swedish Nuclear Power Inspectorate, *Symposium on Verification and Validation of Geosphere Performance Assessment Models Proceedings (GEOVAL '87), April 7–9, 1987, Stockholm, Sweden,* 1:1 [1988].

Levi, H.W., "The 'Project-Safety-Studies Entsorgung' in the Federal Republic of Germany," in IAEA/OECD Nuclear Energy Agency, *Symposium on the Underground Disposal of Radioactive Wastes Proceedings, July 2-6, 1979, Otaniemi, Finland,* International Atomic Energy Agency, IAEA-SM-243/17, 2:437-450 [1980].

McCombie, C., and I. McKinley, "Validation —Another Perspective: Guest Editorial," *Ground Water*, 31(4):530-531 [1993].

Mehta, K., G.R. Sherman, and S.G. King, "Potential Health Hazard of Nuclear Fuel Waste and Uranium Ore," Pinawa, Manitoba, Whiteshell Nuclear Research Establishment, Atomic Energy of Canada, Ltd., AECL-8407, June 1991.

National Research Council, *Ground Water Models—Scientific and Regulatory Applications*, Washington, D.C., National Academy Press, 1990.

Neuman, S.P., "Validation of Safety Assessment Models as a Process of Scientific and Public Confidence-Building," American Nuclear Society/American Society of Civil Engineers, *Proceedings of the Third International Conference: High-Level Radioactive Waste Management, April 12-16, 1992, Las Vegas, Nevada,* 2:1404–1413 [1992].

Niederer, U., "Perception of Safety in Waste Disposal: The Review of the Swiss Project Gewähr 1985," in Swedish Nuclear Power Inspectorate, *Symposium on Verification and Validation of Geosphere Performance Assessment Models Proceedings (GEOVAL '87),* *April 7–9, 1987, Stockholm, Sweden,* 1:11-26 [1988].

Neiderer, U., "In Search of Truth: The Regulatory Necessity of Validation," in Swedish Nuclear Power Inspectorate/OECD Nuclear Energy Agency, *Symposium on Validation of Geosphere Flow and Transport Models Proceedings (GEOVAL '90), May 14-17, 1990, Stockholm, Sweden*, Organization for Economic Cooperation and Development, pp. 29-35 [1991].

Organization for Economic Co-operation and Development/Nuclear Energy Agency/Swedish Nuclear Power Inspectorate, *GEOVAL '94—Validation through Model Testing Symposium Proceedings, October 11-14, 1994, Paris, France,* Organization for Economic Cooperation and Development, Safety Assessment of Radioactive Waste Repository Series, 1995.

Popper, K.R., *The Logic of Scientific Discovery*, London, England, Hutchinson Co., 1959.

Reimus, P.W., *et al.*, "C-Holes Update Report: Reinterpretation of the Reactive Tracer Test in the Bullfrog Tuff and Results of Laboratory Testing," Los Alamos, New Mexico, Los Alamos National Laboratory, Milestone Deliverable No. SP32E2M4SZ, September 15, 1998.

Robinson, B.A., "A Strategy for Validating a Conceptual Model for Radionuclide Migration in the Saturated Zone Beneath Yucca Mountain," *Radioactive Waste Management and Environmental Restoration*, 19:73-96 [1994].

Sagar, B., and G. Wittmeyer, "Role of Natural Analogues in Performance Assessment of Nuclear Waste Repositories, in W.M. Murphy and L.A. Kovach (eds.), "The Role of Natural Analogues in Geologic Disposal of High-Level Nuclear Waste," *[Proceedings of the July 23–25, 1991, Workshop held in San Antonio, Texas]*, Center for Nuclear Waste Regulatory Analyses, CNWRA 93-020, September 1993, pp. 21-28.

Sargent, R.G., "Simulation Model Validation," in T.I. Oren, B.P. Zeigler, and M.S. Elzas (eds.), *Simulation and Model-Based Methodologies: An Integrated View*, New York, Springer-Verlag, 1984,

Schweitzer, D.G., and C. Sastre, "Level of Evidence for Reasonable Assurance Guides to Prediction," Upton, New York, Brookhaven National Laboratory, BNL-52075, April 1987. [Prepared for DOE.]

Singer, D.A., and D.L. Moiser, "A Review of Regional Resource Assessment Methods," *Economic Geology*, 76[5]:1016-1031 [1981].

Swedish Nuclear Power Inspectorate, *Symposium on Verification and Validation of Geosphere Performance Assessment Models Proceedings (GEOVAL '87), April 7–9, 1987, Stockholm, Sweden*, 2 vols.[1988].

Swedish Nuclear Power Inspectorate/OECD/ Nuclear Energy Agency, *International Symposium on Validation of Geosphere Flow and Transport Models (GEOVAL '90), May 14-17, 1990, Stockholm, Sweden,* Organization for Economic Cooperation and Development [1991].

Tacca, J.A., G.G. Wicks, and W.L. Marter, "Relative Activity of DWPF Waste Glass Compared to Natural Uranium Ore," in American Ceramic Society, *93rd Annual Meeting and Exposition of the American Ceramic Society: Annual Meeting Abstracts, April 28–May 2, 1991*, Cincinnati, Ohio, p. 164 [1991].

Tsang, C.-F., "The Modeling Process and Model Validation," *Ground Water*, 29(6):825-831 [1991]. [Also see Konikow (1992), which includes Tsang's reply.]

U.S. Department of Energy, "Site Characterization Plan, Yucca Mountain Site, Nevada Research and Development Area, Nevada," Office of Civilian Radioactive Waste Management, Nevada Operations Office/Yucca Mountain Project Office, Nevada, DOE/RW-0199, 9 vols., December 1988.

U.S. Department of Energy, "Quality Assurance Requirements and Description for the Civilian Radioactive Waste Management

Program," Washington, D.C., Office of Civilian Radioactive Waste Management, DOE/RW/0333P (Rev. 8), June 1998.

U.S. Department of Energy, "Viability Assessment of a Repository at Yucca Mountain—Overview," Las Vegas, Nevada, Office of Civilian Radioactive Waste Management, Yucca Mountain Site Characterization Office, DOE/RW-0508, December 1998a.

U.S. Department of Energy, "Viability Assessment of a Repository at Yucca Mountain—Total System Performance Assessment (Volume 3)," Las Vegas, Nevada, Office of Civilian Radioactive Waste Management, Yucca Mountain Site Characterization Office, DOE/RW-0508/V3, December 1998b.

U.S. Environmental Protection Agency, "Population Risks from Disposal of High-Level Radioactive Wastes in Geologic Repositories," Washington, D.C., EPA 520/3-80-006, 1982a.

U.S. Environmental Protection Agency, "Draft Regulatory Impact Analysis for 40 CFR 191: Environmental Standards for Management and Disposal of Spent Nuclear Fuel, High-Level and Transuranic Radioactive Wastes," Washington, D.C., EPA 520/3-82-024, 1982b.

U.S. Environmental Protection Agency, "Environmental Standards for the Management of Spent Nuclear Fuel, High-Level and Transuranic Wastes [Final Rule]," *Federal Register*, vol. 50, no. 182, September 19, 1985, pp. 38066–38089. [Also see U.S. Environmental Protection Agency, "High-Level and Transuranic Radioactive Wastes: Background Information Document for Final Rule," Washington, D.C., EPA 520/1-85-023, August 1985.]

U.S. Nuclear Regulatory Commission, "10 CFR Part 60, Disposal of High-Level Radioactive Wastes in Geologic Repositories, Technical Criteria [Final Rule]," *Federal Register*, Vol. 48, No. 120, June 21, 1983, pp. 28194–28229.

Voss, C.F., "A Proposed Methodology for Validating Performance Assessment Models for the DOE Office of Civilian Radioactive Waste Management Program," in American Nuclear Society/American Society of Civil Engineers, *High-Level Radioactive Waste Management: Proceedings of the International Topical Meeting, April 8-12, 1990, Las Vegas, Nevada,* 1:359–363 [1990].

Wick, O.J. and M.O. Cloninger, "Comparison of Potential Radiological Consequences from a Spent-Fuel Repository and Natural Uranium Deposits," Battelle Pacific Northwest Laboratories, PNL-3540, September 1980.

Williams, W.A., "Population Risks from Uranium Ore Bodies," U.S. Environmental Protection Agency, Office of Radiation Programs, EPA 520/3-80-009, October 1980.

Winograd, I. J., "Archeology and Public Perception of a Transscientific Problem—Disposal of Toxic Wastes in the Unsaturated Zone," U.S. Geological Survey Circular 990, 1986.

Wittwer, C., *et al.*, "Preliminary Development of the LBL/USGS Three-Dimensional Site-Scale Model of Yucca Mountain, Nevada," Berkeley, California, Lawrence Berkeley National Laboratory, Report LBL-37356, UC-814, June 1995.

Wu, Y.S., *et al.*, "Integrated 3-D Site-Scale Flow Model (Chapter 19)" in G.S. Bodvarsson, T.M. Bandurraga, and Y.S. Wu (eds), "The Site-Scale Unsaturated Zone Model of Yucca Mountain, Nevada, for the Viability Assessment," Berkeley, California, Lawrence Berkeley National Laboratory LBNL-40376, June 1997. [Prepared in collaboration with the U.S. Geological Survey.]

# APPENDIX C
# GLOSSARY

The following definitions are intended to expand on some of the technical concepts introduced earlier in this White Paper. Because these definitions are provided as an aid to the reader's understanding of issues related to model validation, they should not be construed as a U.S. Nuclear Regulatory Commission (NRC)/Swedish Nuclear Power Inspectorate (SKI) staff position on these matters.

***Abstracted model:*** A conceptual model of a component, barrier, or subsystem that is purposely simplified to fit into a model of the overall geologic repository (system). Abstractions may take the form of reduction in dimensionality, elimination of time dependence, a table obtained from more complex models, a response surface derived from the use of more complex models, representation of a continuous process or entity with a few discrete elements, etc. Example: reduction of a three-dimensional (3-D) transport model to a one-dimensional (1-D) stream tube model.

***Auxiliary analysis:*** A quantitative evaluation performed in addition to, and in support of, the main analysis for estimating performance of the overall repository system. Examples would include the quantitative evaluation of conceptual models and their abstractions, analyses to show that results of the performance of the overall repository system model were bounded, and analyses to justify the assumptions and parameter values used.

***Computer code:*** An implementation of a mathematical model on a digital computer generally in a higher-order computer language such as *FORTRAN* or *C*. Computer codes often have names and version numbers for identification. Examples: NRC's total-system performance assessment computer code—TPA Version 3.2 (Mohanty and McCartin, 1999); or *MULTIFLOW,* Version 1.2, for coupled flow, reactive transport under non-isothermal conditions.

***Code verification:*** A process of assuring that the implementation of a mathematical model in the form of a computer code is free of coding errors, that the numerical schemes used are within the bounds of required accuracy, and that the equations were correctly solved. The process consists of following established quality assurance (QA) procedures during the development of the code, comparison of code with analytic solutions, and comparison with results from other codes. Some examples of code verification are found in the processes followed in INTRACOIN (the **Int**ernational **Tr**ansport **Co**de **In**tercomparison project—SKI, 1986) and HYDOCOIN (the **Hydr**ologic **Co**de **In**tercomparison project—The Coordinating Group of the HYDROCOIN Project, 1992) international code comparison exercises.

***Conservatism:*** In developing and applying mathematical models of physical systems, choices can be made regarding assumptions, approximations, data values, and data distributions. If these choices are made so that the resulting models and the estimates produced by them tend to make the estimated performance of a safety system worse than might actually be expected, the choices made are considered *conservative* or *pessimistic*. If the development and application of the model are such that the estimated performance tends to be better than might actually be expected, the choices made are considered *optimistic.*

***Conceptual model:*** A representation of the behavior of a real-world process, phenomenon, or object as an aggregation of scientific concepts, so as to enable predictions about its behavior. Such a model consists of concepts related to geometrical elements of the object (size and shape); dimensionality (1-, 2-, or 3-D); time dependence (steady-state or transient); applicable conservation principles (mass, momentum, energy); applicable constitutive relations; significant processes; boundary conditions; and initial conditions. Examples: representation of a natural geologic system by ***n*** number of 1-D independent vertical rectangular columns with transient flow, using mass conservation and Darcy's equation.

***Expert Judgment:*** Kotra *et al.* (1996, p. 3) suggest that expert judgment is information,[1] provided by a technical expert, in his or her subject matter area of expertise, based on opinion, or on a belief based on reasoning. Questions are usually posed to experts because they cannot be answered by other means. Expert judgments can be evaluations of theories, models, or experiments, or they can be recommendations for further research. Expert judgments may also be opinions that can be analyzed and interpreted, and used in subsequent technical assessments. Expert judgments can be either qualitative or quantitative. Expert judgments can also be judgments about uncertain quantities or judgments about value preferences. Frequently, subjective probabilities are used to quantify expert judgment. Expert judgment has also been called *expert opinion, subjective judgment, expert forecast, best estimate, educated guess*, and, most recently *expert knowledge* (see Meyer and Booker, 1990; p. 3). Regardless of how one defines it, expert judgment ultimately reflects the technical expert's evaluation and interpretation of some scientific knowledge base, to the extent that the knowledge base exists. Moreover, expert judgment does not create knowledge, rather it "...synthesizes disparate and often conflicting sources of information to produce an integrated picture..." (see Hora, 1993).

***Mathematical model:*** A representation of a conceptual model of a system, subsystem, or component through the use of mathematics. Mathematical models can be *mechanistic*, in which the causal relations are based on physical conservation principles and constitutive equations. In *empirical models*, causal relations are based entirely on observations. An example of a *mechanistic* mathematical model is Navier-Stokes' equations for fluid flow. An example of an *empirical* mathematical model is the van Genuchten equation relating permeability to the degree of saturation for an unsaturated porous medium.

***Peer Review:***[2] Much scientific and engineering development is subjected to the normal review process, consisting of critical evaluation by colleagues in various venues. These so-called *peer reviews* are typically documented, critical reviews that evaluate the acceptability and adequacy of some particular form of original research, performed by peers who are independent of the work being reviewed. A peer review can be conducted by obtaining input separately from a number of peers or by convening a panel to conduct the review. Also, discussions among the panel members can generate useful information not available from a set of independent reviews. The most common peer review process (i.e., pre-publication technical review of a scientific journal article) typically uses informal expert judgment to evaluate scientific methods and results. However, in principle, the nature of peer review is sufficiently flexible that its rigor and formality are commensurate with the study being reviewed. For example, the National Academy of Sciences is frequently called on to review reports or conclusions as a group of technical experts (see National Research Council, 1995a and 1995b). Peer reviews can also be conducted using a formal process to review the solution of very important problems.

***Performance assessment:*** A process of quantitatively evaluating the ability of a geologic repository system (at various levels) to contain and isolate high-level radioactive waste (HLW). This process is generally comprised of two main parts: (1) a quantitative estimate of system performance through the use of mathematical models, including auxiliary analyses and sensitivity and uncertainty analyses, followed by a comparison of the results with appropriate standards and criteria; and (2) an adequate documentation of assumptions, data, modeling approaches,

[1] Expert judgment is sometimes referred to as "data" (e.g., for purposes of aggregating the judgments of multiple experts).

[2] This discussion is an expansion of the earlier definition of *peer review* provided by Altman *et al.* (1988, p. 2).

and modeling results. Also see *safety assessment.*

***Performance measure (or safety indicator):*** A variable used to quantitatively evaluate the behavior of a geologic repository system, subsystem, or component to contain, isolate, or retard HLW. One or more performance measures may be defined. They may be in the form of engineering design specifications or requirements, or they can be described as the expected characteristics of the physico-chemical processes or phenomena active within the repository system itself. Some examples are peak dose at the location of some receptor, waste package lifetime, ground-water travel time, and amount of dilution within the saturated zone. Limits placed on acceptable values of performance measures form regulatory standards or criteria.

***Repository (or process) system:*** An aggregation of the engineered and natural components, contained within a postulated boundary, that function together as a cohesive unit to contain and isolate radioactive waste. Features, events and processes originating outside the repository may affect a system's functioning by altering fluxes of mass and energy across the system boundary. The concept has been previously defined by SKI (see Andersson, 1989; p. 44) as "... the complete set of 'deterministic' chemical and physical processes that might influence the release of radionuclides from the repository to the biosphere...."

***Safety assessment:*** Refers to the analysis used "...to predict the performance of the overall geologic repository system...." (IAEA, 1993; p. 11) or "... an analysis to predict the performance of a system or subsystem, followed by a comparison of the results of such [an] analysis with appropriate [safety] standards and criteria.... (Organization of Economic Cooperation and Development/ Nuclear Energy Agency, 1991, p. 14). In general, a safety assessment is a type of performance assessment insofar as the system under consideration is the **overall** repository system and the performance measure of interest is radiological impact or some other global measure of safety. Also see *performance assessment*.

***Sensitivity analysis:*** Because of the complexity of the systems comprising a geologic repository, it is not usually possible to develop exact analytical expressions for the relationship between repository performance (measures) and the input parameters used to formulate mathematical models. To gain this understanding, quantitative (statistical) evaluations are used to describe the change in a performance measure corresponding to a change in the value or probability distribution of a model parameter. Sensitivity analyses are used to rank parameters according to the sensitivity of the performance measure to the parameters.[3] An example is the peak dose to a member of the critical group changing by $x$ percent when the infiltration rate changes by $y$ percent. Also see *uncertainty analysis.*

***Uncertainty:*** Alternative definitions exist for classifying the different types of uncertainty. Generally, there are two types of uncertainty present in any calculation. These are: (1) *stochastic* (or *aleatory*) *uncertainty* caused by the random variability in a process or phenomenon; and (2) *state-of knowledge* (or *epistemic*) uncertainty, which results from a lack of complete information about physical phenomena. *State-of knowledge uncertainty* may be further divided into (i) parameter uncertainty, which results from imperfect knowledge about the inputs to analytical models; (ii) model uncertainty, which is caused by imperfect models of physical systems, resulting from simplifying assumptions or an incomplete identification of the system modeled; or (iii) completeness uncertainty, which refers to the uncertainty as to whether all the significant physical phenomena, relationships (coupling), and events have been considered. Also, see Senior

[3] Types of sensitivity analyses include the Monte Carlo method (Helton, 1970); fractional factorial design (Cochran, 1963); differential analysis (Baybutt *et al.*, 1981); response surface methodology; Fourier amplitude sensitivity (Helton *et al.*, 1991); and the Limit-State Approach (Wu *et al.*, 1992), to name a few.

Seismic Hazard Analysis Committee (1997, pp. 13-14).

*Uncertainty analysis:* A quantitative (statistical) evaluation performed to estimate the uncertainty and error bounds in a performance measure that may be caused by *epistemic uncertainty* (described above). The objective of an uncertainty analysis is to assess the degree of variability in calculated results as a function of the variability in model and input parameters. Example: The uncertainty in the estimates of waste package life time decreases by *x* percent for every *y* percent decrease in the uncertainty in relative humidity around waste packages.

*Validation:* As noted in Section 3.1, there is no internationally agreed-to definition of model validation. The most recent International Atomic Energy Agency (IAEA) definition (1993, p. 48) of validation is:

> "A process carried out by comparison of **model** predictions with field observations and experimental measurements. A **model** is considered validated when sufficient testing has been performed to ensure an acceptable level of predictive accuracy over the range of conditions over which the **model** may be applied. (Note that the acceptable level of accuracy is judgmental and will vary depending on the specific problem or question to be addressed by the **model**.)"

In the regulatory process, it should be noted that a model: (1) may need more or less validation depending on its importance to compliance demonstrations; and (2) is said to be sufficiently validated when it can be used for its intended purpose with some degree of confidence. An example is a flow model used to estimate inflow into geologic repository emplacement drifts is sufficiently validated when it is determined that the calculated inflow for plausible scenarios is within the range of data uncertainties—the validation process may employ theoretical arguments, peer review, laboratory data, field data, and data from natural analogs.

*Verification:* A process of assuring that the implementation of a mathematical model, in the form of a computer code, is free of coding errors, and that the numerical schemes used are within the bounds of required accuracy. The process consists of following established QA procedures during the development of the code, comparison of the code with analytic solutions, and comparison with results from other codes. Some examples are the processes followed in the INTRACOIN and HYDROCOIN international code comparison exercises. Also, see IAEA (1993, pp. 48-49).

**References**

Altman, W.D., *et al.*, "Peer Review for High-Level Waste Repositories," U.S. Nuclear Regulatory Commission, Generic Technical Position, NUREG–1297, February 1988.

Andersson, J. (ed.), "The Joint SKI/SKB Scenario Development Project," Stockholm, Sweden, Swedish Nuclear Power Inspectorate, SKI Technical Report, 89:14, December 1989.

Baybutt, P., D.C. Cox, and R.E. Kurth, "Methodology for Uncertainty Analysis of Light Water Reactor Meltdown Accident Consequences," Battelle Columbus Laboratories, Letter Report, May 31, 1981. [Prepared for the NRC.]

Cochran, W.G., *Sampling Techniques*, New York, John Wiley and Sons, Inc., 1963.

Helton, J.C., "A Retrospective and Prospective Survey of the Monte Carlo Method," *SIAM Review*, 12[1]:1-60 [1970].

Helton, J.C., *et al.*, "Sensitivity Analysis Techniques and Results for Performance Assessment at the Waste Isolation Pilot Plant," Albuquerque, New Mexico, Sandia National Laboratories, SAND90–7103, March 1991.

Hora, S.C., "Acquisition of Expert Judgment: Examples from Risk Assessment," *Journal of Energy Engineering*, 118:136-148 [1993].

International Atomic Energy Agency, *Radioactive Waste Management Glossary (Third Edition)*, Vienna, Austria, 1993.

Kotra, J.P., *et al.*, "Branch Technical Position on the Use of Expert Elicitation in the High-Level Radioactive Waste Program," U.S. Nuclear Regulatory Commission, NUREG–1563, November 1996.

Meyer, M.A. and J.M. Booker, "Eliciting and Analyzing Expert Judgment: A Practical Guide," U.S. Nuclear Regulatory Commission, NUREG/CR–5424, January 1990. [Prepared by the Los Alamos National Laboratory.]

Mohanty, S., and T.J. McCartin (coordinators), "Sensitivity and Uncertainty Analyses for a Proposed HLW Repository at Yucca Mountain, Nevada, Using TPA Computer Code Version 3.1," U.S. Nuclear Regulatory Commission, NUREG–1668, 2 vols., 1999 [in press].

National Research Council, "Ward Valley: An Examination of Seven Issues in Earth Sciences and Ecology," Washington, D.C., National Academy Press, Commission on Geosciences, Environment, and Resources, [May] 1995a.

National Research Council, "Review of U.S. Department of Energy Technical Basis Report for Surface Characteristics, Preclosure Hydrology, and Erosion," Washington, D.C., National Academy Press, Commission on Geosciences, Environment, and Resources [December] 1995b.

Organization for Economic Cooperation and Development/Nuclear Energy Agency, "Review of Safety Assessment Methods," Paris, France, Performance Assessment Advisory Group of the Radioactive Waste Management Committee, Disposal of Radioactive Waste Series, 1991.

Senior Seismic Hazard Analysis Committee, "Recommendations for Probabilistic Seismic Hazard Analysis: Guidance on Uncertainty and Use of Experts," U.S. Nuclear Regulatory Commission, NUREG–1372, 2 vols., April 1997. [Prepared by the Lawrence Livermore National Laboratory for the U.S. Department of Energy, NRC, and the Electric Power Research Institute.]

Swedish Nuclear Power Inspectorate, "INTRACOIN (International Nuclide Transport Code Intercomparison Study)—Final Report Levels 2 and 3 (Model Validation and Uncertainty Analysis)," Stockholm, Sweden, SKI 86:2, May 1986.

The Coordinating Group of the HYDRO-COIN Project, "The International HYDROCOIN Project—Summary Report," Paris, France, HYDROCOIN Project Secretariat/SKI, Paris, France, Organization for Economic Cooperation and Development, 1992.

Wu, Y.T., A.B. Gureghian, and R.B. Codell, "Sensitivity and Uncertainty Analysis Applied to One-Dimensional Transport in a Layered Fractured Rock," San Antonio, Center for Nuclear Waste Regulatory Analyses, CNWRA 92-002, March 1992.

NRC FORM 335
(2-89)
NRCM 1102,
3201, 3202

U.S. NUCLEAR REGULATORY COMMISSION

**BIBLIOGRAPHIC DATA SHEET**

(See instructions on the reverse)

1. REPORT NUMBER (Assigned by NRC, Add Vol., Supp., Rev., and Addendum Numbers, if any.)

NUREG–1636

2. TITLE AND SUBTITLE

Regulatory Perspectives on Model Validation in High-Level Radioactive Waste Management Programs: A Joint NRC/SKI White Paper

3. DATE REPORT PUBLISHED

| MONTH | YEAR |
|---|---|
| March | 1999 |

4. FIN OR GRANT NUMBER


5. AUTHOR(S)

N.A. Eisenberg[a], M.P. Lee[a], M.V. Federline[a], S. Wingefors[b], J. Andersson[b], S. Norrby[b], B. Sagar[c], and G.W. Wittmeyer[c]


6. TYPE OF REPORT

Technical

7. PERIOD COVERED (Inclusive Dates)


8. PERFORMING ORGANIZATION – NAME AND ADDRESS (If NRC, provide Division, Office or Region, U.S. Nuclear Regulatory Commission, and mailing address; if contractor, provide name and mailing address.)

[a]Division of Waste Management, Office of Nuclear Material Safety and Safeguards, U.S. Nuclear Regulatory Commission, Washington, D.C., 20555–0001
[b]Office of Nuclear Waste Safety, Swedish Nuclear Power Inspectorate (Statens Kärnkraftinspektion), Stockholm, Sweden
[c]Center for Nuclear Waste Regulatory Analyses, San Antonio, Texas, USA



9. SPONSORING ORGANIZATION – NAME AND ADDRESS (If NRC, type "Same as above"; if contractor, provide NRC Division, Office or Region, U.S. Nuclear Regulatory Commission, and mailing address.)

Division of Waste Management, Office of Nuclear Material Safety and Safeguards, U.S. Nuclear Regulatory Commission, Washington, D.C., 20555–0001
Office of Nuclear Waste Safety, Swedish Nuclear Power Inspectorate


10. SUPPLEMENTARY NOTES


11. ABSTRACT (200 words or less)

Validation should be an important aspect of the regulatory uses of mathematical models in the safety assessments of geologic repositories for the disposal of spent nuclear fuel and other high-level radioactive wastes. Because models for a geologic repository performance assessment cannot be tested over the spatial scales of interest and long time periods for which the models will make estimates of performance, the usual avenue for model validation—e.g., comparison of model estimates with actual data at the space-time scales of interest—is precluded. Further complicating the model validation process in HLW programs are the uncertainties inherent in describing the geologic complexities of potential disposal sites, and their interactions with the engineered system, with a limited set of generally imprecise data, making it difficult to discriminate between model discrepancy and inadequacy of input data. A successful strategy for model validation should attempt to recognize these difficulties, address their resolution, and document the resolution in a careful manner. The end result of validation efforts should be a documented enhancement of confidence in the model to an extent that the model's results can aid in regulatory decision-making. The level of validation needed should be determined by the intended uses of these models.



12. KEY WORDS/DESCRIPTORS (List words or phrases that will assist researchers in locating the report.)

conceptual models
confidence building
decision-making
geologic repository
high-level radioactive waste
mathematical models
model validation
performance assessment
probabilistic risk assessment
safety assessment


C-7

13. AVAILABILITY STATEMENT

Unlimited

14. SECURITY CLASSIFICATION

(This Page)
Unclassified
(This Report)
Unclassified

15. NUMBER OF PAGES

16. PRICE

NRC FORM 335 (2–89)

Printed on recycled paper

Federal Recycling Program

C-10